

Research Article

Multimodal Blog Sentiment Classification Based on MD-HFCE

Baozhen Yang ^{1,2} and Xuedong Tian ^{1,2}

¹School of Cyber Security and Computer, Hebei University, Baoding 071002, China

²Intelligent Image and Document Information Processing Institute of Hebei University, Hebei University, Baoding 071002, China

Correspondence should be addressed to Xuedong Tian; xuedong_tian@126.com

Received 25 May 2021; Revised 27 July 2021; Accepted 5 September 2021; Published 23 September 2021

Academic Editor: Feiran Huang

Copyright © 2021 Baozhen Yang and Xuedong Tian. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the rapid growth of multimodal information has become an important factor affecting the results of sentiment analysis. However, a few state-of-the-art works take into account the multimodal features and sentiment fuzziness. To this end, a fuzzy method is proposed for assessing sentiment intensity in this paper. Firstly, based on the visual-text conversion network (CNN-LSTM), as well as sentiment optimization through SentiBank and SentiBridge, the visual features are normalized to the text features. At the same time, the emotional features of the extracted audio will be predicted by the random forest algorithm. Subsequently, the sentiment characteristics are processed by dual hesitant fuzzification to form positive and negative sentiment intensity factors. Finally, a classification method, that is, MD-HFCE (multilayer dual hesitant fuzzy comprehensive evaluation), fuzzy comprehensive evaluation method improved by Mamdani fuzzy reasoning, is proposed to realize the multifeature fuzzy sentiment classification based on the comprehensive sentiment dictionary. The classification results are applicable to the topics of sentiment monitoring. The experimental results show that the proposed algorithm can effectively realize feature integration and improve the average sentiment classification accuracy of multimodal blogs to 82.2%.

1. Introduction

With the advent of the information age, enormous data is generated by users on the Internet in real time. It is important to utilize the data for sentiment analysis to achieve public opinion monitoring, stock market prediction, and consumption preference analysis [1]. Due to the diversity of social information, multimodal sentiment analysis has attracted great attention from researchers. To this end, various methods are investigated in this research field.

The traditional dictionary method ignores a lot of multimodal information containing emotions. Although the extended dictionaries can solve the problem to some extent, the performance improvement is still limited. The sentiment analysis approaches based on the emerging machine learning and neural network can effectively utilize the multimodal information. However, they fail to consider the sentiment fuzziness and may lead to long runtime due to processing huge data of images and videos.

To solve these problems, the multilayer dual hesitant fuzzy comprehensive evaluation (MD-HFCE) method is proposed in this paper. It is mainly based on the improved fuzzy comprehensive evaluation model of Mamdani fuzzy reasoning. Moreover, the feature transformation model in the convolutional neural network and long short-term memory (CNN-LSTM) neural network is utilized. The main contributions of this paper are summarized as follows:

- (1) The dual hesitant fuzzy set is used to fuzzify the sentiment intensity, which considers both positive and negative sentiment factors at the same time.
- (2) Video frames are filtered following a two-step scheme of image selection. The visual features are normalized by the text features based on the visual-text conversion network model and sentiment knowledge graph.
- (3) The MD-HFCE method is proposed for sentiment analysis according to the fuzzy sentiment intensity.

The rest of this paper is organized as follows: in Section 2, we briefly introduce previous work on recent sentiment analysis. In Section 3, we describe the overall process of the proposed method, and then, the experimental results are given in Section 4. Finally, we conclude the study in Section 5.

2. Related Work

At present, the main approaches of sentiment analysis can be divided into four categories: sentiment dictionary, machine learning, deep learning, and hybrid methods.

The most common methods of sentiment analysis are to build extended dictionaries [2, 3]. The concept of sentiment is considered based on the dictionary, and sentiment analysis is realized through sentiment embedding [4]. With the automatic construction of the domain dictionary, the context is considered, and thus the performance of the basic dictionary can be improved [5].

The work in [6–10] studied machine learning based sentiment analysis. Support Vector Machine (SVM) is used to achieve the sentiment analysis of images by combining the SentiBank which is formed using adjective-noun pairs [6]. Thelwall et al. [7] considered both positive and negative sentiment and proposed the SentiStrength algorithm to reduce affective disruption. In [8], the naive Bayesian network was established for sentiment evolution experiments. Yuan et al. [9] introduced the SentiStrength algorithm for image sentiment classification, which constructs a sentiment prediction framework based on the image features. Chen et al. [10] expanded the review text using a knowledge map to improve the accuracy of sentiment analysis on online travel review texts.

CNN can effectively integrate multimodal information [11, 12], which improves the sentiment analysis of the long text by increasing the convolutional layers [13]. In addition, the LSTM network pays attention to the semantic environment where the sentiment words are located [14]. As a result, the combination of CNN and LSTM network can improve the classification accuracy by integrating features with time factors [15]. The neural network model combined with an extended dictionary presents better performance than the approaches based on either the sentiment dictionary or the neural network alone [16]. Luo proposed to combine the neural network with Latent Dirichlet Allocation (LDA) model for network text sentiment classification in [17]. Gu et al. [18] analyzed the text sentiment of commodity evaluation, combining the neural network model with the semantic rules, context, and other factors. The improved artificial neural network (ANN) also has great advantages in research such as prediction [19].

There are also many other methods of extracting terms for the specific sentiment classification [20]. Sheik et al. [21] proposed the sentinel method to establish the sentiment circle through the Cartesian coordinate system for the sentiment classification of sentences. Alzubi et al. [22] presented a Collaborative Adversarial Network (CAN) model for paraphrase identification. Bel'tyukov et al. [23] utilized logical formulas and logical reasoning to achieve emotional analysis. By fuzziness of sentiment words, Phan

et al. [24] obtained the fuzzy embedding feature and effectively improved the F_1 value of Twitter sentiment analysis. Vashishtha et al. [25] defuzzified the output to achieve sentiment analysis, based on the sentiment intensity fuzzification and fuzzy inference rules.

3. Materials and Method

3.1. Establishment of Exclusive Fuzzy Dictionary for Microblog. The comprehensive sentiment dictionary consists of seven separate dictionaries, among which the fuzzy dictionary is specifically established for microblogs and other dictionaries have general applicability. The crawled microblog text data is divided into two parts. One is for establishing the exclusive fuzzy dictionary while the other is for data testing. A total of 45,344 items are used to establish the fuzzy dictionary. The statistics for the data are detailed in Figure 1. The positive rate, the neutral rate, and the negative rate are 27%, 29%, and 44%, respectively, which strikes a balance on different emotions.

The crawled data is preprocessed with word segmentation, of which results are used as the input of the TextRank algorithm. The TextRank is a graph-based ranking algorithm. It constructs the network according to the adjacent relationship among words and iteratively calculates the ranking score (i.e., importance) of each node (i.e., word) [26]. The importance of the same word in a post is accumulated. The fuzzy dictionary is finally formed by removing the repetitive words. Table 1 details the pseudocode of the algorithm, where w represents the importance of a word and NUM denotes the total number of the posts.

3.2. Feature Extraction and Processing of Multimodal Data

3.2.1. Feature Extraction of Blog Images. Based on the intermediate features of images, this paper uses the visual-to-text conversion model proposed by Google in 2015 to extract image features in the CNN encoder. The extracted features are input into the LSTM decoder to obtain the text description feature of the image [27]. To adapt to the sentiment classification, the text sentiment is expanded by two sentiment knowledge maps, namely, SentiBank and SentiBridge [28], as shown in Figure 2. In addition, before the feature extraction of the image description, the text features and face features contained in the image are extracted first. The related techniques for extracting the image matching and the facial expression have been studied well. In this paper, the API interface is provided by Baidu to realize the extraction of two image features.

3.2.2. Video Features Extraction. Video features consist of image and audio features. The image feature extraction in the video can follow the method discussed in Section 3.2.1. However, no matter how many seconds the video lasts and how many frames the video has, it can capture a huge number of still images from videos, and most of the obtained images are repetitive with each other. As a result, we filter the captured images with two steps before the image feature extraction starts.

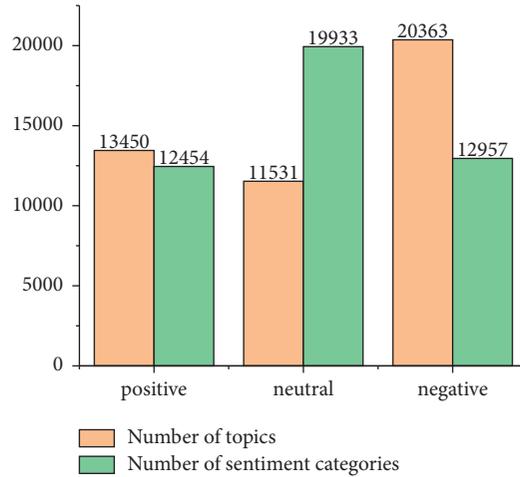


FIGURE 1: The statistics of the crawled microblog text data.

TABLE 1: The algorithm for forming a fuzzy exclusive dictionary.

Input: Blog text T
Output: Fuzzy sentiment dictionary W

Process:

- (1) IF t in T [i]:
- (2) t_new [i] = jieba (t)
- (3) END if
- (4) FOR word in t_new [i]:
- (5) w [i] = TextRank (t_new [i])
- (6) IF word [i] = word [j] and word [i], word [j] both in t_new [i]:
- (7) k [i] = w [i] + w [j]
- (8) END if
- (9) IF $j \leq 5000$:
- (10) FW = k [j] in sort (k [i])
- (11) END IF
- (12) IF word in FW:
- (13) pos = num (word in positive)/NUM
- (14) neg = num (word in negative)/NUM
- (15) other = 1 - pos - neg
- (16) END if
- (17) Remove the common and incorrect words from FW to form a final fuzzy dictionary W with 593 words, as shown in Table 2
- (18) Print W

TABLE 2: The fuzzy dictionary.

Examples	Positive degree	Negative degree	Uncertainty	Importance
Happy	0.604672	0.225517	0.169811	51.05354
Joyful	0.685851	0.111751	0.202398	47.92227
Unwell	0.080226	0.525424	0.39435	6.822836
...
Care	0.5	0	0.5	0.034836
Tired	0	1	0	0.031808
Boredom	0.071429	0.5	0.428571	0.027837

Firstly, a YOLOv3 model is used to filter images containing similar objects. The YOLOv3 model utilizes Darknet53 as the main network. It can identify the subjects and the number of objects in images, especially the small and medium objects [29]. After the images are selected by the YOLOv3 model, the differences among the

rest of the images have little effect on the sentiment classification results. To this end, the second round of image selection is required. The Scale Invariant Feature Transform (SIFT) algorithm is adopted to detect the key points of the images, and the local similarity among the images is subsequently calculated [30]. The two-step

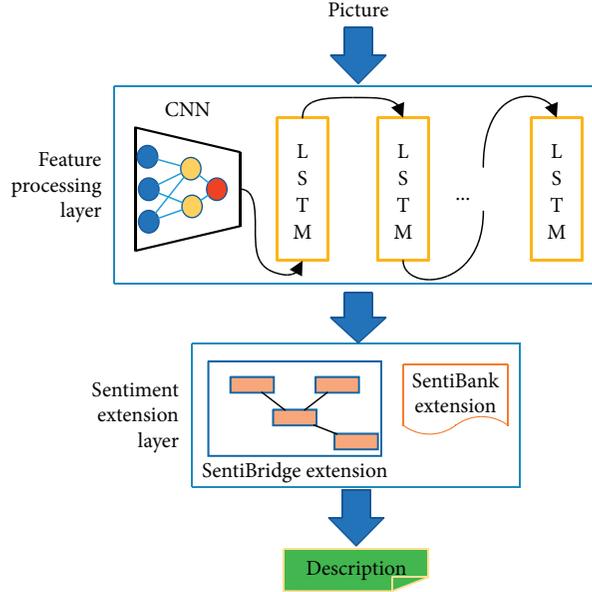


FIGURE 2: The visual-text feature transformation model for blog images.

image selection can greatly reduce the number of images for feature extraction, and thus the efficiency of image processing is improved.

For the audio feature extraction, the Mel-frequency cepstral coefficient (MFCC) is regarded as the most representative feature in literature [31]. In addition to the MFCC, another two audio features are considered in this paper: zero-crossing rate and spectral centroid. Specifically, the zero-crossing rate represents the number of zero-crossings in the signal spectrum, while the spectrum centroid denotes the texture of the audio. The two features can assist MFCC features with distinguishing the sentiment of audio. A 22-dimensional vector is introduced to represent the three considered audio features, by which an optimal random forest model is trained for the sentiment classification of video audio.

3.3. The Ambiguity of the Sentiment Intensity. In this paper, fuzzy inference rules are used to fuzzify the sentiment intensity. Dual hesitant fuzzy is introduced for the sentiment intensity, and general fuzzy is considered as the output of the reasoning model.

3.3.1. Dual Hesitant Fuzzy Sets

Definition 1. Let X be a fixed set; then, a dual hesitant fuzzy set (DHF) D on X is described as [32]

$$D = \{ \langle x, h(x), g(x) \rangle | x \in X \}, \quad (1)$$

in which $h(x)$ and $g(x)$ are two sets of some values in $[0, 1]$, denoting the possible membership degrees and nonmembership degrees of the element $x \in X$ to the set D , respectively, with the conditions: $0 \leq \gamma, \eta \leq 1, 0 \leq \gamma^+ + \eta^+ \leq 1$, where

$\gamma \in h(x), \eta \in g(x)$. Note that this paper only deals with the dual hesitant fuzzification when the number of γ, η is equal to 1, respectively. Moreover, π_D denotes the uncertainty of the element x belonging to D in X , which is called the swing degree in this paper. Then, we have

$$\pi_D = 1 - \gamma - \eta. \quad (2)$$

3.3.2. Sentiment Value Calculation. In the fuzzy dictionary, the sentiment value is twice the strong ambiguity and equal to the weak ambiguity, which is calculated by pos (positive) and neg (negative). For the basic sentiment dictionary, the sentiment value is determined by pos_b and neg_b . Degree adverbs dictionary and negative words dictionary are the strengthening and weakening of sentiment intensity, of which values are denoted as C and N , respectively. Moreover, the text sentiment is calculated using E_pos and E_neg . The facial sentiment is obtained by F_pos and F_neg , while the speech sentiment is determined by A_pos and A_neg . Table 3 presents the calculations of sentiment values in different dictionaries, where k is the existence coefficient, namely, the number of the corresponding sentiment units. k is set to 0 if it does not exist. n and m in Table 3 represent the number of sentiment words with even and odd negative words in the text, respectively. S_{pos} and S_{neg} represent the sum of positive sentiment value and the sum of negative sentiment value calculated by fuzzy dictionary and basic dictionary, respectively. To simplify the determination of the membership function, the positive and negative sentiment values are unified. The total sentiment value of the blog is obtained according to the text and video sentiment values.

TABLE 3: The equations for the sentiment value calculation.

	Positive	Negative
Fuzzy dictionary	$\sum_{i=1}^n (k_1 * C_r * \text{pos}) + \sum_{j=1}^m \sum_{j=1}^m [k_2 * C_r * \text{neg} * (-1)^N]$	$\sum_{i=1}^n (k_1 * C_r * \text{neg}) + \sum_{j=1}^m [k_2 * C_r * \text{pos} * (-1)^N]$
Basic dictionary	$\sum_{i=1}^n (k_1 * C_r * \text{pos}_b) + \sum_{j=1}^m [k_2 * C_r * \text{neg}_b * (-1)^N]$	$-\sum_{i=1}^n (k_1 * C_r * \text{neg}_b) + \sum_{j=1}^m [k_2 * C_r * \text{pos}_b * (-1)^N]$
Text sentiment value	$\sum_{i=0}^2 S_{\text{pos}_i} + \sum_{j=0}^n (k_3 * E_{\text{pos}_i})$	$\sum_{i=0}^2 S_{\text{neg}_i} + \sum_{j=0}^n (k_3 * E_{\text{neg}_i})$
Video sentiment value	$\sum_{i=1}^n (k_7 * F_{\text{pos}_i} * P_i) + k_8 * A_{\text{pos}}$	$-\sum_{i=1}^n (k_9 * F_{\text{neg}_i} * P_i) + k_8 * A_{\text{neg}}$

$$F(x) = \begin{cases} 0, & 0 \leq x \leq 2.5, \\ \frac{x-2.5}{3}, & 2.5 < x \leq 5.5, \\ \frac{8.5-x}{3}, & 5.5 < x \leq 8.5, \\ 0, & x > 8.5, \end{cases} \quad (3)$$

$$F(x) = \begin{cases} 1, & 0 \leq x \leq 2, \\ \frac{5.5-x}{3.5}, & 2 < x \leq 5.5, \\ \frac{x-5.5}{4.5}, & 5.5 < x \leq 10, \\ 1, & x > 10, \end{cases} \quad (4)$$

$$F(x) = \begin{cases} -\frac{1}{8}x + 1, & 0 \leq x \leq 8, \\ 0, & x > 8, \end{cases} \quad (4)$$

$$F(x) = \begin{cases} \frac{1}{100}x^2, & 0 \leq x \leq 10, \\ 1, & x > 10, \end{cases} \quad (4)$$

$$F(x) = \begin{cases} 0, & 0 \leq x \leq 3, \\ \frac{1}{49}x^2 - \frac{6}{49}x + \frac{9}{49}, & 3 < x \leq 10, \\ 1, & x > 10, \end{cases} \quad (5)$$

$$F(x) = \begin{cases} 1, & 0 \leq x \leq 2, \\ -\frac{1}{36}x^2 + \frac{x}{9} + \frac{8}{9}, & 2 < x \leq 8, \\ 0, & x > 8, \end{cases} \quad (5)$$

3.3.3. Membership Function and Nonmembership Function. The sentiment intensity is defined by three different levels, that is, low, middle, and high. As mentioned above, we use dual

hesitant fuzzy sets to fuzzify the sentiment intensity of posts. In this way, the membership and nonmembership functions are corresponded by the positive and negative sentiments for the three-level sentiment intensity, respectively. And they meet the conditions in Section 3.3.1. Specifically, the formula above and the formula below of equations (3)–(5) show the membership and nonmembership functions for the sentiment intensity with middle, low, and high levels, respectively.

3.4. MD-HFCE Sentiment Classification Method

3.4.1. Improved Mamdani Fuzzy Inference Model. Mamdani fuzzy reasoning model is a fuzzy reasoning model based on the IF-THEN rule proposed by Mamdani and Assilian in 1975. It conducts the fuzzy reasoning through the input IF conditions and calculates the membership degree of the final result [33]. Considering the situation of multimodal blog sentiment classification in reality, the fuzzy inference rules on sentiment intensity are given as follows:

IF R_p is r_i and R_n is r_j THEN S is s_m ,

where R_p and R_n represent the membership degrees of the positive and negative sentiment intensity, respectively. r represents the fuzzy set of sentiment intensity, that is, [low-level, middle-level, high-level]. S denotes the last membership degree of multimodal blogs. s is the fuzzy set of blog sentiment classification, that is, [low-positive, strong-positive, neutral, low-negative, strong-negative]. Table 4 elaborates the nine fuzzy inference rules of the Mamdani model. To simplify the calculation, after the membership degree of each category is determined, the neutral membership degree is obtained by comparisons among different categories. Therefore, in the membership function of the Mamdani output, $y = 0$ is the intermediate membership function of the model output results, as shown in Figure 3.

3.4.2. The Improved Fuzzy Comprehensive Evaluation Model. A fuzzy comprehensive evaluation model is a comprehensive fuzzy method based on fuzzy mathematics. It can transform qualitative evaluation into quantitative evaluation using the membership degree theory of fuzzy mathematics. According to the classification characteristics of the sentiment intensity, the two-level fuzzy comprehensive evaluation model is utilized, which includes five steps:

- (1) Establish the two-level factor set: the primary factor set U meets $U = [U_1, U_2, U_3, U_4] = [\text{positive intensity, negative intensity, negatively positive intensity, negatively negative intensity}]$, while the secondary factor set is denoted by $U_1 = U_2 = [\text{low, middle, high}]$ and

TABLE 4: The nine specific Mamdani fuzzy rules.

Rule	Category (positive)	Category (negative)	Category (blog)
r_1	Low	Low	Neutral
r_2	Low	Medium	Low-negative
r_3	Low	High	Strong-negative
r_4	Medium	Low	Low-positive
r_5	Medium	Medium	Neutral
r_6	Medium	High	Low-negative
r_7	High	Low	Strong-positive
r_8	High	Medium	Low-positive
r_9	High	High	Neutral

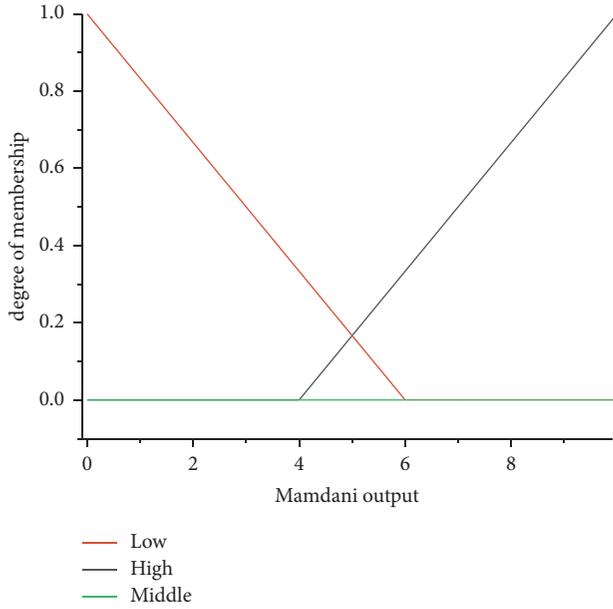


FIGURE 3: The model outputs with various membership functions.

$U_3 = U_4 = [\text{nonlow}, \text{nonmiddle}, \text{nonhigh}]$. That is, the positive intensity can be divided into low positive, middle positive, and high positive. Nonpositive intensity can be divided into nonlow positive, nonmiddle positive, and nonhigh positive. The same goes for negative and nonnegative.

- (2) Establish the two-level weight set: the weight set is the numerical reflection of the degree of influence of various factors on the classification result. The positively positive and negative factor in the primary factor can have a larger influence than the negatively positive and negative factor on the sentiment classification. As a result, the weight set of the primary factor A_1 is set to $A_1 = [0.3, 0.3, 0.2, 0.2]$, and the weight set of the secondary sentiment factor A_2 is determined by its own swing degree, namely, π_D in (2) in Section 3.3.1. The larger the swing degree is, the smaller the influence proportion is, that is,

$$A_2 = 1 - \pi_D. \quad (6)$$

- (3) Establish the evaluation set V : the sentiment classification results of multimodal blogs are the final evaluation results, namely, $V = [\text{low-positive},$

$\text{strong-positive}, \text{neutral}, \text{low-negative}, \text{strong-negative}]$.

- (4) Conduct first-level fuzzy comprehensive evaluation: from the first-level fuzzy comprehensive evaluation of the second-level factors, the preliminary evaluation set j is obtained by

$$j_q = A_2^q u_q, \quad (7)$$

where u_q is the fuzzy possibility matrix of the relevant evaluation set corresponding to the secondary factor set. The matrix is calculated by Mamdani fuzzy reasoning, which is presented by the following.

- (5) Make the second-level fuzzy comprehensive evaluation: the comprehensive evaluation set J is composed of the preliminary evaluation sets j obtained by the first-level fuzzy comprehensive evaluation, by which the final evaluation result set is calculated. The possibility of the neutral evaluation is modified, and the sentiment classification results of multimodal blogs are determined according to the principle of the maximum membership, that is,

$$B = A_1 J = A_1 \begin{bmatrix} j_1 \\ j_2 \\ j_3 \\ j_4 \end{bmatrix}. \quad (8)$$

Equations (9)–(15) are the process of preliminarily calculating the fuzzy set of the category of blog post sentiment through Mamdani, which is the basis for further calculation of the fuzzy possibility matrix u_q . In, pos_i and neg_i correspond to the membership of positive and negative sentiment intensity of rules in Table 4, respectively. SC_{r_i} is the final sentiment classification in Table 4. W_{r_i} is the fuzzy set of the positive factors U_1 and U_2 , and u_1, u_2 are the fuzzy possibility matrix sets formed by W_{r_i} . $N\text{pos}_l, N\text{pos}_m, N\text{pos}_h, N\text{neg}_l, N\text{neg}_m,$ and $N\text{neg}_h$ represent the positive and negative membership with nonlow, nonmiddle, and nonhigh sentiment intensity, respectively. $\text{pos}_l, \text{pos}_m, \text{pos}_h, \text{neg}_l, \text{neg}_m,$ and neg_h represent the positive and negative subordination degrees with low-level, middle-level, and high-level sentiment intensity, respectively. $\langle W_{mN_plow}, W_{hN_plow} \rangle,$ $\langle W_{lN_pmid}, W_{hN_pmid} \rangle,$ and $\langle W_{lN_phig}, W_{mN_phig} \rangle$ are

the fuzzy sets of nonlow, nonmiddle, and nonhigh positive sentiment membership. $\langle W_{mN_nlow_i}, W_{hN_nlow_i} \rangle$, $\langle W_{IN_nmid_i}, W_{hN_nmid_i} \rangle$, and $\langle W_{IN_nhig_i}, W_{mN_nhig_i} \rangle$ denote

the nonlow, nonmiddle, and nonhigh negative sentiments belonging to the fuzzy sets. Among them, i represents the order of the rules in Table 4:

$$W_{r_i} = (\text{pos}_{r_i} \wedge \text{neg}_{r_i}) \wedge SC_{r_i}, \quad (9)$$

$$\langle W_{mN_plow_i}, W_{hN_plow_i} \rangle = \langle \langle (N\text{pos}_l \vee \text{pos}_{m_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{pos}_l \vee \text{pos}_{h_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (10)$$

$$\langle W_{IN_pmid_i}, W_{hN_pmid_i} \rangle = \langle \langle (N\text{pos}_m \vee \text{pos}_{l_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{pos}_m \vee \text{pos}_{h_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (11)$$

$$\langle W_{IN_phig_i}, W_{mN_phig_i} \rangle = \langle \langle (N\text{pos}_h \vee \text{pos}_{l_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{pos}_h \vee \text{pos}_{m_{r_i}}) \wedge \text{neg}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (12)$$

$$\langle W_{mN_nlow_i}, W_{hN_nlow_i} \rangle = \langle \langle (N\text{neg}_l \vee \text{neg}_{m_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{neg}_l \vee \text{neg}_{h_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (13)$$

$$\langle W_{IN_nmid_i}, W_{hN_nmid_i} \rangle = \langle \langle (N\text{neg}_m \vee \text{neg}_{l_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{neg}_l \vee \text{neg}_{h_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (14)$$

$$\langle W_{IN_nhig_i}, W_{mN_nhig_i} \rangle = \langle \langle (N\text{neg}_h \vee \text{neg}_{l_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i}, \langle (N\text{neg}_h \vee \text{neg}_{m_{r_i}}) \wedge \text{pos}_{r_i} \rangle \wedge SC_{r_i} \rangle, \quad (15)$$

$$u_1 = \begin{bmatrix} 0 & 0 & \overline{w_{r_1}} & \overline{w_{r_2}} & \overline{w_{r_3}} \\ \overline{w_{r_4}} & 0 & \overline{w_{r_5}} & \overline{w_{r_6}} & 0 \\ \overline{w_{r_8}} & \overline{w_{r_7}} & \overline{w_{r_9}} & 0 & 0 \end{bmatrix}, \quad (16)$$

$$u_2 = \begin{bmatrix} \overline{w_{r_4}} & \overline{w_{r_7}} & \overline{w_{r_1}} & 0 & 0 \\ \overline{w_{r_8}} & 0 & \overline{w_{r_5}} & \overline{w_{r_2}} & 0 \\ 0 & 0 & \overline{w_{r_9}} & \overline{w_{r_6}} & \overline{w_{r_3}} \end{bmatrix}, \quad (17)$$

$$u_3 = \begin{bmatrix} \overline{\overline{w_{mN_plow_4} + w_{hN_plow_8}}} & \overline{w_{hN_pmid_8}} & \overline{w_{mN_phig_4}} \\ \overline{w_{hN_plow_7}} & \overline{w_{hN_pmid_7}} & 0 \\ \overline{\overline{w_{mN_plow_5} + w_{hN_plow_9}}} & \overline{\overline{w_{IN_pmid_1} + w_{hN_pmid_9}}} & \overline{\overline{w_{IN_phig_1} + w_{mN_phig_5}}} \\ \overline{w_{hN_plow_6}} & \overline{w_{IN_pmid_2}} & \overline{\overline{w_{IN_phig_2} + w_{mN_phig_6}}} \\ 0 & \overline{w_{IN_pmid_3}} & \overline{w_{IN_phig_3}} \end{bmatrix}^T, \quad (18)$$

$$u_4 = \begin{bmatrix} \overline{w_{mN_nlow_8}} & \overline{w_{IN_nmid_4}} & \overline{\overline{w_{IN_nhig_4} + w_{mN_nhig_8}}} \\ 0 & \overline{w_{IN_nmid_7}} & \overline{w_{IN_nhig_7}} \\ \overline{\overline{w_{mN_nlow_5} + w_{hN_nlow_9}}} & \overline{\overline{w_{IN_nmid_1} + w_{hN_nmid_9}}} & \overline{\overline{w_{IN_nhig_1} + w_{mN_nhig_5}}} \\ \overline{\overline{w_{mN_nlow_2} + w_{hN_nlow_6}}} & \overline{w_{hN_nmid_6}} & \overline{w_{mN_nhig_2}} \\ \overline{w_{hN_nlow_3}} & \overline{w_{hN_nmid_3}} & 0 \end{bmatrix}^T. \quad (19)$$

3.5. Application. The first step is to select the monitoring topic, crawl the multimodal blogs under the selected topic in real time, conduct the above-mentioned sentiment analysis on the crawled blogs, and record the classification results. Subsequently, the real-time classification results are analyzed to form the trend curve of the topic sentiment, and the negative sentiment intensity is obtained. Eventually, we can determine whether the intensity of the negative sentiment exceeds the threshold. If the threshold is exceeded, the

negative alerts notify the software to adjust the topic sentiment of public opinions in time.

4. Results and Discussions

4.1. Experimental Dataset. The video-image filtering is based on the COCO image dataset, which has been widely used in object detection, target segmentation, subtitle generation, and other aspects. The dataset has about 330,000 images and

is marked by Microsoft in 2014 [34]. Flickr30K image dataset is used to generate image description features. It contains 30,000 images and each image is marked with five text descriptions [35]. Moreover, for video-audio sentiment classification, 10,500 Chinese items are selected from the emotional speech dataset (ESD) dataset, which owes a large number of English and Chinese sentiment speech [36]. Furthermore, 10,000 posts with multiple topics are crawled in microblog for experimental evaluation, including 4562 positive posts, 3636 negative posts, 1802 neutral posts, 4374 images, and 412 videos.

4.2. Experimental Implementation. In this section, we take the post shown in Figure 4 as an example to show the steps of experiment implementation. The post is composed of three elements, that is, texts, emoticons, and videos, to imply the user's sentiment. Specifically, a total of 41 images are extracted from the video. Subsequently, following the image selection scheme, six of the extracted images are filtered in the preliminary selection, and two of the six images remain in the second round. The two images do not present the text and facial features, and their description features are similar to each other. Finally, the description feature is synthesized with sentiment expansion; that is, a sad dog is lying in a bright bathroom. Furthermore, the audio is analyzed by a random forest algorithm. It can be seen from Table 5 that the positive sentiment value of the multimodal blog is 5.69, and the negative sentiment value is 6.08.

Figure 5 shows the membership degree and swing degree of positive, nonpositive, negative, and nonnegative sentiment intensity for Figure 4. Figure 6 is the results of the u_q calculation. Figure 7 presents the fuzzy comprehensive evaluation results, which can be obtained by the maximum membership degree scheme, the weighted average method, the F distribution method, and other methods. Among them, the maximum membership degree method is the simplest and the most popular method. Therefore, this paper uses the maximum membership degree method to deal with the evaluation results. The final sentiment of the blog is classified as low-level negative. As shown in Figure 7, the low-level negative membership degree has the highest value. Moreover, if the difference between the low-level negative and positive membership degrees is less than the threshold, the blog is regarded as a neutral category. It means the membership degree is 0 and vice versa. Through multiple experiments, it can achieve the best performance if the threshold is set to 0.003.

The content of the blog post: "Very lucky, the kidney of Zheng does not fail, and it has been out of danger. My friend and I are going back to Litang first, as Bubu is waiting for us. A friend in Chengdu will help us to take care of it. It looked aggrieved and unhappy when we said goodbye. The poor dog really suffers but is very tough."

4.3. Experimental Comparisons Based on Different Dictionaries. The proposed comprehensive dictionary is compared with HowNet [37] and National Taiwan



FIGURE 4: The example of a blog post.

University Sentiment Dictionary (NTUSD) [38] in Table 6. As shown in Table 7, HowNet contains 38 propositions, while NTUSD only includes sentiment words. It can be seen from Table 6 that the proposed comprehensive fuzzy dictionary can outperform the HowNet and NTUSD. In addition, the recall rate and F_1 value of HowNet are higher than those of NTUSD for both positive and neutral blogs. However, in terms of the negative blogs, the performance of NTUSD is slightly better than that of HowNet, since the number of the negative words is small.

4.4. Comparison of Methods. To verify the proposed method, four baseline tests are conducted as follows:

- (1) Extended dictionary + semantic rules [2], which combines the extended dictionary with semantic rules, hereinafter referred to as Method 1.
- (2) Improved multilayer CNN network [13], which increases the number of convolution layers, hereinafter referred to as Method 2.
- (3) Dictionary + fuzzification, which integrates the sentiment dictionary with general fuzzification and is referred to as Method 3.
- (4) Fuzzy rules + Mamdani fuzzy reasoning [23], where Mamdani fuzzy reasoning is utilized and the scheme is referred to as Method 4.

Figures 8–11 show all comparative experiments and the experimental results of the method in this paper on the correct rate, recall rate, F_1 value and average results of positive, neutral and negative blog posts. The performance of method 3 is the most unsatisfactory. It can be seen that the general fuzzy rules have too much freedom and cannot improve the classification effect very well. Method 4 with the related fuzzy rules further improves the correctness and stability of the fuzzy methods and it almost reaches the effect of the traditional method. The method in this paper introduces hesitating fuzzy and fuzzy evaluation rules on the basis of the basic fuzzy method, so that the fuzzy limit is well controlled. The experimental results also show that the method in this paper performs best among several methods, and the classification effect is better improved. And the classification result is more stable than method 2 with better accuracy.

TABLE 5: The calculation results of blog sentiment.

Blog	Presence	Positive	Negative
Text	Y	4.59	3.54
Emoji	Y	0	0.6
Image	N	0	0
Video-image	Y	1.1	0.94
Video-audio	Y	0	1

Membership of positive sentiment intensity (U1):
 [0.2887499999999995, 0.9366666666666665, 0.147675510204081571
 Membership of non-positive sentiment intensity (U3):
 [0.323761, 0.04222222222222231, 0.61775]
 uncertainty:
 [0.38748900000000003, 0.02111111111111157, 0.2305494897959185]
 Membership of negative sentiment intensity (U2):
 [0.24, 0.8066666666666666, 0.1935999999999999]
 Membership of non-negative sentiment intensity (U4):
 [0.369664, 0.12888888888888891, 0.5376]
 uncertainty:
 [0.390336, 0.06444444444444444, 0.26880000000000015]

FIGURE 5: The degree of the strength membership.

```

u1:
[[0.      0.      0.      0.32275  0.2276   ]
 [0.274   0.      0.      0.2276   0.        ]
 [0.17721061 0.17721061 0.      0.        0.        ]]
u2:
[[0.274   0.17721061 0.      0.        0.        ]
 [0.17721061 0.      0.      0.32275  0.        ]
 [0.        0.        0.      0.2276   0.2276   ]]
u3:
[[0.3158805 0.274   0.      0.2276   0.        ]
 [0.17721061 0.17721061 0.      0.32275  0.2276   ]
 [0.274   0.      0.      0.4003325 0.2276   ]]
u4:
[[0.17721061 0.      0.      0.3592406 0.32275  ]
 [0.274   0.17721061 0.      0.2276   0.2276   ]
 [0.34988531 0.17721061 0.      0.32275  0.        ]]
    
```

FIGURE 6: The calculation results of u_q .

result:
 low-positive: 0.19871060468052376
 strong-positive: 0.08642821779112575
 Neutral: 0
 low-negative: 0.23851034787725028
 strong-negative: 0.10935218089616358

FIGURE 7: The comprehensive evaluation results.

TABLE 6: Comparison of dictionary experiment results.

	Precision			Recall			F_1		
	NTUSD	HowNet	Ours	NTUSD	HowNet	Ours	NTUSD	HowNet	Ours
Positive blog	0.711	0.767	0.857	0.747	0.76	0.812	0.729	0.764	0.834
Neutral blog	0.694	0.667	0.782	0.548	0.631	0.768	0.612	0.649	0.775
Negative blog	0.799	0.752	0.827	0.73	0.757	0.791	0.763	0.754	0.801

Bold values intuitively compare and emphasize the effectiveness of the fuzzy dictionary established in this article.

TABLE 7: The specific conditions of different dictionaries.

Category	HowNet	NTUSD
Positive words	4566	2812
Negative words	4370	8278
Degree adverbs	219	0
Advocate words	38	0

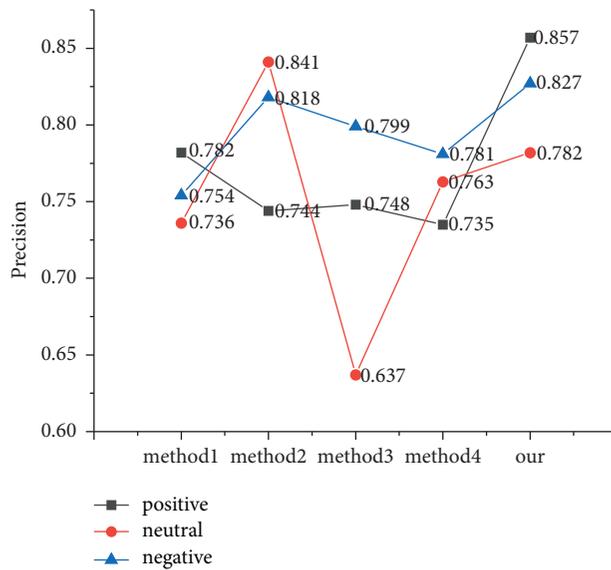


FIGURE 8: The precision comparisons of different methods.

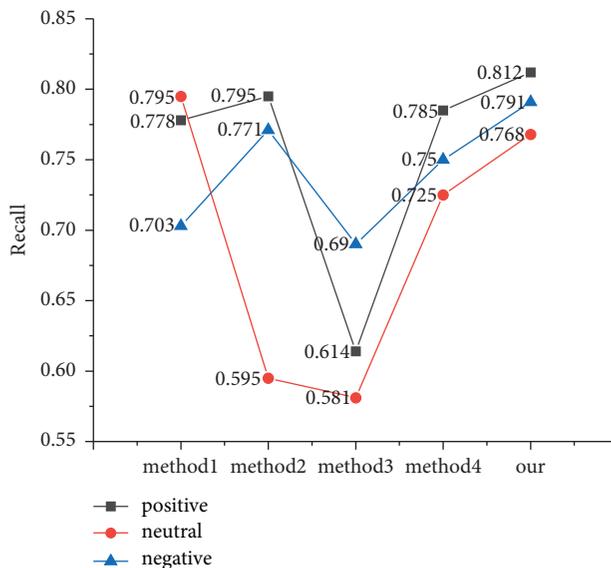


FIGURE 9: The recall rate comparisons of different methods.

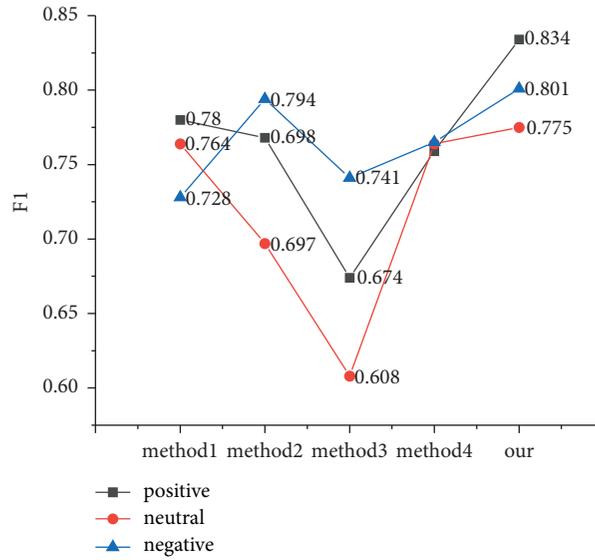


FIGURE 10: The F_1 value comparisons of different methods.

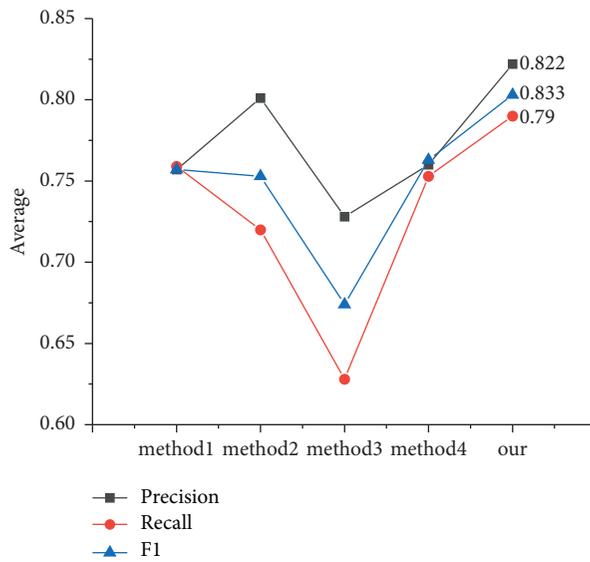


FIGURE 11: The average value comparisons of different methods.

4.5. *Application Experiment.* In this section, the topic “#teachers discriminate students after comparing the income of their parents #” is studied for the sentiment orientation experiment. The blog posts under this topic are crawled every half hour, which obtains 253, 644, 1105, and 1512 blog posts in each crawling. Firstly, each crawled post is processed by the sentiment classification, and the proportions of the positive blogs and negative blogs are obtained, respectively. Then, the proportions of low-level negative blogs and strong-negative blogs are calculated among the negative blogs. In this way, the proportion of negative sentiment blogs is monitored under

the topic, which can decide whether the intervention is required. As shown in Figure 12, the number of positive sentiment blogs under the topic is not more than 12%. Although varying over time, the number of low-negative blogs is much larger than that of strong-negative blogs at each period. It implies that their sentiment intensity is not very strong, even though people have relatively negative attitudes. Moreover, the sentiment intensity is not strengthened over time and maintains the same level. As a result, the app administrator can continue to supervise and analyze the posts of this topic without any intervention.

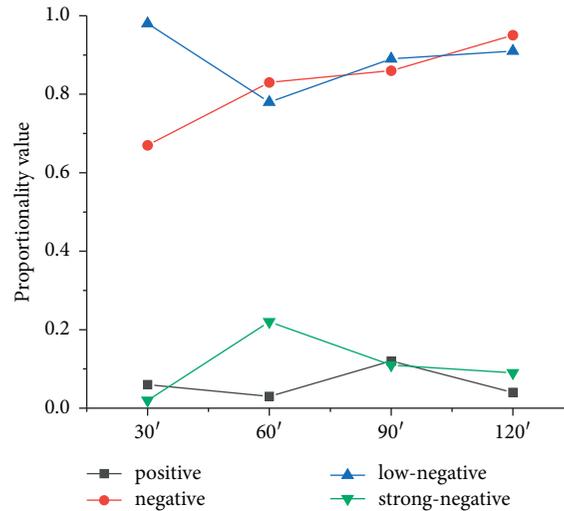


FIGURE 12: The proportions of different sentiment types in topic blogs.

5. Conclusions

This paper proposes an improved sentiment classification method MD-HFCE based on fuzzy improvement. The method normalizes multimodal features and standardizes fuzzification. The main task of this research is to establish a special Weibo fuzzy dictionary and the design of dual hesitation fuzzy inference classification rules. Experiments have verified that the method in this paper has achieved good results.

However, there are still many areas that need to be improved. First of all, the classification results of neutral blog posts are not ideal, and the model output membership function and determination threshold need to be further optimized and improved. Secondly, the emotion classification is single and not detailed enough.

In the future, after the method in this paper is further improved, it can be combined with neural networks and knowledge graphs to form a hesitant fuzzy network for direct conversion of fuzzy information and target recognition; in addition, the method combined with sentiment analysis can also be used in social software community discovery, emotional robot dialogue, and so forth.

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.

Acknowledgments

This work is supported by the Natural Science Foundation of Hebei Province, China (No. F2019201329).

References

- [1] J. Yuan, Y. Wu, X. Lu, Y. Zhao, B. Qin, and T. Liu, "Recent advances in deep learning based sentiment analysis," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1947–1970, 2020.
- [2] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, "Chinese text sentiment analysis based on extended sentiment dictionary," *IEEE Access*, vol. 7, pp. 43749–43762, 2019.
- [3] M. Ahmed, Q. Chen, and Z. Li, "Constructing domain-dependent sentiment dictionary for sentiment analysis," *Neural Computing and Applications*, vol. 32, no. 18, pp. 14719–14732, 2020.
- [4] Y. Wang, G. Huang, J. Li, H. Li, Y. Zhou, and H. Jiang, "Refined global word embeddings based on sentiment concept for sentiment analysis," *IEEE Access*, vol. 9, pp. 37075–37085, 2021.
- [5] Y. Wang, F. Yin, J. Liu, and M. Tosato, "Automatic construction of domain sentiment lexicon for semantic disambiguation," *Multimedia Tools and Applications*, vol. 79, no. 31–32, pp. 22355–22373, 2020.
- [6] D. Borth, R. Ji, T. Chen, T. Breuel, and S. F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *Proceedings of the [ACM Press the 21st ACM International Conference]*, vol. 13, pp. 223–232, Barcelona, Spain, October 2013.
- [7] T. Mike, B. Kevan, P. Georgios, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.
- [8] H. Liang, U. Ganeshbabu, and T. Thorne, "A dynamic Bayesian network approach for analysing topic-sentiment evolution," *IEEE Access*, vol. 8, pp. 54164–54174, 2020.
- [9] J. Yuan, Q. You, S. McDonough, and J. Luo, "Sentribute: image sentiment analysis from a mid-level perspective," *WIS-DOM'13: Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, vol. 10, pp. 1–8, 2013.
- [10] W. Chen, Z. Xu, X. Zheng, Q. Yu, and Y. Luo, "Research on sentiment classification of online travel review text," *Applied Sciences*, vol. 10, no. 15, p. 5275, 2020.

- [11] S. Sharmin and D. Chakma, "Attention-based convolutional neural network for Bangla sentiment analysis," *AI and Society*, vol. 36, no. 1, pp. 381–396, 2020.
- [12] Y.-J. Su, W.-C. Hu, J.-H. Jiang, and R.-Y. Su, "A novel LMAEB-CNN model for Chinese microblog sentiment analysis," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 9127–9141, 2020.
- [13] H. Kim and Y.-S. Jeong, "Sentiment classification using convolutional neural networks," *Applied Sciences*, vol. 9, no. 11, p. 2347, 2019.
- [14] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis," *Cognitive Computation*, vol. 10, no. 4, pp. 639–650, 2018.
- [15] N. K. Senthil Kumar and N. Malarvizhi, "Bi-directional LSTM-CNN combined method for sentiment analysis in part of speech tagging (PoS)," *International Journal of Speech Technology*, vol. 23, no. 2, pp. 373–380, 2020.
- [16] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: sentiment analysis using lexicon integrated two-channel CNN-LSTM family models," *Applied Soft Computing*, vol. 94, Article ID 106435, 2020.
- [17] L.-x. Luo, "Network text sentiment analysis method combining LDA text representation and GRU-CNN," *Personal and Ubiquitous Computing*, vol. 23, no. 3-4, pp. 405–412, 2019.
- [18] T. Gu, G. Xu, and J. Luo, "Sentiment analysis via deep multichannel neural networks with v information bottleneck," *IEEE Access*, vol. 8, pp. 121014–121021, 2020.
- [19] A. Movassagh, A. Alzubi, M. Gheisari et al., "Artificial neural networks training algorithm integrating invasive weed optimization with differential evolutionary model," *Journal of Ambient Intelligence and Humanized Computing*, no. 1, 2021.
- [20] X. Tan, Y. Cai, J. Xu, H.-F. Leung, W. Chen, and Q. Li, "Improving aspect-based sentiment analysis via aligning aspect embedding," *Neurocomputing*, vol. 383, pp. 336–347, 2020.
- [21] R. Sheik, S. Philip, A. Sajeev, S. Sirinivasan, and G. Jose, "Entity level contextual sentiment detection of topic sensitive influential twitterers using SentiCircles," *Data Engineering and Intelligent Computing*, vol. 542, pp. 207–216, 2018.
- [22] A. Alzubi, R. Jain, A. Kathuria, A. Khandelwal, and A. Saxena, "Paraphrase identification using collaborative adversarial networks," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 20, pp. 1–12, 2020.
- [23] A. Bel'tyukov and M. Abbasi, "Logical analysis of emotions in text from natural language," *Vestnik Udmurtskogo Universiteta-Matematika Mekhanika Kompyuternye Nauki*, vol. 29, no. 1, pp. 106–116, 2019.
- [24] H. T. Phan, V. C. Tran, N. T. Nguyen, and D. Hwang, "Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model," *IEEE Access*, vol. 8, pp. 14630–14641, 2020.
- [25] V. Srishti and S. Seba, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, Article ID 112834, 2019.
- [26] R. Mihalcea and P. Tarau, "TextRank: bringing order into texts," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 404–411, Doha, Qatar, October 2014.
- [27] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: lessons learned from the 2015 MSCOCO image captioning challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 652–663, 2016.
- [28] Q. Lu and W. Chen, "Automatically building a large scale dictionary of Chinese entity sentiment expressions," *Journal of Chinese Information Processing*, vol. 32, no. 8, pp. 32–41, 2018.
- [29] J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," 2018, <https://arxiv.org/abs/1804.02767>.
- [30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [31] S. Likitha, S. Gupta, K. Hasitha, and A. Upendra Raju, "Speech based human emotion recognition using MFCC," in *Proceedings of the International Conference on Wireless Communications*, pp. 2257–2260, IEEE, Chennai, India, March 2017.
- [32] B. Zhu, Z. Xu, and M. Xia, "Dual hesitant fuzzy sets," *Journal of Applied Mathematics*, vol. 2012, Article ID 879629, 2645 pages, 2012.
- [33] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Human-Computer Studies*, vol. 51, no. 2, pp. 135–147, 1999.
- [34] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," *Computer Vision*, vol. 8693, pp. 740–755, 2014.
- [35] Y. Peter, L. Alice, H. Micah, and H. Julie, "From image descriptions to visual denotations: new similarity metrics for semantic inference over event descriptions," *Transactions of the Association for Computational Linguistics*, vol. 2, pp. 67–78, 2014.
- [36] K. Zhou, B. Sisman, R. Liu, and H. Liu, "Seen and unseen emotional style transfer for voice conversion with a new emotional speech dataset," in *Proceeding of the 2021 IEEE ICASSP International Conference on Acoustics, Speech and Signal Processing*, IEEE, Toronto, ON, Canada, February 2021.
- [37] HowNet, 2020, http://www.keenage.com/html/e_index.html.
- [38] NTUSD, 2020, <http://nlg.csie.ntu.edu.tw/nlpresource/NTUSD-Fin/>.