

Research Article

Rumor Situation Discrimination Based on Empirical Mode Decomposition Correlation Dimension

Yanwen Xin  and Fengming Liu 

School of Business, Shandong Normal University, Jinan 250014, China

Correspondence should be addressed to Fengming Liu; liufm@sdnu.edu.cn

Received 8 February 2021; Revised 14 March 2021; Accepted 16 April 2021; Published 23 April 2021

Academic Editor: Nishant Malik

Copyright © 2021 Yanwen Xin and Fengming Liu. 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.

To effectively identify network rumors and block their spread, this paper uses fractal theory to analyze a network rumor spreading situation time series, reveal its inner regularity, extract features, and establish a network rumor recognition model. The model is based on an empirical mode decomposition (EMD) correlation dimension and K-nearest neighbor (KNN) approach. Firstly, a partition function is used to determine if the time series of the rumor spreading situation is a uniform fractal process. Secondly, the rumor spreading situation is subjected to EMD to obtain a series of intrinsic mode functions (IMFs), construct the IMF1–IMF6 components containing effective feature information as the principal components, and reconstruct the phase space of the principal components, respectively. Finally, the correlation dimensions of the principal components IMF1–IMF6 as obtained by the Grassberger-Procaccia algorithm are used as feature parameters and are imported into the KNN model for rumor recognition. The experimental results show that the correlation dimension of a spreading situation can better reflect the characteristic information; as combined with the KNN model for identifying rumors, the recognition rate reaches 87.5%. This result verifies the effectiveness of fractal theory in network rumors recognition, expands the thinking for the research of rumors recognition, and provides theoretical support for rumor governance.

1. Introduction

Nowadays, network rumors use the Internet (through mainstream social software and other media platforms) to breed and spread in various forms and channels, seriously affecting people's normal lives and social order and even endangering economic development and national security. For example, the mass disturbance in Shishou, Hubei, seriously jeopardized social stability and disturbed the political order. A rumor regarding a spread of nuclear radiation from Japan triggered "salt grabbing" among the Chinese people, temporarily destroying the economic order. Therefore, quickly recognizing rumors from numerous messages and blocking their spread over time are an important research task for scientific and technological workers aiming to keep networks clear.

Rumor recognition methods are mainly based on feature and model. A recognition process based on feature

classification mainly includes learning and classification; such processes usually extract some significant features from the content, users, or propagation of rumors and apply such features in the recognition. Common classification algorithms include K-nearest neighbor (KNN) algorithms, support vector machines, decision trees, and naive Bayesian models.

Castill et al. [1] extracted 68 rumor recognition features from news topics on Twitter based on messages, users, topics, and propagation and selected the 15 best features for classification using a decision tree classifier. Qazvinian et al. [2] extracted lexical patterns, portions of speech patterns (e.g., 1–2 words), and web-based positive and negative user model characteristics, achieving a high degree of precision. Yang et al. [3] used rumor data on Sina Weibo and, based on existing research, added two new client features for publishing messages and the locations of events mentioned in the message content; they also constructed a support vector

machine classifier for automatically identifying rumors and nonrumors. Sun et al. [4] added multimedia features to the recognition features. Zhang et al. [5] believed that the content features in the shallow layer could not be effectively distinguished and extracted four implicit features based on the content to build a support vector machine classifier for recognizing rumors.

A user's behavioral characteristics can reflect the user's activity and behavioral motivation, so they are often used as a distinguishing feature for recognition. Liang et al. [6] and Zhang et al. [5] extracted features from the perspective of user behaviors as key clues for identifying rumors; both studies achieved a good identification effect.

Propagation characteristics can be regarded as the attitude of users towards the news and also reflect the influence of an event on the Internet. Zhao [7] focused on early rumor detection, used regular expressions to identify skeptical and denying tweets, and employed them as the keys to uncovering rumors. Kwon et al. [8] investigated the temporal patterns of rumor propagation, identified the characteristics of network rumors through the diffusion of time, structure, and language, and proposed a new periodic time series model for proving that rumors on social media networks may fluctuate with the passage of time. Ma et al. [9] believed that only extracting features from the static characteristics of the propagation was insufficient and established a dynamic series time structure to capture the dynamic changes in users, content, and propagation characteristics in the message life cycle. Wu et al. [10] found differences in propagation trees between rumors and nonrumors and constructed a random walk graph kernel model for calculating the similarity between propagation trees to distinguish between them. Ma et al. [11] and Liu et al. [12] also extracted static features and captured dynamic information from propagation trees to detect rumors.

In recent years, scholars have used comprehensive features from various aspects to recognize rumors. Ruchansky et al. [13] found that false information had three basically consistent features: the text of the article, corresponding users received, and original users who promoted it. Therefore, the authors proposed a method for detecting rumors based on combining three features from text content, both sides, and user behaviors. Zhang et al. [14] explored the relationships between authors, themes, post content, and news authenticity through a large number of experiments, seeking explicit and implicit features for judging the credibility of news. Shu et al. [15] indicated that the relationships among news publishers, news content, and social network users help to improve the identification of rumors. Dong Scholar [16] indicated that combining multiple recognition algorithms will lead to an improved rumor recognition effect. Wu et al. [17] fused user credibility, emotional consistency, regional correlation characteristics, a stacking integrated learning method, and an optimized rumor recognition method with different models and provided thorough cross-validation. Thus, as can be seen from the above, feature-based classification methods have been

widely studied and applied in rumor identification, and the features gradually tend to resemble each other. However, a large number of features are often needed to achieve a good recognition effect. This is time-consuming and labor-intensive. Moreover, most of these features are based on static and simple statistical features; without digging for deeper features, as rumor styles become increasingly similar to nonrumors, static features may lose their descriptiveness.

Model-based identification methods mainly include infectious disease models, information dissemination models, and neural network models. Infectious disease models are widely used to explore the evolution, spread, and control of rumors. Jin et al. [18] applied a modeling method for rumor recognition based on using an improved susceptible-exposed-infected-skeptic infectious disease model [19] to identify the spread of rumors and news, obtain conversion parameters between user states to identify rumors and news, and provide new ideas for rumor recognition. Rumor spreading and nonrumor spreading have large differences between user behaviors. Liu et al. [20] constructed a rumor and nonrumor information spreading model based on different user behaviors and calculated the spread of messages. Models are more likely to identify rumors; as such, several typical neural network structure patterns have been extensively studied in the field of rumor recognition. Ma et al. [21] used recurrent neural networks and their variants, that is, a long short-term memory (LSTM) and gated recurrent unit, to identify hidden expressions in topic-based message forwarding comments over time. They then used the hidden expressions to determine whether a topic was a rumor. Zubiag et al. [22] found that the LSTM classifier (which models a time series of information) performs well in rumor recognition. Liu et al. [23] considered the dynamic differences between rumors and nonrumors in spreaders and communication structures and proposed a rumor recognition model based on an LSTM network combined with convolutional neural networks considering the forwarding content, diffuser, and diffusion structure. Chen et al. [24] proposed a recurrent neural network model for identifying network rumors based on potential content changes in a time series. Chen et al. [25] and Srinivasan et al. [26] proposed a new convolutional neural network method and new activation function, respectively, based on using time and language features to jointly identify network rumors. Model-based rumor methods make up for the shortcomings of feature-based classification algorithms to a certain extent. Wu et al. [27–30] proposed a variety of evolutionary algorithms as the basis for information recognition, which provided a strong theoretical basis for rumor control strategies. However, as they are more complex and require a large number of data set parameters, these models have high uncertainty, which will make the recognition effects different.

This study investigates the characteristics of network rumors propagation and finds that there are differences between network rumors and nonrumors in the propagation state. An experiment verifies that a network rumor

propagation situation has fractal characteristics and that the self-similarity and complexity of its time series can be quantitatively described by the fractal dimension(s).

Fractal theory was founded by the French mathematician B. B. Mandelrot. In 1967, he published his paper “How Long is the Coastline of Britain” in the American journal “Science” [31]. Since its birth, fractal theory has developed rapidly and has been widely used in natural science, social science, cognitive science, and other fields. “Fractal” refers to the statistical self-similarity of morphologies, functions, information, and structural characteristics [32]. The emergence and application of fractal theory have given a great impetus to new subjects, and such subjects are developing rapidly. The application of fractal theory in the field of recognition is one example. Fractal theory has promoted the rapid development of recognition technology and facilitated the emergence of new features and methods for generating higher detection and classification recognition rates. Moreover, the application of these methods to a large amount of measured data promotes the development of fractal theory.

The application of fractal theory in the field of target recognition is mainly reflected in one-dimensional time series analyses. Scholars usually extract the fractal dimension of a sequence for recognition. Feng [33] and Si [34] extracted the box-counting dimensions from a time series as features and combined them with classifiers for identification. Yu [35] and Liu [36] used a correlation dimension in a time series to detect and recognize a target and were able to reflect the characteristics of nonlinear signals. For the complex fractal structure of a time series, Yuan et al. [37] and Xiong et al. [38] established multifractal theories and combined them with classifiers to build recognition models. Gintautas [39] combined multiple fractal dimensions into multiple features (or combined with them other statistical parameters) to improve the recognition efficiency when performing a voice emotion classification. To achieve better recognition performance, some scholars have improved the basic fractal dimension algorithm before extracting features for target recognition [40–43].

In the previous studies, there are relatively few studies on rumor recognition based on the nonlinear characteristics of a time series. Based on the nonlinearity of the network rumors situation, this study starts with the situation trend of Internet rumors, independent of the static text and propagation process, focuses on extracting network rumors feature vectors and understanding the characteristics and differences of rumors and nonrumors in a situation, and deeply investigates the changes in the situations of different components based on empirical mode decomposition (EMD). The correlation dimensions of the principal components of each intrinsic mode function (IMF) are obtained as a feature vector, and the vector is imported into a KNN classification algorithm for rumor recognition.

2. Partition Function to Judge Single Fractal

A change in the spread of a rumor is an important and very complex issue. Many factors affect fluctuations in rumor

spreading. If the time series of a network rumor situation has a uniform fractal structure, it can be discussed using a single fractal. If the time series is nonuniform, it cannot be analyzed with a single fractal dimension, and multiple analyses are required to discuss the complexity of the fluctuation law(s).

First, a rumor spreading situation is selected, and a partition function (statistical moment) method is used to judge whether the time series is single-fractal or multifractal, so as to accurately select features for recognition. Based on the “Zhiwei Data” platform, this study examined the global influence of the “plastic laver” network rumor every hour as a trend indicator for analysis. The event started at 00:00 on February 19, 2017, and ended at 23:00 on March 8, 2017, providing a total of 433 data points. The spreading situation for “plastic laver” network rumor is shown in Figure 1.

A partition function method can be used to determine whether a stochastic process is a single-fractal or multifractal process. The specific steps of the determination are [44] as follows:

Step 1: for a time series $\{x(t) | t \in [0, T]\}$, the interval $[0, T]$ is divided into several disjoint subintervals. λ is the maximum value of the ratio between the length of the subinterval and that of the total interval. M is the sum of the object data over the entire range and is determined as follows:

$$M = \sum_{t \in [0, T]} x(t). \quad (1)$$

Step 2: if $N(\lambda, i)$ is the i -th interval sum of the values, the average density of the i -th interval can be calculated as follows:

$$\varepsilon(\lambda, i) = \frac{N(\lambda, i)}{M}. \quad (2)$$

Step 3: a statistical moment function is constructed as follows:

$$M(\lambda, q) = \sum_t \varepsilon(\lambda, i)^q. \quad (3)$$

The statistical moment function $M(\lambda, q)$ can show the size and function of the average density $\varepsilon(\lambda, i)$.

Step 4: if the statistical moment function $M(\lambda, q) \propto \lambda^{\tau(q)}$, if $\tau(q)$ is a partition function or if $\tau(q)$ is a linear function of q , then the time series is a single-fractal process; if $\tau(q)$ is a nonlinear function about q , then the time series is multifractal.

In the time series of the “plastic laver” network rumor situation, the total interval is $[0, 433]$, and the interdistrict length is 1–8 years; thus, $\lambda = 1/433, \dots, \lambda = 8/433$, and the values of q are 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 3.5. The

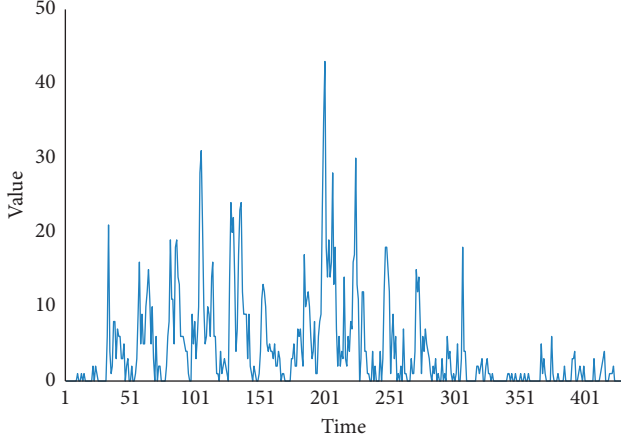


FIGURE 1: Spreading situation of “plastic laver” network rumor.

relationship between the statistical moment of the order $M(\lambda, q)$ and λ can be determined as follows:

$$\tau(q) = \lim_{\lambda \rightarrow 0} \frac{\ln M(\lambda, q)}{\ln \lambda}. \quad (4)$$

By fitting $\ln M(\lambda, q)$ and $\ln \lambda$, the slope of the resulting straight line is $\tau(q)$. The $\tau(q) - q$ relationship is shown in Table 1.

The scatter diagram and fitting line are shown in Figure 2.

It can be seen from the graph that the linear relationship between the partition function $\tau(q)$ and q is very evident. The correlation coefficient reaches 99.9%. Therefore, the network rumors situation time series is considered as a single-fractal process.

A nonrumor topic of “Hong Kong-Macao-Zhuhai Bridge opened to traffic” is used to verify the partition function; the $\tau(q)$ and q scatter plot and fitting line are shown in Figure 3.

As can be seen from Figure 3, the linear relationship between the partition function $\tau(q)$ and q is also very evident, and the correlation coefficient is 99.9%. Thus, it is feasible and reasonable to use a single fractal to identify the features of network rumors and nonrumors.

3. Materials and Methods

3.1. Empirical Mode Decomposition (EMD). EMD is an adaptive time-frequency processing method for a time series. This method does not require selection of a basis function, so it has advantages in processing nonlinear time series [45]. Using the EMD algorithm, a complex time series can be decomposed into a finite number of IMFs. The decomposed IMF components contain the local characteristics of the different time scales of the original time series. Therefore, EMD decomposition is more conducive to distinguishing between the characteristics of network rumors and nonrumors. The specific decomposition steps are as follows [46]:

Step 1: identify the local extremum of the original signal $X(t)$, connect the maximum value to the upper

TABLE 1: The $\tau(q) - q$ relationship.

q	$\tau(q)$
0.0	-0.8077
0.5	-0.4239
1.0	0.0
1.5	0.4252
2.0	0.8395
2.5	1.241
3.0	1.632
3.5	2.013

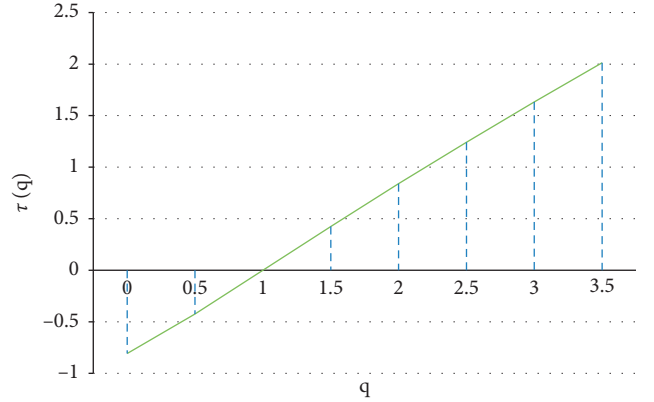


FIGURE 2: Relationship between network rumors $\tau(q) - q$.

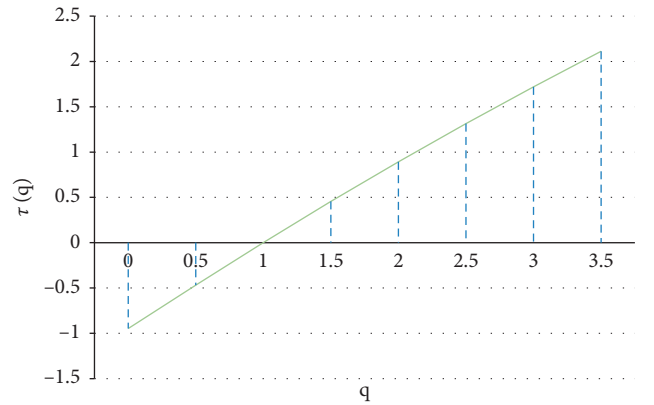


FIGURE 3: Relationship between network nonrumors $\tau(q) - q$.

envelope, connect the minimum value to the lower envelope, and obtain the mean sequence m_1 according to the upper and lower envelope.

$$h_1(t) = X(t) - m_1(t). \quad (5)$$

If h_1 satisfies the definition of an IMF, h_1 can be called an IMF; otherwise, repeat the above iterative steps.

$$h_2(t) = h_1(t) - m_2(t). \quad (6)$$

Repeat this operation until $h(k)$ satisfies the above IMF component condition; $h_k(t) = IMF_1(t)$; this is the first IMF component.

$$h_k(t) = h_{k-1}(t) - m_k(t). \quad (7)$$

Step 2: IMF_1 is separated from the original signal. The remaining amount $r_1(t)$ is used as a new original signal, and step 1 is executed cyclically. After the iterative calculation of n , $r_n(t)$ is obtained as shown. When $r_n(t)$ tends to a monotonic function, the decomposition ends.

$$\begin{cases} r_2(t) = r_1(t) - IMF_2(t), \\ r_3(t) = r_2(t) - IMF_3(t), \\ \dots \\ r_n(t) = r_{n-1}(t) - IMF_n(t). \end{cases} \quad (8)$$

Step 3: the original signal is decomposed to obtain n IMF components and one residual quantity r_n as follows:

$$X(t) = \sum_{j=1}^n IMF_j + r_n. \quad (9)$$

In the above, r_n represents the average trend of the decomposition remainder and signal. Each IMF component corresponds to different characteristic scales, and the component frequency is gradually distributed from high to low. Finally, several IMFs can reflect the fluctuation degree of the original time series on different scales and further reveal the inherent regularity of the series volatility.

3.2. Correlation Dimension. The correlation dimension describes the nonlinear behavior and complexity of the system. The larger the correlation dimension, the more complex the system. At present, the main method for calculating the correlation dimension is the Grassberger-Procaccia (G-P) algorithm [47]. The calculation steps are described below.

The hypothesis $\{x_1, x_2, x_3, \dots, x_k\}$ is the time series obtained for monitoring, where $K = 1, 2, 3, \dots, N$. A pair of phase points are determined after reconstructing the phase space as follows:

$$X_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), \quad (10)$$

$$X_j = (x_j, x_{j+\tau}, \dots, x_{j+(m-1)\tau}). \quad (11)$$

Then, the distance between X_i and X_j in the Euclidean space is r_{ij} , which is determined as follows:

$$r_{ij} = |X_i - X_j| \quad (12)$$

Given the critical distance r , the proportion of point pairs whose distance is less than r_{ij} in all of the point pairs $C(r)$ is as follows:

$$C(r) = \frac{1}{N^2} \sum_{i \neq j} \theta(r - |X_i - X_j|). \quad (13)$$

In the formula, N is the total number of phase points, and $\theta(x)$ is the Heaviside function, which is defined as follows:

$$\theta(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (14)$$

where $C(r)$ is a correlation integral, and it characterizes the probability that the distance r_{ij} between two points in the phase space is less than r ; it is used to appropriately adjust the range of values of r . The correlation integral is $C(r) \propto r^D$, and D is the correlation dimension. At this time, the reconstructed attractor has fractal characteristics. The correlation dimension D is determined as follows:

$$D = \frac{\ln C(r)}{\ln r}. \quad (15)$$

Therefore, to determine the correlation dimension of a system, one can draw the double logarithmic curve of $\ln C(r)$ relative to $\ln r$ and fit the relatively long section of the near-linear part of the curve using a least-squares method. The slope can then be used to calculate the value of the correlation dimension of the system. If the slope of the double logarithmic curve converges to a saturation value as the embedding dimension m increases, then the saturation value is the correlation dimension D .

3.3. *K*-Nearest Neighbor (KNN) Algorithm. The KNN algorithm is stable and effective. The most significant advantages of this algorithm are its simple calculations, high accuracy, and high tolerance for outliers and noise. Its principle is relatively simple: given a test object to be classified, find the K objects that are most similar to the object to be classified, and determine the category of the object to be classified according to the category of the K objects.

In the KNN algorithm, the value of K has a significant influence on the experimental results. A small value of K makes it impossible to observe the characteristics of the object to be classified; moreover, it is easy to introduce objects that are irrelevant to the object to be classified, and, eventually, the classification accuracy will be reduced.

4. Experimental Testing and Result Analysis

4.1. Data Description. To test the performance of the recognition algorithm based on the platform of "Zhiwei Data", 60 network rumors (such as the "Spring Festival Express Shutdown Timetable") and 60 nonrumors (such as the "Beijing Chaoyang Hospital Injury Event") were collected as the time series constituting the data set for recognizing network rumors; the per hour "global influence" was used as the index. The global influence of the "Zhiwei Data" platform is based on the social media and network media data of the

entire network and can describe authoritative indicators of the spread of a single event on the Internet. The data set was divided into a training set and a test set at a ratio of 4 : 1; that is, the training sample data set comprised 96 samples: 48 rumor situation time series and 48 nonrumor situation time series, and the testing data set comprised 24 test samples: 12 rumor situation time series and 12 nonrumor situation time series. This study used MATLAB software for analysis.

4.2. Feature Extraction

4.2.1. EMD Decomposition. According to the EMD decomposition algorithm in Chapter 3, the time series of the spreading situation of 120 pieces of information was decomposed. The time series of the “plastic laver” network rumor situation and “Hong Kong-Macao-Zhuhai Bridge opened to traffic” nonrumor situation time series EMDs are shown in Figures 4 and 5, respectively. In the figure, IMF1–IMF7 are the effect diagrams of the EMD decomposition of the rumor propagation situation in turn. Each order of the IMF components contains different characteristic information.

As can be seen from the exploded diagrams, there is a difference between the frequency components of network rumors and nonrumors after EMD; the decomposed layers are not identical, but the effective components of the two are mainly distributed in several low-order IMF components.

To facilitate follow-up work, the effective IMF components must be extracted as feature vectors. The cross-correlation function represents the degree of correlation between two time series and is used as a criterion for the correlations between IMF components. The propagation situation after EMD comprises using the cross-correlation method to screen the decomposed IMF components, calculate the correlation coefficient between the IMF components and original time series, and select the IMF component with more effective information as the principle components. The value range of the cross-correlation coefficient is $[-1, 1]$; the greater the absolute value, the higher the correlation. The correlation coefficients between some network rumors and nonrumors IMF components and the original time series are shown in Table 2.

As shown, the correlation between the first six orders of IMF components and the original time series is evident; the correlation between IMF7 and the original time series is not large, and thus its analysis is of little significance. Therefore, the IMF1–IMF6 components are used as the principle components for subsequent research.

4.2.2. Calculation of Correlation Dimension. Before resolving the correlation dimension in the reconstructed phase space, the embedding dimension m and embedding delay τ must be determined. This study uses the “C-C” algorithm and MATLAB to calculate the two parameters. The results from randomly selecting rumors and nonrumors are shown in Figures 6 and 7, respectively.

As shown, the algorithm with the optimal embedding delay time can be determined when the left-side $\Delta s(t)$ curve

takes the first minimum value. As can be seen from the above figure, the embedding delay for rumors is $\tau = 3$, and that for nonrumors is $\tau = 2$. The time window width can be determined as when the right-side $S_{\text{cor}}(t)$ curve takes the minimum value; that is, the time window for rumors is $\tau_w = 9$, and the time window for nonrumors is $\tau_w = 8$. According to the formula $\tau_w = \tau(m - 1)$, the embedding dimension for rumors is $m = 4$, and the embedding dimension for nonrumors is $m = 5$.

According to the obtained time delay τ and embedding dimension m , the G-P algorithm introduced in Chapter 3 is used to verify the time delay and embedding dimension; that is, as the embedding dimension m continues to increase, the correlation dimension no longer changes, and, with some convergence, the minimum embedding dimension m can be determined.

Therefore, relationship diagrams are constructed for rumors and nonrumors $\ln C(r) - \ln r$, and the corresponding correlation dimension is obtained by fitting the slope of the straight line. The results are shown in Figures 8 and 9.

As shown, as the embedding dimension m increases, the correlation dimension tends to converge gradually and does not change. For the time series of rumors and nonrumors in the above cases, the correlation dimensions are 1.0366 and 1.2265, respectively. The fractal characteristics of the time series of the rumor and nonrumor spread can be seen based on the correlation dimension, which is a decimal number.

In the discussion on EMD decomposition, the correlation dimension of the IMF1–IMF6 components was taken as the feature for recognition. According to the above algorithm, the average values of the principal components for 60 rumors and 60 nonrumors are obtained, as shown in Table 3.

The line chart is shown in Figure 10.

It can be seen from Table 3 and Figure 10 that there is a relatively evident difference between the correlation dimension for rumors and that for nonrumor. From the linear trend lines, it can be observed that the correlation dimension of the rumors decreases with an increase in the IMF order, indicating that the internal changes of rumors have evident regularity; the correlation dimension for nonrumors is relatively stable and has no great fluctuations. Moreover, the correlation dimension of the rumors is evidently larger than that of the nonrumors, reflecting that there are more effective and complex components in the spread of rumors than those in nonrumors; this is consistent with our conjecture based on reality.

4.3. Rumor Recognition Based on KNN. Through calculation, the correlation dimension of the principal components of each propagation situation is obtained. According to the recognition model, the correlation dimension for the rumor and nonrumor situations is used as the feature vector for training the KNN classification model. A total of 60 rumor (1) and nonrumor (0) situation pieces of data were collected; 48 groups were randomly selected as training samples, and the remaining 12 groups were used as test samples. In this way, 96 sets of training samples and 24 sets of test samples

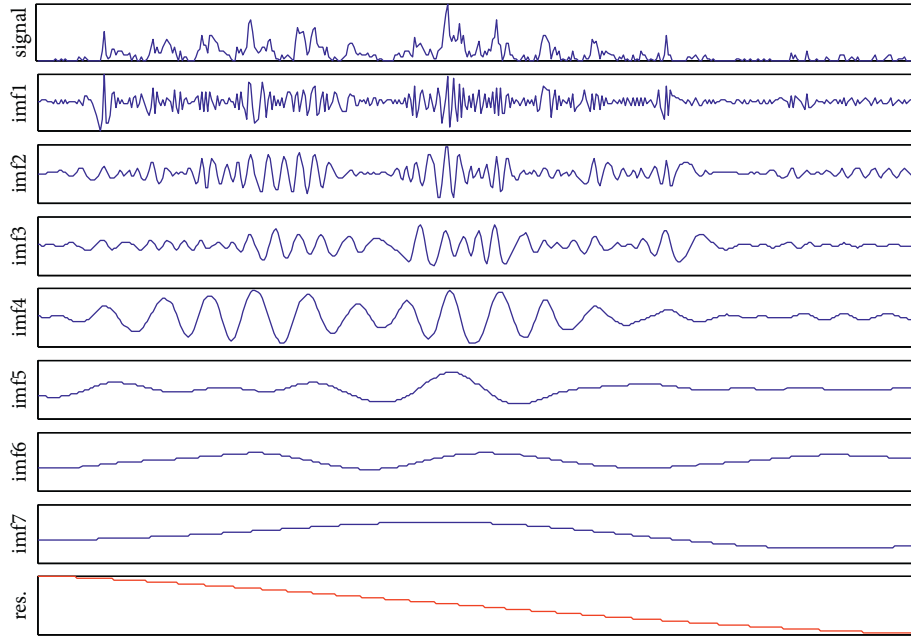


FIGURE 4: Empirical mode decomposition (EMD) exploded view of “plastic laver” network rumor.

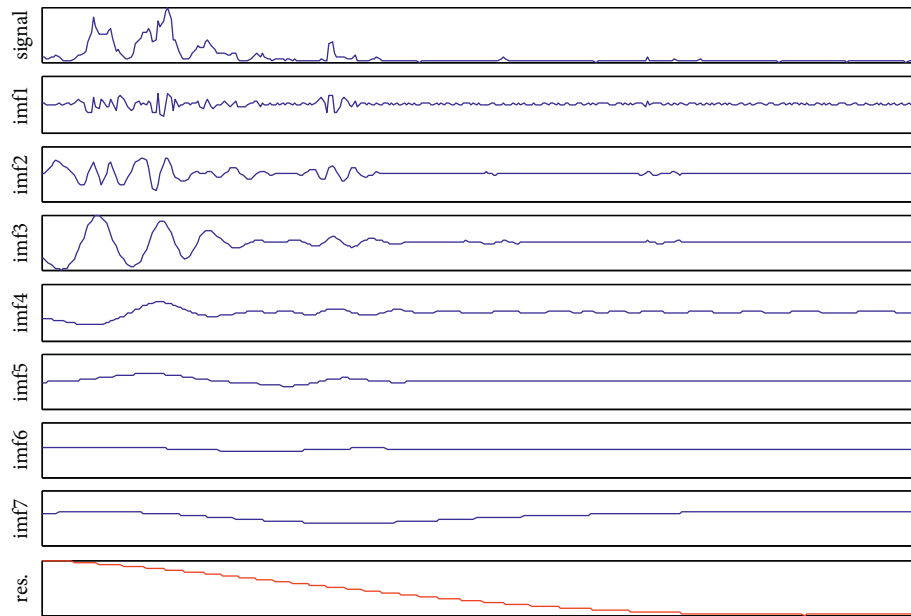


FIGURE 5: EMD exploded view of “Hong Kong-Macao-Zhuhai Bridge opened to traffic” nonrumor.

TABLE 2: Correlation coefficients between intrinsic mode function (IMF) components and original time series.

Category	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7
Rumor	0.0524	0.3837	0.1041	0.1426	0.1465	0.2034	-0.0177
	-0.0632	0.4740	0.2999	0.3642	0.4553	0.3227	-0.0169
	0.4322	0.3762	0.4585	0.5587	0.1736	0.1168	-0.0478
Nonrumor	0.2531	0.4964	0.4703	0.3806	0.1335	0.6107	-0.1901
	0.1020	0.4531	0.3708	0.3952	0.3336	0.2901	0.0967
	-0.1111	0.0388	0.3525	0.1834	0.1870	0.2607	0.0759

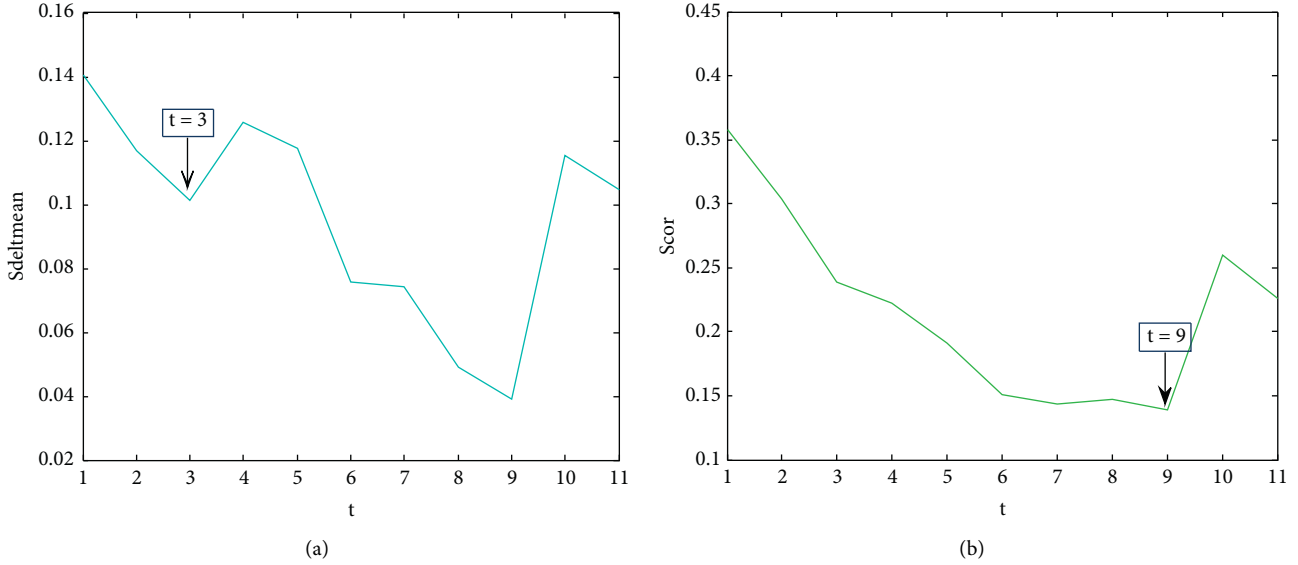


FIGURE 6: Rumor C-C algorithm results.

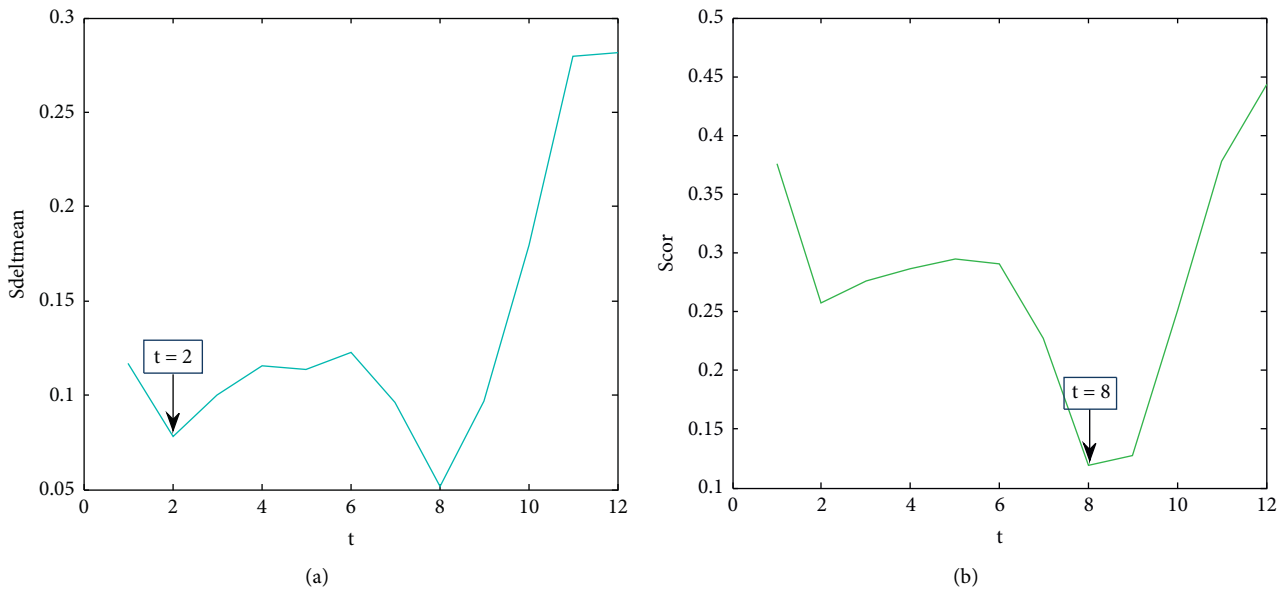


FIGURE 7: Nonrumor C-C algorithm results.

were obtained. The correlation dimension of the principle components of each situation's time series was employed as the input signal, that is, the respective correlation dimensions of IMF1–IMF6. The output signals were rumors and nonrumors; an output of “1” denoted rumors, whereas “0” represented nonrumors.

The process is shown in Figure 11.

4.4. Analysis and Comparison of Experimental Results. The classification effect evaluation index is calculated according to the confusion matrix shown in Table 4.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}.$$

$$\text{Precision} = \frac{TP}{TP + FP}.$$

$$\text{Recall} = \frac{TP}{TP + FN}.$$

F1 is the harmonic mean of precision and recall, and it is determined as follows:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

The larger the value of F1, the better the overall effect of the rumor recognition classifier.

Through experiments, the confusion matrix for the results is determined, as shown in Table 5.

The accuracy of the rumor recognition based on KNN was 87.5%, the precision was 84.6%, the recall rate was 91.67%, and F1 was 87.1%.

The experimental results in this study are compared with those from existing research on rumor recognition, as shown in Table 6.

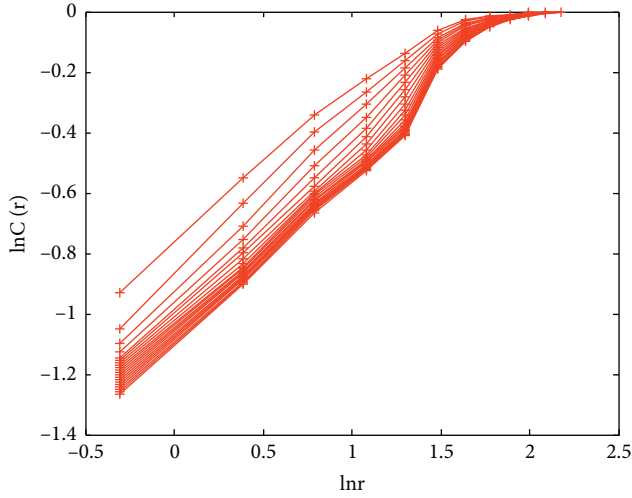


FIGURE 8: Distribution of rumor correlation integral.

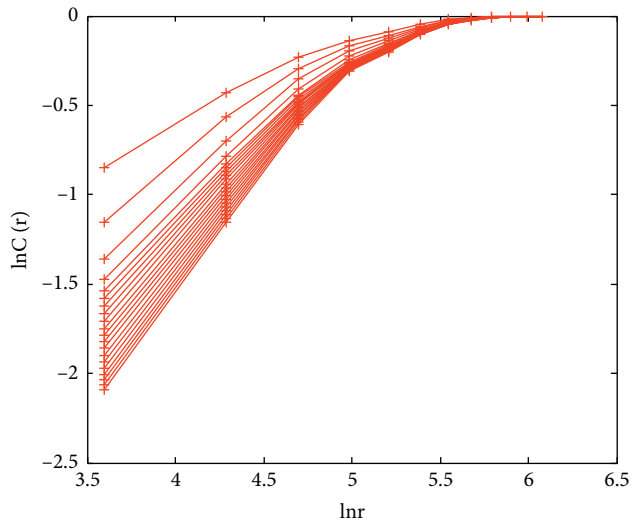


FIGURE 9: Distribution of nonrumor correlation integral.

TABLE 3: Average correlation dimensions of the principal components.

Category	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6
Rumor	2.4049	2.0713	1.8860	1.8587	1.5439	1.3260
Nonrumor	1.3023	1.4567	1.5817	1.7313	1.5009	1.2850

It can be seen from Table 6 that the recognition effect based on the EMD correlation dimension proposed in this study is ideal. In the selection of rumor identification and classification methods, many scholars use support vector machines and decision trees to identify and classify network rumors. Based on the features selected in this study, the same numbers of training samples and test samples were selected, and the recognition results are shown in Table 7.

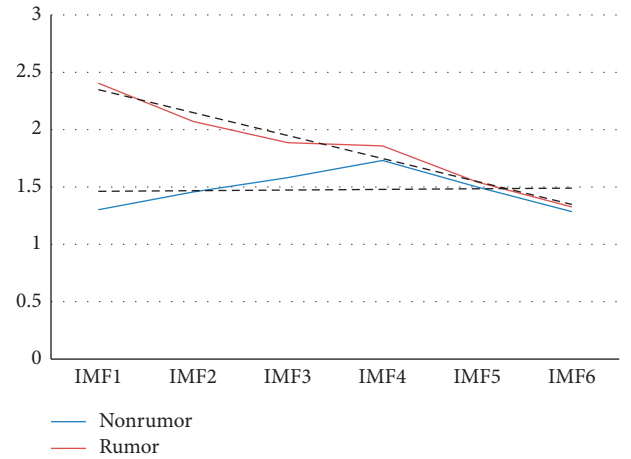


FIGURE 10: Line graph of correlation dimension of intrinsic mode function (IMF) components.

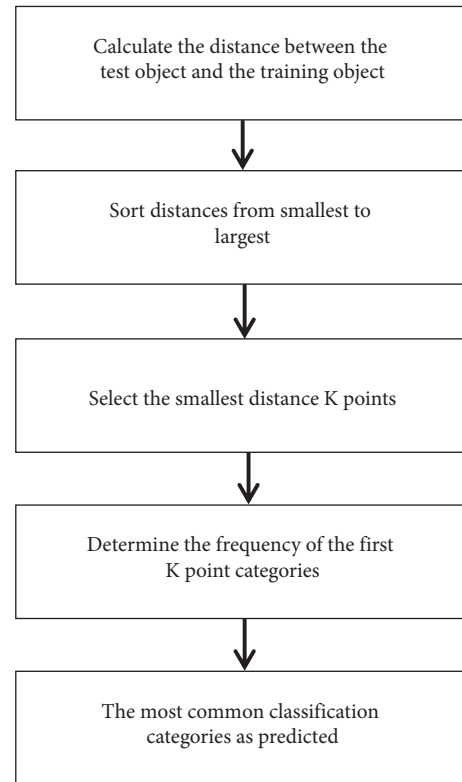


FIGURE 11: K-nearest neighbor (KNN) classification process.

TABLE 4: Confusion matrix.

		Prediction	
		1	0
Actual	1	TP	FN
	0	FP	TN

The confusion matrix is used to calculate the evaluation indexes used in the classification results.

TABLE 5: Confusion matrix of rumor results.

		Prediction	
		1	0
Actual	1	11	1
	0	2	10

TABLE 6: Comparison of experimental results.

Research work	Feature	Classification method	Evaluation index	
			Accuracy (%)	F1
Castill et al. [1]	Content, user, propagation	Decision tree	86	84.9%
Qazvinian et al. [2]	Content, user	Support vector machine (SVM)		89.7%
Yang et al. [3]	Content, user, propagation	Decision tree	78.7	
Sun et al. [4]	Content, user	SVM		70.4%
Zhang et al. [5]	Content, user	SVM		74.4%
	User	SVM		76.9%
Liang et al. [6]	User	Decision tree		85.9%
	User	Naive Bayesian model		77.8%
	Propagation	SVM	87.3	86.7%
Kwon et al. [8]	Propagation	Decision tree	82.1	82.2%
	Propagation	Naive Bayesian model	89.7	87.8%
Yuan et al. [33]	Propagation	SVM	85	84%
Wu et al. [17]	Content, user, propagation	Ensemble learning	80	
Liu et al. [20]	User	Information dissemination model		81.3%
Ma et al. [21]	Content	Recurrent neural network (RNN)	88.1	89.8%
Liu et al. [23]	User, propagation	Long short-term memory (LSTM)	94.8	94.6%
Srinivasan et al. [26]	Content	Convolutional neural network (CNN)	88	92%
This study	The fractal characteristics of propagation	K-nearest neighbor (KNN)	87.5	87.1%

TABLE 7: KNN, decision tree, and SVM experimental results comparison.

Methods	Recognition rate (%)
KNN	87.5
Decision tree	83.3
SVM	66.7

According to Table 7, the recognition rate of this method is slightly higher when the data samples are identical. Therefore, it can be shown that a feature extraction method based on a nonlinear theory combined with the KNN classifier can effectively identify rumors, thereby furthering research on rumor control and governance.

5. Conclusions

In this study, a rumor recognition model is established based on an EMD correlation dimension and KNN approach. It uses the time series of a rumor situation based on a “global influence” index and fractal theory, aiming at the nonlinear characteristics of a rumor spreading process. The recognition model does not rely on the static text and complex propagation process of rumors but only extracts fractal features from the propagation situation for rumors recognition.

To judge whether the time series of a rumor situation is a uniform fractal process, an example is verified according to a partition function method. After fitting, the linear

relationship between $\tau(q)$ and q is evident; thus, it is more reasonable to use a single fractal to extract the features of rumor and nonrumor spreading situations.

To further reveal the internal regularity of the time series fluctuations and study the detailed characteristics of the original signal more accurately, EMD is conducted for the rumor propagation situation before identification, so as to obtain each IMF component containing effective feature information. A cross-correlation method is used to extract the principle components, laying a good foundation for the later research of the correlation dimension.

Next, the phase space of the rumor situation is reconstructed, and the C-C algorithm is used to obtain the time delay and embedding dimension. Then, the G-P algorithm is used to calculate the correlation dimensions of the principal components IMF1–IMF6 as the characteristic parameters. The correlation dimension of an IMF component can more comprehensively highlight the internal information corresponding to rumors and nonrumors, and the discrimination becomes evident. The KNN model is introduced to identify the rumors.

The principal components are taken as the input to the KNN model, and rumors and nonrumors are taken as the outputs of the model. The results show that the correlation dimension and its detailed information regarding the situation, including the change rules, can better reflect the characteristic information. Combined with the KNN model, it can effectively identify and classify a rumor situation, and the recognition rate can reach 87.5%. The experiment shows

that this method is effective in classification and has certain advantages in rumor recognition, thereby providing a new method for rumor recognition.

Data Availability

The data are included in the manuscript.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Acknowledgments

This research was supported in part by the National Natural Science Foundation of China (no. 71701115), National Social Science Foundation of China (no. 14BTQ049), Shandong Natural Science Foundation (no. ZR2017MF058), Special Project for Internet Development of Social Science Planning Special Program of Shandong Province (no. 17CHLJ23), and Social Science Planning Program of Shandong Province (no. 13BTQJ01).

References

- [1] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web ACM*, pp. 675–684, New York, NY, USA, April 2011.
- [2] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, "Rumor has it: identifying misinformation in microblogs," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1589–1599, Association for Computational Linguistics, Edinburgh, UK, July 2011.
- [3] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, August 2012.
- [4] S. Sun, H. Liu, J. He, and X. Du, "Detecting event rumors on sina weibo automatically," in *Proceedings of the Asia-Pacific Web Conference*, pp. 120–131, Springer, Berlin, Germany, November 2013.
- [5] Q. Zhang, S. Zhang, J. Dong, J. Xiong, and X. Cheng, "Automatic Detection of Rumor on Social Network," in *Proceedings of the Natural Language Processing And Chinese Computing*, pp. 113–122, Springer, Berlin, Germany, 2015.
- [6] G. Liang, W. He, C. Xu, L. Chen, and J. Zeng, "Rumor identification in microblogging systems based on users' behavior," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 99–108, 2015.
- [7] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: early detection of rumors in social media from enquiry posts," in *Proceedings of the International World Wide Web Conferences Steering Committee*, pp. 1395–1405, Geneva, Switzerland, May 2015.
- [8] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *Proceedings of the 2013 IEEE 13th International Conference Data Mining*, pp. 1103–1108, IEEE, New York, NY, USA, December 2013.
- [9] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *Proceedings of the 24th ACM International Conference Information Knowledge Management*, pp. 1751–1754, ACM, New York, NY, USA, December 2015.
- [10] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, "Real-time rumor debunking on twitter," in *Proceedings of the 24th ACM Int. Conf. Information Knowledge Management*, pp. 1867–1870, ACM, New York, NY, USA, September 2015.
- [11] J. Ma, W. Gao, Z. Wei, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," in *Proceedings of the 55th Annual Meeting Of the Association For Comput. Linguistics*, pp. 708–717, ACL, Vancouver, Canada, August 2017.
- [12] Y. Liu, X. Jin, H. Shen, and X. Cheng, "Do rumors diffuse differently from Non-rumors? A systematically empirical analysis in Sina Weibo for rumor identification," in *Proceedings of the Pacific-asia Conference Knowledge Discovery and Data Mining*, pp. 407–420, Springer, Berlin, Germany, July 2017.
- [13] N. Ruchansky, S. Seo, Y. Liu, and C. S. I. "A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM Conference Information And Knowledge Management*, pp. 797–806, Association for Computing Machinery, Singapore, December 2017.
- [14] J. Zhang, L. Cui, Y. Fu, and F. Gouza, Fake News Detection with Deep Diffusive Network Model, 2018.
- [15] K. Shu, S. Wang, and H. Liu, "Beyond News Contents: the role of social context for fake news detection," in *Proceedings of the 12th ACM International Conference. Web Search And Data Mining*, pp. 312–320, Melbourne, Australia, February 2019.
- [16] D. Wang and X. Yue, "The weighted multiple meta-models stacking method for regression problem," in *Proceedings of the 38th Chinese Control Conference*, pp. 29–34, Guangzhou, China, July 2019.
- [17] Y. Wu, Y. Zeng, J. Yang, and Z. Zhao, "Weibo rumor recognition based on communication and stacking ensemble learning," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 9352153, 12 pages, 2020.
- [18] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan, "Epidemiological modeling of news and rumors on Twitter," in *Proceedings of the 7th Workshop On Social Network Mining And Analysis*, pp. 1–9, ACM, Chicago, Illinois, August 2013.
- [19] L. M. A. Bettencourt, A. Cintrón-Arias, D. I. Castillo-Chávez, and C. Castillo, "The power of a good idea: quantitative modeling of the spread of ideas from epidemiological models," *Physica A: Statistical Mechanics and Its Applications*, vol. 364, pp. 513–536, 2006.
- [20] Y. Liu and S. Xu, "Detecting Rumors through modeling information propagation networks in a social media environment," *IEEE Transactions on Computational Social Systems*, vol. 3, no. 2, pp. 46–62, 2016.
- [21] J. Ma, W. Gao, P. Mitra et al., "Detecting rumors from microblogs with recurrent neural networks," in *Proceedings of the 25th International Joint Conference On Artificial Intelligence*, pp. 3818–3824, AAAI Press, New York, NY, USA, 2016.
- [22] A. Zubiaga, E. Kochkina, M. Liakata et al., "Discourse-aware rumour stance classification in social media using sequential classifiers," *Information Processing & Management*, vol. 54, no. 2, pp. 273–290, 2018.
- [23] Y. Procter, X. Jin, and H. Shen, "Towards early identification of online rumors based on long short-term memory networks," *Information Processing & Management*, vol. 56, no. 4, pp. 1457–1467, 2019.
- [24] T. Chen, X. Li, H. Yin, and J. Zhang, "Call attention to rumors: deep attention based recurrent neural networks for early

- rumor detection,” in *Proceedings of the 11th International Workshop On Semantic Evaluation*, pp. 40–52, Vancouver, Canada, 2017.
- [25] Y. Chen, Z. Liu, and H. Kao, “IKM at SemEval-2017 task 8: convolutional neural networks for stance detection and rumor verification,” in *Proceedings of the 11th International Workshop On Semantic Evaluation*, pp. 465–469, Vancouver, Canada, 2017.
- [26] S. Srinivasan and L. D. Dhinesh Babu, “A parallel neural network approach for faster rumor identification in online social networks,” *International Journal on Semantic Web and Information Systems*, vol. 15, no. 4, pp. 69–89, 2019.
- [27] W. Deng, J. Xu, Y. Song, and H. Zhao, “Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem,” *Applied Soft Computing*, vol. 100, Article ID 106724, 2020.
- [28] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, “An improved quantum-inspired differential evolution algorithm for deep belief network,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, 2020.
- [29] Y. Song, D. Wu, W. Deng et al., “Multi-population parallel co-evolutionary differential evolution for parameter optimization,” *Energy Conversion and Management*, vol. 228, no. 15, Article ID 113661, 2021.
- [30] W. Deng and J. Xu, “An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 99, 2020.
- [31] B. Mandelbrot, “How long is the coast of Britain? Statistical self-similarity and fractional dimension,” *Science*, vol. 156, no. 3775, pp. 636–638, 1967.
- [32] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory,” *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.
- [33] F.-C. Gu, H.-C. Chang, F.-H. Chen, and C.-C. Kuo, “Partial discharge pattern recognition of power cable joints using extension method with fractal feature enhancement,” *Expert Systems with Applications*, vol. 39, no. 3, pp. 2804–2812, 2012.
- [34] W. Si, G. Yang, X. Chen, and J. Jia, “Gait identification using fractal analysis and support vector machine,” *Soft Computing*, vol. 23, no. 19, pp. 9287–9297, 2019.
- [35] Y. Yao and N. Wang, “Fault diagnosis model of adaptive miniature circuit breaker based on fractal theory and probabilistic neural network,” *Mechanical Systems and Signal Processing*, vol. 142, Article ID 106772, 2020.
- [36] M. Liu, B. Li, J. Zhang, and K. Wang, “An application of ensemble empirical mode decomposition and correlation dimension for the HV circuit breaker diagnosis,” *Automatika*, vol. 60, no. 1, pp. 105–112, 2019.
- [37] Y. Yuan and F. Liu, “Rumor situation recognition recognition based on multifractals,” *Fractals*, vol. 27, no. 3, Article ID 1950027, 2019.
- [38] X. Xiong and W. Zhang, “Optical fiber perimeter vibration signal recognition based on multifractal spectrum,” *Acta Photonica Sinica*, vol. 48, no. 2, Article ID 206001, 2019.
- [39] G. Tamulevičius, R. Karbauskaitė, and G. Dzemyda, “Speech emotion classification using fractal dimension-based features,” *Nonlinear Anal.* vol. 24, no. 5, pp. 679–695, 2019.
- [40] S. Fu, G. Zhang, and L. Yang, “Spectrum sensing defending against PUE attack based on fractal dimension,” *Cluster Computing*, vol. 22, no. S2, pp. 2667–2675, 2019.
- [41] X. Chen, J. Li, and H. Han, “Signal subtle feature extraction algorithm based on improved fractal box-counting dimension,” in *Proceedings of the Conf. Cloud Computing And Security (ICCCS)*, pp. 684–696, Haikou, China, June 2018.
- [42] M. Dalal, M. Tanveer, and R. B. Pachori, “Automated identification system for focal EEG signals using fractal dimension of FAWT-based Sub-bands signals,” *Advances in Intelligent Systems and Computing*, vol. 748, pp. 583–596, 2018.
- [43] R. Uthayakumar and D. Easwaramoorthy, “Epileptic seizure detection in eeg signals using multifractal analysis and wavelet transform,” *Fractals*, vol. 21, no. 2, Article ID 1350011, 2013.
- [44] X. Sun, H. Chen, Z. Wu, and Y. Yuan, “Multifractal analysis of Hang Seng index in Hong Kong stock market,” *Physica A: Statistical Mechanics and Its Applications*, vol. 291, no. 1–4, pp. 553–562, 2001.
- [45] J. Wang, P. Shang, J. Xia, and W. Shi, “EMD based refined composite multiscale entropy analysis of complex signals,” *Physica A: Statistical Mechanics and Its Applications*, vol. 421, pp. 583–593, 2015.
- [46] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
- [47] P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” *Physica D: Nonlinear Phenomena*, vol. 9, no. 1–2, pp. 189–208, 1983.