Developments in Mobile Multimedia Technologies

Lead Guest Editor: Yuanlong Cao
Guest Editors: Jiyan Wu, Cunhua Pan, and Wei Quan

Developments in Mobile Multimedia Technologies

Developments in Mobile Multimedia Technologies

Lead Guest Editor: Yuanlong Cao
Guest Editors: Jiyan Wu, Cunhua Pan, and Wei Quan

Contents
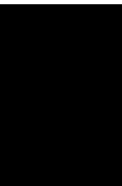
WILEY | Hindawi

Research Article

Performance Analysis of WOFDM-WiMAX Integrating Diverse Wavelets for 5G Applications

Lavish Kansal [iD],[1] **Gurjot Singh Gaba** [iD],[2] **Ashutosh Sharma** [iD],[3] **Gaurav Dhiman** [iD],[4] **Mohammed Baz** [iD],[5] **and Mehedi Masud** [iD][6]

[1]*School of Electronics & Electrical Engineering, Lovely Professional University, Phagwara 144411, India*
[2]*School of Computer Science, Mohammed VI Polytechnic University, Ben Guerir 43150, Morocco*
[3]*Southern Federal University, Rostov-on-Don, Russia*
[4]*Department of Computer Science, Government Bikram College of Commerce, India*
[5]*Department of Computer Engineering, College of Computer and Information Technology, Taif University, PO Box. 11099, Taif 21994, Saudi Arabia*
[6]*Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Mehedi Masud; mmasud@tu.edu.sa

In the 5th generation (5G) and 6th generation (6G) of wireless mobile telecommunication networks, the requests for an elevated data rate with access to stationary as well as portable customers are going to be overwhelming. Mobile worldwide interoperability for microwave access (WiMAX) comes out as a favourable alternative that is intelligibly developed and more matured than wireless fidelity (Wi-Fi). Mobile WiMAX makes use of the orthogonal frequency division multiple access (OFDMA) technology for its two-way communication to enhance the system performance in fading environments making it more suitable for 5G applications. The diverse OFDM forms deliberated here are the fast Fourier transform- (FFT-) based WiMAX and discrete wavelet transform- (DWT-) based WiMAX. The suggested study exhibits the bit error rate (BER) and peak to average power ratio (PAPR) reduction by integrating different wavelet families, i.e., Haar, symlet, coiflet, and reverse biorthogonal over Rayleigh fading channel. The simulation results obtained by MATLAB depicts an improvement in PAPR reduction, and signal to noise ratio (SNR) requirement is also reduced by 6-12 dB by using DWT-incorporated WiMAX at a BER of 10^{-4}.

1. Introduction

WiMAX organizations specifically designed WiMAX system to have unanimity and consistency of the IEEE 802.16 guidelines, as of now called a wireless metropolitan area network (MAN). IEEE 802.16c operating in the 10-66 GHz frequency range for the line of sight (LOS) propagation was introduced by IEEE in 2002. The IEEE 802.16c includes the specifications for connectivity between remote areas and data distribution over a wide area network for both point to point (PPP) and point to multipoint (PMP) communication over microwave frequencies. Further, IEEE introduced a new version, i.e., IEEE 802.16a operating over the frequency range of 2-11 GHz and suitable for nonline of sight (NLOS) signal propagation. IEEE 802.16d, an improved form of IEEE 802.16, was proposed in 2004 for providing wideband connectivity to indoor clients. The IEEE affirmed the 802.16 benchmarks in June 2004, and three working groups were framed to access and rate the guidelines. WiMAX can be termed completely as an inheritor of the Wi-Fi system. The standard when extended to IEEE 802.16e-2005 popularly called as mobile WiMAX standard provides various solutions for wireless broadband data that provides ease for meeting requirements of the mobile and fixed broadband networks

[1, 2]. WiMAX provides a way through which anyone can be in touch with another person regardless of their location anytime and anywhere. 5G has a crucial role in operation of society along with IoT. It needs to think about new trust model [3], increased privacy concerns [4, 5], and IoT-based health care [6].

The modulation level and the minimum BER requirements and the SNR are major factors that impact the spectral effectiveness of the WiMAX scheme. An appropriate selection of the modulation techniques/level is the trade-off between BER and spectral efficiency, i.e., as we keep on increasing the level of modulation from QPSK to 16-PSK and so on the BER performance keeps on decreasing, the increase in the modulation level results in an enhancement in the spectral effectiveness of the module. In addition to this elevated PAPR is a key shortcoming of the modern-day multicarrier transmission systems, i.e., 4G and 5G. As we know, to transmit the signal over a wide coverage area, the power amplifier is being employed at the transmitter side, and to achieve maximum coverage area and adequate signal strength, the OFDM system forced the amplifier to operate in the nonlinear region. The operation of power amplifiers in the linear region will result in high PAPR. High PAPR leads to certain distortions, i.e., out-band and in-band radiation. The outcome of the OFDM/WiMAX system can be improved by using diverse transforms like discrete cosine Stockwell transform (DCST), fractional Fourier transform (FRFT), wavelet Haar transform (WHT), and DWT instead of conventional fast Fourier transform (FFT). DWT offers the frequency and time domain depiction of signals, whereas in contrast, DFT gives the depiction of the signal in the frequency domain [7]. The wavelet characteristics, which comprise demonstration in time and frequency domain, orthogonality by means of a scale, and translation, highlight an entire new viewpoint in wireless mobile communication. The DWT is being used as a signal processing technology in several modern-day wireless communication applications, assimilating MCM and mobile wireless communication. Many researchers have proposed a replacement of FFT by DWT in modern-day wireless telecommunication systems (4G, 5G) due to its inferior time-frequency localization, poor bandwidth efficiency due to utilization of cyclic prefix, poor BER performance, etc. In the wavelet-based WiMAX, the cyclic prefix will not be utilized which will result in efficient utilization of the bandwidth. Technically, the DWT is a highly advanced transform in comparison to the DFT.

1.1. Our Contribution. On carefully studying the investigation work described in the literature survey, the following research gaps are addressed to realize the goal of efficient transmission of information over 5G network:

(1) Diverse wavelets are being utilized to analyse the performance outcome of the WiMAX system

(2) SNR vs. BER and PAPR reductions are the parameters using which the performance of WiMAX system is studied to establish wavelets as a reliable alternative to FFT

(3) The simulation outcome also presents the variation of received signal quality for diverse modulation type/levels

The entire manuscript is structured as follows: Section 1 explains the reasons for using wavelet transforms in the existing wireless communication systems by citing the issues related to WiMAX. A brief insight on the existing work based on FFT-OFDM, WHT-OFDM, FFT-WiMAX, and WHT-WiMAX is provided in Section 2. The basic model description is provided in Section 3, followed by an insight into the simulation parameters and channel models in Section 3. Section 4 presents the interpretations extracted from the MATLAB simulation results, and finally, conclusion remarks for the proposed work are presented in Section 5.

2. Related Work

In the preliminary stages, the DWT-based OFDM was simulated to lessen the influence of ICI/ISI and to elevate the bandwidth efficiency. The DWT-OFDM proved to be more resilient to the ICI and ISI because the wavelet filters enable spectral containment capabilities which in result decrease the ICI and ISI [8]. Moreover, in DWT-OFDM, the guard bands were not required, thereby making DWT-OFDM more bandwidth effective in comparison to FFT-OFDM. DFT-based OFDM was the most popular methodology for the efficient implementation of the OFDM system. But those systems suffer from certain drawbacks such as the requirement of a cyclic prefix, low QOS, ISI, and ICI. A new model for the OFDM system which incorporates the complex wavelet packet transform (WPT), i.e., CWP-OFDM, was proposed [9]. The time-varying Doppler shift affects the orthogonality of the OFDM subcarriers drastically [10]. The DWT-OFDM was proved to be more resilient to the shift and also offers side lobes of far lower magnitude in comparison to FFT-OFDM. The accurate reconstruction of the symbols with much lower complexity was a squeezing characteristic of WHT-OFDM [11]. The WHT-OFDM offered a considerable improvement regarding transmission efficiency and spectrum leakage in comparison to FFT-OFDM over power line channels [12]. A wavelet-based OFDM system that satisfies the impeccable reconstruction property by virtue of their properties of orthonormal bases was proposed. Due to the orthogonal basis and accurate reconstruction of the symbols, the OFDM system incorporating the wavelet transforms, i.e., DWT-OFDM and WPM-OFDM exhibits a significant improvement in BER, was in contrast to Fourier-based OFDM [13, 14]. In addition to the wavelets, the BER assessment of the OFDM system was also be improved by using other transforms like discrete cosine transforms (DCT) and discrete sine transforms (DST) in comparison to FFT. Multiwavelet design also offers a substantial enrichment in the BER performance thereby reducing the minimum SNR required to attain the anticipated BER performance [15–17]. Because of their time-frequency localization property, the utilization of wavelets was having a robust influence on the BER outcome of the OFDM system over both the flat fading and fading channels which are frequency selective in nature [18]. It was presented that the

TABLE 1: Comparative analysis of work presented in literature and the proposed methodology.

Article	Research outcome
BER assessment of FFT-OFDM against WHT-OFDM over different fading channels [30]	In this research article, Haar transform is utilized to showcase the BER improvement in OFDM systems. However, the effects of other wavelet transform families such as symlet, coiflet, and reverse biorthogonal were not studied in this paper.
Comparative analysis of wavelet and OFDM-based systems [31]	In this paper, the appropriateness of constant envelope multicarrier modulation methodology, OFDM, and wavelet-based systems is deliberate. The prime idea of this paper is to do the assessment of wavelet-based and OFDSM-based systems by using BER vs. SNR performance metrics. In our present research, we have proposed an adaptive system which will select any of the optimum wavelets that will give us best PAPR and BER reduction.
On BER assessment of conventional- and wavelet-OFDM over AWGN channel [19]	Performance of OFDM system based on DFT and DWT transform is evaluated in this paper using the BER matric. Diverse modulation schemes (PSK and QAM) and levels (2, 4, 8, and 16) are being utilized to analyse the performance over AWGN channel for both DFT- and DWT-based OFDM system. However, in the present work, Rayleigh fading channel is explored for the assessment of the performance of conventional and proposed system model.

BER outcome was superior in flat fading channel but it degrades drastically in the frequency selective fading channels which are time-varying in nature. The suggested methodology performs exceptionally well in the frequency selective and time-varying channel condition due to the excellent time-frequency localization capabilities of the wavelet-based OFDM systems. ICI and ISI generated by vanishing effect of orthogonality in the OFDM system were significantly condensed by employing wavelet transform instead of DFT in the conventional OFDM-WiMAX system. In addition to the reduction of ICI and ISI, the wavelet-incorporated OFDM also enhances the spectral efficiency contrary to the DFT-incorporated OFDM system [18]. The WHT-OFDM outperforms the FFT-OFDM over the AWGN channels also, mainly due to the robust orthogonality of the subcarriers exhibited by WHT-OFDM in contrast to FFT-OFDM [19]. The impact of exponential power delay profile in conjunction with Rayleigh fading channel was also less on the WHT-OFDM contrary to FFT-OFDM [20]. Wavelet packet modulation (WPM) was also incorporated in the WiMAX, and the performance was evaluated over the AWGN channel, which again shows BER improvement on the use of wavelets in place of Fourier transforms [21]. DVB-T based on DWT-OFDM was also investigated in the literature, and the BER outcome of the wavelet family was superior to that of Fourier transforms. Also, out of all the members of the wavelet family, the Haar transform performed the best-regarding BER improvement for a given value of SNR [22]. It has been observed that precoding techniques [23–29], as well as the use of transform other than FFT [27–49], improve the BER and PAPR reduction performance for OFDM/WiMAX/MIMO/OFDM-MIMO system.

2.1. Research Gaps

(1) The work presented in the literature does not describe the effect of diverse wavelet families on WiMAX system's performance

(2) In most of the existing schemes presented in Table 1, only the Haar wavelet has been utilized to assess the BER and PAPR outcome of the OFDM/WiMAX system. Symlets, coiflets, and reverse biorthogonal wavelet families were not utilized for the analysis purpose in most of the cases

(3) While doing the performance assessment of the receiver section, the signal quality at the input of the demodulator was not presented in most of the recent work presented on OFDM/WiMAX. It is significant to study the behaviour of the signal at the receiver side as it will be useful while modelling the receiver section to give an optimum outcome

3. Model Description, Channel Model, and Signal Detection

The physical layer of the proposed model of WiMAX is presented in Figure 1, and the system parameters are considered as same as that of [32]. The multipath channel that is used for assessing the behaviour of the FFT-incorporated WiMAX and DWT-incorporated WiMAX is Rayleigh fading channel. The nth element of complex OFDM symbol in discrete-time domain is written as

$$X_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j((2\pi kn)/N)}, \ 0 \le n \le N-1, 0 \le k \le N-1. \quad (1)$$

Here, Equation (1) represents the discrete-time complex OFDM symbol. N represents the number of samples, X_k represents the input signal, and X_n represents the OFDM symbols obtained after the Fourier transform. Diverse form of wavelets are comprised in the wavelet family, such as Haar, coiflet, symlet, biorthogonal, and reverse biorthogonal wavelets. However, in the proposed methodology, Haar wavelet is being utilized. The operating principle of the DWT and IDWT is presented in Figure 2. Here, $h(n)$ is the half band impulse response of high-pass filters, and $g(n)$ is the half-band impulse responses of the low-pass filters. The two filters are associated to each other by a quadrature mirror filter (QMF) relationship as follows:
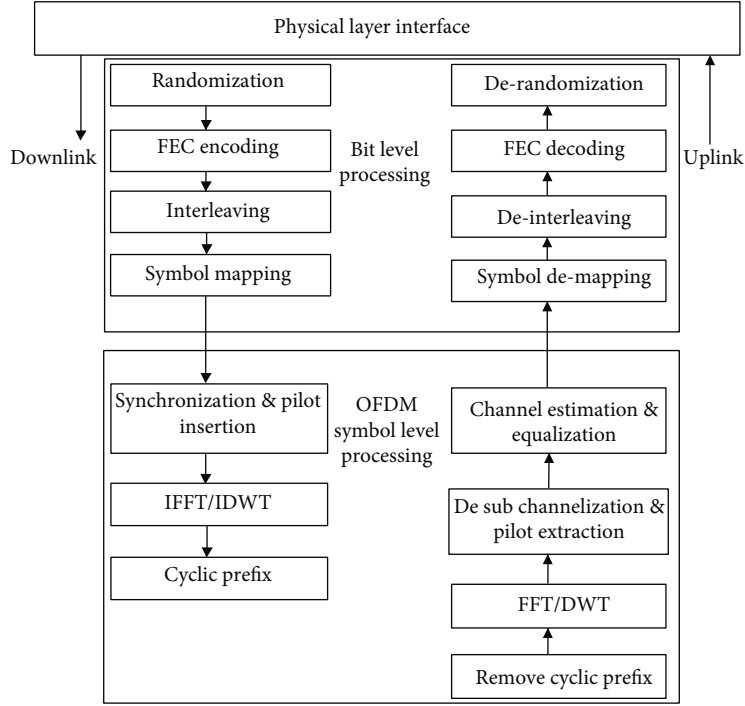
FIGURE 1: Physical layer model of proposed WiMAX system.
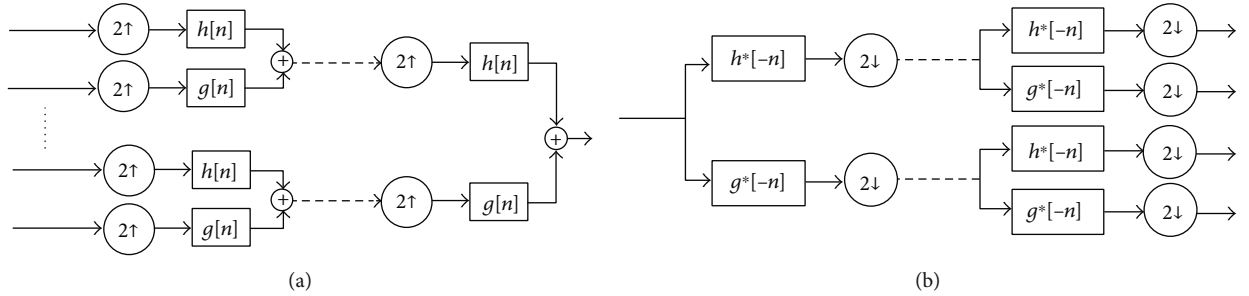
FIGURE 2: (a) IDWT and (b) DWT block diagram.

TABLE 2: Basic parameters of system simulations.

Parameter	Value
N_{FFT}	256
Cyclic prefix	1/8
Wavelets used	Haar, symlet, coiflet, and reverse biorthogonal
Channel coding rate	1/2, 2/3, 3/4
Modulation type	M-PSK and M-QAM
Modulation levels	2, 4, 16, and 64
Channel model	Rayleigh channel
Symbols used for simulation	10^6

$$(-1)^n h(n) = g(L - 1 - n). \qquad (2)$$

These independently modulated subcarriers when summed up coherently produce a large PAPR. The PAPR value for OFDM signal is defined as

$$\text{PAPR}_{dB} = 10 \log \left(\frac{\max [x(t) \, x^*(t)]}{E[x(t)x^*(t)]} \right), \qquad (3)$$

Here, $E[.]$ denotes the expectation operation. The PAPR will increase if we keep on increasing the number of

```
                    ┌─────────┐
                    │  Start  │
                    └─────────┘
                         │
              ┌────────────────────┐
              │     Input data     │
              └────────────────────┘
                         │
              ┌────────────────────┐
              │ Channel encoding and│
              │     modulation      │
              └────────────────────┘
                         │
              ┌────────────────────┐        ┌─────────────────────┐
              │ Apply DWT with a   │◄───────│ Change the wavelet i.e.│
              │  selected wavelet  │        │ haar/coiflet/symlet/ │
              └────────────────────┘        │  reverse biorthogonal │
                         │                  └─────────────────────┘
              ┌────────────────────┐                 ▲
              │ Send the data through│                │
              │   Rayleigh channel  │                │
              └────────────────────┘                 │
                         │                            │
              ┌────────────────────┐                 │
              │ Apply IDWT on received│               │
              │ data with a selected │               │
              │      wavelet        │               │
              └────────────────────┘                 │
                         │                            │
              ┌────────────────────┐                 │
              │ Demodulation and channel│             │
              │      decoding       │               │
              └────────────────────┘                 │
                         │                            │
              ┌────────────────────┐                 │
              │ Calculate the BER and SNR│            │
              │ required to achieve that BER│         │
              └────────────────────┘                 │
                         │                            │
                    ╱─────────╲                       │
                   ╱  BER      ╲      No              │
                  ╱ calculated for╲───────────────────┘
                  ╲ all waveletes ╱
                   ╲            ╱
                    ╲─────────╱
                         │ Yes
              ┌────────────────────┐
              │ Select wavelet as wavelet_opt│
              │ for which SNR required is│
              │      minimum        │
              └────────────────────┘
                         │
                    ┌─────────┐
                    │  Stop   │
                    └─────────┘
```

FIGURE 3: Proposed methodologies for DWT-incorporated WiMAX.

TABLE 3: SNR requirement analysis for WiMAX incorporating diverse wavelet families.

Modulation types/level with variable channel encoding rates	SNR (dB) requirement for the achievement of the minimum BER (10^{-4}) in WiMAX incorporated with FFT and diverse wavelets				
	FFT	Wavelet transform			
		Haar	Symlet	Coiflet	Reverse biorthogonal
BPSK (1/2 CC)	28	15.1	16	16.5	17.5
QPSK (1/2 CC)	37	22	23.5	24	23
QPSK (3/4 CC)	40	27	28	28	29
16-QAM (1/2 CC)	27	16.5	17.5	18.5	18
16-QAM (3/4 CC)	28	21.5	22	22.5	23
64-QAM (2/3 CC)	29.5	22	27	27.5	24
64-QAM (3/4 CC)	35.5	25	25.5	26.5	27

subcarriers. The prime objective of the PAPR reduction methodologies is to decrease the value of max $|x(n)|$. Rayleigh Channel is an additive, and subtractive quality of different paths segments in flat fading channels which are estimated by Rayleigh distribution having none observable pathway which implies if none immediate way is there

amongst sender and receiver. The simplification of the arrived signal can be

$$r(t) = s(t) * h(t) + n(t), \qquad (4)$$

where $r(t)$ is the received signal, $s(t)$ is the transmitted

(a) BPSK (1/2 CC rate)

(b) QPSK (1/2 CC rate)

(c) QPSK (3/4 CC rate)

(d) 16-QAM (1/2 CC rate)

(e) 16-QAM (3/4 CC rate)

(f) 64-QAM (2/3 CC rate)

(g) 64-QAM (3/4 CC rate)
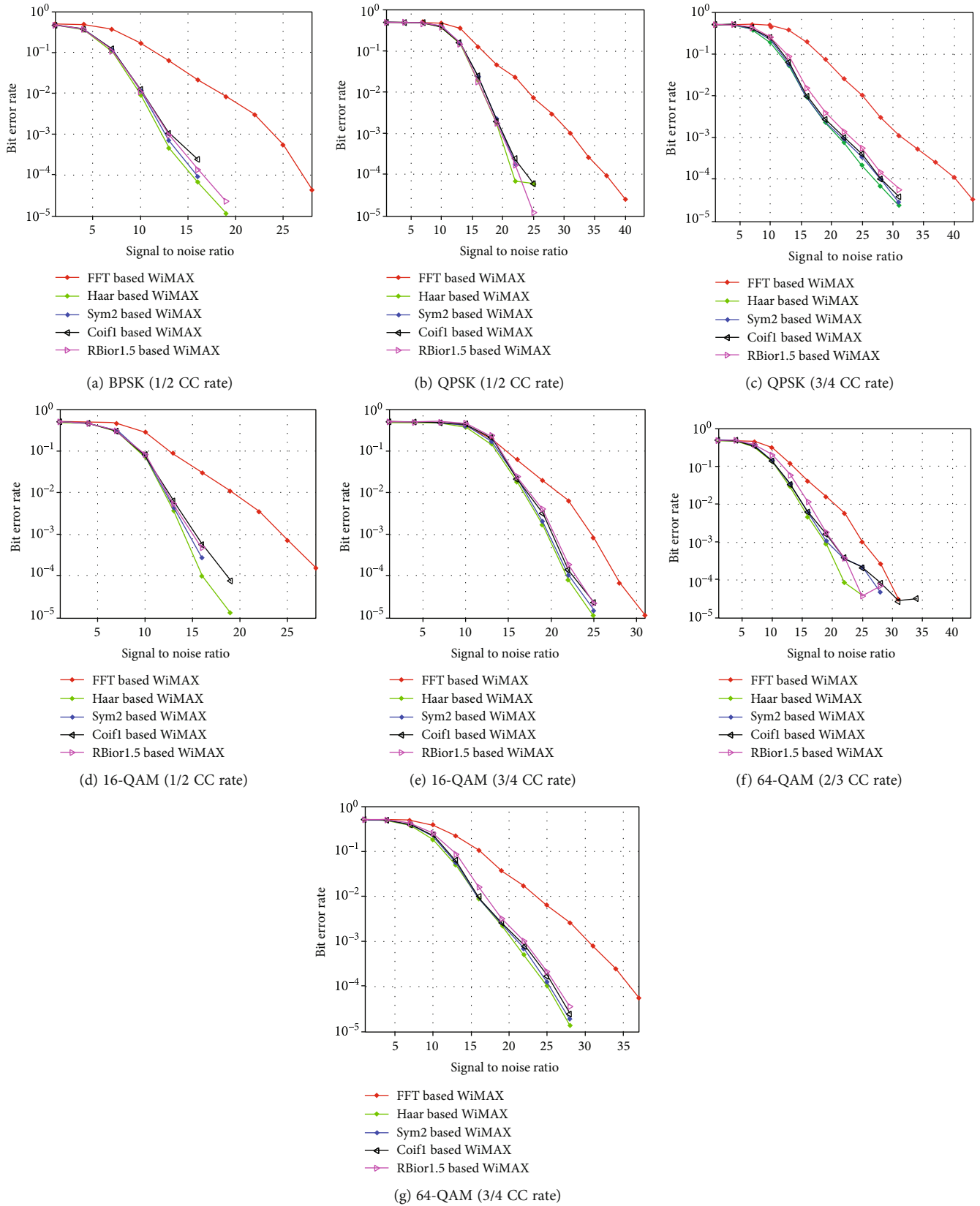
FIGURE 4: (a–g) SNR vs. BER measurements.

(a)

(b)

(c)
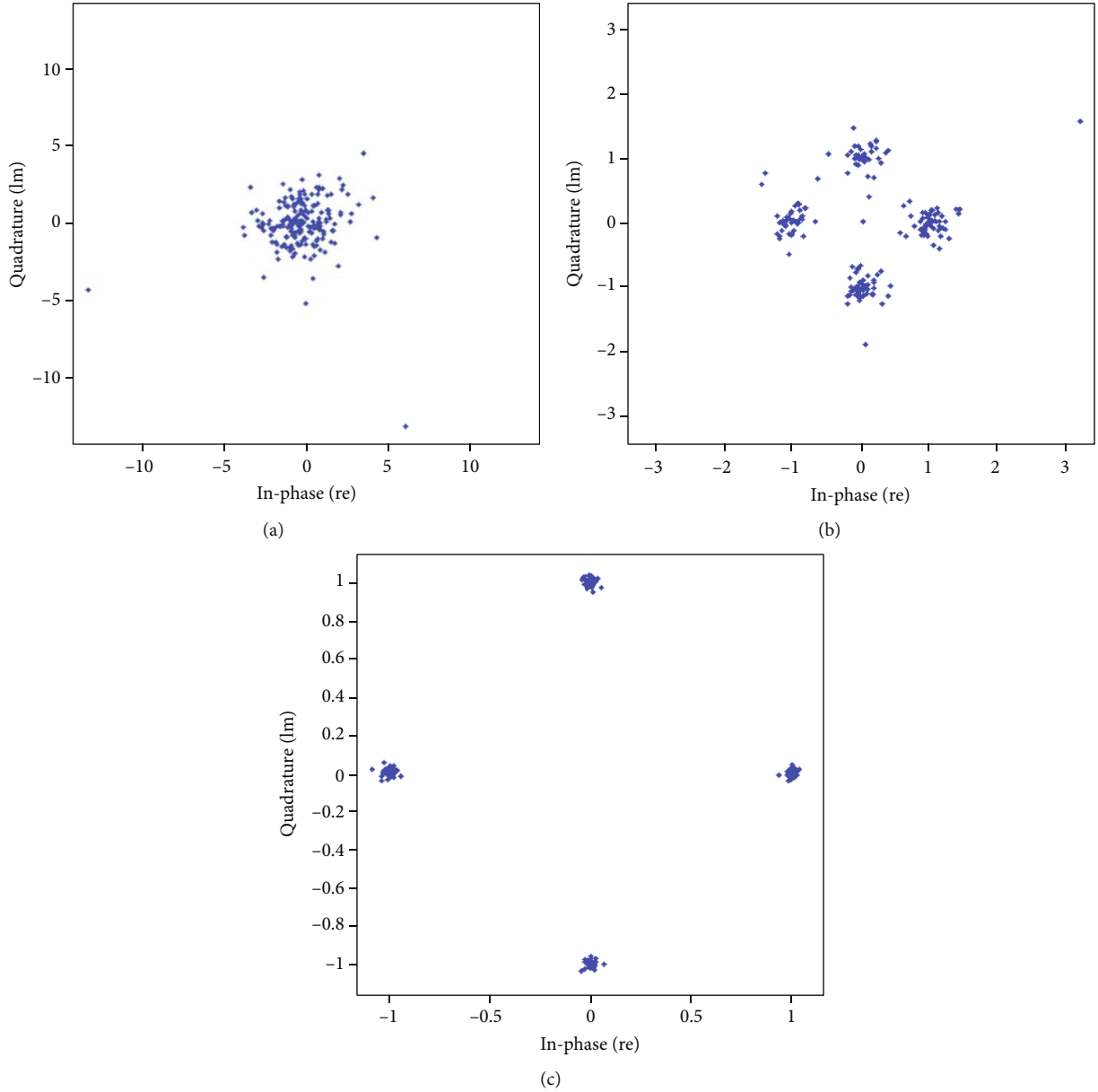
FIGURE 5: (a–c) Input to QPSK demodulator on using "Haar" wavelet for WiMAX system with variable SNR (a) 5 dB, (b) 15 dB, and (c) 30 dB.

signal, $h(t)$ symbolizes the channel matrix, and $n(t)$ symbolizes the AWGN. The probability density function (pdf) of "r" having Rayleigh distribution is expressed as

$$f(r) = \frac{r}{\sigma^2} e^{-r^2/2\sigma^2}, \ 0 \leq r \leq \infty, \tag{5}$$

where received signal's time-average power is represented by σ^2. The phase and the gain components of a channel's deformation are symbolized as a complex number often. Here, it is assumed for the exhibition of Rayleigh fading that the modelling of response's real and imaginary parts is done by individual allocated zero-mean Gaussian processes.

An issue experienced in the plan of receivers for exchanging information digitally is the recognition of information from estimations with the noise of the sent data. For a practical solution, the receiver is because of the disturbance bound making incidental mistakes or errors. Accordingly, planning a receiver with the characteristics of a likelihood of blunder is insignificant engaging practically and hypothetically. Lamentably, such outlines tend to bring out computationally complex receivers, and hence, they are frequently surrendered for computationally less complex yet problematic recipients.

ZF equalizer is a linear detection strategy that is utilized in transmission systems for detection purposes and was introduced by Lucky [33]. In this algorithm, the frequency response of the channel is reversed. It is ideal for a channel without having any noise, whereas, for a channel with noise, the noise also gets amplified greatly at frequency f where the little magnitude is there in channel response $H(j2\pi f)$ in the attempt to reduce the channel completely. At the side of the

(a)

(b)

(c)
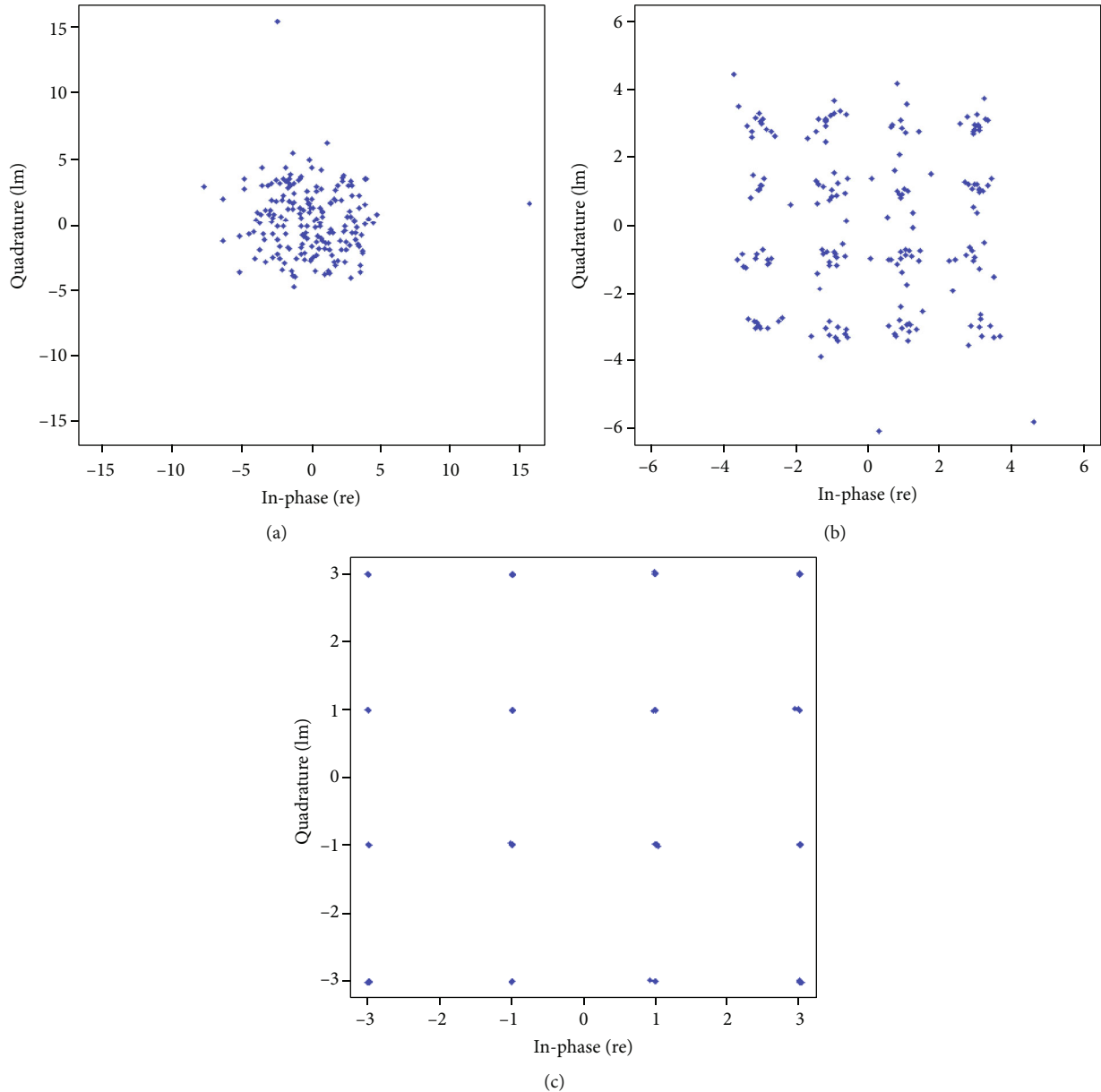
FIGURE 6: (a–c) Input to 16 QAM demodulator on using "Haar" wavelet for WiMAX system with variable SNR (a) 5 dB, (b) 15 dB, and (c) 30 dB.

transmission, if the CSI is known perfectly, then the system capacity can be achieved for the huge number of users by ZF precoding. To attain the complete multiplexing gain, ZF precoding requires the essential feedback overhead. Table 2 presents the basic simulation parameters along with values.

Due to the presence of cascaded 2 tap finite impulse response (FIR) filters and down samplers, the number of complex multiplications in Haar wavelet transform is higher than FFT. Therefore, the computational complexity of DWT-incorporated WiMAX is higher than the FFT-incorporated WiMAX. Besides, DWT performs well with nonstationary signals, whereas FFT is more suited for stationary signals.

The proposed methodology for carrying out the simulation of the WOFDM-WiMAX system incorporating diverse transforms is explained through a flowchart depicted in Figure 3. The proposed methodology in Figure 3 symbolises

the principle of utilizing the optimum wavelet from the family of the wavelets to guarantee the finest BER outcome. The utilization of wavelets results in enhancements in the BER outcome of WiMAX systems [30]. Figure 3 depicts the methodology that are being utilized to do the MATLAB simulations for the DWT-based WiMAX system, in the endeavour to decide the optimum wavelet from the family of the wavelets. Table 3 presents the thorough BER outcome enhancements over Rayleigh channel presented in the form of the SNR that is essential to accomplish a BER of 10^{-4} for DWT-based WiMAX systems.

4. Result Discussion

The outcome of the WiMAX system incorporating diverse wavelet families is analysed over the Rayleigh fading channel

(a)

(b)

(c)
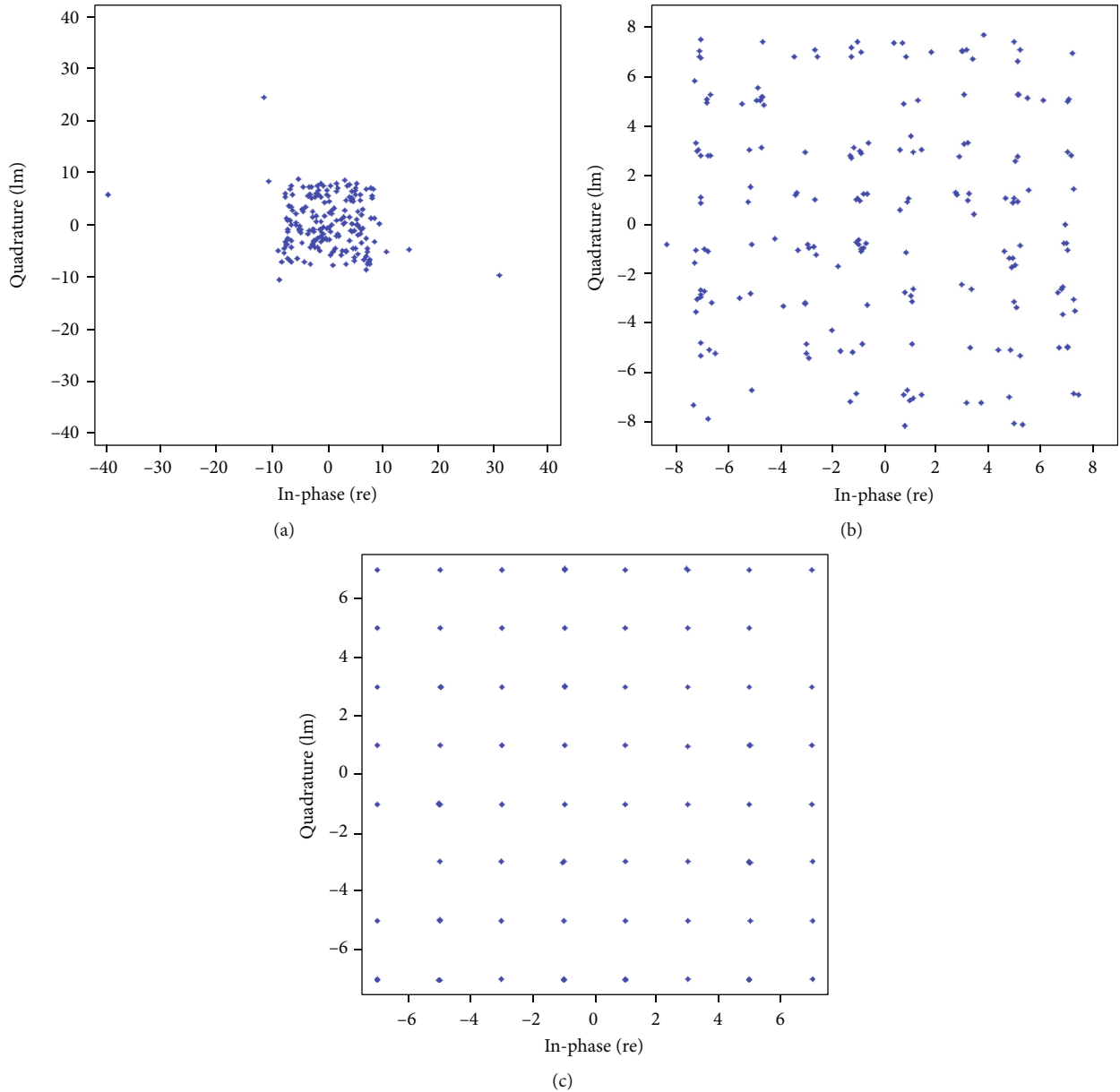
FIGURE 7: (a–c) Input to 64 QAM demodulator on using "Haar" wavelet for WiMAX system with variable SNR (a) 5 dB, (b) 15 dB, and (c) 30 dB.

using the comparison of SNR vs. BER deviations and PAPR reductions through the MATLAB simulations. It is assumed that the required BER for satisfactory performance is 10^{-4}, and the same is used to compare the SNR requirement for FFT-incorporated WiMAX and diverse wavelet-incorporated WiMAX. Figures 4(a)–4(g) depicts the simulation evaluation over the Rayleigh fading channel for both DWT-WiMAX and FFT-WiMAX using SNR vs. BER matrices for variable modulation levels and convolution coding (CC) rates, i.e., 1/2, 2/3, and 3/4. After drawing inferences from the simulation results, it is evident that the DWT-incorporated WiMAX performs superior to the FFT-incorporated WiMAX. DWT ensures that the orthogonality between the subcarriers remains intact even under the severe effects of the multipath fading [19, 20, 30, 32, 34]. It is apparent from Figure 4(a) that FFT-WiMAX necessitates an SNR

of 28 dB; however, this SNR prerequisite falls to 15.1 dB, 16 dB, 16.5 dB, and 17.5 dB for Haar, symlet, coiflet, and reverse biorthogonal wavelet-based WiMAX, respectively, to achieve a desired level of BER over Rayleigh fading channel. Rayleigh fading channel is used for the analysis purpose instead of AWGN channel as its distribution is more close to practical wireless channel. AWGN channel on internal noise component is considered for simulation of channel environment, whereas Rayleigh fading channel random noise and Rayleigh fading distribution are considered to model the channel environment. Also, in Rayleigh fading channel distribution, it is assumed that no direct path is available between transmitter and receiver side, which is quite equivalent to practical channel condition. Similar inferences can also be achieved in carefully analysing Figures 4(b)–4(g).

(a)

(b)

(c)
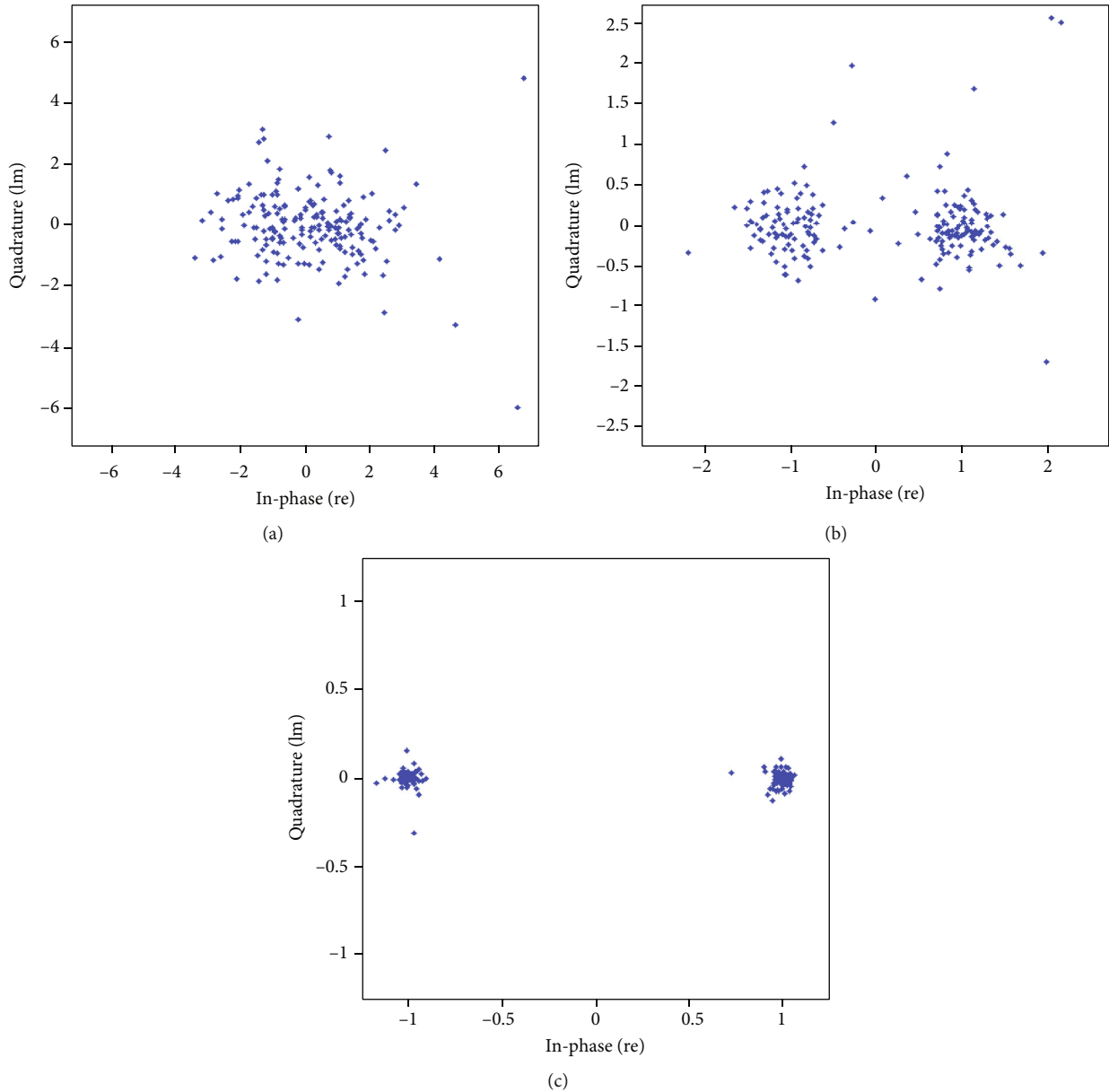
FIGURE 8: (a–c) Input to the BPSK demodulator on using "Haar" wavelet for WiMAX system with variable SNR (a) 5 dB, (b) 15 dB, and (c) 30 dB.

Table 3 provides the tabulated representation of the SNR requirements for the achievement of the minimum BER (10^{-4}) in WiMAX incorporated with FFT and diverse wavelets. It is very much evident from Table 3 that the DWT-incorporated WiMAX propose an enhancement of 10-15 dB of SNR in contrast to FFT-incorporated WiMAX. Also, the Haar transform outperforms the other wavelets, i.e., symlet, coiflet, and reverse biorthogonal wavelets in the wavelet family by a margin of .5 to 5 dB. The Haar wavelet performs better in comparison to other wavelet families due to the fact that it is the modest orthonormal wavelet basis. It comprises of only 2 taps, whereas there are 8 taps in the other wavelet families but with diverse features (orthogonal, biorthogonal, etc.). The Haar wavelet is theoretically modest, memory efficient, exactly changeable, and computationally

inexpensive. The Haar transform does not suffer from the effect of overlapping windows, whereas the Daubechies wavelet uses overlapping windows. Daubechies wavelets are the originating point of the coiflets. However, the coiflets suffer from the effects of overlapping windows and elevated computational complexities.

From the SNR vs. BER plots, it is quite clear that the Haar turns out to be most robust followed by the symlets and then the coiflets and reverse biorthogonal in the last in the wavelet family. The error is still there for all the modulation levels even at high values of SNR. An analysis of received symbols at the input of the demodulator is presented in Figures 5–7. In Figures 8(a)–8(c), the analysis is shown for BPSK modulation using the Haar wavelet transform. It is apparent from Figure 8(b) that even at 15 dB of SNR, the constellation

(a) BPSK (1/2 CC rate)

(b) QPSK (1/2 CC rate)

(c) QPSK (3/4 CC rate)

(d) 16-QAM (1/2 CC rate)

(e) 16-QAM (3/4 CC rate)

(f) 64-QAM (2/3 CC rate)

(g) 64-QAM (3/4 CC rate)
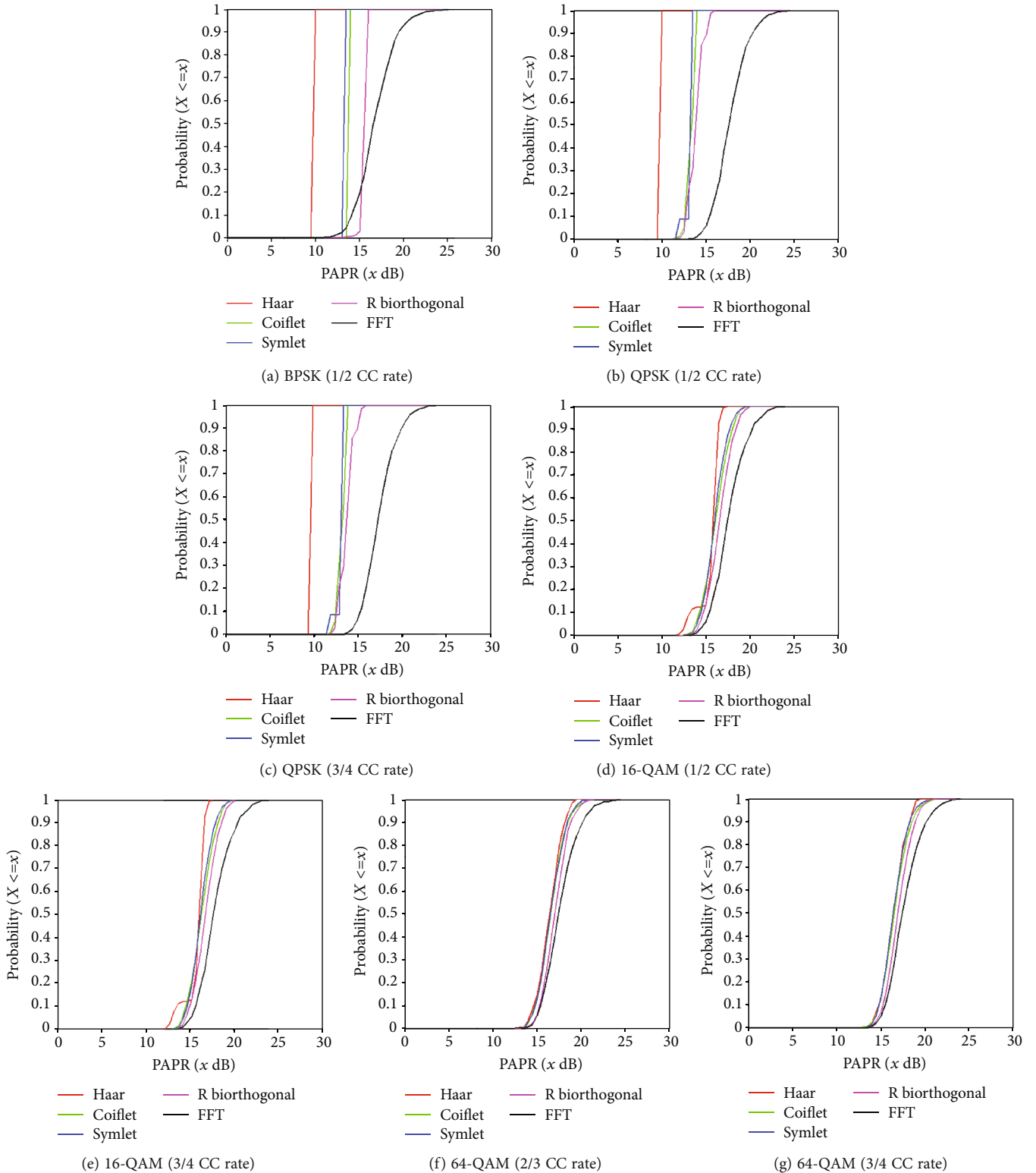
FIGURE 9: (a–g) PAPR evaluation.

points of the transmitted signal do not map to the constellation points of the received signal. Now, if we increase the SNR up to 30 dB as shown in Figure 8(c), still, the point does not exactly map on to the ideal constellation. Further, the same observation can be drawn out of Figures 5(a)–5(c),

6(a)–6(c), and 7(a)–7(c) which displays the constellation diagram of diverse modulation types/levels. The primary purpose for this is that in the wavelet transform, the parts of information being transmitted are contained by upper and lower subbands independently [22].

The PAPR performance of FFT-based WiMAX is matched with the DWT-based WiMAX. It is very much apparent from Figures 9(a)–9(g) that DWT-incorporated WiMAX performs superior to FFT-incorporated WiMAX diverse modulation types/levels. It is very much evident from Figure 9(a) that WiMAX based on diverse wavelets provides reduced PAPR of around 7-13 dB over FFT-based WiMAX for BPSK modulation that is favourable for the operation of RF amplifier. Similar inference can be drawn out after the analysis of subsequent results of PAPR reduction. However, from the overall analysis, it is also very much clear that Haar wavelet offers the highest PAPR reduction among the family of wavelets.

Most of 5G application requires rapid information correspondence for significant distance places with tolerable signal strength; the existence of elevated PAPR affects the excellence of the signal and range of the telecommunication system. Due to elevated PAPR, the distortions in the transmitted signal would increase, and due to condense high PAPR, the power amplifiers situated at the transmitter side must be operated at moderate or below moderate power levels. The consequence of the decreased power level leads to lessening in range of the telecommunication system. Therefore, the PAPR reduction achieved on incorporating the proposed methodology decrease the effects of elevated PAPR and deliver an optimum signal strength at the remote places for all the fixed as well as mobile users.

5. Conclusion

Overall performance analysis clearly depicts that wavelet-based WiMAX offers significant improvements regarding BER and PAPR reduction in comparison to FFT-based WiMAX. It has been observed that Haar wavelet is the most robust in the wavelet family followed by symlets, coiflets, and reverse biorthogonal. The DWT-incorporated WiMAX presents an SNR enhancement of 10-15 dB over the conventional FFT-WiMAX over Rayleigh fading channel links to achieve the desired BER outcome of 10^{-4}. Further, it is also reported that as keep on increasing the modulation level (BPSK, QPSK, 16-QAM, and 4-QAM) or convolution coding rates (1/2, 2/3, 3/4) to achieve the higher data rates, SNR requirements to achieve the required BER also increase. It is also very clear from the simulation results that the signal quality at the receiver side is much improved in the case of DWT-incorporated WiMAX in contrast to FFT-incorporated WiMAX. Furthermore, the results reveal out that wavelet-incorporated WiMAX proffers a reduced level of PAPR in the range of 7-13 dB over FFT-WiMAX for diverse modulation types/levels and is preferred for the operation of RF amplifier.

In the future, the hybrid combination of multiuser MIMO-OFDM/WiMAX system can also be utilized to achieve higher spectral efficiency and robust BER outcome. Also, the suggested methodology can also be utilized to study the outcome of OFDM/WiMAX or MIMO-OFDM/WiMAX system for transmission of multimedia information.

Data Availability

All data are included within this manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

References

[1] Local and Metropolitan Area Networks—Part 16, Air Interface for Fixed Broadband Wireless Access Systems, IEEE Standard, 2004.

[2] Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer in the 5 GHz Band, IEEE Standard, 1999.

[3] P. Singh, A. Kaur, R. S. Batth, G. S. Aujla, and M. Masud, "Service vs protection: a Bayesian learning approach for trust provisioning in edge of things environment," IEEE Internet of Things Journal, 2021.

[4] M. Masud, M. Alazab, K. Choudhary, and G. S. Gaba, "3P-SAKE: privacy-preserving and physically secured authenticated key establishment protocol for wireless industrial networks," Computer Communications, vol. 175, pp. 82–90, 2021.

[5] M. Masud, G. S. Gaba, K. Choudhary, M. S. Hossain, M. F. Alhamid, and G. Muhammad, "Lightweight and anonymity-preserving user authentication scheme for IoT-based healthcare," IEEE Internet of Things Journal, 2021.

[6] M. Masud, G. S. Gaba, S. Alqahtani et al., "A lightweight and robust secure key establishment protocol for internet of medical things in COVID-19 patients care," IEEE Internet of Things Journal, vol. 8, no. 21, 2021.

[7] M. K. Lakshmanan and H. Nikookar, "A review of wavelets for digital wireless communication," Wireless Personal Communication, vol. 37, no. 3-4, pp. 387–420, 2006.

[8] N. Ahmed, Joint Detection Strategies for Orthogonal Frequency Division Multiplexing, [M.S. thesis], RICE University, Houston, TX, USA, 2000.

[9] Z. Xiaodong, X. Pingping, Z. Guoan, and B. Guangguo, "Study on complex wavelet packet based OFDM modulation (CWPOFDM)," Acta Electronica Sinica, vol. 30, pp. 477–479, 2002.

[10] T. Pollet, M. van Bladel, and M. Moeneclaey, "BER sensitivity of OFDM systems to carrier frequency offset and Wiener phase noise," IEEE Transactions on Communications, vol. 43, no. 2/3/4, pp. 191–193, 1995.

[11] Si Jung Chang, Moon Ho Lee, and Ju Yong Park, "A high speed VLSI architecture of discrete wavelet transform for MPEG-4," IEEE Transaction Consumer Electronics, vol. 43, no. 3, pp. 623–627, 1997.

[12] S. Galli, H. Koga, and N. Kodama, "Advanced signal processing for PLCs: wavelet-OFDM," in 2008 IEEE International Symposium on Power Line Communications and Its Applications, pp. 187–192, Jeju, Republic of Korea, 2008.

[13] K. Abdullah, A. Z. Sadik, and Z. M. Hussain, "On the DWT- and WPT- OFDM versus FFT-OFDM," in *5th IEEE GCC Conference & Exhibition*, pp. 1–5, Kuwait, Kuwait, 2009.

[14] R. Asif, R. A. Abd-Al Hameed, O. Oanoh et al., "Performance comparison between DWT-OFDM and FFT-OFDM using time domain zero forcing equalization," in *2012 International Conference on Telecommunications and Multimedia (TEMU)*, pp. 175–179, Heraklion, Greece, 2012.

[15] D. Gupta, V. B. Vats, and K. K. Garg, "Performance analysis of DFT-OFDM, DCT-OFDM, and DWT-OFDM systems in AWGN channel," in *2008 The Fourth International Conference on Wireless and Mobile Communications*, pp. 214–216, Athens, Greece, 2008.

[16] Z. A. Hamid, M. Samir, S. M. Abd el-atty et al., "On the performance of FFT/DWT/DCT-based OFDM systems with chaotic interleaving and channel estimation algorithms," *Wireless Personal Communication*, vol. 78, no. 2, pp. 1495–1510, 2014.

[17] A. H. Kattoush, W. A. Mahmoud, and S. Nihad, "The performance of multiwavelets based OFDM system under different channel conditions," *Digital Signal Processing*, vol. 20, no. 2, pp. 472–482, 2010.

[18] M. Oltean and M. Nafornita, "Wavelet OFDM performance in frequency selective fading channels," in *2010 8th International Conference on Communications*, pp. 343–346, Bucharest, Romania, 2010.

[19] V. Sharma and G. Singh, "On BER assessment of conventional- and wavelet-OFDM over AWGN channel," *Optik*, vol. 125, no. 20, pp. 6071–6073, 2014.

[20] V. Kumbasar and O. Kucur, "Performance comparison of wavelet based and conventional OFDM systems in multipath Rayleigh fading channels," *Digital Signal Processing*, vol. 22, no. 5, pp. 841–846, 2012.

[21] L. Na, G. Yun-Quan, D. Chun-Mei, and H. Xue, "Optimizing performance of WIMAX system based on WPM," in *2012 7th International Conference on Computer Science & Education (ICCSE)*, pp. 372–375, Melbourne, VIC, Australia, 2012.

[22] S. L. Linfoot, "A study of different wavelets in orthogonal wavelet division multiplex for DVB-T," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1042–1047, 2008.

[23] S. H. Han and J. H. Lee, "PAPR reduction of OFDM signals using a reduced complexity PTS technique," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 887–890, 2004.

[24] I. Baig and V. Jeoti, "PAPR analysis of DHT-precoded OFDM system for M-QAM," in *2010 International Conference on Intelligent and Advanced Systems*, pp. 1–4, Kuala Lumpur, Malaysia, 2010.

[25] I. Baig and V. Jeoti, "A novel precoding based hybrid MC/SC radio access system for PAPR reduction in layered OFDMA of LTE-advanced IAENG," *International Journal of Computer Science*, vol. 38, pp. 314–319, 2011.

[26] I. Baig and V. Jeoti, "A new DHT precoding based interleaved-OFDMA uplink system with improved PAPR for 4G cellular networks," *Journal of Engineering Science and Technology*, vol. 6, pp. 685–694, 2011.

[27] I. Baig and V. Jeoti, "A ZCMT precoding based multicarrier OFDM system to minimize the high PAPR," *Wireless Personal Communications*, vol. 68, no. 3, pp. 1135–1145, 2013.

[28] I. Baig and V. Jeoti, "DCT precoded SLM technique for PAPR reduction in OFDM systems," in *2010 International Conference on Intelligent and Advanced Systems*, pp. 1–6, Kuala Lumpur, Malaysia, 2010.

[29] M. M. Hasan, "VLM precoded SLM technique for PAPR reduction in OFDM systems," *Wireless Personal Communication*, vol. 73, no. 3, pp. 791–801, 2013.

[30] L. Kansal, V. Sharma, and J. Singh, "BER assessment of FFT-OFDM against WHT-OFDM over different fading channel," *Wireless Networks*, vol. 23, no. 7, pp. 2189–2196, 2017.

[31] M. Srivastava, R. Ahmad, and K. K. Singh, "Comparative analysis of wavelet and OFDM-based systems," in *Computing Algorithms with Applications in Engineering*, pp. 107–115, Springer, 2020.

[32] L. Kansal, V. Sharma, and J. Singh, "Performance evaluation of FFT-WiMAX against WHT-WiMAX over Rayleigh fading channel," *Optik*, vol. 127, no. 10, pp. 4514–4519, 2016.

[33] R. W. Lucky, "Automatic equalization for digital communication," *Bell System Technical Journal*, vol. 44, no. 4, pp. 547–588, 1965.

[34] L. Kansal, V. Sharma, and J. Singh, "Multiuser massive MIMO-OFDM system incorporated with diverse transformation for 5G applications," *Wireless Personal Communications*, vol. 109, no. 4, pp. 2741–2756, 2019.

[35] M. Oltean and M. Nafornita, "Errors per scale statistics for a wavelet OFDM transmission in flat fading channels," in *2009 IEEE International Symposium on Intelligent Signal Processing*, pp. 119–124, Budapest, Hungary, 2009.

[36] K. Y. Min and V. Jeoti, "A novel signal independent technique for PAPR reduction in OFDM systems," in *2007 International Conference on Signal Processing, Communications and Networking*, pp. 308–311, Chennai, India, 2007.

[37] E. Sun, K. Yi, B. Tian, and X. Wang, "A method for PAPR reduction in MSE-OFDM systems," in *20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA'06)*, p. 4, Vienna, Austria, 2006.

[38] M. Park, H. Jun, J. Cho, N. Cho, D. Hong, and C. Kang, "PAPR reduction in OFDM transmission using Hadamard transform," in *2000 IEEE International Conference on Communications. ICC 2000. Global Convergence Through Communications. Conference Record*, pp. 430–433, New Orleans, LA, USA, 2000.

[39] K. Pachori and A. Mishra, "An efficient combinational approach for PAPR reduction in MIMO–OFDM system," *Wireless Networks*, vol. 22, no. 2, pp. 417–425, 2016.

[40] L. Zhuang, L. Liu, J. Li, K. Shao, and G. Wang, "Discrete sine and cosine transforms in single carrier modulation systems," *Wireless Personal Communication*, vol. 78, no. 2, pp. 1313–1329, 2014.

[41] X. L. Huang, G. Wang, J. Chen, and Q. Q. Sun, "A novel Haar wavelet-based BPSK OFDM system robust to spectral null channels and with reduced PAPR," *Wireless Personal Communication*, vol. 67, no. 3, pp. 599–612, 2012.

[42] F. Mangone, J. He, J. Tang et al., "A PAPR reduction technique using Hadamard transform combined with clipping and filtering based on DCT/IDCT for IM/DD optical OFDM systems," *Optical Fiber Technology*, vol. 20, no. 4, pp. 384–390, 2014.

[43] H. Kaur and M. L. Singh, "Performance analysis of RSCC based WiMAX system for different SUI channel conditions using QPSK and QAM-16," *Wireless Personal Communications*, vol. 88, no. 4, pp. 787–796, 2016.

[44] N. Kumar and B. S. Sohi, "Evaluation of conventional and wavelet based OFDM system for ICI cancellation," *Wireless Personal Communications*, vol. 91, no. 3, pp. 1435–1446, 2016.

[45] F. S. Hasan, M. J. Zaiter, and R. A. Mohammed, "Design and analysis of a wavelet packet modulation based differential chaos shift keying communication system," *Wireless Personal Communications*, vol. 109, no. 4, pp. 2439–2450, 2019.

[46] S. Kumar, A. Singh, and M. Kumar, "Covert communication integrates into wavelet packet transform OFDM system over Rayleigh fading channel," *Wireless Networks*, vol. 26, no. 1, pp. 81–89, 2020.

[47] N. Awasthi and S. Sharma, "Comparative analysis of least square, minimum mean square error and KALMAN estimator using DWT (discrete wavelet transform)-based MIMO-OFDM system," in *Intelligent Communication, Control and Devices*, pp. 233–241, Springer, 2020.

[48] M. Mounir, M. B. El_Mashade, S. Berra, G. S. Gaba, and M. Masud, "A novel hybrid precoding-companding technique for peak-to-average power ratio reduction in 5G and beyond," *Sensors*, vol. 21, no. 4, p. 1410, 2021.

[49] L. Kansal, V. Sharma, and J. Singh, "MIMO-WiMAX system incorporated with diverse tranformation for 5G applications," *Froniter of Optoelectronics*, vol. 12, no. 3, pp. 296–310, 2019.

WILEY | Hindawi

Research Article
Video Scene Information Detection Based on Entity Recognition

Hui Qian [ID],[1] Mengxuan Dai,[1] Yong Ma [ID],[1] Jiale Zhao,[1] Qinghua Liu,[1] Tao Tao,[1] Shugang Yin,[2] Haipeng Li,[3] and Youcheng Zhang[3]

[1]*School of Computer and Information Engineering, Jiangxi Normal University, Nanchang 330022, China*
[2]*Siji Network Security Technology (Beijing) Co., Ltd., Beijing 102209, China*
[3]*Nanjing Unary Information Technology Co., Ltd., Nanjing 210002, China*

Correspondence should be addressed to Yong Ma; may@jxnu.edu.cn

Video situational information detection is widely used in the fields of video query, character anomaly detection, surveillance analysis, and so on. However, most of the existing researches pay much attention to the subject or video backgrounds, but little attention to the recognition of situational information. What is more, because there is no strong relation between the pixel information and the scene information of video data, it is difficult for computers to obtain corresponding high-level scene information through the low-level pixel information of video data. Video scene information detection is mainly to detect and analyze the multiple features in the video and mark the scenes in the video. It is aimed at automatically extracting video scene information from all kinds of original video data and realizing the recognition of scene information through "comprehensive consideration of pixel information and spatiotemporal continuity." In order to solve the problem of transforming pixel information into scene information, this paper proposes a video scene information detection method based on entity recognition. This model integrates the spatiotemporal relationship between the video subject and object on the basis of entity recognition, so as to realize the recognition of scene information by establishing mapping relation. The effectiveness and accuracy of the model are verified by simulation experiments with the TV series as experimental data. The accuracy of this model in the simulation experiment can reach more than 85%.

1. Introduction

With the development of computer network and multimedia technology, the way characters receive information has shifted from traditional words and pictures to video stream. Taking China as an example, in the first half of 2020, the number of online audiovisual users has reached 901 million, with a year-on-year growth of 4.87% (https://new.qq.com/omn/20201014/20201014A05GLY00.html), which also leads to a sharp increase in video data. With the development of 5G technology, video's share of worldwide mobile data traffic will climb from 60% in 2018 to 74% in 2024 (https://blogs.cisco.com/sp/mobile-vni-forecast-2017-2022-5g-emerges). In such an environment with a large amount of video data, understanding video content is an important step for the intelligent system to approach human's understanding ability. It also has a great application value in social services,

national security, and industrial development. However, video data is characterized by nonstructuration, strong redundancy, high dimension, deep information hiding, and understanding difficulties. How to map the complex video information into the semantic space in line with human cognitive habits is a challenge for video information extraction.

In recent years, the extraction and analysis of video information has become an important research content in video processing, which is of great significance in video semantic extraction, video query, and other aspects. Character detection and background detection, which are similar to scene information detection, have been deeply studied and widely applied [1–9]. However, there are not many in-depth researches on video situational information. At present, most of the proposed model is targeted to the recognition of face [10], character [11, 12], or background content [14] of the video, by extracting key frames and recognizing

the character information or the scene information in the frames to realize the extraction of the relationship between characters [13–15] and video scene classification [16, 17]. Zheng and Yu [10] combined the squeeze-and-excitation network (SEN) and residual network (ResNet) to accurately detect the face information in each frame, extract the position of the target face, and then extract face features from adjacent frames through the RNFT model to predict the position of the target face in the next frame. Gong and Wang [16] extracted background audio signals from match shots and recognized the sound of cheering and hitting from the audio signals of each match shot. By combining background audio signals and shot image information, this method realizes a more accurate video classification. Ding and Yilmaz [14] used to analyze whether characters appear in the same video scene, so as to extract the relationship network of the characters in the video. Tran and Jung [15] counted the cooccurrence of characters in video images to extract their relationship. However, most of these methods only take the global character/scene features at the camera level into consideration, ignoring the local features with more information and the relations that exist among them.

Scene detection is also widely used in real life. For example, in the novel coronavirus epidemic which started from 2020, the mode of online meeting and online teaching has become more and more popular, and the video data of meeting and course have also increased. When we process these video data, we find that there is a kind of application condition, that is, in a video, we usually only pay attention to the state of a target person/object under a specific situation. For instance, if a student participates in two consecutive classes in the same classroom, and the surveillance camera in the classroom will shot a video of these two classes. And we would like to analyze the student's attendance in one of the classes to ensure whether he was late or left early or returned after leaving for a period of time. When using the video information processing model mentioned above to analyze it, we found the following problems:

(1) Without more information, it is difficult for the computer to directly judge whether the student is in a changed course or not

(2) The computer is able to recognize all the parts when the student was absent in the whole video, but the process of determining whether the absence occurred in the course we are concerned about usually needs to be done manually

Lei et al. [21] proposed the SSCD method. It realizes the recognition of changing objects in a fixed scene and judges the change of street scene. However, it can not solve the above problems. In the case of lens movement or a large number of personnel changes, the error rate of the model will increase greatly, and it is difficult to deal with the processing of human-centered video. Similarly, there is the method proposed by Santana et al. [22], which can realize the rapid recognition of moving objects from a fixed perspective and judge the scene changes based on the results. However, this method can only obtain the contour map of

moving objects and still can not well solve the above problems. The method proposed by Huang and Liao [23] can realize the scene detection task from the perspective of motion, but it has certain requirements for the consistency of video. At the same time, the method compares frame by frame, which has high requirements for the performance of the machine and insufficient processing speed.

To solve the above problems, a video scene information detection model based on entity recognition is proposed in this paper. This model makes use of more information including global information at video level and partial information at entity level for more information to get more accurate results. Similar to this example, there are many application conditions, such as the situational judgment of meeting process and the abnormal judgment of security video, etc., but existing video processing models are not able to handle such application conditions well.

According to the spatiotemporal features of the video scene, this paper selects the state of the video object as the characteristic to help us analyze and understand the video scene, combines with the state feature of the video subject, and determines the scene feature of the video subject. In this paper, the innovations can be summarized as the following three points:

(1) This paper proposes a new situational information detection model, which can recognize the changes of video situational information with high efficiency

(2) This paper establishes situational features by combining the spatiotemporal continuity between the subject and the object in video content, which enables the model to recognize situational information without semantic information of the video object and achieves good results.

(3) The accuracy of the model proposed in this paper reaches 80%

In this paper, we will explain and verify the above research contents. Section 2 will briefly introduce the existing entity recognition models, such as Yolo, and some mature face recognition models, such as face recognition. At present stage, these models are the premise for the test in this study. In Section 3, we will introduce the models, including their establishment, mathematical basis, and partial content of the pseudocode. Section 4 will present our experimental results and summarize the failed parts, which are also what need to be further discussed in our subsequent research work. In Section 5, we will summarize the research content and briefly introduce the main research directions in the future.

2. Relevant Work

2.1. Yolo. Yolo is a new target detection method [18], which is characterized by rapid detection and high accuracy. Redmon regarded the target detection task as a regression problem of target region prediction and category prediction. In this method, a single neural network is used to directly

predict item boundary and category probability to achieve end-to-end item detection. Yolo is widely used in target detection [19], target tracking [20], and other applications. Zhang et al. [19] used the deep separable convolutional method to optimize the convolution layer of the tiny Yolo model and divided a complete convolution operation into deep convolution and point-by-point convolution, thus reducing the parameters of CNN and improving the operation speed. Mohammed et al. [20] combined the neural network, image-based tracking, and Yolo V3 to solve the problem of intelligent vehicle tracking.

In this paper, Yolo V4 can be used as the target detection network in the entity detection stage. On the basis of Yolo V3, Yolo V4 has made a lot of innovations. The innovation of the input end is mainly the improvement of the input end during training, including Mosaic data enhancement, CMBN, and SAT self-confrontation training. Backbone network combines all kinds of new ways, including CSPDarknet53, Mish activation function, and Dropblock. The neck target detection network often inserts some layers in the backbone and the final output layer, such as the SPP module in the Yolo V4 and FPN+PAN structure. The anchor frame mechanism of the output layer is the same as that of Yolo V3. The main improvement is in the loss function Clou-Loss during training, and the NMS screened by the prediction box is changed into DIOU-nms. Yolo V4 is a major update of the Yolo series, with average accuracy (AP) and frame per second (FPS) in the COCO dataset improved by 10% and 12%, respectively.

2.2. Face Recognition Algorithm.

In the model proposed in this paper, it is also feasible to directly use the face recognition algorithm to replace the target detection network. This method will reduce the accuracy of the model to some extent, but meanwhile, the computing efficiency will be better than the complete target detection network. When only the face recognition algorithm is used for scene information detection, the target object will be replaced by face recognition results, which greatly reduces the computational load of the model.

Face recognition is a powerful, simple, and easy-to-use face recognition open-source project, equipped with integrated development documents and application cases, and compatible with the Raspberry Pi system. You can use Python and command line tools to extract, recognize, and manipulate faces. Face recognition is a deep learning model based on C++ open-source library dlib. The face dataset Labeled Faces in the Wild is used for testing with a 99.38% accuracy. But the recognition accuracy of children and Asian faces has yet to be improved.

SeetaFace2 is a face recognition project written in C++ that supports Windows, Linux, and ARM platforms and does not rely on third-party libraries. This project includes face recognition module FaceDetector, face key point locating module Face Landmarks, and face feature extraction and comparison module Facerecognizer. FaceDetector can achieve a recall rate of over 92% under the condition of 100 false detections on FDDB, it also supports 5-point and 81-point localization of face key points, and its 1-to-N mod-

ule supports face recognition applications with a base of thousands of characters.

3. Model

3.1. Model Description.

The steps of video scene information extraction are as follows: Firstly, the input video is analyzed and preprocessed to obtain the entity target in each frame of the video. The main purpose of this work is to lay a good foundation for the subsequent subject-object labeling and the establishment of spatiotemporal relationship. Secondly, according to the input subject picture, the entity targets are compared and labeled, and the remaining entity targets are labeled as the object. Then, the video subject labeling results are used as scene nodes to extract and analyze the spatiotemporal relationship between the objects and the subjects in the video, so as to judge whether the scene is continuous or not. Finally, the attributes of scene nodes, namely, the scene information of the subject, is determined in the continuous scene.

This paper mainly focuses on the following:

(1) How to establish the relationship between subjects and objects?

(2) How to judge the attributes of scene nodes?

The model in this paper completes the above research contents through three stages of information processing.

3.1.1. The First Stage: Establish the Spatiotemporal Relationship between the Subject and the Object.

In this stage, we lock the current situational information by establishing the relationship between the subject and the object, which is also the situational feature introduced in the model. The spatiotemporal relationship is mainly based on the randomness of object selection. Under the same condition, the mathematical probability that a certain number of randomly selected entities in the initial image of the scene will simultaneously appear abnormal in this period of time and space is very small.

According to the Bayesian probability formula, let the subject be X, and the object set $Y = \{y_1, y_2 \cdots, y_n\}$, $y_1, y_2 \cdots, y_n$ are independent of each other and random; the anomaly probability of Y can be shown as P_y; and the probability of n object anomalies and subject anomalies at the same time can be shown as P.

$$P = P_y{}^n * P_{\text{subject anomaly}}, \tag{1}$$

can be concluded. As shown in Figure 1, when $P_y = 0.3$, the probability of misrecognition is less than 5% when the value of n is greater than 3.

The spatiotemporal relationship is also reflected in the spatiotemporal continuity of the object. In the same scene, the mathematical probability of continuous abnormal occurrence of an entity randomly selected in the initial image of the scene in this period of time and space is also very small.
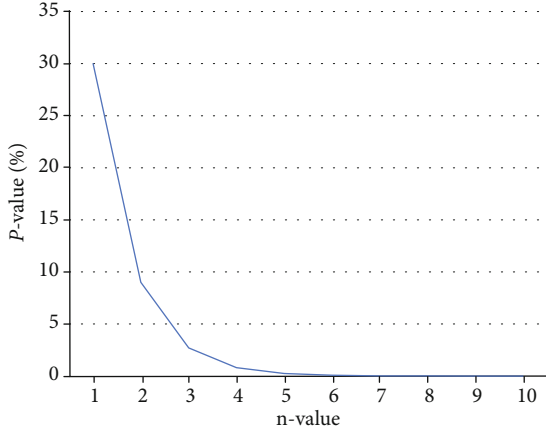
FIGURE 1: Influence curve of n value on model accuracy when $P_{\text{object}} = 0.3$.

FIGURE 2: Influence curve of n value on model accuracy when $P_{\text{object}} = 0.2$.

Since the occurrence of an anomaly in the same entity is an independent event, according to the principle of event independence, the object anomaly probability is assumed to be $P(y)$, and the probability P_{object} of object continuous abnormal n times.

$$P_{\text{object}} = P(y)^n, \tag{2}$$

As shown in Figure 2, in the case of $P_{\text{object}} = 0.2$, when n value is 2, the probability of misrecognition is 4%.

3.1.2. The Second Stage: Recognize whether the Subject and the object Are Abnormal. After establishing the subject-object relationship in the previous stage, we can realize the marking on the clips of the same scene in the video.

The main work in this stage can be divided into three steps. Firstly, each frame image is named according to the video frame order, and the same scene fragments are split one by one. Secondly, the features of the partial images of each entity in each frame of the same scene is extracted and compared with the target image feature to recognize whether the subject is in each video frame of the continuous scene. And the file names of each image that the subject exists are extracted as the subject recognition set. Finally, the features of the partial images of each entity are compared with the recognized object image features to recognize whether the object is in each video frame of the continuous scene, and the file names of each image that the object exists are extracted as the object recognition set.

The scene feature of a video V in a continuous scene is defined as

$$\widehat{V}(X, Y, t) = \frac{V(X, Y, t) - V(X, Y, t_l)}{\sigma(X, Y, t_l)}, \tag{3}$$

where $X \in \{1, 2, \cdots, M\}$, $Y \in \{0, 1, \cdots, N\}$ are scene indices and M and N are the number of subject and object of the video frame, respectively. $t \in \{1, 2, \cdots, T\}$ is the temporal index, and T is number of frames in a video. $V(X, Y, t)$ is

the quantity of subject and object at time t. And t_l is the last time the situation changed.

$$\sigma(X, Y, t_l) = \frac{\sum_{t_l}^{t} V(X, Y, t)}{t - t_l}, \tag{4}$$

is the average value of V of each frame from scene change to present.

Since feature extraction and comparison are independent, tasks at this stage can improve detection efficiency through parallel approach. Similarly, nature video has high correlation among neighboring pixels both in space and time. In order to further improve the processing efficiency, we can also choose to extract a picture every few frames for comparison.

3.1.3. The Third Stage: Calculate the Results of Scene Detection. After the subject and object recognition set of the previous stage is obtained, we can integrate them to obtain scene detection results. In the above work, the method of renaming each frame image and taking each image file name as the result set is to reduce the computational load at this stage, so as to improve the efficiency of result integration.

The work in this stage is mainly divided into two steps. Firstly, the intersection part of each object recognition set is taken, and then, the intersection with the subject recognition set is taken. This part of image has two features: (1) the scene information does not change under the same scene and (2) there is no exception in the body. According to the above two features, we can get the video clips of the corresponding frame of intersection images and ensure that the scene of this video is unchanged and the subject is not abnormal.

$$R_{\text{normal}} = (A \cap B_1) \cup (A \cap B_2) \cdots \cup (A \cap B_n), \tag{5}$$

where A is the result frame set that recognizes the subject and B_n is the result frame set that recognizes the object.

Then, the image filenames of the intersection of the image filenames of the subject recognition set and the object

```
Input: entity target set, number of reference objects
Output: object target set
1:encodings is all entity target codes of first_frame
2:   for i in 0 to length of encodings − 1:
3:      if target code equal to the ith entity code of encodings:
4:      adds I to temp list
5:         end
6:   num is the number of reference objects
7:   for i in 0 to num − 1:
8:      rand is a random integer from 0 to length of encodings − 1
9:      while rand is in temp:
10:        rand is a random integer from 0 to length of encodings − 1
11:      adds rand to temp list
12:      adds the rand element of encodings to the lst
```

ALGORITHM 1: Entity relationship algorithm.

recognition set are compared to obtain the partial images containing the subject but not containing the object. According to the comparison results, we can get the video clips of corresponding frames and determine that the scene changes have been taken place in this video.

$$R_{\text{abnormal}} = \bigcup - R_{\text{normal}}. \qquad (6)$$

3.2. Establishment of Spatiotemporal Relationship. A large number of existing models, such as Yolo and face recognition, have been able to realize fast entity recognition of image information. In this paper, such entity recognition results are directly seen as the entity target of video scenes.

After the entity target result set of the video is obtained by using the above models, the entity relationship algorithm is used to establish the relationship between the subject and the object target and establish scene features.

The entity object of the frame is recognized by the entity recognition method and is used as the input of the relationship algorithm. The number of reference objects is determined by the user, and the corresponding number of objects is arbitrarily selected from the entity object as the reference object of the current scene. According to the naive Bayes theory, there is a great similarity between the arbitrarily selected object and the subject in the same scene, and the more the selected objects, the stronger the relationship in the space and time.

3.3. Judgment of Scene Node Attributes. After the establishment of spatiotemporal relationship, the continuity of scene information is firstly detected. Only in continuous scenes can the judgment of scene node attributes have practical application attributes. After obtaining the continuous video clips of scene information, the attributes of scene nodes are determined according to the state of the main body of the video

The subject target and the output results of the relationship algorithm are taken as the input of the judgment algorithm. According to the relationship between the subject and the object, the entity target in the current frame is tra-

```
Input: subject coding, object target set
Output: Situational node attributes
1:i = 0
2:   encodings is all entity targets encode of frame
3:   for all the encode of scene's entity targets:
4:      if encode is not in encodings:
5:         i = i + 1
6:   if i equals to scene_entity:
7:      scene has been changed
8:   else:
9:      scene has not been changed
10:encodings is all object targets encode of frame
11:   if target_encode is in encodings:
12:   subject of video is in a particular situation
13:   else:
14:   subject of video is not in a particular situation
```

ALGORITHM 2: Scene attribute judgment algorithm.

versed, and the scene attributes are determined from the subject state and object state.

3.4. Video Scene Detection Model. After the completion of entity relationship and scene node attribute judgment, the information of one scene can be detected. However, in general, a video contains multiple scene information. Therefore, on the basis of Algorithm 1 and Algorithm 2, this paper proposes Algorithm 3 to realize the detection of all scene information in a video data.

The content of the first frame of the video is taken as the initial scene information. Algorithms 2 and 3 traverse the video data. When the change of video scene information is detected, the time sequence of the scene change frame is recorded, and the content of the scene change frame is taken as the initial scene information of the subsequent video data to cycle to the end of the video.

4. Experiments

4.1. Experimental Data. The experimental dataset adopted in this paper is a public video dataset; the main content of

```
Input: video data
Output: Scene detection results
1:for i in 0 to total video frames − 1:
2:    ic = i + 1
3:    image is the icth frame of video
4:    if image is the first frame:
5:          determine if the subject of video is in the first frame
6:    else:
7:          if scene has not been changed:
8:                determine if the subject of video is in the frame:
9:                      if result is True
10:                           adds ic to timeImage
11:                           set nextframe to False
12:                      else:
13:                           if time lag between now and the last scene is more than 3 s:
14:                                 adds ic to timeImage
15:                                 set nextframe to False
16:                           else:
17:                                 clear the last record in timeImage
18:                                 set nextframe to True
19:          else:
20:                is_end = True
21:                for i in the number of frames in scene change:
22:                      if scene back to original:
23:                            is_end = False
24:                      end if
26:                end for
25:                if is_end is True:
26:                      adds ic to timeImage
27:                end if
```

ALGORITHM 3: Video scene detection model.

which is TV play Ten Miles of Peach Blossom (the data comes from Tencent video, which is only used for academic research in this paper, and the copyright belongs to Tencent company), Hospital Playlist (the data comes from Netflix and is only used for academic research in this paper; the copyright belongs to Netflix company), Nirvana in Fire (the data comes from Tencent video, which is only used for academic research in this paper, and the copyright belongs to Tencent company), and It started with a Kiss. The average scenario switching time of each dataset is 7-10 seconds.

In this experimental environment, in order to analyze the performance of the algorithm proposed in this paper, the evaluation index used in this study is precision.

$$precision = \frac{correctly\ recognized\ number}{total}. \qquad (7)$$

The values are between 0 and 1, and the closer they are to 1, the better the effect of the model will be.

The hardware configuration information used in the experiment is as follows: CPU R53600, graphics card GTX1660, internal storage 16G, operating system Win10, and development language Python3.

4.2. Experimental Results and Analysis. As mentioned in Section 3, the model proposed in this study is to process and

TABLE 1: The experimental results based on face recognition.

Dataset	Correctly recognized	Total	Wrongly recognized
Dataset1	7	7	0
Dataset2	9	10	1
Dataset3	16	18	0
Dataset4	15	17	0
Dataset5	12	14	0

calculate the entity recognition results in the video, so we will use the existing mature entity recognition algorithms in the experiment. Two existing character recognition algorithms are used to meet the needs of model operation. First, face recognition was used to extract environmental features and human face features; second, SeetaFace2 was used to extract environmental features and human face features. The evaluation criteria are whether the target disappears and whether the scene transforms. In this experiment, scene changes have been manually marked. The precision of marking is seconds, and the precision of model detection is video frames. Due to the inconsistency of precision between manual marking and model detection, when the time axes corresponding to the video frame contained in the detection results are the same as that of manual marks, the results are right.

FIGURE 3: Incorrect results with face recognition.

4.2.1. Experiments Based on Face Recognition. In this paper, we use Face_recognition as the entity recognition part of the model. As the entity to establish scene features in the model, it is used to recognize the characters in the video image. After establishing the association between the entities in the video (Algorithm 1), the model calculates and determines the scene characteristics of each frame based on Formula (3), records the frame number of $\widehat{V}(X, Y, t)$ difference that is abnormal, and determines the corresponding time point on the time axis. Then, we compare it with the change time which is manually marked and get the test result.

These five datasets have been, respectively, tested, and the experimental results are as shown in Table 1.

TABLE 2: The experimental results based on SeetaFace2.

Dataset	Correctly recognized	Total	Wrongly recognized
Dataset1	7	7	0
Dataset2	9	10	0
Dataset3	14	18	0
Dataset4	16	17	1
Dataset5	12	14	0

We have found a part of the unrecognized images, as shown in Figure 3.

We have conducted separate recognition processing and found that some frames of the model could not recognize

FIGURE 4: Incorrect results with SeetaFace2.

their face features, leading to recognition errors during scene detection, which may be caused by decorations on the face.

The experimental results show that face recognition can work well in this dataset. However, face recognition sometimes fails to extract character features because face features are currently used for scene features. The occurrence of the above problems has some serious impacts on the establishment of scene features in the proposed model. Due to the errors of character face recognition, the frame missing a certain entity is wrongly recognized as the changes of scene features during the establishment of scene features. On the whole, the model proposed in this paper does well in scene detection in the case of limited entities.

4.2.2. Experiments Based on SeetaFace2. The experiment is similar to the former one, but we make some changes that we use SeetaFace2 instead of face recognition as the entity recognition part of the model. SeetaFace2 is used to recognize the faces in the video image as the entity to establish scene features of the model and test the same datasets. The experimental results are as shown in Table 2.

We have found a part of the unrecognized images, as shown in Figure 4.

We have found that the SeetaFace2 model recognized faces very sensitively and even can achieve the recognition of supporting characters in the background of photos. And as the camera moves, the number of supporting characters

changes dramatically, leading to the misjudgment of a scene switch.

The experimental results show that the recognition effect of SeetaFace2 is very sensitive, and SeetaFace2 uses the model structure of ResNet50. In this network, multiple residual learning blocks are connected in series, and the deep representation of the deep learning image of the model is utilized, so the recognition effect is very sensitive thus the recognition error occurs. Different from the problems mentioned above, when SeetaFace is combined with the model proposed in this paper, entities will increase abnormally in some frames due to the recognition of background characters. This leads to the errors of the proposed model in establishing scene features, resulting in recognition errors.

4.2.3. Experiment Summary. In addition to the above open dataset experiments, we also used 20 self-made datasets for experiments, and the content of them is meeting recording videos. In order to produce situational changes, the videos have some situations such as characters leaving midway, characters joining midway, and meeting pausing. The results are as follows:

(i) Face recognition: 87%

(ii) SeetaFace2: 85%

The test results meet the expectations. The model proposed in this paper can achieve more accurate scene change detection. It can realize video scene change detection on the premise of using face recognition results as the main entity. The feasibility and universality of the model have been already proved in the experiment. We believe that the accuracy can be further improved if the result including object recognition is introduced as an entity. But in some special cases, such as too many characters in the background, characters turning back, and decorations on the face, it will lead to the failure of scene recognition. In the future, we plan to increase the correct rate of scene transformation recognition by using judgment logic, model recognition, adding background object feature recognition module, and other measures.

5. Conclusion

This paper proposes a video scene information detection based on entity recognition, which can achieve the task of video scene information detection on the premise of entity recognition of video pixel data. The proposed model has strong robustness, and the precision can reach more than 85%. At the same time, it can replace entity recognition with face recognition algorithm as the input of scene information detection without too many impacts on the results of scene information detection.

In this paper, we will explain and verify the above research contents. Section 2 briefly introduces the existing entity recognition models, such as Yolo, and some mature face recognition models, such as face recognition. At present stage, these models are the premise for the test in this study. In Section 3, we introduce the models, including their estab-

lishment, mathematical basis and partial content of the pseudocode. Section 4 presents our experimental results and summarizes the failed parts, which are also what need to be further discussed in our subsequent research work. In Section 5, we summarize the research content and briefly introduce the main research directions in the future.

We take the spatiotemporal relationship of video entities as the basis of situational information detection and creatively put forward the concept of situational features to ensure the logical accuracy of the model. In the process of experiments, we found some existing problems, such as overreliance on the accuracy of entity recognition and difficulties in screening noise information effectively. In the research, we focus on how to better combine the entity recognition model with the model proposed in this paper to improve the detection efficiency of the proposed video scene information detection model.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. Li, J. Wang, H. Wang, M. Zhao, W. Li, and X. Deng, "Visual-texual emotion analysis with deep coupled video and danmu neural networks," *IEEE Transactions on Multimedia*, vol. 22, no. 6, pp. 1634–1646, 2020.

[2] J. Zhang, K. Mei, Y. Zheng, and J. Fan, "Exploiting mid-level semantics for large-scale complex video classification," *IEEE Transactions on Multimedia*, vol. 21, no. 10, pp. 2518–2530, 2019.

[3] T. Prathiba and R. S. S. Kumari, "Eagle eye CBVR based on unique key frame extraction and deep belief neural network," *Wireless Personal Communications*, vol. 116, no. 1, pp. 411–441, 2021.

[4] M. M. Azab, H. A. Shedeed, and A. S. Hussein, "New technique for online object tracking-by-detection in video," *IET Image Processing*, vol. 8, no. 12, pp. 794–803, 2014.

[5] D. Tao, X. Li, S. J. Maybank, and X. Wu, "Human carrying status in visual surveillance," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 1670–1677, New York, NY, USA, 2006.

[6] Z. Jin, Z. Lou, J. Yang, and Q. Sun, "Face detection using template matching and skin-color information," *Neurocomputing*, vol. 70, no. 4-6, pp. 794–800, 2007.

[7] D. Tao, X. Li, X. Wu, and S. Maybank, "Elapsed time in human gait recognition: a new approach," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 2, II-II, Toulouse, France, 2006.

[8] B. Xiao, X. Gao, D. Tao, and X. Li, "A new approach for face recognition by sketches in photos," *Signal Processing*, vol. 89, no. 8, pp. 1576–1588, 2009.

[9] K. Hui, J. Wang, H. He, and W. H. Ip, "A multilevel single stage network for face detection," *Wireless Communications*

and *Mobile Computing*, vol. 2021, Article ID 5582132, 10 pages, 2021.

[10] G. Zheng and Y. Xu, "Efficient face detection and tracking in video sequences based on deep learning," *Information Sciences*, vol. 568, pp. 265–285, 2021.

[11] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient HOG human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011.

[12] P. Bojanowski, F. Bach, I. Laptev, J. Ponce, C. Schmid, and J. Sivic, "Finding actors and actions in movies," in *Proceedings of the IEEE international conference on computer vision*, pp. 2280–2287, Sydney, NSW, Australia, 2013.

[13] J. Lv, W. Liu, L. Zhou, B. Wu, and H. Ma, "Multi-stream fusion model for social relation recognition from videos," in *Multi-Media Modeling. MMM 2018*, vol. 10704, Springer, Cham, 2018.

[14] L. Ding and A. Yilmaz, "Learning relations among movie characters: a social network perspective," in *Computer Vision – ECCV 2010. ECCV 2010*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6314 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2010.

[15] Q. D. Tran and J. E. Jung, "CoCharNet: extracting social networks using character co-occurrence in movies," *Journal of Universal Computer Science*, vol. 21, pp. 796–815, 2015.

[16] X. Gong and F. Wang, "Classification of tennis video types based on machine learning technology," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 2055703, 11 pages, 2021.

[17] Y. Li, "Research on sports video image analysis based on the fuzzy clustering algorithm," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6630130, 8 pages, 2021.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, Las Vegas, NV, USA, 2016.

[19] S. Zhang, Y. Wu, C. Men, and X. Li, "Tiny YOLO optimization oriented bus passenger object detection," *Chinese Journal of Electronics*, vol. 29, no. 1, pp. 132–138, 2020.

[20] M. A. A. al-qaness, A. A. Abbasi, H. Fan, R. A. Ibrahim, S. H. Alsamhi, and A. Hawbani, "An improved YOLO-based road traffic monitoring system," *Computing*, vol. 103, no. 2, pp. 211–230, 2021.

[21] Y. Lei, D. Peng, P. Zhang, Q. Ke, and H. Li, "Hierarchical paired channel fusion network for street scene change detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 55–67, 2021.

[22] M. C. S. Santana, L. A. Passos, T. P. Moreira, D. Colombo, V. H. C. de Albuquerque, and J. P. Papa, "A novel Siamese-based approach for scene change detection with applications to obstructed routes in hazardous environments," *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 44–53, 2020.

[23] Chung-Lin Huang and Bing-Yao Liao, "A robust scene-change detection method for video segmentation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1281–1288, 2001.

WILEY | Hindawi

Research Article

A Multicriteria Selection Framework for Wireless Communication Infrastructure with Interval-Valued Pythagorean Fuzzy Assessment

Shanshan Qiu,[1] **Dan Fu,**[2] **and Xiaofang Deng** [1,3]

[1]*Software School, Jiangxi Normal University, Nanchang 330027, China*
[2]*Foreign Languages College, Jiangxi Normal University, Nanchang 330022, China*
[3]*School of Information and Communication Engineering, Guangzhou Maritime University, Guangzhou 510725, China*

Correspondence should be addressed to Xiaofang Deng; dxf@jxnu.edu.cn

In recent years, interval-valued Pythagorean fuzzy number is playing a more and more important role in decision management. It is a more effective and powerful tool to handle fuzzy information in decision problems. The multicriteria decision-making theory has been widely used in solving practical problems, such as the risk assessment of financial investment, engineering and construction, medical and health care, and information security. The main purpose of this paper is to apply a new interval-valued Pythagorean fuzzy decision-making method to practice and to analyze and solve the problem of wireless communication infrastructure. In this paper, a new interval-valued Pythagorean fuzzy ranking method, extending scope of application of the VIKOR method to interval-valued Pythagorean fuzzy set, is proposed. In order to adapt to actual needs, subjective and objective weights are combined to solve decision-making problems to enhance its practicality, validity, and effectiveness. An example of wireless communication infrastructure problem is provided to illustrate the rationality of this method and verify its advantages.

1. Introduction

In real life, people often cannot describe things accurately. Words like "almost," "probably," and other vague adjectives are often used to do the description. Therefore, fuzzy numbers are widely used, so they play an important role in decision-making. In 1965, Zadeh developed the mathematical expression of fuzzy things, called fuzzy set theory [1] (FS); it is widely used to deal with uncertain information in decision-making process. It is also an important method for fuzzy decision-making. FS expresses two dimensions: degree of certainty (degree of membership) and degree of uncertainty (degree of nonmembership). Yet, a neutral attitude that neither agrees nor opposes cannot be expressed by FS. Subsequently, Atanassov [2] extended fuzzy sets to intuitionistic fuzzy sets (IFS); he firstly took the hesitancy degree into consideration, which analyzes information from three dimensions: the degree of membership u, nonmembership v, and hesitancy π. IFS express support, opposition, and neu-

trality at the same time to overcome the disadvantages of single membership degree of fuzzy sets. Thus, the reliability of decision information is greatly improved. However, there still are cases that cannot be handled by using intuitionistic fuzzy sets, such as $u + v > 1$. To solve this problem, Yager [3] introduced the concept of Pythagorean fuzzy set, stipulating that the sum of the square of membership and nonmembership of Pythagorean fuzzy set is less than or equal to 1. For example, if we describe a problem as follows: the definite metrics (membership degree) is 0.6, indefinite metrics (nonmembership) is 0.8, because $0.6 + 0.8 > 1$, which is beyond the representation of intuitionistic fuzzy sets, but the problem can be well described by Pythagoras fuzzy set. Therefore, Pythagorean fuzzy set extended its expression range greatly on the basis of intuitionistic fuzzy set. It helped fuzzy decision-making to solve more practical problems and enhanced its practicability greatly. With the further study of practical problems, there is also uncertainty among membership index. In 2016, Zhang [4] raised the concept of interval-

valued Pythagorean fuzzy number. This theory followed the earlier concept of interval-valued intuitionistic fuzzy sets and extended Interval-valued Intuitive Fuzzy Set Theory. Besides, it enriched the previous Pythagorean fuzzy set decision method. Hereafter, Garg [5] extended Pythagorean fuzzy set theory to interval Pythagorean fuzzy environment and introduces some aggregation operators related to interval-valued Pythagorean fuzzy sets. Furthermore, Garg [6] extended the TOPSIS method and proposed a score function to determine the relative closeness coefficient in the interval Pythagorean fuzzy environment. Chen [7] proposed a novel function of relative closeness and uses the closeness-based assignment model to determine the ranking of alternatives. Garg [8] provided new Pythagorean fuzzy exponential aggregation operators and methods to solve MCDM problem under the Pythagorean fuzzy environment. A novel concept of linguistic interval-valued Pythagorean fuzzy set (LIVPFS) is presented in [9] to solve more MAGDM problems. More-over, Garg [10] developed a new neutral addition and scalar multiplication operations to solve the MAGDM problems. These methods make fuzzy decision-making more widely used. For example, when we say a watermelon is ripe, there is 60% ~ 80% probability that it is ripe and 40% ~ 50% probability that it is unripe. In this case, researchers use interval values to make a description, and the above situation can be easily expressed as $([0.6, 0.8], [0.4, 0.5])$.

In the process of decision-making, experts often need to choose the best one among candidates. Therefore, ranking method is very important in the decision-making process for multicriteria fuzzy decision-making. It seems that the previous Pythagorean fuzzy number ranking method is effective to solve the multicriteria decision-making problems. Yager [3] defined a set of Pythagorean natural order relationships to compare two fuzzy numbers through their membership and nonmembership. However, there are some Pythagorean fuzzy number pairs which cannot be compared by Yager [3]. To solve this problem, Zhang [11] came up with a strict order relation, which takes the hesitancy into consideration. However, for some Pythagorean fuzzy number pairs, the order relation cannot be clearly distinguished. Therefore, the researchers turned to ranking methods based on score functions and accuracy functions. Zhang and Xu [12] defined a score function for Pythagorean fuzzy numbers by using the square difference between membership and nonmembership. If the membership degree is the same as the nonmembership degree in a pair of Pythagorean fuzzy numbers, they will be considered as no difference by [12], which is inconsistent with the actual situation. Similarly, the ranking method based on score function and accuracy function is also proposed by Ren et al. [13], which is sensitive to disturbance, and even very small disturbance on membership degree could probably result in completely different ranking results. Subsequently, Ma and Xu [14] also proposed similar ranking methods with [13], but the ranking method lacks robustness, and the score function and accuracy function of ranking methods in literature [13] and literature [14] did not take the influence of hesitation into account. This may lead to information loss to some extent. To overcome the defect

in [13, 14], Peng and Dai [15] have defined a new ranking schema by incorporating hesitancy into the consideration of the score function. However, in solving practical problems, the result is different from that of other previous ranking methods. Zhang [4] proposed a new ranking algorithm based on relative closeness. Ranking score is obtained by calculating the distance between the positive and negative ideal solutions, but the ranking method in [4] cannot distinguish the two Pythagorean fuzzy numbers with equal membership degree and nonmembership degree. Considering that Zhang [16] gave an expression to describe the ratio index of Pythagorean fuzzy number, which can sort a set of Pythagorean fuzzy number pairs. However, this method depends too much on the set of Pythagorean fuzzy numbers. Once the number changed, it may lead to different results, lacking reliability and stability.

In view of multicriteria decision-making problem, researchers proposed many decision-making methods from different aspects. Since 1981, based on the thought of "the idea that the smaller the distance from positive ideal solution, or the larger the distance from the negative ideal solution, the better the fuzzy number it will be," Hwang and Yoon proposed the TOPSIS decision-making method [17]. Many researchers have studied and applied this method [12, 18–24]. To solve the problem of multicriteria decision-making with incomplete attribute information, the TODIM decision-making method of incomplete information was proposed in [25], considering that experts have reference dependence and loss avoidance behaviour. Later, many researchers have enriched the method [13, 26–28]. Taking the maximum group utility value and the minimum individual regret value into account, the VIKOR method was proposed by Opricovic, in 1998. It is one of the most effective tools in multicriteria decision evaluation and helpful to compare the relationship between individual and group. Over the past 20 years, VIKOR has been continuously extended and successfully applied to various fields [29–34]. This paper also cites VIKOR methods to determine objective weights in multicriteria decision-making problems.

The theory of multicriteria decision-making is widely used in solving practical problems, such as the risk assessment of financial investment [35–37], engineering and construction [38], medical and health care [39, 40], information security [20], and other areas [41, 42]. Based on the development of 5G technology, this paper did some research on the new situation and problems in the construction and protection of wireless communication infrastructure, focusing on the location of 5G base stations and aiming at the balance of coverage, capacity, cost, and quality of current wireless networks. The proposed options are evaluated by fuzzy decision-making method to provide a better scheme for experts. They are also conducive to the rapid promotion and the use of 5G technology.

The contributions of this paper can be illustrated as follows.

(1) By considering the relationship between membership degree, nonmembership degree, and hesitancy degree, this paper proposed a new ranking method

for interval-valued Pythagorean fuzzy number based on information reliability

(2) To solve the construction and protection of wireless communication infrastructure problem, this paper puts forward a decision-making schema which combines subjective and objective weights to reduce the strong influence of subjectivity on the rationality of decision-making

(3) The proposed method extends VIKOR method to interval-valued Pythagorean fuzzy environment. Based on this compromised VIKOR method, the proposed method makes an objective balance between the maximum group utility value and the minimum individual regret value according to the objective data and solves the 5G base station selection

The rest of this article is divided into six sections. Section 2 provides a brief introduction to fuzzy sets and some related theories. Section 3 analyzes and summarizes the previous ranking methods of Pythagorean fuzzy numbers and interval-valued Pythagorean fuzzy numbers. Section 4 introduces the interval-valued Pythagorean fuzzy number ranking method and the multicriteria decision method based on the ranking rule. Section 5 uses a real case to elaborate the working process of the proposed decision-making method. Section 6 compares the ranking method and decision method proposed in this paper with the previous research results in detail. Finally, in Section 7, we perform sensitivity analysis of the adjustment parameters involved in the proposed decision method.

2. Preliminaries

In this part, we will briefly introduce some basic definitions, properties, and distance formulas related to interval-valued Pythagorean fuzzy sets and then introduce an aggregation operator.

2.1. Pythagorean Fuzzy Set

Definition 1 (see [3]). Let X be a nonempty domain, and the Pythagorean fuzzy set P on X is represented as follows.

$$P = \{ \langle x, (\mu_P(x), \nu_P(x)) \rangle \mid x \in X \}, \quad (1)$$

where $\mu_P(x)$ and $\nu_P(x)$ are the membership and nonmembership functions of element x for set P, respectively, $\mu_P(x) \in [0, 1]$ and $\nu_P(x) \in [0, 1]$ and $0 \le (\mu_P(x))^2 + (\nu_P(x))^2 \le 1$.

The following distance evaluation of two Pythagorean fuzzy numbers can be obtained.

$$d(p_1, p_2) = \frac{1}{2} \left(\left| \left(\mu_{p_1}\right)^2 - \left(\mu_{p_2}\right)^2 \right| + \left| \left(\nu_{p_1}\right)^2 - \left(\nu_{p_2}\right)^2 \right| + \left| \left(\pi_{p_1}\right)^2 - \left(\pi_{p_2}\right)^2 \right| \right). \quad (2)$$

Based on the Pythagorean fuzzy set theory, the researchers make a thorough discussion on the interval-valued Pythagorean fuzzy set theory. The following is a brief introduction to the basic definition, properties, and distance formula of the interval-valued Pythagorean fuzzy set.

2.2. Interval-Valued Pythagorean Fuzzy Sets

Definition 2 (see [4]). Let the X be a nonnull domain, and the interval-valued Pythagorean fuzzy set on the X is $\tilde{P}$, represented as follows.

$$\tilde{P} = \left\{ <x, \left(\left[\mu_P^L(x), \mu_P^U(x) \right], \left[\nu_P^L(x), \nu_P^U(x) \right] \right) > | x \in X \right\}, \quad (3)$$

where $\left[\mu_P^L(x), \mu_P^U(x) \right]$ is the element x for the membership range of the set, $\left[\nu_P^L(x), \nu_P^U(x) \right]$ is the element x for the set of nonmembership value range, $0 \le \mu_P^L(x) < \mu_P^U(x) \le 1$ and $0 \le \nu_P^L(x) < \nu_P^U(x) \le 1$ and $0 \le (\mu_P^U(x))^2 + (\nu_P^U(x))^2 \le 1$.

The distance evaluation of two interval-valued Pythagorean fuzzy numbers can be formulized as follows.

$$d(p_1, p_2) = \frac{1}{4} \left(\left| \left(\mu_{p_1}^L\right)^P - \left(\mu_{p_2}^L\right)^P \right| + \left| \left(\mu_{p_1}^U\right)^P - \left(\mu_{p_2}^U\right)^P \right| \right.$$
$$+ \left| \left(\nu_{p_1}^L\right)^P - \left(\nu_{p_2}^L\right)^P \right| + \left| \left(\nu_{p_1}^U\right)^P - \left(\nu_{p_2}^U\right)^P \right|$$
$$+ \left. \left| \left(\pi_{p_1}^L\right)^P - \left(\pi_{p_2}^L\right)^P \right| + \left| \left(\pi_{p_1}^U\right)^P - \left(\pi_{p_2}^U\right)^P \right| \right). \quad (4)$$

2.3. Aggregation Operator.
Because of the complexity of decision-making problem, there are a lot of interference factors in the decision-making process when experts judge the superiority of the candidates. Therefore, the aggregation operator for interval-valued Pythagorean fuzzy set is also very important. Here, the traditional geometric operator will be employed by the proposed ranking method in the paper.

Definition 3 (see [43]). Let the $z_j = <[\mu_{jL}, \mu_{jU}], [\mu\nu_{jL}, \nu_{jU}] >$ $(j = 1, 2, \cdots, n)$ be the interval-valued Pythagorean fuzzy set. By weighted assembly operator, the matrix is D based on the optional scheme and expressed as follows.

$$g_n = g_\omega^{z_n}(z_1, z_2, \cdots, z_j)$$
$$= \left\langle \left[1 - \prod_{j=1}^m \left(1 - \mu_{jL}\right)^{\omega^j}, 1 - \prod_{j=1}^m \left(1 - \mu_{jU}\right)^{\omega^j} \right], \right.$$
$$\left. \cdot \left[\prod_{j=1}^m \nu_{jL}{}^{\omega^j}, \prod_{j=1}^m \nu_{jU}{}^{\omega^j} \right] \right\rangle. \quad (5)$$

3. Previous Ranking Methods

This section introduced some proposed ranking methods, which contains some natural sequence relationship, knowledge measurement, and information reliability and some other ranking methods of Pythagorean fuzzy numbers and interval-valued Pythagorean fuzzy numbers, and the shortcomings of these methods are analyzed as well.

3.1. Natural Sequence Relationship. Yager [1] defines a set of Pythagorean fuzzy number natural order relations. For a set of fuzzy number $P_i = (\mu_i, \nu_i)$, if $\mu_1 \geq \mu_2$, and $\nu_1 \leq \nu_2$, then $P_1 \geq P_2$, which means P_1 is bigger than or indifferent to P_2, we get the representation which P_1 takes precedence over the P_2. By comparing the membership degree of two fuzzy numbers and the nonmembership degree, this method obtains the size relation. According to the definition in [1], there are some Pythagorean fuzzy number pairs that cannot be compared.

For example, there are two Pythagorean fuzzy numbers P_1 and P_2, where $P_1 = (0.8, 0.4)$ and $P_2 = (0.7, 0.3)$. The comparison shows that $\mu_1 \geq \mu_2$ and $\nu_1 \geq \nu_2$, we can not make a judgment.

To make up for the defect of Yager's method in [1], Zhang [11] puts forward strict order relationship and considers the hesitation degree.

For a set of fuzzy numbers $P_i = (\mu_i, \nu_i)$, $\pi_i^2 = 1 - \mu_i^2 - \nu_i^2$, if $\mu_1 \geq \mu_2$, $\nu_1 \leq \nu_2$, and $\pi_1 \leq \pi_2$, and there is a $P_1 \geq P_2$, which means the priority of P_1 is higher than that of P_2. However, strict order relationships cannot distinguish certain Pythagorean fuzzy numbers.

For example, there are two Pythagorean fuzzy numbers P_1 and P_2, where $P_1 = (0.8, 0.4)$ and $P_2 = (0.7, 0.5)$. It can be seen that $\mu_1 \geq \mu_2$, $\nu_1 \leq \nu_2$, but $\pi_1 \geq \pi_2$, so it can not be judged by this method.

3.2. Ranking Methods for Pythagorean Fuzzy Numbers. Zhang and Xu [12] defined a score function that, let $P_1 = (\mu_1, \nu_1)$ and $P_2 = (\mu_2, \nu_2)$ be two Pythagorean fuzzy numbers, and the function they proposed as follows.

$$S_{\text{zhang}}(\tilde{p}) = \mu^2 - \nu^2, \tag{6}$$

The following ranking rules are defined.

$$\text{if } S_{\text{zhang}}(P_1) < S_{\text{zhang}}(P_2), \quad \text{then} \quad P_1 \prec P_2, \tag{7}$$

$$\text{if } S_{\text{zhang}}(P_1) = S_{\text{zhang}}(P_2), \quad \text{then} \quad P_1 \sim P_2, \tag{8}$$

where $\prec$ means smaller than and $\sim$ means indifferent to.

According to above methods, we find that the ranking method cannot be compared in some cases. For Pythagorean fuzzy numbers $P_1 = (0.8, 0.4)$ and $P_2 = (0.7, 0.1)$, the value can be computed as follows.

$$\begin{aligned} S_{\text{zhang}}(P_1) &= 0.8^2 - 0.4^2 = 0.48 \, ; \\ S_{\text{zhang}}(P_2) &= 0.7^2 - 0.1^2 = 0.48. \end{aligned} \tag{9}$$

Since $S_{\text{zhang}}(P_1) = S_{\text{zhang}}(P_2)$, we got $P_1 \sim P_2$. But it is obvious that the P_1 and P_2 are two different fuzzy numbers, so the judgment method is unreasonable.

To solve the above problems, the method [13] is motivated by [12], and it can be described as follows.

$$S_{\text{peng}}(\tilde{p}) = \mu^2 + \nu^2. \tag{10}$$

Moreover, following ranking rules are defined.

$$\begin{aligned} \text{if } S_{\text{zhang}}(P_1) &< S_{\text{zhang}}(P_2), \quad \text{then} \quad P_1 \prec P_2, \\ \text{if } S_{\text{zhang}}(P_1) &= S_{\text{zhang}}(P_2), \\ \text{if } S_{\text{peng}}(P_1) &< S_{\text{peng}}(P_2), \quad \text{then} \quad P_1 \prec P_2, \\ \text{if } S_{\text{peng}}(P_1) &= S_{\text{peng}}(P_2), \quad \text{then} \quad P_1 \sim P_2. \end{aligned} \tag{11}$$

The method [13] has improved and overcome the shortcomings of [12]; it seems more reasonable, but in fact, there are still some problems. For example, existed two Pythagorean fuzzy numbers P_1 and P_2, $P_1 = (0.8, 0.4)$, $P_2 = (0.7, 0.1)$, from formula (5), $S_{\text{zhang}}(P_1) = S_{\text{zhang}}(P_2)$, we cannot make the right judgment; then, we reuse formula (7) $S_{\text{peng}}(P_1) = 08^2 + 0.4^2 = 0.80 > S_{\text{peng}}(P_2) = 0.50 = 0.7^2 + 0.1^2$, that is to say $P_1 \succ P_2$. When we consider making a small change in P_1 as a P_1', and set P_1' to $P_1' = (0.7999, 0.4)$, and the difference between the original P_1 of the membership value is only 0.0001. This time we get $S_{\text{zhang}}(P_1') < S_{\text{zhang}}(P_2)$, which is contrary to the original result. Therefore, we concluded that this ranking method is less robust and sensitive to perturbations of small values, and the ranking method is unstable.

Zhang [4] focuses on distance measurement, and the ranking result is obtained according to the distance between Pythagorean number and ideal solution.

They defined $A(1, 0)$ and $B(0, 1)$ representation point is A a positive ideal solution point; the point is B a negative ideal solution point, and the distance between the positive and negative ideal solution is expressed as $d(P, A)$ and $d(P, B)$, respectively, and

$$d(P, A) = \frac{1}{2} \left(\left| (\mu_P)^2 - (\mu_A)^2 \right| + \left| (\nu_P)^2 - (\nu_A)^2 \right| + \left| (\pi_P)^2 - (\pi_A)^2 \right| \right). \tag{12}$$

There is a set of Pythagorean fuzzy numbers $\tilde{P} = (\tilde{\mu}, \tilde{\nu})$; by measuring the distance, Zhang [4] defines the following scoring functions.

$$c(\tilde{p}) = \frac{d(\tilde{P}, B)}{d(\tilde{P}, A) + d(\tilde{P}, B)} = \frac{1 - \nu^2}{2 - \mu^2 - \nu^2}. \tag{13}$$

The ranking method is described as follows.

$$\begin{aligned} \text{if } c(P_1) &< c(P_2), \quad \text{then} \quad P_1 \prec P_2, \\ \text{if } c(P_1) &= c(P_2), \quad \text{then} \quad P_1 \sim P_2. \end{aligned} \tag{14}$$

For example, there are two Pythagorean fuzzy numbers $P_1 = (0.7, 0.7)$ and $P_2 = (0.5, 0.5)$. The calculation shows that $c(P_1) = c(P_2) = 0.5$, then $P_1 \sim P_2$. It can be obviously concluded that P_1 and P_2 are two different fuzzy numbers, so the judgment method is unreasonable.

3.3. Knowledge Measurement and Information Reliability. The correlation ranking method about comprehensive knowledge measure and information reliability was proposed
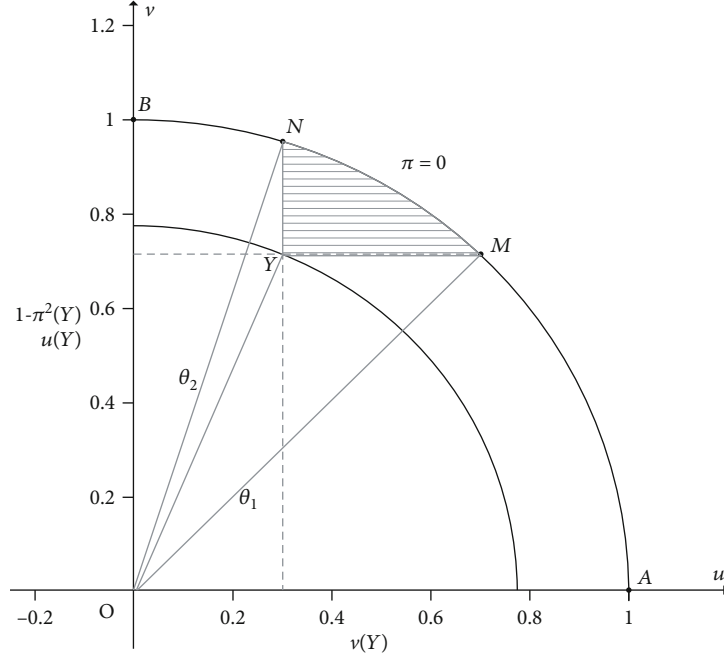
FIGURE 1: Schematic illustration of Pythagorean fuzzy number geometry.

in [44]. If there is a Pythagorean fuzzy number $Y(\mu, \nu, \pi)$, set a preference factor $\delta(0 \leq \delta < 1/2)$. Based on the theory of distance measurement from positive and negative ideal solutions, they introduce two Pythagorean fuzzy numbers $A(1, 0, 0)$ and $B(0, 1, 0)$ and use formulas $d(Y, P_{\text{near}})$ and $d(Y, P_{\text{far}})$ to express a closer or longer distance between Y and A and B.

The expression of the defined knowledge measure function are as follows:

$$K(Y) = \delta \left[1 - \frac{1}{2} \left(\frac{d(Y, P_{\text{near}})}{d(Y, P_{\text{far}})} + \pi^2 \right) \right] + (1 - \delta) d(Y, O), \tag{15}$$

where δ is a preference parameter.

Generally, if there is no extra preference relationship, $\delta = 1/2$ means that expert treats the two components equally.

As shown in Figure 1, the arc AB represents the point of $\mu^2 + \nu^2 = 1$, and the hesitation value of any point on the arc is 0. At this point, we call the Pythagorean fuzzy number 1. The point in the sector area OAB can be described as $0 < \mu^2 + \nu^2 < 1$. The Pythagorean fuzzy number information represented by these points is more or less hesitant. There is a Pythagorean fuzzy number Y located in the regional OAB. The projection of the arc AB in the μ direction and the ν direction is represented by points M and N, respectively. The shaded area MYN can be used to indicate the distance between the point Y, and the arc MN is an uncertain degree of information Y. The smaller the shadow MYN area, the more reliable the information represented by the Pythagorean fuzzy number Y [45].

$$S(Y) = 1 - S_{MYN} = 1 - (S_{MON} - S_{\Delta MOY} - S_{\Delta NOY}) = 1 - \frac{1}{2} \pi^2. \tag{16}$$

Definition 4 (see [45]). There is a set of Pythagorean fuzzy number $P_i = (\mu_i, \nu_i, \pi_i)(i = 1, 2)$. The ranking rules for knowledge measurement and information reliability are as follows.

$$\begin{aligned} &\text{if } K(P_1) < K(P_2), \quad \text{then} \quad P_1 \prec P_2, \\ &\quad\quad \text{if } K(P_1) = K(P_2), \\ &\text{if } S(P_1) < S(P_2), \quad \text{then} \quad P_1 \prec P_2, \\ &\text{if } S(P_1) = S(P_2), \quad \text{then} \quad P_1 \sim P_2. \end{aligned} \tag{17}$$

This method can calculate the Pythagorean correlation ranking results well and completely, but it is impossible to distinguish the interval-valued Pythagorean fuzzy numbers. Therefore, we extend this method to the interval-valued Pythagorean fuzzy environment to expand the scope of practical application.

3.4. Ranking Methods for Interval-Valued Pythagorean Fuzzy Numbers

3.4.1. Closeness Index-Based Score Function. In order to sort two interval-valued Pythagorean fuzzy numbers, the corresponding method was proposed in [4] which makes judgement base on the distance of positive and negative ideal solutions; the main ideas of this method are as follows.

To define an interval-valued Pythagorean fuzzy number $\beta = P([\mu_\beta^L, \mu_\beta^U], [\nu_\beta^L, \nu_\beta^U])$ and o^+, o^- represents its positive and negative ideal solution, which $o^+ = P([1, 1], [0, 0])$ and

$o^- = P([0,0],[1,1])$. Then, the function of β is defined as follows.

$$\zeta(\beta) = \frac{d(\beta, O^-)}{d(\beta, O^-) + d(\beta, O^+)}$$
$$= \frac{2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}{4 - \left(\mu_\beta^L\right)^2 - \left(\mu_\beta^U\right)^2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}. \quad (18)$$

The ranking rules can be described as follows.

$$\text{if } \zeta(P_1) < \zeta(P_2), \quad \text{then} \quad P_1 \prec P_2,$$
$$\text{if } \zeta(P_1) = \zeta(P_2), \quad \text{then} \quad P_1 \sim P_2. \quad (19)$$

Example 1. There are two interval-valued Pythagorean fuzzy numbers P_1 and P_2; $P_1 = ([0.5,0.6],[0.5,0.6])$, and $P_2 = ([0.4,0.7],[0.4,0.7])$. The calculation shows that $\zeta(P_1) = \zeta(P_2) = 0.5$; we obtained that $P_1 \sim P_2$. But, obviously, the fuzzy number P_1 is different from P_2, where the membership and nonmembership ranges are consistent; the Pythagorean fuzzy numbers cannot be compared.

And then, Garg [46] developed an improved score function for interval-valued Pythagorean fuzzy numbers, to define an interval-valued Pythagorean fuzzy number $P = ([\mu_P^L(x), \mu_P^U(x)], [v_P^L(x), v_P^U(x)])$, which is denoted by $P = ([a, b],[c, d])$. And they gave the score function of P as follows.

$$M(P) = \frac{\left(a^2 - c^2\right)\left(1 + \sqrt{1 - a^2 - c^2}\right) + \left(b^2 - d^2\right)\left(1 + \sqrt{1 - b^2 - d^2}\right)}{2}, \quad (20)$$

where $M(P) \in [-1, 1]$.

Based on the definition of $M(P)$, the ranking rules can be described as follows.

$$\text{if } M(P_1) < M(P_2), \quad \text{then} \quad P_1 \prec P_2,$$
$$\text{if } M(P_1) = M(P_2), \quad \text{then} \quad P_1 \sim P_2. \quad (21)$$

Example 2. There are two interval-valued Pythagorean fuzzy numbers P_1 and P_2; $P_1 = ([0,1],[0,0])$, and $P_2 = ([1,1],[0,0])$. The calculation shows that $M(P_1) = M(P_2) = 0$; we obtained $P_1 \sim P_2$. Obviously, P_1 and P_2 are different, but we have come to the conclusion that $P_1 = P_2$. Therefore, Garg's method cannot distinguish P_1 and P_2.

3.4.2. Novel Pearson-Like Correlation-Based TOPSIS Model. Similarly, the Pearson-based correlation measure method and the interval-valued Pythagorean correlation coefficient method were proposed in [47] to solve the ranking problem of interval-valued Pythagorean fuzzy number.

Let P_{i1} and P_{i2} be two values in the interval-valued Pythagorean fuzzy matrix D. The interval-valued Pythagorean fuzzy correlation coefficient of $\gamma^w g(P_{i1}, P_{i2})$ between P_{i1} and P_{i2} is defined as follows.

$$\gamma^w(P_{i1}, P_{i2}) = \frac{1}{6}\left(r_\mu^{w-}(P_{i1}, P_{i2}) + r_\mu^{w+}(P_{i1}, P_{i2}) + r_v^{w-}(P_{i1}, P_{i2}) \right.$$
$$\left. + r_v^{w+}(P_{i1}, P_{i2}) + r_\pi^{w-}(P_{i1}, P_{i2}) + r_\pi^{w+}(P_{i1}, P_{i2})\right), \quad (22)$$

which,

$$r_\mu^-(P_{i1}, P_{i2}) = \frac{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^-\right)^2\right)\cdot\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^-\right)^2\right)}{\sqrt{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^-\right)^2\right)^2}\cdot\sqrt{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^-\right)^2\right)^2}},$$

$$r_\mu^+(P_{i1}, P_{i2}) = \frac{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^+\right)^2\right)\cdot\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^+\right)^2\right)}{\sqrt{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^+\right)^2\right)^2}\cdot\sqrt{\sum_{j=1}^n w_j\left(\left(\mu_{i_ij}^-\right)^2 - \left(\bar{\mu}_j^+\right)^2\right)^2}}. \quad (23)$$

On the basis of the TOPSIS method and the Pearson correlation coefficient, the interval-valued Pythagorean correlation closeness index $\text{CI}(P)$ of interval-valued Pythagorean numbers can be defined as follows.

$$\text{CI}(P) = \eta\cdot\text{CC}_a^w(P_i) + (1 - \eta)\cdot\text{CC}_v^w(P_i), \quad (24)$$

where

$$\text{CC}_a^w(P_i) = \frac{1 - \gamma^w(P_i, P_\#)}{2 - \gamma^w(P_i, P_\#) - \gamma^w(P_i, P_*)},$$
$$\text{CC}_v^w(P_i) = \frac{1 + \gamma^w(P_i, P_\#)}{2 + \gamma^w(P_i, P_\#) + \gamma^w(P_i, P_*)}, \quad (25)$$

where P_i means interval-valued Pythagorean numbers of matrix D, P_* means an characteristics of positive ideal solution, and $P_\#$ means a characteristics of negative ideal solution.

The larger the CI's value, the better the fuzzy number. The $\text{CC}_a^w(P_i)$ and $\text{CC}_v^w(P_i)$ are the interval-valued Pythagorean correlation coefficients based on approximation and avoidance attitude, and the parameter η is the adjustment coefficient of approximation and avoidance, where $0 < \eta < 1$.

Example 3. For two sets of interval-valued Pythagorean fuzzy numbers P_1 and P_2, $p_1([0.2,0.3],[0.4,0.6])$ and $p_2([0.5,0.8],[0.4,0.6])$ are with same nonmembership degree. Our intuitive comparison of membership can be concluded quickly, but the results cannot be obtained by using the above method. Therefore, the method has some defects.

3.4.3. Accuracy Function and Distance Measures. In [5], Garg introduced an accuracy function of interval-valued Pythagorean fuzzy numbers. Let $P = ([\mu_P^L(x), \mu_P^U(x)], [v_P^L(x), v_P^U(x)])$ which is denoted by $P = ([a, b],[c, d])$. And there are three score functions as follows.

$$S(P) = \frac{a^2 + b^2 - c^2 - d^2}{2},$$

$$H(P) = \frac{a^2 + b^2 + c^2 + d^2}{2}, \tag{26}$$

$$M_g(P) = \frac{a^2 - \sqrt{1 - a^2 - c^2} + b^2 - \sqrt{1 - b^2 - d^2}}{2},$$

where $S(P) \in [-1, 1]$, $H(P) \in [-1, 1]$, and $M_g(P) \in [-1, 1]$.

Let P_1 and P_2 be any two interval-valued Pythagorean fuzzy numbers. Then,

$$
\begin{aligned}
&\text{if } S(P_1) < S(P_2), &&\text{then} && P_1 \prec P_2, \\
&\text{if } S(P_1) = S(P_2), &&\text{then} && P_1 \sim P_2, \\
&\text{if } H(P_1) < H(P_2), &&\text{then} && P_1 \prec P_2, \\
&\text{if } H(P_1) = H(P_2), &&\text{then} && P_1 \sim P_2, \\
&\text{if } M_g(P_1) < M_g(P_2), &&\text{then} && P_1 \prec P_2, \\
&\text{if } M_g(P_1) = M_g(P_2), &&\text{then} && P_1 \sim P_2.
\end{aligned}
\tag{27}
$$

Example 4. There are two interval-valued Pythagorean fuzzy numbers P_1 and P_2; $P_1 = ([0.4, 0.6], [0.6, 0.8])$, and $P_2 = ([0.5, \sqrt{0.27}], [\sqrt{0.27}, \sqrt{0.73}])$. The calculation shows that $S(P_1) = S(P_2) = -0.24$, $H(P_1) = H(P_2) = 0.76$, and $M_g(P_1) = M_g(P_2) \approx -0.0864$. We obtained $P_1 \sim P_2$. Obviously, P_1 and P_2 are different, but we have come to the conclusion that $P_1 = P_2$. Therefore, Garg's method cannot distinguish P_1 and P_2.

To overcome this problem, Garg [48] developed a different accuracy function. For an interval-valued Pythagorean fuzzy number $P = ([\mu_P^L(x), \mu_P^U(x)], [v_P^L(x), v_P^U(x)])$ which is also denoted by $P = ([a, b], [c, d])$, an improved accuracy function K is defined by

$$M_{\mathrm{Grag}}(P) = \frac{a^2 + b^2\sqrt{1 - a^2 - c^2} + b^2 + a^2\sqrt{1 - b^2 - d^2}}{2},$$

$$\tag{28}$$

where $M_{\mathrm{Grag}}(P) \in [-1, 1]$.

Based on the function, the ranking rules can be described as follows.

$$
\begin{aligned}
&\text{if } M_{\mathrm{Grag}}(P_1) < M_{\mathrm{Grag}}(P_2), &&\text{then} && P_1 \prec P_2, \\
&\text{if } M_{\mathrm{Grag}}(P_1) = M_{\mathrm{Grag}}(P_2), &&\text{then} && P_1 \sim P_2.
\end{aligned}
\tag{29}
$$

Example 5. There are two interval-valued Pythagorean fuzzy numbers P_1 and P_2; $P_1 = ([0, 1], [0, 0])$, and $P_2 = ([1, 1], [0, 0])$. The calculation shows that $M_{\mathrm{Grag}}(P_1) = M_{\mathrm{Grag}}(P_2) = 0$. We obtained $P_1 \sim P_2$. Obviously, P_1 and P_2 are different, but we have come to the conclusion that $P_1 = P_2$. Therefore, Garg's method cannot distinguish P_1 and P_2. So it still has some disadvantage of those accuracy functions.

Kumar et al. [49] put forward the interval-valued Pythagorean fuzzy ranking method about accuracy function and distance measure; the method can be illustrated as follows.

Set of interval-valued Pythagorean fuzzy number $p = ([\underline{\mu}_p(u), \bar{\mu}_p(u)], [\underline{v}_p(u), \bar{v}_p(u)])$, a transformation of the expression of P, recorded as $p = ([\underline{r}_p, \bar{r}_p], [\underline{d}_p, \bar{d}_p])$, of which $[\underline{r}_p, \bar{r}_p]$ represents the lower and upper strength of interval-valued Pythagorean fuzzy number; $[\underline{d}_p, \bar{d}_p]$ represents the lower and upper strength directions of interval-valued Pythagorean fuzzy numbers. The relationship between $p = ([\underline{\mu}_p(u), \bar{\mu}_p(u)], [\underline{v}_p(u), \bar{v}_p(u)])$ and $[\underline{r}_p, \bar{r}_p], [\underline{d}_p, \bar{d}_p]$ is as follows.

$$\underline{\mu}_p(u) = \underline{r}_p \cos(\underline{\theta}_p), \underline{v}_p(u) = \underline{r}_p \sin(\underline{\theta}_p), \tag{30}$$

$$\bar{\mu}_p(u) = \bar{r}_p \cos(\bar{\theta}_p), \bar{v}_p(u) = \bar{r}_p \sin(\bar{\theta}_p), \tag{31}$$

$$\underline{d}_p = \left(1 - \frac{2\underline{\theta}_p}{\pi}, 1 - \frac{2\bar{\theta}_p}{\pi}\right). \tag{32}$$

According to the above relations, in order to solve the problem of comparison between Pythagorean fuzzy numbers during the period, an extended accuracy function is proposed. For each Pythagorean fuzzy number $p = ([\underline{\mu}_p(u), \bar{\mu}_p(u)], [\underline{v}_p(u), \bar{v}_p(u)])$, the following accuracy functions are defined as follows.

$$T(p) = \frac{2(\underline{r}_p \underline{d}_p + \bar{r}_p \bar{d}_p) - (\underline{r}_p + \bar{r}_p) + 2}{4}, \tag{33}$$

where

$$
\begin{aligned}
\underline{r}_p &= \sqrt{\left(\underline{\mu}_p(u)\right)^2 + \left(\underline{v}_p(u)\right)^2}, \\
\bar{r}_p &= \sqrt{\left(\bar{\mu}_p(u)\right)^2 + \left(\bar{v}_p(u)\right)^2},
\end{aligned}
\tag{34}
$$

$$\underline{d}_p = \left(1 - \frac{2\underline{\theta}_p}{\pi}, 1 - \frac{2\bar{\theta}_p}{\pi}\right). \tag{35}$$

And define the comparison rules as follows.

$$
\begin{aligned}
&\text{if } T(p_1) < T(p_2) \text{ then } p_1 \prec p_2, \\
&\text{if } T(p_1) > T(p_2) \text{ then } p_1 \succ p_2, \\
&\text{if } T(p_1) = T(p_2) \text{ then } p_1 \sim p_2.
\end{aligned}
\tag{36}
$$

4. A New Ranking Method of Interval-Valued Pythagorean Fuzzy Numbers Based on Information Reliability

From what we have discussed in the previous section, there are still some problems with the ordering of two intervals of Pythagorean fuzzy numbers.

(1) If the membership and nonmembership of the interval Pythagorean are equal (equivalent to the Pythagorean fuzzy number), it may not be possible to distinguish between such two intervals with the previous method

(2) For two different intervals of Pythagorean fuzzy numbers, when we use previous method to compare the size, sometimes the two fuzzy numbers are the same, so the size cannot be compared

(3) The use of previous methods may sometimes go against popular perception and lead to wrong conclusions

In order to solve these problems, we propose a new interval Pythagorean fuzzy sorting function.

4.1. Ranking Function. In practical applications, the membership degree and nonmembership degree of a candidate may only be expressed by interval number, which cannot be determined as a certain number. Based on the above considerations, we extend the ranking method to interval-valued Pythagorean fuzzy numbers to expand the available range of the ranking method to solve more practical problems.

Based on the idea of the method proposed by the [50], we make the assumptions as follows.

Two sets of interval-valued Pythagorean fuzzy number $P = ([\mu_p^L, \mu_p^U], [\nu_p^L, \nu_p^U])$, $P' = ([\mu_{p'}^L, \mu_{p'}^U], [\nu_{p'}^L, \nu_{p'}^U])$, the value distribution is shown in Figure 2; the shadow $S1$ indicates the degree of information uncertainty in Pythagorean fuzzy P.

Inspired by the ranking thought of literature [50], an interval Pythagorean fuzzy number $P = \{[\mu_p^L, \mu_p^U], [\nu_p^L, \nu_p^U]\}$ can be expressed in the rectangular coordinate axis as following four points. If we express it in the right axis, we can get the following points $P_1(\mu_p^L, \nu_p^L)$, $P_2(\mu_p^L, \nu_p^U)$, $P_3(\mu_p^U, \nu_p^L)$, and $P_4(\mu_p^U, \nu_p^U)$.

We call P_4 the most credible point. The hesitancy degree of point P_4 can be expressed in the area $y_3 p_4 y_4$ [50] and recorded as S_1. Suppose there are two interval Pythagorean fuzzy numbers P_{A4} and P_{B4}, when these two points take the optimal values P_{A4} and P_{B4}, we can calculate their information reliability. The larger S_1 is, the more hesitant the interval-valued Pythagorean fuzzy number P is, and the less credible the interval-valued Pythagorean fuzzy number P is.

However, when two interval-valued Pythagorean fuzzy numbers have the same degree of hesitation, we may not be able to compare them. To achieve the preciseness of the ranking method, we propose to compare information reliability in μ direction and ν direction. Compared with the two points

$P_1(\mu_p^L, \nu_p^L)$ and $P_3(\mu_p^U, \nu_p^L)$, the μ values are taken as the two ends of the interval $[\mu_p^L, \mu_p^U]$, respectively. The degree of hesitation in point P_1 is expressed in area $S_{y_2 p_1 y_1}$, and the degree of hesitation in point P_3 is expressed in area $S_{y_3 p_3 y_1}$. Then, the difference between those two areas is the hesitation of interval-valued Pythagorean fuzzy numbers in the μ direction and recorded as S_2, which, $S_2 = S_{y_2 p_1 p_3 y_3} = S_{y_2 p_1 y_1} - S_{y_3 p_3 y_1}$, and when the S_2 gets bigger, the hesitation degree of P gets bigger, and the credibility degree gets lower.

Likewise, the interval-valued Pythagorean fuzzy number hesitation degree of P in the ν direction is recorded as S_3, which $S_3 = S_{y_4 p_1 p_2 y_1} = S_{y_2 p_1 y_1} - S_{y_2 p_2 y_4}$, and when the S_3 gets bigger, the hesitation degree of P gets bigger, and its credibility gets lower, and the less credible it is. Inspired by [49], S_1, S_2, and S_3 can be calculated as follows.

$$S1 = S_{y_3 p_4 y_4} = S_{y_3 o y_4} - S_{y_3 o p_4} - S_{y_4 o p_4} = \frac{1}{2}\pi_{p_4}^2, \qquad (37)$$

$$S_2 = S_{y_2 p_1 p_3 y_3} = S_{y_2 p_1 y_1} - S_{y_3 p_3 y_1}$$
$$= \left(S_{y_2 o y_1} - S_{y_2 o p_1} - S_{y_1 o p_1}\right) - \left(S_{y_3 o y_1} - S_{y_3 o p_1} - S_{y_3 o p_1}\right)$$
$$= \frac{1}{2}\pi_{p_1}^2 - \frac{1}{2}\pi_{p_3}^2, \qquad (38)$$

$$S_3 = S_{y_2 p_1 p_3 y_3} = S_{y_2 p_1 y_1} - S_{y_2 p_2 y_4}$$
$$= \left(S_{y_2 o y_1} - S_{y_2 o p_1} - S_{y_1 o p_1}\right) - \left(S_{y_2 o y_4} - S_{y_2 o p_2} - S_{y_4 o p_2}\right)$$
$$= \frac{1}{2}\pi_{p_1}^2 - \frac{1}{2}\pi_{p_2}^2. \qquad (39)$$

In order to compare interval-valued Pythagorean fuzzy numbers, we propose the following ranking methods.

$$\text{if } S_1(p_1) \le S_1(p_2) \quad \text{then} \quad P_1 \succeq P_2, \qquad (40)$$

$$\text{if } S_1(p_1) = S_1(p_2), \qquad (41)$$

$$\text{if } S_2(p_1) \le S_2(p_2) \text{then} P_1 \succeq P_2, \qquad (42)$$

$$\text{if } S_2(p_1) = S_2(p_2), \qquad (43)$$

$$\text{if } S_3(p_1) \le S_3(p_2) \text{then} P_1 \succeq P_2, \qquad (44)$$

$$\text{if } S_3(p_1) = S_3(p_2) \text{then} P_1 \sim P_2. \qquad (45)$$

Apply the new ranking method proposed in this paper to the previous methods described above, and we can get that.

In Example 1, $P_1 = ([0.5, 0.6], [0.5, 0.6])$, and $P_2 = ([0.4, 0.7], [0.4, 0.7])$, we get $S_1(P_1) = 0.14$, and $S_1(P_2) = 0.1$. The rank of those two interval-valued Pythagorean fuzzy number is $P_1 \prec P_2$.

In Examples 2 and 5, $P_1 = ([0, 1], [0, 0])$, and $P_2 = ([1, 1], [0, 0])$, we get $S_1(P_1) = S_1(P_2) = 0$, $S_2(P_1) = 0.5$, and $S_2(P_2) = 0$. The rank of those two interval-valued Pythagorean fuzzy number is $P_1 \prec P_2$.
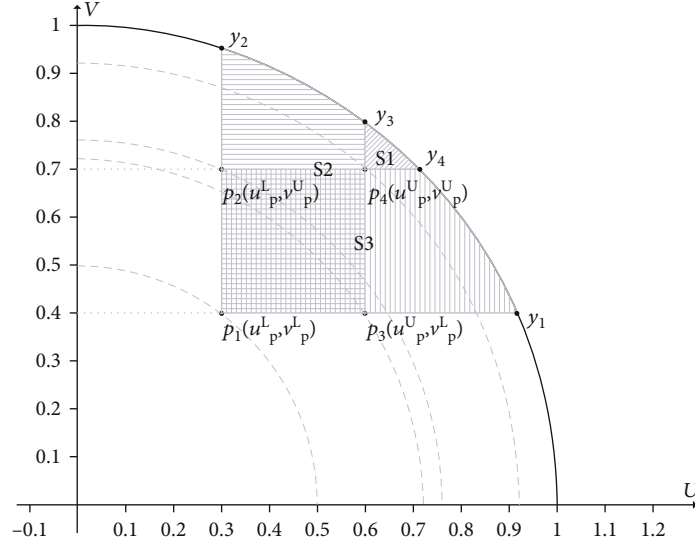
Figure 2: Schematic diagram of interval-valued Pythagorean fuzzy number.

In Example 3, $P_1 = ([0.2,0.3], [0.4,0.6])$, and $P_2 = ([0.5,0.8], [0.4,0.6])$, we get $S_1(P_1) = 0.275$, and $S_1(P_2) = 0$. The rank of those two interval-valued Pythagorean fuzzy number is $P_1 \prec P_2$.

In Example 4, $P_1 = ([0.4,0.6], [0.6,0.8])$, and $P_2 = (0.5, \sqrt{0.27}], [\sqrt{0.27}, \sqrt{0.73}])$, we get $S_1(P_1) = S_1(P_2) = 0$, $S_2(P_1) = 0.1$, and $S_2(P_2) = 0.01$. The rank of those two interval-valued Pythagorean fuzzy number is $P_1 \prec P_2$.

There are some unreasonable problems in the previous ranking methods. We list the defects of these ranking methods as Table 1.

Obviously, we can compare the two interval-valued Pythagorean fuzzy sets with the same membership degree and nonmembership degree. We can also compare the Pythagorean fuzzy numbers between two intervals with the same nonmembership degree. Therefore, the proposed method can solve some problems that cannot be compared before, and this has practical significance.

4.2. Extended VIKOR Decision-Making Methods. Because the research of Pythagorean fuzzy decision-making is still in its infancy and the extension of Pythagorean fuzzy set has the advantage of information representation, it is of practical significance to find a decision-making method for Pythagorean fuzzy set to solve practical problems. VIKOR (Vlse Kriterijuska Optimizacija I Komoromisno Resenje) is a multiattribute decision-making method proposed by Opricovic [51] in 1998. By calculating the positive and negative ideal solutions of each option and combining the preferences of experts, the alternatives are sorted according to the evaluation value of the scheme and the proximity of the ideal scheme. Based on the maximum group utility value and the minimum individual regret value, this study summarizes the current decision methods of Pythagorean fuzzy numbers. It is considered that the extension of multicriteria compromise solution scheduling method (VIKOR) to Pythagorean fuzzy environment will help to solve more practical problems. VIKOR method is a compromise method, which sorts

the finite decision scheme by maximizing group utility and minimizing individual regret.

4.2.1. Determination of Positive and Negative Ideal Solution. Set A, as a collection of m options, represents $A = \{A_1, A_2, \cdots, A_m\}$; C is a collection of n evaluation criteria, which represents $C = \{C_1, C_2, \cdots, C_n\}$. The evaluation criteria C is divided into two categories C^+ and C^-, where $C^+ \subset C, C^- \subset C, C^+ \cap C^- = \phi$, C^+ indicates that the criteria were benefit-oriented, and C^- indicates that the criteria is cost-based. To obtain a set of performance data $a_{ij} = ([\mu_{ij}^L, \mu_{ij}^U][v_{ij}^L, v_{ij}^U])$ $(i = 1, 2, \cdots, m; j = 1, 2, \cdots, n)$ for m options of n evaluation criteria in the form of interval-valued Pythagorean fuzzy numbers, of which $[\mu_{ij}^L, \mu_{ij}^U]$ is the membership value range, which $0 \le \mu_{ij}^L < \mu_{ij}^U \le 1$, $[v_{ij}^L, v_{ij}^U]$ is the value range of nonmembership, which $0 \le v_{ij}^L < v_{ij}^U \le 1 v_{ij}^{U2} + u_{ij}^{U2} \le 1$.

Generally, we can obtain an interval-valued Pythagorean decision matrix D as follows.

$$D = \left(a_{ij}\right)_{m \times n} = \begin{matrix} & C_1 \ C_2 \cdots C_n, \\ A_1 \\ A_1 \\ \vdots \\ A_1 \end{matrix} \begin{bmatrix} Z_{11} & Z_{21} & \cdots & Z_{n1} \\ Z_{21} & Z_{22} & \cdots & Z_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{bmatrix}. \tag{46}$$

As for a Pythagorean fuzzy set, the larger value of μ and the smaller value of v, the bigger the value is. Likewise, interval numbers $([1, 1], [0, 0])$ and numbers $([0, 0], [1, 1])$ can be regarded as positive and negative ideal solutions of interval-valued Pythagorean fuzzy sets. In terms of solving practical problems, it is sometimes unreasonable to simply set a fixed value as a positive and negative ideal solution. Therefore, we propose to calculate the corresponding positive and

TABLE 1: Shortcoming of previous ranking methods.

Method used	Score function	Shortcoming (some fuzzy numbers cannot be judged)
Yager [1]	If $\mu_1 \geq \mu_2$ and $v_1 \leq v_2$, then $P_1 \succeq P_2$	If $\mu_1 \geq \mu_2$ and $v_1 \geq v_2$
Zhang [11]	$\mu_1 \geq \mu_2$, $v_1 \leq v_2$, and $\pi_1 \leq \pi_2$, then $P_1 \succeq P_2$	If $\mu_1 \geq \mu_2$, $v_1 \leq v_2$, $\pi_1 \geq \pi_2$
Zhang and Xu [12]	$S_{\text{zhang}}(\tilde{p}) = \mu^2 - v^2$	e.g., $P_1 = (0.8, 0.4)$, and $P_2 = (0.7, 0.1)$
Peng [13]	$S_{\text{peng}}(\tilde{p}) = \mu^2 + v^2$	Robustness
Zhang [4]	$\zeta(\beta) = \dfrac{d(\beta, O^-)}{d(\beta, O^-) + d(\beta, O^+)} = \dfrac{2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}{4 - \left(\mu_\beta^L\right)^2 - \left(\mu_\beta^U\right)^2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}$	e.g., $P_1 = ([0.5, 0.6], [0.5, 0.6])$, and $P_2 = ([0.4, 0.7], [0.4, 0.7])$
Ho [47]	$\text{CI}(P) = \eta \bullet \text{CC}_a^w(P_i) + (1 - \eta) \bullet \text{CC}_v^w(P_i)$	e.g., $p_1([0.2, 0.3], [0.4, 0.6])$, and $p_2([0.5, 0.8], [0.4, 0.6])$
Garg [5]	$S(P) = \frac{a^2 + b^2 - c^2 - d^2}{2}$ $H(P) = \frac{a^2 + b^2 + c^2 + d^2}{2}$ $M_g(P) = \frac{a^2 - \sqrt{1 - a^2 - c^2} + b^2 - \sqrt{1 - b^2 - d^2}}{2}$	e.g., $P_1 = ([0.4, 0.6], [0.6, 0.8])$, and $P_2 = \left(\left[0.5, \sqrt{0.27}\right], \left[\sqrt{0.27}, \sqrt{0.73}\right]\right)$
Garg [48]	$M_{\text{Grag}}(P) = \dfrac{a^2 + b^2\sqrt{1 - a^2 - c^2} + b^2 + a^2\sqrt{1 - b^2 - d^2}}{2}$	e.g., $P_1 = ([0, 1], [0, 0])$, and $P_2 = ([1, 1], [0, 0])$
Garg [46]	$M(P) = \dfrac{\left(a^2 - c^2\right)\left(1 + \sqrt{1 - a^2 - c^2}\right) + \left(b^2 - d^2\right)\left(1 + \sqrt{1 - b^2 - d^2}\right)}{2}$	e.g., $P_1 = ([0, 1], [0, 0])$, and $P_2 = ([1, 1], [0, 0])$

negative ideal solutions for each criteria according to the actual standard data values.

Let the positive ideal solution be expressed as $Z^+ = \{[\mu_{Z^+}^L, \mu_{Z^+}^U], [v_{Z^+}^L, v_{Z^+}^U]\}$. The standard data column is represented as $C_n = \{Z_{1n}, Z_{2n}, \cdots, Z_{mn}\}$, where $Z_{mn} = \{[\mu_{mn}^L, \mu_{mn}^U], [v_{mn}^L, v_{mn}^U]\}$, According to the theory of positive and negative ideal solution, the maximum value of the upper and lower bounds of the membership degree in all schemes is taken as the upper and lower bounds of the membership degree of the positive ideal solution, and the minimum value of the upper and lower bounds of the nonmembership degree of the positive ideal solution is taken as the upper and lower bounds of the nonmembership degree of the positive ideal solution. In the same way, we hold the minimum of upper and lower bounds of membership as the upper and lower bounds of nonmembership degree of the positive ideal solution and hold the maximum of upper and lower bounds of nonmembership as the upper and lower bounds of nonmembership degree of the positive ideal solution. Z^+ and Z^- are defined as follows.

$$Z^+ = \left\{\left[\max \mu_{mn}^L, \max \mu_{mn}^U\right], \left[\min v_{mn}^L, \min v_{mn}^U\right]\right\}, \quad (47)$$

$$Z^- = \left\{\left[\min \mu_{mn}^L, \min \mu_{mn}^U\right], \left[\max v_{mn}^L, \max v_{mn}^U\right]\right\}. \quad (48)$$

4.2.2. Integrated Assessment Value. A conventional VIKOR method has the following rules.

$$S_i = \sum_{j=1}^{n} \frac{\omega_j \left(z_j^+ - z_{ij}\right)}{z_j^+ - z_j^-}, \quad (49)$$

$$R_i = \max_j \frac{\omega_j \left(z_j^+ - z_{ij}\right)}{z_j^+ - z_j^-}. \quad (50)$$

The value of S_i is the group benefit value, which is to ensure that all the indicators can be considered in the decision-making. The R_i is the largest individual regret value, which is to consider the impact of the largest deviation index on the scheme ranking.

Direct calculations of Pythagorean fuzzy numbers lose too much important raw information. We use the Hamming distance.

$$
\begin{aligned}
d(z_1, z_2) = \frac{1}{4} \Bigg(& \left|\left(\mu_{z_1}^L\right)^p - \left(\mu_{z_2}^L\right)^p\right| + \left|\left(\mu_{z_1}^U\right)^p - \left(\mu_{z_2}^U\right)^p\right| \\
& + \left|\left(v_{z_1}^L\right)^p - \left(v_{z_2}^L\right)^p\right| + \left|\left(v_{z_1}^U\right)^p - \left(v_{z_2}^U\right)^p\right| \\
& + \left|\left(\pi_{z_1}^L\right)^p - \left(\pi_{z_2}^L\right)^p\right| + \left|\left(\pi_{z_1}^U\right)^p - \left(\pi_{z_2}^U\right)^p\right| \Bigg).
\end{aligned}
$$
$$(51)$$

Among them, z_1 and z_2 are two interval-valued Pythagorean fuzzy numbers.

The improved S_i and R_i formulas are as follows.

$$S_i = \sum_{j=1}^{n} \frac{d\left(z_j^+, z_{ij}\right)}{d\left(z_j^+, z_j^-\right)}, \quad (52)$$

$$R_i = \max_j \frac{d\left(z_j^+, z_{ij}\right)}{d\left(z_j^+, z_j^-\right)}. \quad (53)$$
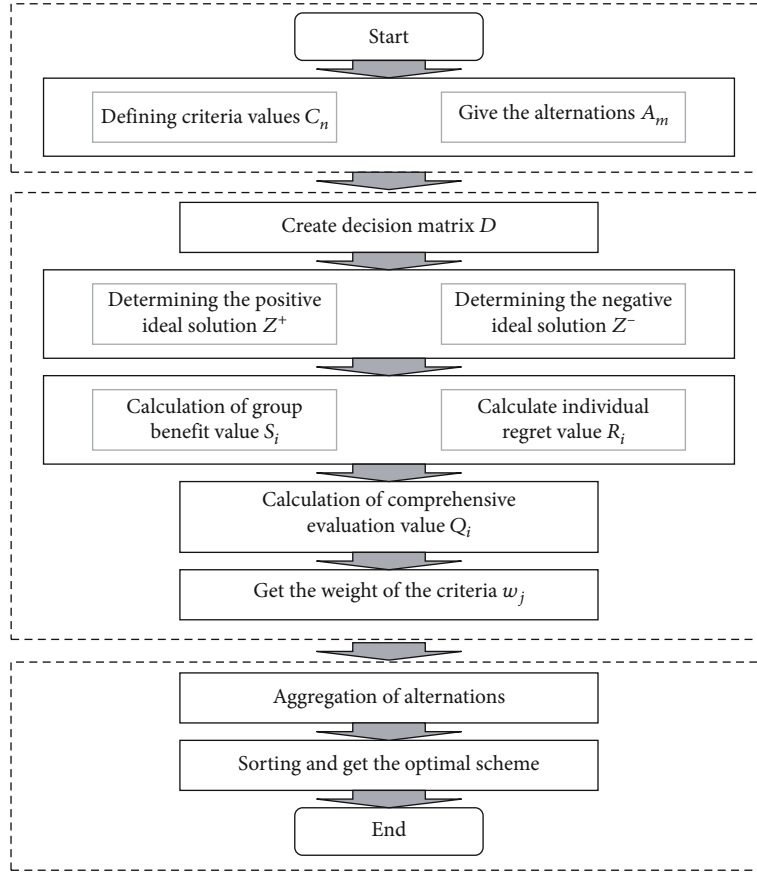
FIGURE 3: Working process of the proposed method.

To be able to take into account the value of group benefit and not ignore the maximum individual regret value, the following comprehensive evaluation methods for S_i and R_i are defined as follows.

$$Q_i = v \frac{S_i - S^-}{S^+ - S^-} + (1 - v) \frac{R_i - R^-}{R^+ - R^-}, \tag{54}$$

where v is the adjustment parameter of decision mechanism. When $v > 0.5$, the decision to maximize the group benefit is adopted; when $v < 0.5$, the decision to maximize the maximum regret of the individual is adopted, and when $v = 0.5$, the compromise decision strategy is adopted, which not only pays attention to the overall situation of the scheme but also takes into account the situation of a single multi-index. By doing this, it pursues the maximization of comprehensive benefits and the minimization of individual losses. In addition, $S^+ = \max_i S_i$, $S^- = \min_i S_i$, $R^+ = \max_i R_i$, and $R^- = \min_i R_i$.

4.2.3. Weight Determination Method.
Based on the consideration of practical problems and the relevance of practical problems, we adopt the method of integrating subjective and objective weight, set up a regulating parameter $\lambda (0 < \lambda < 1)$, and calculate the final weight of each evaluation criteria. The subjective weight given by each evaluation criteria is assumed to be $\omega_i^s = (\omega_1^s, \omega_2^s, \cdots, \omega_n^s)$, while the objective weight ω_i^o is based on data and calculated on the basis of comprehensive evaluation values as follows.

$$\omega_i^o = \frac{Qi}{\sum_{i=1}^n Qi}, \tag{55}$$

where Q_i is a comprehensive evaluation and can be taken to calculate the objective weight. So it can reflect real data information.

Which,

$$\omega_i^o = (\omega_1^o, \omega_2^o, \cdots, \omega_n^o). \tag{56}$$

Integrated,

$$\omega_i = \lambda \cdot \omega_i^0 + (1 - \lambda) \cdot \omega_i^s. \tag{57}$$

The λ is the adjustment parameter of subjective and objective weight. When >0.5, the objective weight is larger and more attention to objective data; when $\lambda < 0.5$, the subjective weight is larger and more attention to expert judgment. As $\lambda = 0.5$, the subjective and objective weight is equally important. ω_i is the final weight of each alternative criteria.

4.3. Multicriteria Decision-Making Method under Interval-Valued Pythagorean Fuzzy Environment.
This section extends VIKOR method and aggregation operator to propose a new multicriteria interval-valued Pythagorean fuzzy decision method to solve practical problems. The proposed solution is shown in Figure 3.
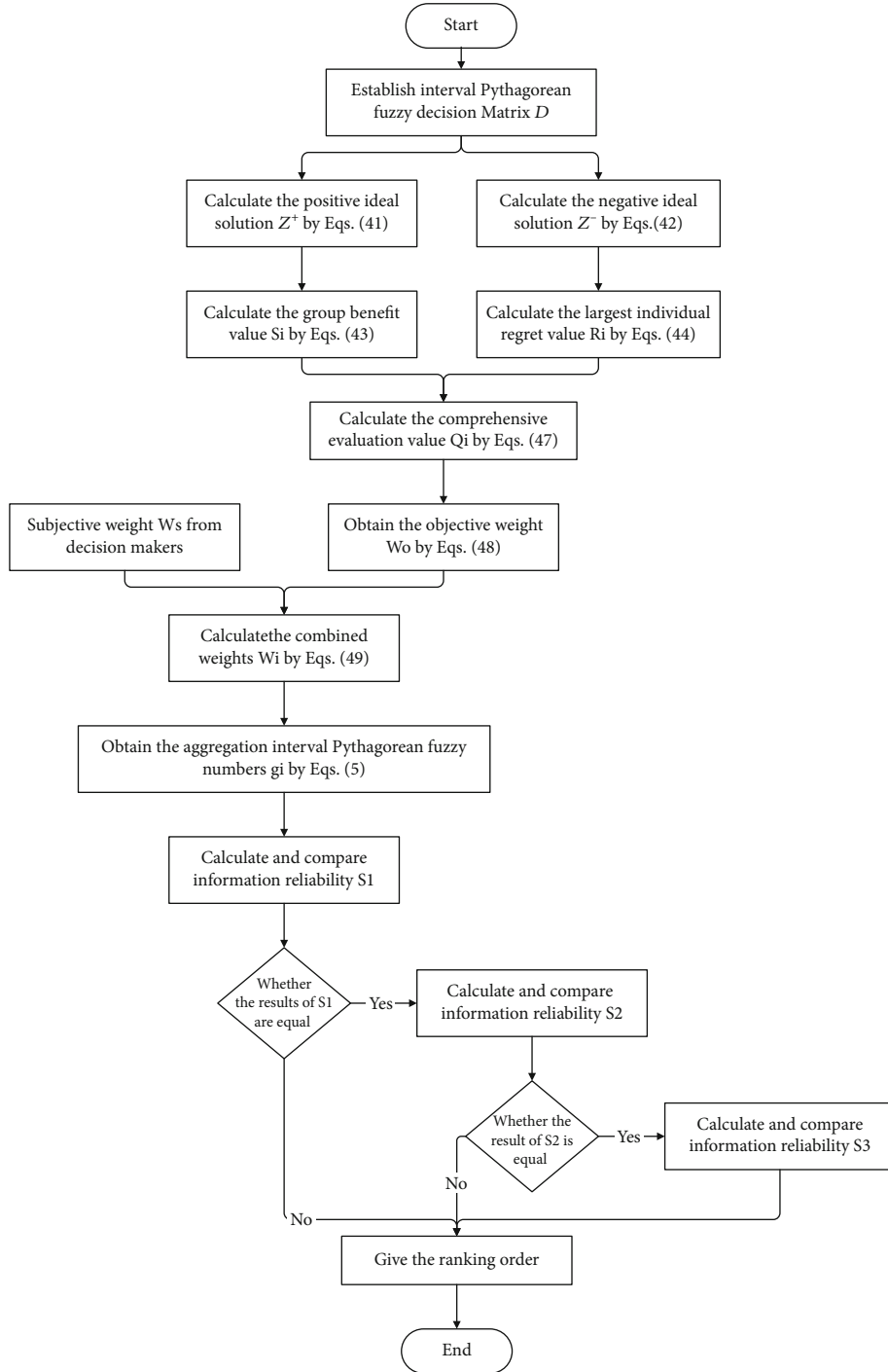
Figure 4: Flow chart of decision process.

A multicriteria decision-making method with interval-valued Pythagorean fuzzy information may be summarized as follows.

Step 1. Establish interval-valued Pythagorean fuzzy decision matrix $D(A_i * C_j)$.

Step 2. Calculate the positive ideal solution Z^+ by Eq. (47) and negative ideal solution Z^- by Eq. (48).

Step 3. Calculate the group benefit value S_i by Eq. (49) and the largest individual regret value R_i by Eq. (50) and make use of S_i and R_i to calculate the comprehensive evaluation value Q_i by Eq. (54).

Step 4. Obtain the objective weight ω_i^o by Eq. (55) and then combine it with the subjective weight ω_i^s from expert to get the combined weights ω_i by Eq. (57).

TABLE 2: Decision matrix.

	C1	C2	C3	C4
A1	([0.5,0.6], [0.2,0.3])	([0.6,0.7], [0.3,0.4])	([0.5,0.8], [0.5,0.6])	([0.3,0.6], [0.3,0.4])
A2	([0.4,0.8], [0.3,0.4])	([0.7,0.9], [0.2,0.3])	([0.6,0.8], [0.3,0.4])	([0.6,0.7], [0.2,0.3])
A3	([0.5,0.6], [0.1,0.3])	([0.6,0.7], [0.2,0.4])	([0.5,0.8], [0.4,0.6])	([0.3,0.6], [0.2,0.4])
A4	([0.3,0.7], [0.3,0.4])	([0.6,0.8], [0.2,0.3])	([0.5,0.7], [0.3,0.4])	([0.5,0.6], [0.2,0.3])

Step 5. Obtain the aggregation interval-valued Pythagorean fuzzy numbers g_j by Eq. (5).

Step 6. Calculate the information reliability S_1, S_2, and S_3 (Eqs. (37)–(39)) of alternatives A_i by using the interval-valued Pythagorean fuzzy number set g_j.

Step 7. The ranking method proposed in this paper is aimed at calculating the score function by Eq. (40) of each scheme, and the better scheme is obtained by ranking rules.

The flow chart of decision process is shown in Figure 4.

5. Application Example

In order to illustrate the feasibility of the proposed method in solving the multicriteria decision problem, this paper takes the construction of wireless communications infrastructure as an example to illustrate the decision analysis process and makes a comparison of this method with other methods.

Due to the rapid development of the Internet and the continuous upgrading of communication technology, we are ushered to the era of high-speed mobile communication network. The development of 5G technology can deal with the growth of more traffic mobile services in the future and massive new services as well as new devices. 5G technology has been studied further. The launch of 5G of mobile communication technology standard industry in 2016 accelerated the pace of 5G technology research and development. China regards information technology as the primary task of science and technology development and thus continuously increases 5G R&D investment. 5G technology satisfies the Internet and people's demand for communication stability. It has been officially commercialized. People enjoy the efficient and convenient life brought by high-speed network. When mobile phones are used to watch TV, video call, watch live broadcast, and on-line courses and do other activities, the efficiency of network transmission and the fluency of viewing has been significantly improved, which has brought better experience for users.

However, with the development of communication network and communication service, there are many new problems in the construction and protection of communication infrastructure. It is also difficult to choose the base station. The site selection unit needs to make a comprehensive programmer on the siting of base station and other relevant parameters on the basis of the environmental and business distribution characteristics of the covered area so as to balance the coverage, capacity, cost, and quality of wireless net-

TABLE 3: Distance matrix from the positive solution.

	C1	C2	C3	C4
A1	0.0941	0.1127	0.0834	0.0947
A2	0.0405	0.0000	0.0000	0.0000
A3	0.0990	0.1199	0.0661	0.0973
A4	0.0673	0.0757	0.0658	0.0602

work in this area. 5g development, base station first. 5G base station is not only an important part of the new infrastructure construction but also the foundation of 5 g technology to bring economic output. Therefore, it is urgent to solve this problem. The siting of 5G base station should be based on the total cost of its construction, the cost of signal coverage, system capacity, and network communication quality. Evaluation criteria should be established as $C_j (j = 1, 2, 3, 4)$. The experts have proposed four plans A_i $(i = 1, 2, 3, 4)$. Based on previous experience, network operators evaluate each scheme and make decisions matrix as shown in Table 2. Experts indicated that four programmers have its own strengths. And the corresponding weight is $\omega_i^s = \{0.25 ; 0.25 ; 0.25 ; 0.25\}$.

According to the expert evaluation and weight information, the expert calculates the scores of the four schemes and obtains the better scheme.

On the basis of the decision matrix, the positive ideal solution group of the criteria $Z_j^+ (j = 1, 2, 3, 4)$ and the negative ideal solution $Z_j^- (j = 1, 2, 3, 4)$ are calculated as follows.

$$Z_j^+ = \{([0.5,0.8], [0.1,0.3]), ([0.7,0.9], [0.2,0.3]),$$
$$([0.6,0.8], [0.3,0.4]), ([0.6,0.7], [0.2,0.3])\},$$

$$Z_j^- = \{([0.3,0.6], [0.3,0.4]), ([0.6,0.7], [0.3,0.4]),$$
$$([0.5,0.7], [0.5,0.6]), ([0.3,0.6], [0.3,0.4])\}. \tag{58}$$

According to formula (50), the distance between each evaluation value and positive ideal solution is calculated as the matrix composed of $d(z_j^+, z_{ij})$ shown in Table 3, and the distance between the positive and negative ideal solutions is obtained from the vector $D = [0.1011, 0.1127, 0.0823, 0.0947]$.

According to formula (52), the group benefit value is calculated as $S_i (i = 1, 2, 3, 4)$

$$S_1 = \sum_{j=1}^{n} \frac{d\left(z_j^+, z_{ij}\right)}{d\left(z_j^+, z_j^-\right)} = \frac{0.0941}{0.1011} + \frac{0.0405}{0.1011} + \frac{0.0990}{0.1011} + \frac{0.0673}{0.1011} = 2.9772,$$

$$S_2 = \frac{0.1127}{0.1127} + \frac{0.0000}{0.1127} + \frac{0.1199}{0.1127} + \frac{0.0757}{0.1127} = 2.7358,$$

$$S_3 = \frac{0.0834}{0.0823} + \frac{0.0000}{0.0823} + \frac{0.0661}{0.0823} + \frac{0.0658}{0.0823} = 2.6163,$$

$$S_4 = \frac{0.0947}{0.0947} + \frac{0.0000}{0.0947} + \frac{0.0973}{0.0947} + \frac{0.0602}{0.0947} = 2.6630.$$

$$(59)$$

According to formula (53), the matrix of individual regret value $R_i(i = 1, 2, 3, 4)$ is showed in Table 4.

Got that $R_1 = 0.9796$, $R_2 = 1.0644$, $R_3 = 1.0137$, and $R_4 = 1.0275$

According to formula (54), the comprehensive evaluation value of each criteria is calculated. The value of v here is $v = 0.5$, which indicates that a compromise decision strategy is adopted, and the overall situation of the scheme and the case of a single multi-index are taken into account. The distance formula in the example adopts Euclidean distance.

$$Q_1 = 0.5 * \frac{2.9772 - 2.6163}{2.9772 - 2.6163} + (1 - 0.5) * \frac{0.9796 - 0.9796}{1.0664 - 0.9796} = 0.5000,$$

$$Q_2 = 0.5 * \frac{2.7358 - 2.6163}{2.9772 - 2.6163} + (1 - 0.5) * \frac{1.0664 - 0.9796}{1.0664 - 0.9796} = 0.6655,$$

$$Q_3 = 0.5 * \frac{2.6163 - 2.6163}{2.9772 - 2.6163} + (1 - 0.5) * \frac{1.0137 - 0.9796}{1.0664 - 0.9796} = 0.2014,$$

$$Q_4 = 0.5 * \frac{2.6630 - 2.6163}{2.9772 - 2.6163} + (1 - 0.5) * \frac{1.0275 - 0.9796}{1.0664 - 0.9796} = 0.3471,$$

$$Q_i = \{0.5000 ; 0.6655 ; 0.2014 ; 0.3471\}, (i = 1, 2, 3, 4).$$

$$(60)$$

According to formula (55) and the comprehensive evaluation value, the objective weight of the alternative scheme is calculated.

$$\omega_1^o = \frac{0.5000}{0.5000 + 0.6655 + 0.2014 + 0.3471} = 0.2917,$$

$$\omega_2^o = \frac{0.6655}{0.5000 + 0.6655 + 0.2014 + 0.3471} = 0.3883,$$

$$\omega_3^o = \frac{0.2014}{0.5000 + 0.6655 + 0.2014 + 0.3471} = 0.1175,$$

$$\omega_4^o = \frac{0.3471}{0.5000 + 0.6655 + 0.2014 + 0.3471} = 0.2025,$$

$$\omega_i^o = \{0.2917 ; 0.3883 ; 0.1175 ; 0.2025\}, (i = 1, 2, 3, 4).$$

$$(61)$$

According to formula (57), the subjective and objective weight is considered synthetically. As for this case, the subjective and objective weight is equally important. The $\lambda = 0.5$, this means the adoption of a more compromise value combination.

Table 4: Total regret matrix.

	C1	C2	C3	C4
A1	0.9316	1.0000	1.0137	1.0000
A2	0.4004	0.0000	0.0000	0.0000
A3	0.9796	1.0644	0.8036	1.0275
A4	0.6656	0.6714	0.799	0.6355

$$\omega_i = \lambda \cdot \omega_i^0 + (1 - \lambda) \cdot \omega_i^s = 0.5 * \{0.2917 ; 0.3883 ; 0.1175 ; 0.2025\}$$
$$+ (1 - 0.5) * \{0.25 ; 0.25 ; 0.25 ; 0.25\}$$
$$= \{0.2709 ; 0.3191 ; 0.1838 ; 0.2263\}, (i = 1, 2, 3, 4).$$

$$(62)$$

Therefore, we can calculate that the combined weights of the four options are $\omega_1 = 0.2709$, $\omega_2 = 0.3191$, $\omega_3 = 0.1838$, and $\omega_4 = 0.2263$.

By formula (5), each scheme is assembled.

$$g_1 = \left\langle \left[1 - (1 - 0.5)^{0.3208} \prod_{j=1}^{m} \left(1 - \mu_{jL}\right)^{\omega^j}, 1 - \prod_{j=1}^{m} \left(1 - \mu_{jU}\right)^{\omega^j}\right], \right.$$
$$\left. \cdot \left[\prod_{j=1}^{m} \nu_{jL}^{\omega^j}, \prod_{j=1}^{m} \nu_{jU}^{\omega^j}\right] \right\rangle,$$

$$g_1 = \langle [0.4975, 0.6787], [0.2953, 0.3986] \rangle,$$

$$g_2 = \langle [0.5927, 0.8243], [0.2405, 0.3419] \rangle,$$

$$g_3 = \langle [0.4975, 0.6787], [0.1883, 0.3986] \rangle,$$

$$g_4 = \langle [0.4899, 0.7187], [0.2405, 0.3419] \rangle.$$

$$(63)$$

Finally, we calculate and analyze the score function of Pythagorean fuzzy number in the above four assembly intervals. According to formula (37), it can be calculated as follows.

$$S_{1(g_1)} = \frac{1}{2}\pi_{P_4}^2 = \frac{1}{2} * \left(\sqrt{1 - 0.6787^2 - 0.3986^2}\right)^2 = 0.3084,$$

$$S_{1(g_2)} = \frac{1}{2}\pi_{P_4}^2 = \frac{1}{2} * \left(\sqrt{1 - 0.8243^2 - 0.3419^2}\right)^2 = 0.2256,$$

$$S_{1(g_3)} = \frac{1}{2}\pi_{P_4}^2 = \frac{1}{2} * \left(\sqrt{1 - 0.6787^2 - 0.3986^2}\right)^2 = 0.3084,$$

$$S_{1(g_4)} = \frac{1}{2}\pi_{P_4}^2 = \frac{1}{2} * \left(\sqrt{1 - 0.7187^2 - 0.3419^2}\right)^2 = 0.3027.$$

$$(64)$$

According to the rules of judgment, the smaller S_1 is, the better the scheme. We conclude that $2 \succ 4 \succ 1 \sim 3$. We can see that the interval-valued Pythagorean fuzzy number $g_1([0.4975, 0.6787], [0.2953, 0.3986])$ and $g_2([0.5927, 0.8243], [0.2405, 0.3419])$ is different. Based on the above ranking rule, when $S_{1(g_1)} = S_{1(g_3)}$, comparative $S_{2(g_1)}$ and $S_{2(g_3)}$. Utilization $S_2 = (1/2)\pi_{P_1}^2 - (1/2)\pi_{P_3}^2$ (37).

TABLE 5: Results from various ranking methods.

Method used	Score function	Alternative	Score value	Ranking order of alternatives
Zhang [4]	$\zeta(\beta) = \dfrac{2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}{4 - \left(\mu_\beta^L\right)^2 - \left(\mu_\beta^U\right)^2 - \left(v_\beta^L\right)^2 - \left(v_\beta^U\right)^2}$	A1	0.5759	$2 \succ 4 \succ 3 \succ 1$
		A2	0.6532	
		A3	0.5830	
		A4	0.5948	
Garg [5]	$S(P) = \frac{a^2 + b^2 - c^2 - d^2}{2}$ $H(P) = \frac{a^2 + b^2 + c^2 + d^2}{2}$ $M_g(P) = \frac{a^2 - \sqrt{1 - a^2 - c^2} + b^2 - \sqrt{1 - b^2 - d^2}}{2}$	A1	0.4771	$2 \succ 1 \succ 4 \succ 3$
		A2	0.6027	
		A3	0.4512	
		A4	0.4656	
Garg [48]	$M_{\text{Grag}}(P) = \frac{a^2 + b^2\sqrt{1 - a^2 - c^2} + b^2 + a^2\sqrt{1 - b^2 - d^2}}{2}$	A1	0.9723	$2 \succ 4 \succ 3 \succ 1$
		A2	1.3712	
		A3	0.9795	
		A4	1.04560	
Grag [46]	$M(P) = \frac{\left(a^2 - c^2\right)\left(1 + \sqrt{1 - a^2 - c^2}\right) + \left(b^2 - d^2\right)\left(1 + \sqrt{1 - b^2 - d^2}\right)}{2}$	A1	0.3895	$2 \succ 4 \succ 3 \succ 1$
		A2	0.6677	
		A3	0.4397	
		A4	0.4882	
Ho [47]	$\text{CI}(P) = \eta \cdot \text{CC}_a^w(P_i) + (1 - \eta) \cdot \text{CC}_v^w(P_i)$	A1	0.4830	$4 \succ 3 \succ 2 \succ 1$
		A2	0.4850	
		A3	0.4923	
		A4	0.4992	
This paper proposed	$S1 = \frac{1}{2}\pi_{p_4}^2$ $S_2 = \frac{1}{2}\pi_{p_1}^2 - \frac{1}{2}\pi_{p_3}^2$ $S_3 = \frac{1}{2}\pi_{p_1}^2 - \frac{1}{2}\pi_{p_2}^2$	A1	S1 0.3084 S2 0.0716	$2 \succ 4 \succ 3 \succ 1$
		A2	0.2256 —	
		A3	0.3084 0.0685	
		A4	0.3027 —	

$$S_{2(g_1)} = \frac{1}{2} * \left(\sqrt{1 - 0.4975^2 - 0.2953^2}\right)^2 - \frac{1}{2} * \left(\sqrt{1 - 0.6787^2 - 0.2953^2}\right)^2 = 0.0716,$$
$$S_{2(g_3)} = \frac{1}{2} * \left(\sqrt{1 - 0.4975^2 - 0.1883^2}\right)^2 - \frac{1}{2} * \left(\sqrt{1 - 0.6787^2 - 0.1883^2}\right)^2 = 0.0685. \tag{65}$$

According to the judgment rule (40), the smaller the S_2, the better the scheme, and we conclude that $3 \succ 1$. To sum up, we can draw $2 \succ 4 \succ 3 \succ 1$.

6. Comparative Analysis

Various ranking methods are employed to rank the candidates in the selection of wireless communication base station, and the results are shown in Table 5.

If there is an interval-valued Pythagorean fuzzy number $p = ([a, b], [c, d])$ and it satisfies the requirements of $c = b$ and $d^2 = 1 - b^2$, then we can simplify the formulas (28)–(30) as follows:

$$S(P) = \frac{a^2 + b^2 - c^2 - d^2}{2} = \frac{a^2 + b^2 - 1}{2},$$

$$H(P) = \frac{a^2 + b^2 + c^2 + d^2}{2} = \frac{a^2 + b^2 + 1}{2},$$

$$M_g(P) = \frac{a^2 - \sqrt{1 - a^2 - c^2} + b^2 - \sqrt{1 - b^2 - d^2}}{2}$$
$$= \frac{a^2 + b^2 - \sqrt{1 - a^2 + b^2}}{2}. \tag{66}$$

There are two interval-valued Pythagorean fuzzy numbers, $p_1 = ([a_1, b_1], [c_1, d_1])$ and $p_2 = ([a_2, b_2], [c_2, d_2])$. Suppose that $a_1 + b_1 = a_2 + b_2 = X$, then there always $S(p_1) = S(p_2) = (X - 1)/2$, $H(p_1) = H(p_2) = (X + 1)/2$, and $M_g(p_1) = M_g(p_2) = (X + \sqrt{1 - X})/2$; therefore, we cannot judge the two interval-valued Pythagorean fuzzy numbers. There are some flaws in this method, so the result in method Garg [5] that $2 \succ 1 \succ 4 \succ 3$ is dubious. On the other hand, according to Table 2, the nonmembership of alternatives A2 and A4 are the same, and the membership of A2 is bigger than that of A4, so, $2 \succ 4$ is supposed to be the conclusion. But, the result in method Ho [47] shows us $4 \succ 3 \succ 2 \succ 1$; it is against the law $2 \succ 4$. So, the accuracy of method Ho [47] is also questionable.
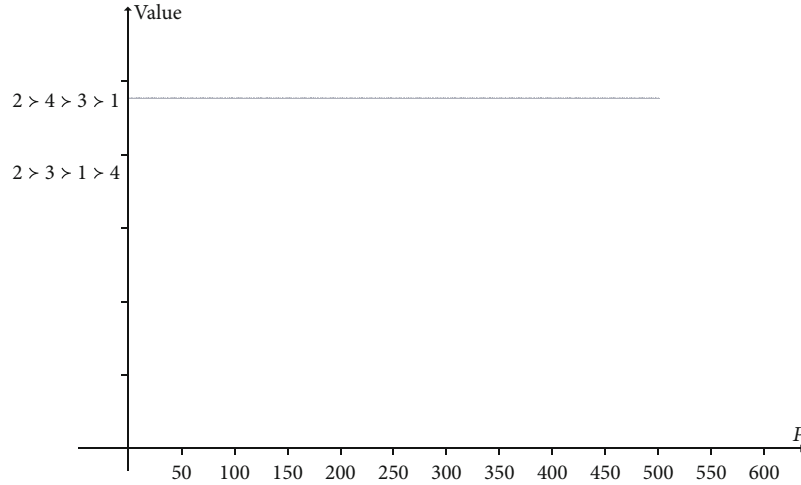
FIGURE 5: Influence of the change of p value.

However, it is easy to be seen from the results in Table 5; the alternatives of numerical example show that methods Zhang [4], Garg [46], and Garg [48] are consistent with the conclusions of our method, i.e., $2 \succ 4 \succ 3 \succ 1$, and that is in line with logic. The results show that the proposed method is reliable. To prove this, a comparison between the ranking methods from previous studies and the ranking method proposed in this paper reveals two advantages of the latter: (1) the way of determining weight in decision-making process is more reasonable, and (2) the proposed method is more robust.

Ho [47] defines the parameter η as the adjustment coefficient; it cannot adjust the subjective and objective weight of the decision data, but only determines the decision attitude of the whole scheme. However, when determining the objective weight, the method proposed in this paper based on the real data given by experts adopts the VIKOR method which considers the maximum group benefit and the minimum individual regret to obtain the comprehensive evaluation value of the alternative scheme. Then, the objective weight is determined, and the parameters are adjusted. The reference to historical experience is also important in solving practical problems because some experienced experts tend to value one or more evaluation criteria more. These more valuable criteria will be given greater weight. This paper combines the objective decision data given by experts, the experience of experts, the thoughts of supervisors, and the analysis of the specific problems so as to sort and select the alternatives more scientifically.

The distance formulas are used by the method in [47] and the proposed method in this paper to analyze and process the decision data. According to the Minkovsky distance formula used in this paper, the ranking results are very stable when the different values of p are taken. For the commonly used distance formula such as Manhattan distance and Euclidean distance, this method is robust. To test the stability of the distance formula used in this paper, we have tested the robustness of the value of p from 1 to 500. As shown in Figure 5, the experiment shows that when the value of p is 2 to 500,
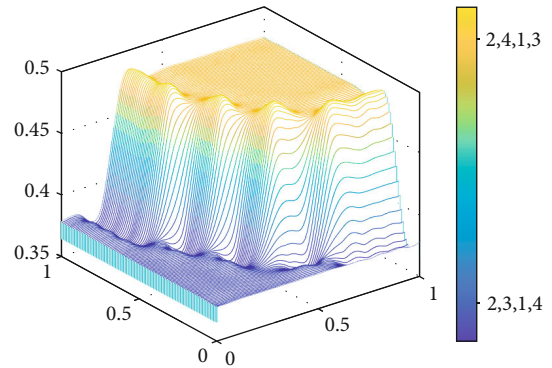
FIGURE 6: Waterfall of sensitivity analysis.
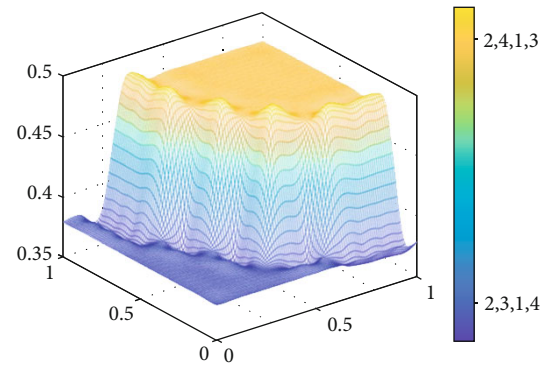
FIGURE 7: Contour lines of sensitivity analysis.

the ranking result is the same, and when the value of p is 1, the ranking result is not like the value we got previously, but, no matter what value the parameter v and λ is, it still gets the same result. So, the distance formula is very stable.

The ranking result of the alternative cannot be changed by the value of p. However, the distance formula used in [47] cannot obtain that result, so the method proposed by Ho [47] is robustness.

In the ranking method proposed in this paper, the interval-valued Pythagorean fuzzy number pair information
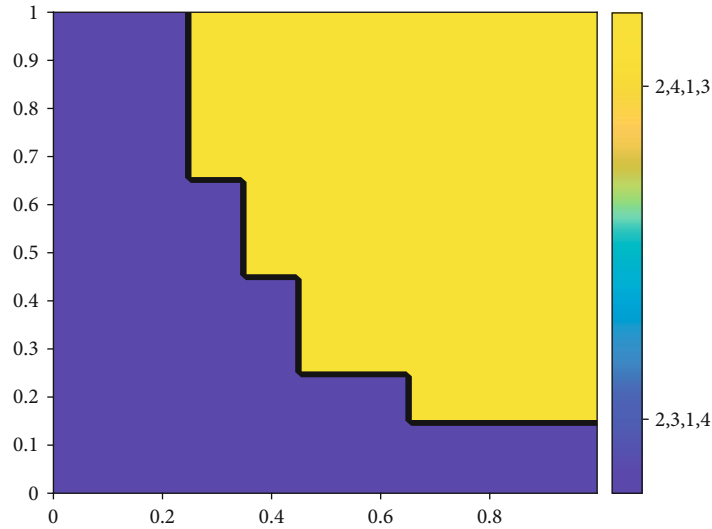
Figure 8: Thermal force diagram of sensitivity analysis.

aggregation is used in the last step, and the same type of interval-valued Pythagorean fuzzy number is finally obtained. The advantage of this method is that both the value range of membership degree and the value range of non-membership degree are considered, which can effectively prevent the loss of decision information and minimize the impact of information aggregation on ranking results. The decision data in Table 1 clearly shows that Scheme 2 is obviously superior to Scheme 4, and the conclusion of the decision method proposed in [47]; the literature is $4 \succ 3 \succ 2 \succ 1$, which is contrary to common sense judgment.

7. Sensitivity Analysis

A ranking method proposed in this paper sets the v and λ of adjusting parameters; parameter v is used to regulate VIKOR relationship between maximizing group benefit and minimizing individual regret. Parameter λ is used to adjust the relationship between objective weight and subjective weight. To achieve the convenience of studying the effect values of parameters v and λ on ranking results, we combine parameter v and parameter λ in the [0,1] interval and observe their influence on the ranking results of alternatives. Experimental results show that with the different values of parameters v and λ, the ranking results also change, and there are obvious laws.

In order to observe the influence of parameter changing on the ranking result of alternative scheme in a quick visual way, we use contour map, waterfall map, and thermodynamic map to express the ranking result.

As shown in Figures 6–8, the three graphs are contour plots, waterfall plots, and thermodynamic plots shown in the ranking results, respectively. It is easy to find that under the joint action of parameters v and λ, there are two main distributions of the result value of the alternative, namely, result 1 ($2 \succ 4 \succ 3 \succ 1$) and result 2 ($2 \succ 3 \succ 1 \succ 4$). The observation shows that as the values of the parameters v and λ are close to 1, the ranking is closer to the result 1. When the values of parameters v and λ are close to 0, the ranking is close to result

2. All the values of v and λ in the interval [0,1] obey this law. Hence, we believe that this result is valid; the ranking method is stable, and the parameters have an effective moderating effect on the ranking results. With the uncertainty of data values in practical problems, the transition position of parameters and the law of variation of ranking results may be slightly different.

With regard to the ranking method proposed in this paper, parameter v affects the weight of group benefit and individual regret preference, thus affecting the value of objective weight determined by VIKOR method. However, the parameter λ affects the preference of subjective and objective, and the two parameters act together to adjust the decision scheme in different aspects. Experts can make more reasonable and meaningful decisions by analysing the actual problems and the needs of the actual situation and adopting different parameter combinations.

8. Conclusions

This paper overcomes some ranking problem that other interval-valued Pythagorean fuzzy sets cannot solve. As a whole, this study focuses on a new ranking function that combines traditional assembly operators with extended VIKOR functions to solve practical problems. The proposed method can effectively obtain a more feasible and practical result and improve the robustness of result. After comparing the results from previous studies with the one proposed in this paper, it can be concluded that the decision-making method in this paper is more stable. According to the previous discussion, some superiorities of the proposed approach are as follows.

(1) A new interval-valued Pythagorean fuzzy decision method not only expands the application scope of fuzzy decision but also adjusts the subjective and objective weight proportion precisely, which is important to solve practical problems

(2) The ranking method proposed in this paper is very stable, and the ranking results will not change because of the change of the distance formula used. Therefore, the method is robust

(3) This paper stands on the shoulders of giants by inheriting and developing the study of previous research and apply the research result to practical usage

Although our proposed approach has played an important role in solving practical problems, this problem is still worthy of our further study. We look forward to proposing more practical decision methods and enriching fuzzy decision theory in the future. We will refer to the current research status of q-rung orthopair fuzzy decision-making and its application [52, 53] and consider expanding more ranking functions in q-rung orthopair fuzzy environment. Then, we integrate the extending function into the decision-making method of multidimensional preference linear programming to solve multicriteria decision-making problems. In the future, we will use this method to try to solve the problems of mobile transmission path selection and the selection of communication base station connection scheme in the field of wireless communication.

Data Availability

The example data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

We declare that there is no conflict of interest regarding the publication of this paper.

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[3] R. R. Yager, "Pythagorean membership grades in multicriteria decision making," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 4, pp. 958–965, 2014.

[4] X. Zhang, "Multicriteria Pythagorean fuzzy decision analysis: a hierarchical QUALIFLEX approach with the closeness index-based ranking methods," *Information Sciences*, vol. 330, pp. 104–124, 2016.

[5] H. Garg, "A novel accuracy function under interval-valued Pythagorean fuzzy environment for solving multicriteria decision making problem," *Journal of Intelligent & Fuzzy Systems*, vol. 31, no. 1, pp. 529–540, 2016.

[6] H. Garg, "A new improved score function of an interval-valued Pythagorean fuzzy set based TOPSIS method," *International Journal for Uncertainty Quantification*, vol. 7, no. 5, pp. 463–474, 2017.

[7] T.-Y. Chen, "An interval-valued Pythagorean fuzzy outranking method with a closeness-based assignment model for multiple criteria decision making," *International Journal of Intelligent Systems*, vol. 33, no. 1, pp. 126–168, 2018.

[8] H. Garg, "New exponential operational laws and their aggregation operators for interval-valued Pythagorean fuzzy multicriteria decision-making," *International Journal of Intelligent Systems*, vol. 33, no. 3, pp. 653–683, 2018.

[9] H. Garg, "Linguistic interval-valued Pythagorean fuzzy sets and their application to multiple attribute group decision-making process," *Cognitive Computation*, vol. 12, no. 6, pp. 1313–1337, 2020.

[10] H. Garg, "Neutrality operations-based Pythagorean fuzzy aggregation operators and its applications to multiple attribute group decision-making process," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 7, pp. 3021–3041, 2020.

[11] X. Zhang, "A novel approach based on similarity measure for Pythagorean fuzzy multiple criteria group decision making," *International Journal of Intelligent Systems*, vol. 31, no. 6, pp. 593–611, 2016.

[12] X. Zhang and Z. Xu, "Extension of TOPSIS to multiple criteria decision making with Pythagorean fuzzy sets," *International Journal of Intelligent Systems*, vol. 29, no. 12, pp. 1061–1078, 2014.

[13] P. Ren, Z. Xu, and X. Gou, "Pythagorean fuzzy TODIM approach to multi-criteria decision making," *Applied Soft Computing*, vol. 42, pp. 246–259, 2016.

[14] Z. Ma and Z. Xu, "Symmetric Pythagorean fuzzy weighted geometric/averaging operators and their application in multicriteria decision-making problems," *International Journal of Intelligent Systems*, vol. 31, no. 12, pp. 1198–1219, 2016.

[15] X. Peng and J. Dai, "Approaches to Pythagorean fuzzy stochastic multi-criteria decision making based on prospect theory and regret theory with new distance measure and score function," *International Journal of Intelligent Systems*, vol. 32, no. 11, pp. 1187–1214, 2017.

[16] X. Zhang, "Pythagorean fuzzy clustering analysis: a hierarchical clustering algorithm with the ratio index-based ranking methods," *International Journal of Intelligent Systems*, vol. 33, no. 9, pp. 1798–1822, 2018.

[17] C.-L. Hwang and K. Yoon, "Methods for multiple attribute decision making," in *Lecture Notes in Economics and Mathematical Systems, vol 186*, pp. 58–191, Springer, Berlin, Heidelberg, 1981.

[18] C.-T. Chen, "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.

[19] T. Chen and C. Tsao, "The interval-valued fuzzy TOPSIS method and experimental analysis," *Fuzzy Sets and Systems*, vol. 159, no. 11, pp. 1410–1428, 2008.

[20] M. F. Ak and M. Gul, "AHP–TOPSIS integration extended with Pythagorean fuzzy sets for information security risk analysis," *Complex & Intelligent Systems*, vol. 5, no. 2, pp. 113–126, 2019.

[21] A. Calik, "A novel Pythagorean fuzzy AHP and fuzzy TOPSIS methodology for green supplier selection in the Industry 4.0 era," *Soft Computing*, vol. 25, no. 3, pp. 2253–2265, 2021.

[22] H. Garg and K. Kumar, "A novel exponential distance and its based TOPSIS method for interval-valued intuitionistic fuzzy sets using connection number of SPA theory," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 595–624, 2020.

[23] H. Garg, A. Keikha, and H. Mishmast Nehi, "Multiple-attribute decision-making problem using TOPSIS and Choquet integral with hesitant fuzzy number information,"

Mathematical Problems in Engineering, vol. 2020, Article ID 9874951, 12 pages, 2020.

[24] H. Garg and R. Arora, "TOPSIS method based on correlation coefficient for solving decision-making problems with intuitionistic fuzzy soft set information," *AIMS Mathematics*, vol. 5, no. 4, pp. 2944–2966, 2020.

[25] L. F. A. M. Gomes and M. M. P. P. Lima, "TODIM: basics and application to multicriteria ranking of projects with environmental impacts," *Foundations of Computing and Decision Sciences*, vol. 16, no. 4, pp. 113–127, 1992.

[26] C. Wei, Z. Ren, and R. M. Rodríguez, "A hesitant fuzzy linguistic TODIM method based on a score function," *International Journal of Computational Intelligence Systems*, vol. 8, no. 4, pp. 701–712, 2015.

[27] D. K. Sen, S. Datta, and S. S. Mahapatra, "Extension of TODIM for decision making in fuzzy environment: a case empirical research on selection of industrial robot," *International Journal of Services and Operations Management*, vol. 26, no. 2, pp. 238–276, 2017.

[28] A. Biswas and B. Sarkar, "Interval-valued Pythagorean fuzzy TODIM approach through point operator-based similarity measures for multicriteria group decision making," *Kybernetes*, vol. 48, no. 3, pp. 496–519, 2019.

[29] M. Gul, E. Celik, N. Aydin, A. Taskin Gumus, and A. F. Guneri, "A state of the art literature review of VIKOR and its fuzzy extensions on applications," *Applied Soft Computing*, vol. 46, pp. 60–89, 2016.

[30] P. Gupta, M. K. Mehlawat, and N. Grover, "Intuitionistic fuzzy multi-attribute group decision-making with an application to plant location selection based on a new extended VIKOR method," *Information Sciences*, vol. 370-371, pp. 184–203, 2016.

[31] A. Hafezalkotob and A. Hafezalkotob, "Interval target-based VIKOR method supported on interval distance and preference degree for machine selection," *Engineering Applications of Artificial Intelligence*, vol. 2017, no. 57, pp. 184–196, 2017.

[32] Z. Hao, Z. Xu, H. Zhao, and R. Zhang, "Novel intuitionistic fuzzy decision making models in the framework of decision field theory," *Information Fusion*, vol. 33, pp. 57–70, 2017.

[33] T.-Y. Chen, "Remoteness index-based Pythagorean fuzzy VIKOR methods with a generalized distance measure for multiple criteria decision analysis," *Information Fusion*, vol. 41, pp. 129–150, 2018.

[34] H. Garg, J. Gwak, T. Mahmood, and Z. Ali, "Power aggregation operators and VIKOR methods for complex q-rung orthopair fuzzy sets and their applications," *Mathematics*, vol. 8, no. 4, p. 538, 2020.

[35] T.-Y. Chen, "A novel risk evaluation method of technological innovation using an inferior ratio-based assignment model in the face of complex uncertainty," *Expert Systems with Applications*, vol. 95, pp. 333–350, 2018.

[36] T.-Y. Chen, "An outranking approach using a risk attitudinal assignment model involving Pythagorean fuzzy information and its application to financial decision making," *Applied Soft Computing*, vol. 71, pp. 460–487, 2018.

[37] D. N. le, G. Nhu Nguyen, H. Garg, Q. T. Huynh, T. Ngoc Bao, and N. Ngoc Tuan, "Optimizing bidders selection of multi-round procurement problem in software project management using parallel max-min ant system algorithm," *Computers Materials & Continua*, vol. 66, no. 1, pp. 993–1010, 2020.

[38] T.-Y. Chen, "An interval-valued Pythagorean fuzzy compromise approach with correlation- based closeness indices for multiple-criteria decision analysis of bridge construction methods," *Complexity*, vol. 2018, Article ID 6463039, 29 pages, 2018.

[39] T.-Y. Chen, "A novel VIKOR method with an application to multiple criteria decision analysis for hospital-based post-acute care within a highly complex uncertain environment," *Neural Computing & Applications*, vol. 31, no. 8, pp. 3969–3999, 2019.

[40] H. Garg, G. Shahzadi, and M. Akram, "Decision-making analysis based on Fermatean fuzzy Yager aggregation operators with application in COVID-19 testing facility," *Mathematical Problems in Engineering*, vol. 2020, Article ID 7279027, 16 pages, 2020.

[41] H. Garg, "Guest Editorial: Neutrosophic decision making and applications in knowledge management," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, p. 67, 2020.

[42] S. K. Giri, T. Garai, H. Garg, and S. Islam, "Possibilistic mean of generalized non-linear intuitionistic fuzzy number to solve a price and quality dependent demand multi-item inventory model," *Computational and Applied Mathematics*, vol. 40, no. 4, pp. 1–24, 2021.

[43] X. U. Ze-shui, "Methods for aggregating interval-valued intuitionistic fuzzy information and their application to decision making," *Control and Decision*, vol. 2, pp. 215–219, 2007.

[44] K. Rahman, "Approaches to some induced Einstein geometric aggregation operators based on interval-valued Pythagorean fuzzy numbers and their application," *New Mathematics and Natural Computation*, vol. 16, no. 2, pp. 211–230, 2020.

[45] S.-P. Wan, Z. Jin, F. Wang, and Z. Jin, "A new ranking method for Pythagorean fuzzy numbers," in *2017 12th International Conference on Intelligent Systems And Knowledge Engineering*, pp. 1–6, Nanjing, China, November 2017.

[46] H. Garg, "A linear programming method based on an improved score function for interval-valued Pythagorean fuzzy numbers and its application to decision-making," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 26, no. 1, pp. 67–80, 2018.

[47] L.-H. Ho, Y.-L. Lin, and T.-Y. Chen, "A Pearson-like correlation-based TOPSIS method with interval-valued Pythagorean fuzzy uncertainty and its application to multiple criteria decision analysis of stroke rehabilitation treatments," *Neural Computing & Applications*, vol. 32, no. 12, pp. 8265–8295, 2020.

[48] H. Garg, "A novel improved accuracy function for interval valued Pythagorean fuzzy sets and its applications in the decision-making process," *International Journal of Intelligent Systems*, vol. 32, no. 12, pp. 1247–1260, 2017.

[49] T. Kumar, R. Bajaj, and M. D. Ansari, "On accuracy function and distance measures of interval-valued Pythagorean fuzzy sets with application to decision making," *Scientia Iranica*, vol. 27, no. 4, pp. 2127–2139, 2020.

[50] S.-P. Wan, Z. Jin, and J.-Y. Dong, "A new order relation for Pythagorean fuzzy numbers and application to multi-attribute group decision making," *Knowledge and Information Systems*, vol. 62, no. 2, pp. 751–785, 2020.

[51] S. Opricovic, "Multicriteria optimization of civil engineering systems," *Faculty of Civil Engineering, Belgrade*, vol. 2, no. 1, pp. 5–21, 1998.

[52] H. Garg, "CN-q-ROFS: connection number-basedq-rung orthopair fuzzy set and their application to decision-making process," *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3106–3143, 2021.

[53] D. Rani and H. Garg, "Complex intuitionistic fuzzy preference relations and their applications in individual and group decision-making problems," *International Journal of Intelligent Systems*, vol. 36, no. 4, pp. 1800–1830, 2021.

WILEY | Hindawi

Research Article

An Adaptive Video Transmission Mechanism over MEC-Based Content-Centric Networks

Longzhe Han ,[1] Jia Zhao ,[1] Xuecai Bao ,[1] Guangming Liu ,[1] Yan Liu ,[2] and Taras Maksymyuk [3]

[1]*Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang Institute of Technology, Nanchang 330099, China*
[2]*College of Computer and Software Engineering, East China Normal University, Shanghai 200062, China*
[3]*Department of Telecommunications, Lviv Polytechnic National University, Lviv 79013, Ukraine*

Correspondence should be addressed to Longzhe Han; longzhehan@gmail.com
and Taras Maksymyuk; taras.maksymyuk@gmail.com

The rapid growth of video traffic poses serious challenges to the current Internet. Content-Centric Networking (CCN) as a promising candidate has been proposed to reengineer the Internet architecture. The in-network caching and named content communication model of CCN can enhance the video streaming applications and reduce the network workload. Due to the bandwidth-consuming characteristic of video streaming, the aggressive transmission of video data will cause a reduction of overall network efficiency. In this paper, we present an adaptive video transmission mechanism over Mobile Edge Computing- (MEC-) based CCN. The computation and storage resources of the MEC server are utilized to facilitate the video delivery. Our mechanism adopts a scalable video coding scheme to adaptively control transmission rate to cope with the network condition variation. To analyse the equilibrium property of the proposed mechanism, an analytical model is deduced by using network utility function and convex programming. We also take into account the packet loss in wired and wireless links and present a MEC assistant loss recovery algorithm. The experiment results demonstrate the performance improvement of our proposed mechanism.

1. Introduction

With the continuous progress of wireless communication technology and mobile devices, emerging multimedia services (e.g., mobile TV, user-created video, video game, and mobile video calling) have gradually become people's daily applications, and dominated mobile Internet traffic [1–3]. According to Cisco's report, 79 percent of global mobile IP traffic will be video by 2022. The explosive growth of video traffic will bring huge pressure to the mobile network operators [4]. Although several technical solutions (e.g., peer-to-peer network and content delivery network) have been carried out to relieve the network transmission burden, the problem cannot be solved completely [5]. The current Internet protocols, TCP/IP, were designed in the 1960s. The principle of TCP/IP is to interconnect two hosts across multiple physical links and exchange data. At that time, the main applications were for resource sharing, such as FTP and Telnet. The evolution of Internet applications makes TCP/IP inefficient for the new requirement, which is content sharing [6].

To address the shift of Internet use, Content-Centric Networking (CCN) has been proposed as a new network architecture of the Internet [7]. The major difference between TCP/IP and CCN is changing the protocol core from IP to named content. Instead of connecting two end hosts, the key design goal of CCN is to deliver the requested content to the user [8]. In CCN, every node is equipped with a data cache. The user announces the name of interested
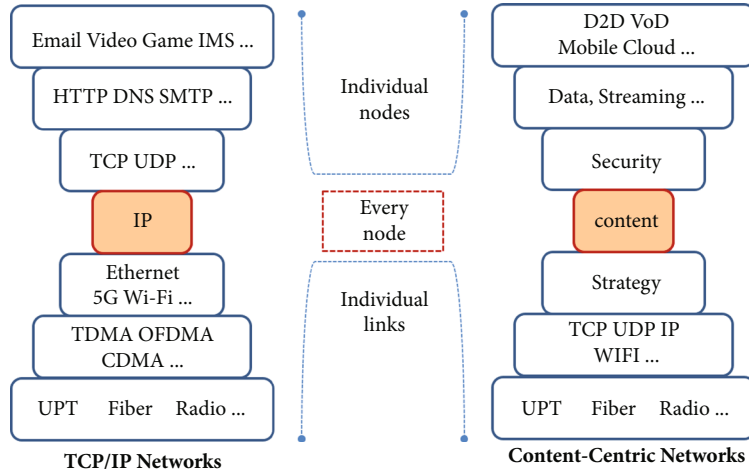
FIGURE 1: The comparison of TCP/IP protocols with CCN protocols.

information to the network, and any node possessing the data can respond [9]. Focusing on content sharing, CCN is able to dramatically improve the network transmission efficiency. Due to its promising characteristics, plentiful research efforts have been dedicated to video streaming over CCN. In [10–12], adaptive video streaming schemes in a wireless network environment are presented. The management of the data cache directly affects the availability of content; efficient cache management algorithms have been proposed [13–15].

Another crucial technique to cope with the challenge of the Internet is Mobile Edge Computing (MEC). In contrast to centralized cloud computing, MEC adopts a distributed architecture. To improve the data transmission rate, reduce network latency, and the workload of backhaul, MEC migrates the computing and storage capability of the central cloud to the edge of the network. By deployment of resource close to the end user, MEC can provide location-aware and high-speed data services. In [16, 17], algorithms are proposed to dynamically adjust the quality of experience for video streaming application in MEC. To utilize the computing resource of the MEC server, adaptive bitrate streaming approaches are presented in [18–20]. By the estimation of wireless channel and assistance of the MEC server, video quality is adapted to the wireless channel variations.

In this paper, we present an adaptive video transmission mechanism, which considering the advantages of both CCN and MEC. The main contributions of the proposed mechanism are as follows. To explore the in-network caching capability and coexist with different network applications, we introduce a transmission control algorithm to dynamically regulate the transmission rate. To adapt the network condition variation, the scalable video coding (SVC) technique is used. The encoded video has a layered structure, and various video qualities (e.g., frame rate, resolution, and fidelity) can be provided by selectively extracting enhancement layers. An analytical model is presented to verify the proposed transmission control algorithm. In addition, we identify the packet loss in wired and wireless links and propose a recovery algorithm for overcoming wireless link error.

The remainder of this paper is organized as follows. Section 2 discusses the main differences between IP network and CCN. Section 3 describes our proposed adaptive transmission control mechanism. In Section 4, the experiment environment and result analysis are presented. Finally, the summary of our work and our future research plan are described in Section 5.

2. TCP/IP Network vs. Content-Centric Network

The TCP/IP protocol suit is the de facto Internet standard [20, 21]. The success of the TCP/IP protocol suit owes to its preeminent network architecture design as depicted in Figure 1. The TCP/IP protocol suit adopts a layered structure, and intricate tasks for the data communication process are encapsulated into each layer [22, 23]. Each layer provides certain functions to its neighbour layers, such as routing, flow control, error detection, and session management. The interactions between adjacent layers are through predefined interfaces. The layered encapsulation makes the TCP/IP protocols can flexibly deal with the constant change of communication techniques [24, 25]. There are several protocols in each layer for different communication purposes, for example, TCP and UDP in the transport layer. However, the network layer only runs IP protocol, which unifies the packet forward process and builds the basis for the protocols in higher layers [26, 27].

CCN inherits the layered design principle of TCP/IP protocol suit as shown in Figure 1 [7]. The most important difference is that CCN has replaced the IP at the network layer with the named content. In TCP/IP, the function of the IP address is to virtually bind sender, receiver, and transmission packets. According to the IP address, the IP routers can forward the packets from the sender to the receiver. Nevertheless, the focus of CCN is to deliver the content to the request instead of connecting the sender and receiver. The named content is the core communication component, and CCN unbinds the content from its location. The advantage of unbinding is that the requested content can be retrieved from any node that holds the content [28].
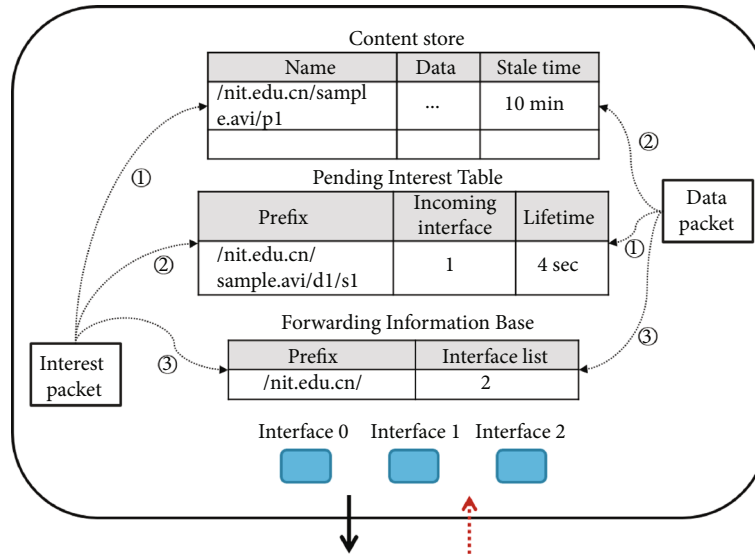
Figure 2: The comparison of TCP/IP protocols with CCN protocols.

In order to achieve the goal, CCN introduces new packet types and node structure as presented in Figure 2. A series of interest packets are generated and sent into the network by the receiving node, when the user wants to obtain certain content, for example, listening to a music or watching a movie. The content is encapsulated into data packets. Any CCN node holding the content can send corresponding data packets back to the request node. The CCN packets are processed only based on the name field in the packet header. The content name is recommended to take a form of tree structure, such as the form of Uniform Resource Identifier [29].

A typical CCN node has three data modules: Pending Interest Table (PIT), Content Store (CS), and Forwarding Information Base (FIB). The PIT records every interest packet that the node forwarded and the forwarded interface. When the interest packet arrives at the content holder node, the PIT entries of the forwarding nodes set up a virtual path for the data packet traveling back. The PIT also has a role of transmission rate control and loss recovery. Since content is unbended with its location, the CCN node can reutilize the forwarded data packets other than eliminate them [30]. The CS is a memory space of a CCN node for buffering data packets. To improve the efficiency, memory replacement algorithms are used to manage the CS. The FIB has the same function as the routing table in the IP router. The CCN node references the FIB to forward interest packets to successive nodes and periodically updates the FIB according to routing algorithms.

Figure 2 shows three steps when a CCN node receives an interest packet. The first step is to search the CS by using the name field as the key [7]. If the matched data packet is found, then the data packet is sent back through the incoming network interface, and the interest packet is abandoned. This means that the content has been transmitted by this node and the content is reutilized. If there are no matched data packets in the CS, the second step is searching the PIT. If there is a matched entry in the PIT, then the same request has been forwarded before. The CCN node appends the incoming network interface to the existing entry and waits the return of the data packet. In the case of no matched entry in the PIT, the third step is to look up the FIB. If an entry with the same name is found in FIB, then the interest packet is forwarded to the subsequent node. Otherwise, the interest packet is abandoned [30].

There are also three steps for the incoming data packet. The first step is to look up the PIT. If there is a match, then the data packet has been requested. The CCN node sends the data packet to each network interface listed in the entry, and the entry is removed from the PIT after sending. In case of no match in the PIT, the data packet is abandoned and the process is finished. The reason is that malicious nodes may broadcast junk content into the network, which causes the contamination of the CS and useful content cannot be stored. In the second step, the content is buffered in the CS for future reuse. The third step is to check the FIB by using the name of the data packet. If there is no match, the name and incoming interface are added into the FIB. The third step can help the routing algorithm to update the FIB in a distributed manner. After that, the whole process is finished [31].

The content caching capability of CCN nodes is crucial for video streaming applications. In [32], the authors present a progressive caching algorithm for video streaming over CCN. The algorithm generates a metafile that includes the content priority information, and the CCN nodes can decide the cache policy based on the metafile. In [8], a multisource video streaming algorithm is proposed. Due to the caching capability of CCN nodes, the receivers might receive the data from different nodes. By considering the quality of experience, the proposed algorithm switches to different sources. [13] utilizes a scalable video coding scheme and proposes caching algorithms to improve the video delivery services. The content centric communication model is used not only for static networks but also for high dynamic networks [33–35]. In [36], the authors show the feasibilities of video streaming over CCN-based vehicular networks.
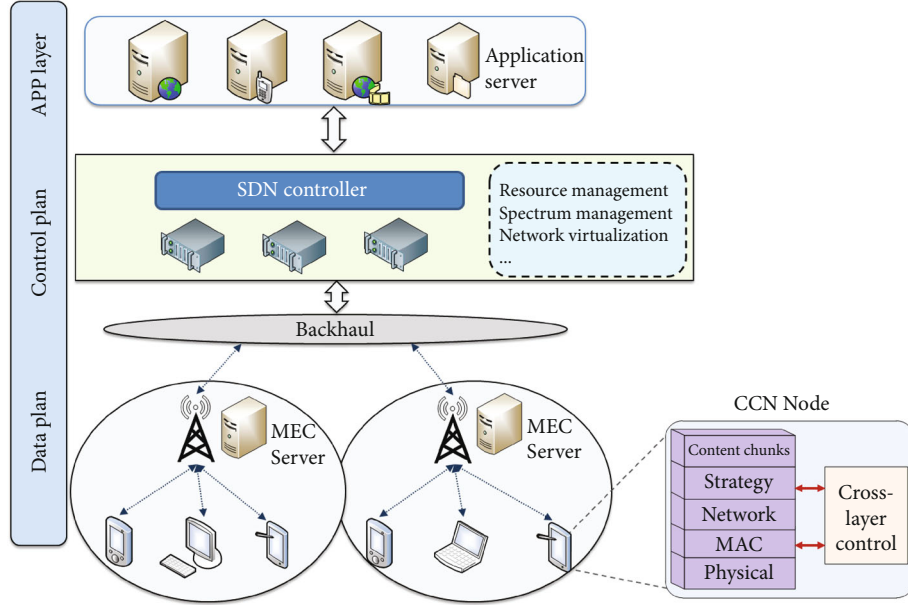
FIGURE 3: The system architecture of MEC-based Content-Centric Network.

3. Adaptive Video Transmission Mechanism

3.1. Architecture of Video Streaming System. The system architecture of our proposed mechanism is presented in Figure 3. For the flexible configuration of network resources, the proposed video streaming system integrates the SDN scheme. SDN consists of the control plan, data plan, and application layer. The application layer includes various types of application servers, for example, video streaming server, web server, and mail server. The different applications have different requirements for quality of service. In addition, CCN is independent from particular transport protocols; any protocol with the function of packet transmission, such as TCP, UDP, IP, and P2P, can be used. The SDN control plan uniformly and flexibly schedules network resources and provides a unified programming interface for the management program to support programmable control. The data plane including the backhaul network and access network is only responsible for data forwarding. Through the encapsulation of control functions and the abstraction of applications and networks, SDN treats the network as a logical or virtual entity to form an architecture similar to a computer operating system. The data plane provides a standardized and open interface for the control plane. In this way, the control plane can achieve more flexible control capabilities based on the global network view, and the data plan and application layer can be flexibly and independently expanded according to actual requirements to meet changing needs.

3.2. Adaptive Transmission Control Algorithm. CCN follows the receiver-driven model, where the receiver takes a major role during the video transmission. Because CCN does not assume the transport layer protocols offering reliable data communication service, flow control, congestion control, etc. It is the receiver's responsibility to regulate its communi-

cation policy to respond to network condition variation. Our proposed transmission control algorithm is based on the receiver-driven model and TCP flow control and congestion control scheme. According to the video bitrate, the receiver periodically sends interest packets for retrieving video data. The ongoing interest packets are stored in the PIT. For each PIT entry, there is a timer with a value called LifeTime. If no data packet is received within LifeTime, either interest or data packet is lost. We consider the time between sending an interest packet and receiving the corresponding data packet as the round trip time (RTT) of the transmission. The control variable W_{PIT} indicates the length of the PIT, which is the number of ongoing interest packets allowed to be sent. When the number of PIT entries reaches W_{PIT}, the sending of interest packets is suspended until receiving a data packet. Initially, the W_{PIT} is set to satisfy the required bitrate of the base layer denoted W_{base}. When a data packet is received, the W_{PIT} is increased proportionally to its current value.

As shown in Figure 4, the communication path between the receiver and the video server contains wired and wireless links. The loss of interest and data packet can happen in any links. In wired links, the CCN nodes use queue management algorithms that actively dropping packets to prevent network congestion, for instance, random early detection, random exponential marking, and stochastic fair BLUE. In wireless links, the packet loss is mainly caused by transmission errors. To distinguish between wired and wireless link loss, the MEC server traces the wireless loss rate denoted as $P_{wireless}$ and sets into the data packet header. The receiver can obtain the total packet loss rate P_{total}, by observing the event of the PIT timeout. If there is a timeout and P_{total} is equal to $P_{wireless}$, it means that packet loss is at wireless links and W_{PIT} remains unchanged. If P_{total} is larger than $P_{wireless}$, then certain nodes along the links are expected to encounter congestion. The receiver needs to reduce the transmission rate; therefore,
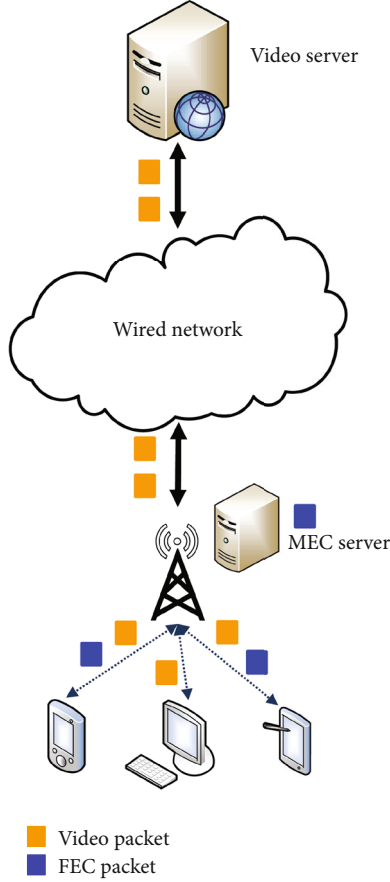
FIGURE 4: Data offloading and computation offloading for video streaming over MEC-based Content-Centric Network.

W_{PIT} is set to W_{base}. We can use ordinary differential equation to model the dynamics of the proposed transmission algorithm as

$$\dot{W}_{\text{PIT}} = \frac{x(t)(1 - P_{\text{total}} + P_{\text{wireless}})\alpha}{W_{\text{PIT}}} - x(t)(P_{\text{total}} - P_{\text{wireless}})W_{\text{base}},$$

(1)

where $x(t)$ is the transmission rate of the receiver and $x(t) = W_{\text{PIT}}/\text{RTT}$. α is the increasing rate of W_{PIT}. With the replacement, the derivative of the transmission rate is

$$\dot{x}(t) = \frac{(1 - P_{\text{total}} + P_{\text{wireless}})\alpha}{\text{RTT}^2} - x(t)(P_{\text{total}} - P_{\text{wireless}})x_{\text{base}},$$

(2)

where $x_{\text{base}} = W_{\text{base}}/\text{RTT}$ is the transmission rate of the base layer.

3.3. Analytical Model of Proposed Algorithm. To analyse the equilibrium behaviour of the proposed transmission control algorithm, we introduce a network utility function $U_i(x_i)$ for the ith receiver. The network utility function is a mapping of user benefit to the transmission rate. By increasing the transmission rate, the user can acquire more enhancement

```
1: if arriving data packet then
2:   obtain packet loss rate P_wireless from packet header
3:   W_PIT = (1 + α)W_PIT
4:   clear timeout counter n = 0.
5: end if
6: if PIT timeout then
7:   calculate total packet loss rate P_total.
8:   increase timeout counter n = n + 1
9:   if P_total > P_wireless then
10:    if W_PIT > W_base then
11:      W_PIT = W_base
12:    else
13:      W_PIT = β^n W_base
14:    end if
15:  else
16:    calculate number of FEC packet W_FEC
17:    send FEC interest packets to overcome loss
18:  end if
19: end if
```

ALGORITHM 1: Adaptive Transmission Control Algorithm.

layers, and the user benefit is also increased. Hence, the network utility function is assumed to be concave and differentiable. Consider a wireless network with N receivers and K links, the transmission control problem is presented as network utility maximization [37]:

$$\text{maximize}_{x_r > 0} \sum_{i=1}^{N} U_i(x_i),$$

$$\text{subject to} \sum_{i=1}^{K} d_{ji}x_i \leq C_j,$$

(3)

where C_j is the link capacity and the inequality constraint means that the total transmission rate should be not greater than the link capacity. d_{ji} is the portion of transmission flow carried by the link j. Let λ_j be the Lagrange multiplier for the capacity constraint of the link j. The Lagrangian is defined as

$$L(\mathbf{x}, \lambda) = \sum_{i=1}^{N} U_i(x_i) - \sum_{j=1}^{K} \lambda_j \left(\sum_{i=1}^{N} d_{ji}x_i - C_j \right),$$

(4)

where the last term can be considered as the penalty if the total transmission rate exceeds the link capacity. Because the object function of (3) is concave and the inequality constraint is linear, the network utility maximization forms a convex optimization problem. From Karush–Kuhn–Tucker (KKT) conditions, we can derive from taking the derivative of $L(\mathbf{x}, \lambda)$ respect to x_i:

$$U_i'(x_i^*) - \sum_{j=1}^{K} \lambda_j^* d_{ji} = 0.$$

(5)

At the equilibrium point, the derivative of the transmission rate (2) is zero. The packet loss caused by the congestion
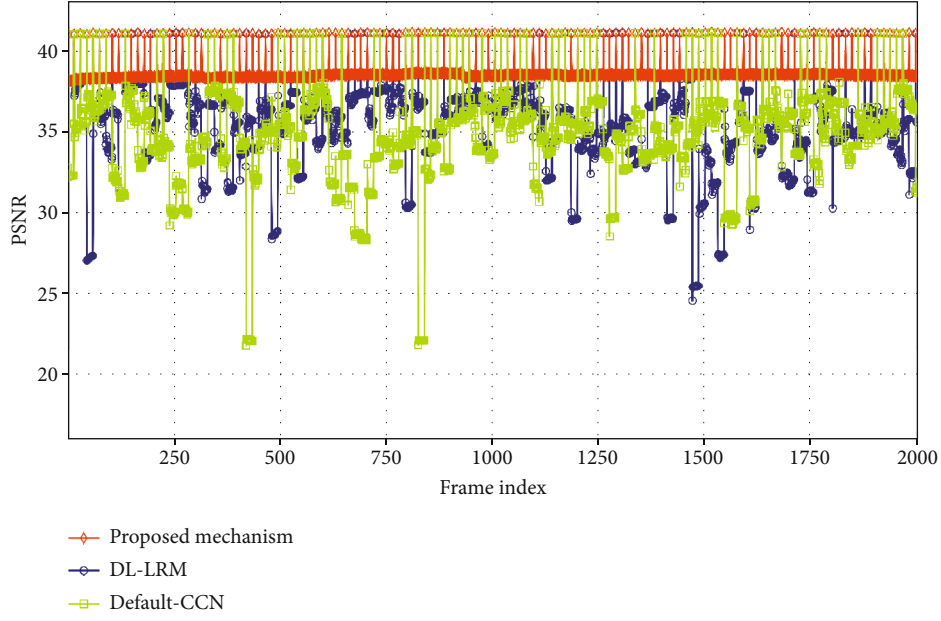
FIGURE 5: Video quality with no wireless link error.
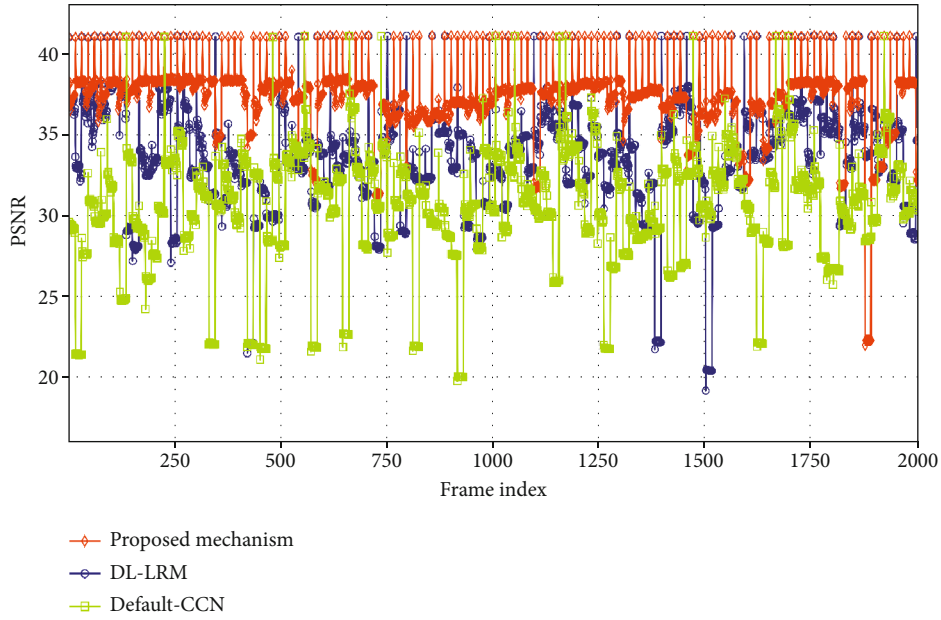
FIGURE 6: Video quality with 8% wireless link error rate.

is also a penalty; with (5), we can obtain

$$U_i'(x_i) = \frac{1}{\alpha + \mathrm{RTT}^2\, x_{\mathrm{base}} x_i}. \tag{6}$$

Then, the utility function is

$$U_i(x_i) = \frac{\log\left(\alpha + \mathrm{RTT}^2\, x_{\mathrm{base}} x_i\right)}{\mathrm{RTT}^2\, x_{\mathrm{base}}}. \tag{7}$$

Although there are different optimization algorithms [38–40], with the analytical expression of the utility function, the convex optimization algorithm is carried out to find the optimal solution.

3.4. MEC Assistant Link Loss Recovery. As explained in the previous section, the packet loss can happen in wired and wireless links. Usually, the loss in the wired link is caused by the routing nodes proactively dropping packets to prevent the link congestion. On the other hand, the loss in the wireless link is due to wireless link error. The packet loss can be
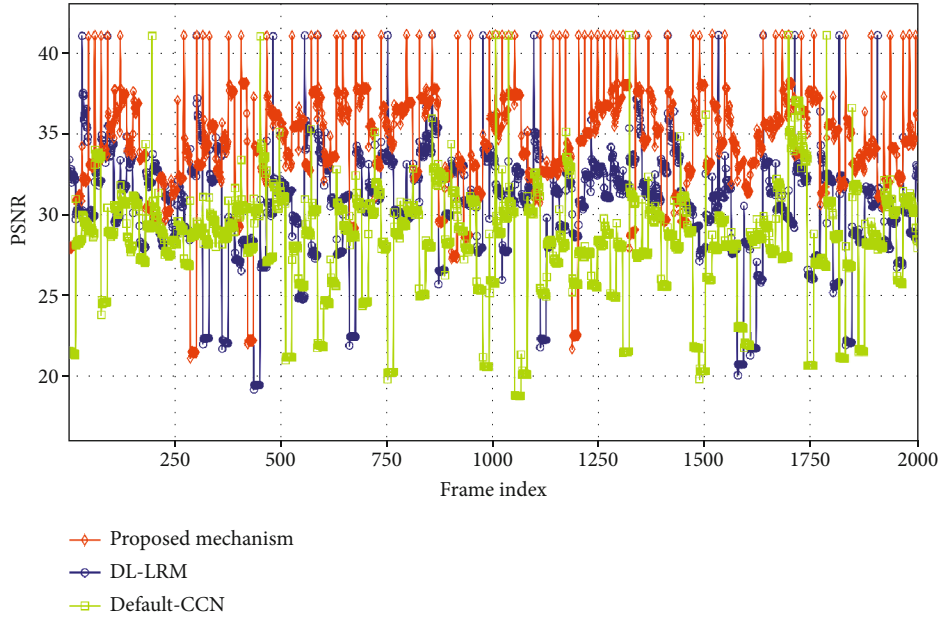
FIGURE 7: Video quality with 12% wireless link error rate.
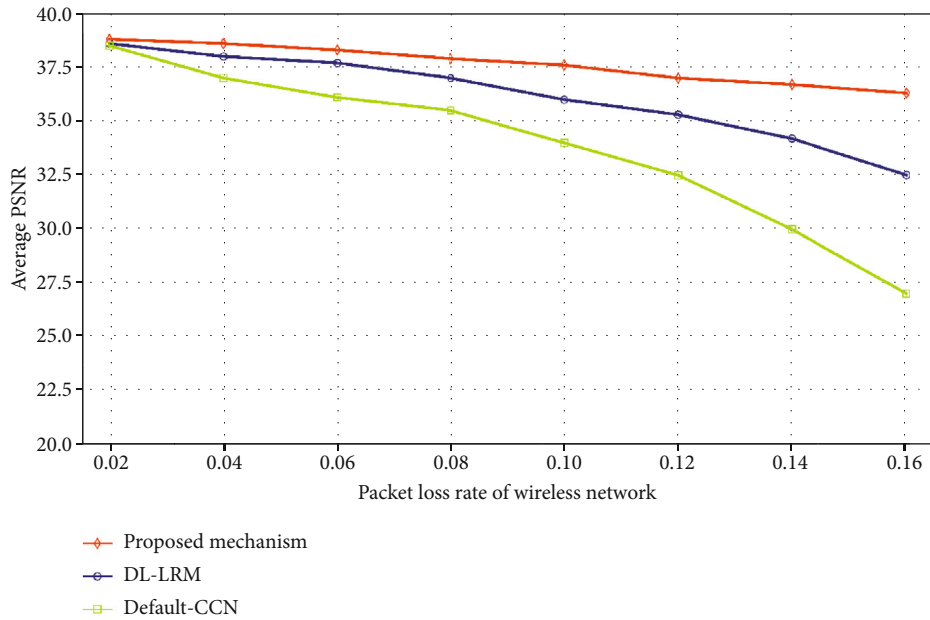
FIGURE 8: Average video quality with various wireless link error rates.

recovered by Automatic Repeat reQuest (ARQ) or forward error correction. In our proposed system scheme as shown in Figure 4, the MEC server, routing nodes, and the video server are able to store the video content. The receiver may obtain data packets from different nodes; hence, the ARQ is improper. Our proposed mechanism adopts the FEC approach to recover packet loss in the wireless link. The generation of the FEC packet is offloaded to the MEC server from the video server. Since all data packets must go through the wireless base station to the receiver. The MEC server collects the video data and generates FEC packets. The FEC packet is retrieved from the MEC server instead of the video server; the workload of wired links is vastly reduced. When the receiver gets the wireless loss rate from the MEC server, the minimum number of FEC packets required to overcome the loss is

$$W_{\text{FEC}} = \frac{W_{\text{PIT}} P_{\text{wireless}}}{1 - P_{\text{wireless}}}, \quad (8)$$

where $W_{\text{PIT}} P_{\text{wireless}}$ is the expected number of packet loss. Depending on the current transmission rate W_{PIT}, the receiver sends interest packets for FEC. The detailed steps of our proposed mechanism are presented in Algorithm 1.

4. Experiment Results and Analysis

To analyse the performance of our proposed mechanism, a series of experiments were carried out with NS-3 simulator [41]. We used H.264/scalable video coding to compress the raw video sequences "Elephants Dream" [42]. The proposed mechanism was compared with DL-LRM [35] and CCN default approach [7] with the same network conditions. The encoded video consisted of one base layer and three enhancement layers. The network topology of experiments included wired and wireless links. The links from the video server to the base stations were wired and the bandwidth was 1 Gigabit per second. The base stations communicated with end nodes through wireless link, and the bandwidth was 100 Megabits per second. There were 200 end nodes, and the nodes were divided into 10 groups. Each group node requested the same video content to simulate network congestion. The experiments were conducted under different packet loss rates to simulate wireless link error.

Figure 5 shows the video quality of receivers with no wireless link error. Because the DL-LRM and CCN default approach have no control of transmission rate, even the network is congested, the two approaches still aggressively send interest packets. This results in numerous packets being dropped by routing nodes and decreasing video quality.

With increasing wireless link error rate, the number of packet loss is also increased. As described in Section 2, each PIT entry has a timer which value is called LifeTime. The LifeTime is set by the receiver, and it should be larger than the upper bound of RTT samples. The reason is that every node in CCN has a PIT. If a timeout arises at an end node, the node may retransmit the interest packet. If a timeout arises at a routing node, the PIT entry is deleted. Later, if the data packet arrives, and there is no corresponding PIT entry, then the data packet is abandoned. On account of the in-network caching function of CCN, the data packets may be retrieved from different nodes and the RTT value has high fluctuation. The default CCN approach for packet loss is ARQ, and the LifeTime is set to a constant value. The video streaming application is delay sensitive. If the loss packet is recovered, nonetheless it exceeds the playtime, then the packet is useless. From experiment results shown in Figures 6 and 7, the performance of the default CCN approach drops dramatically with the increase of the error rates.

Our proposed mechanism distinguishes the cause of packet loss. If the network is congested, then the transmission is limited to the request for the base layer only to relieve the congestion. If the loss is caused by wireless link error, FEC packets are requested. The DL-LRM does not identify the reason of packet loss and just sends FEC packets to respond to the PIT timeout. As presented in Figure 8, our proposed mechanism outperforms the DL-LRM and default CCN approach under different network conditions.

5. Conclusions

In this paper, we have presented an adaptive video transmission mechanism in Mobile Edge Computing- (MEC-) based CCN. The main contribution of our mechanism is to adaptively control transmission rate to adapt to the network conditions variation, instead of greedily acquiring video content. Therefore, it can friendly coexist with different types of applications and improve overall network efficiency. We adopt scalable video coding (SVC) scheme to balance the video quality and transmission rate. In addition, the MEC assistant loss recovery algorithm is suggested to reduce the burden of the video streaming server and backhaul network. The experiment results show that our proposed mechanism outperforms the existing approaches. For future work, we will study the scalability issue and analyse the performance in a large network environment.

Data Availability

The data used to support the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A subjective and objective study of stalling events in mobile streaming videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 183–197, 2019.

[2] Y. Cao, L. Zeng, Q. Liu, G. Lei, M. Huang, and H. Wang, "Receiver-assisted partial-reliable multimedia multipathing over multi-homed wireless networks," *IEEE Access*, vol. 7, pp. 177675–177689, 2019.

[3] H. Wang, J. Liang, and C. C. J. Kuo, "Overview of robust video streaming with network coding," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 1, pp. 36–50, 2010.

[4] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022*, White Paper, 2019.

[5] Y. Zhang, C. Gao, Y. Guo et al., "Proactive video push for optimizing bandwidth consumption in hybrid CDN-P2P VoD systems," *IEEE INFOCOM*, vol. 2018, pp. 2555–2563, 2018.

[6] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin, and F. R. Yu, "A survey of mobile information-centric networking: research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2353–2371, 2018.

[7] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pp. 1–12, Rome, Italy, 2009.

[8] L. Han, X. Bao, W. Wang, X. Feng, Z. Liu, and W. Tan, "A receiver-driven loss recovery mechanism for video dissemination over information-centric VANET," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 7, 2017.

[9] M. N. Sadat, R. Dai, L. Kong, and J. Zhu, "QoE-aware multi-source video streaming in content centric networks," *IEEE Transactions on Multimedia*, vol. 22, no. 9, pp. 2321–2330, 2020.

[10] R. Jia, Z. Liu, X. Wang, X. Gan, X. Wang, and J. J. Xu, "Modeling dynamic adaptive streaming over information-centric networking," *IEEE Access*, vol. 4, pp. 8362–8374, 2016.

[11] C. Xu, P. Zhang, S. Jia, M. Wang, and G. Muntean, "Video streaming in content-centric mobile networks: challenges and solutions," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 157–165, 2017.

[12] S. Lederer, C. Mueller, C. Timmerer, and H. Hellwagner, "Adaptive multimedia streaming in information-centric networks," *IEEE Network*, vol. 28, no. 6, pp. 91–96, 2014.

[13] Z. Zhang, J. Dai, M. Zeng, D. Liu, and S. Mao, "Scalable video caching for information centric wireless networks," *IEEE Access*, vol. 8, pp. 77272–77284, 2020.

[14] W. Li, S. M. A. Oteafy, and H. S. Hassanein, "Rate-selective caching for adaptive streaming over information-centric networks," *IEEE Transactions on Computers*, vol. 66, no. 9, pp. 1613–1628, 2017.

[15] T. Zhang, X. Xu, Z. Le, X. Jiang, and J. Loo, "Cache Space Efficient Caching Scheme for Content-Centric Mobile Ad Hoc Networks," *IEEE Systems Journal*, vol. 13, no. 1, pp. 530–541, 2019.

[16] A. Younis, T. X. Tran, and D. Pompili, "On-demand video-streaming quality of experience maximization in mobile edge computing," in *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks*, pp. 1–9, Washington DC, United States, 2019.

[17] M. F. Tuysuz and M. E. Aydin, "QoE-based mobility-aware collaborative video streaming on the edge of 5G," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 11, pp. 7115–7125, 2020.

[18] J. Luo, F. R. Yu, Q. Chen, and L. Tang, "Adaptive video streaming with edge caching and video transcoding over software-defined Mobile networks: a deep reinforcement learning approach," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1577–1592, 2020.

[19] Y. Guo, F. R. Yu, J. An, K. Yang, C. Yu, and V. C. M. Leung, "Adaptive bitrate streaming in wireless networks with transcoding at network edge using deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3879–3892, 2020.

[20] Y. Li, D. Leith, and R. N. Shorten, "Experimental evaluation of TCP protocols for high-speed networks," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1109–1122, 2007.

[21] Y. Cao, J. Chen, Q. Liu, G. Lei, H. Wang, and I. You, "Can multipath TCP be robust to cyber attacks with incomplete information?," *IEEE Access*, vol. 8, pp. 165872–165883, 2020.

[22] K. Sasaki, M. Hanai, K. Miyazawa, A. Kobayashi, N. Oda, and S. Yamaguchi, "TCP fairness among modern TCP congestion control algorithms including TCP BBR," in *2018 IEEE 7th International Conference on Cloud Networking*, pp. 1–4, Tokyo, Japan, 2018.

[23] K. Ratna Pavani and N. Sreenath, "Performance evaluation of TCP-Reno, TCP-Newreno and TCP-Westwood on burstifica-

tion in an OBS network," in *2012 18th International Conference on Advanced Computing and Communications*, pp. 19–24, Bangalore, India, 2012.

[24] R. Amrutha and V. Nithya, "Performance analysis of TCP incast with TCP Lite and Abstract TCP," in *2015 Global Conference on Communication Technologies*, pp. 18–22, Thuckalay, India, 2015.

[25] P. Chaudhary and S. Kumar, "Comparative study of TCP variants for congestion control in wireless network," in *2017 International Conference on Computing, Communication and Automation*, pp. 641–646, Greater Noida, India, 2017.

[26] M. Polese, R. Jana, and M. Zorzi, "TCP and MP-TCP in 5G mmWave networks," *IEEE Internet Computing*, vol. 21, no. 5, pp. 12–19, 2017.

[27] Y. Cao, M. Collotta, S. I. Xu, L. Huang, X. Tao, and Z. Zhou, "Towards adaptive multipath managing: a lightweight path management mechanism to aid multihomed mobile computing devices," *Applied Sciences*, vol. 10, no. 1, 2020.

[28] N. Kamiyama, "Analyzing impact of introducing CCN on profit of ISPs," *IEEE Transactions on Network and Service Management*, vol. 12, no. 2, pp. 176–187, 2015.

[29] A. Aboodi, T. Wan, and G. Sodhy, "Survey on the incorporation of NDN/CCN in IoT," *IEEE Access*, vol. 7, pp. 71827–71858, 2019.

[30] R. Jmal and L. Chaari Fourati, "Content-centric networking management based on software defined networks: survey," *IEEE Transactions on Network and Service Management*, vol. 14, no. 4, pp. 1128–1142, 2017.

[31] Y. Wang, Z. Li, G. Tyson, S. Uhlig, and G. Xie, "Design and evaluation of the optimal cache allocation for content-centric networking," *IEEE Transactions on Computers*, vol. 65, no. 1, pp. 95–107, 2016.

[32] H. Noh and H. Song, "Progressive caching system for video streaming services over content centric network," *IEEE Access*, vol. 7, pp. 47079–47089, 2019.

[33] F. Khan and H. Li, "Ensuring trust and confidentiality for adaptive video streaming in ICN," *in Journal of Communications and Networks*, vol. 21, no. 6, pp. 539–547, 2019.

[34] Z. Zhang, C. Lung, M. St-Hilaire, and I. Lambadaris, "An SDN-based caching decision policy for video caching in information-centric networking," *in IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 1069–1083, 2020.

[35] X. Maksymyuk, J. Zhao, and Y. Liu, "Deep learning based loss recovery mechanism for video streaming over mobile information-centric network," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 9, pp. 4572–4586, 2019.

[36] Z. Zhang, C. Lung, M. St-Hilaire, and I. Lambadaris, "Smart proactive caching: empower the video delivery for autonomous vehicles in ICN-based networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7955–7965, 2020.

[37] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low, "Congestion control for high performance, stability, and fairness in general networks," *IEEE/ACM Transactions on Networking*, vol. 13, no. 1, pp. 43–56, 2005.

[38] J. Pan, N. Liu, S. Chu, and T. Lai, "An efficient surrogate-assisted hybrid optimization algorithm for expensive optimization problems," *Information Sciences*, vol. 561, pp. 304–325, 2021.

[39] L. Lv, J. Zhao, J. Wang, and T. Fan, "Multi-objective firefly algorithm based on compensation factor and elite learning," *Future Generation Computer Systems*, vol. 91, pp. 37–47, 2019.

[40] J. Z. Xie, L. Lv, H. Wang, H. Sun, and X. Yu, "Firefly algorithm with deep learning," *Acta Electronica Sinica*, vol. 46, no. 11, pp. 2633–2641, 2018.

[41] G. F. Riley and T. R. Henderson, *The NS-3 Network Simulator*, Modeling and Tools for Network Simulation, Springer, Berlin, Heidelberg, 2010.

[42] *Elephants Dream* Video Test Media Collection https://media .xiph.org/video/derf/.

Research Article

Robust Graph Structure Learning for Multimedia Data Analysis

Wei Zhou [iD],[1,2] **Zhaoxuan Gong,**[1] **Wei Guo,**[1] **Nan Han,**[3] **and Shaojie Qiao** [iD][4]

[1]*School of Computer, Shenyang Aerospace University, Shenyang 110136, China*
[2]*Shenyang Institute of Computing Technology Co. Ltd., CAS, Shenyang 110168, China*
[3]*School of Management, Chengdu University of Information Technology, Chengdu 610103, China*
[4]*School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China*

Correspondence should be addressed to Wei Zhou; zhouweineu@outlook.com and Shaojie Qiao; sjqiao@cuit.edu.cn

With the rapid development of computer network technology, we can acquire a large amount of multimedia data, and it becomes a very important task to analyze these data. Since graph construction or graph learning is a powerful tool for multimedia data analysis, many graph-based subspace learning and clustering approaches have been proposed. Among the existing graph learning algorithms, the sample reconstruction-based approaches have gone the mainstream. Nevertheless, these approaches not only ignore the local and global structure information but also are sensitive to noise. To address these limitations, this paper proposes a graph learning framework, termed Robust Graph Structure Learning (RGSL). Different from the existing graph learning approaches, our approach adopts the self-expressiveness of samples to capture the global structure, meanwhile utilizing data locality to depict the local structure. Specially, in order to improve the robustness of our approach against noise, we introduce $l_{2,1}$-norm regularization criterion and nonnegative constraint into the graph construction process. Furthermore, an iterative updating optimization algorithm is designed to solve the objective function. A large number of subspace learning and clustering experiments are carried out to verify the effectiveness of the proposed approach.

1. Introduction

With the rapid growth of information technology and computer network technology, a large number of multimedia data can be collected from many research fields such as computer vision, image processing, and natural language processing. However, most of the multimedia data are represented by the high dimension and complex structures [1, 2]. Therefore, how to accurately analyze these data becomes a vital problem. Inspired by the pattern recognition and machine learning techniques, many multimedia data analysis approaches based on subspace learning and clustering have been put forward recently [3–6]. However, learning or constructing a valuable graph to describe the pairwise similarity or relationship among the samples is a key issue to multimedia data analysis [7].

Nowadays, a series of graph learning approaches have been proposed in which the heat-kernel function is the most widely used graph construction manner, such as k-nearest-neighborhood graph (k-NN graph) or ε-nearest-neighborhood graph (ε graph). The edges of vertexes are computed based on the Euclidean distance among samples and then the weights of the edge between two vertexes are estimated by the heat kernel [8]. However, there are two main limitations in these approaches [9]. First, the choice of parameters in these approaches, such as the neighbor number k or radius ε, is very challenging, which can impact the final performance of the task. Second, the processes of neighbor selection and weight calculation are independent, which are sensitive to noise and often cannot well reveal the real similarities of samples [10].

To overcome these drawbacks, sparse representation (SR) based graph construction approach has been proposed, which is often called l_1-graph or sparse graph. In l_1-graph [11], each sample is regarded as the query sample and the rest of samples are considered as the dictionary to represent the query sample. Therefore, the similarities between the query sample and the remainder samples can

be measured. Since l_1-graph employs l_1-norm constraint on the regression model for selecting a few important samples, it has a better discriminability and more robustness to deal with noise. In the past decades, a series of excellent learning approaches based on l_1-graph have been designed and successfully applied in different areas [12]. Although l_1-graph can reveal the linear relationship between a single point and other points, there are still some limitations as follows. First, l_1-graph strictly assumes the dictionary of regression should be overcompleted, which is unsatisfied in many real applications especially for the graph learning. Second, l_1-graph pays too much attention on the sparsity while it neglects the correlations between the samples, so it cannot offer a smooth data representation. Therefore, SR is not a good choice for graph construction. To overcome the aforementioned problems suffered by SR, Zhang et al. [13] introduced a Collaborative Representation (CR) linear regression approach by employing the l_2-norm rather than l_1-norm sparsity regularization. Compared to SR, CR provides more relaxation for regression coefficients and obtains a smoother data representation.

Considering that both SR and CR usually reveal the linear relationship between a single data point and other data points, the global structure of the data is ignored. To address this problem, Liu et al. [14] suggested Low-Rank Representation (LRR) for subspace clustering. The main purpose of LRR is to find a coefficient matrix Z by trying to reconstruct each data point as a linear combination of all the other data points, which is called self-representation. Different from the traditional similarity measurement approaches based on distance, i.e., k-nearest neighborhood or ε-nearest neighborhood, the representation-based approaches, such as SR, CR, or LRR, measure the similarity between data by solving an optimization problem. These approaches improve the image structure to achieve better classification and clustering performance overall. However, the objective function of LRR is not differentiable which has a high computation complexity on solving the rank minimization problem. To efficiently solve the limitation of LRR, Lu et al. [15] proposed Least Squares Regression (LSR) by grouping the highly correlated data together, which is robust to noise. Compared with LRR, LSR is simpler and more efficient.

In recent years, researchers found that the relationships between data in real applications usually show high dimension nonlinear, so the aforementioned linear representation approaches can hardly achieve good performance. Many researchers paid more attention on revealing the nonlinear relationship between data points of interests [16–27]. For example, Wang et al. [28] explored the criterion of Locally Linear Embedding (LLE) and used it to construct the graph by computing the weights between the pairs of samples. Wei and Peng [29] adopted a similar criterion to that LLE to construct a neighborhood-preserving graph for semisupervised dimensionality reduction. Furthermore, Yu et al. [30] have found that the nonzero coefficients of the sparse coding always are assigned to the neighbor samples of the query sample. To encourage the coding to be locality, some local feature-based coding approaches have been proposed, which achieve excellent performance for the classification

and clustering tasks [31]. With the usage of the merits of local constraints, Peng et al. [32] put forward Locality-Constrained Collaborative (LCC) representation, which achieves better classification performance than those nonlocal approaches. Chen and Yi [33] took the local constraint and LSR into consideration and designed Locality-Constrained LSR (LCLSR) for subspace clustering. LCLSR explores both the global structure of data points and the local linear relationship for data points, forcing the representation to prefer the selection of neighborhood points. Although LCLSR considers the locality structure of data, there are still some limitations on the graph structure. On the one hand, the objective function of LCLSR is based on l_2-norm, which is very sensitive to noise; on the other hand, the process of sample reconstruction ignores the relationships between sample representations. For example, similar original samples should generate similar coding vectors, and this process weakens the effectiveness of graph learning approaches.

To combat these issues, we design a novel graph learning approach, named Robust Graph Structure Learning (RGSL). Specifically, the self-expressiveness of samples and adaptive neighbor selection approach are introduced to preserve both the local and global structures of data. For enhancing the robustness of graph construction, we introduce the $l_{2,1}$-norm constraint and nonnegative constraint on the adjacency graph weight matrix to reduce the influence of noise points in graph construction. Therefore, the proposed approach can estimate the graph from data alone by self-expressiveness of samples and data locality, which is independent of a priori affinity matrix. We assess the benefits of the proposed approach on the subspace learning and clustering tasks. Extensive experiments verify the effectiveness of the proposed approach over other state-of-the-art approaches. The framework of the proposed approach is shown in Figure 1.

The outline of this paper is as follows. Section 2 reviews some related work briefly. Section 3 gives the proposed approach in details. Section 4 shows extensive experiments to prove the effectiveness of the proposed approach. Section 5 presents some conclusions.

2. Related Work

In this section, first, many classic and widely used graph construction approaches are introduced. Then, two kinds of multimedia data analysis techniques including subspace learning and spectral clustering are presented in detail accordingly.

2.1. Graph Construction Approaches. Recently, many graph construction approaches have been proposed for multimedia data analysis. In this subsection, we will review some graph construction approaches related to our work as below.

Liu et al. [14] proposed a Low-Rank Representation (LRR) graph construction approach, in which each sample can be represented by a linear combination of all samples, and meanwhile, a low-rank constraint of coefficient matrix is imposed. Given a high dimensional database $X = [x_1, x_2, \cdots, x_N] \in R^{d \times N}$ in which d is the data dimensionality and N
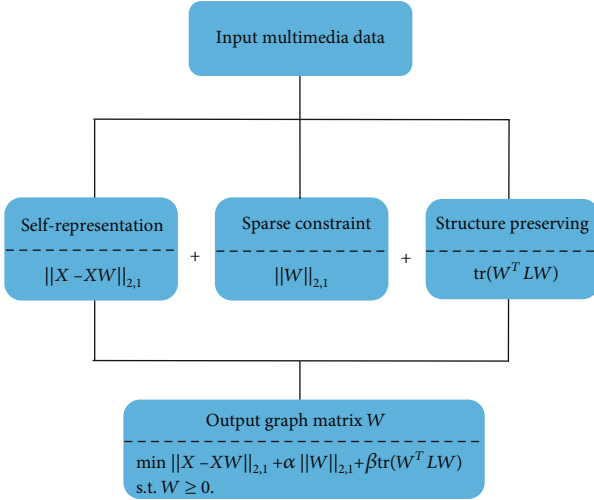
FIGURE 1: The framework of the proposed approach.

is the number of samples. The LRR graph can be obtained by optimizing the following problem:

$$\min_{W} \quad \|W\|_*$$
$$\text{s.t.} \quad X = XW, \tag{1}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix, i.e., the sum of the singular values of the matrix. W denotes the coefficient matrix of data X with the lowest rank.

Although LRR graph can obtain the global structure of data, it is very time-consuming to solve the problem of optimal nuclear norm. Hence, Lu et al. [15] utilized the l_2-graph based on Frobenius norm in place of nuclear norm for fast computing the weight matrix. The LSR graph is defined as

$$\min_{W} \quad \|X\|_F$$
$$\text{s.t.} \quad X = XW, \quad \text{diag}(W) = 0, \tag{2}$$

where $\|\cdot\|_F$ is the Frobenius norm. diag $(\cdot)$ denotes the diagonal operation of a matrix.

In order to make full use of the advantage of locality constraints, Chen and Yi [33] combined LSR and the locality constraints into a unified framework and proposed the LCLSR approach for graph construction. The objective function of LCLSR is

$$\min_{W} \frac{1}{2} \|X - XW\|_F^2 + \frac{\beta_1}{2} \|W\|_F^2 + \frac{\beta_2}{2} \|W \odot D\|_F^2, \tag{3}$$

where β_1 and β_2 are two balance parameters and the symbol $\odot$ represents the Hadamard product. $D = [d_{ij}]_{N \times N} = [e^{\text{dist}(x_i, x_j)}]_{N \times N}$ denotes the distance matrix between samples where the function $\text{dist}(x_i, x_j)$ is a distance metric, such as the Euclidean distance.

2.2. Subspace Learning.

Locality Preserving Projection (LPP) [34] is a well-known subspace learning approach which is used to discover the geometric property of high-dimensional feature space. Suppose that the adjacency graph weight matrix W is given, LPP as aims at ensuring that if the original high-dimensional samples x_i and x_j are "close," then the low-dimensional representations y_i and y_j should be close as well. With the usage of weight matrix W_{ij} as a penalty, LPP is to minimize the following objective function:

$$\frac{1}{2} \sum_{i,j=1}^{N} \left\| y_i - y_j \right\|_2^2 W_{ij} = \text{tr}\left(W^T X L X^T W \right), \tag{4}$$

where $L = D - W$ is the Laplacian matrix, in which D is a diagonal matrix with diagonal elements $d_{ii} = \sum_{j=1}^{N} W_{ij}$. tr$(S)$ is the trace of matrix S. d_{ii} is used to measure the local density around x_i and the bigger d_{ii} indicates that y_i is more important. Hence, a nature constraint can be imposed as $Y^T DY = I$. Based on the equation $Y = P^T X$, the LPP model can be rewritten as

$$\min_{P} \quad \text{tr}\left(P^T X L X^T P \right)$$
$$\text{s.t.} \quad P^T X D X^T P = I. \tag{5}$$

The projection matrix P is constructed by the eigenvectors associated with d smallest nonzero eigenvalues, which can be solved by

$$X L X^T p = \lambda X D X^T p. \tag{6}$$

For a new high-dimensional data x, with the usage of the obtained projection matrix P, we can obtain a low-dimensional data representation by $y = P^T x$.

2.3. Spectral Clustering.

Spectral clustering is a popular clustering approach that uses eigenvectors of a symmetric matrix derived from the distance between data points [35, 36]. Given a data set consisting of N data points $X \in R^{N \times D}$, spectral clustering approach is aimed at partitioning X into K disjoint clusters by exploiting the top K eigenvectors of the normalized graph Laplacian L. Suppose that the graph matrix W is obtained by graph construction approaches and the new representation $Q \in R^{C \times N}$ can be acquired by optimizing the following objective function:

$$\min_{QQ^T = I} Tr\left(Q L_W Q^T \right), \tag{7}$$

where $L_W = D - ((W + W^T)/2)$ is the Laplacian matrix of W, in which $D \in R^{N \times N}$ is the diagonal matrix with $d_{ii} = \sum_j (w_{ij} + w_{ji})/2$. C is the number of selected clusters. Each column in Q represents the new discriminative representation of the corresponding original sample.

Finally, data clustering can be accomplished by performing K-means on the new representation Q.

3. Proposed Method

In this section, some notations are introduced first. Second, we give some detailed descriptions of the proposed RGSL approach. At last, an iterative update algorithm is designed to solve our RGSL approach.

3.1. Notations. Let $X = [x_1, x_2, \cdots, x_N] \in R^{D \times N}$ be the given high-dimensional original data matrix, where D is the dimensionality of samples and N corresponds to the total number of samples. For a matrix $B \in R^{D \times N}$, the definitions of Frobenius norm and $l_{2,1}$-norm are as follows: $\|B\|_F = \sqrt{\sum_{i=1}^{D} \|b^i\|_2^2} = \sqrt{\sum_{j=1}^{N} \|b_j\|_2^2}$ and $\|B\|_{2,1} = \sum_{i=1}^{D} \sqrt{\sum_{j=1}^{N} b_{ij}^2}$, in which b^i and b_j are the i-th row and the j-th column of B, respectively.

3.2. Objective Function. First, in order to enhance the robustness of graph learning algorithm to noise and obtain the discriminative graph structure, the $l_{2,1}$-norm measure criterion on the traditional LSR is introduced, which is defined as

$$\min \sum_{i=1}^{N} \left(\|x_i - Xw_i\|_2 + \alpha \|w_i\|_2 \right) = \|X - XW\|_{2,1} + \alpha \|W\|_{2,1}$$
$$= \operatorname{tr}\left((X - XW)^T G (X - XW) \right)$$
$$+ \alpha \operatorname{tr}\left(W^T Q W \right), \tag{8}$$

where $\|\cdot\|_2$ and $\|\cdot\|_{2,1}$ denote l_2-norm and $l_{2,1}$-norm, respectively. α is a balance parameter. G and Q are diagonal matrices whose diagonal elements are, respectively, defined as $G_{ij} = 1/(\|x_i - Xw_j\|_2 + \varepsilon)$ and $Q_{ii} = 1/(\|w_i\|_2 + \varepsilon)$. ε is a small nonnegative constant for preventing the value of the denominator from being zero.

Second, the relationship between representation coefficients is ignored in the sample reconstruction, i.e., similar original samples should generate similar coding vectors, weakening the effectiveness of graph learning. To solve the abovementioned issue, a modified manifold constraint based on the $l_{2,1}$-norm is designed, which is defined as

$$\min_W \sum_{i,j=1}^{N} \|w_i - w_j\|_2 s_{ij} = \sum_{i,j=1}^{N} \operatorname{tr}\left((w_i - w_j)^T R (w_i - w_j) \right)$$
$$= \operatorname{tr}\left(W^T (D - R) W \right) = \operatorname{tr}\left(W^T L W \right), \tag{9}$$

where s_{ij} denotes the similarity weight value between sample x_i and sample x_j. The elements in matrix R are defined as $R_{ij} = s_{ij}/(\|w_i - w_j\|_2 + \varepsilon)$, and D is a diagonal matrix whose

diagonal elements are $D_{ii} = \sum_{j=1} R_{ij}$. L is the Laplacian of the graph matrix W.

At last, the nonnegative constraint is also imposed on the representation coefficients, and the final objective function of the proposed approach is

$$\min \quad \|X - XW\|_{2,1} + \alpha \|W\|_{2,1} + \beta \operatorname{tr}\left(W^T L W \right)$$
$$= \operatorname{tr}\left((X - XW)^T G (X - XW) \right) + \alpha \operatorname{tr}\left(W^T Q W \right)$$
$$+ \beta \operatorname{tr}\left(W^T L W \right)$$
$$\text{s.t.} \quad W \geq 0, \tag{10}$$

where α and β are two positive balance parameters.

3.3. Optimization. In this section, we give the optimization procedures for the objective function of the proposed approach in Equation (10). From Equation (10), we can observe that the objective function is related to $l_{2,1}$-norm. Thus, the variable W in the objective function is nonconvex and a closed form solution to Equation (10) cannot be given. With regard to this limitation, an iterative update algorithm is designed to optimize the objective function.

3.3.1. Fix G, Q, and R, Update W. First, we fix matrices G, Q, and R. After removing the irrelevant terms, the optimization problem with respect to W in Equation (10) can be simplified to

$$\min \quad \operatorname{tr}\left(X^T G X - 2 W^T X^T G X + W^T X^T G X W \right)$$
$$+ \alpha \operatorname{tr}\left(W^T Q W \right) + \beta \operatorname{tr}\left(W^T L W \right)$$
$$\text{s.t.} \quad W \geq 0. \tag{11}$$

The Lagrangian function of Equation (11) is represented as

$$\ell(W) = \operatorname{tr}\left(X^T G X - 2 W^T X^T G X + W^T X^T G X W \right)$$
$$+ \alpha \operatorname{tr}\left(W^T Q W \right) + \beta \operatorname{tr}\left(W^T L W \right) + \lambda \operatorname{tr}(\Lambda W)$$
$$\text{s.t.} \quad W \geq 0. \tag{12}$$

By computing the derivative of Equation (12) with respect to W and setting it equals to zero

$$\frac{\partial \ell(W)}{\partial W} = -2 X^T G X + 2 X^T G X W + 2 \alpha Q W + 2 \beta L W + \lambda \Lambda = 0. \tag{13}$$

According to the KKT condition [37], we update the solution for W as below:

$$W_{ij} = W_{ij} \frac{\left[X^T G X \right]_{ij}}{\left[\alpha Q + \beta L + X^T G X \right]_{ij}}. \tag{14}$$

1: **Input:** the data matrix $X = [x_1, x_2, \cdots, x_N]$, two balance parameters α and β.

2: **Initialize:** set G_t and Q_t to be the identity matrices, random matrix W, and sample similarity matrix S, $t = 1$.

3: **Repeat**

4:      Update the matrix $R_{t+1} = \{R_{t+1,ij}\}_{N \times N} = s_{ij}/(\|w_{t,i} - w_{t,j}\|_2 + \varepsilon)$

5:      Compute the matrix $D_{t+1,ii} = \sum_{j=1}^{N} R_{t+1,ij}$ and the Laplacian matrix $L_{t+1} = D_{t+1} - R_{t+1}$

6:      Update the matrix $W_{ij} = W_{ij}([X^T G X]_{ij}/[\alpha Q + \beta L + X^T G X]_{ij})$

7:   Update the matrix $G_{t+1} = \{G_{t+1,ij}\}_{N \times N} = 1/(\|x_i - Xw_{t+1,j}\|_2 + \varepsilon)$

8:      Update the matrix $Q_{t+1} = \{Q_{t+1,ii}\}_{N \times N} = 1/(\|w_{t+1,i}\|_2 + \varepsilon)$

9:      $t = t + 1$

10: **Until** convergence

11: **Output:** the graph matrix W

ALGORITHM 1: RGSL.

3.3.2. Fix W, Update G, Q, and R.

When W is fixed and all the irrelevant terms are removed, the solution can be formulated as

$$G_{ij} = \frac{1}{\|x_i - Xw_j\|_2 + \varepsilon}, \quad (15)$$

$$Q_{ii} = \frac{1}{\|w_i\|_2 + \varepsilon}, \quad (16)$$

$$R_{ij} = \frac{s_{ij}}{\|w_i - w_j\|_2 + \varepsilon}. \quad (17)$$

3.4. Algorithm.

In conclusion, the proposed optimization algorithm for RGSL can be summarized as below.

In Algorithm 1, the convergence condition is defined as the change of the value of objective function in Equation (10) which is less than a threshold or a predefined maximum iteration number is reached.

4. Experiment and Results

In this section, first, we will introduce the used databases in our experiment. Next, some graph learning comparison approaches are given. At last, subspace learning and clustering tasks are employed for verifying the effectiveness of the proposed approach.

4.1. Databases.

Four commonly used multimedia databases from the Internet including Yale [38], AR [39], CMU PIE [40], and Extended YaleB [41] are used for verifying the effectiveness of the proposed approach. The detailed statistical information about the four different databases is depicted in Table 1.

Yale database: it contains 165 face images captured from 15 different subjects. Each subject has 11 different images with the varied facial expressions, under different illumination conditions, and wearing glasses or not. Some example images of the Yale database are depicted in Figure 2(a).

AR database: it consists of over 4000 facial images obtained from 70 male and 56 female faces. Images of each person were captured with 26 frontal face images with anger, smiling, and screaming, under varied illumination condi-

TABLE 1: Description of databases.

Databases	Image size	Instances	Classes	Instances of per class
Yale	32×32	165	15	11
AR	32×32	1400	100	14
CMU PIE	32×32	1632	68	24
Extended YaleB	32×32	2432	38	64

tions, and with sunglass and scarf occlusions. Some examples of the AR database are shown in Figure 2(b).

CMU PIE database: there are 41,368 face images of 68 different subjects. Images of each person are captured under 43 different illumination conditions with 13 different poses and 4 different expressions. Here, we employ a subset of CMU PIE which consists of 24 images per subject. A part of example images is illustrated in Figure 2(c).

Extended YaleB database: there are 38 individuals and each individual has 64 images. For each individual, the face images are taken from different illumination conditions with small changes in head pose and facial expression. Example images from this database are shown in Figure 2(d).

Although we can obtain the graph structure from the proposed approach, it is intractable to assess the graph learning approaches using the estimated graph alone. Hence, we will assess the quality of the learned graph by two kinds of multimedia data analysis tasks including subspace learning and spectral clustering. In our experiments, we first vary the graph construction approaches by fixing the graph learning task and then observe the obtained performance associated with subspace learning and spectral clustering tasks.

4.2. Comparison among Several Graph Learning Approaches.

To investigate the performance of our approach on subspace learning and clustering, several state-of-the-art graph learning approaches are chosen to compare in our work, which are shown as below:

(i) KNN graph [8]: the graph edges connected by two vertexes can be generated by the Euclidean distance-based K-nearest neighbor and the heat

(a) Yale

(b) AR

(c) CMU PIE

(d) Extended YaleB
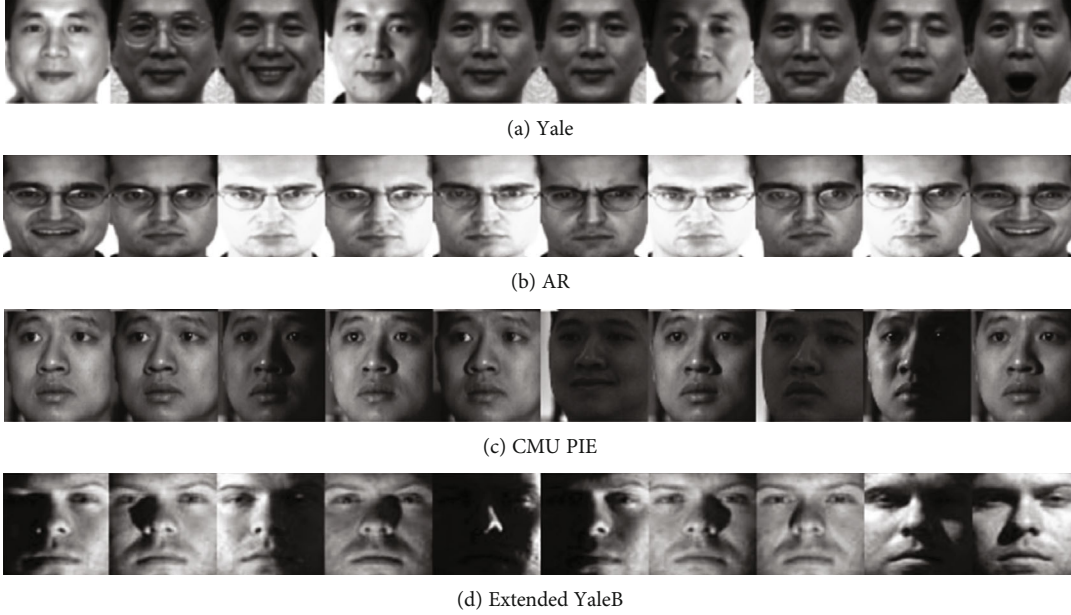
FIGURE 2: Some of the images from the different databases employed in our experiments.

kernel function is used to measure the weight of an edge

(ii) LLE graph [28]: each sample is linearly reconstructed by its neighbors within a local area to preserve the local manifold structure

(iii) L1 graph [11]: the locality structure of data by using L1 sparse representation optimization

(iv) LRR graph [14]: based on self-expressive property, a low-rank graph is obtained

(v) LSR graph [15]: self-expressive property and Frobenius norm are used for fast computing the weight matrix

(vi) LCLSR graph [33]: it combines the locality constraint and LSR together to explore both the global structure of data points and the locality linear relationship of data points

(vii) SGLS graph [42]: it integrates manifold constraints on the unknown sparse codes as a graph regularizer

(viii) Our proposed RGSL graph: our approach takes the global and local structure information into consideration and also introduces the $l_{2,1}$-norm regularization criterion and nonnegative constraint into graph construction process to enhance its robustness

4.3. Subspace Learning Experiment and Analysis. In this section, we employ an unsupervised subspace learning approach represented by Locality Preserving Projections (LPP) to verify the effectiveness of the proposed approach. In our experiments, different graphs are employed as W in LPP approach for subspace learning and then the classification accuracy is used for performance comparison. For each database, we randomly select l images from each class as the training samples. The remaining images are treated as the test samples. The values of l for Yale, AR, CMU PIE, and Extended YaleB databases are set as $\{4, 5, 6\}$, $\{4, 5, 6\}$, $\{6, 8, 10\}$, and $\{10, 15, 20\}$, respectively. In order to more effectively and fairly test the performance of the proposed approach, the random sample selection is repeated 20 times and the average classification accuracy and standard deviation are regarded as the final results for comparison. In this work, we employ the nearest neighbor classifier with Euclidean distance for classification due to its simplicity. To compare the performances of different approaches, the classification accuracy rate is chosen as the evaluation criterion, which is defined as

$$\text{Class\_accuracy} = \frac{N_{t\_\text{correct}}}{N_{\text{total}}} \times 100\%, \qquad (18)$$

where N_{t_correct} is the number of test samples which are correctly classified using the nearest neighbor classifier. N_{total} is the total number of the test samples.

All the experiments are conducted using MATLAB 2016b on a 3.60 Hz with 8 G RAM. In order to acquire the optimal parameters of different approaches, we employ the grid-search manner in our experiments. Tables 2–5 depict the average classification accuracy rates and standard deviations of the proposed approach on the Yale, AR, CMU PIE, and Extended YaleB databases, respectively. Note that the brackets in Tables 2–5 mean the data dimensionality when achieving the best classification accuracy rates.

From the results depicted in Tables 2–5, we can clearly observe that most of the graph learning approaches perform better than the KNN graph and LLE graph. It indicates that

Table 2: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the Yale database.

Method	$l = 4$	$l = 5$	$l = 6$
KNN+LPP	74.86 ± 2.70 (56)	77.56 ± 2.39 (72)	79.07 ± 3.27 (80)
LLE+LPP	77.71 ± 4.22 (56)	79.78 ± 1.55 (56)	80.13 ± 3.23 (80)
L1+LPP	81.52 ± 2.30 (44)	84.79 ± 2.19 (64)	85.07 ± 3.31 (80)
LRR+LPP	79.43 ± 2.70 (56)	82.33 ± 2.59 (72)	82.93 ± 4.06 (76)
LSR+LPP	84.29 ± 4.07 (56)	86.00 ± 2.58 (72)	86.40 ± 5.02 (80)
LCLSR+LPP	82.57 ± 3.84 (52)	85.56 ± 2.92 (72)	85.73 ± 4.71 (80)
SGLS+LPP	83.33 ± 4.15 (56)	85.33 ± 3.18 (72)	85.20 ± 3.74 (76)
RGSL+LPP	84.95 ± 4.12 (56)	86.89 ± 1.72 (64)	86.53 ± 4.14 (64)

Table 3: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the AR database.

Method	$l = 4$	$l = 5$	$l = 6$
KNN+LPP	60.76 ± 1.92 (320)	61.40 ± 1.99 (400)	60.11 ± 1.41 (400)
LLE+LPP	66.79 ± 2.28 (320)	67.56 ± 2.15 (380)	66.39 ± 3.17 (400)
L1+LPP	74.12 ± 1.90 (280)	74.96 ± 1.47 (360)	72.46 ± 2.64 (400)
LRR+LPP	77.66 ± 1.16 (380)	78.93 ± 0.95 (400)	82.56 ± 1.53 (400)
LSR+LPP	79.39 ± 1.80 (320)	80.14 ± 0.83 (400)	83.66 ± 1.43 (400)
LCLSR+LPP	80.23 ± 1.54 (380)	81.76 ± 1.01 (400)	85.36 ± 1.06 (400)
SGLS+LPP	79.43 ± 2.06 (380)	84.18 ± 0.97 (400)	85.35 ± 1.01 (400)
RGSL+LPP	83.28 ± 1.25 (380)	85.20 ± 1.06 (400)	90.54 ± 1.23 (380)

Table 4: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the CMU PIE database.

Method	$l = 6$	$l = 8$	$l = 10$
KNN+LPP	85.37 ± 0.81 (400)	86.56 ± 0.70 (400)	86.63 ± 0.99 (400)
LLE+LPP	87.55 ± 0.92 (400)	88.05 ± 0.91 (400)	87.95 ± 0.80 (400)
L1+LPP	88.95 ± 0.86 (380)	88.97 ± 0.73 (400)	88.09 ± 1.13 (380)
LRR+LPP	90.08 ± 0.37 (380)	90.18 ± 0.87 (400)	90.11 ± 0.99 (400)
LSR+LPP	89.90 ± 0.55 (400)	90.26 ± 0.81 (400)	90.21 ± 0.89 (320)
LCLSR+LPP	90.32 ± 0.44 (400)	90.68 ± 0.73 (400)	90.68 ± 0.83 (400)
SGLS+LPP	89.72 ± 0.68 (400)	90.08 ± 0.57 (280)	90.09 ± 0.93 (380)
RGSL+LPP	90.46 ± 0.81 (340)	92.72 ± 0.63 (380)	93.60 ± 0.48 (400)

Table 5: The average classification accuracy rates (%) and standard deviations (%) of different approaches on the Extended YaleB database.

Method	$l = 10$	$l = 15$	$l = 20$
KNN+LPP	65.38 ± 2.33 (360)	69.62 ± 1.12 (400)	64.49 ± 1.54 (400)
LLE+LPP	73.93 ± 1.71 (360)	75.00 ± 1.18 (400)	72.05 ± 1.17 (400)
L1+LPP	75.77 ± 4.10 (280)	75.97 ± 1.84 (340)	67.70 ± 1.69 (380)
LRR+LPP	82.67 ± 1.02 (360)	85.97 ± 0.71 (400)	84.53 ± 1.19 (380)
LSR+LPP	84.20 ± 0.66 (360)	86.19 ± 0.84 (360)	85.14 ± 0.71 (400)
LCLSR+LPP	85.59 ± 0.75 (360)	88.06 ± 0.75 (400)	87.62 ± 0.62 (380)
SGLS+LPP	84.83 ± 0.73 (360)	86.64 ± 0.73 (380)	87.09 ± 0.94 (340)
RGSL+LPP	86.14 ± 0.65 (360)	90.99 ± 0.48 (400)	92.68 ± 0.54 (400)

graph learning based on Euclidean distance is very sensitive to noise points, weakening the classification performance. Besides, compared to L1 graph learning approach, LRR graph, LSR graph, SGLS graph, and LCLSR graph take the locality structure of data into consideration during the pro-cess of graph construction to get more excellent perfor-mance. At last, the proposed RGSL approach performs best among all of the compared approaches. The main reasons are as follows: first, both the global structure and local struc-ture are essential to the graph learning. Second, $l_{2,1}$-norm

(a) Yale

(b) AR

(c) CMU PIE

(d) Extended YaleB
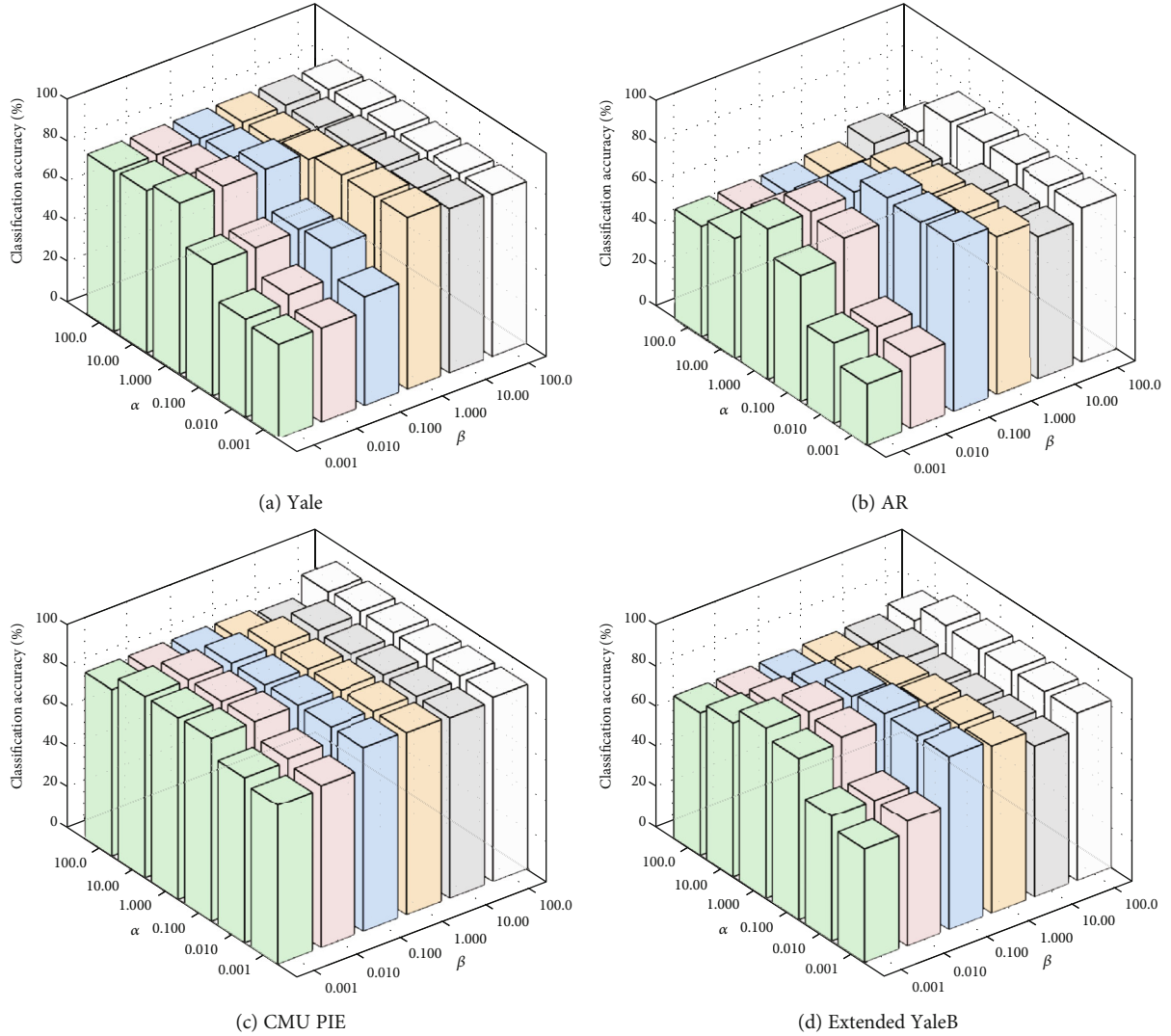
FIGURE 3: The classification accuracy rates of the proposed RGSL approach with different values of parameters α and β on the four different databases.

regularization criterion and nonnegative constraint are introduced into graph construction process to improve the robustness of our approach against noise. Therefore, our approach can improve the classification performance further.

There are two parameters, i.e., α and β in the objective function of our proposed approach. Hence, how to appropriately set their values is very important for our approach. In this study, we tune the values of parameters α and β by searching the grid $\{0.001, 0.01, 0.1, 1, 10, 100\}$ in an alternate manner. The best results of different parameter values on the four databases are shown in Figure 3.

As we can see from Figure 3, when the values of parameters α and β are relatively small, the performance of the proposed approach is relatively small. With the increase of parameters α and β, the performance of the proposed approach will be improved. However, after it achieves its best classification result, the performance of the proposed approach dramatically decreases with the increase of the two parameters. Therefore, the proposed approach can

obtain its best classification results when the values of parameters α and β are set as neither too large nor too small. At last, the convergence curves of our RGSL on the four databases are shown in Figure 4. In this figure, the x-axis and the y-axis are, respectively, denoted as the iteration number and the value of the objective function. As seen from Figure 4, the value of the objective function declines at each iteration and converges very fast on all of the databases.

4.4. Clustering Experiment and Analysis. In spectral clustering, the initialization has a major impact on the performance of the K-means clustering algorithm. Therefore, we carry out the process of clustering 50 times with different random initializations. Then, the average clustering results with standard deviations are used as the final results. In the experiments, three widely employed clustering evaluation indicators including Accuracy (ACC), Normalized Mutual Information (NMI), and Purity are used to evaluate the performance of the proposed approach.

(a) Yale

(b) AR

(c) CMU PIE
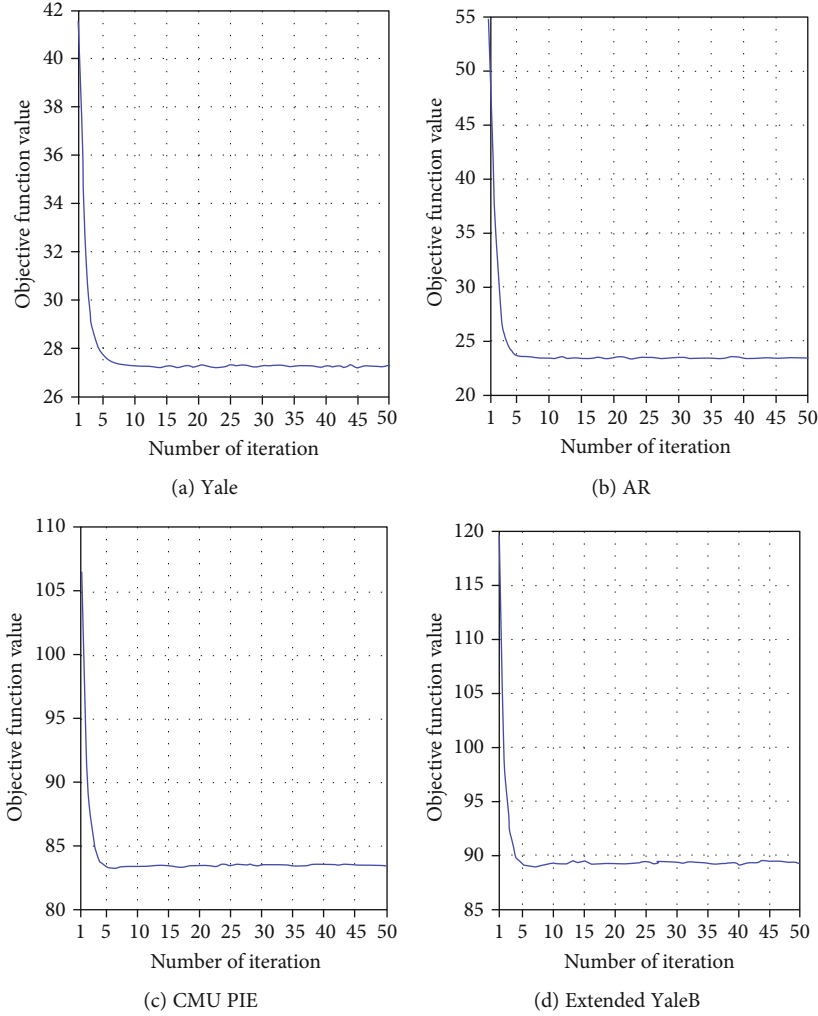
(d) Extended YaleB

FIGURE 4: The convergence curves of RGSL on the four databases: (a) Yale, (b) AR, (c) CMU PIE, and (d) Extended YaleB.

For a given sample x_i, supposing that the obtained clustering result is p_i and true label is t_i, the clustering accuracy is calculated as

$$ACC = \frac{\sum_{i=1}^{N} \delta(t_i, m(p_i))}{N}, \quad (19)$$

where $\delta(x, y) = 1$ if $x = y$, $\delta(x, y) = 0$ otherwise. The function $m(\cdot)$ maps the clustering result to the corresponding ground truth label. N is the number of samples. The Kuhn-Munkres algorithm [37] is employed to find the best mapping result.

Assuming that P and T are, respectively, the clustering result and the true label set obtained by different approaches, the Mutual Information (MI) is defined as

$$MI(P, T) = \sum_{p_i \in Q, t_i \in T} Q(p_i, t_i) \cdot \log_2 \frac{Q(p_i, t_i)}{Q(p_i) \cdot Q(t_i)}, \quad (20)$$

where $Q(p_i)$ and $Q(t_i)$ represent the probabilities that a sample is randomly selected from the dataset belonging to p_i and t_i, respectively. $Q(p_i, t_i)$ represents the joint probability of a sample randomly being selected from the dataset belonging to p_i and t_i.

Let $H(P)$ and $H(T)$ be the entropies of P and T, respectively. The Normalized Mutual Information (NMI) is calculated as

$$NMI = \frac{MI(P, T)}{\max(H(P), H(T))}. \quad (21)$$

Purity is defined as follows:

$$Purity = \frac{1}{k} \sum_{i=1}^{k} \frac{|C_i^d|}{|C_i|}, \quad (22)$$

where k represents the number of clusters, $|C_i^d|$ is the number of elements in the most numerous category in cluster C_i, and $|C_i|$ is the number of elements in cluster C_i.

Tables 6–9 show the best values of ACC, NMI, and Purity of eight approaches, respectively, on the Yale, AR, CMU PIE, and Extended YaleB databases. According to the results as shown in Tables 6–9, the following conclusions can be

TABLE 6: The clustering results obtained by different approaches on the Yale database.

Method	ACC	NMI	Purity
KNN	0.6887 ± 0.0173	0.7098 ± 0.0069	0.6753 ± 0.0155
LLE	0.7028 ± 0.0032	0.7147 ± 0.0048	0.7085 ± 0.0028
L1	0.6998 ± 0.0191	0.7396 ± 0.0145	0.7145 ± 0.0160
LRR	0.7320 ± 0.0119	0.7580 ± 0.0102	0.7358 ± 0.0111
LSR	0.7509 ± 0.0123	0.7790 ± 0.0092	0.7520 ± 0.0122
LCLSR	0.7782 ± 0.0220	0.7848 ± 0.0183	0.7865 ± 0.0202
SGLS	0.7353 ± 0.0284	0.7811 ± 0.0201	0.7550 ± 0.0219
RGSL	0.8532 ± 0.0243	0.8665 ± 0.0321	0.8563 ± 0.0304

TABLE 7: The clustering results obtained by different approaches on the AR database.

Method	ACC	NMI	Purity
KNN	0.3907 ± 0.0060	0.6556 ± 0.0032	0.4023 ± 0.0062
LLE	0.3908 ± 0.0072	0.6645 ± 0.0043	0.4090 ± 0.0064
L1	0.6451 ± 0.0123	0.8057 ± 0.0068	0.6816 ± 0.0096
LRR	0.6763 ± 0.0121	0.8182 ± 0.0082	0.7130 ± 0.0096
LSR	0.7236 ± 0.0165	0.8703 ± 0.0078	0.7547 ± 0.0144
LCLSR	0.8108 ± 0.0147	0.9050 ± 0.0052	0.8326 ± 0.0103
SGLS	0.8130 ± 0.0151	0.9186 ± 0.0056	0.8361 ± 0.0121
RGSL	0.8301 ± 0.0124	0.9243 ± 0.0036	0.8485 ± 0.0132

TABLE 8: The clustering results obtained by different approaches on the CMU PIE database.

Method	ACC	NMI	Purity
KNN	0.4825 ± 0.0118	0.7186 ± 0.0059	0.5244 ± 0.0082
LLE	0.5884 ± 0.0124	0.7520 ± 0.0049	0.6108 ± 0.0090
L1	0.6284 ± 0.0168	0.7484 ± 0.0091	0.6644 ± 0.0130
LRR	0.7805 ± 0.0193	0.8694 ± 0.0065	0.7992 ± 0.0151
LSR	0.8077 ± 0.0146	0.8774 ± 0.0049	0.8165 ± 0.0096
LCLSR	0.8159 ± 0.0114	0.8733 ± 0.0059	0.8262 ± 0.0084
SGLS	0.8276 ± 0.0147	0.8717 ± 0.0069	0.8380 ± 0.0109
RGSL	0.8399 ± 0.0095	0.8886 ± 0.0070	0.8490 ± 0.0081

obtained. First, since KNN graph and LLE graph are based on Euclidean distance, they are very sensitive to the noise points, outliers, and parameter values. So the clustering performance based on KNN graph and LLE graph is lower than that based on other compared approaches. Second, the performance of LRR graph, LSR graph, SGLS graph, and LCLSR graph is superior to that of L1 graph because of taking the locality structure of data into consideration during the process of graph construction. However, these objective functions are all based on l_2-norm, so it is very sensitive to the noise data.

TABLE 9: The clustering results obtained by different approaches on the Extended YaleB database.

Method	ACC	NMI	Purity
KNN	0.4195 ± 0.0074	0.5191 ± 0.0069	0.4311 ± 0.0061
LLE	0.4611 ± 0.0090	0.5460 ± 0.0070	0.4757 ± 0.0070
L1	0.4793 ± 0.0168	0.5981 ± 0.0087	0.5082 ± 0.0123
LRR	0.7146 ± 0.0147	0.7603 ± 0.0085	0.7181 ± 0.0139
LSR	0.7043 ± 0.0181	0.7496 ± 0.0102	0.7086 ± 0.0174
LCLSR	0.7157 ± 0.0226	0.7656 ± 0.0140	0.7186 ± 0.0219
SGLS	0.7220 ± 0.0235	0.7768 ± 0.0122	0.7256 ± 0.0226
RGSL	0.7402 ± 0.0276	0.7897 ± 0.0163	0.7487 ± 0.0194

Besides, the relationship between the representation coefficients is ignored in the sample reconstruction, i.e., similar original samples should generate similar coding vectors, weakening the effectiveness of graph learning. To overcome these disadvantages, our RGSL approach combines $l_{2,1}$-norm with manifold constraints on the coding coefficients to learn a locality and smoothness representation. Therefore, the performance of the proposed approach is superior to that of all of the comparison approaches.

Similar to the subspace learning experiment, we also tune the values of parameters α and β by searing the grid $\{0.001, 0.01, 0.1, 1, 10, 100\}$ in an alternate manner. From the objective function, we can learn that there are three terms. When the values of parameters α and β are set as small, the effectiveness of the second and third terms in the objective function will be weakened, and the role of the first term will be overemphasized. On the contrary, the second and third terms in the objective function will play a major role, reducing the effect of the first term. Therefore, the proposed RGSL approach can achieve the best performance when parameters α and β are set as moderate values, which is similar to the discussions of subspace learning.

5. Conclusions

This paper proposes a novel graph learning framework, named Robust Graph Structure Learning (RGSL) for effective multimedia data analysis. In order to preserve both local and global structures of data, we employ the data self-representativeness to capture the global structure and adaptive neighbor approach to describe the local structure. Furthermore, we also introduce the $l_{2,1}$-norm regularization criterion and nonnegative constraint into graph learning to improve the robustness of the model against noise. Extensive experimental results associated with subspace learning and clustering tasks show that the proposed approach performs better performance than the state-of-the-art graph learning approaches. Since our proposed approach will be affected by the graph construction when the dimensionality of data is high, in the future, we will take the dimensionality reduction, subspace learning, and graph learning into a united framework to address this issue.

Data Availability

The data are derived from public domain resources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Nonnegative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.

[2] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.

[3] W. Zhou, C. Wu, J. Wang, X. Yu, and Y. Yi, "Double regularized matrix factorization for image classification and clustering," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 1, Article ID 49, 2018.

[4] W. Wei, H. Dai, and W. Liang, "Regularized least squares locality preserving projections with applications to image recognition," *Neural Networks*, vol. 128, pp. 322–330, 2020.

[5] S. Affeldt, L. Labiod, and M. Nadif, "Spectral clustering via ensemble deep autoencoder learning (SC-EDAE)," *Pattern Recognition*, vol. 108, p. 107522, 2020.

[6] Y. Yi, Y. Chen, J. Wang, G. Lei, J. Dai, and H. Zhang, "Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization," *Signal Processing: Image Communication*, vol. 89, p. 115984, 2020.

[7] Z. Kang, C. Peng, Q. Cheng et al., "Structured graph learning for clustering and semi-supervised classification," *Pattern Recognition*, vol. 110, p. 107627, 2021.

[8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[9] F. Dornaika, L. Weng, and Z. Jin, "Structured sparse graphs using manifold constraints for visual data analysis," *Neurocomputing*, vol. 315, pp. 107–114, 2018.

[10] C. Tang, M. Bian, X. Liu et al., "Unsupervised feature selection via latent representation learning and manifold regularization," *Neural Networks*, vol. 117, pp. 163–178, 2019.

[11] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with ℓ^{1}-graph for image analysis," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 858–866, 2010.

[12] P. Chen, L. Jiao, F. Liu, Z. Zhao, and J. Zhao, "Adaptive sparse graph learning based dimensionality reduction for classification," *Applied Soft Computing*, vol. 82, p. 105459, 2019.

[13] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *2011 International Conference on Computer Vision*, pp. 471–478, Barcelona, Spain, November 2011.

[14] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proceedings of the 27-Th International Conference on Machine Learning*, pp. 663–670, Haifa, Israel, 2010.

[15] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *European Conference on Computer Vision*, pp. 347–360, Springer, 2012.

[16] X. Peng, J. Feng, S. Xiao, W. Y. Yau, J. T. Zhou, and S. Yang, "Structured autoencoders for subspace clustering," *IEEE Transaction on Image Processing*, vol. 27, no. 10, pp. 5076–5086, 2018.

[17] Y. Chen, L. Zhang, and Z. Yi, "Subspace clustering using a low-rank constrained autoencoder," *Information Sciences*, vol. 424, pp. 27–38, 2018.

[18] K. G. Dizaji, A. Herandi, C. Deng, W. Cai, and H. Huang, "Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 5747–5756, Venice, Italy, October 2017.

[19] W. Zhou, Y. Yi, J. Bao, and W. Wang, "Adaptive weighted locality-constrained sparse coding for glaucoma diagnosis," *Medical & Biological Engineering & Computing*, vol. 57, no. 9, pp. 2055–2067, 2019.

[20] L. Zhang, W. Lu, X. Liu, W. Pedrycz, and C. Zhong, "Fuzzy c-means clustering of incomplete data based on probabilistic information granules of missing values," *Knowledge-Based Systems*, vol. 99, pp. 51–70, 2016.

[21] H. Zuo, G. Zhang, W. Pedrycz, V. Behbood, and J. Lu, "Fuzzy regression transfer learning in Takagi-Sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 6, pp. 1795–1807, 2017.

[22] W. Zhou, S. Qiao, Y. Yi, N. Han, Y. Chen, and G. Lei, "Automatic optic disc detection using low-rank representation based semi-supervised extreme learning machine," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 1, pp. 55–69, 2020.

[23] Z. Shu, X. Wu, C. Hu, C. Z. You, and H. H. Fan, "Deep semi-nonnegative matrix factorization with elastic preserving for data representation," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 1707–1724, 2021.

[24] W. Zhou, C. Wu, D. Chen, Y. Yi, and W. du, "Automatic microaneurysm detection using the sparse principal component analysis-based unsupervised classification method," *IEEE Access*, vol. 5, pp. 2563–2572, 2017.

[25] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2019.

[26] S. Huang, Z. Kang, I. W. Tsang, and Z. Xu, "Auto-weighted multi-view clustering via kernelized graph learning," *Pattern Recognition*, vol. 88, pp. 174–184, 2019.

[27] P. Jing, Y. Su, Z. Li, J. Liu, and L. Nie, "Low-rank regularized tensor discriminant representation for image set classification," *Signal Processing*, vol. 156, pp. 62–70, 2019.

[28] J. Wang, F. Wang, C. Zhang, H. C. Shen, and Long Quan, "Linear neighborhood propagation and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 9, pp. 1600–1615, 2009.

[29] J. Wei and H. Peng, "Neighbourhood preserving based semi-supervised dimensionality reduction," *Electronics Letters*, vol. 44, no. 20, pp. 1190–1192, 2008.

[30] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *Advance in Neural Information Processing Systems, NIPS*, pp. 2223–2231, 2009.

[31] D. Arpit, G. Srivastava, and Y. Fu, "Locality-constrained low rank coding for face recognition," in *International Conference on Pattern Recognition*, pp. 1687–1690, Tsukuba, Japan, 2012.

[32] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognition*, vol. 47, no. 9, pp. 2794–2806, 2014.

[33] Y. Chen and Z. Yi, "Locality-constrained least squares regression for subspace clustering," *Knowledge-Based Systems*, vol. 163, pp. 51–56, 2019.

[34] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[35] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern Recognition*, vol. 41, no. 1, pp. 176–190, 2008.

[36] K. Song, X. Yao, F. Nie, X. Li, and M. Xu, "Weighted bilateral K-means algorithm for fast co-clustering and fast spectral clustering," *Pattern Recognition*, vol. 109, p. 107560, 2021.

[37] H. Zhu and M. C. Zhou, "Efficient role transfer based on Kuhn-Munkres algorithm," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 2, pp. 491–496, 2012.

[38] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.

[39] A. Martinez, *The AR Face Database*, CVC, Luxembourg City, Luxembourg, 1998, Technical Report 24.

[40] Terence Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.

[41] Y. Yi, C. Bi, X. Li, J. Wang, and J. Kong, "Semi-supervised local ridge regression for local matching based face recognition," *Neurocomputing*, vol. 167, pp. 132–146, 2015.

[42] F. Dornaika and L. Weng, "Sparse graphs with smoothness constraints: application to dimensionality reduction and semi-supervised classification," *Pattern Recognition*, vol. 95, pp. 285–295, 2019.

WILEY | Hindawi

Research Article

Reliable Reputation Review and Secure Energy Transaction of Microgrid Community Based on Hybrid Blockchain

Zilong Song, Xiaohong Zhang (ID), and Miaomiao Liang

School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China

Correspondence should be addressed to Xiaohong Zhang; xiaohongzh@jxust.edu.cn

A growing number of prosumers have entered the local power market in response to an increase in the number of residential users who can afford to install distributed energy resources. The traditional microgrid trading platform has many problems, such as low transaction efficiency, the high cost of market maintenance, opaque transactions, and the difficulty of ensuring user privacy, which are not conducive to encouraging users to participate in local electricity trading. A blockchain-based mechanism of microgrid transactions can solve these problems, but the common single-blockchain framework cannot manage user identity. This study thus proposes a mechanism for secure microgrid transactions based on the hybrid blockchain. A hybrid framework consisting of private blockchain and consortium blockchain is first proposed to complete market transactions. The private blockchain stores the identifying information of users and a review of their transactions, while the consortium blockchain is responsible for storing transaction information. The block digest of the private blockchain is stored in the consortium blockchain to prevent information on the private blockchain from being tampered with by the central node. A reputation evaluation algorithm based on user behavior is then developed to evaluate user reputation, which affects the results of the access audit on the private blockchain. The higher a user's reputation score is, the more benefits he/she can obtain in the transaction process. Finally, an identity-based proxy signcryption algorithm is proposed to help the intelligent management device with limited computing power obtain signcryption information in the transaction process to protect the transaction information. A system analysis showed that the secure transaction mechanism of the microgrid based on the hybrid blockchain boasts many security features, such as privacy, transparency, and imtamperability. The proposed reputation evaluation algorithm can objectively reflect all users' behaviors through their reputation scores, and the identity-based proxy signcryption algorithm is practical.

1. Introduction

The Energy Internet (EI), a distributed sharing network that combines the Internet and distributed energy resources (DER), can connect many kinds of distributed energy nodes to achieve the two-way flow of energy. Energy is used to provide light, heat, power, and other necessities to human beings. With continual scientific and technological progress, a variety of devices are now available to easily convert electric energy into various kinds of energy needed for human production and living. Therefore, the two-way flow of electric energy will form the core of future research on EI.

Currently used forms of primary energy include fossil energy, light energy, wind energy, and water energy [1], whereas electric energy needs to be obtained through conver-

sion from primary energy. The traditional method of conversion is thermal power generation, that is, generating electricity through the combustion of fossil fuels. However, this method is inefficient and causes serious environmental pollution. In 2010, carbon dioxide emissions from energy production, such as the production of electricity, accounted for 76% of global emissions [2]. Considering the importance of environmental protection, research on new methods of conversion has gained momentum. Renewable energy sources (RES) such as light, wind, and water are widely used in the world through primary energy conversion devices. By the end of 2018, the installed capacity of hydropower in China was 352 GW, that of wind power was 184 GW, and that of solar power was 174 GW [3]. In addition, as the number of residential and industrial users who can afford DER

deployment, in the form of solar photovoltaic panels, biomass generators, microwind turbines, and diesel engines, grows each year, a growing number of DER are being deployed at the industrial and residential scales [4]. Although DER has the characteristics of low loss, little pollution, and good system economy, it still has problems that need to be solved. First, the distributed generator (DG) that uses the RES for power generation has a small capacity and is limited by external conditions, because of which the electricity generated by it is intermittent and random. This significantly reduces the reliability of the power supply [5]. Moreover, when a large number of invisible and uncontrollable power generated by DER directly flow into the power grid, the overall power supply line is prone to overshooting the power flow, which jeopardizes the safety and reliability of the power system [6]. Finally, the relationship between supply and demand in the power market is a major obstacle to the development of the DER, and consumers' acceptance of DER power generation needs to be considered.

To solve the above problems of DER, two technologies have been proposed: the virtual power plant (VPP) and the microgrid (MG) [7]. The VPP leverages advanced coordinated control technologies, smart metering technologies, and information and communication technologies to interact with participants in EI, thus making full use of the large-scale and multiregional DER. Due to the limitation of the available power transmission technology, long-distance power transmission causes partial power loss. For industrial and residential users who have DG installed, close-range MG technology is a better choice. MG focuses on regional balance of distributed load and power supply to achieve energy autonomy. VPP focuses on realizing the maximum benefit of the main body and has the derivative function of participating in the power market and auxiliary service market [8]. Liu et al. [9] have provided a distributed robust energy management scheme for a system composed of multiple MGs. Uncertain factors in the operation of the MG have been dealt with by tunable robust optimization technology to optimize the total operating cost of the MG, and studies have verified the effectiveness of the method in a four-MG system. Zhang et al. [10] proposed a networked physical–social system for DER management in the MG that has the capability of parallel learning and can promote the emergence of high-quality DER optimization strategies through human–computer interactive learning. A case study was used to show that this technique can yield a DER optimization strategy more quickly than other heuristic algorithms. Ranjbar et al. [11] proposed an MG protection method in which the short-time Fourier transform (STFT) is used to pretreat the voltage waveform within a period, and the features of disturbance are extracted accordingly. These features are fed to a decision tree algorithm to identify fault events in the MG. The results of simulations showed that depending on the type of event, only two or six features were needed to detect any fault.

The above literature has mainly focused on solving the technical and economic problems of the MG, but it needs to be further developed to solve issues with its management. The prevalent mode of energy operation mostly uses centralized third-party management organization to manage trans-

actions. This mode of management has the following problems: First, with an increase in the number of DER transactions within its jurisdiction, the operating cost of the trading center increases, transaction efficiency is significantly reduced, and it is difficult to ensure the effective operation of the microgrid in real time. Second, in the energy trading process, the trading center and the trading side cannot achieve complete trust, which imposes a significant annual cost on the trading centers to maintain trust. Moreover, there is no open and transparent trading and information platform in the MG, because of which the security and effectiveness of the transaction cannot be guaranteed, and its cost is high. Finally, the centralized trading center is prone to a single point of failure; that is, the trading center causes the entire system to collapse once it is attacked, and the disclosure or tempering of trading information damages the property and violates the privacy of both parties to the transaction.

Since 2016, Bitcoin, a decentralized digital currency, has gained considerable attention from the financial community due to an increase in its economic value. Academics have found that in addition to the economic value of Bitcoin itself, its core supporting technology, namely, the blockchain, has significant research value. The blockchain has the characteristics of decentralization, trustlessness, openness, and transparency. With progress in research, the scope of applications of the blockchain is no longer limited to the financial field. Adding blockchain technology to the transaction process of the MG may provide a new solution to the abovementioned management problems. Research on combining the microgrid energy market with blockchain technology is still in its preliminary stage. To prove the feasibility of this combination, many scholars have carried out a series of studies, and the results show that the blockchain has the ability to support energy transactions within a certain range [12–14]. Based on this assessment of theoretical feasibility, a growing number of papers have been published in the area. Di Silvestre et al. [15] discussed the loss in the distribution of energy transactions of blockchains when applied to the MG and proposed two indicators of loss distribution to solve this problem. The feasibility of these indicators was verified in two operating scenarios of a medium-voltage microgrid. Di Silvestre et al. [16] considered the provision of voltage regulation technology based on the blockchain for the MG, mainly by solving for reactive power optimization power flow and reactive power compensation. The former was intended to ensure optimal economic planning in reactive power production and the latter to evaluate the contribution of voltage regulation. Hassan et al. [17] proposed an energy transaction auction mechanism called differential privacy auction to provide moderately costly but secure and private energy auctions for the MG based on consortium blockchain. Experimental comparisons showed that this mechanism was superior to the VCG mechanism. van Leeuwen et al. [18] designed an integrated energy management platform based on the blockchain that is composed of three parts: a physical layer, economic layer, and information layer. It can facilitate the trade of energy in the microgrid community through a bilateral transaction mechanism and optimize energy flow by solving optimal power flow problems. Meeuw et al. [19]

studied the impact of limitations of hardware and the communication infrastructure of applications on the blockchain system. Based on the conditions of the Swiss blockchain-based Walenstadt microgrid, the researchers artificially adjusted the bandwidth between nodes to simulate the bandwidth of the communication infrastructure. They found that a communication network with a bandwidth of less than 1000 kbit/s leads to insufficient system throughput. To solve the problems of default risk and demand uncertainty in designing a renewable energy microgrid based on the blockchain, a method based on robust two-type fuzzy programming was proposed by Tsao et al. [20], and its effectiveness was proved by a case study. The above literature has examined the blockchain-based microgrid system from different technical aspects, but a safe method to protect energy transactions in the MG remains elusive.

In this paper, a secure microgrid transaction mechanism based on the blockchain is proposed. The main contributions are as follows:

(1) Blockchain-based microgrid trading platforms can solve the problems of trust and transparency in microgrid energy trading, but most schemes proposed in the literature are based on a single blockchain. In application, a single blockchain struggles to provide effective user identity management, and this makes it easier for malicious actors to infiltrate the system. This paper proposes a microgrid energy transaction framework based on the hybrid blockchain containing a trading consortium blockchain and N private blockchains for identity management, where N is the number of microgrids in the network. Only users verified by the private blockchain can conduct transactions on the consortium blockchain

(2) To ensure good market trading behavior, a reputation evaluation algorithm based on user behavior is proposed. Because there are two kinds of identities, buyer and seller, in energy trading, this algorithm contains separate algorithms to assess buyer and seller behaviors. Whether a user can be authenticated by the private blockchain depends on their own reputation: when the reputation has a score of zero, the user cannot use the energy transaction function. In addition, the energy in the consortium blockchain is mainly auctioned by using the continuous double auction algorithm based on reputation. The higher the reputation score of a user is, the more benefit from the transaction they can draw

(3) When users participate in energy transactions, they need to communicate with the microgrid continuously. To ensure the security of the information shared during transaction-related communication, an identity-based proxy signcryption algorithm is proposed that is suitable for users with smart home manager. Proxy signcryption allows the smart home manager with a limited amount of computing power and storage to delegate its data processing rights to the powerful energy manager to participate in energy

trading. The identity-based proxy signcryption algorithm solves the defect whereby the typical proxy signcryption algorithm needs to store a large number of certificates

The remainder of this paper is arranged as follows: Section 2 introduces some preliminary information, and the system as a whole and its detailed framework are introduced in Section 3. In Section 4, we describe the steps of implementation of the proposed scheme, such as details of the buyer and seller reputation evaluation algorithms, the identity-based proxy signcryption process, and the process of generation of new blocks. Section 5 is devoted to a performance analysis and evaluation of the proposed scheme, and we summarize our findings in Section 6.

2. Preliminaries

In this section, we review some preliminary knowledge, such as the structure of the microgrid, the nature of the bilinear pairing involved, and the principle of proxy signcrytion.

2.1. Microgrid. The earliest concept of the microgrid was proposed by the United States Consortium for Electric Reliability Technology Solutions (CERTS) [21] and remains the most authoritative one. The CERTS microgrid assumes that the set of loads and DER operate as a single system. A critical feature is that it can autonomously exist in the distribution system as a self-controlling entity. In other words, it is impossible to distinguish the MG from legitimate customer sites in the grid. The initial work by the CERTS was based on small-scale micropower sources with a capacity lower than 500 kW, and the basic structure of the MG developed by the CERTS is shown in Figure 1.

The power system in the diagram consists of three feeders (A, B, C) and a set of loads. The entire power supply network has a radial shape. In terms of load distribution, both feeders A and B contain sensitive loads while feeder C contains traditional loads. The installation of micropower supply is based on the user's load demand and only in case of sensitive loads are the microsources installed on the feeder. This system contains two kinds of microsources for installation: a microturbine and a fuel cell. The microturbine is installed on feeder A, and two kinds of microsources are installed on feeder B. The fuel cell can produce electric energy as well as a large amount of heat energy when burning, which endows feeder B with the ability of cogeneration. To adjust the power and feeder flow of the microsources, each is equipped with a power and voltage controller managed by the energy manager of the MG or the local system of users.

In addition, the microgrid shown in the figure has two operating modes: the networking mode and the island mode [21]. When the distribution network is disturbed, feeders A and B can use a separation device to separate themselves from the power grid, thereby minimizing interference in the inductive load. If local power generation is not sufficient to meet the demand of sensitive loads, the islands are rendered meaningless. Feeder C is left in the interference, mainly to eliminate the interference trip caused by traditional loads.
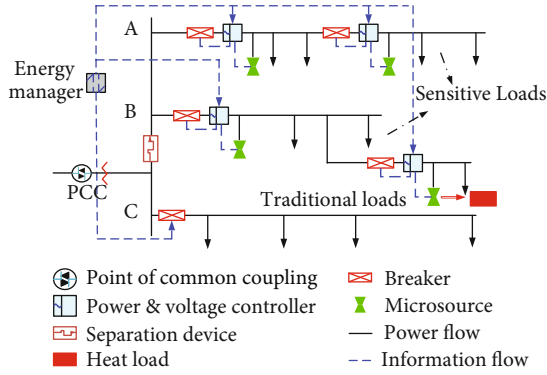
FIGURE 1: The basic structure of the microgrid proposed by CERTS.

With the development of power electronic equipment, microsources that can be installed in the MG are no longer limited to the two mentioned above, and the micropower supply based on clean energy, such as solar and wind energy, has been widely incorporated into the MG. The control strategy of the MG has improved, and the most commonly used one is hierarchical control. The hierarchical control strategy is divided into three layers. The first layer consists of DER and the local self-control of loads, the second layer consists of the management control of the MG, and the third layer features distribution network management control. Through the hierarchical architecture, the electrical magnitudes of the MG at different time scales can be controlled.

2.2. Bilinear Pairing. Suppose there are three cyclic groups G_1, G_2, and G_3. The order of the cyclic group is p, and the generator of the cyclic group is g. Based on bilinear pairing, there is a mapping relationship among these three cyclic groups called $e : G_1 \times G_2 \longrightarrow G_3$ that satisfies the following properties:

(i) Bilinearity: For any generators $g_1 \in G_1$, $g_2 \in G_2$, and $x, y \in Z_p^*$, there always exists $e(xg_1, yg_2) = e(g_1, g_2)^{xy}$

(ii) Nondegeneracy: There always exists $g_1 \in G_1$, $g_2 \in G_2$, such that $e(g_1, g_2) \neq 1$

(iii) Computability: There is an algorithm that renders e (g_1, g_2) computable under the condition $\forall g_1 \in G_1$, $g_2 \in G_2$

This is an asymmetric bilinear pairing. The commonly used bilinear pairing is symmetric, that is, $G_1 = G_2$. Furthermore, the bilinear pairing commonly referred to is based on the prime order. The composite-order bilinear pairing proposed by Boneh et al. [22] is still undergoing improvement. For a detailed classification of bilinear pairing, the interested reader can see Ref. [23].

2.3. Proxy Signcrytion. In 1999, Gamage et al. [24] proposed proxy signcryption as a cryptographic primitive that is generated on the basis of signcryption [25] and the proxy signature [26] and inherits the characteristics of both. In the proxy signcryption scheme, the owner of the original data can entrust the authority for processing them to a person, that is, the proxy signcrypter; then, the agent can replace the owner of the original data to perform the signcryption operation. In a computation-constrained smart device, it would be a significant burden for the intelligent device to constantly consume computing power to perform signcryption. Proxy signcryption can solve this problem. It entrusts the signcryption operation to the agent with strong computing power to relieve the computing pressure on intelligent devices. The proxy signcryption scheme consists of five parts: system setup, key generation, delegation generation, proxy signcrypt, and proxy unsigncrypt. The process is as follows:

System Setup: Given a security parameter λ, the algorithm outputs the system parameter params.

Key Generation: Given system parameters params, the algorithm outputs the public/private key pairs (PK_o, SK_o) of the original data owner, pair (PK_p, SK_p) of the proxy signcrypter, and the pair (PK_r, SK_r) of the message receiver.

Delegation Generation: Given system parameters params, the data owner's private key is SK_o, and the warrant is ω; the algorithm outputs a delegation σ_ω and sends (ω, σ_ω) to the proxy signcrypter.

Proxy Signcrypt: Given system parameters params, the warrant ω, the delegation σ_ω, the proxy signcrypter's private key SK_p, the receiver's public key PK_r the message M, and the algorithm output ciphertext σ.

Proxy Unsigncrypt: Given system parameters params, warrant ω, ciphertext σ, the receiver's private key SK_r, data owner's public key PK_o and the proxy signcrypter's public key PK_p, if the ciphertext σ is legal, the algorithm outputs message M, otherwise, outputs the error symbol $\perp$.

3. System Model

In this section, we laid out the structure of the hybrid blockchain, where the private blockchain is responsible for identity authentication, and the consortium blockchain is responsible for energy transactions.

3.1. Data Storage and Sharing Model of MG Based on Hybrid Blockchain. The blockchain can be divided into public blockchain, consortium blockchain, and private blockchain according to the different modes of participation of its nodes [27–29]. The public blockchain allows all nodes to participate in the network and has the highest security. However, its deployment comes at the cost of a large amount of resources, and its characteristics of low extensibility and weak data throughput do not support the application of the public blockchain to commercial transactions involving large amounts of data. The private blockchain has the disadvantage of too high a degree of centralization, which renders it suitable only for information sharing within a single entity but not for storing transaction information involving multiple entities. The consortium blockchain is a compromise between the public and the private blockchains. It retains their advantages and is free of their major disadvantages. The consortium blockchain is the most commonly used blockchain in applications.

In the design considered in this paper, we use the consortium blockchain as the transaction blockchain to store

transaction data and call it the transaction consortium block-chain (TCB). In contrast to other literatures, we set up an identity chain outside the trading chain to authenticate and manage the identities of traders. During the operation of the MG, an MG community can be regarded as an entity, and transactions are usually carried out within it so that each MG community can establish its own identity chain to manage traders. Based on the characteristics of operation of the MG, we use the private blockchain as the identity chain of managing traders and call it the identity private blockchain (IPB). Therefore, the overall framework of this paper is a 1 + N hybrid blockchain framework; that is, it contains a transaction consortium blockchain and N identity private blockchains, and N is determined by the number of MGs. The framework is shown in Figure 2.

As is shown in Figure 2, the MG community communicates with the authority for verification off the blockchain. Once the verification has passed, each MG community generates an exclusive IPB that is responsible for authenticating the user's identity, assessing their reputation, and storing the corresponding identity information. When users in the MG community want to conduct energy transactions, those participating in the transactions need to communicate with the MG and send their demands to it. According to the different needs of each user, participants in the transaction can be divided into prosumers and consumers. Prosumers are users that have DER installed and can sell their surplus electricity. Consumers are regular users who need to purchase electricity to meet their needs. When the total remaining electricity among prosumers in the MG is not enough to meet the demand of the consumers, the MG purchases the required electricity from the power trading center. When the total amount of electricity left over by prosumers in the MG exceeds the total electricity demand of the community and the storage is full, the MG sells the excess electricity to the power trading center. Because power transactions involve the transfer of user property, all transactions occur on the TCB, and, accordingly, transaction data are stored on it. The TCB is jointly maintained by all network nodes and has only a TCB, which can provide adequate security for the TCB. To ensure that users have acceptable market trans-action behavior, the MG uses the reputation-based auction algorithm to auction energy. The user reputation required by the auction is stored in the IPB while the algorithm used to assess the user's reputation requires the user transaction data stored on the TCB. Based on this scenario, we allow the TCB and IPB to interact with each other through smart contracts. The degree of centralization in a private block-chain is too high, and the data stored on it are at risk of being tampered with by the central node. Therefore, we store the hash digest of the IPB on the TCB to ensure the security of data in the private blockchain by relying on the security of the consortium blockchain.

3.2. Identity Authentication Model Based on IPB.

When a user wants to make an energy transaction, he needs to be authenticated on the IPB. Only an authenticated user can obtain the transaction license; otherwise, he cannot use the MG power transaction platform. Community users who are new to the platform first need to register their identity so that they can join the IPB. The process of building and joining the IPB is shown in Figure 3.

Figure 3 shows seven steps, which 1–3 show of the construction of the IPB and 4–7 show the joining process of nodes. In the construction process, in step 1, the MG manager sends an IPB build request to the authority that contains the identity of the MG community, maximum power limit, and the jurisdiction to be divided. After receiving the MG's application, the authority reviews it and, after approving it, returns the information to the MG and invites it to build an IPB exclusively for its community, as in step 2. In step 3, the MG and the authority jointly set up a private blockchain. Once the IPB has been built, users of the community can apply to join the private blockchain. In the first step of the joining process, namely, step 4, community users need to send registration information to the authority. Different from consumers, the registration information of prosumers contains their identity information as well as detailed information on the deployed DER. Upon receiving the registration information from community users, the authority verifies the information, and if verified, authority will encrypt and upload the information to the private blockchain of the community to which the given user belongs. After that, the authority will generate the exclusive key for the user and distribute it. Steps 5 and 6 show this process. As shown in step 7, when a user receives a private key, they can officially join the IPB to which they belong according to this key.

Successfully joining the IPB does not mean that energy transactions can be conducted on the MG energy trading platform. Energy transactions can be officially conducted only on the TCB after obtaining a trading license on the IPB. To obtain a trading license, a user's identity must be authenticated out on the IPB. A flowchart of identity authentication is shown in Figure 4.

The system first checks whether the user node ID exists in the blacklist BL; if it does, the user is denied the use of the energy transaction function and can otherwise continue to the next step. The user node then selects the roles it plays on the TCB, where only seller and buyer roles are available. When it chooses to be a seller, the system checks whether the user node has a DER certified by the given authority. If not, the system returns to the previous step, and the user node reselects its role. In this case, the system checks the seller reputation value R_j of the user to whom the node belongs. If the reputation value is zero, the node reenters the role selection process. If the reputation value is not zero, the node is issued a seller license. When selecting a buyer, the system needs to only check the buyer reputation value R_i of the user to whom the node belongs. If the reputation value is zero, the user ID is added to the BL, and the user node is denied energy transactions. If the reputation value is not zero, a buyer license is issued to the user node. Finally, users with seller or buyer licenses can use them to trade energy on the TCB. Prosumers can choose to trade as either sellers or buyers. So, in each round of trading, the IPB provides identity authentication for each user only once. After the transaction, the system recalculates the reputation score of each node based on its performance in the transaction. All new nodes
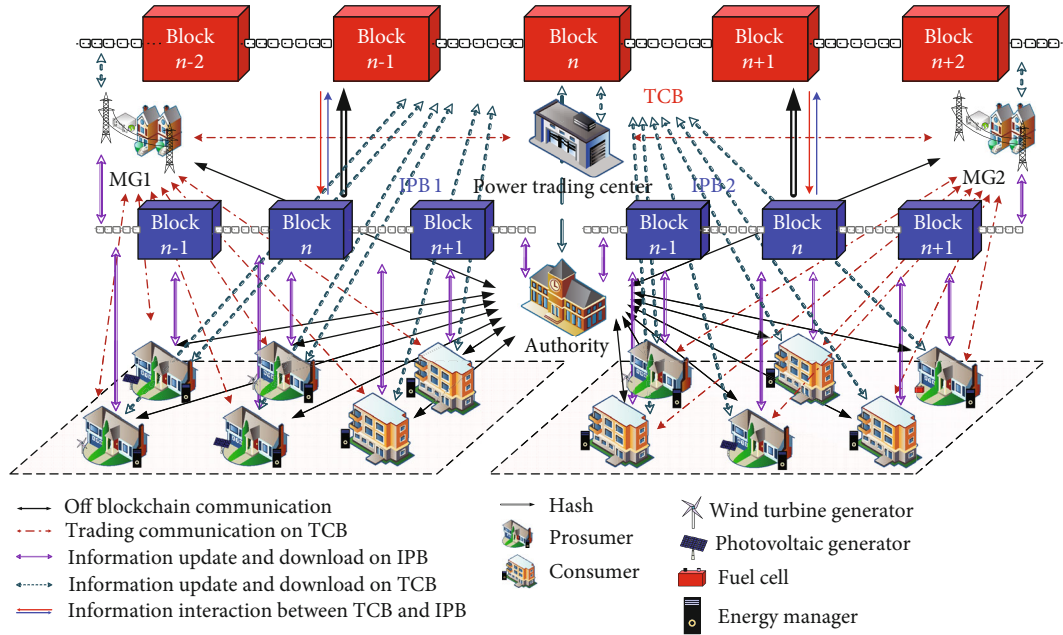
FIGURE 2: Data storage and sharing model of microgrid based on the hybrid block chain.
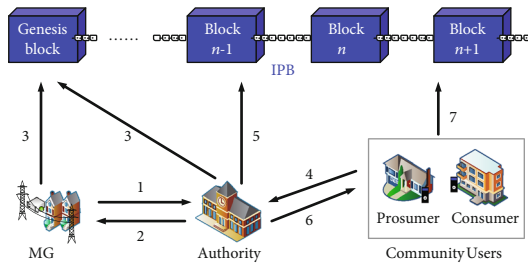

FIGURE 3: Flowchart of building and joining the IPB.

are assigned the same initial reputation score by the system, and each node has the reputation scores of the seller and the buyer, but only prosumers can use the former score. Prosumers with a seller reputation of zero cannot conduct energy transactions on the TCB as sellers and can only purchase energy as ordinary consumers. Users in the blacklist cannot reregister with the authority to obtain a new node identity. The means of obtaining another transaction authority is formulated by the given authority.

3.3. Secure Trading Model Based on TCB. Forms of DER energy trading on the MG energy trading platform can be divided into two types: P2P trading and centralized clearing. A P2P transaction is a direct transaction between individuals that is executed automatically according to the corresponding contract. However, such a transaction is disordered and can easily affect the power system. Centralized clearing requires a third-party platform for unified trading under optimized scheduling to reduce the risk of system disorder. Due to the problem of trust of the third-party platform, the mode of transaction of centralized clearing has been criticized. The emergence of the blockchain provides a new

solution to the problem of trust. The TCB uses continuous double auction (CDA) as method for transactions, where this is a kind of centralized clearing. The CDA allows both parties to a contract to modify their quotations continuously during the auction to maximize the interests of the traders [30]. The secure transaction model is shown in Figure 5.

The traditional CDA has only three trading entities: the auctioneer, buyer, and seller. In our model, MG is the auctioneer, the prosumer is the seller, and the consumer is the buyer. Besides, the model features another power trading center acting as an energy balancer. The greater the number of nodes in the blockchain network is, the stronger the security of the blockchain is. For security-related reasons, we should deploy as many nodes as possible in the blockchain network. The block is the basic unit in the formation of a blockchain and collection of data. Each block is composed of a header and a body [31]. The block header contains a transaction information hash, a block hash, and a time stamp while detailed transaction information is stored in the block body. The byte size of the block header is smaller than that in the block body. Therefore, nodes that store only block header information are called light nodes, and those that store information on the entire block are called full nodes. For the smart home manager or the DER with limited computing and storage capabilities, being a light node in the blockchain network can not only reduce its own storage pressure but also enhance the security of the network. As for the problem of limited computing power, the pressure can be shared by an energy manager with powerful computing power, which exists in the form of a full node in the network. As a result, the buyers in this scheme can be accurate to smart home managers such as smart appliances, smart lighting and smart windows, and doors, while the sellers are the DER devices deployed by prosumer.
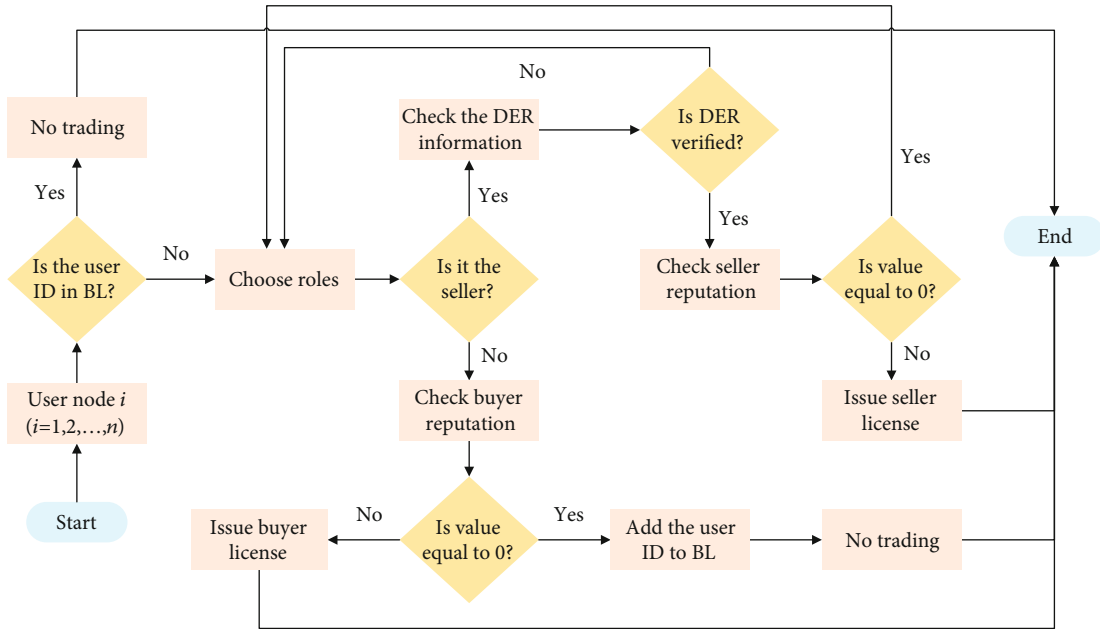
FIGURE 4: Flowchart of identity authentication.
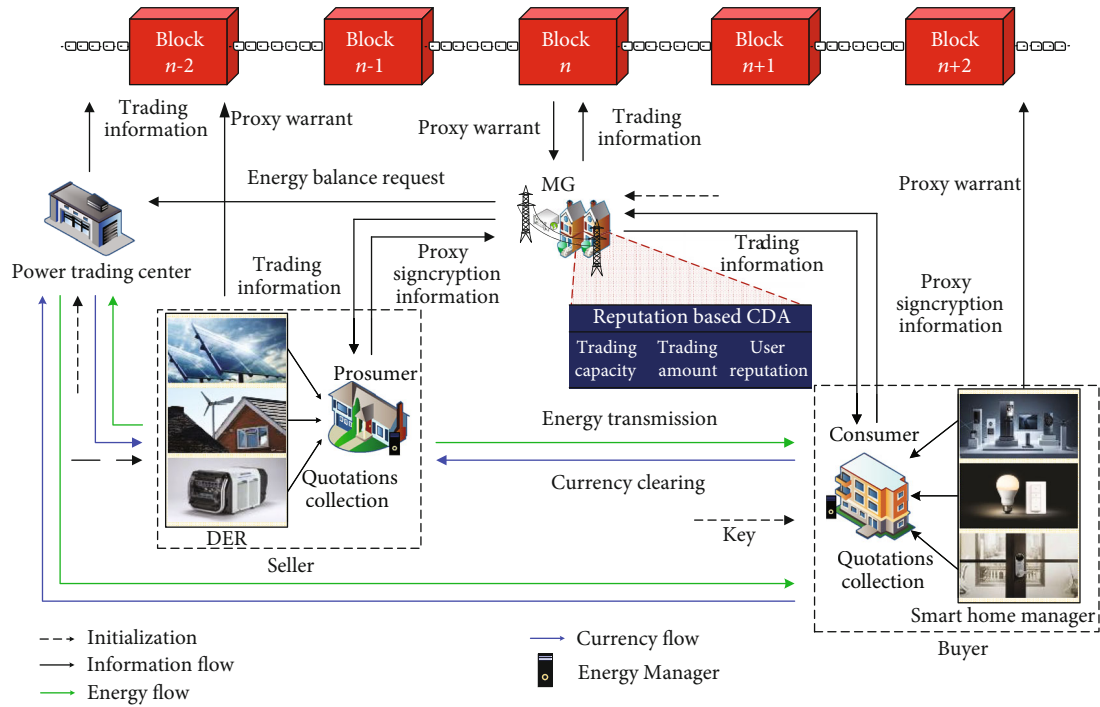
FIGURE 5: Secure trading model based on the TCB.

The basic flow of the secure trading model is as follows:

(1) Initialization: Once users of the community have passed IPB authentication, they can conduct identity transactions on the TCB according to their identity licenses. Users with the seller's license act as sellers, and those with the buyer's license act as buyers. The MG manager node acts as auctioneer, and the keys required for the transactions are generated by the authority

(2) Quotation collection: Although the smart home manager or DER, as a light node, can generate the quotation and signcrypt it to the MG manager node by itself, frequent signcrypts require a large amount of computing power. In addition, multiple energy quotes may belong to the same entity in the auction list, which increases the workload of the auction and causes unnecessary waste. As a full node, the energy manager can first integrate the quotation

information of the user's home manager, signcrypt this information through the identity-based proxy signcryption algorithm, and send the signcrypted information to the MG manager node. In this way, the effect of the light node's participation in the transaction can be obtained and problems incurred by this participation can be solved

(3) Energy auction: After receiving the signcryption information from buyers and sellers, the MG manager node decrypts them. Auction matching is then carried out according to the bidding price. To increase the user's attention to reputation, the CDA auction mechanism based on reputation is used for auctioning; it divides users' grades according to their scores. The higher the grade is, the wider is the range of options to which the corresponding user can match. The matching rule of "price first, reputation first, time first" is used. At the end of the auction, if energy balance has not been attained within the MG community, an energy transaction is conducted with the power trading center as is appropriate

(4) Transaction and clearing: Community users who have been successfully matched check the transaction information and then conduct energy transaction according to the confirmed transaction contract. During the transaction, the default users are punished financially, and default behaviors will lead to a decline in their reputation scores. Transaction clearing needs to be carried out through the unique energy coin of the system that is generated by the power trading center. When each user joins the TCB for the first time, they can get a certain value of energy coin for free through their IDs. Each ID can be collected only once

4. System Implementation

To implement the secure transaction model of the MG proposed here, the most important factors to consider are the reputation evaluation of users, secure collection of quotations required for MG auction, and generation of new blocks on the blockchain. The section is thus composed of three parts, namely, a reputation evaluation algorithm based on user behavior, an identity-based proxy signcryption algorithm, and a data block generation algorithm. These three parts correspond to the implementation of the above functions. The variables involved in system implementation are shown in Table 1.

4.1. Reputation Evaluation Algorithm Based on User Behavior.
In the traditional centralized power supply mode, users' power consumption behavior changes with time and leads to the emergence of peak and valley periods of power consumption. The peak period refers to the duration when the power consumption is concentrated and the power supply is limited. The valley period is the opposite duration, when there is less activity and supply is plentiful. The detailed peak–valley time is divided according to the local season and when the peak–

TABLE 1: Symbol definitions.

Symbol	Definition
$\eta(T)$	Time adaptive weight parameter
P	The peak period of power consumption
A	The average period of power consumption
V	The valley period of power consumption
$R_{i/j}^{\mathrm{con}}$	Buyer/seller's contract reputation
$R_{i/j}^{C}$	Buyer/seller's consensus reputation
R_{i}^{DR}	Buyer's demand response reputation
R_{j}^{F}	Seller's feedback reputation
$R_{i/j}$	Buyer/seller's reputation
R_{MG}	MG's reputation
GP	System global parameters
MSK	System master key
$\mathrm{PK_{ori}, SK_{ori}}$	Original signcrypter's public and private keys
$\mathrm{PK_{proxy}, SK_{proxy}}$	Proxy signcrypter's public and private keys
$\mathrm{PK_{MG}, SK_{MG}}$	MG's public and private keys
$\mathrm{PSK_{op}}$	Proxy key
ω	Warrant
M	Plaintext information
σ	Proxy signcryption information

valley load appears. The peak–valley time varies slightly in different regions; therefore, only the peak P, average A, and the valley V are defined in this paper, and no detailed time division is given. To better control the user's trading behavior, we define a time-adaptive weight parameter $\eta(T)$ according to the peak–valley interval.

$$\eta(T) = \begin{cases} 1.5, & T \in P, \\ 1.25, & T \in V, \\ 1, & T \in A. \end{cases} \tag{1}$$

In the formula, regardless of whether the trading time T is in the peak or the valley period, its weight is greater than the weight of the average period. This is done to enhance users' attention to the two periods and reduce the probability of poor trading behavior.

4.1.1. Algorithm to Assess Reputation of Buyer Based on Behavior. Buyers in the MG community are mainly composed of consumers who need to buy energy. When prosumers have a seller reputation score of zero or the DER power supply cannot meet their needs, they are also buyers. In the algorithm to assess the reputation of the buyer, three reputation events affect the buyer's reputation: default events, demand response events, and block generation events.

When the buyer and the seller reach the intention to engage in a transaction through an auction and sign a contract, the seller begins to transmit electricity to the buyer.

When the electricity consumption of the buyer in the agreed time slots exceeds the agreed transaction capacity, the seller's interests are undermined because the number of coins stipulated in the contract is fixed.

Therefore, this paper uses the default contract as an evaluation index for the buyer's reputation. The buyer's contract reputation R_i^{con} is assessed as shown in Formula (2):

$$R_{i,t-1}^{con} = \begin{cases} -\eta(T_i), & Q_{ij}^i > Q_{ij}^{con}, \\ 0, & Q_{ij}^i \leq Q_{ij}^{con}, \end{cases} \tag{2}$$

where $R_{i,t-1}^{con}$ is the contract reputation of buyer i at the end of round $t - 1$, T_i is the transaction time of buyer i, Q_{ij}^i is the electricity consumption of i in the trading time slot, and Q_{ij}^{con} is the amount of electricity agreed in the contract. When the buyer violates the contract, they must compensate the seller according to the price of electricity of the power trading center.

A demand response event is used to balance the energy demand of the power supply system when power is in short supply. Each MG issues the demand of reducing energy consumption to its community and announces the total value by which the energy consumption needs to decrease. Buyers who are qualified to respond reduce their electricity consumption according to the agreed response amount. The buyer's demand response reputation R_i^{DR} is shown in Formula (3).

$$R_{i,t-1}^{DR} = \begin{cases} \dfrac{T_{i,d.5\%} + 0.7T_{i,d.5\%\sim25\%} + 0.5T_{i,d.25\%\sim50\%}}{T_{i,t-1}^{DR}}, & r = 1, \\ -0.5, & r = 0, \end{cases} \tag{3}$$

where $R_{i,t-1}^{DR}$ is the demand response reputation of buyer i at the end of round $t - 1$, $T_{i,d.5\%}$ indicates the time when the deviation between the response capacity and the agreed capacity of buyer i is less than 5%, $T_{i,d.5\%\sim25\%}$ indicates the time when the deviation is in the range 5% ~25%, $T_{i,d.25\%\sim50\%}$ indicates the time when the deviation is in the range 25% ~50%, and $T_{i,t-1}^{DR}$ indicates the total time buyer i needs to respond in round $t - 1$ of the transaction. If r is 1, i responds as required, and if r is 0, i does not respond.

The generation of blocks is inseparable from the consensus algorithm. To encourage nodes on the network to participate in the consensus process, the system assigns a certain reputation to nodes that participate in the consensus process. The consensus reputation R_i^C is calculated as follows:

$$R_{i,t-1}^C = k_i \alpha, \tag{4}$$

where $R_{i,t-1}^C$ is the consensus reputation of buyer i at the end of round $t - 1$, and k_i represents the number of times that buyer i participates in the consensus process in round $t - 1$. α is always greater than zero, and its value depends on the rate of block generation. The two are inversely proportional.

To sum up, before the buyer prepares to conduct a transaction in round t, we calculate their reputation according to Formula (5):

$$R_{i,t} = \begin{cases} 50, & t = \text{init}, \\ R_{i,t-1} + R_{i,t-1}^{con} + RR_{i,t-1}^{DR} + R_{i,t-1}^C, & t \neq \text{init}, \end{cases} \tag{5}$$

where $R_{i,t}$ is the reputation of buyer i at the end of round $t - 1$, $R_{i,t-1}$ is the reputation of i in round $t - 1$, and R is a judgment function. When i is qualified to respond, R is 1 and is otherwise 0. When buyer i is participating in the transaction for the first time, their initial reputation is 50.

4.1.2. Algorithm to Assess Reputation of Seller Based on Behavior. In the MG community, only prosumers can be sellers. The algorithm to assess their behavior as seller uses three reputation events as indicators: a default event, a feedback event, and a block generation event.

When the supply of electricity provided by the seller fails to reach the trading capacity agreed with the buyer in the contract, the buyer needs to purchase electricity from the power trading center to meet their electricity demand. The price of electricity offered by the power trading center is often higher than the transaction price in the MG community and causes losses for the buyer.

Therefore, whether a contract is default can be used as an evaluation index to assess the seller's reputation, the buyer's contract reputation R_j^{con} is as shown in Formula (6).

$$R_{j,t-1}^{con} = \begin{cases} -\eta(T_j), & Q_{ij}^j < Q_{ij}^{con}, \\ 0, & Q_{ij}^j \geq Q_{ij}^{con}, \end{cases} \tag{6}$$

where $R_{j,t-1}^{con}$ is the contract reputation of seller j at the end of round $t - 1$, T_j is the transaction time of j, and Q_{ij}^j is their electricity consumption in the trading time slot. When seller j defaults, they compensate buyer i for the extra cost of purchasing electricity from the power trading center.

A feedback event is one where after a transaction between buyer i and seller j, and i needs to provide feedback to j on the quality of the power supply service, which is expressed by F_{ij}. The range of F_{ij} is [-0.5,0.5], and the seller's feedback reputation R_j^F is shown in Formula (7):

$$R_{j,t-1}^F = \frac{\sum_{i=1}^L (F_{ij}E_i)}{\sum_{i=1}^L E_i}. \tag{7}$$

In the formula, $R_{j,t-1}^F$ is the feedback reputation of seller j at the end of round $t - 1$, L is the number of buyers who give feedback to seller j, and E_i is the feedback equilibrium function used to balance the feedback value.

The value of E_i can be only 0 or 1. To use E_i, the average value F_a and standard deviation F_{sd} of the buyer's feedback

need to be calculated first. They are calculated, respectively, by Formulae (8) and (9):

$$F_a = \frac{\sum_{i=1}^{L} F_{ij}}{L}, \tag{8}$$

$$F_{sd} = \sqrt{\frac{\sum_{i=1}^{L} \left(F_{ij} - F_a\right)^2}{L}}. \tag{9}$$

Then, the equilibrium parameter ε is obtained from F_a and F_{sd} as shown in Formula (10):

$$\varepsilon = F_a - F_{sd}. \tag{10}$$

Finally, E_i is assigned a value by Formula (11).

$$E_i = \begin{cases} 1, & F_{ij} \geq \varepsilon, \\ 0, & F_{ij} < \varepsilon. \end{cases} \tag{11}$$

The block generation events of the seller and the buyer are identical, and the seller's consensus reputation R_j^C is calculated as follows:

$$R_{j,t-1}^C = k_j \alpha, \tag{12}$$

where $R_{j,t-1}^C$ is the consensus reputation of seller j at the end of round $t-1$, and k_j represents the number of times that seller j participated in the consensus process in round $t-1$.

We can then obtain the seller's reputation in round t:

$$R_{j,t} = \begin{cases} 50, & t = \text{init}, \\ R_{j,t-1} + R_{j,t-1}^{\text{con}} + R_{j,t-1}^F + R_{j,t-1}^C, & t \neq \text{init}, \end{cases} \tag{13}$$

where $R_{j,t}$ is the reputation of seller j at the end of round $t-1$, and $R_{j,t-1}$ is their reputation in round $t-1$. When j participates for the first time in the transaction, their initial reputation is 50.

For both buyers and sellers, the range of values of the reputation score is [0, 100]. Therefore, after calculating the reputation of the user for participating in a given round using the above formulae, their reputation score should be recalculated through Formula (14) to restrict the range of reputation scores:

$$R_{i/j,t} = \begin{cases} 100, & R_{i/j,t} > 100, \\ R_{i/j,t}, & 0 \leq R_{i/j,t} \leq 100, \\ 0, & R_{i/j,t} < 100, \end{cases} \tag{14}$$

where $R_{i/j,t}$ is the reputation of the buyer/seller in round t. When the seller's reputation is zero, they can participate only in the transaction as a buyer. When the buyer's reputation is zero, they are blacklisted and forbidden from participating in round t.

4.2. Identity-Based Proxy Signcryption Scheme.

The identity-based proxy signcryption (IDPSC) algorithm [32] is an improvement over the proxy signcryption algorithm. The core idea is the same; that is, the right to signcryption of the data can be entrusted to the proxy signcrypter, who can then signcrypt the data instead of the original signcrypter. After receiving the proxy signcryption data, the receiver can regard the proxy signcrypter as the original signcrypter, who becomes responsible for the data. The original proxy signcryption algorithm uses traditional public key facilities and encounters the problem of authenticating the user's identity when using the key generated by this facility to communicate. It thus needs a trusted third party to issue the user's identity certificate. When there are too many users, the problem of certificate management becomes significant. The IDPSC does not encounter this problem and thus is more efficient.

In the MG community, the auctioning algorithm is often used for transaction matching. It involves a large number of communication processes. To ensure the confidentiality and nonrepudiation of the communicated information, the information needs to be signcrypted. In the proposed model, the participants are DER installed by prosumers and smart home manager installed by consumers, but these are unsuitable for data signcryption because of their hardware. To solve this problem, we propose a secure IDPSC algorithm for MG energy trading consisting of five subalgorithms: those for system setup, key generation, proxy key generation, proxy signcryption, and unsigncryption. The program description is as follows:

4.2.1. System Setup.

The primary function of the phase is that the government initializes the system in order to provide basic conditions for the operation of subsequent phases. The core algorithm of this phase is $\text{Setup}(1^\ell) \longrightarrow (\text{GP}, \text{MSK})$, which is controlled by the authority. The authority first enters the security parameter ℓ into the system, which generates two additive cyclic groups G and G_T of order p according to the parameter ℓ, and defines four hash functions $H_1 : \{0,1\}^* \longrightarrow G$, $\quad H_2 : \{0,1\}^* \longrightarrow Z_p^*$, $\quad H_3 : G_T \longrightarrow \{0,1\}^n$, and $H_4 : \{0,1\}^n \times G_T \longrightarrow Z_p^*$, where n is the byte length of message M. For the cyclic group G, g is the generator; G and G_T satisfy the bilinear mapping relation $e : G \times G \longrightarrow G_T$. The authority then randomly selects element $\lambda \in Z_p^*$ as the system's master key and computes the system's public key $\text{PK}_{\text{pub}} = \lambda g$. Finally, the master key MSK and the system global parameter GP are output:

$$\begin{cases} \text{MSK} = \lambda, \\ \text{GP} = \left\{n, e, g, \text{PK}_{\text{pub}}, H_1, H_2, H_3, H_4\right\}. \end{cases} \tag{15}$$

MSK is kept secret by the authority while GP is published to the entire network. All nodes can access this information.

4.2.2. Key Generation.

The primary function of the phase is to generate public and private key pairs for users, and the core algorithm is $\text{KGen}(\text{GP}, \text{MSK}, \text{ID}) \longrightarrow (\text{PK}_{\text{ID}}, \text{SK}_{\text{ID}})$. The

algorithm needs input a user ID; the authority audits it. If ID fails to pass the audit, the authority refuses to generate the key for the user. If the audit is passed, the authority outputs the public–private key pair (PK_{ID}, SK_{ID}) corresponding to the ID:

$$\begin{cases} PK_{ID} = H_1(ID), \\ SK_{ID} = \lambda PK_{ID.} \end{cases} \tag{16}$$

During the key generation process, each user can obtain multiple public–private key pairs. The energy managers and smart home managers deployed in the consumers' houses obtain exclusive public–private key pairs. In addition to obtaining exclusive public–private key pairs for the energy managers and smart home managers, prosumers also obtain public–private key pair for the DER deployed by them. To better explain our scheme, we define the public–private key pair of the smart home manager and DER as (PK_{ori}, SK_{ori}), which is the key of the original signcrypter. The key of the energy manager is (PK_{proxy}, SK_{proxy}), which is also the key of the proxy signcrypter. The recipient's key is represented by the public–private key pair (PK_{MG}, SK_{MG}) of the MG.

4.2.3. Proxy Key Generation. This phase is the core of IDPSC algorithm, and its main function is to entrust the data signcryption right to the proxy signcrypter, which is represented by $PKGen(GP, \omega, SK_{ori}, SK_{proxy}) \longrightarrow PSK_{op}$. The process of proxy key generation is performed by the user and can be divided into three steps. The first step is performed by the original signcrypter, such as smart home manager and DER, who uploads a warrant $\omega_{ori} \in (0, 1)^*$ (which records the proxy expiration date, proxy content permissions, and the identities of the original signcrypter and the proxy signcrypter) to the consortium blockchain. At this time, the MG needs to update the value of ω_{ori} of each node in the storage list in real time to verify the validity of the information during unsigncryption.

The second step is to generate the delegation D_ω. Both the private key SK_{ori} of the original signcrypter and the private key SK_{proxy} of the proxy signcrypter belong to the same user. In this step, the user can choose for the delegation to be generated by either the original signcrypter or the proxy signcrypter (energy manager). When the user has too many original signcrypters, it is recommended that the original signcrypter generate the delegation D_ω. The process is as follows:

$$\begin{aligned} s &= H_2(\omega_{ori}), \\ D_\omega &= sSK_{ori}. \end{aligned} \tag{17}$$

In the third step, the proxy signer generates the proxy key.

$$PSK_{op} = SK_{proxy} + D_\omega. \tag{18}$$

The informational interaction between the proxy signcrypter and the original signcrypter takes place on the home

LAN, and so there is no need to verify the information with the original signer.

4.2.4. Proxy Signcryption. In this phase, signcryption is performed on plaintext, the result of which is the ciphertext after signature and encryption, and the core algorithm is $PSigc(GP, \omega_{proxy}, M, PK_{MG}, PSK_{op}, SK_{proxy}) \longrightarrow \sigma$. When a user wants to participate in an auction, their proxy signcrypter needs to collect the quotation information $M \in (0, 1)^n$ from the original signcrypter and organize it. The proxy signer randomly selects an element $x \in Z_p^*$ and computes the symmetric encryption key K:

$$K = H_3(e(PK_{pub}, PK_{MG})^x). \tag{19}$$

Then, the proxy signer calculates the symmetric encrypted ciphertext C.

$$C = K \oplus M. \tag{20}$$

The proxy key PSK_{op} is then used to perform the proxy signcryption.

$$\begin{cases} V = e(g, PK_{pub})^x, \\ \mu = H_4(C, V), \\ S = xPK_{pub} - (\mu SK_{proxy} + PSK_{op}). \end{cases} \tag{21}$$

Finally, the proxy signcrypter outputs the proxy signcryption information $\sigma = (\omega_{proxy}, \mu, C, S)$ and sends it to the MG. ω_{proxy} represents the original signcrypter's warrant forwarded by the proxy signcrypter. Placing ω_{proxy} in the proxy signcryption information σ helps the MG quickly find the original signcrypter corresponding to the proxy signcryption information.

4.2.5. Unsigncryption. The primary function of this phase is to help the information receiver recover the real and effective plaintext information from the ciphertext. The algorithm is expressed as $UnSigc(GP, \sigma, SK_{MG}, PK_{MG}, PK_{ori}, PK_{proxy})$. When the MG conducts an energy auction, it needs the quotation information M from the original signcrypter, which exists in the received proxy signcryption information σ. The MG thus needs to perform the unsigncryption process. If the warrant ω_{proxy} in the proxy signcryption information σ is inconsistent with that sent by the original signcrypter, an error symbol $\perp$ is returned. Otherwise, the MG performs the following tasks to check the validity of the ciphertext C in the proxy signcryption information σ:

$$\begin{cases} s = H_2(\omega_{ori}), \\ V' = e(g, S)e(PK_{pub}, PK_{proxy})^{\mu+1}e(PK_{pub}, PK_{ori})^s. \end{cases} \tag{22}$$

Only when $\mu = H_4(C, V')$ is true does the MG receive the ciphertext C; it then calculates the symmetric encryption key:

$$K' = H_3\left(e(S, \text{PK}_{\text{MG}})e(\text{PK}_{\text{proxy}}, \text{SK}_{\text{MG}})^{\mu+1}e(\text{PK}_{\text{ori}}, \text{SK}_{\text{MG}})^s\right).$$

$$(23)$$

Finally, the MG obtains the original signcrypter's quotation information $M = K' \oplus C$. If the content of M does not fall within the scope specified in the warrant ω_{ori}, the algorithm outputs an error symbol $\perp$.

4.2.6. Proof of Correctness. We prove the correctness of this scheme by proving the correctness of V' and K'.

First, we prove the correctness of V'.

$$
\begin{aligned}
V' &= e(g, S)e(PK_{pub}, PK_{proxy})^{\mu+1}e(PK_{pub}, PK_{ori})^s \\
&= e(g, S)e(\lambda g, PK_{proxy})^{\mu+1}e(\lambda g, sPK_{ori}) \\
&= e(g, S)e(g, \lambda PK_{proxy})^{\mu}e(g, \lambda PK_{proxy})e(g, s\lambda PK_{ori}) \\
&= e(g, S)e(g, \mu SK_{proxy})e(g, SK_{proxy})e(g, sSK_{ori}) \\
&= e(g, S)e(g, \mu SK_{proxy} + SK_{proxy} + D_\omega) \\
&= e(g, xPK_{pub} - (\mu SK_{proxy} + PSK_{op}))e(g, \mu SK_{proxy} + PSK_{op}) \\
&= e(g, xPK_{pub}) = e(g, PK_{pub})^x = V.
\end{aligned}
$$

$$(24)$$

Then, we can prove the correctness of K'.

$$
\begin{aligned}
K' &= H_3\left(e(S, \text{PK}_{\text{MG}})e(\text{PK}_{\text{proxy}}, \text{SK}_{\text{MG}})^{\mu+1}e(\text{PK}_{\text{ori}}, \text{SK}_{\text{MG}})^s\right) \\
&= H_3\left(\frac{e(S, \text{PK}_{\text{MG}})e(\text{PK}_{\text{proxy}}, \lambda\text{PK}_{\text{MG}})^{\mu}}{e(\text{PK}_{\text{proxy}}, \lambda\text{PK}_{\text{MG}})e(s\text{PK}_{\text{ori}}, \lambda\text{PK}_{\text{MG}})}\right) \\
&= H_3\left(e(S, \text{PK}_{\text{MG}})e(\lambda\text{PK}_{\text{proxy}}, \text{PK}_{\text{MG}})^{\mu}e(\lambda\text{PK}_{\text{proxy}}, \text{PK}_{\text{MG}})e(s\lambda\text{PK}_{\text{ori}}, \text{PK}_{\text{MG}})\right) \\
&= H_3\left(e(S, \text{PK}_{\text{MG}})e(\mu\text{SK}_{\text{proxy}}, \text{PK}_{\text{MG}})e(\text{SK}_{\text{proxy}}, \text{PK}_{\text{MG}})e(s\text{SK}_{\text{ori}}, \text{PK}_{\text{MG}})\right) \\
&= H_3\left(e(S, \text{PK}_{\text{MG}})e(\mu\text{SK}_{\text{proxy}} + \text{SK}_{\text{proxy}} + D_\omega, \text{PK}_{\text{MG}})\right) \\
&= H_3\left(e(x\text{PK}_{\text{pub}} - (\mu\text{SK}_{\text{proxy}} + \text{PSK}_{\text{op}}), \text{PK}_{\text{MG}})e(\mu\text{SK}_{\text{proxy}} + \text{PSK}_{\text{op}}, \text{PK}_{\text{MG}})\right) \\
&= H_3\left(e(x\text{PK}_{\text{pub}}, \text{PK}_{\text{MG}})\right) \\
&= H_3\left(e(\text{PK}_{\text{pub}}, \text{PK}_{\text{MG}})^x\right) = K
\end{aligned}
$$

$$(25)$$

The $e(.,.)$ in all of the above formulas refers to the bilinear algorithm in Section 2.2.

4.3. Generation of Data Blocks. In the blockchain network, the generation of blocks is closely related to the consensus algorithm used on the blockchain, where consensus algorithms used by different types of blockchains are different. The system used in this paper is a $1 + N$ hybrid blockchain system, which is essentially a transaction consortium blockchain and N identity private blockchains. Because there are two different types of blockchains, this system uses two consensus algorithms to generate blocks at the same time, namely, the Raft consensus algorithm [33] and the PBFT consensus algorithm [34].

4.3.1. Consensus Algorithms of IPB. The private blockchain built by the MG itself is highly centralized and has central nodes, which coincides with the strong leadership of the Raft consensus algorithm. We thus use the Raft consensus algorithm on identity private blockchain. The strong leadership of the Raft consensus algorithm is mainly manifested in the fact that all log entries flow only from the leader server to the backup server, which simplifies the management of replicated logs [33]. There are three identities of the leader, candidate, and follower in the Raft consensus mechanism. In the identity private blockchain constructed by the MG, the leader is a node chosen by the MG from among its manager nodes, other unselected MG manager nodes and authority nodes are candidates, and followers are community user nodes in the identity private blockchain. A candidate is a candidate for leader. When the leader fails to operate normally, a new leader is selected from among the candidates. Only when all MG manager nodes fail does the authority node temporarily act as leader. The process of generating blocks using the Raft consensus algorithm can be simplified into three steps. In the first step, the leader node reviews the reputation data of the community users. Once the review is passed, it is sent to the energy manager nodes of each community user for reinspection. In the second step, the energy manager node reinspects the data sent by the leader node and returns the result to the leader node. In the third step, the leader node packages the data passed by both validation and revalidation into blocks and uploads them to the local private blockchain. To prevent the MG manager node on the private blockchain from tampering with the data, the hash value of the blocks on the private blockchain is uploaded to the trading consortium blockchain. In addition, the Raft consensus algorithm does not affect the user's reputation.

4.3.2. Consensus Algorithms of TCB. The PBFT consensus algorithm is used in the transaction consortium blockchain. A modified form of the original is used in this paper by changing the mechanism of establishing the consensus committee. In the original algorithm, the consensus committee is composed of preselected consensus nodes, the number of nodes in the consensus committee is fixed, and the consensus nodes do not change. In our modified PBFT algorithm, the consensus nodes that form the consensus committee are constantly changing. At the beginning of each round of transaction, the system selects the consensus nodes to form a new consensus committee, and the working time of each consensus committee is one trading cycle as planned by the system. To encourage MG trading platforms to maintain good internal trading behavior, the system constructs a consensus committee according to the principle of "reputation first, quantity first." The essence is to select the highest ranked MG community according to the selection principle. Reputation first in the selection rules refers to the selection of the MG community with the highest reputation score. Reputation depends on its internal users and is calculated as follows:

$$R_{\text{MG}} = \frac{\sum_{i=1}^{l} R_i + \sum_{j=1}^{m} R_j}{l + m}, \quad (l + m) > H, \qquad (26)$$

where l represents the number of buyers in the MG community at the time, m represents the number of sellers, and H is the fixed number of nodes of the consensus committee. The community with more users is preferred if two MG communities have the same reputation score. The same MG community cannot be selected consecutively; when this happens, the second-ranked MG community is responsible for block production.

Once the consensus committee has been built, members of the committee follow the same consensus process as in the original PBFT consensus algorithm to generate the blocks, as shown in Figure 6.

The distributed consensus process is divided into five stages: request, preprepare, prepare, commit, and reply. At the beginning of the request phase, the authority sends the client's request to the MG manager node. After receiving the request, the algorithm enters the preprepare phase; this stage requires the MG manager node to broadcast the sequence of execution of the transaction to the user node inside the consensus committee. In the prepare stage, the user node has two behaviors for the received information: one is to receive and forward the received information to the nodes, and the other is to do nothing. Nodes in the second state are called Byzantine nodes, such as producer 2 in Figure 6. The trigger for the commit phase is to receive $(H - f)$ identical requests for information. If this condition is satisfied, the commitment information is broadcast to the entire network. f is the number of Byzantine nodes in the consensus committee and needs to meet the condition $f \leq (H - 1)/3$. In the reply phase, the consensus nodes also collect $(H - f)$ identical commitment information items before feeding it back to the authority. The verified information is packaged into blocks and uploaded to the trading consortium blockchain. Except for the MG manager node, all consensus nodes can receive reputation rewards after the block has been generated.

5. Analysis and Evaluation

5.1. Analysis of System Performance. To verify the security of the MG energy trading mechanism based on the hybrid blockchain, we theoretically analyzed the implementation of various security features of this scheme. We also conducted a feature comparison with some proposals in the literature, and the results are shown in Table 2. It is clear that the proposed scheme outperformed the other schemes.

(1) Privacy and confidentiality: In this paper, two types of blockchains, the transaction consortium blockchain and the identity private blockchain, were used to store information. The user's identity-related information is stored in the identity private blockchain. When the user transacts in the trading consortium blockchain, the attacker cannot learn their identity from the transaction information. To ensure the confidentiality of the transaction information, we use an identity-based proxy signcryption algorithm in which the quotation submitted by users is encrypted with a symmetric encryption key. Nodes
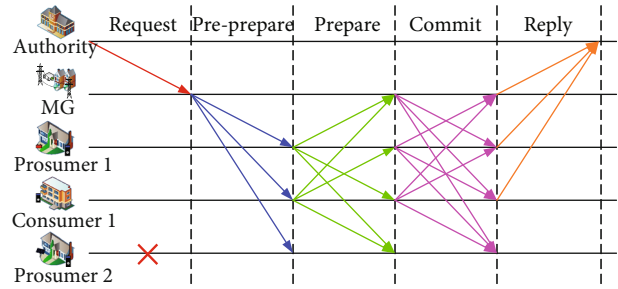

FIGURE 6: Consensus flowchart of the simplified PBFT algorithm.

TABLE 2: Comparison of security features.

Security features	Ref. [35]	Ref. [36]	Ref. [37]	Our scheme
Privacy	√	×	×	√
Confidentiality	×	×	×	√
Transparency	√	√	√	√
Traceability	√	√	√	√
Imtamperability	√	√	√	√
Nonrepudiation	√	×	×	√

in the consortium blockchain cannot understand the quotation information contained in the ciphertext without obtaining the symmetric encryption key

(2) Transparency and traceability: As blockchain is a shared ledger, all nodes share the same data, and the transaction records generated by each node are public. The generation of blocks in the blockchain also follows transparent consensus rules. The consensus node processes the transaction information and generates new blocks according to the specific consensus to render the data transparent. Blocks on the blockchain are generated in chronological order, and all transactions are open due to the transparency of the blockchain. When there is doubt about a transaction, the information on it can be traced according to the above conditions

(3) Imtamperability: The difficulty of data tampering on the blockchain is related to the consensus algorithm used. In this paper, the Raft consensus algorithm was used on the private blockchain and the PBFT consensus algorithm on the consortium blockchain. The central node of the private chain is powerful. To prevent the central node from tampering with the data, we store the block digest of the private blockchain in the consortium blockchain. Therefore, the scheme's resistance to being tampered with is implemented by the PBFT consensus algorithm on the consortium blockchain. The PBFT algorithm is the most commonly used consensus algorithm on the consortium blockchain because of its high scalability and low power consumption. When it is used, the system can still work normally even if 33% of the nodes in the system are Byzantine nodes [38]. In

addition, the consensus nodes in this paper were selected from among nodes with good reputation, which yield more benefits. It is unrealistic for more than a third of nodes with good reputation to go against their own interests to jeopardize the stability of the system

(4) Nonrepudiation: The nonrepudiation of information is realized by a digital signature that is broadcast and verified between nodes before being stored in the blockchain. When a trade dispute arises, the nonrepudiation of the given trade can be realized by tracing the trade signature in the blockchain. The identity-based proxy signcryption scheme used in this paper can not only encrypt the information but can also sign it. Because the transaction entity does not have enough computing power to carry out frequent signcryption, we entrust the right of signcryption of the data to a proxy signcrypter with strong computing power. The signcryption information of the proxy signcrypter is identical to that of the original signcrypter

5.2. Analyzing Validity of Algorithm to Assess Reputation Based on User Behavior. Algorithms to assess reputation based on user behavior can be divided into those based on buyer behavior and seller behavior. The validity of the algorithm considered here thus needs to be analyzed from the perspectives of both the buyer and the seller. Because the contract reputation scores of the buyers and sellers are closely related to the peak and trough periods of trading time, this paper used the peak–valley period division table as shown in Table 3 for a more concise analysis of the validity of the algorithm. Each trading cycle is 30 minutes long.

5.2.1. Analyzing the Validity of Algorithm to Assess Buyer Behavior. To verify the validity of the algorithm to assess buyer behavior, we considered a scenario in which three buyers performed different behaviors over 24 hours. As Figure 7 shows, when the buyer did not trigger a reputation event, their score remained the same. From 3:00 to 3:30, buyer A triggered a default event; as this was a valley period, buyer A's score dropped by 1.25. From 6:00 to 6:30, buyer C triggered a default event. At this was the average period, buyer C's score dropped by 1. From 8:00 to 11:00, the user entered the peak period of electricity consumption. When power consumption was in short supply during the peak period, the MG issued a demand response event. Both buyers A and C activated the demand response event in this period. From 8:00 to 10:00, buyer A continuously activated four demand response events and maintained a capacity deviation of less than 5%. Buyer A thus added 4 scores in total in this period. From 8:30 to 10:30, buyer C also activated four demand response events in a row, but the capacity deviation the first two times was in the range of 5%–25% and was less than 5% for the last two instances. Buyer C thus added 3.4 scores in total. From 10:00 to 10:30, the MG community was selected to lead the block generation process, and buyer C successfully triggered the block generation event and earned a score of 0.45. From 14:30 to 15:00, buyer C triggered the default event again, in the average period, and lost

TABLE 3: Peak and valley time division.

Period (hour)	Peak period	Average period	Valley period
(0-6]	0	0	1
(6-8]	0	1	0
(8-11]	1	0	0
(11-16]	0	1	0
(16-21]	1	0	0
(21-22]	0	1	0
(22-24]	0	0	1

1: current time belongs to this period; 0: current time does not belong to this period.
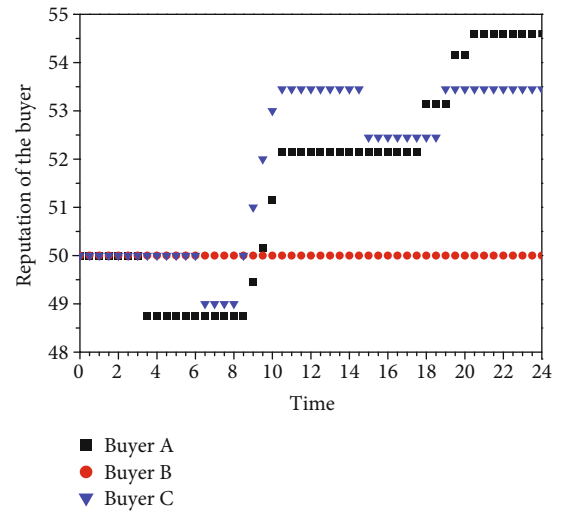

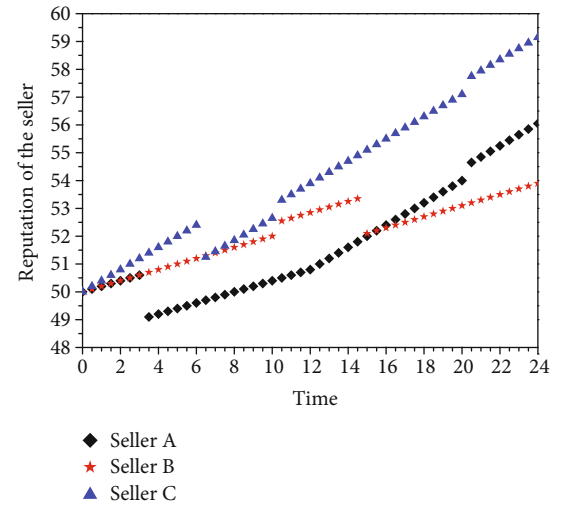FIGURE 7: Reputation scores of three buyers.

FIGURE 8: Reputation scores of three sellers.

1 point again. From 16:00 to 21:00, during the peak period of electricity consumption, buyer A activated demand response events two times, from 17:30 to 18:00 and from 19:00 to 19:30 and obtained 2 scores; buyer C activated only once, from 18:30 to 19:00, and earned 1 reputation score.

TABLE 4: Comparison of computational overheads.

Scheme	PKGen	PSigc	UnSigc	Total
Ref. [39]	$2T_M$	$5T_M + T_E$	$T_M + T_E + 5T_P$	$8T_M + 2T_E + 5T_P$
Ref. [40]	$2T_M + T_E + T_P$	$3T_M + T_E$	$T_M + T_E + 4T_P$	$6T_M + 3T_E + 5T_P$
Ref. [41]	$2T_E + 2T_P$	$T_M + 2T_E$	$4T_P$	$T_M + 4T_E + 6T_P$
Our scheme	T_M	$2T_M + 2T_E$	$4T_E + 2T_P$	$3T_M + 6T_E + 2T_P$

T_M: time needed for scalar multiplication operation on G; T_E: time needed for exponential operation on G_T; T_P: time needed for bilinear pairing operation.

From 20:00 to 20:30, the MG community was again eligible for block generation, and buyer A triggered this block generation event to obtain 0.45 scores. Buyer B neither initiatively triggered the default event nor the demand response event. When the MG community obtained the block generation qualification, buyer B's reputation was not among the top H-1; so, B could not trigger the block generation event, and its reputation score remained unchanged at the initial value of 50. The results show that all behaviors of buyers are objectively reflected in their reputation scores.

5.2.2. Analyzing the Validity of Algorithm to Assess Seller Behavior. To verify the validity of the algorithm to assess seller behavior, we considered another scenario in which three sellers performed different behaviors in 24 hours. As Figure 8 shows, assuming sellers A and B had the same quality of power supply service, and seller C supplied better service than them, and when the seller did not trigger a default event, the reputation scores of 0.1 were obtained for A and B by the feedback event at the end of each round. C obtained 0.2 reputation points through the feedback event. When the seller triggered the default event, both A and B lost 0.25 reputation points for the feedback event and C lost 0.15 reputation points. The occurrence of consensus events did not affect the score of the feedback events. From 3:00 to 3:30, seller A triggered a default event; as this was a valley period, seller A's score dropped by 1.5. From 6:00 to 6:30, seller C triggered a default event in the average period, and his/her score dropped by 1.15. From 10:00 to 10:30, the MG community was selected to dominate the block generation process, sellers B and C triggered the block generation event together, and seller B scored 0.55 while seller C scored 0.65. At 12:00, and seller A updated the equipment. Following this, A and C had the same quality of service, and the reputation scores provided by the feedback event were identical to those for C. From 14:30 to 15:00, seller B triggered the default event in the average period, and his/her reputation score decreased by 1.25. From 20:00 to 20:30, the MG community was again eligible for block generation, and A and C triggered block generation events together; both received 0.65 points. The results show that all behaviors of sellers were objectively reflected in their reputation scores.

5.3. Assessing the IDPSC Algorithm. The computational cost of the proxy signcryption algorithm consists mainly of three operations: proxy key generation, proxy signcryption, and unsigncryption. In this section, our scheme is compared with those proposed in Refs. [39–41] from the perspective of com-
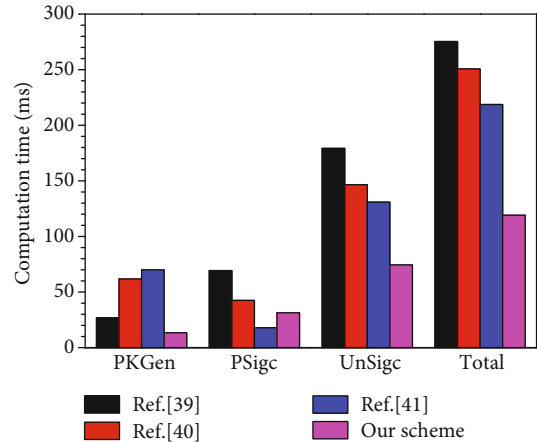

FIGURE 9: Comparison of computational overhead.

putational cost, and the results are shown in Table 4. Many key parameters in the IDPSC algorithm can be used all the time after one calculation. These parameters are calculated in advance by the system and thus are not included in the comparison of computational costs of the algorithms considered here. In Table 4, T_M represents the duration of operation of scalar multiplication on G, T_E represents that of the exponential operation on G_T, and T_P represents the time required for the bilinear pairing operation. To show the computational cost of each scheme more clearly, we refer to the operation time defined by He et al. [42] for calculation; that is, the time required for the scalar multiplication operation on G was 13.405 ms, that for exponential operation on G_T was 2.249 ms, and the time required for the bilinear pairing operation was 32.713 ms.

Figure 9 shows a comparison of the calculation costs of the proposed scheme with certain other schemes. It is clear that our scheme delivered the best performance on the proxy key generation algorithm and the unsigncryption algorithm. In the proxy signcryption algorithm, although our scheme was not the best, only the one proposed by Yu et al. [41] was superior to it. In terms of overall overhead, our method was the best. Our overall overhead accounts for 43.27% of Ref. [39], 47.51% of Ref. [40], and 54.62% of Ref. [41]. In general, it was more useful in practical application scenarios.

6. Conclusion

To address the problems of data storage and identity management and transaction in the microgrid, this paper proposed a secure transaction mechanism for it based on a

hybrid blockchain. A combination of the identity private blockchain and the transaction consortium blockchain is used to store users' identity information and trade information separately to guarantee user privacy. Blockchain-based features such as transparency and traceability provide a transparent and open energy trading platform for users of the MG community. In the process of energy transactions in the microgrid community, both parties to the transaction may have dishonest behaviors, and the occurrence of dishonest behaviors will result in property losses, which will reduce the participation of users. A reputation evaluation algorithm based on user behavior is used to constrain users' MG trading behavior on the identity private blockchain and is committed to creating a favorable atmosphere for the energy trading market. The smart home manager or DER, as a light node, cannot afford the computing power required for frequent signcryption. This paper proposes an identity-based proxy signcryption algorithm to guarantee the confidentiality of user quotations and the nonrepudiation of transactions. A system analysis showed that the reputation evaluation algorithm proposed here can objectively reflect all the behaviors of users, and the identity-based proxy signcryption scheme has advantages over competitors in data sharing.

Data Availability

The authors approve that data used to support the finding of this study are included in the article.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

Conceptualization was done by Z. S. and X. Z. Algorithm and simulation were done by Z. S. Writing-original draft and review were done by Z. S and X. Z. Security analysis was done by X. Z. Paper polish, editing, and revision were done by M. L. Visualization was done by Z. S. and X. Z.

Acknowledgments

References

[1] W. A. Hermann, "Quantifying global exergy resources," *Energy*, vol. 31, no. 12, pp. 1685–1702, 2006.

[2] K. N. Khaqqi, J. J. Sikorski, K. Hadinoto, and M. Kraft, "Incorporating seller/buyer reputation-based system in blockchain-enabled emission trading application," *Applied Energy*, vol. 209, pp. 8–19, 2018.

[3] D. Han and H. Xie, "China Renewable Energy Development Report 2018," *Water Power*, vol. 45, no. 8, p. 46, 2019.

[4] A. Y. Saber and G. K. Venayagamoorthy, "Plug-in vehicles and renewable energy sources for cost and emission reductions," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 4, pp. 1229–1238, 2011.

[5] A. Ahmad and J. Y. Khan, "Real-Time load scheduling and storage Management for Solar Powered Network Connected EVs," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1220–1235, 2020.

[6] W. Shao, W. Xu, Z. Xu, B. Liu, and H. Zou, "A grid connection mechanism of large-scale distributed energy resources based on blockchain," in *2019 Chinese Control Conference (CCC)*, pp. 7500–7505, Guangzhou, China, July 2019.

[7] A. Khanjanzadeh, S. Soleymani, and B. Mozafari, "A decentralized control strategy to bring back frequency and share reactive power in isolated microgrids with virtual power plant," *Bulletin of the Polish Academy of Sciences-Technical Sciences*, vol. 69, no. 1, article e136190, 2021.

[8] M. Royapoor, M. Pazhoohesh, P. J. Davison, C. Patsios, and S. Walker, "Building as a virtual power plant, magnitude and persistence of deferrable loads and human comfort implications," *Energy & Buildings*, vol. 213, article 109794, 2020.

[9] Y. Liu, Y. Li, H. B. Gooi et al., "Distributed robust energy management of a multimicrogrid system in the real-time energy market," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 396–406, 2019.

[10] X. Zhang, T. Yu, Z. Xu, and Z. Fan, "A cyber-physical-social system with parallel learning for distributed energy management of a microgrid," *Energy*, vol. 165, pp. 205–221, 2018.

[11] S. Ranjbar, A. R. Farsa, and S. Jamali, "Voltage-based protection of microgrids using decision tree algorithms," *International Transactions on Electrical Energy Systems*, vol. 30, no. 4, article 2274, 2020.

[12] J. J. Sikorski, J. Haughton, and M. Kraft, "Blockchain technology in the chemical industry: Machine-to-machine electricity market," *Applied Energy*, vol. 195, pp. 234–246, 2017.

[13] G. Kim, J. Park, and J. Ryou, "A study on utilization of blockchain for electricity trading in microgrid," in *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 743–746, Shanghai, China, January 2018.

[14] E. Mengelkamp, J. Gärttner, K. Rock, S. Kessler, L. Orsini, and C. Weinhardt, "Designing microgrid energy markets: a case study: the Brooklyn microgrid," *Applied Energy*, vol. 210, pp. 870–880, 2018.

[15] M. L. Di Silvestre, P. Gallo, M. G. Ippolito, E. R. Sanseverino, and G. Zizzo, "A technical approach to the energy blockchain in microgrids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4792–4803, 2018.

[16] M. L. Di Silvestre, P. Gallo, M. G. Ippolito et al., "Ancillary services in the energy blockchain for microgrids," *IEEE Transactions on Industry Applications*, vol. 55, no. 6, pp. 7310–7319, 2019.

[17] M. U. Hassan, M. H. Rehmani, and J. Chen, "DEAL: differentially private auction for blockchain-based microgrids energy trading," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 263–275, 2020.

[18] G. van Leeuwen, T. AlSkaif, M. Gibescu, and W. van Sark, "An integrated blockchain-based energy management platform with bilateral trading for microgrid communities," *Applied Energy*, vol. 263, article 114613, 2020.

[19] A. Meeuw, S. Schopfer, A. Woerner et al., "Implementing a blockchain-based local energy market: Insights on communication and scalability," *Computer Communications*, vol. 160, pp. 158–171, 2020.

[20] Y. C. Tsao and V. V. Thanh, "Toward blockchain-based renewable energy microgrid design considering default risk and demand uncertainty," *Renewable Energy*, vol. 163, pp. 870–881, 2021.

[21] R. Lasseter, A. Akhil, C. Marnay et al., *Integration of distributed energy resources. The CERTS Microgrid Concept*, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2002.

[22] D. Boneh, E. J. Goh, and K. Nissim, "Evaluating 2-DNF Formulas on Ciphertexts," in *Theory of Cryptography. TCC 2005. Lecture Notes in Computer Science, vol 3378*, J. Kilian, Ed., pp. 325–341, Springer, Berlin, Heidelberg, 2005.

[23] D. Freeman, S. Michael, and T. Edlyn, "A taxonomy of pairing-friendly elliptic curves," *Journal of Cryptology*, vol. 23, no. 2, pp. 224–280, 2010.

[24] C. Gamage, J. Leiwo, and Y. Zheng, "An efficient scheme for secure message transmission using proxy-signcryption," in *The Twenty Second Australasian Computer Science Conference*, pp. 18–21, Auckland, New Zealand, 1999.

[25] Y. Zheng, "Digital signcryption or how to achieve cost (signature & encryption) ≪ cost (signature) + cost (encryption)," in *Advances in Cryptology — CRYPTO '97. CRYPTO 1997. Lecture Notes in Computer Science, vol 1294*, B. S. Kaliski, Ed., pp. 165–179, Springer, Berlin, Heidelberg, 1997.

[26] M. Mambo, K. Usuda, and E. Okamoto, "Proxy signatures: delegation of the power to sign messages," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 79, no. 9, pp. 1338–1354, 1996.

[27] G. Sun, M. Dai, F. Zhang, H. Yu, X. Du, and M. Guizani, "Blockchain-enhanced high-confidence energy sharing in internet of electric vehicles," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 7868–7882, 2020.

[28] X. Zhang and D. Wang, "Adaptive traffic signal control mechanism for intelligent transportation based on a consortium blockchain," *IEEE Access*, vol. 7, pp. 97281–97295, 2019.

[29] J. Kang, R. Yu, X. Huang, S. Maharjan, Y. Zhang, and E. Hossain, "Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3154–3164, 2017.

[30] P. Wang, Y. Li, S. Zhao, H. Chen, Y. Jin, and Y. Ding, "Key technologies of distributed energy trading based on blockchain," *Automation of Electric Power Systems*, vol. 43, no. 14, pp. 53–64, 2019.

[31] X. Zhang and X. Chen, "Data security sharing and storage based on a consortium Blockchain in a vehicular ad-hoc network," *IEEE Access*, vol. 7, pp. 58241–58254, 2019.

[32] H. Zhu, Y. Wang, C. Wang, and X. Cheng, "An efficient identity-based proxy signcryption using lattice," *Future Generation Computer Systems*, vol. 117, pp. 321–327, 2021.

[33] D. Ongaro and J. Ousterhout, "In search of an understandable consensus algorithm," in *2014 {USENIX} Annual Technical Conference ({USENIX}{ATC} 14)*, pp. 305–319, Philadelphia, PA, USA, June 2014.

[34] M. Castro and B. Liskov, "Practical byzantine fault tolerance," in *the Third Symposium on Operating Systems Design and Implementation (OSDI'99)*, pp. 173–186, New Orleans, Louisiana, USA, 1999.

[35] N. Z. Aitzhan and D. Svetinovic, "Security and privacy in decentralized energy trading through multi-signatures, blockchain and anonymous messaging streams," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 5, pp. 840–852, 2018.

[36] W. Zhao, J. Lv, X. Yao et al., "Consortium blockchain-based microgrid market transaction research," *Energies*, vol. 12, no. 20, article 3812, 2019.

[37] Z. Xu, D. Yang, and W. Li, "Microgrid group trading model and solving algorithm based on bockchain," *Energies*, vol. 12, no. 7, article 1292, 2019.

[38] M. Castro and B. Liskov, "Practical byzantine fault tolerance and proactive recovery," *ACM Transactions on Computer Systems*, vol. 20, no. 4, pp. 398–461, 2002.

[39] C. Zhou, "Identity based generalized proxy signcryption in standard model," *Journal of Cryptologic Research*, vol. 3, no. 3, pp. 307–320, 2016.

[40] H. Yu, Z. Wang, J. Li, and X. Gao, "Identity-based proxy signcryption protocol with universal composability," *Security and Communication Networks*, vol. 2018, 11 pages, 2018.

[41] H. Yu and Z. Wang, "Construction of certificateless proxy signcryption scheme from CMGs," *IEEE Access*, vol. 7, pp. 141910–141919, 2019.

[42] D. He, H. Wang, L. Wang, J. Shen, and X. Yang, "Efficient certificateless anonymous multi-receiver encryption scheme for mobile devices," *Soft Computing*, vol. 21, no. 22, pp. 6801–6810, 2017.

WILEY | Hindawi

Research Article

Dynamic Transmission Rate Control for Multi-Interface IoT Devices: A Stochastic Optimization Framework

Yuming Zhang [iD],[1] **Bohao Feng** [iD],[1] **Aleteng Tian,**[1] **Chengxiao Yu,**[1] **Zhiruo Liu,**[2] **and Hongke Zhang** [iD][1]

[1]*School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*China Academy of Information and Communication Technology, China*

Correspondence should be addressed to Bohao Feng; bhfeng@bjtu.edu.cn

Recent advances in the Internet of Things (IoT) technologies have enabled ubiquitous smart devices to sense and process various kinds of data. However, these innovations also raise the concern of efficient data transmission. Tackling the above issue is nontrivial since the resource constraints and environmental randomness in IoT require a lightweight transmission scheme while guaranteeing system stability. In this paper, we formulate the transmission scheduling problem of multi-interface IoT devices as a concave optimization, aimed at accommodating the randomness of the IoT environment within the network capacity. By applying the Lyapunov optimization technique, we divide the stochastic problem into a series of low-complex subproblems, which can be individually solved per time slot, and develop a dynamical control algorithm that does not require a priori knowledge such as link states. Theoretical analysis shows that our algorithms nicely bound the average queue length and are asymptotically optimal. Finally, extensive simulation results verify the theoretical conclusions and validate the effectiveness of the proposed algorithm.

1. Introduction

With the ubiquitously deployed smart devices, the Internet of Things (IoT) technology has been facilitating the intelligence of residential daily activities by providing advanced services for transportation, agriculture, industrial manufacturing, etc. [1–3]. However, the continuous growth of IoT applications with the proliferation of various devices has resulted in an unprecedented explosion of network data traffic [4]. According to related report [5], the global IoT cellular traffic is expected to grow to 1.7 exabytes per month by 2022, a two-fold increase over 2020. This huge amount of data traffic poses a critical challenge to the current networks [6, 7], making it impractical to provide transmission guarantees.

Recent innovations of scalable communication systems can empower the IoT devices to connect to heterogeneous networks concurrently, enabling IoT devices to transmit data through different networks in parallel [8]. In particular, an IoT device equipped with multiple interfaces can use multi-

ple channels simultaneously in the physical link layer [9] and deliver packets through Multipath TCP (MPTCP) in the network layer [10]. As a result, the IoT devices with multiple data flows can be treated as a transmission scheduler with a many-to-many traffic pattern. Namely, an IoT device can adopt different kinds of data and category them into different types of flows, depending on priorities or requirements. Then, the transmission scheduler can choose to deliver these data flows through either one link or multiple links. In this regard, the IoT device performs like an input-queued switch for data transmission. Moreover, this design can also bring other advantages for IoT applications, such as performance improvement and scalability support [11].

The novel transmission paradigm introduced by [8] also raises the concerns of transmission scheduling for achieving the desired transmission rate. For example, how many packets of different flows should be transmitted through which link, and how to adjust transmission rates for each flow with system stability guaranteed. These concerns are

further complicated with IoT devices' mobility, as it brings dynamical communication environments and stochastic link conditions. Few efforts like [12] consider the stochastic multipath IoT scenarios, as it is not easy to determine the optimal scheduling schemes. The essential reason lies in two aspects. On the one hand, foresightedly optimize the scheduling problems requires the full knowledge of environmental information. However, because of the randomness of packet arrival rates and link states, it is almost impractical to apply this kind of solution to IoT scenarios. Although predicting the network state may be an alternative approach, it is inefficient in facing emergencies, i.e., traffic bursts, which would result in system instability. On the other hand, most IoT devices are constrained with limited computation and energy resources, incapable of operating complex algorithms. Besides, it is also unnecessary to spare a large number of resources for determining data transmission decisions. Thus, efficient online scheduling solutions with low computation complexity are more suitable for IoT devices.

This paper introduces a novel stochastic optimization framework for multi-interface IoT devices to simultaneously achieve optimal transmission scheduling. We first formulate the transmission scheduling problem as a stochastic optimization with the objective of maximizing the long-term transmission utility within the network capacity. Then, we leverage the Lyapunov optimization technique to develop a low-complexity scheduling algorithm. Our main contributions are summarized as follows.

(1) We characterize the multi-interface IoT devices with multiple traffics as a multiqueue model and decouple the scheduling process into two subproblems: admission control and output rate control. Then, we formulate the transmission scheduling problem as a stochastic concave optimization, which includes the randomness of network states and system stability constraint

(2) By leveraging the *drift-plus-penalty* framework, we divide the proposed stochastic problem into three deterministic subproblems, which can be further separated based on their linearly coupling characteristics. Then, we propose a low-complex transmission scheduling algorithm and prove that it can provide an upper bound of the queue length and achieve a $[O(V), O(1/V)]$ trade-off between the queue length and the transmission utility

(3) Extensive simulation results demonstrate the efficiency of our algorithms, which explicitly outperform the benchmarks in terms of system stability, average delay, and network utility

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 describes the considered model and formulates the scheduling problem. In Section 4, we develop the scheduling algorithm and present the theoretical analysis. Experimental results are shown in Section 5, and the conclusion is given at last in Section 6.

2. Related Works

As transmission rate control plays a crucial issue in the research literature of communication, there have been extensive efforts in developing customized schemes for diverse network paradigms in different aspects. For example, to deal with the variation traffic requirements, the authors of [13] consider the situation that the resources can be dynamically shifted between cells and develop a dynamic resource allocation protocol. For device to device communication systems, the authors of [14] try to maximize the weighted sum transmission rate and use a two-step approach to solve the nonconvex mixed-integer problem of resource allocation and subchannel assignment. In [15], the authors focus on the information-centric network and present a multipath-aware ICN rate-based congestion control algorithm to calculate per-link rates for the multipath scenario. The authors of [16] address the bandwidth sharing issues in a software-defined network and design a distributed resource allocation algorithm that can provide a trade-off between fairness and cost. For data center networks, the authors of [17] formulate the multiple rate control issue as a convex optimization problem and use their proposed transmission protocol to achieve efficient bandwidth allocation.

In addition, there are also many research works jointly considering the assignment of transmission and other kinds of resources. In [18], the authors study the trade-off between data rate performance and energy consumption in heterogeneous networks and introduce an energy-efficient scheduling scheme to improve system performance. The authors of [19] propose a low-complexity algorithm by difference-of-convex programming to simultaneously optimize service level selection and transmission resource allocation in mobile edge computing systems. By jointly considering task assignment, transmission, and computing resources allocation, the authors of [20] propose a multilayer data flow process system that can provide low latency services for real-time applications. The authors of [21] consider a three-node relay system and provide analytical solutions to the proposed optimization problem for power assignment and relay location. The authors of [22] formulate a delay-sensitive data offloading algorithm to optimize the computing and communication resources to minimize the execution delay and transmission delay concurrently for fog networks.

3. System Model and Problem Formulation

3.1. Multi-Interface System Model. In this paper, we consider the transmission scheduling problem in a typical multi-interface IoT scenario. Each IoT device is equipped with multiple antennas and can deliver data through different links concurrently. Their collected data are categorized into different types, forming transmission flows, respectively. The transmission scheduler operates at the IoT device and makes transmitting decisions according to current network condition. Hence, the multi-interface IoT system can be viewed as a model of a single node with multiple uplink channels. Table 1 summarizes the notations used in this paper.

TABLE 1: Notations used in this paper.

Notation	Description
M	Number of data flows
N	Number of available links
$\alpha^m(t)$	Arrived packet number of flow m at time t
$\mu_n(t)$	The transmission rate of link n at time t
$a_n^m(t)$	Number of packets from data flow m allocated to link n at t
$u_n^m(t)$	The transmission rate of link n assigned to data flow m at t
$Q_n^m(t)$	Queue backlog of data flow m scheduled to link n at t
$\bar{a}_n$	Time-average admitted packets of link n
$\Phi_n(\cdot)$	The utility function of link n
ω_n	Weight factor of link n
$Y_n(t)$	Length of virtual queue for link n
$\gamma_n(t)$	The auxiliary variable of link n at t
V	Nonnegative penalty parameter

To facilitate the analysis of the above model and deal with the time-varying link states, we assume that the system operates in discrete time with unit time slots $t \in \{0, 1, 2 \cdots \}$. At every time slot, packets randomly enter the transmission scheduler. We define $\alpha^m(t)$ as the amount of packets of flow m (in the unit of packet number) that arrive at time t. For link n, we use $\mu_n(t)$ to denote its maximum allowable rate, the number of packets that can be transmitted at time t. To make our model practical, we only assume $\alpha^m(t)$ and $\mu_n(t)$ are independent and identically distributed (i.i.d) over time and rate-convergent, which means that equations (1) and (2) hold with probability 1. Note that, in our work, the scheduler does not need to know the average arrival rate $\bar{\alpha}^m$ and transmission rate $\bar{\mu}_n$ previously.

$$\bar{\alpha}^m = \lim_{t \longrightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t} \alpha^m(\tau) < \infty, \tag{1}$$

$$\bar{\mu}_n = \lim_{t \longrightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t} \mu_n(\tau) < \infty. \tag{2}$$

In addition, we assume a maximum arrival rate $\alpha^{m,\max}$ and a maximum transmission rate $\mu_{n,\max}$, regardless of the time and the channel state, so that

$$0 \le \alpha^m(t) \le \alpha^{m,\max}, 0 \le \mu_n(t) \le \mu_{n,\max}. \tag{3}$$

Our goal in this work is to design a dynamic, optimal algorithm for the transmission scheduler to make the following decisions strategically: (1) *scheduling decision*: how many packets of flow m should be transmitted through link n at each time slot? (2) *Rate control*: how does the device allocate the transmission rate while ensuring system stability?

We next propose a multiqueue model to characterize the scheduling problem and then develop an optimization

framework to solve the first problem. After that, we introduce a virtual queue to cope with the second issue.

3.2. Queue Model and Optimization Objective

3.2.1. Queue Model. According to the above system, we assume that the transmission scheduler holds multiple queues for each transmitting link and flow, respectively, as shown in Figure 1. Let $Q_n^m(t)$ represent the queue backlog of flow m scheduled to link n on time t. At each slot t, the scheduler observes the arrived packets $\alpha_m(t)$ and then chooses an admission schedule policy $\mathbf{a_m}(t) = \{a_1^m(t), a_2^m(t), \cdots, a_n^m(t)\}$. $a_n^m(t)$ denotes the number of packets of flow m allocated to link n. Besides, it also determines a transmission vector $\mathbf{u_n}(t) = \{u_n^1(t), u_n^2(t), \cdots, u_n^m(t)\}$ based on the current link condition, where $u_n^m(t)$ represents the transmission rate of link n assigned to flow m. Hence, the queue length $Q_n^m(t)$ evolves according to the following equation.

$$Q_n^m(t+1) = \max \{Q_n^m(t) - u_n^m(t), 0\} + a_n^m(t). \tag{4}$$

To ensure system stability, we use the definition of queue stability in [23], which is given as follows.

$$\bar{Q} \triangleq \lim_{t \longrightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t} \sum_n \sum_m Q_n^m(t) \le \infty. \tag{5}$$

By definition, the queue length can be bounded by a positive constant, implying that the average queuing latency is also constrained.

Additionally, the admission decision $\mathbf{a_m}(t)$ is made subject to the constraint, $\alpha^m(t) = \sum_n a_n^m(t)$, implying that all packets should be admitted. Similarly, $\mathbf{u_n}(t)$ must satisfy $0 \le \sum_m u_n^m(t) \le \mu_n(t)$, which means that the delivered packet rate could not exceed the link capacity at any time.
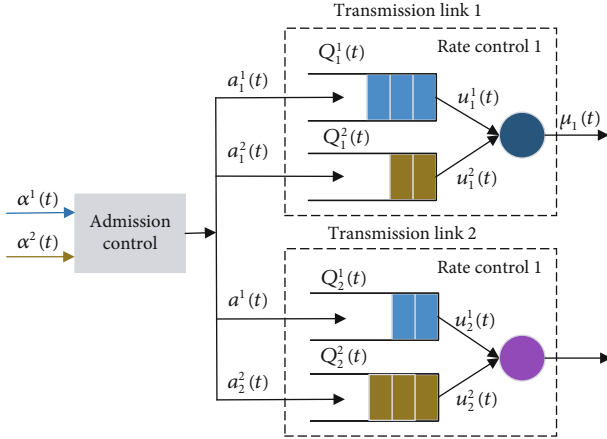
Figure 1: Illustration of the queue model at the IoT device.

3.2.2. Optimization Objective. To introduce the optimization objective, we define the time-average admitted packets on link n, $\bar{a}_n$.

$$\bar{a}_n = \lim_{t \to \infty} \frac{1}{t} \sum_{\tau=0}^{t} \sum_m a_n^m(t). \tag{6}$$

After that, we introduce a continuous, concave, and non-decreasing utility function, $\Phi_n(\bar{a}_n)$, to represent our optimization target. According to [23], this kind of utility function can be used to measure network fairness. Intuitively, this utility can be the profit gained by transmitting packets through link n or the reciprocal of transmission cost of link n. An example of $\Phi_n(\cdot)$ can be given by $\Phi_n(x) = \log(1 + x)$, which is also used in our simulation experiments. Hence, to achieve long-time utility maximization, we choose to maximize the weighted sum of all transmission utilities. Then, the formulated scheduling problem can be as follows.

$$\max \sum_n \omega_n \Phi_n(\bar{a}_n), \tag{7a}$$

$$\text{s.t.} \quad \alpha^m(t) = \sum_n a_n^m(t), \quad \forall t, n, m, \tag{7b}$$

$$0 \le \sum_m u_n^m(t) \le \mu_n(t), \quad \forall t, n, m, \tag{7c}$$

$$\text{All queues } Q_n^m(t) \text{ are stable}, \tag{7d}$$

where ω_n is the weight factor which can be used to dominate the transmission fairness. Additionally, we assume that $\Phi_n(\cdot)$ satisfies the Lipschiz condition, $\Phi_n(x) - \Phi_n(y) \le D|x - y|$, where D is a constant.

Intuitively, the above optimization problem is an integer programming problem. It is hard to derive the optimal solution directly. Moreover, it requires the full knowledge of state information of all time, which is almost impossible in the realistic environment.

In the next section, we will leverage the Lyapunov optimization technique to decompose the scheduling problem and

provide an online algorithm to approximate the optimal solution with system performance guaranteed.

4. Dynamic Scheduling Algorithm

4.1. Problem Transformation with Virtual Queue. To cope with the aforementioned transmission scheduling problem, we introduce a vector of auxiliary variables, $\gamma(\mathbf{t}) = \{\gamma_1(t), \gamma_2(t), \cdots, \gamma_n(t)\}$, with constraints $0 \le \gamma_n(t) \le \mu_{n,\max}$. Then, the problem ((7a), (7b), (7c), and (7d)) can be transformed as follows.

$$\max \sum_n \omega_n \Phi_n(\bar{\gamma}_n), \tag{8a}$$

$$\text{s.t.} a_n^m(t), u_n^m(t) \in \Omega, \quad \forall t, n, m, \tag{8b}$$

$$\bar{\gamma}_n \le \bar{a}_n, \quad \forall n, \tag{8c}$$

$$0 \le \gamma_n(t) \le \mu_{n,\max}, \quad \forall t, n, \tag{8d}$$

$$\text{All queues } Q_n^m(t) \text{ are stable}, \tag{8e}$$

where Ω denotes the state space given by (7a) and (7b) that arbitrary stationary algorithms can achieve. According to [24], the auxiliary variables, $\gamma(\mathbf{t})$, are introduced to decouple the variables from the optimization objective. Besides, in this paper, the auxiliary variables can also simplify the optimization problem by replacing multiple variables with a single variable.

The explicit explanation of the problem transformation can be as follows. According to the definition of Ω, the constraint (8a), which is equivalent to (7a) and (7b), stands for the feasible region of the problem ((7a), (7b), (7c), and (7d)). Then, if we always choose $\gamma_n(t) = \sum_m a_n^m(t)$, (8b) and (8c) are always satisfied, making these two problems the same. This is because the queue stability requirement constrains that the average arrival rate cannot exceed the maximum transmission rate. Besides, as $\Phi_n(\cdot)$ is nondecreasing, if the optimal solution of the problem ((8a), (8b), (8c), (8d), and (8e)) is obtained when constraint (8c) holds with inequality, it provides a utility that is at least as good as the optimal value of (7a), (7b), (7c), and (7d). Therefore, the scheduling policy determined by (8a), (8b), (8c), (8d), and (8e) also solves (7a), (7b), (7c), and (7d). Readers interested in the proof of the transformation can refer to [24–26].

To satisfy (8c), we introduce a virtual queue $Y_n(t)$ for each link n, with an arrival rate of $\gamma_n(t)$ and a departure rate of $\sum_m a_n^m(t)$, which evolves as

$$Y_n(t+1) = \max \left\{ Y_n(t) - \sum_n a_n^m(t) + \gamma_n(t), 0 \right\}. \tag{9}$$

According to [23], if this virtual queue is stable, the time-average value of $\sum_n a_n^m(t)$ is greater than or equal to the time-average value of $\gamma_n(t)$, which ensures (8c), and so the optimization objective will also be large.

4.2. Problem Decomposition via Drift-plus-Penalty Minimization. Next, we solve the problem ((8a), (8b), (8c), (8d), and (8e)) via the drift-plus-penalty framework, aimed at developing a dynamic scheduling algorithm that can preserve the queue stability and solution optimality.

Let $\mathbf{\Theta}(t) \triangleq [\mathbf{Q}(t), \mathbf{Y}(t)]$ be the collective vector of all $Q_n^m(t)$ and $Y_n(t)$ queues. Then, we define the following quadratic *Lyapunov function*.

$$L(\mathbf{\Theta}(t)) = \frac{1}{2}\left[\sum_n \sum_m (Q_n^m(t))^2 + \sum_n (Y_n(t))^2\right]. \quad (10)$$

Denote $\Delta(\mathbf{\Theta}(t)) = L(\mathbf{\Theta}(t+1)) - L(\mathbf{\Theta}(t))$ as the *one-step conditional Lyapunov drift*. According to [23], we can determine our scheduling policy by minimizing the following *drift-plus-penalty* expression at each time slot.

$$\Delta(\mathbf{\Theta}(t)) - V\mathbb{E}\left\{\sum_n \omega_n \Phi_n(\gamma_n(t))|\mathbf{\Theta}(t)\right\}, \quad (11)$$

where V is a nonnegative penalty parameter that will affect the utility delay trade-off. Instead of minimizing equation (11), our algorithm minimizes its upper bound, which has the following expression.

$$
\begin{aligned}
\Delta(\mathbf{\Theta}(t)) &- V\mathbb{E}\left\{\sum_n \omega_n \Phi_n(\gamma_n(t)) \mid \mathbf{\Theta}(t)\right\} \leq B \\
&- V\mathbb{E}\left\{\sum_n \omega_n \Phi_n(\gamma_n(t)) \mid \mathbf{\Theta}(t)\right\} \\
&+ \sum_n \sum_m Q_n^m(t)\mathbb{E}\{(a_n^m(t) - u_n^m(t)) \mid \mathbf{\Theta}(t)\} \\
&+ \sum_n Y_n(t)\mathbb{E}\left\{\left(\gamma_n(t) - \sum_m a_n^m(t)\right)\mid\mathbf{\Theta}(t)\right\},
\end{aligned} \quad (12)
$$

where B is a positive constant satisfying

$$B \geq \frac{1}{2}\left[\sum_n \sum_m (a_n^m(t)^2 + u_n^m(t)^2) + \sum_n \left(\gamma_n(t)^2 + \left(\sum_m a_n^m(t)\right)^2\right)\right]. \quad (13)$$

As $a_n^m(t)$, $u_n^m(t)$, and $\gamma_n(t)$ are all bounded, it is easy to prove that such a constant b always exists for all t. For example, an upper bound of B can be defined as $1/2(1 + |M|)(|N||M|\alpha_{max}^2 + |N|\mu_{max}^2))$, where α_{max} and μ_{max} denote the maximum value of the arrival and transmission rate, respectively.

Our dynamic algorithm is given to make scheduling decisions by minimizing the right hand of equation (12) in each time slot t, based on the observed queue states, $Q_n^m(t)$ and $Y_n(t)$. Furthermore, by observing the form of equation (12), the scheduling decisions for $a_n^m(t)$, $u_n^m(t)$, and $\gamma_n(t)$ are linearly coupled. Thus, we can decompose the scheduling problems into the following three subproblems *auxiliary variable selection*, *admission control*, and *transmission rate allocation*.

(1) *Auxiliary Variable Selection*. The auxiliary variables can be chosen by minimizing the following expression:

$$
\begin{aligned}
\min \quad & \sum_n (Y_n(t)\gamma_n(t) - V\omega_n\Phi(\gamma_n(t))) \\
\text{s.t.} \quad & 0 \leq \gamma_n(t) \leq \mu_{n,max}.
\end{aligned} \quad (14)
$$

As $\gamma(t)$ is linearly coupled, $\gamma_n(t)$ can be separable with each other. Besides, given that $\Phi_n(\gamma_n(t))$ is continuous and concave, we can provide a closed-form solution of (14), $[\Phi_n'^{-1}(Y_n(t)/(V\omega_n))]_0^{\mu_{n,max}}$, where $\Phi_n'^{-1}(\cdot)$ is the inverse of $\Phi_n(\cdot)$'s derivative and $[x]_y^z$ denotes $\min(\max(x,y),z)$.

(2) *Admission Control*. By omitting the terms containing $a_n^m(t)$, the admission control can be determined by solving the following problem:

$$
\begin{aligned}
\min \quad & \sum_m \left[\sum_n (Q_n^m(t) - Y_n(t)a_n^m(t))\right] \\
\text{s.t.} \quad & \alpha^m(t) = \sum_n a_n^m(t).
\end{aligned} \quad (15)
$$

In the case of the separable objective, problem (15) can be divided into m independent subproblems, each for one flow. The optimal solution can be as follows.

$$a_n^m(t) = \begin{cases} \alpha^m(t), & \arg\min(Q_n^m(t) - Y_n(t)), \\ 0, & \text{otherwise}, \end{cases} \quad (16)$$

(3) *Transmission Rate Allocation*. Similarly, the transmission rate allocation variable, $u_n^m(t)$, can be obtained by maximizing the following expression:

$$
\begin{aligned}
\max \quad & \sum_m \sum_n Q_n^m(t)u_n^m(t) \\
\text{s.t.} \quad & 0 \leq \sum_m u_n^m(t) \leq \mu_n(t).
\end{aligned} \quad (17)
$$

Intuitively, the transmission rate allocation algorithm always trends to serve the longest queue.

The scheduling problem can be described by Algorithm 1. At every time slot t, the scheduler receives the packets, $a^m(t)$, and observes the current states of all queues, $Q_n^m(t)$ and $Y_n(t)$. Then, it makes scheduling decisions by solving (14), (15), and (17). After that, it updates the virtual queues $Y_n(t)$ according to equation (9) and the actual queues $Q_n^m(t)$ according to equation (4), with the derived $\gamma_n(t)$, $a_n^m(t)$, and $u_n^m(t)$.

Initialization: Set $t = 0$ and choose a nonnegative penalty parameter V; set up transmission queues $Q_n^m(0) = 0$ and virtual queues $Y_n(0) = 0$ for all n and m;
In each time slot t:
1: The scheduler receives the arrived packets $a^m(t)$ and observes the current queue states $Q_n^m(t)$ and $Y_n(t)$;
2: Solve (14), (15), and (17) to obtain the decision variable $\gamma_n(t)$, $a_n^m(t)$, and $u_n^m(t)$ as follows;
3: **for** $n \in N$ **do**
4: $\gamma_n(t) \longleftarrow [\Phi_n'^{-1}(Y_n(t)/(V\omega_n))]_0^{\mu_{n,max}}$;
5: $u_n^m(t) \longleftarrow \mu_n(t)$, if $m = \text{argmax}(Q_n^m(t))$;
6: **end for**
7: **for** $m \in M$ **do**
8: $a_n^m(t) \longleftarrow \alpha^m(t)$, if $n = \text{argmin}(Q_n^m(t) - Y_n^m(t))$
9: **end for**
10: Assign packets to queue $Q_n^m(t)$ according to $a_n^m(t)$, and deliver packets through link n based on $u_n^m(t)$;
11: Update all queues according to (4) and (9).

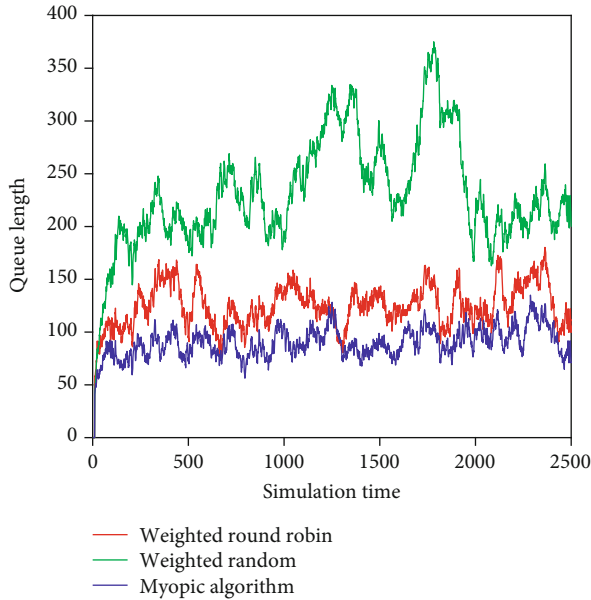ALGORITHM 1: Pseudocode of the proposed algorithm.

FIGURE 2: Queue length of the weighted round robin, weighted random, and myopic algorithm.

4.3. Performance Analysis. Here, we provide the performance of our dynamic algorithm, with respect to the queue length and utility.

Theorem 1. *Suppose the problem ((8a), (8b), (8c), (8d), and (8e)) is feasible, all queues are initially set to be zero, and the above algorithm is used in each time slot t with a fixed V; then:*

(a) All queues $Q_n^m(t)$ are stable for all t. Specifically, the following equation holds:

$$\limsup_{T \longrightarrow \infty} \frac{1}{T} \sum_{t=0}^{T} \sum_n \sum_m Q_n^m(t) \leq \frac{B + V|W|DR}{\varepsilon}, \quad (18)$$

where D is the Lipschiz constant, R represents the region of γ, ε

is a positive constant, and $|W|$ denotes the 1-norm of the weight factor.

(b) The utility achieved by the proposed algorithm satisfies

$$\liminf_{t \longrightarrow \infty} \left(\Phi^{opt} - \sum_n \omega_n \Phi(\bar{a}_n) \right) \leq \frac{B}{V}, \quad (19)$$

where Φ^{opt} is the maximum utility of the problem ((7a), (7b), (7c), and (7d)).

Proof. According to Theorem 4.5 in [23], the feasibility of the problem ((7a), (7b), (7c), and (7d)) implies that for any $\delta, \delta > 0$, there exists a scheduling policy $\widehat{\gamma_n}(t)$, $\widehat{a_n^m}(t)$, and $\widehat{u_n^m}(t)$, which yields

$$\mathbb{E}\left\{ \widehat{a_n^m}(t) - \widehat{u_n^m}(t) \right\} \leq \delta,$$

$$\mathbb{E}\left\{ \widehat{\gamma_n}(t) - \sum_m \widehat{a_n^m}(t) \right\} \leq \delta, \quad (20)$$

$$\Phi^{opt} - \Phi_n\left(\widehat{\gamma_n}(t)\right) \leq \delta.$$

Using $\delta = 0$ and plugging the above into equation (12) yield

$$\Delta(\Theta(t)) - V\mathbb{E}\left\{ \sum_n \omega_n \Phi_n(\gamma_n(t)) \right\} \leq B - V\Phi^{opt}. \quad (21)$$

Summing up the above equation of all t, $t \leq T$, and dividing both side by T, we can have

$$\frac{1}{T} \sum_{t=0}^{T} \left(\Phi^{opt} - \mathbb{E}\left\{ \sum_n \omega_n \Phi_n(\gamma_n(t)) \right\} \right) \leq \frac{B}{V} + \frac{\mathbb{E}\{L(\Theta(0))\}}{VT}. \quad (22)$$
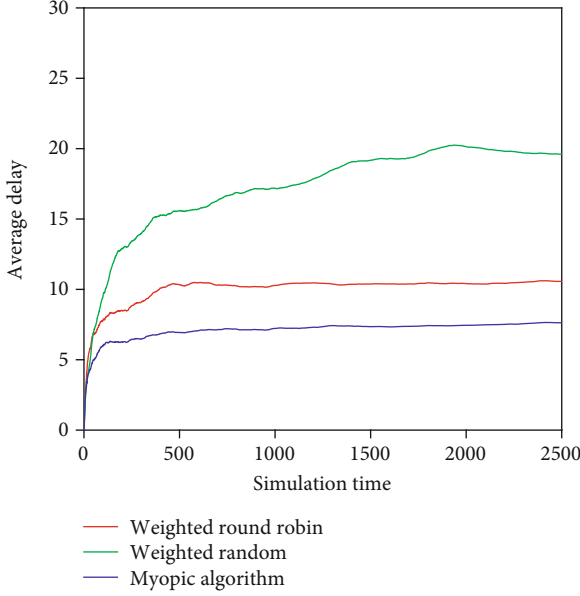
FIGURE 3: Average delay of the weighted round robin, weighted random, and myopic algorithm.
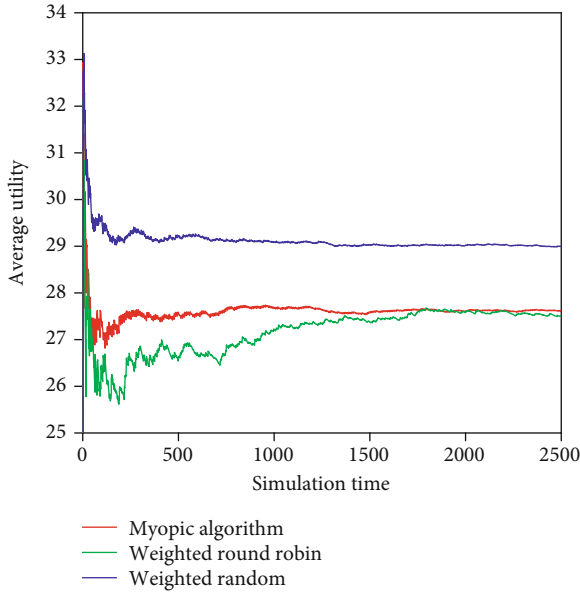

FIGURE 4: Average utility of the weighted round robin, weighted random, and myopic algorithm.

As $\Phi_n(\cdot)$ is concave and nondecreasing, based on Jensen's inequality, the following equation can be derived.

$$\frac{1}{T}\sum_{t=0}^{T}\sum_{n}\omega_n\Phi_n(\gamma_n(t)) \leq \sum_{n}\omega_n\Phi_n(\bar{\gamma}_n) \leq \sum_{n}\omega_n\Phi_n(\bar{a}_n). \quad (23)$$

Plugging the above equation into equation (22) and taking $\liminf_{T\rightarrow\infty}$ of both sides prove part (b).

To prove part (a), we can use $\widehat{\gamma_n}(t) = \mathbb{E}\{\sum_m \widehat{a_n^m}(t)\}$. Then, according to the queue stability constraint, we have the following equations.

$$\mathbb{E}\left\{\widehat{a_n^m}(t) - \widehat{u_n^m}(t)\right\} \leq -\varepsilon,$$
$$\mathbb{E}\left\{\Phi_n\left(\widehat{\gamma_n}(t)\right)\right\} = \Phi_n(\varepsilon). \quad (24)$$

Plugging these two equations in equation (12), we can have

$$\Delta(\Theta(t)) - V\mathbb{E}\left\{\sum_{n}\omega_n\Phi_n(\gamma_n(t))\right\} \leq B - V\Phi_n(\varepsilon) - \varepsilon\sum_{n}\sum_{m}Q_n^m(t). \quad (25)$$

Rearranging the terms of the above equation and using telescoping yield

$$\varepsilon\sum_{t=0}^{T}\sum_{n}\sum_{m}Q_n^m(t) \leq BT - V\sum_{t=0}^{T}\sum_{n}\omega_n\Phi_n(\varepsilon) + V\sum_{t=0}^{T}\mathbb{E}\left\{\sum_{n}\omega_n\Phi_n(\gamma_n(t))|\Theta(t)\right\}. \quad (26)$$

Consider that $\Phi_n(\cdot)$ satisfies the Lipschitz condition. Then, the following equation can be derived.

$$\sum_{n}\omega_n(\Phi_n(\gamma_n(t)) - \Phi_n(\varepsilon)) \leq D\sum_{n}\omega_n|\gamma_n(t) - \varepsilon| \leq |W|DR, \quad (27)$$

Using this in equation (26) and dividing both sides by εT, we can derive for all $t > 0$

$$\frac{1}{T}\sum_{t=0}^{T}\sum_{n}\sum_{m}Q_n^m(t) \leq \frac{B + V|W|DR}{\varepsilon} + \frac{L(\Theta(0))}{\varepsilon T}. \quad (28)$$

Finally, taking $\limsup_{T\rightarrow\infty}$ of both sides proves part (a).

Theorem 1 demonstrates that the control parameter V dominates the performance of our algorithm. In particular, the upper bound of the queue length grows linearly with V. This situation implies we can choose a small V to achieve a short queue length, which corresponds to a short queuing delay according to *Little's theory*. Besides, it also confirms that the gap between the achieved utility and the optimal one is bounded by $O(1/V)$. The gap can be made arbitrarily small by increasing V, which declares that increasing V can approximate the optimal solution of (8a), (8b), (8c), (8d), and (8e). Overall, we can conclude that our algorithm achieves the trade-off between the queuing delay and the transmission utility with $[O(V), O(1/V)]$. This feature shows that improving the transmission utility is at the expense of increasing the queuing delay.

Additionally, consider that (14), (15), and (17) are independent. All these subproblems can be solved in parallel. This situation indicates that the overall complexity of our algorithm depends on that of each subproblem. Furthermore, by observing their forms, the complexity of solving (14),

(a) Queue length vs. arrival rate

(b) Average delay vs. arrival rate

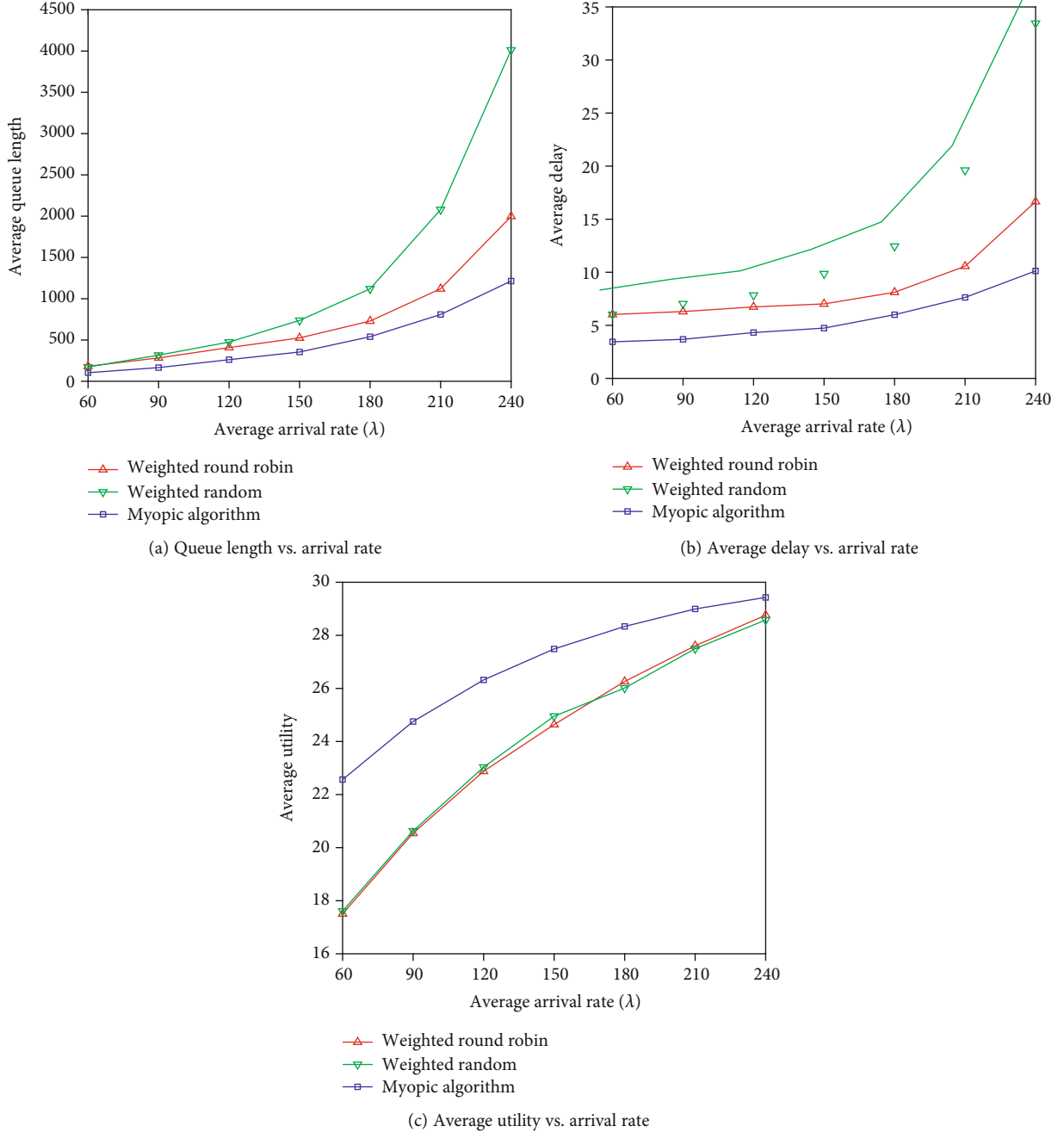(c) Average utility vs. arrival rate

Figure 5: Average queue length, average delay, and average utility with different average arrival rates (varying from 60 to 240).

(15), and (17) is related to the number of data flows and links, M and N. In particular, the computation complexity is bounded by $O(MN)$. Normally, N is constant in practical scenarios, and M varies with the number of data types, which is relatively small. Thus, the complexity of our algorithm is acceptable.

5. Simulation Results

This section presents the simulation results of the proposed algorithm, named myopic algorithm. For comparison, we also simulate two other heuristic algorithms as benchmarks:

(1) *weighted round robin*: the scheduler admits and sends packets of different flows in turn and the time slice of each flow is determined based on its average arrival and transmission rates, and (2) *weighted random*: the scheduler admits and transmits packets of different flows according to a given probability distribution, proportional to the average arrival and transmission rates.

5.1. Experimental Setting. In this work, we consider a 5×3 wireless transmission scheduler with five input ports, each admitting one certain flow, and three output ports, each delivering packets through one wireless link [8]. Hence, there
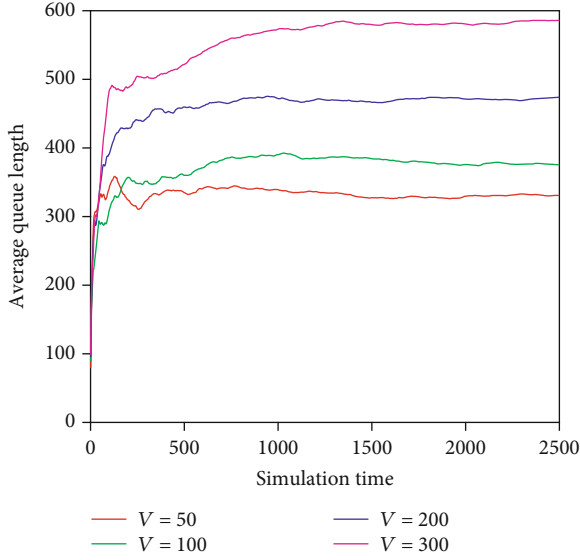
FIGURE 6: Queue length with different V (varying from 50 to 300).
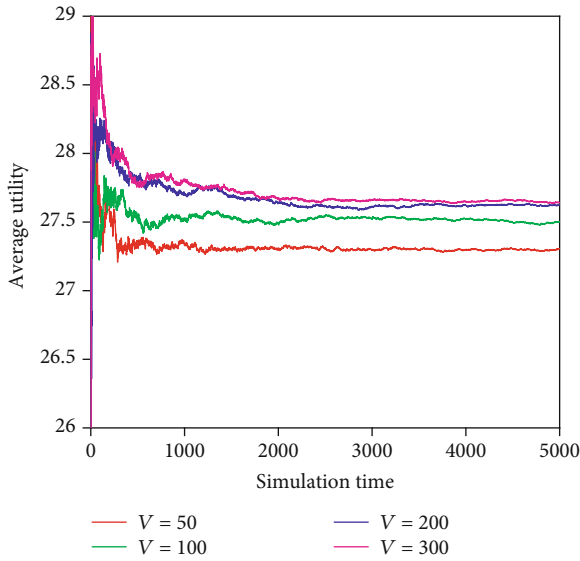
FIGURE 7: Average utility with different V (varying from 50 to 300).

are fifteen queues $Q_n^m(t)$, representing that packets arrived from input port m must be delivered to output port n, for $m \in \{1, 2, 3, 4, 5\}$ and $n \in \{1, 2, 3\}$. The arrival processes follow uniform distributions, i.i.d. over time slots, varying from 60 to 240. Considering that link states are time-varying, transmission rates of all links are arbitrarily distributed, i.i.d. over time slots with average rates $\{30,90,150\}$. We use $\log(1 + x)$ as the utility function for all links and, respectively, set the maximum arrival rate and transmission rate to twice the average rates of the two. We implement a software scheduler with Python and deploy all algorithms in it. To better evaluate the proposed algorithms, we run these three algorithms simultaneously with the same instantaneous arrival and transmission rates. Note that all simulations are

over 5000 time slots, and each data point in the figures is averaged based on 10 times independent runs.

5.2. System Performance. In this section, we verify the system performance with respect to queue length, average delay, and achieved utility, with $V = 100$.

Figure 2 plots the queue length of the three algorithms over time slots. The results confirm that our algorithm outperform the other two algorithms. In particular, the weighted random algorithm varies dramatically over time, while the other algorithms are relatively stable. This is because only the weighted random algorithm schedules the packet randomly, without considering the environment state. Definitely, it can only guarantee that the queue length cannot grow indefinitely.

Figure 3 compares the average delay between the three algorithms. We can see that the myopic algorithm achieves the lowest average delay, followed by myopic weighted round robin, and the weighted random algorithm shows the worst performance. This situation is consistent with the queue length. *Little's theory*, $N = \lambda T$, can provide a reasonable explanation that the average delay grows proportionally with the arrival rate.

In our simulation, we also measure the achieved utility, shown in Figure 4, calculated by the number of packets. Obviously, all these three algorithms converge to a stable value, and our proposed algorithms show a better performance than the benchmarks. The former situation demonstrates the system's stability. The latter is because our proposed algorithms leverage the utility function as targets while the benchmarks can only be designed to ensure system stability. Moreover, the weighted random algorithm converges the slowest. Similar to the queue length, this phenomenon is also induced by the weighted random algorithm's randomicity.

5.3. Impact of Algorithm Parameters. In this section, we investigate the impact of the arrival rate and V on system performance.

Figure 5 reveals how the queue length, average delay, and achieved utility vary with the arrival rate. Figure 5(a) shows that the queue length increases with the increase in λ. In particular, the weighted random algorithm grows the fastest, followed by the weighted round robin algorithm, and our proposed algorithm has the shortest queue lengths. Figure 5(b) shows the same trend as Figure 5(a). These two subfigures demonstrate that our proposed algorithm can always provide queuing delay guarantees. As for Figure 5(c), it reveals that the achieved utility also grows as λ rises. However, the increasing rates of the proposed algorithm decrease with λ. This trend can be explained as follows. When λ is small, the primary objective is to maximize the utility. However, when λ is large, system stability becomes essential. Otherwise, the queue length will grow infinitely. As a comparison, the weighted round robin and weighted random algorithms show an increasing trend almost linearly.

Figure 6 shows the average queue length under different values of V. Obviously, for all cases, the queue length converges to be stable. In particular, we can see that, when V

grows, the stable queue length also increases. Besides, the increasing rates of the stable queue length and V have a linear approximation relationship. Figure 7 depicts the average utility under different values of V. Similar to the queue length, all average utilities finally remain stable and increase with the growth of V. However, the growth rate of the average utility is inversely proportional to that of V. Hence, these two figures together confirm the correctness of the aforementioned theorems.

Another observation from these two figures is that a larger V results in a higher convergence rate. As shown in Figure 6, the convergence time of $V = 300$ is more than 1500 time slots, which is almost three times that of $V = 50$. Moreover, in Figure 7, when $V = 300$, the average utility remains stable after the 3000th time slots, and when $V = 50$, the algorithm reaches its stable value around the 1000th time slots. Therefore, it is necessary to choose V carefully when deploying our algorithms in practical systems.

6. Conclusion

In this paper, we investigate the transmission scheduling in IoT. We propose a generic optimization problem and solve it via the Lyapunov optimization technique. Both theoretical proofs and simulation results confirm that our approach can achieve the optimal utility while guaranteeing system stability and constraining average delay. In the future, we will continue our work in two aspects. (1) Extend our algorithms into more scenarios, i.e., delay-sensitive scenarios, where the delay requirement of a specific flow must be satisfied for all times. (2) Improve the performance of our proposed algorithm, i.e., investigating how to choose parameters dynamically to provide real-time transmission resource allocation.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A vision of IoT: applications, challenges, and opportunities with China perspective," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 349–359, 2014.

[2] A. H. H. Ngu, M. Gutierrez, V. Metsis, S. Nepal, and M. Z. Sheng, "IoT middleware: a survey on issues and enabling technologies," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 1–20, 2017.

[3] Y. Zhang, C. Xu, H. Li, K. Yang, N. Cheng, and X. Shen, "PROTECT: efficient password-based threshold single-sign-on authentication for mobile users against perpetual leakage," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2297–2312, 2021.

[4] K. Peng, H. Huang, S. Wan, and V. C. M. Leung, "End-edge-cloud collaborative computation offloading for multiple mobile users in heterogeneous edge-server environment," *Wireless Networks*, vol. 7, no. 4, pp. 2622–2629, 2020.

[5] G. Forecast, "Cisco visual networking index: global mobile data traffic forecast update, 2017–2022," *Update*, vol. 2017, 2019.

[6] H. Zhang, W. Quan, H. Chao, and C. Qiao, "Smart identifier network: a collaborative architecture for the future internet," *IEEE Network*, vol. 30, no. 3, pp. 46–51, 2016.

[7] Y. Zhang, C. Xu, N. Cheng, H. Li, H. Yang, and X. Shen, "Chronos+: an accurate blockchain-based time-stamping scheme for cloud storage," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 216–229, 2019.

[8] H. Zhang, P. Dong, S. Yu, and J. Song, "A scalable and smart hierarchical wireless communication architecture based on network/user separation," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 18–24, 2017.

[9] X. Deng, J. Luo, L. He, Q. Liu, X. Li, and L. Cai, "Cooperative channel allocation and scheduling in multi-interface wireless mesh networks," *Peer-to-peer Networking and Applications*, vol. 12, no. 1, pp. 1–12, 2019.

[10] C. Yu, W. Quan, N. Cheng, S. Chen, and H. Zhang, "Coupled or uncoupled? multi-path TCP congestion control for high-speed railway networks," in *2019 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 612–617, Changchun, China, 2019.

[11] P. Dong, T. Zheng, S. Yu, H. Zhang, and X. Yan, "Enhancing vehicular communication using 5G-enabled smart collaborative networking," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 72–79, 2017.

[12] M. Wang, C. Xu, X. Chen, H. Hao, L. Zhong, and D. O. Wu, "Design of multipath transmission control for information-centric Internet of things: a distributed stochastic optimization framework," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9475–9488, 2019.

[13] L. P. Do and R. Lehnert, "Dynamic resource allocation protocol for large PLC networks," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, pp. 41–46, Beijing, China, 2012.

[14] H. Tang and Z. Ding, "Mixed mode transmission and resource allocation for D2D communication," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 162–175, 2016.

[15] M. Mahdian, S. Arianfar, J. Gibson, and D. Oran, "MIRCC: multipath-aware ICN rate-based congestion control," in *Proceedings of the 3rd ACM Conference on Information-Centric Networking (ACM-ICN '16)*, pp. 1–10, New York, USA, 2016.

[16] Z. Allybokus, K. Avrachenkov, J. Leguay, and L. Maggi, "Multi-path alpha-fair resource allocation at scale in distributed software-defined networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 12, pp. 2655–2666, 2018.

[17] L. Chen, Y. Feng, B. Li, and B. Li, "Promenade: proportionally fair multipath rate control in datacenter networks with

random network coding," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 11, pp. 2536–2546, 2019.

[18] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. M. Leung, "Energy-efficient resource allocation in NOMA heterogeneous networks," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 48–53, 2018.

[19] Y. Gao, Y. Cui, X. Wang, and Z. Liu, "Optimal resource allocation for scalable mobile edge computing," *IEEE Communications Letters*, vol. 23, no. 7, pp. 1211–1214, 2019.

[20] P. Wang, C. Yao, Z. Zheng, G. Sun, and L. Song, "Joint task assignment, transmission, and computing resource allocation in multilayer mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2872–2884, 2019.

[21] Y. Jia and A. Vosoughi, "Transmission resource allocation for training based amplify-and-forward relay systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 450–455, 2011.

[22] M. Mukherjee, S. Kumar, M. Shojafar, Q. Zhang, and C. X. Mavromoustakis, "Joint task offloading and resource allocation for delay-sensitive fog networks," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–7, Shanghai, China, 2019.

[23] M. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[24] L. Georgiadis, M. Neely, and L. Tassiulas, *Resource Allocation and Cross-Layer Control in Wireless Networks*, Foundations and Trends in Networking, Hanover, MA, 2006.

[25] M. J. Neely, "Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1489–1501, 2006.

[26] M. Neely, "Delay-based network utility maximization," *IEEE/ACM Transactions on Networking*, vol. 21, no. 1, pp. 41–54, 2013.

WILEY | Hindawi

Research Article

Adaptive-Weighted Multiview Deep Basis Matrix Factorization for Multimedia Data Analysis

Shicheng Li [iD],[1] **Qinghua Liu** [iD],[1] **Jiangyan Dai** [iD],[2] **Wenle Wang** [iD],[1] **Xiaolin Gui** [iD],[1] **and Yugen Yi** [iD][1]

[1]*School of Software, Jiangxi Normal University, Nanchang 330022, China*
[2]*School of Computer Engineering, Weifang University, Weifang 261061, China*

Correspondence should be addressed to Qinghua Liu; 004799@jxnu.edu.cn and Yugen Yi; yiyg510@jxnu.edu.cn

Feature representation learning is a key issue in artificial intelligence research. Multiview multimedia data can provide rich information, which makes feature representation become one of the current research hotspots in data analysis. Recently, a large number of multiview data feature representation methods have been proposed, among which matrix factorization shows the excellent performance. Therefore, we propose an adaptive-weighted multiview deep basis matrix factorization (AMDBMF) method that integrates matrix factorization, deep learning, and view fusion together. Specifically, we first perform deep basis matrix factorization on data of each view. Then, all views are integrated to complete the procedure of multiview feature learning. Finally, we propose an adaptive weighting strategy to fuse the low-dimensional features of each view so that a unified feature representation can be obtained for multiview multimedia data. We also design an iterative update algorithm to optimize the objective function and justify the convergence of the optimization algorithm through numerical experiments. We conducted clustering experiments on five multiview multimedia datasets and compare the proposed method with several excellent current methods. The experimental results demonstrate that the clustering performance of the proposed method is better than those of the other comparison methods.

1. Introduction

With the rapid development of computer technology, the collected multimedia data from many research fields, such as computer vision, image processing, and natural language processing, always have features with high dimension and complex structures. These high-dimensional data can not only provide abundant information but also bring some problems such as the "curse of dimensionality" [1, 2]. Therefore, how to effectively deal with high-dimensional data has become a widespread concern [3]. Dimensionality reduction is an efficient way to solve this issue, which can map the original data to a low-dimensional space and obtain a low-dimensional representation derived from the hidden information in the original data [4].

In recent years, many dimensionality reduction methods have been proposed for multimedia data [5]. The matrix factorization method has become one of the research hotspots

owing to its simple theoretical basis and easy implementation. Principal component analysis (PCA) [6], independent components analysis (ICA) [7], vector quantization (VQ) [8], etc. are well-known matrix factorization methods that can obtain a low-rank approximation matrix by decomposing a high-dimensional data matrix, and they can effectively extract a low-dimensional representation from high-dimensional data. However, these methods do not utilize any constraints on the matrix elements during the process of matrix decomposition. It means that the results allow negative elements, which give rise to the loss of physical meaning in low-dimensional representations. To solve this problem, Lee et al. added nonnegative constraints into matrix decomposition and proposed a nonnegative matrix factorization (NMF) [9] method. The low-dimensional feature representations obtained by NMF method are part-based so that they have strong interpretability. Consequently, NMF has attracted the wide attention of researchers. There are a large

number of improved algorithms based on NMF have been emerged, which have achieved great success in computer vision, natural language processing, speech recognition, DNA sequence analysis, and other areas [10–13].

NMF decomposes the original nonnegative data matrix into the product of a nonnegative basis matrix and a nonnegative coefficient matrix (also called low-dimensional feature matrix). The original data can be expressed as a linear combination of basis matrices, and the combination coefficients can form the coefficient matrix. Since NMF uses nonnegative constraints, it reflects the intuitive notion of combining parts to form a whole and has better interpretability than other methods. The obtained experimental results indicate that NMF has achieved good performance on image and document clustering tasks. Nevertheless, the traditional NMF method only considers the nonnegativity constraints of the elements, which may result in the obtained basis matrix having poor sparseness and independence. To solve the above problems, researchers have imposed additional constraints on the basis matrix or the coefficient matrix and proposed a series of improved methods. For instance, Hoyer [14] designed a sparsity measurement criterion and proposed an NMF variant with sparsity constraints (NMF-SC). Moreover, to enhance the independence of the obtained basis matrices and low-dimensional representation, Choi [15] proposed orthogonal nonnegative matrix factorization (ONMF), which imposed orthogonal constraints on the basis matrix and the coefficient matrix. However, the above methods have nonnegative limitations on the original data, thereby limiting the applicability of these NMF-based algorithms. Therefore, Ding et al. [16] proposed a semi-nonnegative matrix factorization (SNMF). Different from traditional NMF, SNMF relaxed the limitations on the original data and coefficient matrix and only imposed a nonnegative constraint on the basis matrix. The methods mentioned above have better capabilities than their predecessors for feature extraction and achieved better results in real-world tasks, but they only extracted shallow features [17].

In recent years, deep learning has exhibited outstanding performance in feature representation tasks [18–20]. Therefore, many researchers have introduced deep learning into matrix factorization and proposed a large number of deep feature representation methods [21–27]. Ahn et al. [21] proposed multilayer nonnegative matrix factorization (MNMF). Different from traditional NMF-based approaches, MNMF decomposed the coefficient matrix several times to obtain an underlying part-based representation that can extract deep hierarchical features from the original data. In addition, to expand the application scope, Trigeorgis et al. [22] integrated deep factorization and semi-NMF to propose a deep semi-nonnegative matrix factorization (deep semi-NMF) method. However, both MNMF and deep semi-NMF only considered the deep decomposition of the coefficient matrix for the training data. For the new test data, the basis matrix was used to obtain the deep low-dimensional representation. Therefore, the basis matrix directly affected the results of the deep low-dimensional representation. To obtain a more accurate deep low-dimensional representation of the original

data matrix, Zhao et al. [23] applied deep factorization to the basis matrix and proposed a deep NMF method based on basis image learning.

With the rapid development of the Internet and data collection technology, a large amount of multiview multimedia data can be easily acquired [28–30]. For example, an object can be shot from different views. An image can be described with different types of features such as color, texture, and shape. These multiview multimedia data can provide different information for each view, but they also contain potential correlations among these different views. Furthermore, they contain more information than single-view data. It is possible to simply integrate multiview data into single-view data, which ignores the differences and potential correlations between the various views of the data [28–30].

Consequently, extensive multiview data dimensionality reduction methods have been proposed [31–33]. Liu et al. [34] proposed a multiview NMF (multi-NMF) method which established the relationship between different perspectives by learning the common coefficient matrix among different views. Subsequently, Chang et al. [35] introduced a new regularization term into the multi-NMF and used it for clothing image clustering. Inspired by ONMF, Liang et al. [36] proposed NMF with coorthogonal constraints (NMFCC) for multiview multimedia data clustering. Additionally, to consider the correlations between multiple views, Zhan et al. [37] jointly optimized the graph matrix and concept factorization process and proposed an adaptive structure concept factorization (ASCF) method for multiview clustering. Although the above methods can handle multiview multimedia data well, they still belong to the class of feature representation method based on shallow factorization [38, 39]. The underlying deep features in the multiview data are still not available. Therefore, Zhao et al. [40] maximized the mutual information between various views, which forced the nonnegative representation of the last layer in each view to be as similar as possible. Then, the deep semi-NMF method was applied to multiview multimedia data clustering. Different from the existing studies, to adaptively provide feature weights for different perspectives in the multiperspective deep feature representation procedure, Huang et al. introduced an adaptive-weighted framework into the multiview deep semi-NMF and proposed an adaptive-weighted multiview clustering method based on deep matrix factorization [41]. Unlike the literature [40], it can adaptively assign weights to different views in a multiview deep feature representation. However, these methods still consider only the deep decomposition of the coefficient matrix. Therefore, an adaptive-weighted multiview deep basis matrix factorization (AMDBMF) is proposed for multimedia data clustering in this paper. Different from the above methods, AMDBMF first decomposes the basis matrix using a deep way on the data of each view simultaneously and then integrates the low-dimensional features of all view through the adaptive weighting mechanism to extract more accurate multiperspective deep low-dimensional representations. The flowchart of the proposed AMDBMF approach is shown in Figure 1. At last, we perform extensive experiments on five publicly
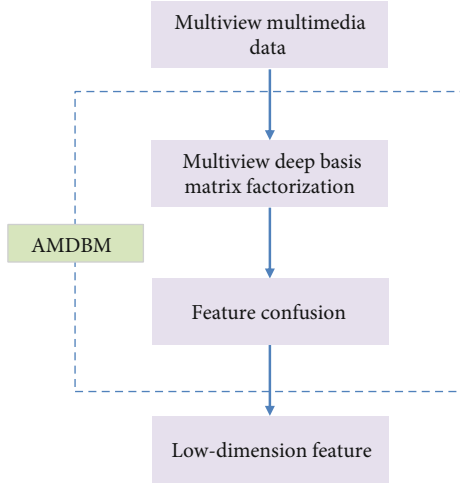
FIGURE 1: The flowchart of the proposed AMDBMF approach.

available multiview multimedia datasets. These experimental results show that the proposed AMDBMF approach outperforms the existing related approaches.

The remainder of this paper is organized as follows. "Related Works" describes the related algorithms including NMF and deep semi-NMF briefly. "Adaptive-Weighted Multiview Deep Basis Matrix Factorization" introduces the adaptive-weighted multiview deep basis matrix factorization (AMDBMF) algorithm in detail. The experimental results and analysis are discussed in "Experiments and Analysis." Finally, the conclusions are given in "Conclusions and Future Work."

2. Related Works

2.1. Nonnegative Matrix Factorization. Suppose that the given multimedia data can be represented as $X = [x_1, \cdots, x_N] \in R_+^{D \times N}$, where D is the dimensionality of the data and N is the number of samples. Each sample can be represented as a D-dimensional feature vector $x_j (1 \leq j \leq N)$. NMF is aimed at finding two low-ranking nonnegative matrices $W = [w_1, \cdots, w_d] \in R_+^{D \times d}$ and $H = [h_1, \cdots, h_N] \in R_+^{d \times N} (k \ll N \text{ and } d \ll D)$ that fulfill $X \approx WH$. After obtaining W and H, the original data can be expressed as $x_j \approx \sum_{i=1}^{d} w_i h_{ij}$, that is, each sample can be expressed as a linear combination of the basis matrix $W = [w_1, \cdots, w_d]$, and the coefficient vector is h_j. Therefore, the matrices W and H are called the basis matrix and coefficient matrix, respectively. The objective function of NMF is defined as follows:

$$\min_{W,H} \|X - WH\|_F^2$$
$$\text{s.t.} \ W \geq 0, H \geq 0, \tag{1}$$

where $\|\|_F$ is the Frobenius norm operation.

According to the Karush-Kuhn-Tucker (KKT) condition, the update formulas for variables W and H are as follows:

$$W_{il} = W_{il} \frac{(XH^T)_{il}}{(WHH^T)_{il}},$$
$$H_{lj} = H_{lj} \frac{(W^T X)_{lj}}{(W^T WH)_{lj}}. \tag{2}$$

2.2. Deep Nonnegative Matrix Factorization. The traditional NMF method can remove redundant information and reveal the hidden semantic features of multimedia data, but it cannot learn an effective feature representation for the data. For example, a facial image contains various changes such as posture, lighting, and expression changes. Therefore, Trigeorgis et al. [22] pointed out that the coefficient matrix, as a low-dimensional representation of high-dimensional data, should be able to continue to be decomposed so that more abstract low-dimensional features can be obtained. Thus, these processes of deep factorization are defined as

$$X \approx W_1 H_1,$$
$$H_1 \approx W_2 H_2,$$
$$\cdots\cdots$$
$$H_{l-1} \approx W_l H_l, \tag{3}$$

where W_i and H_i represent the factorization results of the i-th layer. It can be seen from Eq. (3) that deep NMF performs a procedure of matrix factorization at each layer and uses the decomposed coefficient matrix as the input data of the next layer to continue decomposing. Consequently, the process of deep matrix factorization performed on the data is expressed as

$$X \approx W_1 W_2 \cdots W_l H_l. \tag{4}$$

The objective function of deep NMF is defined as follows:

$$\min \|X - W_1 W_2 \cdots W_l H_l\|_F^2. \tag{5}$$

Similar to that of NMF, the update formula can be defined as follows:

$$W_i = (\Psi^T \Psi)^{-1} \Psi^T X \tilde{H}_i^T (\tilde{H}_i \tilde{H}_i^T)^{-1},$$
$$H_i = H_i \odot \sqrt{\frac{[\Psi^T X]^{\text{pos}} + [\Psi^T \Psi]^{\text{neg}} H_i}{[\Psi^T X]^{\text{neg}} + [\Psi^T \Psi]^{\text{pos}} H_i}}, \tag{6}$$

where $\Psi = W_1 \cdots W_{i-1}$, $\tilde{H}_i$ denotes the reconstruction of the i-th layer's feature matrix, and the symbol $\odot$ represents the dot product of matrices. $[A]^{\text{pos}} = |A| + A/2$ represents a matrix operation that restrains all the negative elements to zeros and keeps the positive elements unchanged. On the contrary, $[A]^{\text{neg}} = |A| - A/2$ turns the positive elements to be zeros while the negative elements are to be nonnegative.

3. Adaptive-Weighted Multiview Deep Basis Matrix Factorization

First, an adaptive-weighted multiview deep basis matrix factorization (AMDBMF) method is proposed, which incorporates the nonnegative matrix factorization and deep learning into a unified framework. Next, an optimization algorithm with an iterative updating rule is designed to solve the objective function of AMDBMF. Then, an adaptive-weighted fusion mechanism is provided. Finally, we provide the complexity analysis of the proposed algorithm.

Suppose that $X = [X_1, \cdots, X_N]$ denotes a multimedia data set which contains N samples. Each sample $x_i (i = 1, \cdots, N)$ is described by M views. Thus, the m-th view's features for this sample can be represented as $x_i^m (m = 1, \cdots, M)$. The features of all samples in this view can be represented as $X^m = [x_1^m, \cdots, x_N^m]$.

3.1. Objective Function.
First, matrix factorization is performed on the features in each view of the multimedia data, and the objective function can be defined as

$$\min_{W^m, H^m} \|X^m - W^m H^m\|_F^2$$
$$\text{s.t.} W^m \geq 0, H^m \geq 0, \tag{7}$$

where W^m and H^m denote the basis matrix and the coefficient matrix of the m-th view's features, respectively.

Then, the deep factorization is performed on W^m. The process is defined as follows:

$$X^m \approx W_1^m H_1^m,$$
$$W_1^m \approx W_2^m H_2^m,$$
$$\cdots \cdots \tag{8}$$
$$W_{l-2}^m \approx W_{l-1}^m H_{l-1}^m,$$
$$W_{l-1}^m \approx W_l^m H_l^m,$$

where $W_1^m, W_2^m, \cdots, W_{l-1}^m, W_l^m$ and $H_1^m, H_2^m, \cdots, H_{l-1}^m, H_l^m$ denote the basis matrices and coefficient matrices for each layer in the m-th view, respectively.

By combining it with Eq. (8), Eq. (7) is further rewritten as follows:

$$\min_{W_l^m, H_l^m} \|X^m - W_l^m H_l^m H_{l-1}^m \cdots H_2^m H_1^m\|_F^2$$
$$\text{s.t.} W_l^m \geq 0, H_l^m \geq 0. \tag{9}$$

Finally, to fuse the data from multiple perspectives, the final objective function is defined as

$$\min_{W_l^m, H_l^m} \sum_{m=1}^{M} \|X^m - W_l^m H_l^m H_{l-1}^m \cdots H_2^m H_1^m\|_F^2$$
$$\text{s.t.} W_l^m \geq 0, H_l^m \geq 0. \tag{10}$$

3.2. Optimization.
From Eq. (10), we can find that the objective function is nonconvex for all variables, but it is convex for each of them on their own. Therefore, we design an iterative update algorithm to find the local optimal solution of the objective function. To solve this problem, one variable is updated while the other variables are fixed. The detailed updating rules are described as follows.

The optimal objective function for variables W_l^m and H_l^m can be defined as

$$\min_{W_l^m, H_l^m} \|X^m - W_l^m H_l^m H_{l-1}^m \cdots H_2^m H_1^m\|_F^2$$
$$\text{s.t.} W_l^m \geq 0, H_l^m \geq 0. \tag{11}$$

Let $\Lambda_l^m = H_{l-1}^m \cdots H_2^m H_1^m$, and then Eq. (11) can be simplified as

$$\min_{W_l^m, H_l^m} \|X^m - W_l^m H_l^m \Lambda_{l-1}^m\|_F^2$$
$$= tr\left((X^m - W_l^m H_l^m \Lambda_{l-1}^m)^T (X^m - W_l^m H_l^m \Lambda_{l-1}^m) \right)$$
$$= tr\left(-2X^{mT} W_l^m H_l^m \Lambda_{l-1}^m + \Lambda_{l-1}^m{}^T H_l^m{}^T W_l^m{}^T W_l^m{}_i H_l^m \Lambda_{l-1}^m \right)$$
$$\text{s.t.} W_l^m \geq 0, H_l^m \geq 0. \tag{12}$$

The Lagrangian function of Eq. (12) is expressed as

$$\varphi(W_l^m, H_l^m) = \|X^m - W_l^m H_l^m \Lambda_{l-1}^m\|_F^2 + tr(\gamma_l^m H_l^m) + tr(\eta_l^m W_l^m), \tag{13}$$

where γ_l^m and η_l^m are Lagrange multipliers.

Taking the partial derivatives of Eq. (13) with respect to W_l^m and H_l^m, and setting these derivatives to zero, we have

$$\frac{\partial \varphi(W_l^m, H_l^m)}{\partial W_l^m} = \frac{\partial \|X^m - W_l^m H_l^m \Lambda_{l-1}^m\|_F^2}{\partial W_l^m} + \frac{\partial tr(\eta_l^m W_l^m)}{\partial W_l^m} = 0,$$
$$\frac{\partial \varphi(W_l^m, H_l^m)}{\partial H_l^m} = \frac{\partial \|X^m - W_l^m H_l^m \Lambda_{l-1}^m\|_F^2}{\partial H_l^m} + \frac{\partial tr(\gamma_l^m H_l^m)}{\partial H_l^m} = 0. \tag{14}$$

According to the KKT condition $(\gamma_l^m H_l^m)_{ij} = 0$ and $(\eta_l^m W_l^m)_{ij} = 0$ [41], the update rules of variables W_l^m and H_l^m are as follows:

$$W_l^m = W_l^m \odot \frac{X \Lambda_{l-1}^m{}^T H_l^m{}^T}{W_l^m H_l^m \Lambda_{l-1}^m \Lambda_{l-1}^m{}^T H_l^m{}^T},$$
$$H_l^m = H_l^m \odot \frac{W_l^m{}^T X^m \Lambda_{l-1}^m{}^T}{W_l^m{}^T W_l^m H_l^m \Lambda_{l-1}^m \Lambda_{l-1}^m{}^T}, \tag{15}$$

where the symbol $\odot$ represents the dot product of matrices.

Finally, the algorithmic steps of the proposed method are given in Algorithm 1. To make it easier to understand,

Input:
Multiview nonnegative matrix $X = [X^1, X^2, \cdots, X^m, \cdots, X^M]$
The hidden features of each layer $\{d_1, d_2, \cdots d_l, \cdots, d_L\}$
Initialize:
The basis matrix for each view $\{W_l^m\}, l = 1, 2, \cdots, L, m = 1, 2, \cdots, M$
The coefficient matrix for each view $\{H_l^m\}, l = 1, 2, \cdots, L, m = 1, 2, \cdots, M$
Pretrain:
for $m = 1 : M$ do
  for $l = 1 : L$ do
$W_l^m, H_l^m = NMF(W_{l-1}^m, k_l)$
  end for
  end for
Update:
Repeat
  for $m = 1 : M$ do
  for $l = 1 : L$ do
  if $l == 1$
$\Lambda_l^m = I$
  else
$\Lambda_l^m = H_{l-1}^m \cdots H_2^m H_1^m$
  end
    Update $W_l^m = W_l^m \odot X \Lambda_{l-1}^{m\,T} H_l^{m\,T} / W_l^m H_l^m \Lambda_{l-1}^m \Lambda_{l-1}^{m\,T} H_l^{m\,T}$
    Update $H_l^m = H_l^m \odot W_l^{m\,T} X^m \Lambda_{l-1}^{m\,T} / W_l^{m\,T} W_l^m H_l^m \Lambda_{l-1}^m \Lambda_{l-1}^{m\,T}$
  end for
  end for
Until Reach the convergence condition or the maximum number of iterations
Output:
The basis matrix for each view $\{W_l^m\}, l = 1, 2, \cdots, L, m = 1, 2, \cdots, M$
The coefficient matrix for each view $\{H_l^m\}, l = 1, 2, \cdots, L, m = 1, 2, \cdots, M$

ALGORITHM 1: The optimization algorithm.

Figure 2 depicts the block diagram of the proposed optimization algorithm.

3.3. Feature Confusion. After obtaining the basis matrix and coefficient matrix of each layer for each view through the optimization algorithm, an adaptive-weighted fusion mechanism is adopted to obtain a low-dimensional representation of the multiview data, and the weight calculation is

$$\alpha^m = \frac{1}{2\sqrt{\|X^m - W_l^m H_l^m H_{l-1}^m \cdots H_2^m H_1^m\|_F^2 + \varepsilon}}, \quad (16)$$

where ε is a small constant.

Then, α^m is normalized by Eq. (17)

$$\alpha^m = \frac{\alpha^m}{\sum_{m=1}^M \alpha^m}. \quad (17)$$

Finally, since the low-dimensional representation of each view is expressed as $H^m = H_l^m \Lambda_{l-1}^m$, the fusion of the low-dimensional features derived from the multiview data can be expressed as

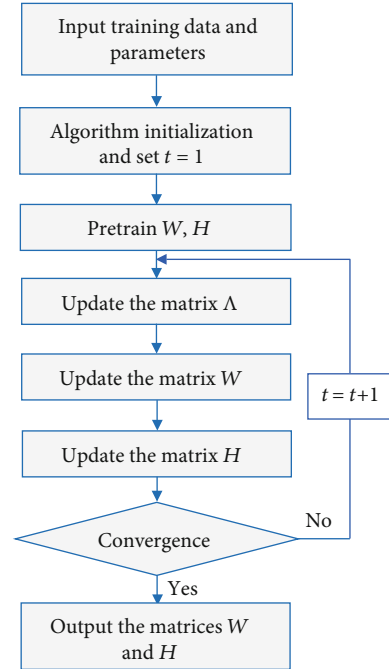$$H^* = \sum_{m=1}^M \alpha^m H^m. \quad (18)$$

FIGURE 2: The block diagram of the proposed optimization algorithm.

TABLE 1: Statistical information about the different datasets.

Datasets	Size (N)	Classes (C)	Views (V)	Feature dimension of each view				
				d_1	d_2	d_3	d_4	d_5
3sources	169	6	3	3560	3068	3631	—	—
BBC	685	5	4	4659	4633	4465	4684	—
BBC Sport	544	5	2	3183	3203	—	—	—
Reuters	1200	6	5	12000	12000	12000	12000	12000
Wikipedia	693	10	2	128	10	—	—	—

(a)

(b)

(c)
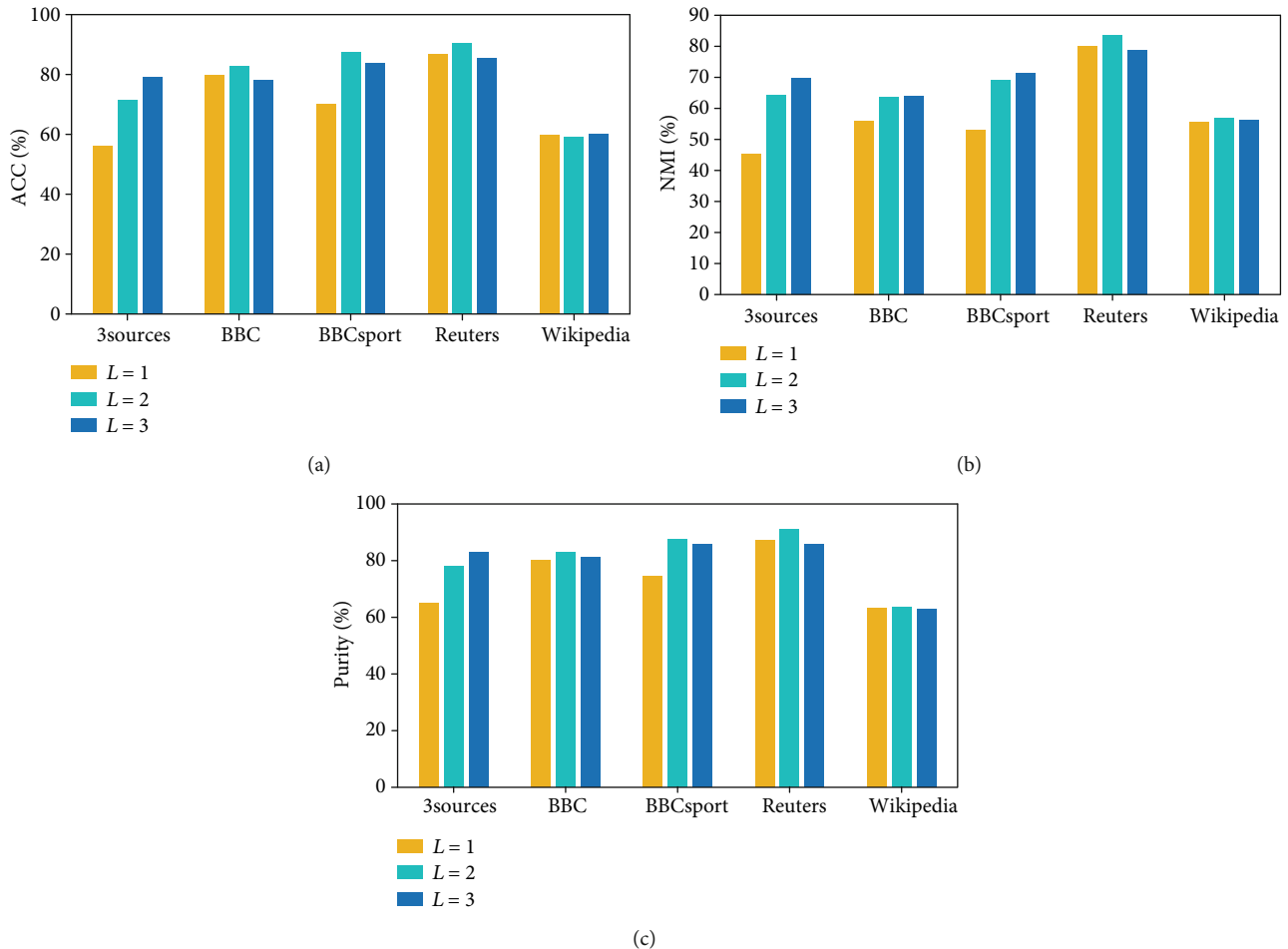
FIGURE 3: The influence of factorization layer L on the clustering results, (a) ACC, (b) NMI, and (c) purity.

3.4. Complexity Analysis. Clearly, the proposed algorithm can be divided into two stages: pretraining and fine-tuning. For convenience, suppose that the number of iterations is T, M is the number of data views, and L is the number of layers. The number of features for all views is D, and the number of low-dimensional representations for each layer is K. In the pretraining process, the complexity of a single view is O(TDNK). Therefore, the complexity of the whole pretraining process is O(TMLDNK). For the fine-tuning part, the main computational complexity is derived from updating Λ_{l-1}^m, W_l^m, and H_l^m, which requires $O(\text{TM}(L\text{-}1)\text{NK}^2)$, O(TMLDNK), and O(TMLDNK) complexity, respectively.

Since $D >> K$, the total computational complexity of the proposed algorithm is O(TMLDNK).

4. Experiments and Analysis

4.1. Datasets. Five commonly used multiview multimedia datasets from the Internet are used in the experiments to verify the effectiveness of the proposed method.

4.1.1. 3sources. This dataset includes a collection of 416 news events and 948 related news reports from February to April 2009 from three well-known news media outlets, including
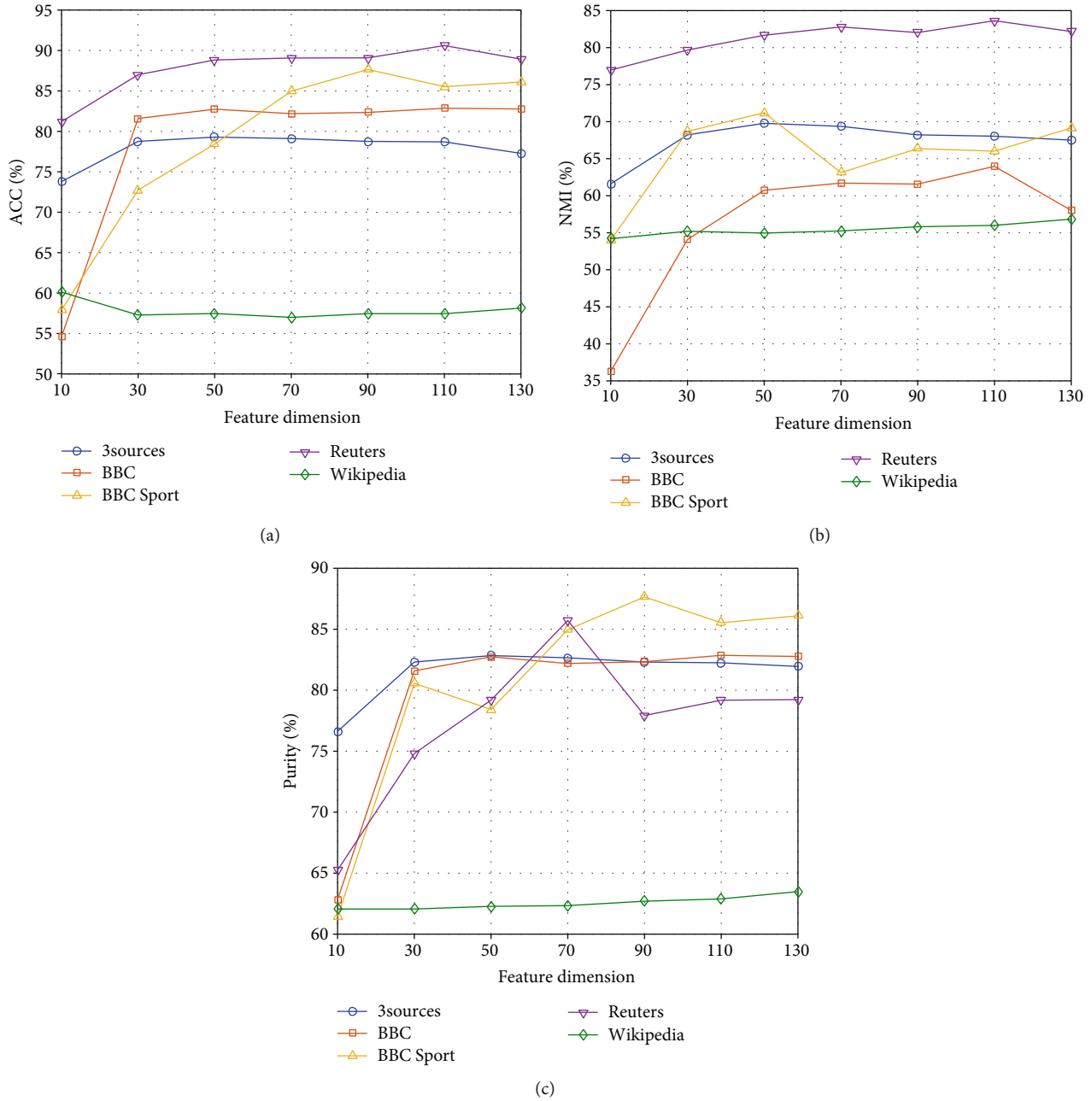
(a)

(b)

(c)

FIGURE 4: The influences of feature dimension D on clustering results, (a) ACC, (b) NMI, and (c) purity.

BBC, Reuters, and Guardian. In the experiments, 169 news items reported by all three news media outlets are used. These news events include six categories: business, entertainment, health, politics, sports, and technology (http://mlg.ucd .ie/datasets/3sources.html).

4.1.2. **BBC** *[42].* This dataset contains 685 news articles collected from the BBC News Network between 2004 and 2005. Each article is divided into four parts, and the data consist of five kinds of news: business, entertainment, politics, sport, and technology (http://mlg.ucd.ie/datasets/segment .html).

TABLE 2: The optimal parameter groups in our proposed algorithm.

Datasets	ACC	NMI	Purity
3sources	{3, 50}	{3, 50}	{3, 50}
BBC	{2, 110}	{3, 110}	{2, 110}
BBC Sport	{2, 90}	{3, 50}	{2, 90}
Reuters	{3, 110}	{2, 110}	{2, 70}
Wikipedia	{3, 10}	{2, 130}	{2, 130}

Table 3: The ACCs (%) of the AMDBMF and single-view methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
NMF	55.03	46.39	61.75	73.57	56.71
DBMF	76.98	63.84	80.24	67.67	57.11
AMDBMF	*79.29*	*82.86*	*87.65*	*90.61*	*60.12*

Table 4: The NMIs (%) of the AMDBMF and single-view methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
NMF	51.90	18.74	40.45	66.84	51.23
DBMF	67.16	56.71	66.80	59.47	54.89
AMDBMF	*69.78*	*63.97*	*71.21*	*86.63*	*56.84*

Table 5: The purities (%) of the AMDBMF and single-view methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
NMF	72.01	48.66	64.67	75.34	60.75
DBMF	81.72	73.74	83.62	69.98	59.60
AMDBMF	*82.84*	*82.86*	*87.65*	*91.13*	*63.49*

Table 6: The ACCs (%) of different multiview methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
MVCF	69.94	63.18	66.82	63.42	54.04
DeepMVC	65.68	72.12	72.12	76.25	60.46
GMC	69.23	69.34	80.70	66.58	44.88
NMFCC	69.82	73.40	83.09	76.04	*61.46*
AMDBMF	*79.29*	*82.86*	*87.65*	*90.61*	60.12

4.1.3. BBC Sport [42]. This dataset includes 737 news articles from the BBC Sport network from 2004 to 2005. These news articles cover six fields, such as track, field, cricket, football, rugby, and tennis (http://mlg.ucd.ie/datasets/segment.html).

4.1.4. Reuters [43]. This is a dataset that includes 1200 English articles from six types of samples, and each article has been translated into French, German, Italian, and Spanish (http://lig-membres.imag.fr/grimal/data.html).

4.1.5. Wikipedia [43]. This dataset consists of specific Wikipedia material with 2669 articles in 29 categories (http://www.svcl.ucsd.edu/projects/crossmodal/). In the experiments, we select a subset of the 10 most popular categories containing a total of 693 samples. The detailed statistical information about the different datasets is given in Table 1.

4.2. Metrics. In the experiments, we select three commonly used clustering evaluation indicators [44]: accuracy (ACC), normalized mutual information (NMI), and purity to evaluate the performance of the proposed method.

Table 7: The NMIs (%) of different multiview methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
MVCF	60.64	45.33	46.01	65.92	49.24
DeepMVC	56.50	52.58	52.58	73.76	55.45
GMC	54.80	48.52	*72.26*	70.43	36.14
NMFCC	61.91	53.95	67.38	73.92	56.00
AMDBMF	*69.78*	*63.97*	*71.21*	*83.63*	*56.84*

Table 8: The purities (%) of different multiview methods on different datasets.

Methods	3sources	BBC	BBC Sport	Reuters	Wikipedia
MVCF	78.52	64.64	68.36	68.84	57.17
DeepMVC	75.15	72.12	72.12	78.00	63.64
GMC	74.56	69.34	84.38	67.25	48.20
NMFCC	78.70	73.45	83.09	78.96	*64.86*
AMDBMF	*82.84*	*82.86*	*87.65*	*91.13*	63.49

Assuming that the clustering result of x_i is l_i and that the corresponding true label is t_i, then the clustering accuracy (ACC) [45] is defined as

$$\text{ACC} = \frac{\sum_{i=1}^{N} \delta(t_i, \text{map}(l_i))}{N}, \tag{19}$$

where the function $\delta(\cdot)$ is defined as follows:

$$\begin{cases} \delta(x, y) = 1 & x = y \\ \delta(x, y) = 0 & x \neq y. \end{cases} \tag{20}$$

The function $\text{map}(\cdot)$ maps the clustering result to the corresponding true label. The Kuhn-Munkres algorithm [46] is employed to find the best mapping result.

Assume that L and T are the clustering result and the true label set, respectively. The mutual information (MI) between them is defined as

$$\text{MI}(L, T) = \sum_{l_i \in L, t_i \in T} p(l_i, t_i) \cdot \log_2 \frac{p(l_i, t_i)}{p(l_i) \cdot p(t_i)}, \tag{21}$$

where $p(l_i)$ and $p(t_i)$ represent the probabilities that a sample is randomly selected from the dataset belonging to l_i and t_i, respectively. $p(l_i, t_i)$ represents the joint probability of a sample randomly being selected from the dataset belonging to l_i and t_i. Let $H(L)$ and $H(T)$ represent the entropies of L and T, respectively. Since the value range of mutual information is between 0 and $\max(H(L), H(T))$, the normalized mutual information (NMI) is defined as
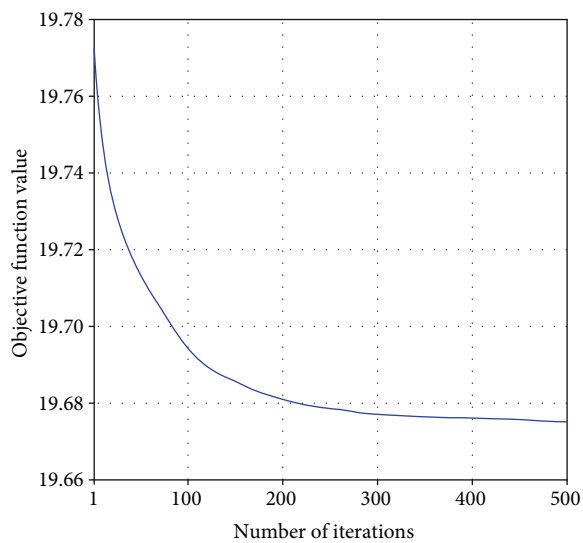
$$\text{NMI} = \frac{\text{MI}(L, T)}{\max(H(L), H(T))}. \tag{22}$$
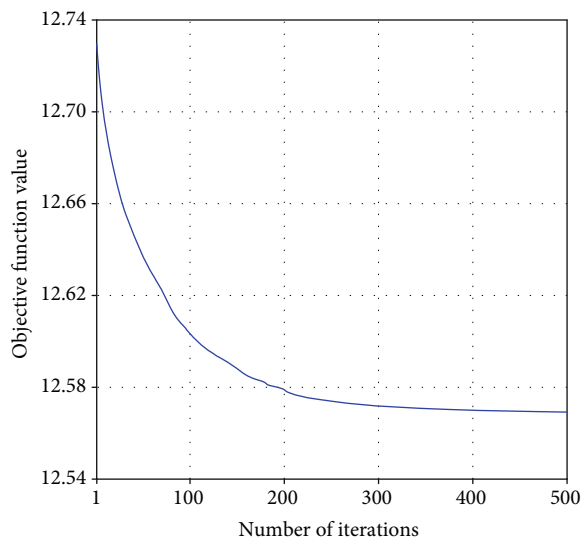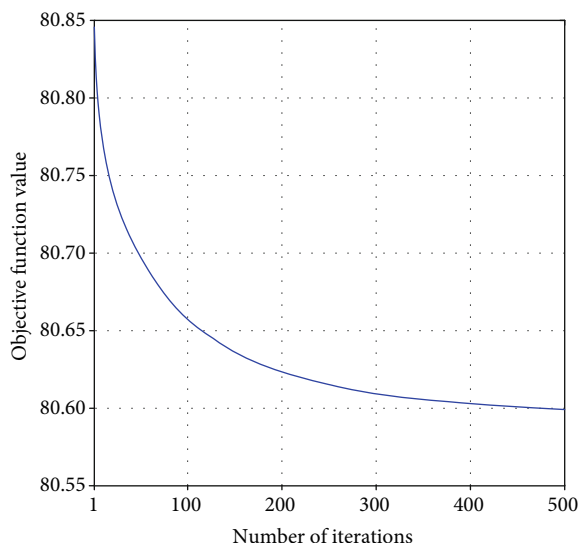
(a)
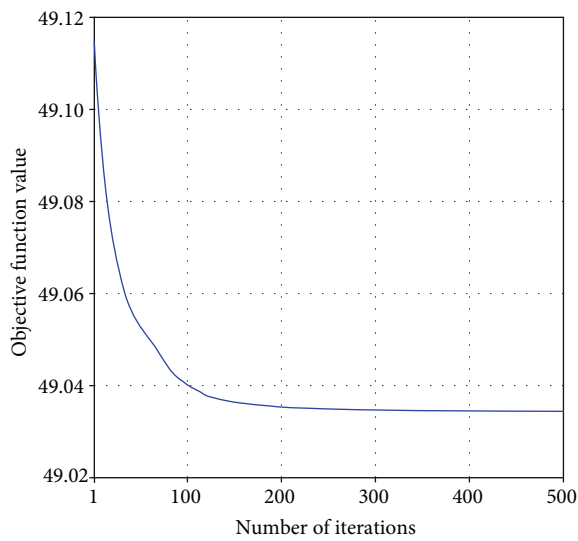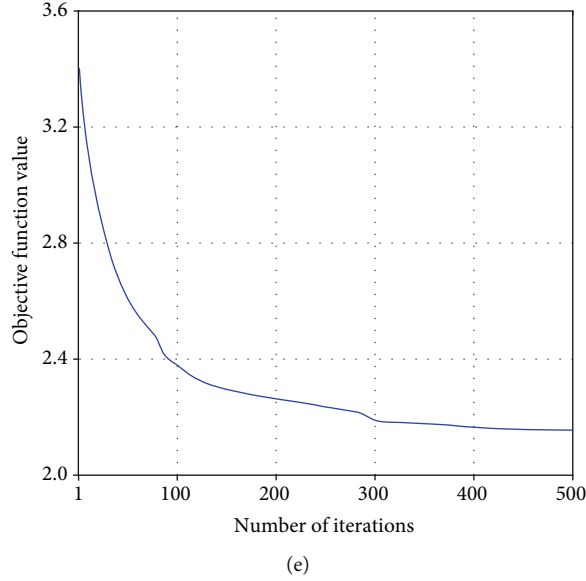
(b)

(c)

(d)

Figure 5: Continued.

(e)

FIGURE 5: The convergence curves of the proposed method on five datasets: (a) 3sources, (b) BBC, (c) BBC Sport, (d) Reuters, and (e) Wikipedia.

Purity is a straightforward and transparent evaluation method that is defined as follows:

$$\text{purity} = \frac{1}{k}\sum_{i=1}^{k}\frac{\left|C_i^d\right|}{\left|C_i\right|}, \tag{23}$$

where k represents the number of clusters, $\left|C_i^d\right|$ is the number of elements in the most numerous category in cluster C_i, and $\left|C_i\right|$ is the number of elements in cluster C_i.

4.3. Experimental Results and Analysis. In the first experiment, to test the influences of the parameters on the proposed method, we set the number of factorization layers L and the feature dimension D of each layer to $\{1, 2, 3\}$ and $\{10, 30, 50, 70, 90, 110, 130\}$, respectively. Furthermore, we adopt a grid search to find the optimal parameter value. In the experiment, the low-dimensional features obtained by the proposed algorithm are clustered by the K-means algorithm. Since the initialization of the K-means algorithm has an impact on the clustering results, we repeat the random initialization process with 10 times and report the mean value. First, the optimal feature dimension of each layer is fixed, and the numbers of layers are changed. As shown in Figure 3, in most cases, when the number of layers is set to 1, the result of each measure is poorer than the rest. However, as the number of layers increases, the performance also increases. It shows that the deep factorization helps to improve the performance of the proposed method.

Then, the numbers of layers are fixed, and the dimension of the feature is changed. The result is shown in Figure 4. It can be seen that as the dimensionality increases, the clustering performance also improves in most cases. However, this trend is not always maintained, and the clustering perfor-

mance decreases or remains stable as the dimensionality increases once the performance reaches the optimal level. The details of the optimal parameter groups in our proposed algorithm are listed in Table 2.

The second experiment is conducted to verify that the fusion of multiview information is beneficial for improving the clustering performance of the proposed method. First, we perform traditional NMF and deep basis matrix factorization (DBMF) for the data of each view. Then, we obtain the low-dimensional features of the multiview data by fusing the features of different views with equal weight. Finally, the proposed AMDBMF method is compared with the above two methods. The comparison results are listed in Tables 3–5. According to the tables, the performance of the DBMF method is better than that of the traditional NMF method, which indicates that more abstract features can be obtained through the deep factorization. The performance of the proposed AMDBMF method is better than that of the DNBMF method, which verifies that the adaptive fusion of different views is beneficial for extracting more robust low-dimensional features from multiview data.

The third experiment compares the performance of the proposed AMDBMF method with those of some currently popular multiview algorithms, including MVCF [37], DeepMVC [41], GMC [47], and NMFCC [36]. MVCF utilized the correlation information between the views obtained by jointly optimizing the graph matrix of the data of each view. DeepMVC used a nonparameterized adaptive learning method to obtain the weights between views. NMFCC introduces orthogonal constraints into the basis matrix and coefficient matrix. The best results yielded by the different multiview learning methods on different datasets are shown in Tables 6–8. It can be seen that the performances of the proposed method are significantly better than those of the other comparison methods in most cases. Since these

methods use different mechanisms to fuse multiview data information, all the methods present different performances on different databases. Therefore, how to effectively integrate fusion mechanisms is still an open problem.

The final experiment verifies the convergence of the proposed optimization algorithm. The convergence curves of the proposed method on different datasets are given in Figure 5. As seen from the figures, the iterative update rules in Algorithm 1 decrease the objective function value obtained by our proposed method. Moreover, we can also see that our proposed method converges very quickly on these datasets.

5. Conclusions and Future Work

To efficiently learn the feature representations of multiview multimedia data, this paper proposes a new deep nonnegative matrix factorization method with multiview learning. Unlike traditional methods, the proposed method deeply decomposes the basis matrix, so it not only can learn the component representation of the original data but also can learn more abstract deep features. Furthermore, to effectively fuse the available multiview data information, this paper introduces an adaptive feature fusion mechanism.

To solve the shortcoming of information fusion for multiview data, a large number of fusion mechanisms have been proposed, and they achieve different performances on different datasets. Therefore, how to effectively integrate different mechanisms to improve the feature representation ability of a given approach is one of the key research tasks to be addressed in the future. Moreover, we will apply our method to other fields such as medical image procession and medical text analysis [48].

Data Availability

The data are derived from public domain resources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] Y. Yi, Y. Chen, J. Wang, G. Lei, J. Dai, and H. Zhang, "Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization," *Signal Processing: Image Communication*, vol. 89, article 115984, 2020.

[2] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Nonnegative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.

[3] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna et al., "Analysis of dimensionality reduction techniques on big data," *IEEE Access*, vol. 8, pp. 54776–54788, 2020.

[4] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.

[5] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data," *Information Fusion*, vol. 59, pp. 44–58, 2020.

[6] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[7] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[8] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Transactions on Communications*, vol. 28, no. 1, pp. 84–95, 1980.

[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[10] Y. X. Wang and Y. J. Zhang, "Nonnegative matrix factorization: a comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.

[11] A. A. Jamali, A. Kusalik, and F. X. Wu, "MDIPA: a microRNA–drug interaction prediction approach based on nonnegative matrix factorization," *Bioinformatics*, vol. 36, no. 20, pp. 5061–5067, 2020.

[12] P. Chalise, Y. Ni, and B. L. Fridley, "Network-based integrative clustering of multiple types of genomic data using nonnegative matrix factorization," *Computers in Biology and Medicine*, vol. 118, p. 103625, 2020.

[13] M. Hou, J. Li, and G. Lu, "A supervised non-negative matrix factorization model for speech emotion recognition," *Speech Communication*, vol. 124, pp. 13–20, 2020.

[14] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, no. 9, pp. 1457–1469, 2004.

[15] S. Choi, "Algorithms for orthogonal nonnegative matrix factorization," in *2008 IEEE International Joint Conference on Neural Networks*, pp. 1828–1832, Hong Kong, China, June 2008.

[16] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2008.

[17] J. Fan and J. Cheng, "Matrix completion by deep matrix factorization," *Neural Networks*, vol. 98, pp. 34–41, 2018.

[18] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on*

Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1798–1828, 2013.

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[20] G. Zhong, L. N. Wang, X. Ling, and J. Dong, "An overview on data representation learning: from traditional feature learning to recent deep learning," *The Journal of Finance and Data Science*, vol. 2, no. 4, pp. 265–278, 2016.

[21] J. H. Ahn, S. Kim, J. H. Oh, and S. Choi, "Multiple nonnegative-matrix factorization of dynamic PET images," in *Proceedings of Asian Conference on Computer Vision*, pp. 1009–1013, Jeju, Korea, 2004.

[22] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 417–429, 2016.

[23] Y. Zhao, H. Wang, and J. Pei, "Deep non-negative matrix factorization architecture based on underlying basis images learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 1897–1913, 2021.

[24] Y. Meng, R. Shang, F. Shang, L. Jiao, S. Yang, and R. Stolkin, "Semi-supervised graph regularized deep NMF with bi-orthogonal constraints for data representation," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 9, pp. 3245–3258, 2020.

[25] M. Tong, Y. Chen, L. Ma, H. Bai, and X. Yue, "NMF with local constraint and deep NMF with temporal dependencies constraint for action recognition," *Neural Computing and Applications*, vol. 32, no. 9, pp. 4481–4505, 2020.

[26] J. Li, G. Zhou, Y. Qiu, Y. Wang, Y. Zhang, and S. Xie, "Deep graph regularized non-negative matrix factorization for multi-view clustering," *Neurocomputing*, vol. 390, pp. 108–116, 2020.

[27] Z. Shu, X. Wu, C. Hu, C. Z. You, and H. H. Fan, "Deep semi-nonnegative matrix factorization with elastic preserving for data representation," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 1707–1724, 2021.

[28] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: recent progress and new challenges," *Information Fusion*, vol. 38, pp. 43–54, 2017.

[29] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 10, pp. 1863–1883, 2018.

[30] T. Hussain, K. Muhammad, W. Ding, J. Lloret, S. W. Baik, and V. H. C. de Albuquerque, "A comprehensive survey of multi-view video summarization," *Pattern Recognition*, vol. 109, p. 107567, 2021.

[31] X. Xu, Y. Yang, C. Deng, and F. Nie, "Adaptive graph weighting for multi-view dimensionality reduction," *Signal Processing*, vol. 165, pp. 186–196, 2019.

[32] R. Zhang, F. Nie, X. Li, and X. Wei, "Feature selection with multi-view data: a survey," *Information Fusion*, vol. 50, pp. 158–167, 2019.

[33] P. Luo, J. Peng, Z. Guan, and J. Fan, "Dual regularized multi-view non-negative matrix factorization for clustering," *Neurocomputing*, vol. 294, pp. 1–11, 2018.

[34] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*, pp. 252–260, Austin, TX, USA, 2013.

[35] W. Y. Chang, C. P. Wei, and Y. C. F. Wang, "Multi-view non-negative matrix factorization for clothing image characteriza-

[36] N. Liang, Z. Yang, Z. Li, W. Sun, and S. Xie, "Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints," *Knowledge-Based Systems*, vol. 194, p. 105582, 2020.

[37] K. Zhan, J. Shi, J. Wang, H. Wang, and Y. Xie, "Adaptive structure concept factorization for multiview clustering," *Neural Computation*, vol. 30, no. 4, pp. 1080–1103, 2018.

[38] S. Wei, J. Wang, G. Yu, C. Domeniconi, and X. Zhang, "Multi-view multiple clusterings using deep matrix factorization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 4, pp. 6348–6355, 2020.

[39] W. Zhao, C. Xu, Z. Guan, and Y. Liu, "Multiview concept learning via deep matrix factorization," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 814–825, 2021.

[40] H. Zhao, Z. Ding, and Y. Fu, "Multi-view clustering via deep matrix factorization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[41] S. Huang, Z. Kang, and Z. Xu, "Auto-weighted multi-view clustering via deep matrix decomposition," *Pattern Recognition*, vol. 97, article 107015, 2020.

[42] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proceedings of the 23rd international conference on Machine learning*, pp. 377–384, New York, NY, USA, 2006.

[43] G. Bisson and C. Grimal, "An architecture to efficiently learn co-similarities from multi-view datasets," in *Neural Information Processing. ICONIP 2012. Lecture Notes in Computer Science, vol 7663*, T. Huang, Z. Zeng, C. Li, and C. S. Leung, Eds., pp. 184–193, Springer, Berlin, Heidelberg, 2012.

[44] L. Fu, P. Lin, A. V. Vasilakos, and S. Wang, "An overview of recent multi-view clustering," *Neurocomputing*, vol. 402, pp. 148–161, 2020.

[45] Y. Yi, Y. Shi, H. Zhang, J. Wang, and J. Kong, "Label propagation based semi-supervised non-negative matrix factorization for feature extraction," *Neurocomputing*, vol. 149, pp. 1021–1037, 2015.

[46] H. Zhu and M. C. Zhou, "Efficient role transfer based on Kuhn–Munkres algorithm," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 2, pp. 491–496, 2011.

[47] H. Wang, Y. Yang, and B. Liu, "GMC: graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1116–1129, 2019.

[48] X. Huang, B. Zhong, Y. Cao, Y. Yi, and M. Gu, "Chest X-ray lung Chinese description generation based on semantic labels and hierarchical LSTM," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1020–1023, Seoul, Korea (South), December 2020.

tion," in *2014 22nd International Conference on Pattern Recognition*, pp. 1272–1277, Stockholm, Sweden, August 2014.

WILEY | Hindawi

Research Article

CDCN: A New NMF-Based Community Detection Method with Community Structures and Node Attributes

Zhiwen Ye,[1] Hui Zhang,[1,2] Libo Feng [ID],[3,4] and Zhangming Shan[1]

[1]*State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China*
[2]*Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China*
[3]*Engineering Research Center of Cyberspace, Yunnan University, Kunming 650500, China*
[4]*School of Software, Yunnan University, Kunming 650500, China*

Correspondence should be addressed to Libo Feng; fenglibo@buaa.edu.cn

Community discovery can discover the community structure in a network, and it provides consumers with personalized services and information pushing. It plays an important role in promoting the intelligence of the network society. Most community networks have a community structure whose vertices are gathered into groups which is significant for network data mining and identification. Existing community detection methods explore the original network topology, but they do not make the full use of the inherent semantic information on nodes, e.g., node attributes. To solve the problem, we explore networks by considering both the original network topology and inherent community structures. In this paper, we propose a novel nonnegative matrix factorization (NMF) model that is divided into two parts, the community structure matrix and the node attribute matrix, and we present a matrix updating method to deal with the nonnegative matrix factorization optimization problem. NMF can achieve large-scale multidimensional data reduction processing to discover the internal relationships between networks and find the degree of network association. The community structure matrix that we proposed provides more information about the network structure by considering the relationships between nodes that connect directly or share similar neighboring nodes. The use of node attributes provides a semantic interpretation for the community structure. We conduct experiments on attributed graph datasets with overlapping and nonoverlapping communities. The results of the experiments show that the performances of the F1-Score and Jaccard-Similarity in the overlapping community and the performances of normalized mutual information (NMI) and accuracy (AC) in the nonoverlapping community are significantly improved. Our proposed model achieves significant improvements in terms of its accuracy and relevance compared with the state-of-the-art approaches.

1. Introduction

The science of networks is a modern discipline spanning the natural, social, and computer sciences, as well as engineering. There are different kinds of networks in the real world, such as citation networks, social networks, and collaboration networks [1]. Community detection algorithms are important methods for analyzing the networks' structure and understanding the node semantics, which play important roles in the era of network intelligence [2–5]. First, analyzing the community structure of the network is helpful for people to study the composition and evolution of the whole network, and it can better explain the intrinsic characteristics and

causes of the network. Furthermore, community detection algorithms are the key to understanding complex network systems and have important applications in different networks in various fields. For example, they are very useful for social networks to recommend friends and groups to users by analyzing the inherent structure and characteristics of their social network and clustering the user nodes. Community detection is very useful in the application of disease spread analysis, e.g., it can be used in reality epidemic spreading [6].

Most existing community detection algorithms analyze the network by using the original network topology information [7–15]. Girvan and Newman [7] analyzed the

community structures in social and biological networks. Newman and Leicht [8] analyzed the mixture models of networks. Rosvall and Bergstrom [9] revealed the community structure of complex networks utilizing the maps of random walks. Xie et al. [1] proposed a method for uncovering overlapping communities in social networks named SLPA via a speaker-listener interaction dynamic process. Coscia et al. [12] proposed a local first discovery method for overlapping communities. He [13] utilized the Markov random field approach for community detection in a specific network. Clauset et al. [14] and Li et al. [15] analyzed the community structure in large scale networks. Cui and Wang [16] used the key bicommunity and intimate degree uncover the overlapping community structures in bipartite networks.

However, for some networks, there are not only network topology information but also node attribute information that is a semantic interpretation of the community structure. For example, papers in citation networks contain titles, abstracts, and keywords that may be represented using binary-valued vectors. We binarize the categorical input so that they can be thought of as vectors in Euclidean space (we call this embedding the vector in Euclidean space).

Such networks with node attributes are named attributed graphs [17, 18]. It is a great challenge to discover the community structure with node attributes in an effective way. To characterize a community, the existing community detection methods mainly rely on the original network topology. The missing and meaningless information in the network topology often leads to poor results. The node attributes of a network may carry essential community information that is complementary to the network topology information. Therefore, even though two nodes are not directly connected, they may belong to the same community according to the node attributes. Several algorithms that consider both structural and attribute information have been proposed in [17, 19–22]. Yang and Leskovec [17] used the nonnegative matrix factorization approach to find the overlapping communities in large scale networks. Atzmueller et al. [19] proposed an exhaustive subgroup discovery method for description-oriented community detection. Wang et al. [20] proposed a semantic community identification method to find the community structures in large attribute networks. Huang et al. [21] analyzed the attributes of community networks. Yang et al. [22] proposed a discriminative approach that combined links and content for community detection. However, all of those methods assign each edge of the attribute graphs the same value. This will lose information about the network. For example, edges that form densely connected subgraphs are much more likely to be in the same community than edges that connect separate subgraphs. Thus, utilizing the original network topology directly causes indiscriminate penalizing of node pairs, whether they are in densely connected structures or not. It means that we should assign the different characteristic nodes various values.

Nonnegative matrix factorization (NMF) is an effective method in community detection. Some scholars have studied it. Luo et al. [23] proposed a symmetric NMF method via pointwise mutual information-incorporated that has highly accurate. Lu et al. [24] used the NMF method to improve density peak clustering in community detection. Zhang and Zhou [25] studied the structure of deep NMF in community detection. Wang et al. [26] used the constraint NMF to detect the community in dynamic networks. These studies laid the foundation for our experiments.

The state-of-the-art methods (e.g., CDE [27]) use both a community structure matrix and a node attribute matrix in the NMF framework, and CDE also considers the densely connected subgraphs. However, CDE only considers the relationship between two nodes directly connected while they analyze the community structure matrix. This behavior will lose information about the community structure.

Scholars have proposed some important methods for large-scale community detection such as neighborhood, maximal subgraph, intimate degree, and core-vertices [28]; these studies provide important ideas for our paper. More importantly, the main contributions of this paper are as follows.

(1) We propose a novel method that generates the community structure matrix, which retains the relationship between two nodes that are directly connected or share the same neighbors

(2) We combine node attribute information and community structure information in an effective way. Then, we propose our method, named Community Detection with Community Structure and Node Attributes (CDCN), to identify the network communities with semantic annotation and community structures using nonnegative matrix factorization framework [29, 30]

(3) Extensive experiments were conducted on public datasets to demonstrate the effectiveness of CDCN, and its accuracy and performance were better than those of the state-of-the-art methods

The remainder of the paper is organized as follows. Section 2 briefly summarizes the three different types of community detection models. Section 3 describes the community detection model, deduces its theory and formula, and describes the solution algorithm for our model. Section 4 conducts extensive comparative experiments to evaluate the effectiveness of our proposed CDCN model on real graph datasets with the ground-truth communities delineated. Section 5 presents the conclusions of this paper and discusses future research directions.

2. Related Works

This section will briefly summarize the three different types of community detection models that use different information to determine the network information. We briefly summarize the three types of community detection models in Table 1.

The first type of community detection method focuses on the original network topology. GN [7] was built around the idea of using centrality indices to find community boundaries. NMM [8] was used to find the mechanism of

TABLE 1: A brief summarization of community detection.

Types	Representative works
Original network topology	GN [7], NMM [8], InfoMap [9], CPM [10], SLPA [11], DEMON [12], SNMF [31], NetMRF [13], SBM [32], and SPAEM [33]
Node attributes	CAN [34], SMR [35], and NC [36]
Both original network topology and node attributes	COMODO [19], PCL-DC [22], SCI [20], BIGCLAM [17], and CDE [27]

probabilistic mixture models and the expectation maximization algorithm to understand the structure of networks. CPM [10] is a clique percolation method, and it consists of two steps. The first step is to construct the vertices of the k-clique graph, and the second is to find the connected components and set-union vertices within each connected component to get a new community. SLPA [11] was presented as general framework for detecting and analyzing both individual overlapping nodes and entire communities. In it, the nodes exchange labels according to dynamic interaction rules. The stochastic block model (SBM [32]) is the simplest node-based community detection model. The nodes of the network randomly fall into K communities, which are denoted as $z_i \in \{1, 2, \cdots, k\}$, and the edges are independently generated at a probability $w_{z_i z_j}$. McDaid et al. [37] improved Bayesian inference for the stochastic block model for large networks.

However, the above community detection methods directly utilized the original network topology and ignore the inherent community structures (e.g., node attributes). The missing and meaningless information in the network topology often leads to poor results. Therefore, the second type of method focuses on node attributes, and it includes some classical or state-of-the-art clustering methods. Strictly speaking, those methods are not community methods, but they could use node attributes information to discover communities. Thus, in this paper, we also regard them as related work. CAN [34] was proposed as a clustering model to learn the data similarity matrix by assigning the adaptive and optimal neighbors for each data point based on the local distances. SMR [35] uses for the kernelized random walks on the global KNN graph and the Smooth Representation Clustering to improve the clustering result. NC [5, 36] uses the eigenvectors of the matrix representations of the network to solve the community detection problem.

The third type of community detection method considers both the original network topology and node attribute information. Several algorithms that consider both structural and attribute information have been proposed in [17, 19–22]. However, all of those methods assign each edge of the attributed graph the same value. This will lose information about the network. The state-of-the-art method SCI [20] uses both the community structure matrix and the node attribute matrix in NMF framework, and CDE [27] encodes the inherent community structures for community detection via the underlying community memberships. However, they only considered the relationship between two directly connected nodes when they analyze the community structure matrix.

Nonnegative matrix factorization (NMF) [38] is an effective means of data dimensionality reduction. It can discover the hidden information and the relationship between multi-

dimensional data and lay the foundation for data mining and knowledge discovery. At present, NMF has been widely used in data mining [39], image retrieval [40], community discovery [41], hotspot prediction [42], social network privacy protection [43], signal processing [44], and other fields. In terms of community discovery [45, 46], NMF can find the associations between networks based on network node attributes, which is an important community detection method [47]. Many scholars have studied the application of NMF in community detection and provided some ideas for the research of our paper.

Different from all those methods, we combine node attribute information and community structure information by generating the community structure matrix, which retains the relationship between two directly connected nodes or nodes that share the same neighbors. This method can more accurately find the relationships between networks. The topology diagram of the three methods can be seen in Figure 1.

Figure 1 shows an unweighted graph with two communities, where the different shapes stand for the nodes of different communities. Figure 1(a) is the result of using the adjacency matrix directly, and Figure 1(b) is the community structure embedding matrix of the CDE model. They both focus on the relationship between two directly connected nodes. Figure 1(c) shows the community structure matrix of our models, where the dotted lines are the correlations of the nodes that share the same neighbors but are not directly connected. In other words, compared with existing models, we can not only express the node relationship using a continuous numerical value, but we can also describe the relationship between nonadjacent nodes. Otherwise, the isolated nodes (cold nodes) have no neighbor nodes. It increases the scale of the system and can be ignored when constructing the adjacent matrix. In this paper, the influence of the isolated node on the system is not considered.

3. CDCN: The Community Detection Model

In this section, we propose a novel algorithm for community detection that combines the community structure and node attribute information. We will introduce the community structure, node attributes, the overall model, and the algorithm in detail.

3.1. Community Structure Part. The community structure part models the network structure. Given an undirected network $G = (P, E)$ with n nodes P and e edges E, we could get a binary-valued adjacency matrix A from G. If node i and node

(a) Using adjacency matrix

(b) Embedding matrix of CDE
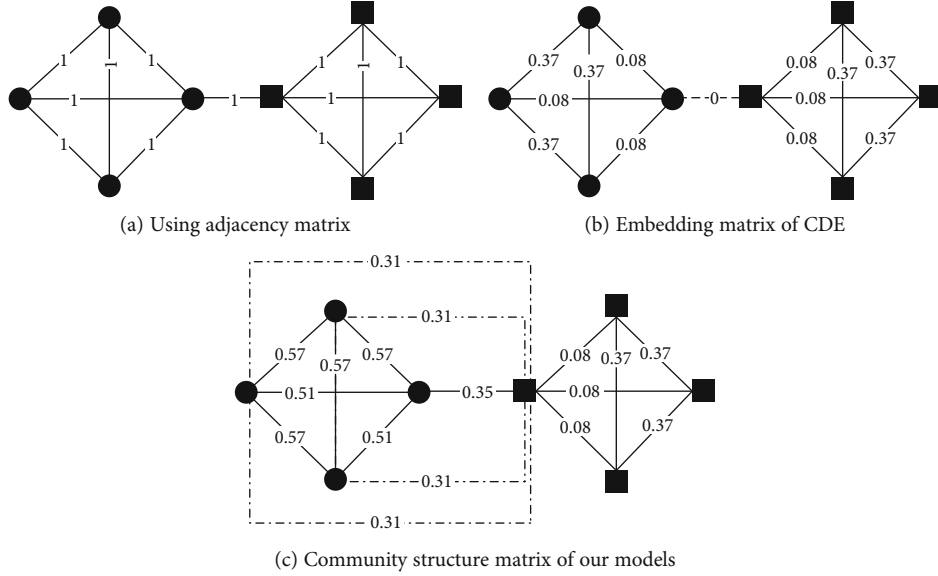

(c) Community structure matrix of our models

Figure 1: The community structure matrix of different methods.

j have a direct connection, the value of A_{ij} is 1, and otherwise, it is 0, where $i = (1, 2, \cdots, n)$ and $j = (1, 2, \cdots, n)$.

As one of the state-of-the-art methods, SCI directly regards the binary-valued adjacency matrix A as community structure matrix. It will degrade the effectiveness of the community detection model without embedding the adjacency matrix due to the sparsity of A. The CDE model is proposed as a novel community structure embedding method to quantify the structural closeness of nodes to offer a good depiction of the inherent community structures in graphs. Though CDE solved the problem of the sparsity of the adjacency matrix, it still has limitations in that CDE only considers the relationship between two directly connected nodes when it analyze the network structure. Both SCI and CDE will lose the relationship information between nodes that are not directly connected.

We start with a concise and reasonable observation regarding whether two nodes belong to the same community. They may be surrounded by the similar environment that means the two nodes may share the similar neighbor nodes.

Therefore, we measure the similarity in a community memberships as follows:

$$\text{Similarity}(i, j) = \log \left(\frac{P(i, j)}{P(i)P(j)} \right), \tag{1}$$

where $P(i, j)$ can be expressed by $\text{Jaccard}(i, j)/D$, and $\text{Jaccard}(i, j)$ is the Jaccard index value if node i and node j. $P(i) = d_i/D$, where $d(i) = \sum_{j=1}^{J} A_{i,j}$ is the degree of node i and $D = \sum_{i=1}^{I} d_i$ is the total degree for network G. The Jaccard index value can be expressed as follows:

$$\text{Jaccard}(i, j) = \frac{|N(i)| \cap |N(j)|}{|N(i)| \cup |N(j)|}, \tag{2}$$

where $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, n$; and $N(i)$ is the neighboring nodes of node i.

According to the similarity of the node community member ships, we could get a community structure matrix $S \in R^{n \times n}$, where $S_{i,j} = \text{Similarity}(i, j)$.

In our daily lives, if two people like the same movie, can we think that they share similar hobbies? That may be right but insufficient if the movie is popular and everyone would like it. Thus, if the movie is unfashionable, we could be sure that the two people share similar hobbies. Obviously, this is also suitable for community detection. If two nodes have a same cold neighboring node (few nodes are connected to it), it will make a great contribution to the similarity between the two nodes. In other words, we should add a penalty to the hot neighboring nodes (many nodes are connected to it). By replacing Jaccard(i, j) with $L(i, j)$, $P(i, j)$ can be expressed as $L(i, j)/D$, $L(i, j) = (\sum_{p=1}^{P} 1/(1 + N(t_p)))/|N(i)| \cup |N(j)|$, where P is the count of $|N(i)| \cap |N(j)|$, $t_p \in |N(i)| \cap |N(j)|$, and $p = (1, 2, \cdots, P)$; $N(t_p)$ is the count of t_p. If $N(t_p) = 0$, Jaccard(i, j) and $L(i, j)$ will be the same.

Figure 1(a) illustrates an unweighted graph with two communities. It uses the SCI method. The edge value is 1 if two nodes are directly connected, and it cannot describe which edges are more important when detecting two communities' structure. Figure 1(b) is the result of the embedding method of CDE. It can assign more weights to the edges that form densely connected subgraphs while assigning less weight to the connection between two communities. Figure 1(c) is the display of our proposed community structure matrix. We show more information about the network structure by considering the relationships between nodes that are directly connected or share similar neighboring nodes.

We define $U \in R^{N \times K}$ as the probability distribution matrix between nodes and communities. U_{ij} stands for the propensity of node i belonging to community j, where $i = (1, 2, \cdots, N)$ and $j = (1, 2, \cdots, K)$. The community structure

matrix $S \in R\, n \times n$ can be approximately decomposed into the multiplication between the probability distribution matrix U and its transpose. For a formal model, the following holds:

$$\min_{U \geq 0} \left\| S - UU^T \right\|_F^2, \tag{3}$$

where $S_{i,j} = \sum_{k=1}^K U_{i,k} U_{k,j}^T$. The process implies that if node i and node j have similar community memberships, they have a high similarity.

3.2. Community Node Structure Part. We define $T \in R^{N \times F}$ as the node attribute matrix, where N is the count of the nodes in network and F is the feature dimension of a node. The attributes of a node are in the form of an N-dimensional binary-valued vector, and T_{i*} represents the vector of node i, where $i = (1, 2, \cdots, N)$. The node attribute function is as follows:

$$\min_{U \geq 0, M \geq 0} \left\| T - UM \right\|_F^2, \tag{4}$$

where $M \in R^{K \times F}$ and $T_{i,j} = \sum_{k=1}^K U_{i,k} M_{k,j}$.

The node attribute matrix T is decomposed into two matrixes, U and M. As mentioned above, U is the probability distribution matrix between nodes and communities, $U_{i,k}$ stands for the propensity of node i belonging to community k, M is the probability distribution matrix between nodes feature and communities, and $M_{k,j}$ is the weight of the j-th node attribute feature for community k.

In this way, we can use the node attribute information to divide the communities.

3.3. The CDCN Model. In this subsection, we will elaborate the overall model of our CDCN method. There are two parts of our method, which include the community structure part and the node attribute part.

We combine the community structure part in equation (3) and the node attribute part in equation (4) together. Therefore, our proposed model is written as follows:

$$\min_{U \geq 0, M \geq 0} \left\| S - UU^T \right\|_F^2 + \alpha \left\| T - UM \right\|_F^2, \tag{5}$$

where α is a positive parameter that adjusts the weight for the two items. If $\alpha > 1$, it means that we are more inclined to the node attribute information; in contrast, we are more inclined to the network topology information.

To make the model more generalized, we do not strictly restrict the symmetric decomposition of S. In regard to the original model in equation (3), it can be equivalently converted into the following:

$$\min_{U \geq 0, V \geq 0} \left\| S - UV^T \right\|_F^2, s.t. U = V, \tag{6}$$

where V is a surrogate variable for U. For further relaxation, equation (5) can be changed to the following:

$$L(U, V, M) = \min_{U \geq 0, V \geq 0, M \geq 0} \left\| S - UV^T \right\|_F^2 \\ + \alpha \left\| T - UM \right\|_F^2 + \beta \left\| U - V \right\|_F^2, \tag{7}$$

where β is a positive parameter to adjust the closeness between U and V. The higher it is, the closer the two variables are.

In practice, it is common to set β to a moderate value for real applications. In this way, we can enhance the generalization ability of the model. This means that we do not need to limit the community structure matrix S to be decomposed into the same matrix. In other words, we relax the matrix decomposition condition.

In summary, we use the same variable U to combine the two parts, the community structure matrix and the node attribute matrix, and to make the model more generalized, we do not strictly restricted the symmetric decomposition of S. In this way, we get the optimization function of the model.

As for the detection of overlapping communities, we identify that node v_i belongs to the j-th community when U_{ij} is higher than a predefined threshold ε. Following CDE, we set $\varepsilon = 0.1$ in our paper.

3.4. The Algorithm for CDCN. In this subsection, we will share the solution algorithm for our proposed model. The learning process algorithm for CDCN can be seen in Algorithm 1.

The matrix U_{mk}, V_{mk}, and M_{kf} in Algorithm 1 can be derived from the following derivation. According to equation (7), take the derivatives of $L(U, V, M)$ with respect to U, V, and M, we can get the formulas (8), (9), and (10), respectively:

$$\frac{\partial L(U, V, M)}{\partial U} = 2UV^T V - 2SV + \alpha \left(2U^T MM^T - 2TM^T \right) \\ + \beta (2U - 2V),$$
$$\tag{8}$$

$$\frac{\partial L(U, V, M)}{\partial V} = 2VU^T U - 2SU + \beta (2V - 2U), \tag{9}$$

$$\frac{\partial L(U, V, M)}{\partial M} = \alpha \left(2U^T UM - 2U^T T \right). \tag{10}$$

Based on this, the updating rules for the variables are given as follows:

$$U_{mk} \longleftarrow U_{mk} - \rho_{mk} \frac{\partial L(U, V, M)}{\partial U_{mk}}, \tag{11}$$

$$V_{mk} \longleftarrow V_{mk} - \theta_{mk} \frac{\partial L(U, V, M)}{\partial V_{mk}}, \tag{12}$$

$$M_{kf} \longleftarrow M_{kf} - \phi_{kf} \frac{\partial L(U, V, M)}{\partial M_{kf}}, \tag{13}$$

where ρ_{mk}, θ_{mk}, and ϕ_{kf} denote the step sizes for the (mk)th element of matrix U, the (mk)th element of matrix V and the

Input: network graph G, node attributes matrix T, hyper-parameters α and β, number of communities K and maximum number of iterations *maxIter*.
Output: the probability distribution matrix U.
Begin
According to network graph G and Eq.(1), generate the community structure matrix and randomly initialize the probability distribution matrix
$U^{(0)} \sim (0, 1)^{N \times K}$, $V^{(0)} \sim (0, 1)^{N \times K}$
$\qquad M^{(0)} \sim (0, 1)^{K \times F}$; $i = 0$
While $i \leq maxIter$ **do**
$U_{mk} \longleftarrow U_{mk}([SV + \alpha TM^T + \beta V]/[UV^TV + \alpha U^TMM^T + \beta U])$;
$V_{mk} \longleftarrow V_{mk}([SU + \beta U]/[VU^TU + \beta V]))$;
$M_{kf} \longleftarrow M_{kf}([U^TT]/[U^TUM])$;
End While.
End

ALGORITHM 1: The learning process: CDCN.

(kf)th element of matrix M, respectively, in the gradient descent methods. If we set $\rho_{mk} = U_{mk}/2(UV^TV + \alpha U^TMM^T + \beta U)$, $\theta_{mk} = V_{mk}/2(VU^TU + \beta V)$, and $\phi_{kf} = M_{kf}/2\alpha U^TU M$, then the following holds:

$$U_{mk} \longleftarrow U_{mk} \frac{[SV + \alpha TM^T + \beta V]}{[UV^TV + \alpha U^TMM^T + \beta U]}, \qquad (14)$$

$$V_{mk} \longleftarrow V_{mk} \frac{[SU + \beta U]}{[VU^TU + \beta V]}, \qquad (15)$$

$$M_{kf} \longleftarrow M_{kf} \frac{[U^TT]}{[U^TUM]}. \qquad (16)$$

The algorithm for the optimization (7) is summarized in Algorithm 1. Note that the updating with variable step sizes will naturally maintain the nonnegative constraints. Because the number of nodes N is far greater than the number of node features F, the time complexity is $O(KN^2)$.

4. Experiments and Analysis

In this section, we have conducted extensive comparative experiments to evaluate the effectiveness of our proposed CDCN model on real graph datasets with ground-truth communities.

4.1. Datasets. We consider 7 widely accepted network ground-truth community datasets, i.e., Karate, Polbooks, Football (http://www-personal.umich.edu/ mejn/netdata/), Citeseer, WebKB (http://linqs.cs.umd.edu/projects/projects/lbc/), Facebook (http://snap.stanford.edu/data/ego-Facebook.html), and HEP-TH (http://snap.stanford.edu/data/cit-HepTh.html). The network statistics are reported in Table 2. The Citeseer network consists of 3312 scientific publications with 4732 edges. The number of node attribute features in Citeseer is 3703. The WebKB network includes 4 subnetworks (i.e., Cornell, Texas, Washington, and Wisconsin), and each subnetwork consists of 5 communities. There

are 877 web pages with 1608 edges, and each webpage is annotated by 1703-dimensional binary-valued word attributes. Karate, Polbooks, and Football are nonoverlapping communities without node attributes. The HEP-TH (high energy physics theory) citation graph is from the ePrint arXiv and covers all the citations within a dataset of 27770 papers with 352807 edges, and we believe that the papers published in the same journal belong to the same community. There are some communities that contain very few nodes; therefore, we exclude these communities (less than 10 nodes) and then get the dataset of 20048 papers with 236230 edges.

The node attributes of Cornell, Texas, Washington, Wisconsin, Citeseer, and Facebook are binary vectors where the elements are either 0 or 1. The node attributes of HEP-TH are dense vector with a dimension of 300. We extract the paper titles and abstracts and then the train word vector model [41] to get the vector.

We compare different methods in these networks to prove the effectiveness of our community structure matrix. The detailed information of the datasets can be seen in Table 2.

4.2. Evaluation Methods. The compared methods may include nonoverlapping and overlapping communities, and so we choose different evaluation metrics.

(i) For nonoverlapping communities:

In terms of the measures to evaluate the quality of nonoverlapping communities, we use two evaluation metrics. We adopt the same evaluation procedure used in [17] that every detected community is matched with its most similar ground-truth community.

The first metric is the accuracy (AC [48]). Given a network containing $|V|$ nodes, for each node, $predict_i$ is the community label we obtain by applying different algorithms, and $real_i$ is the ground-truth label provided by the datasets. The accuracy is defined as follows:

$$C = \frac{\sum_{i=1}^{n} \delta(real_i, map(predict_i))}{|V|}, \qquad (17)$$

TABLE 2: Dataset statistics. $|V|$: number of nodes; $|E|$: number of edges; f: number of node attribute features; K: number of communities.

Datasets	$|V|$	$|E|$	f	K
Karate	34	156	—	2
Polbooks	105	882	—	3
Football	115	1226	—	12
Cornell	195	283	1703	5
Texas	187	280	1703	5
Washington	230	366	1703	5
Wisconsin	265	459	1703	5
Citeseer	3312	4536	3703	6
Facebook	4039	88243	10	193
HEP-TH	20048	236230	300	537

where $\delta(x, y)$ is the function δ that equals 1 if $x = y$ and it equals 0 otherwise, and $\text{map}(\text{predict}_i)$ is the mapping function that maps each community label real_i to the equivalent label from the datasets. The best mapping can be found by using the Kuhn-Munkres algorithm [49].

The second metric is the normalized mutual information (NMI [48]). In clustering applications, mutual information is used to measure the similarity of two sets of clusters. Given the discovered communities C of the results of community detection methods and a set of ground-truth communities C^*, their mutual information metric $\text{NMI}(C, C^*)$ is defined as follows:

$$\text{NMI}(C, C^*) = \sum_{c_i \in C, c_j^* \in C^*,} p\left(c_i, c_j^*\right).\log \frac{p\left(c_i, c_j^*\right)}{p(c_i).p\left(c_j^*\right)}, \quad (18)$$

where $p(c_i)$ and $p(c_j^*)$ denote the probabilities that a node arbitrarily selected from the network belongs to the community c_i and c_j^*, respectively, and $p(c_i, c_j^*)$ denotes the joint probability that this arbitrarily selected node belongs to clusters c_i and c_j^* at the same time.

(ii) For overlapping communities:

We compare a set of detected communities M with the ground-truth communities $M*$ as in [15], and the evaluation function is as follows:

$$\frac{1}{2|C*|} \sum_{c_i^* \in C^*} \max_{c_j \in C} \delta\left(c_i^*, c_j\right) + \frac{1}{2|C|} \sum_{c_i \in C} \max_{c_j \in C} \delta(c_i^* \in C^*), \quad (19)$$

where $\delta(c_i^*, c_j)$ is the similarity measure between communities c_i^* and c_j. We consider a standard metric $\delta(.)$ to quantify the similarity between a pair of communities, and the similarity score will be between 0 and 1.1 indicates the perfect recovery of the ground-truth communities.

4.3. Parameter Sensitivity Analysis.

In this section, we perform the parameter sensitivity analysis of CDCN on the Wisconsin and Washington dataset. The number of nodes in these two is appropriate, which makes it easier to see the effect. Our algorithm has two hyperparameters: α is a nonnegative constant that controls the balance between the original network topology information and node attribute information, and β is a nonnegative constant for the relaxation of matrix U. For each hyperparameter value, we repeated the experiments ten times and took the average of the ten results. The results of other datasets are similar. These two parameters are also applicable to other datasets because they have similar topological structures and network characteristics. The results of the two parameters can be seen in following figures.

Figures 2 and 3 illustrate that CDCN achieves better performances in the range of $\alpha = 0.15$ through $\alpha = 0.3$, and specifically, CDCN achieves the highest AC and NMI scores when $\alpha = 0.2$. This indicates the different importance of the original network topology information and the node attribute information. In terms of the parameter β, we set $\alpha = 0.2$ and vary β from 0 to 0.05.

Figures 4 and 5 illustrate that CDCN achieves the highest AC and NMI scores when $\beta = 0.02$. These results indicate that if we use both node attribute information and community structure information, we will get better results for real networks.

4.4. Experimental Setups.

We compared our algorithm against six topology based methods, i.e., SNMF, SLPA, DEMON, CPM, Louvain, and InfoMap; three node attributes based methods, i.e., CAN, SMR, and NC; and three methods that consider both network topologies and node attributes, i.e., PCL-DC, SCI, and CDE.

As with the experiments in [27], it is hard to compare the quality of community results when the numbers of communities are different for baseline methods. Therefore, we set the number of detected communities K as the number of ground-truth communities. We applied our proposed method and other baseline methods on the public datasets, repeat the tests ten times, and take the average of the ten results.

For all baseline methods, we set their parameters by default to achieve the best results for those methods. For example, for CDE, we set $\alpha = 1$, $\beta = 2$, and $\kappa = 5$, and for SCI, we set $\alpha = 50$ and $\beta = 1$. For more information, please refer to their papers. Regarding the parameters of our CDCN approach, $maxIter$ is set to100 to achieve convergence, and the hyperparameters α and β are set as 0.2 and 0.02, respectively. Our algorithms are implemented in python, and all experiments are performed on a PC with Windows 7, Intel(R) Core(TM) i7-4790 CPU @ 3.60 GHz and 24 GB of main memory.

4.5. Evaluation on Nonoverlapping Communities.

In this subsection, we evaluate the results on nonoverlapping communities. We report the ACs and NMIs of all methods in Table 3. The results indicate that CDCN outperforms all
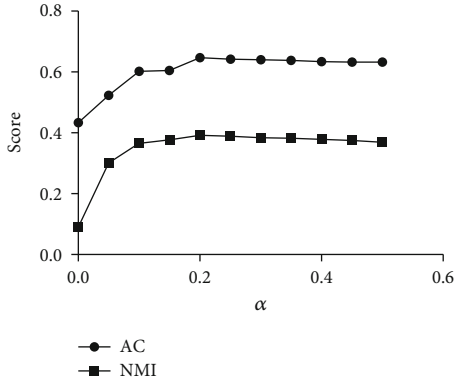
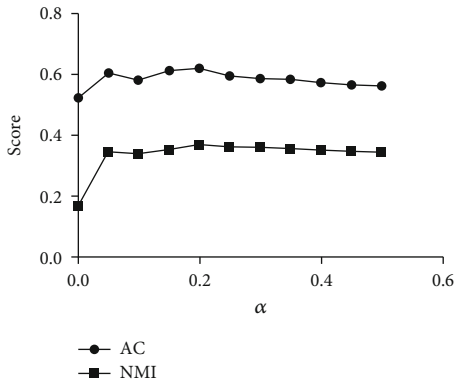FIGURE 2: Fixing $\beta = 0.01$ then varying α from 0 to 0.5 on the Wisconsin dataset.

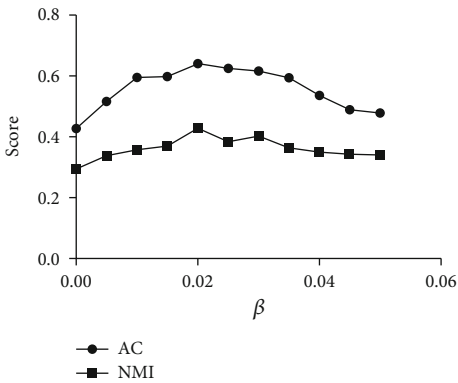FIGURE 3: Fixing $\beta = 0.01$ then varying α from 0 to 0.5 on the Washington dataset.

FIGURE 4: Fixing $\alpha = 0.2$ then varying β from 0 to 0.05 on the Wisconsin dataset.

comparison algorithms for the nonoverlapping community detection task.

The baseline comparison methods include InfoMap, CPM, SLPA, Louvain, DEMON, SNMF, CAN, SMR, NC, PCL-DC, SCI, and CDE. The real datasets include Cornell, Texas, Washington, Wisconsin, and Citeseer. All the datasets are independent of each other, and there is no connection
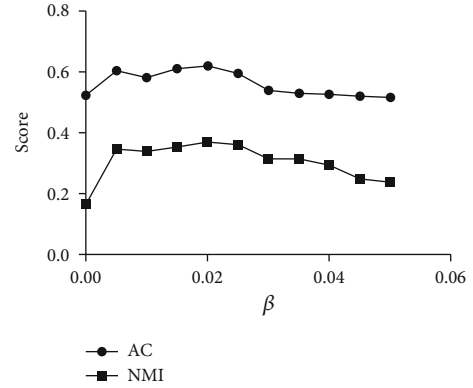

FIGURE 5: Fixing $\alpha = 0.2$ then varying β from 0 to 0.05 on the Washington dataset.

between them; therefore, they are nonoverlapping communities. We apply our method to the above datasets by using the different baseline methods.

Compared with the algorithms that focus on the original network topology or node attributes, the results show that combining both the original network topology and the inherent community structure information together will result in making great improvements. For example, among the algorithms that focus on the original network topology or node attributes, the highest ACs on Cornell, Texas, Washington, Wisconsin, and Citeseer are, respectively, 0.446, 0.545, 0.508, 0.471, 0.314, and 0.168, and ACs on our methods are 0.569, 0.641, 0.695, 0.694, 0.544, and 0.215, respectively. The values increased by 0.123, 0.096, 0.187, 0.223, 0.230, and 0.047, respectively. The same as the AC, our method also greatly improved the NMI.

The experiments results can be seen in Table 3. From Table 3, we can see that compared with the PCL-DC, SCI, and CDE models that combine both original network topology information and node attribute information, obviously, PCL-DC, SCI, and CDE will lose the information on the relationships between nodes that are not directly connected, and our model considers the relationships between nodes that share the same neighbors. The results in Table 3 prove this. In Table 3, CDE gets the best results compared with PCL-DC and SCI. The highest ACs of CDE on Cornell, Texas, Washington, Wisconsin, Citeseer, and HEP-TH are, respectively, 0.499, 0.498, 0.568, 0.645, 0.474, and 0.201. Our model improves the AC values by 0.07, 0.143, 0.127, 0.049, 0.07, and 0.014, respectively. Regarding the AC, compared with the methods that combine both the original network topology information and node attribute information, there are also huge improvements in the results.

4.6. Evaluation on Overlapping Communities. Some scholars have proposed effective methods in the detection of overlapping communities, such as subspace decomposition, maximal cliques, maximal subgraph, and the clustering coefficient. Li et al. [50] proposed a method to measure the performance of the overlapping communities. They are the state-of-the-art methods in recent years. With reference to

TABLE 3: The performances of different community detection algorithms on nonoverlapping communities measured by the AC and NMI.

Metric	Datasets	InfoMap	CPM	SLPA	Louvain	DEMON	SNMF	CAN	SMR	NC	PLC-DC	SCI	CDE	CDCN
	Cornell	0.2	0.446	0.239	0.266	0.377	0.371	0.446	0.415	0.317	0.348	0.354	0.449	0.569
	Texas	0.214	0.471	0.523	0.269	0.475	0.496	0.545	0.470	0.540	0.369	0.488	0.498	0.641
	Washington	0.143	0.469	0.434	0.204	0.381	0.410	0.491	0.508	0.456	0.408	0.401	0.568	0.695
AC	Wisconsin	0.152	0.471	0.262	0.223	0.430	0.386	0.471	0.471	0.422	0.354	0.396	0.645	0.694
	Citeseer	0.053	0.178	0.090	0.238	0.208	0.309	0.212	0.211	0.314	0.452	0.327	0.474	0.544
	HEP-TH	0.041	0.123	0.091	0.135	0.113	0.168	0.135	0.162	0.143	0.193	0.184	0.201	0.215
	Cornell	0.147	0.05	0.138	0.109	0.051	0.061	0.045	0.061	0.084	0.081	0.073	0.311	0.358
	Texas	0.102	0.077	0.040	0.059	0.043	0.097	0.021	0.09	0.115	0.068	0.097	0.252	0.328
NMI	Washington	0.117	0.006	0.158	0.087	0.079	0.035	0.046	0.117	0.038	0.103	0.086	0.341	0.406
	Wisconsin	0.111	0.027	0.106	0.078	0.047	0.070	0.037	0.070	0.077	0.071	0.069	0.406	0.427
	Citeseer	0.214	0.199	0.214	0.228	0.166	0.096	0.003	0.007	0.003	0.221	0.083	0.208	0.263
	HEP-TH	0.035	0.092	0.084	0.112	0.093	0.137	0.114	0.123	0.114	0.152	0.145	0.179	0.191

these methods, we proposed evaluation metrics under overlapping communities.

For the overlapping communities, we use the F1-Score and Jaccard-Similarity to evaluate the partitioned results of all the methods, except the clustering methods that could not discover the overlapping communities. The tested network is the complete Facebook data, and it contains 10 different ego-networks with manually identified circles. We select 4 representative ego-networks from them. The experiment results can be seen in Table 4.

The baseline comparison methods include InfoMap, CPM, SLPA, Louvain, DEMON, SNMF, PCL-DC, SCI, and CDE. The real datasets include FaceBook Ego-network 107, FaceBook Ego-network 698, FaceBook Ego-network 1912, FaceBook Ego-network 3908, and FaceBook. There are some intersections between datasets, and some overlapping areas appear; therefore, there are overlapping communities. We apply our method to the above datasets by using the different baseline methods.

Ego-network 107 has the most nodes, Ego-network 698 has the fewest nodes, Ego-network 1912 has the highest intensive degree, and Ego-network 3908 has lowest intensive degree.

As shown in Table 4, CDCN achieves the best performances on all the tested networks. In addition, it also shows that our model greatly improves community detection by combining both the original network topology and the inherent community structure information together. For instance, the highest F1-Score among the methods that focus on network topology information is 0.517, and the highest F1-Score among the methods that combine both the original network topology and the inherent community structures information together is 0.474 on FaceBook ego-network 107. Compared with those methods, our model achieves an F1-Score of 0.539. The other ego-network results are similar with those of FaceBook ego-network 107. On the complete Facebook data, our model gets an F1-Score of 0.372, which is greater than the best results of the other methods. For the Jaccard-Similarity, the results in Table 4 also indicate that CDCN outperforms all comparison algorithms.

4.7. Evaluation on Nonnode Attributes Communities. There are some communities without node attributes, and it is hard to divide these communities using the majority methods that use node attributes. However, it is easy to deal with the problem using CDCN since we could use only the community structure part of our method. Therefore, our method will be simplified as follows:

$$L(U, V) = \min_{U \geq 0, V \geq 0} \left\| S - UV^T \right\|_F^2 + \beta \|U - V\|_F^2. \quad (20)$$

The update formulas are changed into the following:

$$U_{mk} \longleftarrow U_{mk} \frac{[SV + \beta V]}{\left[UV^T V + \beta U\right]}, \quad (21)$$

$$V_{mk} \longleftarrow V_{mk} \frac{[SU + \beta U]}{\left[VU^T U + \beta V\right]}. \quad (22)$$

To prove the usefulness of our community structure matrix, we add two more baselines, which are called Adj-Mat and Emb-Mat. Adj-Mat just replaces the community structure matrix S in equation (20) with the adjacency matrix of network. Similarly, Emb-Mat just replaces the community structure matrix in equation (20) with the embedding matrix of CDE. Then, the optimization algorithm is changed, as shown in Algorithm 2.

The baseline comparison methods include InfoMap, CPM, SLPA, Louvain, DEMON, SNMF, Adj-Mat, and Emb-Mat. We assessed the NMI and AC values on the karate, football, and polbooks datasets.

In addition, in this part, we compare our method with the methods that focus on the original network topology and that do not node attribute information on four datasets. The four datasets are nonoverlapping communities, and so, we use AC and NMI to evaluate the partitioning result of all the methods. The results can be seen in Table 5.

As shown in Table 5, the results indicate that CDCN outperforms all comparison algorithms for the nonoverlapping community detection task. The NMI reached 1.0, 0.916,

TABLE 4: The performances of different community detection algorithms on overlapping communities measured by the F1-Score and Jaccard-Similarity.

Metric	Datasets	InfoMap	SLPA	Louvain	DEMON	SNMF	PCL-DC	SCI	CDE	CDCN
	FaceBook ego-network 107	0.448	0.510	0.264	0.517	0.378	0.384	0.405	0.474	0.539
	FaceBook ego-network 698	0.636	0.628	0.588	0.576	0.612	0.345	0.239	0.574	0.640
F1-Score	FaceBook ego-network 1912	0.372	0.323	0.366	0.328	0.378	0.312	0.316	0.322	0.379
	FaceBook ego-network 3908	0.579	0.528	0.567	0.387	0.410	0.421	0.388	0.471	0.580
	FaceBook	0.330	0.351	0.321	0.214	0.134	0.224	0.213	0.324	0.372
	FaceBook ego-network 107	0.372	0.410	0.205	0.421	0.267	0.304	0.294	0.369	0.432
	FaceBook ego-network 698	0.556	0.529	0.482	0.464	0.487	0.256	0.141	0.441	0.571
Jaccard-Similarity	FaceBook ego-network 1912	0.286	0.272	0.251	0.229	0.249	0.200	0.202	0.215	0.290
	FaceBook ego-network 3908	0.432	0.441	0.421	0.313	0.287	0.310	0.268	0.352	0.441
	FaceBook	0.206	0.213	0.242	0.178	0.169	0.182	0.168	0.224	0.262

Input: network graph G, hyper-parameters β, number of communities K and maximum number of iterations *maxIter*.
Output: the probability distribution matrix U.
Begin
According to the network graph G and Eq.(1), generate the community structure matrix and randomly initialize the probability distribution matrix
$U^{(0)} \in (0, 1)^{N \times K}$, $V^{(0)} \in (0, 1)^{N \times K}$; $i = 0$
While $i \leq maxIter$ **do**
Update $U^{(i+1)}$ according to Eq.(21);
Update $V^{(i+1)}$ according to Eq.(22);
End While
End

ALGORITHM 2: The optimization process of CDCN.

TABLE 5: The performances of the different community detection algorithms on nonoverlapping as community measured by the AC and NMI.

Metric	Datasets	InfoMap	CPM	SLPA	Louvain	DEMON	SNMF	Adj-Mat	Emb-Mat	CDCN
	Karate	0.581	0.652	0.658	0.569	0.429	0.836	0.825	0.912	1.0
NMI	Football	0.901	0.855	0.582	0.713	0.463	0.894	0.875	0.903	0.909
	Polbooks	0.412	0.538	0.498	0.547	0.383	0.508	0.487	0.524	0.570
	Karate	0.894	0.888	0.820	0.741	0.688	0.970	0.962	0.970	1.0
AC	Football	0.918	0.897	0.613	0.716	0.546	0.891	0.867	0.911	0.923
	Polbooks	0.695	0.819	0.785	0.823	0.740	0.743	0.724	0.798	0.838

and 0.570 on the karate, football, and polbooks datasets, respectively, which are better than the above methods. The AC value reached 1.0, 0.909, and 0.838 on karate, football, and polbooks datasets, respectively, which are also better than the above methods. Furthermore, the results compared with the Adj-Mat and Emb-Mat also show the great usefulness of our community structure matrix.

5. Conclusions

Community detection has been widely used in recommendation systems, social networks, and network security. Efficient and fast community detection algorithms contribute to the development of intelligent networks. Based on the analysis of the network characteristics, in this paper, in order to solve

the problem of community detection in attributed graphs, we propose a novel method to generate the community structure matrix, which retains the relationship between two directly nodes connected or nodes that share the same neighbors, and named it CDCN. We combine node attribute information and community structure information in an effective way in the nonnegative matrix factorization framework. We used two indicators named AC and NMI on nonoverlapping communities and two indicators named F1-Score and Jaccard-Similarity on overlapping communities to evaluate our method. On nonoverlapping communities, the AC and NMI values of CDCN are better than those of other methods. On overlapping communities, the F1-score and Jaccard-Similarity value of CDCN are better than those of other methods. The extensive experimental results demonstrated

that our algorithm can effectively discover the communities in real networks.

Data Availability

The original dataset used in this work is available from the corresponding author on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks," *Acm computing surveys (csur)*, vol. 45, no. 4, pp. 1–35, 2013.

[2] C. Pizzuti, "Evolutionary computation for community detection in networks: a review," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 3, pp. 464–483, 2018.

[3] S. Fortunato and D. Hric, "Community detection in networks: a user guide," 2016, https://arxiv.org/abs/1608.00163/.

[4] Z. Yang, R. Algesheimer, and C. J. Tessone, "A comparative analysis of community detection algorithms on artificial networks," 2016, https://arxiv.org/abs/1608.00763/.

[5] M. E. J. Newman, "Spectral methods for network community detection and graph partitioning," 2013, https://arxiv.org/abs/1307.7729/.

[6] G. Ren and X. Wang, "Epidemic spreading in time-varying community networks," *Chaos*, vol. 24, no. 2, article 023116, 2014.

[7] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proceedings of the national academy of sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[8] M. E. J. Newman and E. A. Leicht, "Mixture models and exploratory analysis in networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 23, pp. 9564–9569, 2007.

[9] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.

[10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, article P10008, no. 10, 2008.

[11] J. Xie, B. K. Szymanski, and X. Liu, "Slpa: uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process," in *2011 IEEE 11th International Conference on Data Mining Workshops*, pp. 344–349, Vancouver, BC, Canada, December 2011.

[12] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Demon: a local-first discovery method for overlapping communities," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 615–623, Beijing, China, August 2012.

[13] D. He, X. You, Z. Feng, X. Y. Di Jin, and W. Zhang, "A network-specific markov random field approach to community detection," in *Thirty-Second AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, February 2018.

[14] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, article 066111, 2004.

[15] Y. Li, K. He, D. Bindel, and J. E. Hopcroft, "Uncovering the small community structure in large networks: a local spectral approach," in *Proceedings of the 24th international conference on world wide web*, pp. 658–668, Florence, Italy, May 2015.

[16] Y. Cui and X. Wang, "Uncovering overlapping community structures by the key bi-community and intimate degree in bipartite networks," *Physica A: Statistical Mechanics & Its Applications*, vol. 407, pp. 7–14, 2014.

[17] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 587–596, Rome, Italy, February 2013.

[18] X. Huang, H. Cheng, and J. X. Yu, "Dense community detection in multi-valued attributed networks," *Information Sciences*, vol. 314, pp. 77–99, 2015.

[19] M. Atzmueller, S. Doerfel, and F. Mitzlaff, "Description-oriented community detection using exhaustive subgroup discovery," *Information Sciences*, vol. 329, pp. 965–984, 2016.

[20] X. Wang, D. Jin, X. Cao, Y. Liang, and W. Zhang, "Semantic community identification in large attribute networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 265–271, Phoenix, Arizona, USA, 2016.

[21] X. Huang, H. Cheng, and J. X. Yu, "Attributed community analysis: global and ego-centric views," *IEEE Database Engineering Bulletin*, vol. 39, no. 3, pp. 29–40, 2016.

[22] T. Yang, R. Jin, Y. Chi, and S. Zhu, "Combining link and content for community detection: a discriminative approach," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 927–936, Paris, France, 2009.

[23] X. Luo, Z. Liu, M. Shang, J. Lou, and M. C. Zhou, "Highly-accurate community detection via pointwise mutual information-incorporated symmetric non-negative matrix factorization," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 1, pp. 463–476, 2021.

[24] H. Lu, X. Sang, Q. Zhao, and J. Lu, "Community detection algorithm based on nonnegative matrix factorization and improved density peak clustering," *IEEE Access*, vol. 8, pp. 5749–5759, 2020.

[25] M. Zhang and Z. Zhou, "Structural Deep Nonnegative Matrix Factorization for community detection," *Applied soft computing*, vol. 97, no. Part B, article 106846, 2020.

[26] S. Wang, G. Li, G. Hu, H. Wei, Y. Pan, and Z. Pan, "Community detection in dynamic networks using constraint nonnegative matrix factorization," *Intelligent Data Analysis*, vol. 24, no. 1, pp. 119–139, 2020.

[27] Y. Li, C. Sha, X. Huang, and Y. Zhang, "Community detection in attributed graphs: an embedding approach," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on*

Educational Advances in Artificial Intelligence (EAAI-18), pp. 338–345, New Orleans, Louisiana, USA, 2018.

[28] X. Wang and J. Li, "Detecting communities by the core-vertex and intimate degree in complex networks," *Physica A: Statistical Mechanics & Its Applications*, vol. 392, no. 10, pp. 2555–2563, 2013.

[29] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13, Papers from Neural Information Processing Systems (NIPS)*, pp. 556–562, Denver, CO, USA, 2000.

[30] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 155–173, 2007.

[31] F. Wang, T. Li, X. Wang, S. Zhu, and C. H. Q. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, 2011.

[32] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," 2010, https://arxiv.org/abs/1008.3926/.

[33] W. Ren, G. Yan, X. Liao, and L. Xiao, "Simple probabilistic algorithm for detecting community structure," *Physical review. E, Statistical, nonlinear and soft matter physics*, vol. 79, no. 3, article 036111, 2009.

[34] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'14*, pp. 977–986, New York, NY, USA, 2014.

[35] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," in *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014*, pp. 3834–3841, Columbus, OH, USA, 2014.

[36] Jianbo Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[37] A. F. McDaid, T. B. Murphy, N. Friel, and N. J. Hurley, "Improved bayesian inference for the stochastic block model with application to large networks," *Computational Statistics & Data Analysis*, vol. 60, pp. 12–31, 2013.

[38] Y. Chen, H. Zhang, X. Zhang, and R. Liu, "Regularized semi-non-negative matrix factorization for hashing," *IEEE Transactions on Multimedia*, vol. 20, no. 7, pp. 1823–1836, 2018.

[39] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, "Experimental explorations on short text topic mining between LDA and NMF based schemes," *Knowledge-Based Systems*, vol. 163, pp. 1–13, 2019.

[40] Y. Chen, H. Zhang, Y. Tong, and M. Lu, "Diversity regularized latent semantic match for hashing," *Neurocomputing*, vol. 230, pp. 77–87, 2017.

[41] C. He, X. Fei, H. Li, Y. Tang, H. Liu, and S. Liu, "Improving NMF-based community discovery using distributed robust nonnegative matrix factorization with SimRank similarity measure," *The Journal of Supercomputing*, vol. 74, no. 10, article 2500, pp. 5601–5624, 2018.

[42] J. Wang, Y. Fan, L. Feng, Z. Ye, and H. Zhang, "Research hotspot prediction and regular evolutionary pattern identification based on NSFC grants using NMF and semantic retrieval," *IEEE Access*, vol. 7, pp. 123776–123787, 2019.

[43] D. Yu, N. Chen, F. Jiang, B. Fu, and A. Qin, "Constrained NMF-based semi-supervised learning for social media spammer detection," *Knowledge-Based Systems*, vol. 125, pp. 64–73, 2017.

[44] J. Gao, Y. Lu, J. Qi, and L. Shen, "A radar signal recognition system based on non-negative matrix factorization network and improved artificial bee colony algorithm," *IEEE Access*, vol. 7, pp. 117612–117626, 2019.

[45] N. Y. Chen, Y. Liu, and H.-C. Chao, "Overlapping community detection using non-negative matrix factorization with orthogonal and sparseness constraints," *IEEE Access*, vol. 6, pp. 21266–21274, 2018.

[46] X. Li, Z. Hu, and H. Wang, "Combining non-negative matrix factorization and sparse coding for functional brain overlapping community detection," *Cognitive Computation*, vol. 10, no. 6, article 9585, pp. 991–1005, 2018.

[47] H. Zhang, X. Niu, I. King, and M. R. Lyu, "Overlapping community detection with preference and locality information: a non-negative matrix factorization approach," *Social Network Analysis and Mining*, vol. 8, no. 1, 2018.

[48] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, pp. 3111–3119, Lake Tahoe, NV, USA, 2013.

[49] X. Cai and F. Sun, "Supervised and constrained nonnegative matrix factorization with sparseness for image representation," *Wireless Personal Communications*, vol. 102, no. 4, article 5325, pp. 3055–3066, 2018.

[50] J. Li, X. Wang, and J. Eustace, "Detecting overlapping communities by seed community in weighted complex networks," *Physica A: Statistical Mechanics & Its Applications*, vol. 392, no. 23, pp. 6125–6134, 2013.

Research Article

Pavement Crack Detection Method Based on Deep Learning Models

Guo X. Hu [iD],[1] **Bao L. Hu,**[2] **Zhong Yang** [iD],[3] **Li Huang,**[4] **and Ping Li**[1]

[1]*School of Software, Jiangxi Normal University, Nanchang 330027, China*
[2]*Foreign Language College, Jiangxi Normal University, Nanchang 330027, China*
[3]*College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[4]*School of Education, Jiangxi Normal University, Nanchang 330027, China*

Correspondence should be addressed to Zhong Yang; yz.nuaa@163.com

Severe weather and long-term driving of vehicles lead to various cracks on asphalt pavement. If these cracks cannot be found and repaired in time, it will have a negative impact on the safe driving of vehicles. Traditional artificial detection has some problems, such as low efficiency and missing detection. The detection model based on machine learning needs artificial design of pavement crack characteristics. According to the pavement distress identification manual proposed by the Federal Highway Administration (FHWA), these categories have three different types of cracks, such as fatigue, longitudinal crack, and transverse cracks. In the face of many types of pavement cracks, it is difficult to design a general feature extraction model to extract pavement crack features, which leads to the poor effect of the automatic detection model based on machine learning. Object detection based on the deep learning model has achieved good results in many fields. As a result, those models have become possible for pavement crack detection. This paper discusses the latest YOLOv5 series detection model for pavement crack detection and is to find out an effective training and detection method. Firstly, the 3001 asphalt crack pavement images with the original size of 2976 × 3978 pixels are collected using a digital camera and are randomly divided into three types according to the severity levels of low, medium, and high. Then, for the dataset of crack pavement, YOLOv5 series models are used for training and testing. The experimental results show that the detection accuracy of the YOLOv5l model is the highest, reaching 88.1%, and the detection time of the YOLOv5s model is the shortest, only 11.1 ms for each image.

1. Introduction

Asphalt pavement is damaged by natural disasters such as long-term exposure to the sun, rain erosion, and natural weathering. And asphalt pavement is also damaged by human error such as rolling of vehicles, pavement materials, construction quality, and later maintenance level. All of these causes have different degrees of impact on pavement performance [1]. If these damaged roads cannot be found and maintained in time, the service life of the highway will be shortened, the service level of the highway will be reduced, and even traffic accidents may be caused [2].

At present, pavement detection is mainly manual detection that has some disadvantages, such as taking a long time, requiring more manpower, obstructing the highway, intimidating the safety of inspectors, and interfering with the detection results by human factors. With the rapid development of a highway, it is difficult to meet the detection requirements of a large-scale highway and has completely failed to meet the needs of highway development [3].

In order to improve the service level of a highway and realize the automatic detection of damaged pavement, some researchers propose the research of automatic detection of pavement based on visual technology. In the past, people used digital image processing technology for crack detection [4]. This method mainly detects by comparing the gray value difference between the crack road and its background. However, the detection rate is low due to the complex background of pavement, various lighting, and the diversity of crack types [5, 6]. Therefore, this method can only be used as auxiliary detection.

In recent years, with the extensive use of machine learning, especially deep learning in the industrial field, it is possible to use deep learning models for automatic detection of crack pavement [7–9]. Cha et al. [10] proposed a crack pavement detection algorithm based on the convolutional neural network (CNN). The algorithm constructs an object classifier based on CNN and uses two sliding windows to scan the image to detect the crack area. However, due to the complex background interference of the photo image, the detection model cannot detect the internal features of the crack. To solve the problems existing in the above detection methods, Chen and Jahanshahi [11] improved the traditional convolutional neural network and proposed a crack detection algorithm combining the convolutional neural network with naive Bayesian (NB-CNN) data fusion. In this method, the convolution neural network and naive Bayes are integrated, which greatly increases the complexity of the model and the number of parameters and makes the model more difficult to train. It can detect the target objects in different scale images because the FCN detection model [12] combines the image features of different volume layers. So, Yang et al. [13] proposed using the full convolution network (FCN) with encoding and decoding structures for pavement crack detection. FCN can collect features of different layers, where the shallow features can concentrate on spatial information and deep features can locate the objects, and finally fuse different features to achieve a damage prediction map. On this basis, many detection models are proposed for crack pavement detection [14–17]. All of the above methods can achieve good results for crack pavement with single background and simple topology, but not for complex pavement.

Crack pavement detection belongs to the field of vision-based automatic detection. In order to improve the accuracy and robustness of crack pavement automatic detection and realize the industrial automatic detection of crack pavement, we use the excellent detection models in other fields and transfer them to the detection of crack pavement. Zhou et al. [18] propose an effective method to automatically perform the recognition and location of concealed cracks based on 3D ground-penetrating radar (GPR) and deep learning models. This proposed method by paper [18] using YOLOv4 is feasible for the detection of concealed cracks. YOLOv5, officially released in June 2020, has become one of the most popular object detection models and is the state-of-the-art model in many fields [19]. To see the effectiveness of YOLOv4 in pavement detection, this paper discusses the feasibility and implementation method of the models in the detection of crack pavement based on YOLOv5 series models.

First of all, we classify the crack pavement by analyzing the damage situation and damage degree and collect the data according to this category. We collected more than 3000 samples and randomly divided them into the training set, test set, and verification set.

Then, we test the detection models of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x in the YOLOv5 series and find that the recognition rate of all detection models is above 85%. However, as YOLOv5s is a lightweight detection model, its detection speed is far better than the other three. In terms of both recognition rate and efficiency, we think that YOLOv5s can meet the needs of practical engineering.

The paper is organized as follows. In Section 2, we introduce related works. Thereafter in Section 3, we demonstrate the detection model. Experiments are presented in Section 4. We concluded with a discussion in Section 5.

2. Related Works

2.1. Automatic Detection Based on Image Processing. The early methods [20–22] usually assumed that the crack pixel is generally darker than the surrounding and then used various threshold processing algorithms to extract the gray of the crack area. Because the threshold segmentation method is simple and fast, it has been widely used in early image segmentation. Early researchers proposed a variety of automatic detection algorithms based on threshold segmentation from different views. Li and Liu [21] proposed the threshold technique of adjacent difference histogram for automatic recognition of cracks in images. This method maximizes the difference between the two types of pixels (crack and noncrack) and achieves better experimental results than the traditional threshold method. Paper [22] detected crack pavement by processing the binary image obtained by the connected domain algorithm (directional segmentation expansion algorithm) and got good results.

Because the threshold segmentation method only considers the gray information of the image and ignores the spatial information of the image and is sensitive to noise, this algorithm is often combined with other methods to improve the segmentation accuracy. Gavilán et al. [23] found that the noncrack features in the image will present false positives. Therefore, in order to obtain the crack area, it is proposed to eliminate the false positive cracks in the noncrack image by calculating the average gray value of the pixels corresponding to the inner and outer contours of the linear object in the image. Li and Mao [24] firstly segmented the image into several complementary overlapping subimage regions. Then, the neighborhood difference histogram was used to segment and fuse the cracks in each subimage region. Lastly, the crack region is obtained in the images. This method is effective in a small range of complex fractures, but not in a large area of complex background.

The traditional threshold extraction method lacks the description of global information and is sensitive to noise. The detection effect depends on the selection of the threshold. However, in practical application, the road background
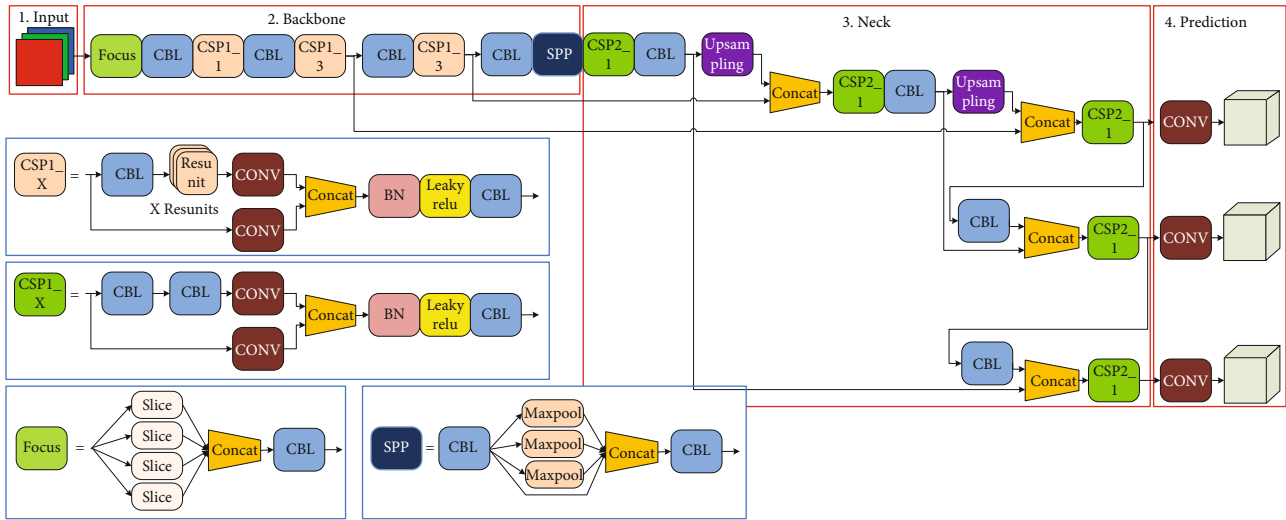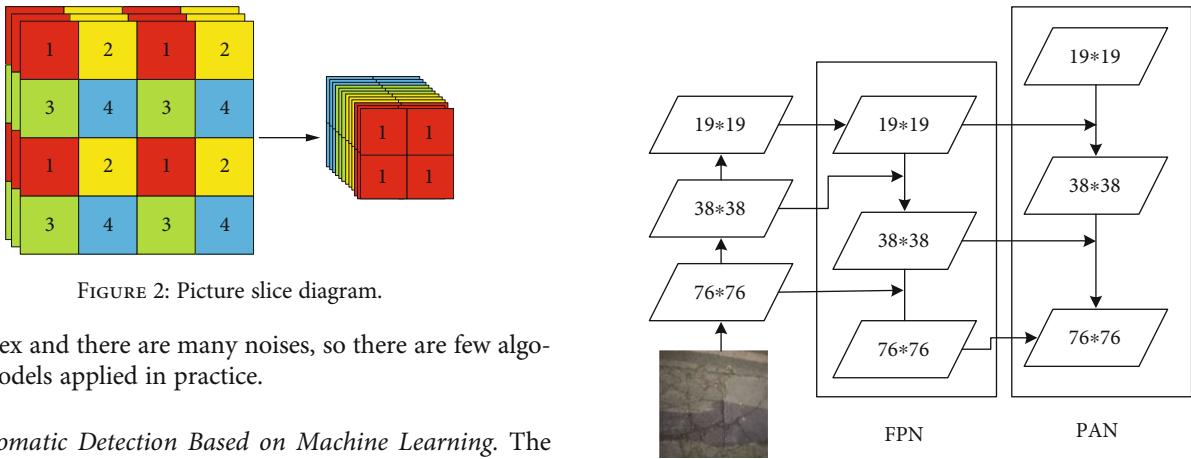
Figure 1: YOLOv5 structure.

Figure 2: Picture slice diagram.

is complex and there are many noises, so there are few algorithm models applied in practice.

2.2. Automatic Detection Based on Machine Learning. The machine learning method based on feature engineering has been successfully applied in many fields. The crack pavement has significant texture characteristics. Lots of researchers obtain the texture features of images from different views and use the classification technology of machine learning to detect the crack pavement automatically. Hu et al. [25] proposed a new automatic crack pavement detection method based on texture analysis and shape description. It is found that the texture features of the crack pavement are uneven. Therefore, that paper proposed to use six texture features and two translation invariant shape descriptors to describe the irregular texture and uneven illumination features of the image and then use the SVM classifier to classify the image into cracks and noncracks. Cord and Chambon [26] proposed a general crack detection method based on supervised learning, which can be applied to all types of defects in those images. The paper thought that the pavement with cracks presented strong texture information. The texture features of the crack pavement fluctuated in the local range and showed uniformity in the global range. Therefore, this paper uses linear and nonlinear filters to describe the image texture information. The image characteristics of different scales are analyzed by using morphological transformation, linear filtering, and nonlinear filtering. Finally, the AdaBoost classi-

Figure 3: FPN&PAN structure.

fier will be used to learn and classify the above texture information, so as to obtain the pavement damage area. The experiment shows that this method can improve the performance of crack detection to a certain extent but cannot complete the fine extraction of cracks. Shi et al. [27] proposed crack forest, a method for asphalt crack pavement extraction using random structure forest. The detection framework was structured based on the representative and distinguishing overall channel features. And it combined the framework with random forest. This model can be trained in small datasets with full supervision. In addition, the proposed two feature histograms are used to express the cracks and eliminate the noise labeled as cracks. Although this method can overcome a small part of the road noise interference, it is still unable to meet the real complex background crack extraction. Hoang and Nguyen [28] used the support vector machine (SVM), artificial neural network (ANN), and random forest (RF) to train and verify the performance of the machine learning algorithm in the dataset. The feature set composed of projective integral and fracture properties can get the most ideal results.
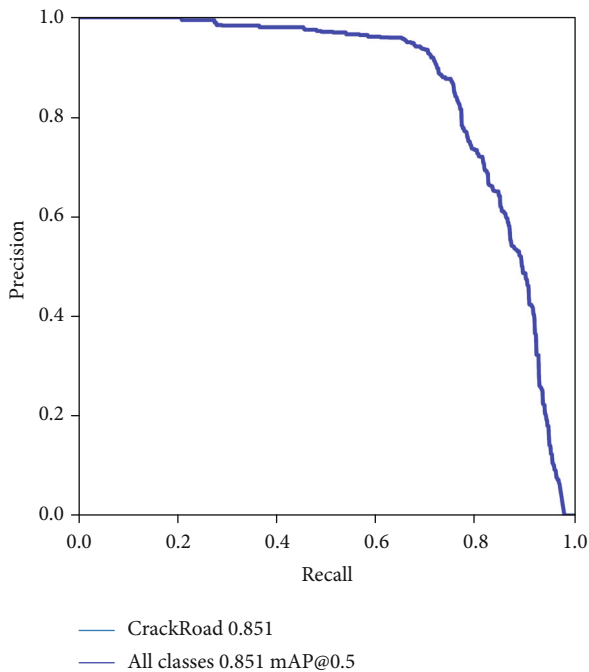
Figure 4: The *P-R* curve of the YOLOv5s model.

CrackRoad 0.851

All classes 0.851 mAP@0.5

Figure 5: Crack pavement datasets.

Table 1: Basic concept in the evaluation indicators.

Symbol	Description
TP	TP refers to the number of cracks that are correctly classified as cracks
TN	TN refers to the number of backgrounds that are correctly classified as backgrounds
FP	FP refers to the number of backgrounds that are incorrectly identified as backgrounds
FN	FN refers to the number of cracks that are incorrectly identified as the background

different scenes, different lighting, and so on. Asphalt road is widely distributed, and the road surface includes all kinds of debris and other noises. So, it is difficult to extract effective features with a unified feature model for cracks in complex and changeable road environment, which leads to poor robustness of the detection model. All in all, the detection model based on machine learning can only be applied in a small scope, not universal.

2.3. Automatic Detection Based on Deep Learning. In recent years, with the development of deep learning technology, it is possible to detect crack pavement automatically based on the deep learning model. Combined with DCNN, by learning different crack samples, the performance of automatic crack detection has been significantly improved. Some people used the methods of object detection [29, 30] or image segmentation [31, 32] to complete the crack extraction. These methods cannot complete the pixel-level detection of cracks and also
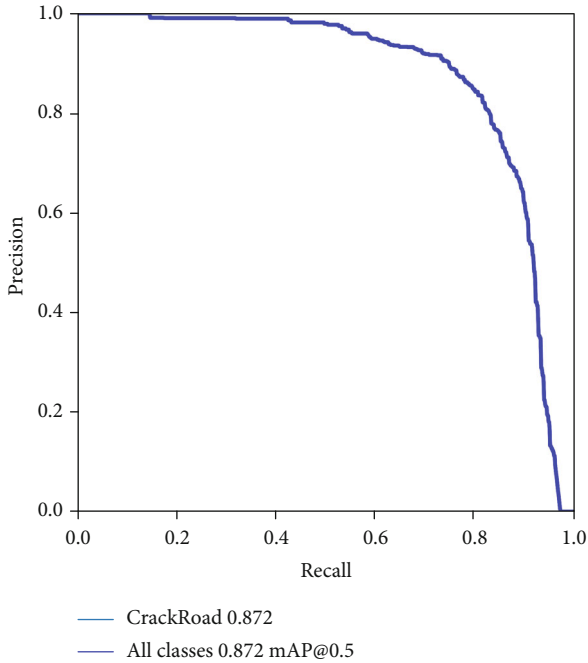
In a word, the detection model based on the machine learning algorithm greatly depends on the manually extracted image features including texture and color. And different feature extraction models need to be designed for

Figure 6: The *P-R* curve of the YOLOv5m model.

Figure 7: The *P-R* curve of the YOLOv5l model.

Figure 8: The *P-R* curve of the YOLOv5x model.

cannot accurately determine the damage category and severity in the subsequent measurement and evaluation. Zhang et al. [33] proposed a pavement automatic detection system named crack net based on the convolution neural network. The system is aimed at the pixel-level extraction of cracks and realized the automatic detection of 3D asphalt crack pavement. Unlike conventional CNN, crack net did not have any pooling layer to reduce the output of the previous layer. Crack net ensured the accuracy of crack extraction by using the constant image width and height technology in all network layers. Compared with the traditional crack detection method based on machine learning, the extraction accuracy of this method is obviously better than the traditional method. Inspired by crack net, Fei et al. [34] proposed an efficient deep network named crack net-v based on crack net, which was used for pixel-level crack automatic detection of asphalt pavement 3D image. Compared with the original crack net, crack net-v had a deeper structure and fewer parameters, which improves the calculation accuracy and efficiency. Crack net-v used the same space size for all layers, so that supervised learning could be carried out at the pixel level. The efficiency of crack net-v further revealed the advantages of deep learning technology in pixel-level pavement crack automatic detection. Zou et al. [35] proposed an end-to-end trainable deep convolution neural network for automatic crack detection, named deep crack, by learning the advanced features of crack representation. The multiscale deep convolution features learned from different convolution layers are fused together to form a linear structure. The image features obtained by this method had more detailed representation characteristics in large-scale feature maps and more comprehensive representation characteristics in small-scale feature maps. The deep crack network which is constructed on the encoder-decoder architecture of SEG net
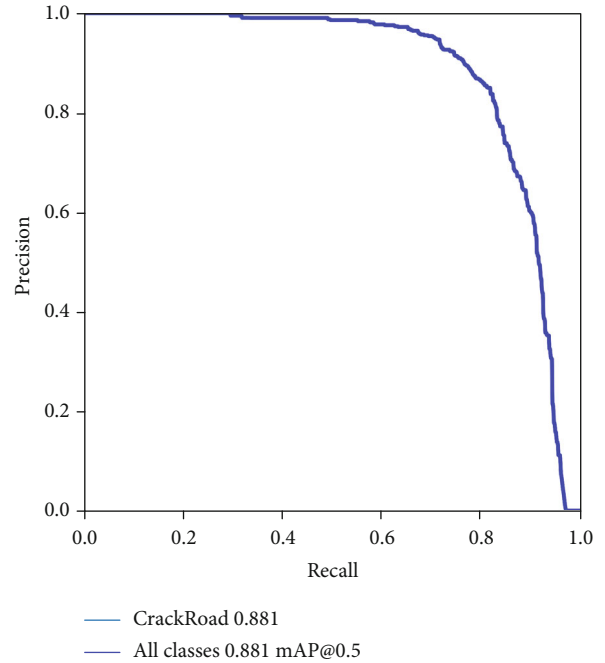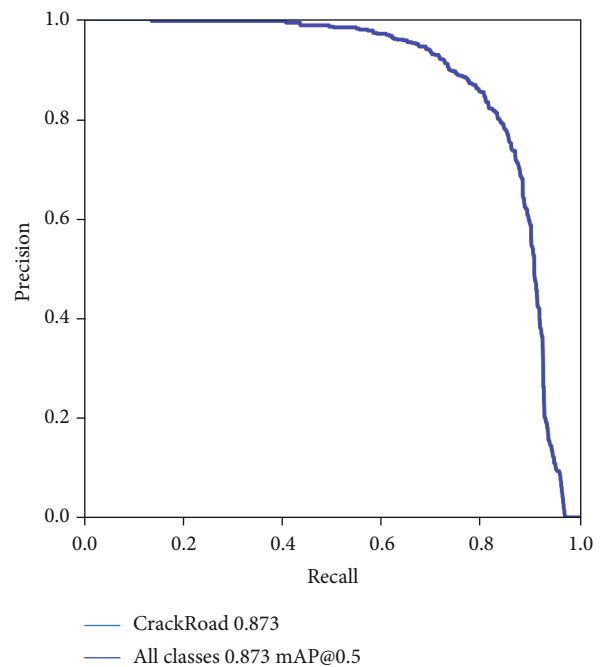
fused the convolution features generated in the encoder network and decoder network with the same scale. The deep crack could complete the pixel-level crack extraction.

In a word, the detection effect of the detection model based on deep learning is much better than that of the machine learning model based on feature engineering. Combined with the current best object detection model, and applying it to the detection of crack pavement, it will greatly improve the efficiency of pavement detection.
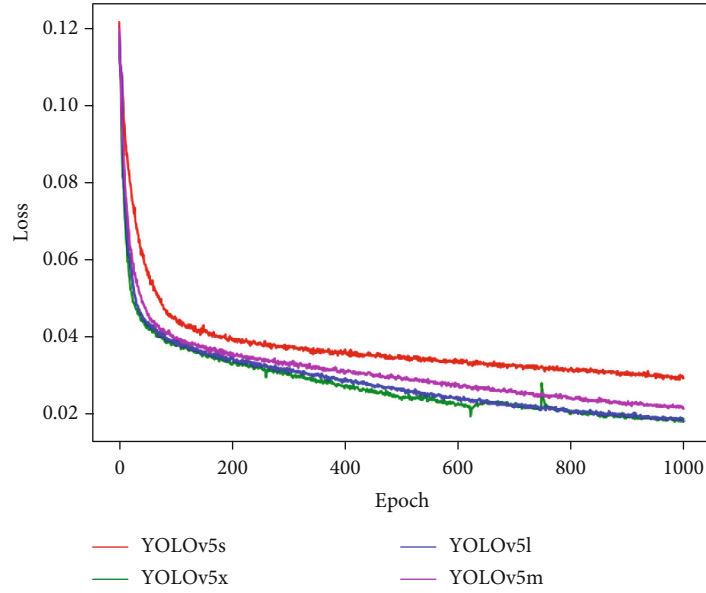
FIGURE 9: Loss figure for YOLOv5 models in training.

3. Model Introduction

At present, object detection models based on deep learning can be divided into two schools.

(1) Two-stage model: generating candidate regions and classifying candidate regions by CNN (RCNN Series [36])

(2) One-stage model: categorizing objects and locating them in one step (YOLO Series)

Following RCNN [36], Fast RCNN [37], and Faster RCNN [38], YOLO is another framework proposed by Ross Girshick for the problem of object detection speed. It has been updated to the fifth edition.

YOLOv5 consists of four parts: input, backbone, neck, and prediction. The details are shown in Figure 1 [39]. Compared with the well-known YOLOv3 [40], YOLOv5 has made the following improvements in the above four parts:

(1) In the input module, mosaic data enhancement and adaptive anchor frame calculation are added

(2) Focus and CSP structure are added in the backbone module

(3) FPN and PAN structure is added in the neck module

(4) GIOU_Loss is used in the prediction module

There are four versions of YOLOv5, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The original author thinks that the network structure of the above four versions of the model is from small to large, the corresponding detection accuracy is from low to high, but the detection speed is from fast to slow. That is to say, YOLOv5s has the smallest network, the least speed, and the lowest AP accuracy. If the

detected objects are mainly large objects, the network can quickly detect the detected objects. YOLOv5s cannot detect small objects if there are a lot of small cracks in the crack pavement.

In order to find out which model has the best combination of detection accuracy and efficiency, this paper uses the four models in YOLOv5 for road detection.

3.1. Input Module

3.1.1. Mosaic Data Enhancement. Yun et al. [41] proposed a new data enhancement method, CutMix, to enhance the diversity of samples by mixing two images and clipping them. The mosaic data enhancement method proposed by YOLOv5 mixes and stitches four images based on CutMix to produce a new image. Learning the new mosaic image is equivalent to learning four pictures at the same time, which improves the learning efficiency. Meanwhile, because the BN calculation is also four pictures at the same time, the minibatch in the training can be set to a smaller value. In a word, YOLOv5 can use one GPU to speed up the training without changing the detection accuracy.

The implementation steps of mosaic data enhancement are as follows:

(1) Four images are randomly read from the dataset at a time

(2) Flip left and right, zoom in and out, and change the brightness, saturation, and hue gamut of the four images

(3) The transformed image is combined with pictures and bounding boxes. The first picture is placed on the top left, the second picture is placed on the bottom left, the third picture is placed on the bottom right, and the fourth picture is placed on the top right

TABLE 2: Performance comparison table of 4 detection models.

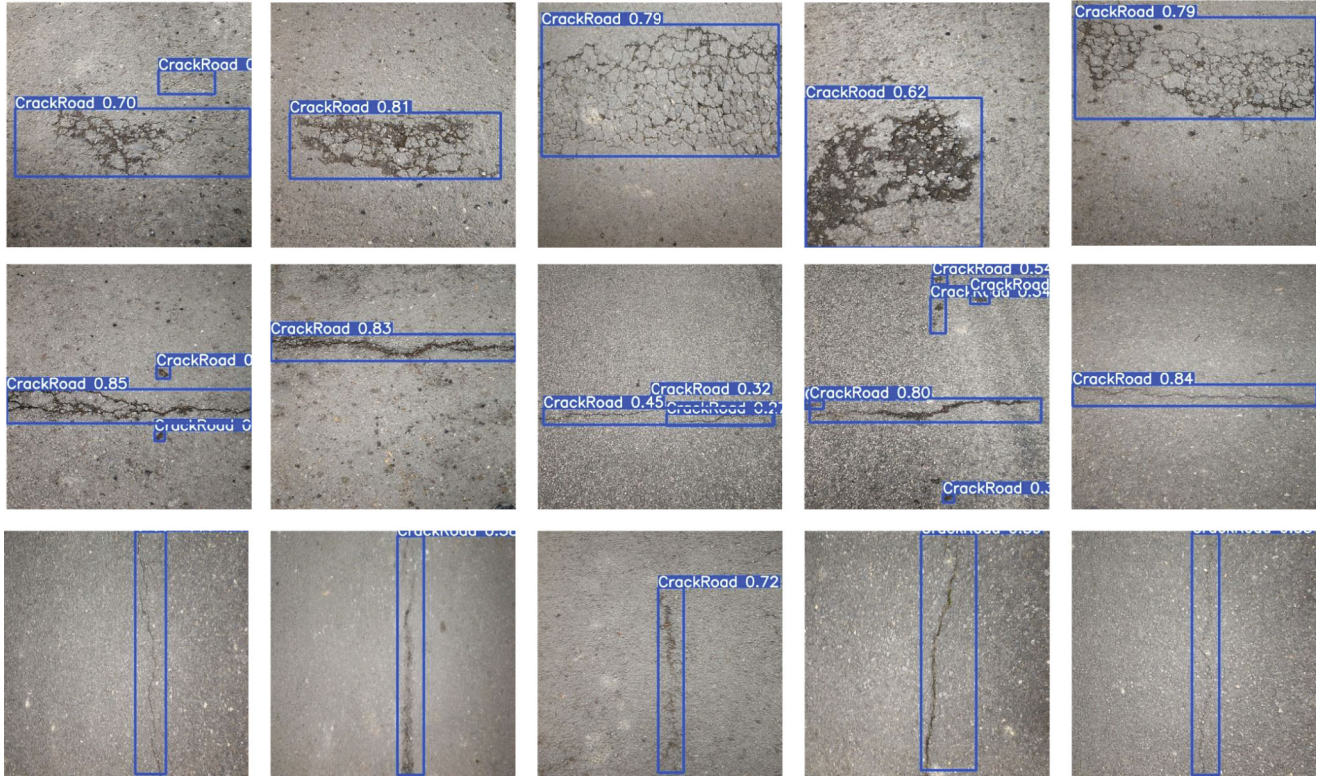| Model	Precision	Recall	mAP0.5	Time	Model parameters	Net layers
YOLOv5s	0.891	0.512	0.851	11.1 ms	7,063,542	283
YOLOv5m	0.908	0.556	0.872	24.2 ms	21,056,406	391
YOLOv5l	0.908	0.569	0.881	40 ms	46,631,350	499
YOLOv5x	0.9	0.582	0.873	67.7 ms	87,244,374	607

FIGURE 10: The detection chart of the YOLOv5s model.

in four directions. They are spliced into a new picture, which contains a bounding box and other contents

(4) When the bounding box (or the picture itself) of a picture exceeds the dividing line between two pictures, we need to remove the part that exceeds the dividing line or the part of the picture for edge processing

3.1.2. Adaptive Anchor Calculation. YOLOv5 sets the initial anchor for each image. In the process of network training, the prediction bounding box is output based on the initial anchor. And then, the bounding box is compared with the real ground truth. The gap between those is calculated. Lastly, the network parameters are updated backward and iterated.

3.1.3. Adaptive Image Scaling. Since different images are different in length and width, the common way is to scale the original image to a standard size and then send it to the detection network for the traditional object detection algorithms.

The YOLOv5 model believes that many pictures have different aspect ratios. So, the sizes of black edges at both ends are different after the picture is scaled and filled. If there are more black edges filled, there will be information redundancy, which will affect the training speed.

Therefore, the author improves the traditional fixed-size scaling and adopts the adaptive method of adding the least black edges for the YOLOv5 model. The specific methods are as follows:

(1) Calculate the scale. If the ratio of length to width of the original image is not fixed, there may be two scaling ratios; the smaller one is chosen

(2) The length and width of the scaled image are calculated, respectively

(3) Calculate the number of the black edge filled. YOLOv5 is filled with gray pixels instead of black pixels, that is (114,114,114). At the same time, YOLOv5 does not use the way of reducing the black edge in training but uses the traditional way of filling,
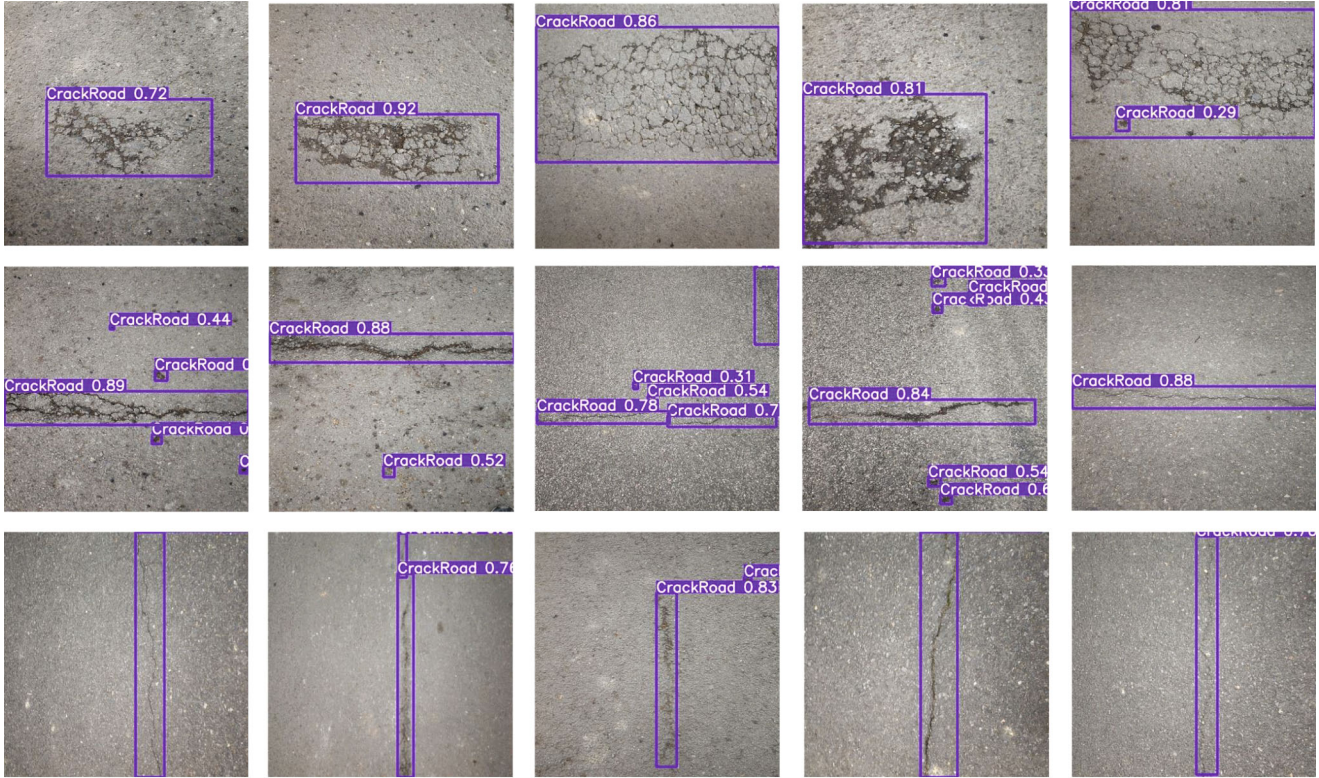
FIGURE 11: The detection chart of the YOLOv5m model.

that is, reducing image to the size of 416×416. In the test phase, the black edge reduction method is adopted to improve the speed of object detection and reasoning when the model reasoning is used

3.2. Backbone Module

3.2.1. Focus Structure.
Focus is mainly used for image slicing, as shown in Figure 2. We slice the $4 \times 4 \times 3$ image into a $2 \times 2 \times 12$ feature map.

For YOLOv5, the $608 \times 608 \times 3$ image is sliced into a $304 \times 304 \times 12$ feature map. Then, the feature map is convoluted by 32 convolution cores once and finally becomes a $304 \times 304 \times 32$ feature map.

3.2.2. CSP Structure.
The author thinks that the problem of excessive reasoning is caused by the repetition of gradient information in network optimization. Cross stage partial network (CSPnet) [42] integrates the change of gradient into the feature graph from beginning to end, which can reduce the amount of calculation and ensure accuracy.

Two CSP structures are designed in YOLOv5. Taking the YOLOv5s model as an example, the CSP1_X structure is applied to the backbone network; another kind of CSP2_X structure is used in the neck. After adding CSP, the training speed of the model is improved. The main advantages of the CSP structure are as follows:

(1) Enhance the learning ability of CNN and make it lightweight while maintaining accuracy

(2) Reduce computing bottlenecks

(3) Reduce memory cost

3.3. Neck Module.
For object detection, in order to better extract fusion features, some layers are usually inserted in the backbone and output layer, which is called the neck. YOLOv5 adopts the FPN+PAN structure in the neck module.

The high-level convolution layer has abstract description meaning for a large object. However, due to the lack of pixel information, object features cannot be described for small objects. FPN [43] combines convolution features of different layers for images, which can not only satisfy the abstract description of large objects but also ensure the feature details of small objects. YOLOv5 uses FPN in the neck layer to ensure that objects are not missed.

In order to improve the detection accuracy of small objects, FPN uses the low-level features of the image. Although low-level features help to detect objects, it is more and more difficult to accurate the position of the object using the features combining the underlying structure feature with the top-level feature. Path aggregation network (PANet) [44] shorts the information path between low-level and top-level features through the bottom-up path enhancement and the whole feature level that is enhanced by using accurate low-level positioning signals. The model connects the feature grid with all feature layers, so that the useful information in each feature layer can be directly propagated to the following subnetwork.
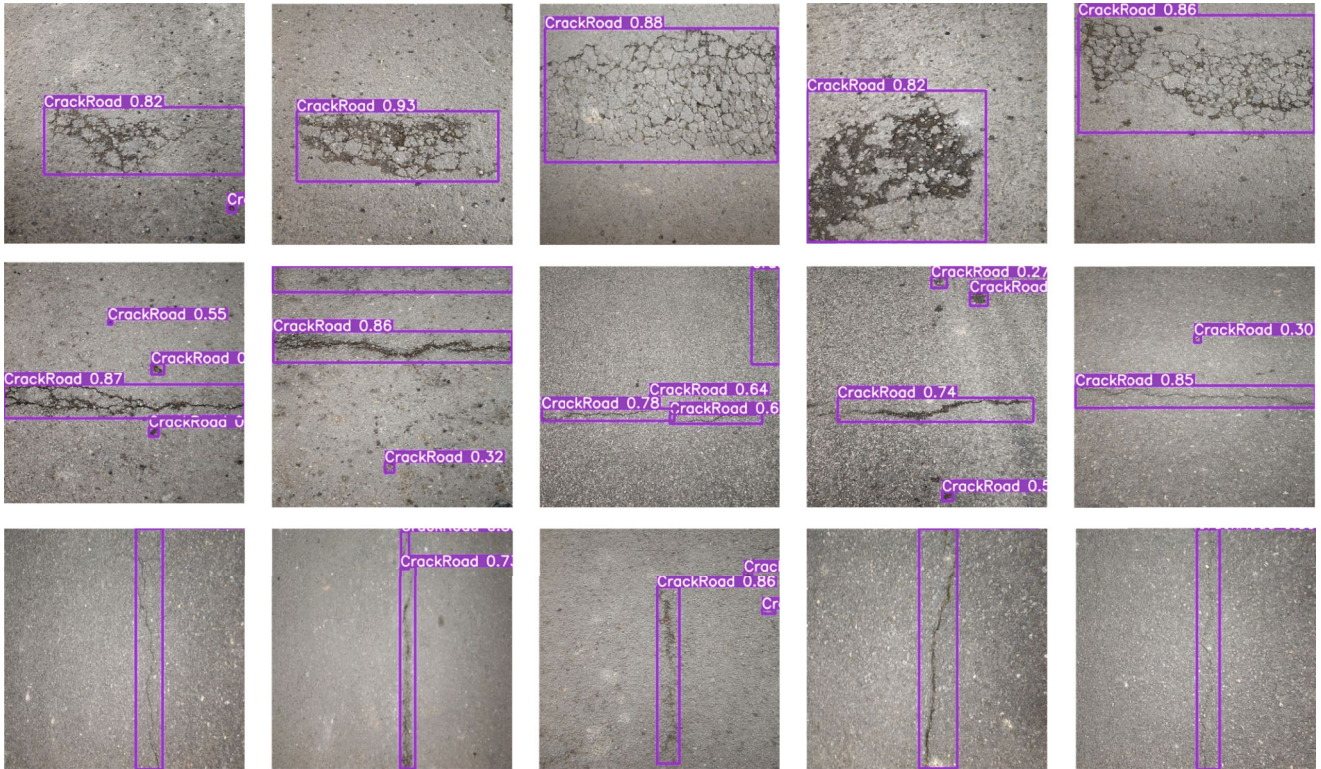
FIGURE 12: The detection chart of the YOLOv5l model.

YOLOv5 uses FPN and PANet in the neck to ensure detection accuracy and positioning accuracy. The details are shown in Figure 3.

3.4. Prediction Module. In YOLOv5, Leaky ReLU is used in the middle/hidden layer of the model. The final detection layer uses sigmoid as the activation function. In order to increase the detection of occluded overlapped objects, YOLOv5 uses DIOU-nm. The default optimization method of the model is the gradient descent method. The loss function of the model consists of three parts: object score loss, category probability loss, and bounding box regression loss. Among them, logits loss is used for object score loss, cross-entropy loss is used for category probability loss, and GIOU [45] loss is used for bounding box regression loss.

4. Experimental Analysis

4.1. Dataset Acquisition. The Federal Highway Administration (FHWA) divides pavement cracks into three categories, including fatigue crack, longitudinal crack, and transverse crack. When we take pictures of crack pavement, we pay attention to various types of sample collection. At the same time, for each type of crack degree is different, we will take samples of crack severity that is divided into high, medium, and low. In practical engineering, there are lots of interference in the pavement that is often accompanied by sign lines, fallen leaves, oil, garbage, and light. These interferences have a great impact on detection accuracy. In order to ensure that the training data samples cover all kinds of conditions as far

as possible, we also consider the above interference factors in the process of data acquisition.

Under the condition of combining all the above crack types and interference factors, we took 3001 photos of crack pavement with a Nikon camera. The picture is 2976×3968 pixels. Considering the existence of various types of crack and interference in practical application, we did not conduct separate model training for each type of crack data in the training process. The data was randomly divided into 1920 training datasets, 480 verification datasets, and 601 test datasets. The specific sample data is shown in Figure 4.

In Figure 5, the first line is fatigue crack type data, the second line is transverse crack type data, and the third line is longitudinal crack type data. For the fatigue crack, due to the large damage area, there is ambiguity in many places. So, this type of data is difficult for machine detection. For transverse and longitudinal cracks, these types of data have the characteristics of thin and long, and many damaged areas are difficult to distinguish by human eyes. Therefore, the detection model is another challenge for the detection of a small crack.

4.2. Experimental Environment. All the algorithms in this paper were programmed by Python3. The models were implemented by using the classical deep learning-based pytorch framework and trained on the Ubuntu experimental platform, Intel® Xeon® Gold 5218 CPU@ 2.30 GHz processor, 260 GB RAM, and NVIDIA Tesla P100-PCIE 16 GB GPU. Pytorch was born to carry out quick experimentation, Nvidia driver version is 4450.51.05, CUDA version is 11.0, and it is also able to quickly convert its ideas
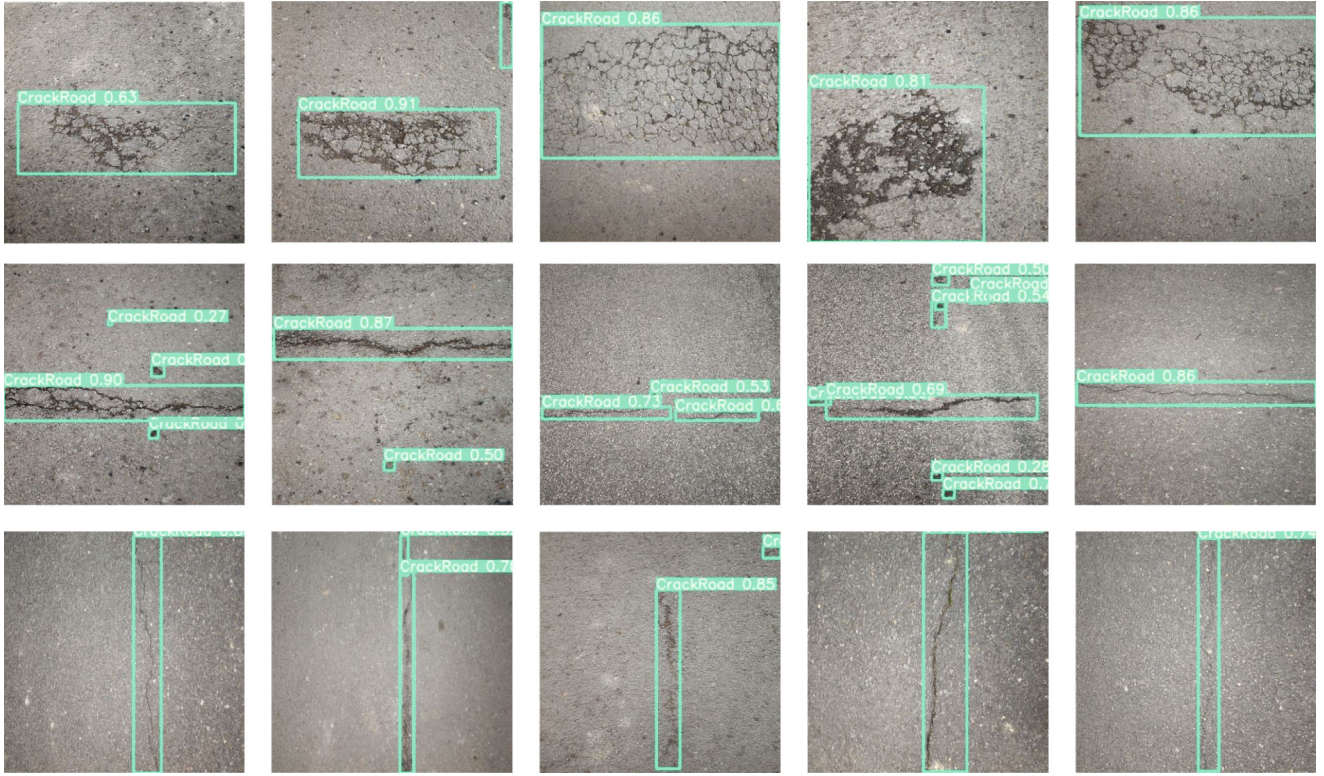
FIGURE 13: The detection chart of the YOLOv5l model.

into results. For this reason, pytorch can quickly and efficiently compare the visualization results of different crack detection algorithms.

4.3. Model Testing. In order to test the performance of the YOLOv5 model in crack pavement detection, we test it on the dataset described in Section 4.1. In this paper, accuracy, recall, and F1 score are used to quantify different crack classification algorithms. These three indicators were calculated with TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively. The specific definitions are shown in Table 1.

The formula for precision, recall, and F1 score is as follows:

$$
\begin{aligned}
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
\text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
\text{F1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
\end{aligned}
\tag{1}
$$

Since there are many types of crack pavement and the degree of different types of crack is different, it is easy to produce imbalance between different types of datasets under the premise of certain datasets. Both precision and recall in the *P-R* curve consider the detection rate of positive samples, which reduces the error between evaluation criteria due to data imbalance. The detection effect of the model is described by the *P-R* curve (see Figures 4 and 6–8 for details).

In the process of training, four pretraining weights (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) provided by YOLOv5 are used for training. In the test process, ensure that the superparameter settings and training weights are the same, such as the learning rate and momentum parameters. The training losses of the four models are shown in Figure 9. It can be seen from the figure that in the process of the same number of iterations, the model with deep layers has better training effect and faster convergence.

The network layers of v5s, v5m, v5l, and v5x, which are provided by YOLOv5, increase from small to large, and the network parameters also increase from small to large. In order to obtain the influence of different size models on the detection effect, we use four models to compare and analyze the crack datasets collected in this paper. The details are shown in Table 2.

According to Table 2, it can be seen that the model with the highest detection accuracy is YOLOv5l; mAP reaches 0.881. However, the detection accuracy between all the four models is not very different, and the maximum difference is within 0.03. The smallest network model is the YOLOv5s model, which has only 283 layers. Because the model level of YOLOv5s is the smallest, the parameters are the least among all models, only 7,063,542. Due to the different sizes of the model, in the detection process, the small model takes less time. On the contrary, the large model takes more time. Therefore, for the four models mentioned above, the least time-consuming model is YOLOv5s. For each $640 * 640$ image, the detection practice only takes 11.1 ms, while YOLOv5l takes 40 ms. The most time-consuming is YOLOv5x, which reaches 67.7 ms. For the above four models,

in order to test the detection effect of the models on the actual project, we use the above four models to test in the datasets shown in Figure 5, and the effect is shown in Figures 10–13.

5. Conclusions

Pavement crack is very difficult to detect because of many categories and larger influence of the surrounding environment. The existing pavement crack detection models fail to detect the cracks because the features of objects are difficult to be extracted by many convolution and pooling operations. Combined with one of the state-of-art object detection models YOLOv5, we discuss the possibility of transferring this model to crack pavement detection. The experimental results show that the detection accuracy of the YOLOv5 series models is above 85%. The shortest time-consuming YOLOv5s model only needs 11.1 ms to detect 640×640 pixels. Therefore, if we pay attention to the detection rate in the actual project, we can choose to use YOLOv5l. If we need to consider both the detection rate and detection efficiency, we can choose the YOLOv5s model. However, no matter which model is used, it can only be used as an aid.

Automatic crack pavement detection is one of the difficult research contents in the field of object detection. The detailed features of the crack pavement have semantic relevance with its surrounding road surface. And the subsequent semantic segmentation technology can be combined to further improve the model detection accuracy. Meanwhile, YOLO is an object detection framework based on anchors. Although YOLO has done a lot of optimization in reducing the amount of computation compared with an anchor-free model, YOLO still needs a lot of computing resources in the training process. The detection rate of the traditional detection model based on anchors is higher than that of the anchor-free model. How to improve the detection rate and reduce the resources is a problem that many engineering applications need to consider. Centernet2 [46] adopts an anchor-free framework, and the detection rate of this model is higher than that of all existing models based on anchors. The proposal of centernet2 provides direction for our next work. In order to meet the requirements of a lightweight model in practical engineering, we will study the detection based on anchor-free in the future.

Data Availability

The datasets we used in this paper enable us to shoot with a Nikon camera. We took 3001 photos of crack pavement. The picture is 2976×3968 pixels. Considering the existence of various types of crack and interference in practical application, we did not conduct separate model training for each type of crack data in the training process. The data was randomly divided into 1920 training datasets, 480 verification datasets, and 601 test datasets.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] V. P. Tran, T. S. Tran, H. J. Lee, K. D. Kim, J. Baek, and T. T. Nguyen, "One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects," *Journal of Civil Structural Health Monitoring*, vol. 11, no. 1, pp. 205–222, 2020.

[2] X. Z. Xiang, Y. Q. Zhang, and A. El Saddik, "Pavement crack detection network based on pyramid structure and attention mechanism," *IET Image Processing*, vol. 14, no. 8, pp. 1580–1586, 2020.

[3] H. N. Nguyen, T. Y. Kam, and P. Y. Cheng, "An automatic approach for accurate edge detection of concrete crack utilizing 2D geometric features of crack," *Journal of Signal Processing Systems*, vol. 77, no. 3, pp. 221–240, 2014.

[4] Z. Y. Yang, "The design of glass crack detection system based on image preprocessing technology," in *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, pp. 39–42, Chongqing, China, 2014.

[5] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.

[6] Y. Yi, Y. Chen, J. Wang, G. Lei, J. Dai, and H. Zhang, "Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization," *Signal Processing-Image Communication*, vol. 89, article 115984, 2020.

[7] D. Ma, H. Y. Fang, B. H. Xue, F. M. Wang, M. A. Msekh, and C. L. Chan, "Intelligent Detection Model Based on a Fully Convolutional Neural Network for Pavement Cracks," *CMES-Computer Modeling in Engineering & Sciences*, vol. 123, no. 3, pp. 1267–1291, 2021.

[8] Y. Wang, K. Song, J. Liu, H. Dong, Y. Yan, and P. Jiang, "RENet: rectangular convolution pyramid and edge enhancement network for salient object detection of pavement cracks," *Measurement*, vol. 170, article 108698, 2021.

[9] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Nonnegative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.

[10] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.

[11] F. C. Chen and M. R. Jahanshahi, "NB-CNN: deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2018.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, 2015.

[13] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1090–1109, 2018.

[14] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1525–1535, 2020.

[15] H. Chen, H. Lin, and M. Yao, "Improving the efficiency of encoder-decoder architecture for pixel-level crack detection," *IEEE Access*, vol. 7, pp. 186657–186670, 2019.

[16] Z. Fan, C. Li, Y. Chen et al., "Ensemble of deep convolutional neural networks for automatic pavement crack detection and measurement," *Coatings*, vol. 10, no. 2, p. 152, 2020.

[17] X. Z. Xiang, Y. Q. Zhang, and A. E. Saddik, "Pavement crack detection network based on pyramid structure and attention mechanism," *IET Image Processing*, vol. 14, no. 8, pp. 1580–1586, 2019.

[18] S. Zhou, Y. Bi, X. Wei et al., "Automated detection and classification of spilled loads on freeways based on improved YOLO network," *Machine Vision and Applications*, vol. 32, no. 2, 2021.

[19] E. U. Rahman, Y. Zhang, S. Ahmad, H. I. Ahmad, and S. Jobaer, "Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs," *Sensors*, vol. 2, no. 3, 2021.

[20] Q. Li, Q. Zou, D. Zhang, and Q. Mao, "FoSA: F∗ seed-growing approach for crack-line detection from pavement images," *Image and Vision Computing*, vol. 29, no. 12, pp. 861–872, 2011.

[21] Q. Q. Li and X. L. Liu, "Novel approach to pavement image segmentation based on neighboring difference histogram method," in *2008 Congress on Image and Signal Processing*, pp. 792–796, Sanya, China, 2008.

[22] F. Liu, G. Xu, Y. Yang, X. Niu, and Y. Pan, "Novel approach to pavement cracking automatic detection based on segment extending," in *2008 International Symposium on Knowledge Acquisition and Modeling*, pp. 610–614, Wuhan, China, 2008.

[23] M. Gavilán, D. Balcones, O. Marcos et al., "Adaptive road crack detection system by pavement classification," *Sensors*, vol. 11, no. 10, pp. 9628–9657, 2011.

[24] Q. Q. Li and Q. Z. Mao, "Land-borne pavement rapid test and measurement," *Journal of Transport Information and Safety*, vol. 27, no. 1, pp. 7–10, 2009.

[25] Y. Hu, C. X. Zhao, and H. N. Wang, "Automatic pavement crack detection using texture and shape descriptors," *IETE Technical Review*, vol. 27, no. 5, pp. 398–405, 2010.

[26] A. Cord and S. Chambon, "Automatic road defect detection by textural pattern recognition based on AdaBoost," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 4, pp. 244–259, 2012.

[27] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.

[28] N. D. Hoang and Q. L. Nguyen, "A novel method for asphalt pavement crack classification based on image processing and machine learning," *Engineering with Computers*, vol. 35, no. 2, pp. 487–498, 2019.

[29] H. F. Li, J. P. Zong, J. J. Nie, Z. L. Wu, and H. Y. Han, "Pavement crack detection algorithm based on densely connected and deeply supervised network," *IEEE Access*, vol. 9, pp. 11835–11842, 2021.

[30] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection using deep neural networks with images captured through a smartphone," 2018, http://arxiv.org/abs/1801.09454.

[31] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," in *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp. 917–924, Banff, AB, Canada, 2017.

[32] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, p. 3452, 2018.

[33] A. Zhang, K. C. P. Wang, B. Li et al., "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp. 805–819, 2017.

[34] Y. Fei, K. C. P. Wang, A. Zhang et al., "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 273–284, 2020.

[35] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "DeepCrack: learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2019.

[36] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 2014.

[37] R. Girshick, "Fast R-CNN," in *The 15th IEEE International Conference on Computer Vision (ICCV'15)*, pp. 1440–1448, Santiago, Chile, 2015.

[38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[39] D. B. Jiang, "Easy to understand Yolo series: Yolov5 core basic knowledge complete explanation," https://blog.csdn.net/nan355655600/article/details/107852353.

[40] R. Joseph and F. Ali, "YOLOv3: an incremental improvement," 2018, http://arxiv.org/abs/1804.02767.

[41] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6022–6031, Seoul, Korea, 2019.

[42] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: a new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1571–1580, Seattle, WA, USA, 2020.

[43] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, Honolulu, HI, USA, 2017.

[44] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, Salt Lake City, UT, USA, 2018.

[45] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666, Long Beach, CA, USA, 2019.

[46] X. Y. Zhou, V. Koltun, and P. Krähenbühl, "Probabilistic two-stage detection," 2021, http://arxiv.org/abs/2013.07461.