International Journal of Innovative Computing, Information and Control
Volume **21**, Number **3**, June **2025**

ICIC International ©2025 ISSN 1349-4198
pp. **791–812**

A PERSONALIZED MODEL BASED ON SIMPLE RECURRENT UNIT FOR ACCURATE PREDICTION OF STOCK PRICE TREND

Wenjiang Bai[1], Shuai Yuan[2], Yan Lu[3], Mingyan Xu[3], Shuoru Chen[4]
Feng Zhao[4,5] and Feng Kang[2,*]

[1]Intelligence and Information Engineering Department
Taiyuan University
No. 7, Fendong Street, Tanghuai Industrial Park, Taiyuan 030032, P. R. China
baiwen1979@tyu.edu.cn

[2]Information Engineering College
Yantai Institute of Technology
No. 100, Gangcheng East Street, Laishan, Yantai 264005, P. R. China
yuanshuai@yitsd.edu.cn; *Corresponding author: kangfeng@yitsd.edu.cn

[3]School of Statistics
[4]School of Computer Science and Technology
[5]Yantai Key Laboratory of Big Data Modeling and Intelligent Computing
Shandong Technology and Business University
No. 191, Haibin Middle Road, Laishan, Yantai 264005, P. R. China
{ 2020420100; 2022410036; 201513350 }@sdtbu.edu.cn; xumingyan@siheinfo.com

Received June 2024; revised October 2024

Abstract. *Stock is a vital component of the financial market, and how to accurately predict stock price trend is a popular topic and open problem. In order to achieve higher prediction accuracy, we propose a personalized model based on simple recurrent unit (PSRU), which can effectively eliminate the impact of noise and diversity of stock price series. Specifically, we firstly introduce variational mode decomposition (VMD) in terms of parameter determination for denoising parameters and perform denoising on the time series, effectively reducing the noise in the data and improving the prediction accuracy. Then, to better capture the characteristics of diverse time series, we cluster the time series into multiple classes, and each class fits an optimal prediction model. Especially, to improve the clustering, we introduce the dynamic multi-perspective personalized similarity measurement (DMPSM), which can help avoid singularities, time shifts and warping in time series. Finally, SRU is adopted as the prediction model which can achieve high prediction accuracy with excellent mapping and parallel processing capabilities. The experimental results show that PSRU outperforms other commonly used time series forecasting methods on multiple performance metrics, demonstrating the effectiveness and superiority of PSRU in stock price prediction, especially in dealing with nonlinear and non-stationary time series data.*
Keywords: Stock price trend prediction, Variational mode decomposition, Simple recurrent unit, Personalized similarity measurement

1. Introduction.

As one of the most popular ways of financial management, stock investigation has attracted more and more institutional and individual investors [1]. Accurate stock price trend prediction can not only reduce the risk of investment decision-making [2], but also effectively promote the stability of financial market and the healthy development of economy. However, the stock price is highly noisy, nonlinear and non-stationary due to many factors in real-world financial market, such as exchange rate, inflation rate,

DOI: 10.24507/ijicic.21.03.791

economic policy, and political emergencies [3]. Certainly, the fluctuation trends of different stock price series are also very different and diverse. Therefore, accurate prediction of stock price trend is still a challenging task and open problem [4].

To date, many methods have been proposed for stock price trend prediction, which can be roughly grouped into three categories. The first category is the statistical methods. Stock prices over time are essentially time series [5], and the most commonly used prediction methods of time series include autoregressive integrated moving average (ARIMA) [6], autoregressive conditional heteroskedasticity (ARCH) [7], generalized autoregressive conditional heteroskedasticity (GARCH) [8], etc. Ariyo et al. applied ARIMA to published stock data, and the results revealed that ARIMA had a strong potential for short-term prediction [9]. Zheng et al. optimized and improved ARIMA algorithm by combining intervention analysis technology [10]. Kristjanpoller and Hernández used the GARCH model extended with exogenous variables to predict the price volatility, which can obtain better forecasting [11]. However, these statistical methods still cannot completely capture the characteristics of the stock price data [12]. The nonlinearity, non-stationary and even seasonally volatility of stock data also make the basic assumptions of these methods barely hold in real-world practice [13].

The second category is the machine learning methods. With the rapid development of artificial intelligence techniques in recent decades, numerous machine learning methods have been employed for financial time series analysis [14,15]. For example, Kim applied support vector machine (SVM) [16] to stock price prediction and achieved better results than statistical prediction models [17]. Support vector regression (SVR) [18] is proposed on the basis of SVM in order to better solve the regression problem in stock price prediction. Huang optimized the SVR model using genetic algorithm, and verified its effectiveness on the data from Taiwan stock market [19]. Decision tree regression (DTR) is also a popular supervised learning method for target value prediction [20]. Rathan et al. utilized DTR to predict crypto-currency's price which got acceptable performance [21]. Despite the effectiveness of the above machine learning methods, they are not able to deal with massive data and reflect more complex mapping relationships. Therefore, it is necessary to develop more advanced methods to better handle the stock data.

The third category is the deep learning methods. With the advent of big data and high-performance computation technology, deep learning has made great achievements in many fields, such as image classification [22], speech recognition [23], and object detection [24]. In view of this, researchers have applied deep learning to the stock market for more accurate prediction [25]. Nikou et al. used artificial neural network (ANN), SVR, random forest (RF) and long short-term memory (LSTM) to predict the closing price of stock respectively, and the experimental results showed that LSTM had the highest precision [26]. Li et al. achieved effective prediction of industry rotation by constructing a multi-dimensional factor dataset and employing an attention LSTM model [27]. Cho et al. firstly proposed gated recurrent unit (GRU) with higher training efficiency than LSTM [28]. Li et al. adopted GRU to predict time series data, and the empirical research on stock data confirmed the superiority of GRU in terms of accuracy and speed [29]. Alsheebah and Al-Fuhaidi proposed a deep learning model based on the gated recurrent unit (GRU) algorithm, which significantly improved the accuracy of predicting the next-day closing price of the stock market by incorporating exogenous variables [30]. On the basis of LSTM and GRU, simple recurrent unit (SRU) [31] was proposed in 2017, which further improved the operation efficiency by parallelizing some operations. Motivated by the idea that the integration of two deep learning models as a hybrid model can achieve better prediction performance, Kim and Kim combined LSTM and convolutional neural network (CNN) which was used to predict stock prices from the perspectives of time series and chart images [32]. Moreover, several methods choose to reconstruct prediction series through decomposition algorithms and similarity measurement of time series before using the deep learning model. Cao et al. established a new hybrid model by combining complete ensemble empirical mode decomposition (CEEMD) and LSTM, which achieved a better performance than a single LSTM model and other hybrid models such as CEEMD-SVR, and EMD-LSTM [33]. Niu et al. introduced a hybrid stock price index forecasting model based on variational mode decomposition (VMD) and LSTM, which shows obvious advantage over some single models as well as the EMD-based or VMD-based hybrid models [34]. Wang et al. proposed a novel hybrid model for wind power interval prediction based on GRU and VMD, and validated the effectiveness of the method [35]. Liu et al. combined VMD and ANN to realize ensemble forecasting of product prices [36]. Xiang et al. proposed the dynamic multi-factor similarity measurement (DMFSM). This method effectively addresses the issues of data singularity and correlation in stock price series by integrating dynamic time warping (DTW) with Mahalanobis distance, while also considering the weights of each node in multidimensional time series. As a result, it enhances the accuracy of stock price prediction [37]. With competence in handling big data and modeling nonlinear relationship between attributes, deep learning methods have been verified the superiority over conventional statistical and machine learning methods in many studies on stock market prediction [38].

In summary, most of the above methods have achieved excellent results, but they still have these three deficiencies. 1) Stock price prediction is influenced by a variety of factors, including market sentiment, company performance, investor behavior, as well as external uncertatinties such as macroeconomic conditions, national policies, natural disasters, and social events. This can lead to a significant amount of noise in the input data, greatly affecting the accuracy of predictions. 2) For stock price series with different fluctuation trends, training only one prediction model will inevitably result in the inaccurate prediction results. The price trends of different stocks are quite different and diverse, making it difficult for a single prediction model to capture all the characteristics of them. When constructing a reasonable similarity measurement method to consider that the similarity among different time series may vary in weight. However, traditional clustering methods are not capable of handling data with high complexity and are ineffective in dealing with the peculiarities, time shifts, and warps in time series data. There is a need for more efficient clustering algorithms to enhance the effectiveness of clustering. 3) Traditional RNN architectures, including LSTM networks, have been widely used for time series data analysis due to their ability to capture temporal dependencies. However, these models often struggle with capturing long-term dependencies effectively, as they are prone to issues like vanishing and exploding gradients. This can lead to suboptimal performance, especially in complex, long-range sequential data such as stock price trends. Moreover, the computational cost of training these models can be high, and their data processing efficiency is comparatively slow.

To address these two issues, we propose a personalized model based on simple recurrent unit (PSRU) for accurate prediction of time series in terms of stock closing price in this paper. Specifically, the proposed framework consists of two steps. 1) To reduce the noise in the input time series data, we propose an improved variational mode decomposition (VMD) with parameter determination, and apply it to denoising. VMD has shown better robustness and effectiveness in denoising than many other methods [39], and it can effectively separate the useful components of a signal from the noise. Compared with other denoising methods such as empirical mode decomposition (EMD), VMD can avoid the problem of mode aliasing, and its adaptive nature allows for the dynamic adjustment of the number of decompositions based on the characteristics of the data, thereby further enhancing the denoising effect. To further improve its performance, we propose a method where the number of decompositions in VMD can be adaptively determined by calculation of the average sample entropy and cross-correlation. 2) Clustering can help us identify different feature patterns in time series, thereby exploring the characteristics of diverse time series and fitting the optimal prediction model for each pattern. We cluster the time series into multiple classes, and each class fits an optimal prediction model. In particular, we employ dynamic multi-perspective personalized similarity measurement (DMPSM) clustering, which is based on our previous work [40]. DMPSM can adaptively adjust the similarity measurement, clustering sequences with similar features together, thereby fitting the optimal prediction model for each clustering category and enhancing the overall forecasting accuracy. DMPSM demonstrates superior performance in addressing issues such as singularities, time shift, and warping in time series data, and is more effective than traditional Euclidean distance and Canberra distance [40]. Therefore, choosing DMPSM as the clustering method enhances the model's ability to recognize different features of time series. 3) After clustering, SRU is adopted as the prediction model considering its excellent mapping and parallel processing capability. SRU is particularly well-suited for processing time series data due to its structural design, which effectively captures long-term dependencies and complex dynamic patterns within sequences. Additionally, the SRU model has a computational efficiency advantage, allowing it to process larger datasets more quickly compared to traditional RNNs and LSTMs.

Overall our work makes the following contributions. 1) A novel adaptive parameter determination method based on average sample entropy and cross-correlation is proposed. It can effectively improve the decomposition accuracy of VMD, while reducing computational load and enchancing the model's robustness in noisy environments. 2) A novel personalized similarity measurement is used for time series clustering. DMPSM can better solve the problems of singularities, time shifts and warping and personalized characteristics in time series, thus greatly improving the effect of clustering. By employing a rational clustering strategy, the model is able to more accurately capture the characteristics of each sequence category. 3) SRU is applied to stock price trend prediction for the first time. To the best of our knowledge, few literature has applied SRU to the financial field at present. Therefore, it is of scientific significance to investigate whether SRU is suitable for stock price trend prediction with high prediction accuracy and efficiency. To verify the effectiveness of the proposed PSRU, three experiments are conducted on 285 stocks from the Shanghai Stock Exchange. The experimental results show that PSRU can achieve better performance than other popular time series prediction methods.

2. Methods and Work-Flow.

In this section, we describe relevant methods and work-flow of our PSRU model. The first three sections include the introductions of VMD and parameter determination (see Section 2.1), DMPSM (see Section 2.2) and SRU (see Section 2.3). The last section describes the specific work-flow of the proposed model (see Section 2.4).

2.1. Variational mode decomposition (VMD) and parameter determination.

Variational mode decomposition (VMD), proposed by Dragomiretskiy and Zosso [41], is a powerful non-recursive signal decomposition method. VMD could concurrently decompose an input data series into multiple sub-series (i.e., modes) by obtaining the optimal solution of the constrained variational model [42]. At present, VMD has been employed in many fields, such as image segmentation [43], crude oil price prediction [44], and economic and financial time series prediction [45].

VMD decomposes a complete original signal $f(t)$ into K principal modes (u_k), called intrinsic mode function (IMF), and each mode has a limited bandwidth and a unique center frequency (ω_k) in the frequency domain. Then, the decomposition process via VMD can be formulated as the optimization problem of variational model. The construction of variational model is facilitated through minimizing the total bandwidth of all modes. Meanwhile, the sum of all modes must be equal to the original signal $f(t)$, which is the constraint.

To estimate the bandwidth of each mode, three steps are executed. 1) Each mode is transformed by Hilbert transform to obtain its analytical signal and the corresponding unilateral frequency spectrum. 2) The frequency spectrum of each mode is modulated to respective baseband by applying an exponential term about the estimated center frequency. 3) The bandwidth of each mode is estimated by L2-norm of the demodulated signal gradient.

After the estimation of the bandwidth, the constrained variational problem can be formulated as

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}$$
$$\text{s.t. } \sum_{k=1}^{K} u_k(t) = f(t) \tag{1}$$

where $f(t)$ is the original signal; $\{u_k\} = \{u_1, u_2, \ldots, u_k\}$ is the set of all modes where u_k represents the kth mode; $\{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_k\}$ are the corresponding center frequencies of the modes; K is the total number of modes; $\delta(t)$ denotes the Dirac distribution; j is the imaginary unit; $*$ denotes the convolution operator.

To solve the above constrained variational problem, Lagrange multiplier and a quadratic penalty term are introduced to transform the constrained problem into an unconstrained one which can be expressed as

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^{K} \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{2}$$

where $\lambda(t)$ is the Lagrange multiplier used to tighten the constraint; α is the balancing parameter of the data fidelity constraint; $\left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2$ is a quadratic penalty term to accelerate convergence [46].

The optimization problem in Equation (2) can be solved by utilizing alternate direction method of multipliers (ADMM). In this way, we can obtain the iterative formulas for updating u_k and ω_k. In addition, the iterative convergence condition is $\sum_{k=1}^{K} \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2 \Big/ \left\| \hat{u}_k^n \right\|_2^2 < \gamma$, where γ denotes the convergence tolerance and n represents the number of iterations. Hence, the solutions for u_k, ω_k and $\lambda(t)$ are calculated as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k, i=1}^{K} \hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega} \tag{4}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \tau\left[\hat{f}(\omega) - \sum_{k=1}^{K}\hat{u}_k^{n+1}(\omega)\right] \tag{5}$$

where $\hat{u}_k^{n+1}(\omega)$, $\hat{f}(\omega)$, $\hat{u}_i(\omega)$ and $\hat{\lambda}(\omega)$ are the Fourier transforms of $u_k^{n+1}(t)$, $f(t)$, $u_i(t)$ and $\lambda(t)$, respectively, and τ is the time step of the dual ascent.

VMD can capture the characteristics of nonlinear signals, and it has been proved to be more effective than traditional algorithms such as empirical mode decomposition (EMD) and wavelet decomposition (WD) [47]. However, VMD lacks a selection rule for the number of decomposition modes K, which affects the accuracy of decomposition [48]. If K is too small, then the original signal is not completely decomposed, and the modes with insufficient information may reduce the prediction accuracy. On the contrary, if K is too large, the original signal will be over decomposed. Too many decomposed modes not only lead to the decline of prediction accuracy, but also cause a lot of unnecessary computation [49]. Therefore, the determination of decomposition mode number K is critical.

At present, a widely used way to determine K utilizes EMD to estimate the parameter according to the frequency distribution of modes after decomposition [50]. However, when the mode aliasing of EMD is serious, it is difficult to distinguish the independent frequency components from the frequency distribution to accurately estimate K. Other methods such as genetic algorithm [51], and signal-energy based rule [49] are not effective to determine K.

Thus to improve the accuracy of decomposition, we propose a novel automatic parameter determination method for VMD, which can adaptively determine the value of K through the average sample entropy and cross-correlation. Next, we describe the proposed method in detail.

First, VMD is applied to the time series with various numbers of decomposition modes. For each configuration, sample entropy is calculated for each intrinsic mode function (IMF), with the highest entropy IMF treated as noise while the remaining IMFs are retained as signal modes. Then, ASE of the retained modes is computed, where a lower ASE indicates a higher decomposition quality. Cross-correlation is further calculated between the signal and noise modes to assess how well different K values preserve the original signal. When cross-correlation is minimized, it suggests a lower similarity between noise and signal, supporting the current K value. By comparing ASE and cross-correlation across different values, the K that minimizes both is selected as optimal.

Compared to traditional empirical mode decomposition (EMD), this dual-metric approach significantly enhances decomposition accuracy. EMD often struggles with mode aliasing, making it difficult to separate noise from signal effectively. By combining ASE and cross-correlation as criteria, this method reduces noise more precisely and avoids iterative trial-and-error in determining K, thus improving computational efficiency. This optimized approach better preserves key signal characteristics, providing more accurate inputs for subsequent prediction tasks.

Sample entropy is effective to measure the complexity of time series. Suppose a time series y with N data, i.e., $y = [y_1, y_2, \ldots, y_N]$. VMD is used to decompose y into K modes, and set the kth mode as $u_k = [u_k(1), u_k(2), \ldots, u_k(N)]$. We form the $N - m + 1$ vectors $U(i)$ as

$$U(i) = [u_k(i), u_k(i+1), \ldots, u_k(i+m-1)] \tag{6}$$

where $i = 1, 2, \ldots, N-m+1$ and m is the length of series to be compared. The maximum distance between the two vectors is defined as

$$d_m[U(i), U(j)] = \max_{l=0,1,\ldots,m-1} |u_k(i+l) - u_k(j+l)| \tag{7}$$

where $j = 1, 2, \ldots, N - m$ and $i \neq j$. Then, we define the function

$$B_i^m(r) = \frac{B_i}{N - m - 1} \tag{8}$$

where B_i is the number of $d_m[U(i), U(j)] < r$, and r is the tolerance for accepting matrices. The average value of B_i^m is calculated as

$$B^m(r) = \sum_{i=1}^{N-m} \frac{B_i^m(r)}{N - m} \tag{9}$$

where $B^m(r)$ represents the probability that two series will match for m points. Similarly, we can derive another expression

$$B^{m+1}(r) = \sum_{i=1}^{N-m} \frac{B_i^{m+1}(r)}{N - m} \tag{10}$$

where $B^{m+1}(r)$ represents the probability that two series will match for $m + 1$ points. Finally, the sample entropy of mode series is defined as

$$SampEn(m, r) = \lim_{N \to \infty} \left\{ -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \right\} \tag{11}$$

which is estimated when the value of N is limited

$$SampEn(m, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \tag{12}$$

From Equation (12), the sample entropy is related to the parameters m and r. According to [52], the results obtained when m is 1 or 2 and r is 0.1-0.25STD (STD is the standard deviation of series data) have reasonable statistical characteristics, and the changing trend of sample entropy is not affected by the parameters. Therefore, we set $m = 2$, $r = 0.2$STD in this paper.

After decomposition by VMD, each mode gets its own center frequency, which indicates that the complexity of each mode becomes low and the corresponding sample entropy is small. For an optimal number of decompositions, the sample entropy of each mode after removing the residual modes (noise component and the last component) reaches the minimum. Then the average sample entropy (ASE) obtained by averaging the sample entropy of these modes is smaller than the ASE of all modes under other decomposition numbers. Therefore, ASE can be used as an indicator to reflect whether the value of K is reasonable or not.

The cross-correlation function can measure the correlation between two time series at any different time. Suppose that the sum series of the residual modes obtained after decomposition by VMD is $c(t)$ and the new series reconstructed by other series is $g(t)$. The cross-correlation function of the two series is calculated as

$$R_{cg}(\tau) = \sum_{t=0}^{N} g(t)c(t - \tau) \tag{13}$$

Normalize $R_{cg}(\tau)$ to obtain $\rho_{cg}(\tau)$

$$\rho_{cg}(\tau) = R_{cg}(\tau) \Big/ \sqrt{R_{gg}(0)R_{cc}(0)} \tag{14}$$

where $R_{gg}(0)$ and $R_{cc}(0)$ are the autocorrelation functions of two series at the same time. The autocorrelation function of series $v(t)$ is

$$R_{vv}(\tau) = \sum_{t=0}^{N} v(t)v(t-\tau) \tag{15}$$

The standard cross-correlation function ranges from 0 to 1. A higher cross-correlation value indicates higher similarity. When the number of decompositions is optimal, the cross-correlation between the sum series of the residual modes and the reconstructed series reaches the minimum. Thus, we use cross-correlation value to measure the rationality of the K value selection. When the cross-correlation value is small, the number of decompositions is better. Otherwise, the K value is inferior.

2.2. **Dynamic multi-perspective personalized similarity measurement (DMPSM).** Euclidean distance is widely used in time series analysis due to its fast computing speed and low complexity. However, its biggest drawback is that it is very sensitive to singularities, which negatively impacts the measurement of similarity in time series. To overcome the drawbacks of Euclidean distance, many methods have been proposed. One of the most effective is Canberra distance. Canberra distance is a dimensionless measure and is insensitive to singularities. However, its biggest drawback is its inability to effectively handle time shifts and distortions. To address the issues of time shifts and distortions, many methods have been developed in recent years, among which the most commonly used is dynamic time warping (DTW). DTW measures the similarity between time series of different lengths by constructing an optimal warping path that minimizes the distance between the two segmented sequences. Unlike traditional similarity measurement methods, DTW allows for data matching through a one-to-many mechanism, effectively handling time shifts and distortions. Although DTW addresses the point-to-point matching issue in time series, it still employs Euclidean distance to construct the distance matrix, which means it cannot reduce the negative impact of singularities.

To accurately measure the similarity between a pair of time series, Zhao et al. proposed the dynamic multi-perspective personalized similarity measurement (DMPSM) [40]. The DMPSM consists of two key steps which will be elaborated as follows.

The first step of DMPSM is to assign different weights to the segmented time series according to the principle that as the factor approaches the current time, the assigned weight increases gradually. Specifically, suppose the weights assigned to the time series $X = (x_1, x_2, \ldots, x_n)$ are $W = (\omega_1, \omega_2, \ldots, \omega_n)$, where $\omega_1 < \omega_2 < \cdots < \omega_n$ and $\omega_1 + \omega_2 + \cdots + \omega_n = 1$, i.e., $X' = (\omega_1 x_1, \omega_2 x_2, \ldots, \omega_n x_n) = (x'_1, x'_2, \ldots, x'_n)$. The setting of weights fully reflects the characteristics of time, i.e., the closer the factor is to the current time, the higher the information value of the series.

The second step is to embed Canberra distance [53] into the dynamic time warping (DTW) [54] to reflect the similarity of time series. Suppose $Y = (y_1, y_2, \ldots, y_m)$ is a time series different from X, and $Y' = (y'_1, y'_2, \ldots, y'_m)$ is its weighted series. The DTW matrix is constructed between X' and Y' by using Canberra distance, i.e.,

$$D'(p_k) = \begin{bmatrix} d^C(x'_1, y'_1) & \cdots & d^C(x'_1, y'_m) \\ \vdots & \ddots & \vdots \\ d^C(x'_n, y'_1) & \cdots & d^C(x'_n, y'_m) \end{bmatrix}_{n\times m} \tag{16}$$

where $d^C\left(x'_i, y'_j\right) = \frac{|x'_i - y'_j|}{|x'_i| + |y'_j|}$ $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$. $P = (p_1, p_2, \ldots, p_k)$ $(1 \leq k \leq K$, $\max(n, m) \leq K \leq n + m - 1$, $K \in Z)$ is a warping path, which means the mapping information derived from X' and Y'. p_k is the matching relation between x'_i and y'_j, i.e., $d'(p_k) = d^C\left(x'_i, y'_j\right)$. Then, the final calculation formula of DMPSM can be described as

$$d^{DMPSM} = \min\left\{\sum_{k=1}^{K} d'(p_k)\right\} \tag{17}$$

According to the work-flow of DMPSM, we can summarize three advantages of it over other similarity measurements in time series clustering. To show these points more intuitively, Figure 1 compares four different similarity distances between two time series in three cases, namely Euclidean distance, Canberra distance, DTW and DMPSM. 1) In Figure 1(a), it can be seen that series 2 has a singularity, and the distance value calculated by DMPSM is less than the other three similarity measurements, which indicates that DMPSM can effectively eliminate the negative influence of singularities. 2) Figure 1(b) shows that the two series have time shifts and warping, while the distance obtained by DMPSM is still the smallest. This indicates that DMPSM can deal with time shifts and warping, and realize point-to-point matching of time series. 3) In Figure 1(c), the first half of the two series is more similar than the second half, and the comparison of the calculation results shows that DMPSM can well reflect the personalized characteristics of time series.

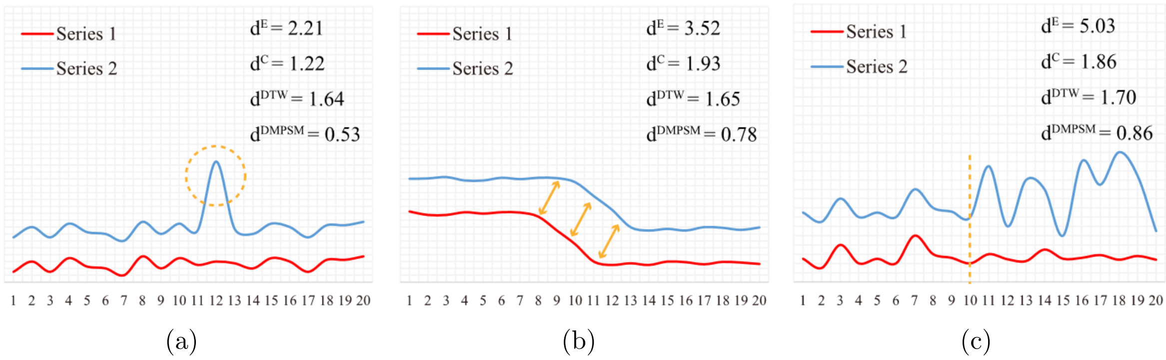


FIGURE 1. Comparison of four similarity measurements in three cases

2.3. **Simple recurrent unit (SRU).** Recurrent neural network (RNN) is a widely-used model for processing time series data. However, it cannot effectively explore historical information. Additionally, it suffers the problem of gradient vanishing or gradient exploding. To overcome these limitations, researchers have developed several variants of RNN among which long short-term memory (LSTM) [55] and gated recurrent unit (GRU) [28] stand out. These variants have been applied to stock series prediction and achieved good performance. However, although LSTM and GRU have made significant progress in capturing long-term dependencies, their parallelization capabilities during training are limited, and they are much slower than convolutional and attention-based models [56].

In 2017, Lei and Zhang proposed simple recurrent unit (SRU) [31], which not only maintains the performance, but also carries out parallel processing to improve the running speed. The SRU neuron module contains a forget gate and a similar reset layer including skip connections. Figure 2 illustrates the internal structure of SRU neuron module. The calculation formulas of internal information states in SRU are as follows.

$$\tilde{x}_t = W x_t \tag{18}$$
$$f_t = \sigma(W_f x_t + b_f) \tag{19}$$
$$r_t = \sigma(W_r x_t + b_r) \tag{20}$$
$$c_t = f_t \odot c_{t-1} + (1 - f_t) \odot \tilde{x}_t \tag{21}$$
$$h_t = r_t \odot \tanh(c_t) + (1 - r_t) \odot x_t \tag{22}$$

where x_t denotes the input of the module at time t, f_t represents the output of the forget gate at time t, r_t is the output of the reset gate at time t, c_t denotes state information of the module at time t, and h_t represents the output of the module at time t. $\odot$ represents the element-wise product. W, W_f and W_r are weight vectors. b_f and b_r are bias vectors. σ is the sigmoid activation function, which is defined as $\sigma(x) = \frac{1}{1+e^{-x}}$. tanh is hyperbolic tangent activation function, and its expression is $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

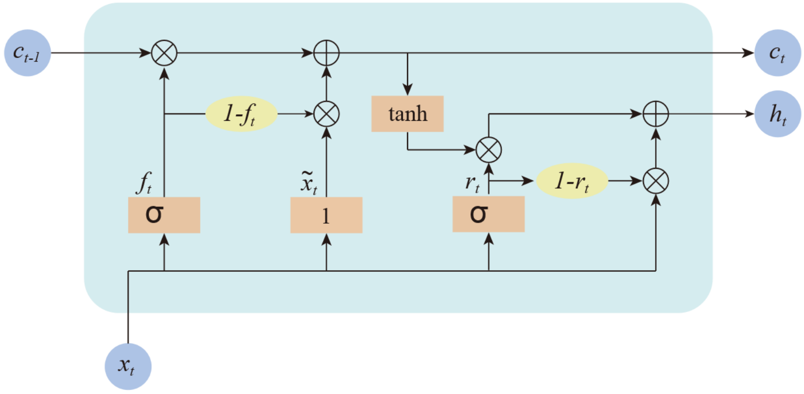


FIGURE 2. Structure of SRU

From Figure 2 and the above formulas, it can be derived that the operations performed by Equation (18), Equation (19) and Equation (20) only depend on the current input rather than the previous output state, which implies that these operations can be calculated in parallel. Therefore, SRU can exhibit a running time similar to convolution, and achieve 5-10x speed-up over an optimized LSTM. At the same time, it can also maintain or slightly improve the performance.

2.4. **Work-flow of PSRU.** To address the issues described in the introduction and predict stock price trend more accurately, we propose a novel time series prediction model, named as PSRU. Figure 3 illustrates the work-flow of the proposed prediction model. There are four steps which are described in detail as follows.

Step 1) Segmentation. Each stock price series is segmented into many small pieces of time series. We preset the length of the segmented series (p) and the step size during sliding segmentation (q).

Step 2) Denoise. All the segmented time series are decomposed using the improved VMD, and the new time series are reconstructed after removing the residual modes. The selection of the number of decomposition modes has been introduced in Section 2.1. Firstly, we calculate the sample entropy of each mode under different decomposition numbers according to Equation (12), so as to obtain the residual modes and reserved modes in each case. Then, ASE can be calculated by averaging the sample entropy of reserved modes, and cross-correlation can be calculated by Equation (14). When these two values reach the minimum, the corresponding decomposition reaches the optimal value. Finally, with the optimal decomposition, the sum series of the reserved modes is the denoised time series.

Step 3) Clustering. We cluster the reconstructed time series into multiple classes by measuring the personalized similarity between any two series. The specific introduction of DMPSM and its advantages in time series clustering are given in Section 2.2. According to Equation (17), the distance between any pair of series can be calculated, and then clustering can be carried out according to the distance.

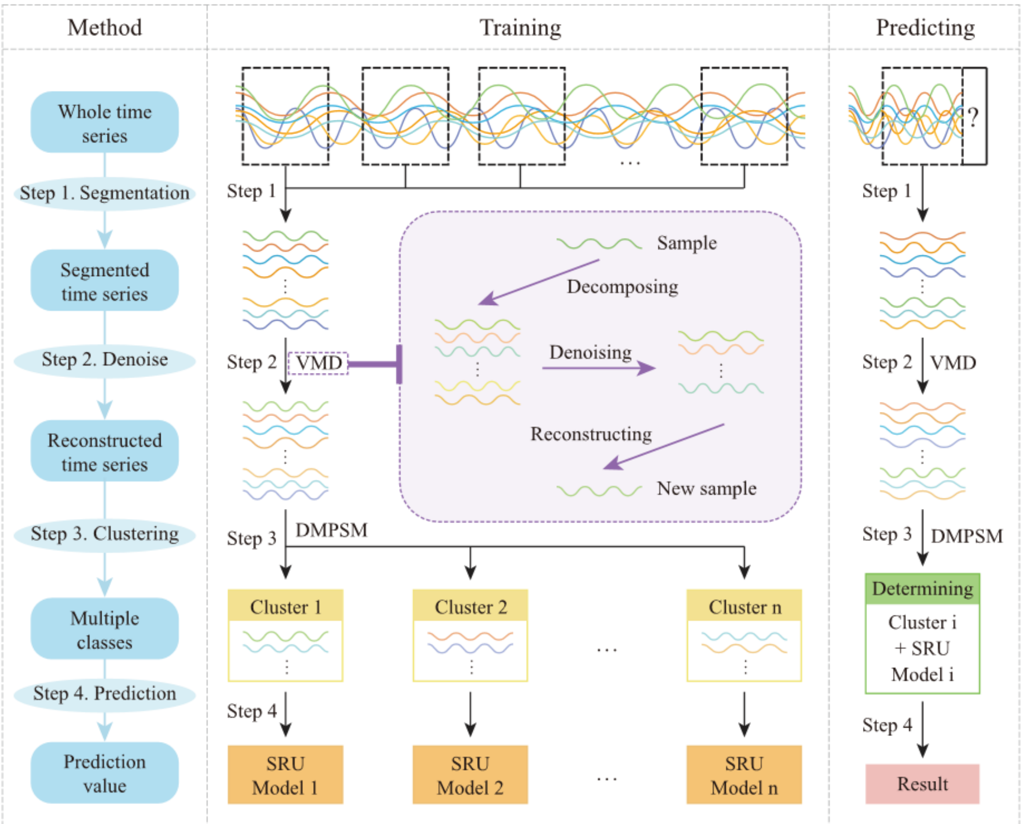


FIGURE 3. Work-flow of PSRU

Step 4) Prediction. The optimal SRU model in each class is trained on the basis of denoised time series data for prediction. The internal structure of SRU model has been provided in detail in Section 2.3, in which the specific information state has been shown in Equations (18) to (22). We train the model to get the most appropriate parameters for series of each class, so as to establish the corresponding optimal SRU model.

After training, we can do prediction of stock price series with the following operations. Assume there are t time points to be predicted. Firstly, the last segmented series of each stock price is intercepted, and the series length is $p-t$. Then, the improved VMD is used to process the segmented series into new denoised ones. Next, the similarity is calculated by DMPSM to determine the class that a time series belongs to. Specifically, the mean series of multiple denoised series in each class is taken as the representative series, and the similarity between the series to be predicted and each representative series is calculated to determine the category. Finally, the trained SRU model in this class is used to obtain the prediction results.

Based on the work-flow of the proposed prediction model, we can make the following analyses. 1) The time series segmented into small pieces can increase the sample size, and reflect their own characteristics. Moreover, due to the frequent historical repetition of stock price series, segmentation is more conducive to the subsequent clustering processing. 2) After decomposition and reconstruction by VMD, the denoised time series are helpful to train the subsequent model. The proposed parameter determination method in VMD can effectively improve the decomposition accuracy and reduce computation cost. 3) Clustering strategy can make the model effectively extract the characteristics of series in each class, which improves the prediction accuracy of the model and reduces the model complexity. Meanwhile, DMPSM can handle singularities, time shifts and warping and personalized characteristics in time series, which greatly promotes the effect of clustering. 4) The reason why SRU is selected as the final prediction model is due to its high prediction accuracy and efficiency. In particular, the fast prediction speed of the SRU model greatly makes up for the increase of prediction time brought by clustering strategy.

3. **Experiments and Analysis.** In this section, we carry out three experiments based on the stock data. Firstly, we introduce the experiment data in Section 3.1. Then, we conduct a comparative experiment on the prediction performance of different methods in Section 3.2. Next, Section 3.3 shows the specific process of determining decomposition number. Finally, the comparison of prediction performance under different clustering settings is conducted in Section 3.4. The results and analysis of all experiments are also shown in their respective sections.

3.1. **Data acquisition and preprocessing.** The closing prices of 285 stocks in 2016 from the Shanghai Stock Exchange are used in our experiments. The date range for data collection is the full trading day of the whole year. Since the overall price range of each stock is different, we preprocess the data as follows.

(a) Data normalization. We normalize the data by using the growth rate of each stock, which is calculated as

$$x'_{i+1} = \frac{x_{i+1} - x_i}{x_i} \quad i = 1, 2, \dots \tag{23}$$

where x_i denotes the closing price of the stock on the ith day, x_{i+1} denotes the closing price of the stock on the $i+1$th day, and x'_{i+1} represents the growth rate of stock.

The growth rate-based normalization method was chosen due to the high volatility and large price differences across stocks. By standardizing with growth rates, the model can focus on capturing relative growth trends rather than absolute price fluctuations, effectively removing the impact of different price scales among stocks. This method is particularly suitable for time series prediction, as it preserves each stock's relative trend, ensuring comparability across various stocks.

For example, consider two stocks: Stock A, priced between 20 and 50 units, and Stock B, priced between 200 and 500 units. Without normalization, the model might focus more on the larger fluctuations of Stock B due to its higher price range. However, by using growth rate normalization, both stocks' relative price changes are comparable. For instance, if Stock A increases from 30 to 32, the growth rate is $(32-30)/30 = 6.67\%$, while Stock B's growth from 300 to 318 results in a growth rate of $(318 - 300)/300 = 6.00\%$. This allows the model to treat both stocks equally, regardless of their price levels.

Additionally, using growth rate normalization reduces the effect of extreme price variations, improving the model's generalization across diverse stock data, leading to more stable and accurate predictions. This approach retains essential dynamic features within the training data, enhancing the model's ability to capture future price trends more effectively.

(b) Data screening. Stocks only open on weekdays, while weekend data are presented using Friday's closing price. Therefore, we take out the value of zero as the weekend growth rate. In this way, we can obtain 365 growth rate data for each stock.

3.2. **Prediction performance of different methods.** To explore the performance of PSRU, we compare performance of different methods in terms of the prediction error and running time. These competing methods include PSRU, VMD-DMPSM-GRU, VMD-DMPSM-LSTM, EMD-DMPSM-SRU, EMD-DMPSM-GRU, and EMD-DMPSM-LSTM. All the methods are tested on the same benchmark, and the specific setups of the experiment are as follows. Firstly, all stock price series are sliced into small pieces. We set the length of the segmented series to 20 and the step size during sliding segmentation to 5, i.e., $p = 20$ and $q = 5$. At the same time, we take the last segmented series of each stock price as the test series, and we predict the last one step closing price of all stocks. At the same time, we split the train, validation and test set in a ratio of 6 : 2 : 2. Then, after all the segmented series have been denoised by the improved VMD or EMD, the series for training are divided into five classes to train different prediction models, respectively. Finally, each test series directly uses the corresponding prediction model to complete the prediction after its class has been determined.

In our experiment, mean absolute error (MAE) and root mean square error (RMSE) are adopted as the performance evaluation metrics. Their expressions are given in Table 1, where $Y = (y_1, y_2, \ldots, y_n)$ denotes the real values, and $\hat{Y} = (\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$ denotes the predicted values. Smaller values of these metrics indicate higher prediction accuracy. In addition, prediction time is also adopted as an evaluation metric.

Table 1. Expressions of evaluation metrics

Metrics	Expressions
MAE	$\frac{1}{n}\sum_{i=1}^{n} \lvert\hat{y}_i - y_i\rvert$
RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$

Table 2 shows the prediction performance of different methods, in which the rows represent different prediction methods and the columns represent three indexes. Bold digits represent the minimum value, i.e., the best performance.

Table 2. Prediction performance of different methods

	Denoising	Forecasting	MAE	RMSE	Time (ms)
PSRU (our proposed)	VMD		**0.006658**	**0.009319**	**0.008524**
EMD-DMPSM-SRU	EMD	SRU	0.008931	0.017656	0.008563
SRU	×		0.012233	0.027449	0.008577
VMD-DMPSM-GRU	VMD		0.006803	0.009473	0.008803
EMD-DMPSM-GRU	EMD	GRU	0.009010	0.017989	0.008795
GRU	×		0.012400	0.027703	0.008755
VMD-DMPSM-LSTM	VMD		0.007898	0.010812	0.009366
EMD-DMPSM-LSTM	EMD	LSTM	0.010352	0.020068	0.009341
LSTM	×		0.015785	0.029545	0.009300

From Table 2, we can make the following observations. 1) The prediction error and prediction time of our PSRU model are the smallest among all methods, indicating its superiority over other methods. 2) The prediction model combined with the improved VMD (PSRU, VMD+DMPSM+GRU, VMD+DMPSM+LSTM) performs better than the prediction model combined with EMD (EMD+DMPSM+SRU, EMD+DMPSM+GRU, EMD+DMPSM+LSTM) and the prediction model without denoising method (SRU, GRU, LSTM), which shows the effectiveness of data denoising using the improved VMD, because it helps to eliminate noise and provide a clearer dataset for the subsequent training of the prediction model. 3) Compared to LSTM and GRU model, SRU-based models

(PSRU, EMD+DMPSM+SRU, SRU) demonstrate faster prediction speeds. This indicates that SRU has higher predictive efficiency, which can be attributed to its parallel processing capabilities. The prediction times for SRU-based models are consistently lower than those for LSTM and GRU, confirming the speed advantage of SRU.

Based on the above observations, we can make the following analysis. Firstly, after the decomposition and reconstruction of the time series by the improved VMD, most of the noise has been eliminated, and the denoised data is beneficial for the training of the subsequent prediction model. Meanwhile, VMD is better than EMD in solving some problems of mode aliasing. Therefore, we can find that the performance of the prediction model with the VMD is improved over the baseline. In addition, the faster prediction speed of the SRU model, as evidenced by the lower time values in the table, is a result of its parallel processing architecture. This characteristic makes SRU particularly suitable for applications where prediction speed is critical. With these combined advantages, the proposed PSRU model achieves the best performance among all these methods, offering a balance of high accuracy and rapid prediction speeds.

3.3. **Determination of decomposition number.** In order to clearly show how the number of decompositions is determined, we randomly select a time series to describe the process in detail. Firstly, the original series is decomposed by VMD, and the sample entropy is calculated for each mode under different decomposition numbers. Figure 4 shows the sample entropy of each mode when the number of decompositions is 2, 3, 4, 5, 6, 7 and 8, respectively. From Figure 4, we can get the mode with the largest sample entropy under each decomposition number, that is, the noise mode. Accordingly, we can obtain the residual modes and reserved modes under each decomposition number, and the specific situation is shown in Table 3.

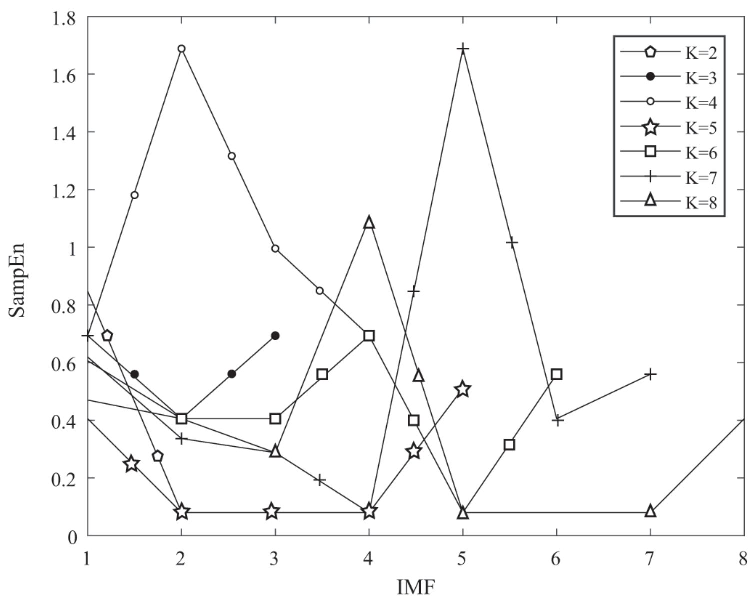


FIGURE 4. Sample entropy of modes under different K

Then, according to the results in Table 3, the average sample entropy (ASE) and cross-correlation can be calculated to determine the optimal number of decompositions. ASE is obtained by averaging the sample entropy of reserved modes, and cross-correlation is obtained by calculating the cross-correlation value between the sum series of residual modes and the sum series of reserved modes. Figure 5 and Figure 6 illustrate ASE and

Table 3. Residual modes and reserved modes under different K

K	Residual modes	Reserved modes
2	IMF2	IMF1
3	IMF3	IMF1+IMF2
4	IMF2+IMF4	IMF1+IMF3
5	IMF5	IMF1+IMF2+IMF3+IMF4
6	IMF4+IMF6	IMF1+IMF2+IMF3+IMF5
7	IMF5+IMF7	IMF1+IMF2+IMF3+IMF4+IMF6
8	IMF4+IMF8	IMF1+IMF2+IMF3+IMF5+IMF6+IMF7

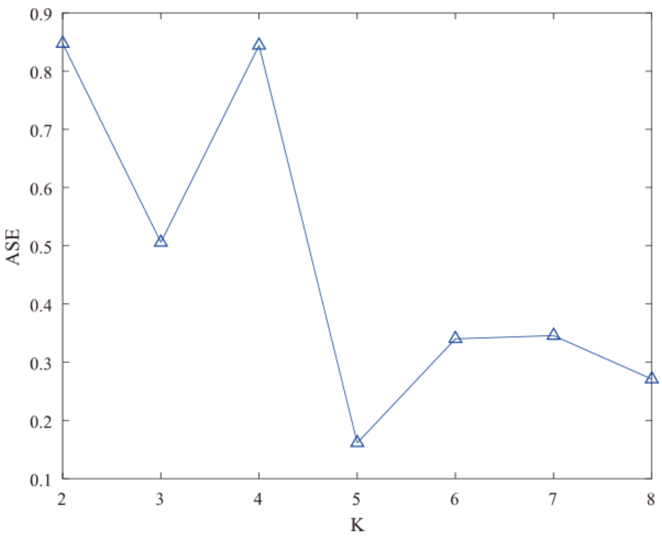


Figure 5. ASE under different K

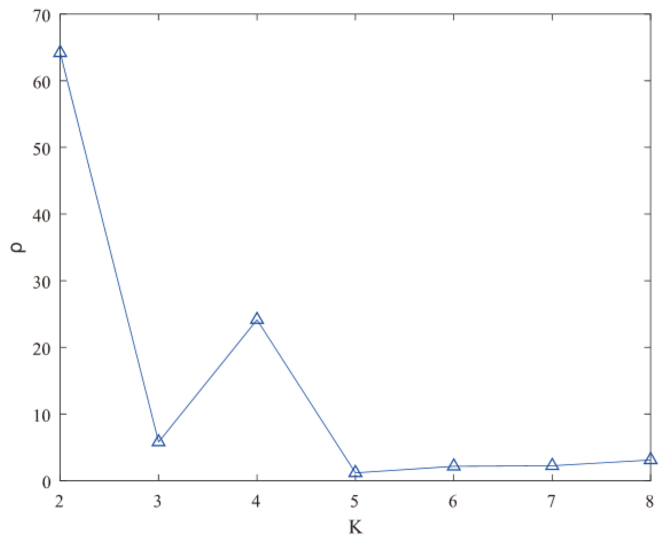


Figure 6. Cross-correlation under different K

cross-correlation under different decomposition numbers. Note that all values of cross-correlation in Figure 6 are multiplied by 10,000 for easier comparison.

Finally, it can be seen from Figure 5 and Figure 6 that when $K = 5$, ASE and cross-correlation reach the minimum value. Therefore, the original series is decomposed into five modes, all of which are shown in Figure 7. Meanwhile, we can also add the reserved modes to obtain the corresponding reconstructed series. The original series and the reconstructed series are shown in Figure 8. From Figure 8, the new reconstructed series alleviates the influence of some singular points and makes the original series more stable, which indicates that the improved VMD is effective for denoising the data.

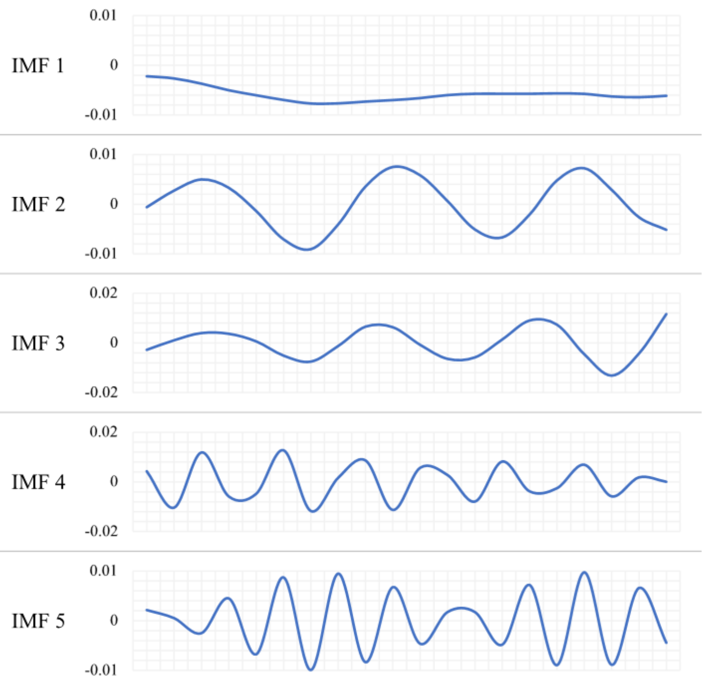


FIGURE 7. Five decomposition modes of the original series

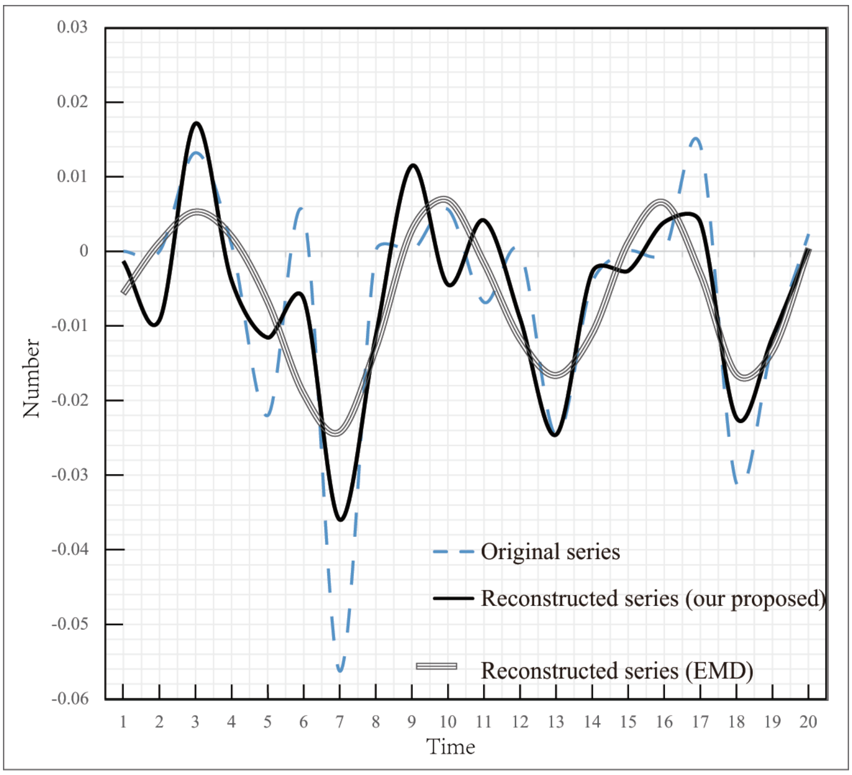


FIGURE 8. Original series and reconstructed series

To further test the proposed parameter determination method in VMD, we compare it with the method of using EMD to estimate the number of decompositions. The above series is also used as an example for detailed description. EMD directly decomposes the original series into 4 IMF, so it is considered that the number of decompositions in VMD is 4. After the parameter is determined, the formal decomposition is carried out. Similarly, after removing the residual mode, the remaining modes are reconstituted into a new series, which is also shown in Figure 8.

From the comparison between the bold line and the three line in Figure 8, the reconstructed series obtained by using the proposed parameter determination method is more consistent with the original series than the reconstructed series obtained by using EMD to estimate the parameter, which also eliminates the influence of some extreme values. It indicates that the VMD based on our proposed parameter determination method can effectively denoise and preserve the information of the original series. This can verify that the decomposition and reconstruction effect of our proposed improved VMD can effectively generate denoised data, thus improving the prediction accuracy.

3.4. **Prediction performance under different clustering settings.** In order to highlight the effectiveness of clustering strategy in our proposed method, we first compare the prediction errors in different cluster numbers. We list the MAE and RMSE of the prediction results when the cluster numbers are 4, 5, 6, 7, 8, 9 and no clustering, respectively, in Table 4. The rows represent two error indexes, while the columns represent different clustering situations. Bold digits in each row represent the minimum value of the indexes, i.e., the best prediction performance.

Table 4. Prediction performance in different cluster numbers

	No clustering	4	5	6	7	8	9
MAE	0.008349	0.007483	**0.006658**	0.007154	0.007576	0.007256	0.007226
RMSE	0.013721	0.009998	**0.009319**	0.010191	0.010415	0.010361	0.010168

From Table 4, 1) the effect of clustering prediction is significantly better than that without clustering, which highlights the effectiveness of clustering strategy. 2) When the number of clusters is 5, the prediction error is the smallest. It means that finding an optimal number of clusters can also increase the prediction accuracy to a certain extent.

The reason for this result is that clustering strategy makes the prediction model capture the characteristics of diverse and complex series well. The optimal prediction model trained by similar series in each class typically performs better than the optimal prediction model trained by all series. Moreover, too few or too many clusters also affect the prediction results. Specifically, too few clusters cannot fully reflect the characteristics of series in each class, and cause loss of information and decline of prediction accuracy. On the contrary, too many clusters usually cause unnecessary calculations without obvious accuracy improvement. Therefore, it is essential to find an appropriate number of clusters.

To show the advantages of DMPSM in clustering, we compare the prediction errors under clustering methods with different similarity measurements. In Table 5, we list the MAE and RMSE of the prediction results using Euclidean distance, Canberra distance, DTW and DMPSM in clustering. The meaning of the rows and columns in the table is the same as that in the previous table.

Based on Table 5, we can see that the final prediction error after clustering with DMPSM is smaller than that after clustering with the other similarity measurements, which shows the superiority of using DMPSM in clustering. This result can be explained by the advantages of DMPSM in measuring the similarity between two series, which have

Table 5. Prediction performance under different clustering methods

	Euclidean distance	Canberra distance	DTW	DMPSM
MAE	0.007723	0.007261	0.006966	**0.006658**
RMSE	0.010687	0.010396	0.009931	**0.009319**

been confirmed in detail in Section 2.2. In view of the complexity of stock series, the personalized similarity measurement is more suitable for clustering than other general distance measurements. It can effectively deal with singularities, time shifts and warping and personalized characteristics in stock series, thus obtaining higher prediction accuracy.

4. **Conclusion.** In this paper, we propose a novel time series prediction method, i.e., PSRU, to accurately predict the closing price of stocks. Firstly, VMD is used to decompose and reconstruct the segmented series for data denoising. For better denoising result, we propose a new method to determine the number of decompositions in VMD by comprehensively considering the average sample entropy and cross-correlation of decomposition modes. Then, all the denoised series are clustered by measuring the personalized similarity between any two series, and each class trains an optimal SRU model for the final prediction. Our experiments verify the superiority of PSRU, which is embodied in the following three points. 1) The application of SRU model to stock price trend prediction greatly improves the prediction speed with comparable prediction accuracy with state-of-the-art methods. 2) The adaptive parameter determination method in VMD improves the decomposition accuracy and the denoising result, which is helpful for training the prediction model. 3) The clustering strategy enables the prediction model to better capture the characteristics of different series, thus improving the prediction accuracy. Especially, DMPSM can deal with most of the problems in time series, which greatly improves the clustering performance.

Despite the good performance of the proposed method, there is still some room for improvements in the future. First, the length of the segmented series and the step size during sliding segmentation are two important parameters that may influence the final prediction results. Therefore, a further study on these two factors will be conducted in the future. Second, our method can be used in other related tasks, such as portfolio selection [57]. In addition, the application of our time series prediction method may not be limited to stock data. Therefore, the series data in other fields can be used to further verify the applicability and reliability of the proposed method.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (62176140, 61972235, 61976124, 61873117, 82001775, 61976125), the Central Guidance on Local Science and Technology Development Fund of Shandong Province (YDZX2022093) and Technology Innovation Program of Yantai City (2023JCY J004, 2023XDRH001).

REFERENCES

[1] V. Cho, MISMIS – A comprehensive decision support system for stock market investment, *Knowledge-Based Systems*, vol.23, no.6, pp.626-633, 2010.

[2] Y. Baek and H. Y. Kim, ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module, *Expert Systems with Applications*, vol.113, pp.457-480, 2018.

[3] F. Zhou, H.-M. Zhou, Z. Yang and L. Yang, EMD2FNN: A strategy combining empirical mode decomposition and factorization machine based neural network for stock market trend prediction, *Expert Systems with Applications*, vol.115, pp.136-151, 2019.

[4] J. Ayala, M. García-Torres, J. L. V. Noguera, F. Gómez-Vela and F. Divina, Technical analysis strategy optimization using a machine learning approach in stock market indices, *Knowledge-Based Systems*, vol.225, 107119, 2021.

[5] A. Hogenboom, A. Brojba-Micu and F. Frasincar, The impact of word sense disambiguation on stock price prediction, *Expert Systems with Applications*, vol.184, 115568, 2021.

[6] M. Khashei and M. Bijari, A novel hybridization of artificial neural networks and ARIMA models for time series forecasting, *Applied Soft Computing*, vol.11, no.2, pp.2664-2675, 2011.

[7] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica: Journal of the Econometric Society*, pp.987-1007, 1982.

[8] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, vol.31, no.3, pp.307-327, 1986.

[9] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, Stock price prediction using the ARIMA model, *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp.106-112, 2014.

[10] T. Zheng, J. Farrish and M. Kitterlin, Performance trends of hotels and casino hotels through the recession: An ARIMA with intervention analysis of stock indices, *Journal of Hospitality Marketing & Management*, vol.25, no.1, pp.49-68, 2016.

[11] W. Kristjanpoller and E. Hernández, Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors, *Expert Systems with Applications*, vol.84, pp.290-300, 2017.

[12] K. K. Yun, S. W. Yoon and D. Won, Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process, *Expert Systems with Applications*, vol.186, 115716, 2021.

[13] C.-M. Hsu, A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming, *Expert Systems with Applications*, vol.38, no.11, pp.14026-14036, 2011.

[14] B. M. Henrique, V. A. Sobreiro and H. Kimura, Literature review: Machine learning techniques applied to financial market prediction, *Expert Systems with Applications*, vol.124, pp.226-251, 2019.

[15] O. Bustos and A. Pomares-Quimbaya, Stock market movement forecast: A systematic review, *Expert Systems with Applications*, vol.156, 113464, 2020.

[16] V. N. Vapnik, An overview of statistical learning theory, *IEEE Transactions on Neural Networks*, vol.10, no.5, pp.988-999, 1999.

[17] K.-J. Kim, Financial time series forecasting using support vector machines, *Neurocomputing*, vol.55, nos.1-2, pp.307-319, 2003.

[18] A. Levis and L. Papageorgiou, Customer demand forecasting via support vector regression analysis, *Chemical Engineering Research and Design*, vol.83, no.8, pp.1009-1018, 2005.

[19] C.-F. Huang, A hybrid stock selection model using genetic algorithms and support vector regression, *Applied Soft Computing*, vol.12, no.2, pp.807-818, 2012.

[20] H. Rezaei, H. Faaljou and G. Mansourfar, Stock price prediction using deep learning and frequency decomposition, *Expert Systems with Applications*, vol.169, 114332, 2021.

[21] K. Rathan, S. V. Sai and T. S. Manikanta, Crypto-currency price prediction using decision tree and regression techniques, *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp.190-194, 2019.

[22] W. Jiang and L. Zhang, Edge-SiamNet and Edge-TripleNet: New deep learning models for handwritten numeral recognition, *IEICE Transactions on Information and Systems*, vol.103, no.3, pp.720-723, 2020.

[23] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates et al., Deep Speech: Scaling up end-to-end speech recognition, *arXiv Preprint*, arXiv: 1412.5567, 2014.

[24] Z.-Q. Zhao, P. Zheng, S.-T. Xu and X. Wu, Object detection with deep learning: A review, *IEEE Transactions on Neural Networks and Learning Systems*, vol.30, no.11, pp.3212-3232, 2019.

[25] E. Chong, C. Han and F. C. Park, Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, *Expert Systems with Applications*, vol.83, pp.187-205, 2017.

[26] M. Nikou, G. Mansourfar and J. Bagherzadeh, Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms, *Intelligent Systems in Accounting, Finance and Management*, vol.26, no.4, pp.164-174, 2019.

[27] J. Li, G. Guo and X. Hu, Prediction algorithm of industry rotation based on attention LSTM model, *International Journal of Innovative Computing, Information and Control*, vol.18, no.6, pp.1969-1977, 2022.

[28] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, Learning phrase representations using RNN Encoder-Decoder for statistical machine translation, *arXiv Preprint*, arXiv: 1406.1078, 2014.

[29] L. Li, A.-X. Chen, W.-S. Li, W.-Q. Liang and S.-T. Yang, GRU neural network's prediction of stock closing price, *Computer and Modernization*, no.11, 103, 2018.

[30] F. M. M. Alsheebah and B. Al-Fuhaidi, Emerging stock market prediction using GRU algorithm: Incorporating endogenous and exogenous variables, *IEEE Access*, vol.12, pp.132964-132971, 2024.

[31] T. Lei and Y. Zhang, Training RNNS as fast as CNNS, *arXiv Preprint*, arXiv: 1709.02755, 2017.

[32] T. Kim and H. Y. Kim, Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data, *PLOS ONE*, vol.14, no.2, e0212320, 2019.

[33] J. Cao, Z. Li and J. Li, Financial time series forecasting model based on CEEMDAN and LSTM, *Physica A: Statistical Mechanics and Its Applications*, vol.519, pp.127-139, 2019.

[34] H. Niu, K. Xu and W. Wang, A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network, *Applied Intelligence*, vol.50, pp.4296-4309, 2020.

[35] R. Wang, C. Li, W. Fu and G. Tang, Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction, *IEEE Transactions on Neural Networks and Learning Systems*, vol.31, no.10, pp.3814-3827, 2020.

[36] W. Liu, C. Wang, Y. Li, Y. Liu and K. Huang, Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks, *Chaos, Solitons & Fractals*, vol.146, 110822, 2021.

[37] Z. L. Xiang, R. Wang, X. R. Yu et al., Experimental analysis of similarity measurements for multivariate time series and its application to the stock market, *Applied Intelligence*, vol.53, no.21, pp.25450-25466, 2023.

[38] W. Jiang, Applications of deep learning in stock market prediction: Recent progress, *Expert Systems with Applications*, vol.184, 115537, 2021.

[39] S. Lahmiri, A variational mode decompoisition approach for analysis and forecasting of economic and financial time series, *Expert Systems with Applications*, vol.55, pp.268-273, 2016.

[40] F. Zhao, Y. Gao, X. Li, Z. An, S. Ge and C. Zhang, A similarity measurement for time series and its application to the stock market, *Expert Systems with Applications*, vol.182, 115217, 2021.

[41] K. Dragomiretskiy and D. Zosso, Variational mode decomposition, *IEEE Transactions on Signal Processing*, vol.62, no.3, pp.531-544, 2013.

[42] Z. Zhang and W.-C. Hong, Application of variational mode decomposition and chaotic grey wolf optimizer with support vector regression for forecasting electric loads, *Knowledge-Based Systems*, vol.228, 107297, 2021.

[43] M. Chouksey and R. K. Jha, A multiverse optimization based colour image segmentation using variational mode decomposition, *Expert Systems with Applications*, vol.171, 114587, 2021.

[44] Y. Huang and Y. Deng, A new crude oil price forecasting model based on variational mode decomposition, *Knowledge-Based Systems*, vol.213, 106669, 2021.

[45] S. Lahmiri, A variational mode decomposition approach for analysis and forecasting of economic and financial time series, *Expert Systems with Applications*, vol.55, pp.268-273, 2016.

[46] G. Shi, C. Qin, J. Tao and C. Liu, A VMD-EWT-LSTM-based multi-step prediction approach for shield tunneling machine cutterhead torque, *Knowledge-Based Systems*, vol.228, 107213, 2021.

[47] S. Lahmiri, Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices, *IEEE Systems Journal*, vol.11, no.3, pp.1907-1910, 2015.

[48] R. Bisoi, P. Dash and A. Parida, Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis, *Applied Soft Computing*, vol.74, pp.652-678, 2019.

[49] Y. Liu, C. Yang, K. Huang and W. Gui, Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network, *Knowledge-Based Systems*, vol.188, 105006, 2020.

[50] X. Jiang, C. Shen, J. Shi and Z. Zhu, Initial center frequency-guided VMD for fault diagnosis of rotating machines, *Journal of Sound and Vibration*, vol.435, pp.36-55, 2018.

[51] Y. Huang, Y. Gao, Y. Gan and M. Ye, A new financial data forecasting model using genetic algorithm and long short-term memory network, *Neurocomputing*, vol.425, pp.207-218, 2021.

[52] S. M. Pincus, Assessing serial irregularity and its implications for health, *Annals of the New York Academy of Sciences*, vol.954, no.1, pp.245-267, 2001.

[53] C. Ananth, M. Karthikeyan, N. Mohananthini and G. Yamuna, Adaptive and robust multiple image watermarking using Canberra distance and dual tree complex wavelet transform, *Journal of Computational and Theoretical Nanoscience*, vol.16, no.4, pp.1234-1240, 2019.

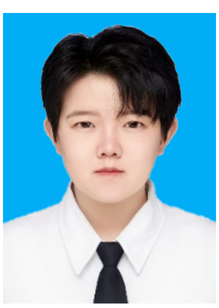
[54] A. Sharabiani, H. Darabi, S. Harford, E. Douzali, F. Karim, H. Johnson and S. Chen, Asymptotic dynamic time warping calculation with utilizing value repetition, *Knowledge and Information Systems*, vol.57, pp.359-388, 2018.
[55] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol.9, no.8, pp.1735-1780, 1997.
[56] Y. Zhang, G. Xu, Y. Wang, X. Liang, L. Wang and T. Huang, Empower event detection with bi-directional neural language model, *Knowledge-Based Systems*, vol.167, pp.87-97, 2019.
[57] H. Guan and Z. An, A local adaptive learning system for online portfolio selection, *Knowledge-Based Systems*, vol.186, 104958, 2019.

Author Biography

Wenjiang Bai received his M.S. degree in Computer Software and Theory from Chongqing University in China in 2010. He is currently a lecturer at Taiyuan University. His research interests lie in the areas of cloud computing and machine learning. He has published multiple papers in well-known journals.

Shuai Yuan received his Master's degree from Yantai University, Yantai, China, in 2008. He is currently an associate professor at the Information Engineering College of Yantai Institute of Technology. His research interests include image segmentation and machine learning. He has published over 10 papers in well-known journals and has also led several scientific research projects.

Yan Lu received a B.S. degree in Information and Computing Science from Nanjing Forestry University in 2020, and an M.S. degree in Applied Statistics from Shandong Technology and Business University in 2023, both in China. She is currently a data analyst at a travel company. Her research interests are in the area of big data analysis. She has published several papers in well-known journals.



Mingyan Xu received a B.S. degree in Applied Statistics from Tiangong University in 2020, and an M.S. degree in Applied Statistics from Shandong Technology and Business University in 2024, both in China. She is currently employed at Harbin Sihe Information Technology Co., Ltd. Her research interests are in the areas of big data analysis and artificial intelligence.

Shuoru Chen received a Bachelor's degree in Engineering from Zaozhuang University in 2022, specializing in Internet of Things Engineering. She is currently a graduate student in Computer Science and Technology at Shandong Technology and Business University, focusing on Financial Big Data. Her research interests include processing complex data and data analysis.

Feng Zhao received the M.S. degree in Computer Science from Xidian University in 2004, and the Ph.D. degree in Computer Science from Xidian University in 2008, both in China. He is currently a professor at Shandong Technology and Business University. His research interests are in the areas of big data analysis and machine learning. He has published over 30 papers in well-known journals and has led multiple scientific research projects, including two funded by the National Natural Science Foundation of China.

Feng Kang received an M.S. degree in Microelectronics and Solid-State Electronics from South China Normal University in 2008. She is currently an associate professor at Yantai Institute of Technology. Her research interests are in the areas of machine learning and micro-nano photonic devices. She has published over 10 papers in well-known journals.