

# A method for representing stock time series features based on trend and inclination angle turning points

Lei Han and Xuedong Gao

School of Economics and Management, University of Science and Technology Beijing  
100083 Beijing, China  
D202210486@xs.ustb.edu.cn  
gaoxuedong@manage.ustb.edu.cn

**Abstract.** Stock time series data mining faces significant challenges in terms of time and space. Time series feature representation is an important means of reducing the dimensionality of time series. This study proposes a stock time series feature representation method based on a combination of slope angle changes and trend turning point screening. This method alleviates the limitations of single methods, such as local feature overfitting, loss of global trends, and sparse long-segment features, through a multi-level collaborative mechanism of screening, filtering, and supplementation. Experimental results show that compared with four other methods, the proposed method effectively preserves the original features of stock time series, achieving good results in both similarity metrics and fitting errors.

**Keywords:** stock time series, feature representation, inclination angle turning point, trend analysis.

## 1. Introduction

With the emergence of new concepts and technologies such as cloud computing, big data, the Internet of Things, and mobile cloud computing, the rate of data growth is rapidly increasing [7]. At the same time, the rapid advancement of computing power and massive storage technology has enabled the systematic acquisition and storage of the complex and heterogeneous data generated by human society in its daily operations and scientific exploration. In this context, time series data, which is organized strictly according to time sequence, has become an important research subject in the field of data analysis because it can describe the dynamic evolution of the state or phenomenon of an object. Efficient processing and in-depth mining of time series data are key means of revealing its internal patterns, correlation laws, and predicting future trends, and have significant practical value. The stock market is a typical application scenario for time series data. Compared with traditional economic data, stock prices exhibit significant volatility and high complexity, which poses unique challenges for the analysis and prediction of stock time series. How to effectively extract useful information from stock time series data and provide scientific guidance to investors has become a hot topic in scientific research.

Due to the inherent high-dimensional characteristics and strong sequence dependency of stock time series, it is very difficult to perform data mining or similarity measurement directly on the original time series. First, at the storage level, uncompressed high-dimensional data occupies a large amount of physical space, resulting in significant storage costs. Second, at the computing level, the computational complexity of traditional

algorithms in high-dimensional spaces often increases exponentially, leading to a sharp decline in processing efficiency. More importantly, raw data usually contains observational noise and measurement errors, and direct processing can easily cause the model to overfit noise patterns, damaging the robustness and generalization ability of the results. The benefits of time series feature representation in improving the efficiency and effectiveness of time series mining have received widespread attention. For example, in areas such as short-term electricity load forecasting [28], carbon price forecasting [39], and electroencephalographic signal detection [2], the high dimensionality of the data and the complexity of the features often result in unsatisfactory prediction results that fail to meet expectations.

By performing feature representation on stock time series, the dimension can be significantly reduced. Using data mining methods, it becomes easier to discover potential patterns in the data, thereby improving data interpretability and prediction accuracy.

The core idea of time series feature representation is to map the original sequence from a high-dimensional space to a low-dimensional feature space or symbolic sequence through dimensionality reduction transformation or symbolic abstraction, thereby achieving the effects of compressing data scale, filtering high-frequency noise, and extracting discriminative features. This preserves the key dynamic patterns and structural information of the time series and enhances the ability to represent real information.

When dealing with stock time series feature representation issues, segmented linear representation, extreme point extraction, and trend analysis have always been the core technical approaches that attract the most attention. However, existing methods often face limitations such as local feature overfitting, loss of global trends, and sparse long segment features. Therefore, this study proposes a new method for feature representation of stock time series data using a multi-level collaborative mechanism. This method initially screens important change points in stock time series using angle changes, then further screens trend turning points using distance metrics, and finally supplements the turning points with error thresholds to ensure trend continuity. Experimental results demonstrate that the method proposed in this study can significantly reduce the volume of raw data, enhance sensitivity to fluctuations in raw stock time series, and better characterize stock price trends, providing important reference value for stock investors.

## 2. Related work

Two key aspects of time series data mining are time series feature representation and time series similarity measurement. With the iterative evolution of information technology and the deep penetration of machine learning methods, related research paradigms have shown significant disciplinary transfer effects, gradually expanding to diverse application scenarios such as financial modelling [23], eco-driving recognition [40], emotion recognition [36], and biomedical diagnosis [1].

Given the high-dimensional, massive, and complex nature of time series data, Faloutsos et al. [9] propose constructing a low-dimensional, robust feature representation space and applying lossy dimensionality reduction compression to the original sequence, which yields more significant effects on data mining. In the process of time series feature representation, the core lies in mapping the multidimensional observations of the original sequence to a low-dimensional space composed of discriminative feature vectors. This

mapping must achieve two objectives: (1) significantly compress the data scale to reduce storage and computational overhead; (2) maximize the preservation of the global evolutionary trends and local structural patterns of the sequence during the compression process.

Therefore, the design of feature representation methods must satisfy three evaluation criteria:

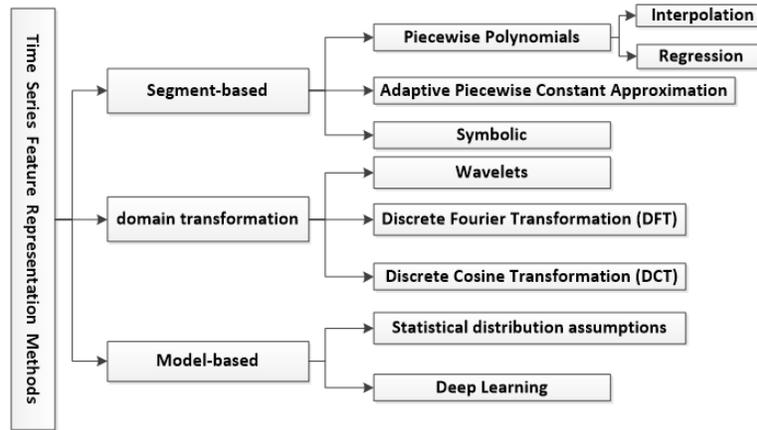
- High dimensionality reduction efficiency: significantly reduce the size of the original sequence data;
- Strong feature fidelity: retain the overall trend pattern and key local features of the sequence;
- Balance between computational efficiency and representation accuracy: achieve high-precision feature encoding with limited computational resources.

In recent years, many excellent methods have been proposed to address the problem of time series feature representation. These methods can be broadly categorized into segment-based feature representation methods, domain transformation-based feature representation methods, and model-based feature representation methods, as shown in Fig1. To assess the similarity of time series or their segmented representations, various metrics have been applied to time series similarity measurement, including Euclidean distance, triangular distance, envelope-based metrics, distance based on dynamic time warping, and distance calculated based on the longest common subsequence of two time series.

## 2.1. Time series feature representation methods

**2.1.1 Segment-Based Feature Representation Methods** The Piecewise Linear Representation (PLR) method was first proposed by Wu et al. in 2004. They used the results of PLR to study data similarity and stock price prediction [38]. In this method, the selection of segment points is particularly critical [22]. The PLR method is simple in form and intuitive in interpretation, but early algorithms had large fitting errors. As the scale of time series data continues to expand, researchers are paying more attention to segmented linear representations based on important point strategies, and trend research has gradually become a hot topic. Wen et al. [32] proposed a bidirectional segmented linear representation method, using a set of linear fitting functions to represent the original time series, which can achieve dimension reduction while maintaining its dynamic characteristics. Hung et al.[13]. used the mean and slope to perform piecewise approximation, but this evenly spaced piecewise result easily distorts the overall trend of the time series. Sun et al.[33] combined the trend distance index with the mean to construct a symbolic description, in order to more comprehensively characterize the overall evolutionary trend of the sequence segment. Zhang et al. [41] designed a symbol representation method based on trend feature enhancement, which retains the mean while integrating trend distance factors and trend pattern factors to jointly encode the dynamic characteristics of the sequence. Zhou et al. [45] proposed a segmented linear representation combining turning points and maximum absolute deviation in stock time series to extract sequence features. Liu et al.[43] proposed a turning point extraction strategy based on angle change thresholds, which calculates the directional angle change between adjacent segments and combines trend segment features to evaluate and screen for significant turning points. Liao et al.[20]

designed a distance-driven threshold discrimination method, which determines the importance of candidate points by comparing them with a preset distance threshold. However, its underlying basic inflection point screening relies solely on the numerical relationship between three adjacent points, resulting in limited robustness. Chen et al. [4] proposed a linear representation method for time series key points, which comprehensively considers the time span and amplitude changes and can efficiently extract key points from time series.



**Fig. 1.** Time Series Feature Representation Methods

Based on PLR, several classic algorithms such as piecewise aggregate approximation (PAA) [16], top-down (TD) [19], and bottom-up (BU) [30] piecewise linear representation methods have been proposed. The PAA algorithm achieves dimension reduction by dividing time series into equal intervals and using the average value of each segment to represent the original sequence. This method does not take into account the differences between adjacent samples, and the results are often unsatisfactory. Without addressing this shortcoming, optimization research on PAA has primarily focused on two directions: feature representation and segmentation strategies. On one hand, Huang et al. [14] enhanced the mean representation by integrating auxiliary features such as slope and variance. On the other hand, Keogh et al. [17] proposed the Adaptive Partitioning Constant Approximation (APCA), which utilizes dynamic programming to achieve sequence-adaptive optimal segmentation, resulting in non-uniformly partitioned segments. Additionally, to simultaneously improve dimension reduction efficiency and enhance the representation of sequence macro trends, researchers proposed the Symbolic Aggregate Approximation (SAX) method [21] and its improved frameworks [15, 12, 44].

Information granularity, as a key extension of piecewise linear representation, focuses on abstracting sequences into granular structures with semantic meaning. Based on information granularity theory [31], Duan et al. [8] developed a time series granularity model and constructed a granularity-driven clustering framework based on this model. Guo et al. [11] designed a trend feature-guided granularity method and verified its representational

validity in clustering tasks, confirming its advantages in representing time series structural features.

**2.1.2 Feature representation methods based on domain transformation** The core of feature representation based on domain transformation lies in mapping time series to a transformation domain independent of the original data, where the transformation operator and feature selection do not depend on the inherent characteristics of the data. Typical transformation spaces include the frequency domain, spectral space, etc. Commonly used methods include the Discrete Fourier Transform (DFT)[18] and the Discrete Wavelet Transform (DWT) [3].

DFT converts time-domain sequences into complex frequency-domain representations that characterize periodicity and frequency structure using trigonometric basis functions. DFT preserves data length and is suitable for smooth sequences. However, it has weak ability to capture local sudden changes. DWT retains local characteristics in both time and frequency through multi-scale decomposition, and the wavelet coefficients it generates can accurately describe the transient modes of non-stationary signals. Compared with DFT, DWT demonstrates superior feature representation robustness in non-stationary time series such as highly volatile financial data. Panda et al[29]. achieved good results in tumor classification by constructing models using wavelet-based feature learning and deep learning-based feature learning. De et al[37]. proposed an end-to-end framework for stock price prediction called the Multi-level Wavelet Decomposition Interaction Network (MWDINet). MWDINet employs multi-scale wavelet decomposition interaction modules (MWDI-Block) and Heron moving average modules (HMA-Block) to extract frequency-domain and time-domain information from data, respectively, demonstrating significant potential in stock price prediction.

**2.1.3 Model-based feature representation methods** Model-based feature representation methods assume that time series are generated by specific mathematical models, and extract features by constructing corresponding models. Their advantages lie in the flexibility of model selection and their derivative capabilities, but they have limitations such as strong assumption dependence, high domain knowledge thresholds, and high computational complexity.

Based on statistical distribution assumptions, the current mainstream methods include autoregressive models (ARM), orthogonal polynomial regression models [10], hidden Markov models (HMM)[24], and kernel models. Such methods require domain-specific prior knowledge to support model selection and are difficult to uniformly capture the coupled effects of long-term dependencies, cyclical patterns, and random fluctuations. Methods based on neural network models utilize deep networks to automatically learn feature representations, giving rise to methods such as CNN-RNN hybrid architectures [26], WDCNN-LSTM[5], multi-channel deep convolutions, and hierarchical bidirectional learning [34]. Lu et al[25]. developed a time series-based recurrent neural network (TRNN) model specifically optimized for the characteristics of financial time series data. The model achieves efficient data dimensionality reduction by intelligently extracting key points of market trends and turning signals, and reinforces the impact of recent trading volume on prices through a dynamic weighting mechanism, significantly improving prediction accuracy while ensuring computational efficiency. Wang et al[35].

proposed an enhanced interval value decomposition integral model based on comprehensive feature extraction and optimized deep learning, utilizing interval variational modal decomposition (FIVMD) with a feedback mechanism to extract internal features, achieving satisfactory prediction results. Although these methods have overcome the assumption constraints of statistical models and enhanced the representational capabilities of big data, the low interpretability, high computational overhead, and weak generalization performance of neural networks in financial time series data remain critical issues that require optimization.

## 2.2. Time series similarity measurement methods

After time series feature representation, time series similarity measures can be used to obtain the degree of similarity between sub-sequences and between sub-sequences and the original sequence. Such measures can usually be defined by a difference function, also known as a distance function[27]. Euclidean distance has very efficient computational performance, making it a commonly used method for measuring the distance between time series. Euclidean distance metrics require sequences to be of equal length to achieve “one-to-one” matching calculations between data, which also limits its scope of application. Dynamic Time Warping (DTW) is a highly accurate and robust time series distance measurement method originally applied in the field of speech recognition. Compared to traditional Euclidean distance, DTW can match data points in time series by warping the time domain of the time series, not only achieving better shape measurement effects but also enabling the measurement of two time series of unequal length.

In recent years, to alleviate the computational bottleneck of early time series similarity measurement methods, Choi et al. [6] proposed the fast Sakoe-Chiba DTW (SC-DTW) and the fast incremental DTW (I-DTW), which achieved a significant improvement in computational speed compared to the traditional DTW. Zhang et al. [42] proposed a time series similarity measurement method that combines sequence decomposition and DTW. This method has lower computational complexity than DTW and demonstrates significant efficiency in classification tasks. Additionally, for the granular representation of time series information, some scholars have adopted the integral area method to measure the distance between the same linear segments.

In the highly complex stock market, stock price sequences are influenced by the coupled effects of multiple heterogeneous factors, coupled with significant interference from high-frequency noise, resulting in their pronounced volatility and inherent instability. To more effectively identify and analyze the underlying trend patterns within such sequences, existing research primarily relies on linear transformation methods and feature representation techniques such as piecewise aggregation approximation. However, the effectiveness of these methods is highly dependent on pre-set fixed or adaptive parameter thresholds. This parameter dependency not only introduces significant subjective risks but may also lead to information loss or false pattern recognition due to improper threshold selection, thereby facing adaptive limitations when handling highly complex, non-stationary financial time series. Therefore, this study focuses on how to retain the main trend features of stock time series while mitigating the limitations caused by segmented linear fitting, such as local feature overfitting, global trend loss, and long segment feature sparsity. It also employs appropriate similarity measurement methods to compare the advantages and disadvantages of different feature representation methods.

### 3. Preliminary

This section provides definitions related to the methods proposed in this study.

**Definition 1 (time series):** A time series is an ordered set of observations collected in chronological order. Recorded as  $X = \{(x_1, t_1), (x_2, t_2), \dots, (x_n, t_n)\}$ . Where  $(x_i, t_i)$  represents the observed value at time  $t_i$  and satisfies  $t_1 < t_2 < \dots < t_n$ . If the sampling interval is constant, i.e.,  $\Delta t = t_{i+1} - t_i$  is a constant, then the time series can be simplified to a numerical sequence form:  $X = \{x_1, x_2, \dots, x_n\}$ . For stock time series, stock price data comes from every transaction in the stock market, and there is a fixed time interval between each transaction within the specified trading time. In the stock market, daily transaction data is not continuous in time, but due to the fixed trading hours, stock price data at different time granularities can be represented as:  $P = \{p_1, p_2, \dots, p_n\}$ .

**Definition 2 (Segmented Linear Representation of Stock Time Series):** Divide the stock time series  $P$  into  $m$  segments, each of which is approximated by a straight line between the endpoints. The formal definition is:

$$P_{PLR} = \{(p_{s1}, p_{e1}), (p_{s2}, p_{e2}), \dots, (p_{sm}, p_{em})\} \quad (1)$$

Where,  $p_{sk}, p_{ek} \in P, ek = sk + 1$ . Since the segment points are interconnected, the sequence represented by the segments can be expressed as:  $P_{PLR} = \{p_1^*, p_2^*, \dots, p_{m+1}^*\}$ .

**Definition 3 (fitting error):** The fitting error is obtained by linearly interpolating line  $P_{PLR}$  to obtain the approximate value  $P' = \{p_1', p_1', \dots, p_n'\}$  of the entire sequence.

$$Error = \sqrt{\sum_{i=1}^n (p_i - p_i')^2} \quad (2)$$

**Definition 4 (inclination angle turning points):** Given a stock price time series  $P = \{p_1, p_2, \dots, p_n\}$ , its slope turning point series  $P^s = \{p_1^s, p_2^s, \dots, p_m^s\}$ . The absolute angle change value  $\alpha_i$  at point  $p_i$  must satisfy:

$$\alpha_i = \begin{cases} \pi - (|\theta^-| - |\theta^+|), & \text{if } \theta^- \times \theta^+ < 0 \\ |\theta^- - \theta^+|, & \text{otherwise} \end{cases} \geq \delta \quad (3)$$

Where,  $\theta^- = \arctan\left(\frac{p_i - p_{i-1}}{\Delta t}\right)$  is the direction angle of the front line segment,  $\theta^+ = \arctan\left(\frac{p_{i+1} - p_i}{\Delta t}\right)$  is the direction angle of the rear line segment, and  $\delta$  is the angle change threshold.

As defined in Definition 4, the smaller the angular change value of a point, the closer the lines connecting the three points before and after that point as the midpoint approach a straight line, indicating a smaller trend change. Consequently, the probability of that point being a trend inflection point is smaller, while the opposite is true for a larger angular change value. As shown in Fig2, Fig2(a) displays the K-line chart and closing prices of the stock (SZ300059) from February 6, 2025, to March 5, 2025. Fig2(b) shows the sequence of angle change values for the stock's closing prices. As can be seen, the stock price time series exhibits significant volatility, and angle change values can be used to efficiently select segmentation points and achieve the preliminary goal of reducing the number of time series.

Definition 5 (Stock Time Series Trend): The stock time series trend is determined by connecting adjacent inclination angle turning points. Given a stock time series inclination angle turning point sequence  $P^s = \{p_1^s, p_2^s, \dots, p_m^s\}$ , the trend is defined as  $T = \{t_1, t_2, \dots, t_j, \dots, t_{m-1}\}$ . Here,  $t_j$  represents the connecting line segment between turning points  $p_j$  and  $p_{j+1}$ .

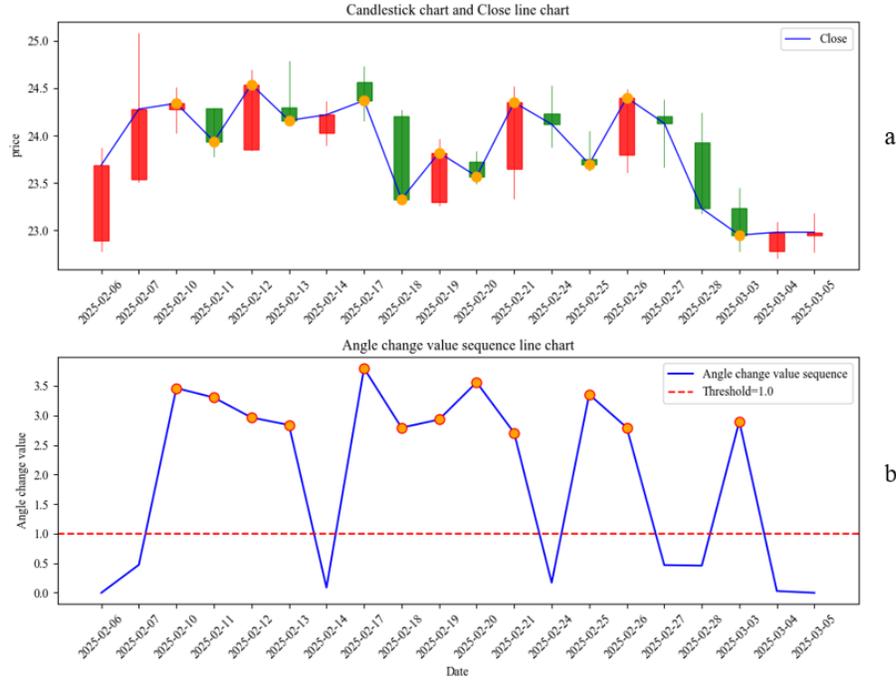


Fig. 2. Example diagram of angle change values

Definition 6 (Stock Time Series Trend Turning Point): For the inclination angle turning point sequence  $P^s = \{p_1^s, p_2^s, \dots, p_m^s\}$ , take the three consecutive points  $p_{j-1}^s, p_j^s$  and  $p_{j+1}^s$  as the minimum pattern for trend formation. Use the vertical distance  $D_v$  between the line segments connecting  $p_j^s$  to  $p_{j-1}^s$  and  $p_{j+1}^s$  as the standard for measuring its importance. The greater the distance, the greater the importance of  $p_j^s$  in the trend turning point, and the more likely it is to become an important node in the trend turning point. Form the trend turning point sequence:  $P^t = (p_1^t, p_2^t, \dots, p_x^t)$ .

$$D_v = \left| p_{j-1}^s + \frac{(p_{j+1}^s - p_{j-1}^s)(t_2 - t_1)}{t_3 - t_1} - p_j^s \right| \quad (4)$$

Definition 7 (maximum error point): For sub-sequences formed by trend turning points in stock time series, if the length of the sub-sequence segment exceeds the threshold or the fitting error exceeds the threshold, calculate the absolute error between the fitted line

and the original data points:  $E = \max(|p_i - f(x_i)|)$ . Select the maximum error point as the supplementary trend turning point.

## 4. A method for representing stock time series characteristics based on trends and inclination angle turning points

### 4.1. Method description

The stock time series feature representation method based on trends and inclination angle turning points proposed by this research institute uses a multi-level collaborative mechanism supplemented by inclination angle turning points, trend turning points, and maximum error points to greatly reduce the amount of data while retaining the trend characteristics of the original time series data to the greatest extent possible, laying the foundation for time series analysis. The specific algorithm steps are as follows: Given a stock time series  $P = \{p_1, p_2, \dots, p_n\}$ , angle change threshold  $\delta$ , distance threshold  $D$ , maximum error threshold  $\eta$ , and maximum segment length  $L$ .

- I. Calculate the angle turning points of the stock time series. According to Definition 4, calculate the absolute angle change value of each data point in order, filter out the data points greater than the angle change threshold, add the point to the inclination angle turning point sequence  $P^s = \{p_1^s, p_2^s, \dots, p_m^s\}$ , and form the trend segment  $T = \{t_1, t_2, \dots, t_j, \dots, t_{m-1}\}$  according to Definition 5.
- II. Based on Definition 6, calculate the perpendicular distance between A and B and the line connecting C, and select points greater than the distance threshold  $D$  as trend turning points.
- III. Based on the sequence length threshold  $L$  select the sequence segment greater than  $L$  in II or the sequence segment with a fitting error greater than the maximum error threshold, calculate the absolute error  $E$  between the fitting line and the original data points in the sequence segment, and select the largest error point as the supplementary trend turning point.
- IV. Trend continuity test. For the trend turning point sequence formed in II and III, calculate the slope of adjacent trend segments, merge trend segments with the same change trend, and ensure that the trend segments are always in a state of alternating rises and falls.

The algorithm flowchart is shown in Fig3. The algorithm pseudocode is shown in Fig4.

**Parameter Description:** The algorithm in this paper involves four key parameters: angle change threshold  $\delta$ , distance threshold  $D$ , maximum error threshold  $\eta$ , and maximum segment length  $L$ .  $\delta$  and  $D$  jointly control the sensitivity of trend inflection point detection. Setting the threshold too low may cause redundant local fluctuations to be misjudged as trend changes, while setting it too high may result in important trend inflection points being missed.  $\eta$  and  $L$  jointly regulate the granularity distribution of trend segments. By constraining the upper limit of segmentation error and the maximum interval length, they ensure feature density while avoiding over-segmentation issues. When optimizing parameters, it is recommended to prioritize balancing the values of  $\delta$  and  $D$ , while  $\eta$  and  $L$  can be appropriately relaxed to improve the continuity of long trends.

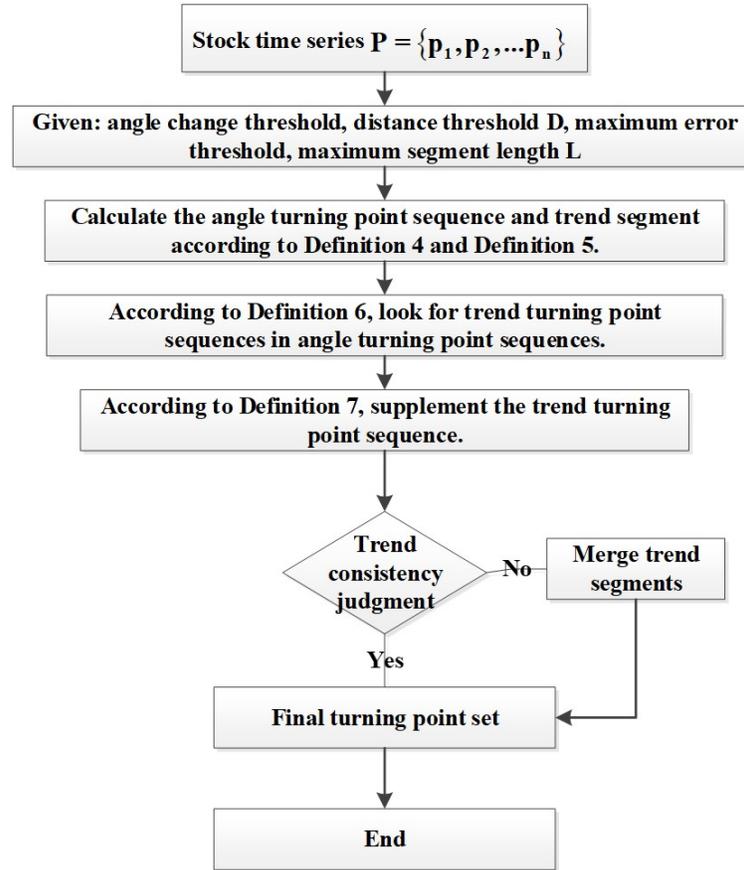


Fig. 3. Algorithm flowchart

## 4.2. Evaluation indicators

Similarity measurement is an important part of evaluating the quality of time series feature representation methods. This study uses Euclidean distance and integral distance to evaluate the fitting error and morphological similarity between the feature-represented sequence and the original sequence data.

**4.2.1 European distance** European distance refers to the distance between two objects in  $m$ -dimensional space. Due to its highly efficient computational performance, it is commonly used to measure the distance between time series. The calculation formula is shown in Eq5:

European distance in two-dimensional space:

$$Dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (5)$$

---

**Input:** A set of stock time series data  $P$ ; angle change threshold; distance threshold  $D$ ; maximum error threshold; maximum segment length  $L$ .

**Output:** Stock Time Series Trend Turning Point Sequence  $X$

---

$P = p_1, p_2, \dots, p_i$   
 Define set  $X$  to store trend turning points, initialized as empty.  
 Define error set  $E$ , used to store the fitting error of the trend turning point, initialized as empty.  
 Definition function find\_Angular inflection point ( $P, \delta$ )  $P^s$  // Searching for angle turning point sequences  
 Definition function find\_Trend turning points ( $P^s$ )  $P^t$  // Finding trend turning point sequences  
 Definition function Additional turning point ( $P^t$ )  $E$  // Supplemental turning point sequence.  
 while True  
   if  $E = \max(|p_i - f(x_i)|) > \eta$   
      $P^t$  append  $p_i$   
     if  $K(p_{i-1}^t, p_i^t) * K(p_i^t, p_{i+1}^t) > 0$   
        $P^t = P^t \cup (p_{i-1}^t, p_{i+1}^t)$   
     Else  
        $P^t = P^t$   
   Else  
      $P^t = P^t$   
      $X = P^t$   
   Break  
 Return  $X$

---

**Fig. 4.** Algorithm pseudocode

**4.2.2 Area distance** Area-based similarity measurement methods can measure the morphological similarity between feature representations and original sequences after linear interpolation. For stock time series, the more similar the morphology, the more important it is for guiding future investment decisions. Therefore, this study uses the area distance method to measure the degree of morphological similarity between feature representations and original sequences.

Given the original stock time series  $A$ , the sequence  $B$  obtained by linear interpolation after feature representation is used, where  $n = m$ . There are two types of relationships between the two sequences: intersecting and non-intersecting, as shown in Fig5. As shown in Fig5(a), the integral distance when two sequences do not intersect is:

$$Dist = \int_{t_1}^{t_2} [(k_1t + b_1) - (k_2t + b_2)] dt \quad (6)$$

As shown in Fig5(b), the integral distance between the two sequences relative to  $t^*$  is:

$$Dist = \int_{t_1}^{t^*} [(k_1t + b_1) - (k_2t + b_2)] dt + \int_{t^*}^{t_2} [(k_2t + b_2) - (k_1t + b_1)] dt \quad (7)$$

## 5. Analysis and evaluation of experimental results

### 5.1. Experimental Process

**5.1.1 Experimental Environment** When implementing the algorithm for identifying extreme points in stock time series trends, we selected a series of development components. The specific selection of these components is shown in Table1.

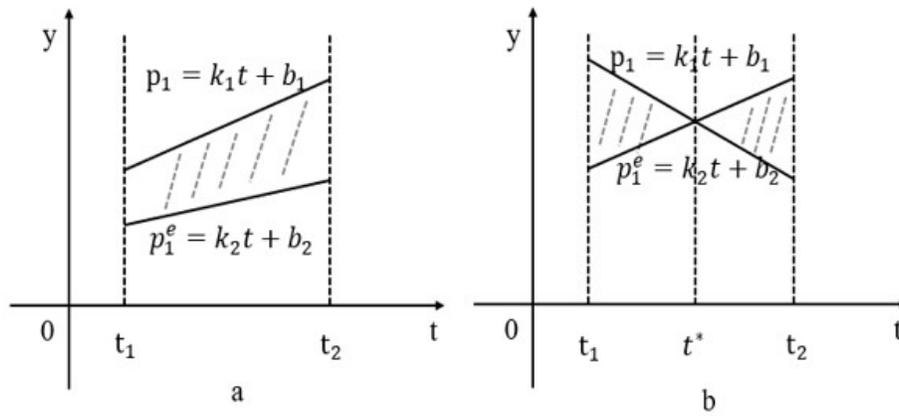


Fig. 5. Relationships between sequences

Table 1. Experimental environment setting

Name	Configurations
Processor CPU	Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 2.00 GHz
Memory RAM	DDR4 8GB 2400MHz
Hard disk	128GB Solid State Drive (Toshiba)
Operating System OS	Windows 10 Home Edition 19043.1889
Programming Language	Python3.5.0
IDE	JetBrains PyCharm 2018.3.2
Toolkits	Pandas NumPy et.al

**5.1.2 Dataset** To ensure that the research findings possess strong industry representativeness and external validity, this study adhered to strict industry coverage and representativeness principles when designing the experimental dataset. Specifically, the primary dataset selected leading listed companies from different industry sectors traded on China's two major securities exchanges—the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE)—as the research subjects. The time span is from March 1, 2024, to February 28, 2025, which is a relatively complete annualized operational cycle. This avoids the influence of extreme short-term noise while reflecting market dynamics and structural characteristics with a certain degree of continuity. The stock data is at a 30-minute interval, which, compared to traditional daily data, can more precisely reveal intraday price behavior patterns, short-term volatility characteristics, and potential microstructural effects. The data source is Wind Financial Terminal (Wind), which is a widely recognized and authoritative professional financial data and analysis service provider in China's financial markets. Its database boasts high completeness, timeliness, and accuracy, making it a commonly used and reliable data source for empirical financial research in both academic and professional circles domestically and internationally. Specific stock information is shown in Table 2.

**Table 2.** Data set stock information

Num	Code	Name	Data volume	Industry
1	000063SZ	ZTE	1928	Industrial Communication Device, 5G
2	000333SZ	Midea Group	1928	Electrical and Mechanical Industries
3	001979SZ	China Merchants Shekou Industrial Zone Holdings	1928	Real Estate Development Sector
4	002371SZ	NAURA Technology Group	1928	Semiconductor and electronic components sector
5	002594SZ	BYD	1928	Automotive Manufacturing Sector
6	600030SH	Citic Securities	1928	Big finance, securities
7	600519SH	KweichowMoutai	1928	Distillery Industry
8	601138SH	Foxconn Industrial Internet	1928	Industrial Internet
9	601985SH	China National Nuclear Power	1928	Power industry
10	688256SH	Cambricon Technologies Corporation Limited	1928	Artificial Intelligence Chip

Data preprocessing is an important part of data mining. Through analysis of the original dataset, the stock data selected for this study shows no missing values between March 1, 2024, and February 28, 2025. The 30-minute line data consists of eight data points for each trading day, totaling 241 trading days. To fully reflect the original state of the stock time series, this study did not standardize the original series.

## 5.2. Analysis of experimental results

In order to verify the validity of the stock time series feature representation method based on inclination angle and trend turning points proposed in this study, we selected the stock time series feature representation method based on trend feature points (TFP) [17], the time series trend feature extraction method based on turning points and trend segments (TPTS)[21], PLR-ITTP [15], and PLR-KP [12]for comparison.

**5.2.1 European distance** After feature representation, the method proposed in this study uses linear interpolation to fill in the time series. The other four methods also use linear interpolation to fill in the gaps, and then calculate the European distance between the filled-in sequence and the original sequence. Table3 shows the European distance of different methods at a compression rate of 70%.

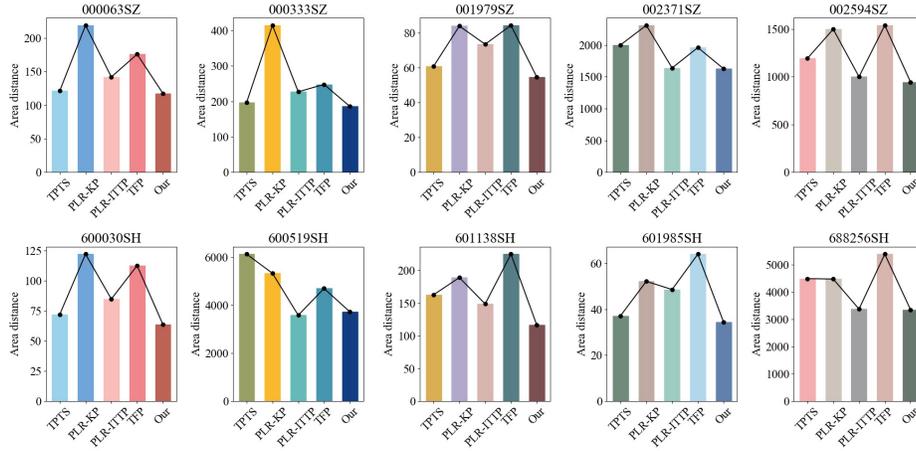
**Table 3.** European distance for different methods at a compression rate of 70%

	TPTS	PLR-KP	PLR-ITTP	TFP	Our
000063SZ	4.2697	9.7442	5.1305	6.8978	<b>4.1358</b>
000333SZ	7.0874	17.013	7.9028	9.5392	<b>6.6985</b>
001979SZ	2.2585	3.2936	2.4531	3.1167	<b>2.0417</b>
002371SZ	78.7373	99.5045	57.5954	77.276	<b>57.3243</b>
002594SZ	40.202	66.0142	35.5037	44.3144	<b>32.8873</b>
600030SH	2.6216	6.0502	2.9709	5.8952	<b>2.3209</b>
600519SH	263.6387	237.6315	<b>130.4544</b>	184.0867	130.567
601138SH	4.265	7.8261	5.1287	7.3047	<b>4.1947</b>
601985SH	1.3405	2.0641	1.6546	2.077	<b>1.2407</b>
688256SH	176.7896	203.3061	116.9405	155.6698	<b>114.68</b>

As can be seen from Table3, except for the 600519SH, where the method proposed in this study is slightly greater than PLR-ITTP in terms of European distance, the method proposed in this study has the smallest value in terms of European distance measurement for other stocks, while the PLR-KP method has the largest value for most stocks. This indicates that considering only the key points and time spans of the time series is insufficient for accurately representing stock time series in an effective linear manner. In contrast, the three-level collaborative method proposed in this study achieves excellent results in the segmented linear representation of stock time series. Because the method proposed in this study uses two parameters, the maximum error threshold  $\eta$  and the maximum segment length  $L$ , to split long time series that other methods fail to focus on, the length of the sub-segments after feature representation is relatively uniform. In this case, the measured European distance will be smaller than that of other methods, which also shows that the method proposed in this paper can more accurately characterize the original stock time series.

**5.2.2 Area distance** Based on European distance measurement, the formula is used to calculate the area distance in order to measure the morphological similarity between the

feature representation sequence and the original sequence. Fig6 shows the area distance of different methods at a compression rate of 70%.



**Fig. 6.** Area distance of different methods at a compression ratio of 70%

As can be seen from the experimental results in Fig6, there are significant differences among different methods in terms of area distance metrics. The method proposed in this study achieved the minimum area distance for all stocks except 600519SH, indicating that after using the method proposed in this paper for stock time series feature representation, the resulting trend lines of turning points maintain the highest degree of similarity with the original series, better aligning with the original stock time series pattern. While the PLR-ITTP method achieved the minimum area distance for the 600519SH stock, its performance on other stock sequences was unstable. The experimental results indicate that the method proposed in this study has good adaptability and demonstrates robust performance in stock time series feature representation.

**5.2.3 Fitting error** Based on Eq2, the fitting error between the linear interpolation sequence after feature representation and the original sequence was calculated. The calculation indicators included root mean square error (RMSE) and mean absolute error (MAE). The experimental results are shown in Table 4.

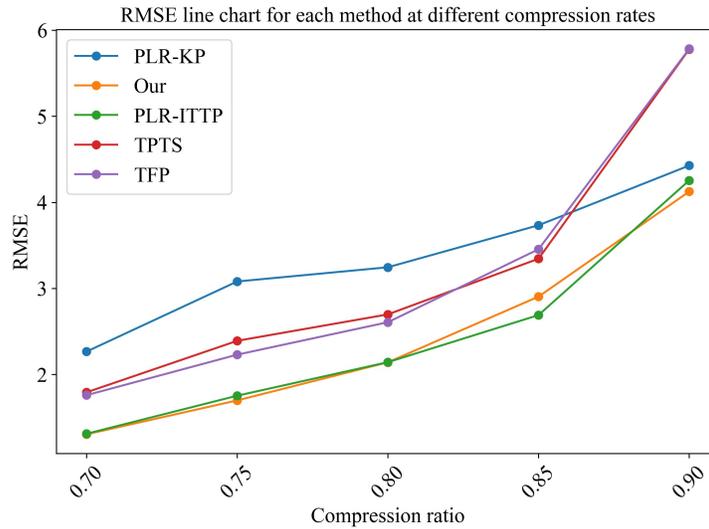
In Table4, bolded data indicate the smallest error. As shown in Table4, among the 10 stock time series, the method proposed in this study achieved the smallest fitting error across all time series, while the PLR-KP method performed poorly in most time series. This indicates that by employing multi-level collaborative screening and supplementation of turning points, it is possible to avoid the premature rejection of important turning points to the greatest extent. By preserving the overall trend of the time series and adequately reducing dimensionality, the method mitigates the limitations of local feature overfitting, loss of global trends, and sparsity of long-term features, thereby reducing fitting errors.

**Table 4.** Fitting error of different methods at a compression ratio of 70%

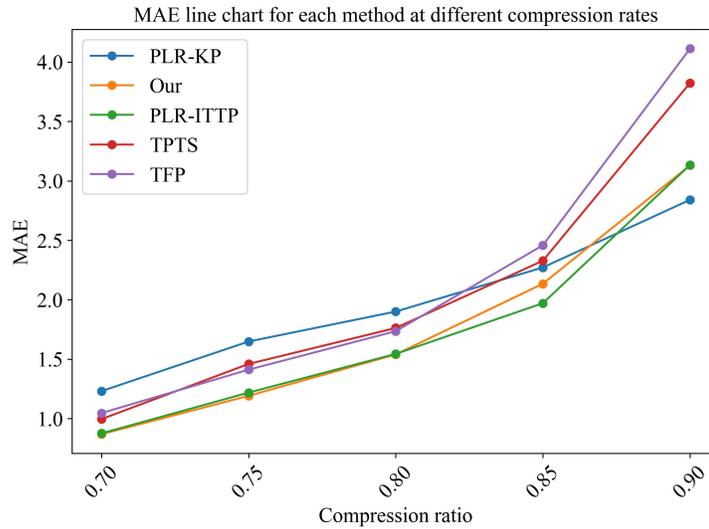
	TPTS		PLR-KP		PLR-ITTP		TFP		Our	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
000063SZ	0.0972	0.066	0.2219	0.1174	0.1168	0.0757	0.1571	0.0931	<b>0.0942</b>	<b>0.0633</b>
000333SZ	0.1614	0.1082	0.3875	0.2221	0.18	0.1212	0.2172	0.1328	<b>0.1526</b>	<b>0.1008</b>
001979SZ	0.0514	0.0331	0.075	0.0449	0.0559	0.0389	0.071	0.0448	<b>0.0465</b>	<b>0.0291</b>
002371SZ	1.7932	0.9954	2.2662	1.2308	1.3117	0.8753	1.7599	1.0462	<b>1.3055</b>	<b>0.8684</b>
002594SZ	0.9156	0.5388	1.5034	0.8026	0.8086	0.5343	1.0092	0.6022	<b>0.749</b>	<b>0.5053</b>
600030SH	0.0597	0.039	0.1378	0.065	0.0677	0.0449	0.1343	0.0599	<b>0.0529</b>	<b>0.0342</b>
600519SH	6.0042	3.0027	5.4119	2.8604	2.971	1.9852	4.1925	2.5142	<b>2.9736</b>	<b>1.9845</b>
601138SH	0.0971	0.065	0.1782	0.1013	0.1168	0.0793	0.1664	0.0939	<b>0.0955</b>	<b>0.0632</b>
601985SH	0.0305	0.0203	0.047	0.0281	0.0377	0.0258	0.0473	0.0284	<b>0.0283</b>	<b>0.0184</b>
688256SH	4.0263	2.0858	4.6302	2.4118	2.6632	1.7871	3.5453	1.9661	<b>2.6118</b>	<b>1.7689</b>

This further demonstrates the effectiveness and superiority of the method proposed in this study.

**5.2.4 Fitting error of each method at different compression rates** After calculating the European distance, area distance, and fitting error of various methods under the same compression ratio, the effectiveness of the method proposed in this study was verified. In this section, we examined the fitting error of the five methods for stock 002371SZ at compression ratios of 70%, 75%, 80%, 85%, and 90%. The experimental results are shown in Fig7 and Fig8.



**Fig. 7.** RMSE line charts for each method at different compression rates



**Fig. 8.** MAE line charts for each method at different compression rates

As shown in Fig7 and Fig8, the proposed method achieves the minimum RMSE and MAE values before a compression rate of 80%. When the compression rate exceeds 80%, the proposed method still achieves relatively low RMSE and MAE values. PLR-ITTP and PLR-KP perform well at higher compression rates. Meanwhile, as the compression rate increases, the growth of RMSE and MAE is relatively stable, indicating the stability of the proposed method in terms of fitting performance.

## 6. Conclusions

Feature representation of stock time series data is a process aimed at transforming raw high-dimensional data into a low-dimensional form. Its core objective is to significantly improve the efficiency of subsequent data mining, similarity measurement, and other related research tasks while maximizing the retention of key information from the original sequence. Based on the segmented linear representation of time series, this study proposes a stock time series feature representation method based on a combination of inclination angle changes and trend turning point screening. Through a multi-level collaborative mechanism, this method alleviates the limitations of local feature overfitting, global trend loss, and long segment feature sparsity in a single method. The study reached the following conclusions:

- The stock time series feature representation method proposed in this study, which combines inclination angle changes with trend turning point screening, has achieved good results in terms of European distance, area distance, and fitting error evaluation indicators. It can effectively retain the original features of stock time series and maintain good stability.

- The method proposed in this study can more accurately capture trend features in stock time series analysis, providing investors with more reliable and precise decision-making basis. This research finding has important implications for understanding the stock market and making investment decisions.

However, this study still has some limitations: first, although the proposed multi-level collaborative processing mechanism improves the depth of analysis of the original sequence, the increased algorithm complexity may affect its computational efficiency on large-scale datasets. Second, the current validation is only applicable to specific types of stock time-series data, and the generalization ability of the method needs further investigation.

Future work will focus on the following directions: (1) reducing computational complexity through algorithm optimization to improve processing efficiency; (2) validating the method's adaptability in broader data scenarios (such as high-frequency trading data and cross-market indices); and (3) exploring its integration with deep learning models to enhance its stability and predictive capabilities in noisy market environments. These improvements will help advance the practical application of this method in complex financial scenarios.

## References

1. Abineza, C. and Balas, V., Nsengiyumva, P.: Deep learning model for early subsequent copd exacerbation prediction. *Studies in Informatics and Control* 32(3), 99–107 (2023)
2. Bae, H.J., Park, J.S., Choi, J.H., Kwon, H.Y.: Learning model combined with data clustering and dimensionality reduction for short-term electricity load forecasting. *Scientific Reports* 15(1), 3575 (2025)
3. Buhler, J., Tompa, M.: Finding motifs using random projections. In: *Proceedings of the fifth annual international conference on Computational biology*. pp. 69–76 (2001)
4. Chen, S., Lv, X., Qi, R., Wang, L., Yu, L.: Linear representation method based on key points for time series. *Computer Science* 43(05), 234–237 (2016)
5. Chen, Z., Zhao, J.: Bearing fault diagnosis based on wide deep convolutional neural network and long short term memory. *Tehnički vjesnik* 30(1), 265–273 (2023)
6. Choi, W., Cho, J., Lee, S., Jung, Y.: Fast constrained dynamic time warping for similarity measure of time series data. *IEEE Access* 8, 222841–222858 (2020)
7. Dars, M.A., Liu, Q., Chaozhu, Z., Shaikh, S., Farid, G.: A context-aware recommendation system based on hadoop big data. *STUDIES IN INFORMATICS AND CONTROL* 34(1) (2025)
8. Duan, L., Yu, F., Pedrycz, W., Wang, X., Yang, X.: Time-series clustering based on linear fuzzy information granules. *Applied Soft Computing* 73, 1053–1067 (2018)
9. Faloutsos, C. and Ranganathan, M., Manolopoulos, Y.: Fast subsequence matching in time-series databases. *ACM Sigmod Record* 23(2), 419–429 (1994)
10. Fuchs, E., Gruber, T., Pree, H., Sick, B.: Temporal data mining using shape space representations of time series. *Neurocomputing* 74(1-3), 379–393 (2010)
11. Guo, H., Wang, L., Liu, X., Pedrycz, W.: Trend-based granular representation of time series and its application in clustering. *IEEE Transactions on Cybernetics* 52(9), 9101–9110 (2021)
12. Hu, Y., Guan, P., Zhan, P., Ding, Y., Li, X.: A novel segmentation and representation approach for streaming time series. *IEEE Access* 7, 184423–184437 (2018)
13. Hung, N., Anh, D.: An improvement of paa for dimensionality reduction in large time series databases. In: *Pacific rim international conference on artificial intelligence*. pp. 698–707 (2008)

14. Hung, N., Anh, D.: An improvement of paa for dimensionality reduction in large time series databases. In: Pacific rim international conference on artificial intelligence. pp. 698–707. Springer (2008)
15. Ji, C., Liu, S., Yang, C., Wu, L., Pan, L., Meng, X.: A piecewise linear representation method based on importance data points for time series data. In: 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD). pp. 111–116. IEEE (2016)
16. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and Information Systems* 3(3), 263–286 (2001)
17. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Locally adaptive dimensionality reduction for indexing large time series databases. In: Proceedings of the 2001 ACM SIGMOD international conference on Management of data. pp. 151–162 (2001)
18. Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Locally adaptive dimensionality reduction for indexing large time series databases. In: Proceedings of the 2001 ACM SIGMOD international conference on Management of data. pp. 151–162 (2001)
19. Keogh, E., Pazzani, M.: An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In: *Kdd*. vol. 98, pp. 239–243 (1998)
20. Liao, J., Zhou, Z., Kou, Y., Luo, H.: Time series piecewise linear representation based on trend transition point. *Computer Engineering and Applications* 46(30), 50–53+81 (2010)
21. Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery* 15(2), 107–144 (2007)
22. Liu, S., Jiang, H.: A review on time series representation for similarity-based pattern search. *Computer Engineering Apply* 27, 53–57 (2004)
23. Liu, Z. and Liu, X., Zhou, L.: A hybrid cnn and lstm based model for financial crisis prediction. *Tehnički vjesnik* 31(1), 185–192 (2024)
24. Liu, Z., Wang, S.: Decoding chinese stock market returns: Three-state hidden semi-markov model. *Pacific-Basin Finance Journal* 44, 127–149 (2017)
25. Lu, M., Xu, X.: Trnn: An efficient time-series recurrent neural network for stock price prediction. *Information Sciences* 657, 119951 (2024)
26. Lu, X., Tian, J., Liao, Q., Xu, Z., Gan, L.: Cnn-lstm based incremental attention mechanism enabled phase-space reconstruction for chaotic time series prediction. *Journal of Electronic Science and Technology* 22, 100256 (2024)
27. Medrano, C.L., Alonso, G.D.C., Lafarga, O.V., Cervantes, J.L.V.: Mcclusteringsm: An approach for the multicriteria clustering problem based on a credibility similarity measure. *Computer Science and Information Systems* 21(3), 1147–1177 (2024)
28. Meena, H.K., Sharma, R., Tailor, A., Verma, H., Saini, R.: Detection of epilepsy using graph signal processing of eeg signals with three features. In: *Pattern Recognition and Data Analysis with Applications*, pp. 569–578. Springer (2022)
29. Panda, N., Muduli, D., Parija, S., Sharma, S.: Enhanced multi-class brain tumor classification model using mri images: A wavelet and deep learned features based approach. In: 2025 International Conference on Microwave, Optical, and Communication Engineering (ICMOCE). pp. 1–5. IEEE (2025)
30. Park, S. and Lee, D., Chu, W.: Fast retrieval of similar subsequences in long sequence databases. In: Proceedings 1999 Workshop on Knowledge and Data Engineering Exchange (KDEX'99)(Cat. No. PR00453). pp. 60–67. IEEE (1999)
31. Pedrycz, W., Homenda, W.: Building the fundamentals of granular computing: A principle of justifiable granularity. *Applied Soft Computing* 13(10), 4209–4218 (2013)
32. Shi, W., Azzopardi, G., Karastoyanova, D., Huang, Y.: Bidirectional piecewise linear representation of time series with application to collective anomaly detection. *Advanced Engineering Informatics* 58, 102155 (2023)

33. Sun, Y., Li, J., Liu, J., Sun, B., Chow, C.: An improvement of symbolic aggregate approximation distance measure for time series. *Neurocomputing* 138, 189–198 (2014)
34. Wang, H., Zhang, Y., Liang, J., Liu, L.: Dafa-bilstm: Deep autoregression feature augmented bidirectional lstm network for time series prediction. *Neural Networks* 157, 240–256 (2023)
35. Wang, J. and Liu, J., Jiang, W.: An enhanced interval-valued decomposition integration model for stock price prediction based on comprehensive feature extraction and optimized deep learning. *Expert Systems with Applications* 243, 122891 (2024)
36. Wang, X., Chuncao, L., Liu, Y., Liang, W., Kuanching, L., Poniszewska-Maranda, A.: A spatio-temporal graph neural network for eeg emotion recognition based on regional and global brain. *Computer Science and Information Systems* (00), 53–53 (2025)
37. Wen, D., Zhao, T., Fang, L., Zhang, C., Li, X.: Mwdinet: A multilevel wavelet decomposition interaction network for stock price prediction. *Expert Systems with Applications* 238, 122091 (2024)
38. Wu, H. and Salzberg, B., Zhang, D.: Online event-driven subsequence matching over financial data streams. In: *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*. pp. 23–34 (2004)
39. Xu, W., Wang, J., Zhang, Y., Li, J., Wei, L.: An optimized decomposition integration framework for carbon price prediction based on multi-factor two-stage feature dimension reduction. *Annals of Operations Research* 345(2), 1229–1266 (2025)
40. Yan, L., Jia, L., Guo, J., Lu, S.: A simulation study on the identification of eco-driving behaviour. *International Journal of Simulation Modelling* 21(3), 489–500 (2022)
41. Zhang, B., Ying, L., Yu, Y.: Novel symbolic aggregate approximation approach for time series based on trend features. *Journal of Computer Applications* 42(S1), 123–129 (2022)
42. Zhang, Q., Zhang, C., Cui, L., Han, X., Jin, Y., Xiang, G., Shi, Y.: A method for measuring similarity of time series based on series decomposition and dynamic time warping. *Applied Intelligence* 53(6), 6448–6463 (2023)
43. Zhang, Y., Li, J., Bai, H., Wang, Z.: Trend feature extraction method for time series based on turning point and trend segment. *Computer Applications* 40(S1), 92–97 (2020)
44. Zhao, D., Liu, S., Miao, Z., Zhang, H., Wei, Y., Xiao, S.: A novel feature extraction approach for mechanical fault diagnosis based on esax and bow model. *IEEE Transactions on Instrumentation and Measurement* 71, 1–11 (2022)
45. Zhou, M., Yi, J., Yang, J., Sima, Y.: Characteristic representation of stock time series based on trend feature points. *IEEE Access* 8, 97016–97031 (2020)

**Lei Han** is a PhD student at the School of Economics and Management, University of Science and Technology Beijing, with research interests in financial time series analysis and data mining.

**Xuedong Gao** is a professor at the School of Economics and Management, University of Science and Technology Beijing, with research interests in management process optimization, financial time series analysis, and data mining.

*Received: July 10, 2025; Accepted: October 16, 2025.*