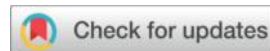




Securities Trading Prediction Using Feature Subsequence

Learning and Fuzzy Temporal Similarity



Aihua Zhu¹, Yaoxin Zhang¹, Haote Zhang¹,
Wanchun Sun¹, Dingkun Zhu^{1*}

^{1*}School of Computer Engineering, Jiangsu University of Technology,
Zhonglou, Changzhou, 213001, Jiangsu, China.

Contributing authors: may.zhu@vip.126.com; zyx_ysh@jsut.edu.cn;
zhangrichard168@gmail.com; sunwanchun@gmail.com;
zhudingkun@jsut.edu.cn

Abstract

This research introduces an innovative framework for securities trading prediction, designed to address the high-dimensional, nonlinear, and non-stationary characteristics of securities market data. The framework comprises three principal components: a self-supervised contrastive learning methodology for extracting multi-granularity event subsequences; a fuzzy temporal sequence similarity recognition algorithm that efficiently retrieves historical patterns; and a clustering-based prediction model capable of generating both short-term and long-term sequential forecasts. The empirical investigation focuses on the mid-to-high-end liquor sector stock group and the stock Yanghe shares as case studies. Experimental evaluations demonstrate that our approach surpasses existing benchmark models in predictive accuracy, exhibiting particular efficacy in multivariate time series similarity recognition and extended sequence forecasting tasks. Furthermore, the proposed methodology demonstrates superior performance in dimensionality reduction, temporal feature selection, computational efficiency, and result interpretability. This study establishes that the accurate identification and utilization of fuzzy temporal event features within historical trading data substantially enhances the prediction accuracy of securities price trends, thereby offering a promising new technical approach for quantitative investment decision-making.

Keywords: Securities Trading Prediction; Event Subsequence Extraction; Fuzzy Temporal Sequence Similarity

1 Introduction

Securities trading prediction represents a central research domain in financial market analysis, attracting significant attention from investors, financial institutions, and regulatory authorities. Securities markets exhibit quintessential complexity and profound uncertainty, characterized primarily by price volatility and numerous

interconnected influencing factors, making the forecasting of price movements exceptionally challenging.

Concomitantly, securities trading data inherently exhibits marked complexity, characterized by high dimensionality, non-linearity, and non-stationarity. High-dimensional data manifests as multivariate time series encompassing price metrics, transaction volumes, market depth indicators, and diverse technical parameters with intricate interdependencies. The relationships among these variables frequently transcend basic linear correlations, rendering traditional linear models inadequate for capturing their sophisticated internal structures. Furthermore, the statistical properties of securities data evolve over time, making conventional prediction methodologies based on stationary distribution assumptions often ineffective in achieving satisfactory predictive performance.

To address these challenges, researchers have systematically explored time series analysis methods. Feature representation has advanced from early global encoding approaches (e.g., variational autoencoders) to local feature extraction techniques, with Shapelet methods[1], [2], [3] establishing a paradigm for discriminative subsequence extraction that has recently expanded into unsupervised learning. Similarity recognition research has progressed from basic Euclidean distance metrics to Dynamic Time Warping (DTW) algorithms[4], [5] and subsequently incorporated fuzzy set theory[6], [7] to manage uncertainty. Despite these advances, current methodologies face significant limitations in practical applications: computational complexity hampers the efficient processing of extended time series, while existing similarity measures show limited capability in recognizing complex market patterns.

Based on this analysis, we propose a securities trading prediction framework using fuzzy temporal sequence similarity recognition designed to overcome existing limitations. This framework integrates three core technological components. First, We propose a Self-supervised Unsupervised Model for Common Feature Extraction (SUMCFE) based on contrastive learning methodology. This framework effectively extracts both common features shared across similar trading entities and unique features specific to individual entities, facilitating multi-level characterization of complex trading patterns within financial markets. Second, we design a fuzzy temporal sequence similarity recognition algorithm that encodes multidimensional temporal data using fuzzy set theory and combines this with efficient distance measurement techniques to facilitate rapid and accurate retrieval of similar historical patterns. Finally, we implement a clustering-based multi-step prediction model that supports forecasting from single-step to multi-step through systematic analysis of subsequent events in historically similar subsequences.

This research's primary contribution is a unified framework for feature subsequence extraction, fuzzy temporal event encoding, and similarity analysis. This framework enhances both the discriminative power and computational efficiency of securities trading prediction while providing a novel methodology for addressing financial time series complexity. Our findings advance multivariate time series analysis techniques theoretically while offering effective tools for quantitative investment and risk management in practice. The main contributions of this paper include:

(1) Development of a self-supervised multi-granularity subsequence extraction mechanism based on contrastive learning that effectively captures market patterns with significant predictive value through a common-distinctive feature separation strategy.

(2) Design of a temporal sequence similarity recognition algorithm incorporating fuzzy set theory, achieving efficient dimensionality reduction and rapid retrieval of high-dimensional non-linear data, thus providing high-quality reference samples for prediction.

(3) Proposal of a comprehensive FS-FTSP prediction framework that effectively addresses the complexity challenges of securities market data through systematic integration of feature extraction, similarity recognition, and clustering prediction modules.

2 related work

Multivariate time series analysis plays a crucial role in securities trading prediction. As financial markets grow increasingly complex and data dimensionality expands, traditional analytical methods face significant challenges, including the curse of dimensionality, computational inefficiency, and difficulties in pattern recognition. Recent research has advanced along two fundamental directions: multivariate temporal feature representation and temporal similarity identification, both aimed at enhancing the accuracy and reliability of securities trading predictions. This chapter systematically reviews research progress in these directions, analyzes their strengths and limitations, and establishes a theoretical foundation for subsequent investigations.

2.1 Multivariate Temporal Feature Representation

A core challenge in securities trading prediction is effectively representing multivariate time series data. As financial market data becomes increasingly high-dimensional and complex, traditional analytical methods face dimensionality problems and computational limitations, prompting researchers to explore more sophisticated feature representation techniques.

Early research on securities trading data with multivariate time series (MTS) characteristics employed encoding techniques for dimensionality reduction while preserving critical information, primarily focusing on global feature representation. However, conventional autoencoders exhibit information degradation when representing multivariate time series. To address this limitation, Li et al. introduced the Mutual Information Variational Autoencoder (MI-VAE), which extracts essential features through latent space learning and incorporates a mutual information term to optimize the loss function, enhancing the model's expressive capacity[8]. Despite MI-VAE's improved representational capabilities, it performs inadequately when handling missing values—a common issue in financial datasets. Addressing this practical challenge, Bianchi et al. developed a recurrent neural network-based autoencoder

architecture capable of simultaneously processing missing values while generating high-fidelity representations, substantially improving classification performance[9]. Although these encoding methodologies have advanced global feature representation, they struggle to effectively address uncertainty and fuzziness within time series data. To overcome this limitation, Liu Fang et al. proposed the Adaptive Fuzzy Population Coding (AFPC) method, which optimizes encoding parameters in Spiking Neural Networks (SNN), thereby enriching the distributional information of training data within spike time series[10]. This approach significantly enhances classification performance and improves generalization capability, offering novel insights into implementing global feature representation encoding through fuzzy set methodologies.

Global features often fail to capture discriminative local patterns in time series, an issue attracting increasing attention. To address this challenge, researchers have proposed subsequence feature extraction methods that identify critical local patterns. This approach effectively addresses high-dimensionality challenges of multivariate time series through feature extraction during preprocessing. Yoon et al. pioneered the CLeVer method based on principal component analysis for feature subset selection[11]. This method preserves correlation information in multivariate time series data, improving classification accuracy while reducing processing time. However, CLeVer requires predefined subsequence lengths, limiting its applicability. To overcome this limitation, Huang et al. proposed a data-driven time series segmentation method that adaptively partitions time series into varying-length subsequences with low computational complexity[12]. Nevertheless, extracting discriminative features from subsequences remains a primary challenge in current research.

The introduction of the Shapelet concept marked a significant breakthrough in subsequence feature identification. Ye and Keogh first defined Shapelets as subsequences that maximally distinguish between different time series classes, pioneering a new paradigm for discriminative subsequence-based feature extraction[13]. Hills et al. further advanced this approach by proposing a single-scan shapelet algorithm that improved classification accuracy while enhancing data interpretability through post-transformation clustering[3]. However, the Shapelet method requires exhaustive enumeration to search for optimal subsequences, resulting in high computational complexity that limits its application to large-scale datasets.

To address the computational complexity of Shapelets, Zhang et al. proposed an unsupervised multivariate time series representation learning framework based on Shapelets[14]. This framework innovatively extends the Shapelet concept to unsupervised learning by adaptively extracting discriminative Shapelets and incorporating self-supervised learning strategies, significantly reducing computational costs while improving model generalization. This approach demonstrated superior performance across various tasks including classification, clustering, and anomaly detection, validating its effectiveness as a general representation learning framework. Building upon this work, Cai et al. further refined the Shapelet selection process[15]. Their SE-Shapelets method employs Significant Subsequence Chains (SSC) and Linear Discriminant Selection (LDS) algorithms to effectively filter irrelevant subsequences

and identify representative temporal shape patterns, thereby improving time series clustering accuracy.

2.2 Multivariate Time Series Similarity Recognition

Building upon multivariate time series representation methods, the identification and measurement of temporal similarity has emerged as a critical issue in securities trading prediction. Oswiecimka et al. systematically investigated self-similar patterns in financial market dynamics, demonstrating their ubiquitous presence and multifractal properties[16], [17]. This finding provides important theoretical support for similarity-based stock prediction. However, effectively measuring similarity between time series remains challenging.

Early similarity measurement methods primarily relied on simple metrics such as Euclidean distance. However, these methods failed to handle temporal elasticity in time series, resulting in inaccurate identification results. To address this issue, Berndt and Clifford proposed the Dynamic Time Warping (DTW) algorithm[18]. This algorithm allows nonlinear alignment of time series along the temporal axis, effectively addressing temporal deformation problems. Nevertheless, DTW exhibits limitations when processing high-dimensional nonlinear data from financial markets, particularly when complex interactions exist among variables. To overcome this limitation, Moser et al. systematically investigated Multivariate Dynamic Time Warping (MDTW) methods, providing a more comprehensive solution for similarity measurement in multivariate time series[5]. Despite advances in handling temporal elasticity, MDTW still falls short in capturing high-dimensional nonlinear patterns.

As understanding of local feature importance deepened, researchers began exploring subsequence-based similarity measurement methods. Li et al. transformed key point subsequences from stock price series into graph structures, enhancing similarity measurement capabilities for multidimensional temporal data and achieving the highest average net value in trading simulations[19]. This method precisely captures local structural features while successfully modeling complex relationships between nodes, improving similarity identification accuracy. Concurrently, Dorr et al. proposed an algorithm based on similar continuous subsequence relationships[20]. By constructing directed acyclic graphs (DAGs) to capture similar patterns between multivariate sequences and form pattern aggregates, this algorithm demonstrates the advantages of utilizing subsequence relationships for multivariate temporal similarity measurement.

Faced with limitations of traditional methods, fuzzy approaches have emerged as an important direction in similarity identification research. Liu et al. proposed a prediction model based on fuzzy information granularity support vector machines to address high uncertainty in financial crisis contagion prediction[10]. This model improves prediction accuracy by estimating stock index boundaries through fuzzy granularity and SVM. However, the method shows limited performance when confronted with inconsistent data distributions. Addressing this limitation, Behbood et al. developed a domain adaptation algorithm based on fuzzy refinement[7]. This

algorithm modifies target instance labels through fuzzy systems and similarity/dissimilarity concepts, effectively addressing the decline in prediction accuracy caused by inconsistent data distributions.

Fuzzy methods have demonstrated greater potential in high-dimensional time series prediction. Bitencourt et al. combined data embedding transformation with Fuzzy Time Series (FTS) methods to propose the Embedding FTS (EFTS) methodology[21]. This approach significantly enhances prediction accuracy for high-dimensional time series while maintaining model parsimony. From an ensemble learning perspective, Papageorgiou et al. proposed a method integrating evolutionary Fuzzy Cognitive Maps (FCMs), artificial neural networks, and hybrid structures, achieving higher prediction accuracy in time series forecasting[22].

3 Materials and Methods

In this section, we propose the Feature Subsequence-based Fuzzy Temporal Sequence Prediction model (FS-FTSP), as shown in Fig. 1. The first module is the special transaction subsequence extraction module, which identifies and extracts special transaction event subsequences with operational value from historical transaction data, primarily developed by improving upon the concept of time series shapelets. The second module is the event sequence similarity identification module, designed to capture sequences similar to the target event sequence in historical data, ensuring quick and accurate identification of all similar transaction event subsequences. The third module is the multivariate time series prediction module, implemented based on statistical analysis of subsequent events following historically similar subsequences.

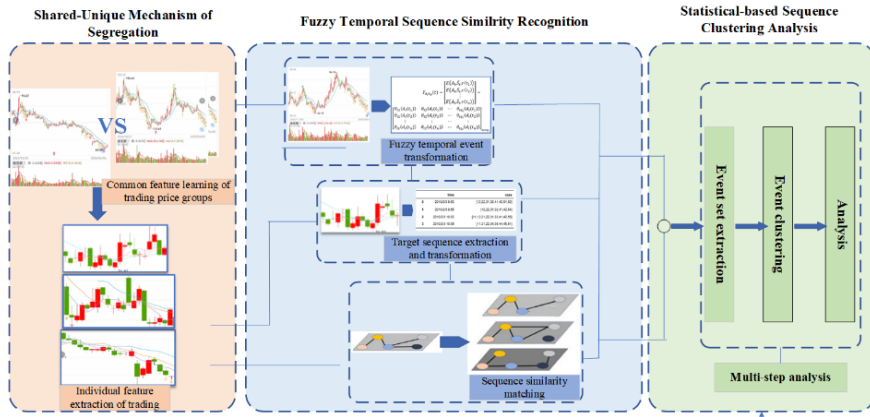


Fig.1 The framework of our method.

3.1 Special transaction subsequence extraction

The transaction subsequence extraction mechanism is the core component of this research, aimed at extracting transaction patterns and event segments with significant predictive value from massive historical transaction data. Inspired by time series shapelets proposed by Ye and Keogh, this mechanism has been systematically

improved to extend from univariate to multivariate time series domains. We have constructed a common and Unique features Unsupervised learning Model for Securities (SUMS) applicable to securities transaction data, capable of effectively extracting both common features and unique features among similar trading objects. This mechanism is implemented through two key methodological components: (1) identifying significant common features in multivariate time series, constructing standardized representations of recurrent transaction patterns, and developing corresponding feature description models; and (2) comparative analysis identifying anomalous sequence segments that deviate from global common patterns, mining unique features of individual time series to capture the differential characteristics among trading partners.

Given that the dataset of any object in the input group is a multivariate time series dataset X_i , where X_i contains a total of L temporal events, each event having C dimensions (variables), i.e., $X_i \in R^{C \times L}$. A sliding window operation is performed on X_i , with window size w and sliding step s , where $L - w + 1$ represents the number of windows that can be extracted from X_i . Any time series captured by a window can be regarded as a subsequence $X_i^{(t)}$ of sequence X_i .

1) Common feature extraction

The system randomly selects feature subsequences (Shapelets) and calculates the distance between each time series window $X_i^{(t)}$ and the selected Shapelets. Those

$X_i^{(t)}$ with minimum distances to all iterated Shapelets are chosen as significant common features to describe that type. For distance measurement, three metrics are selected based on the model proposed by Liang et al.[14]: Euclidean distance $D_{euclidean}$, cosine similarity D_{cosine} , and cross-correlation $D_{cross-correlation}$. Feature subsequences are extracted using these three different distance measurements and then concatenated to form the final significant common feature matrix, denoted as F_{common} :

$$F_{shared} = \text{concat}(F_{euclidean}, F_{cosine}, F_{cross-correlation}) \quad (1)$$

2) Unique feature extraction

Feature extraction is performed on the common features $F_{shared,train}$ of the training data to capture the global feature distribution of similar objects. The mean vector and covariance matrix are calculated:

$$\mu_{shared} = \frac{1}{N} \sum_{i=1}^N F_{shared,train}^i \quad (2)$$

$$\Sigma_{shared} = \frac{1}{N-1} \sum_{i=1}^N (F_{shared,train}^i - \mu_{shared})(F_{shared,train}^i - \mu_{shared})^T \quad (3)$$

The individuality based on the group's common features is calculated by quantifying the deviation degree of the common features of the test individual sample $F_{shared,test}$ using the Mahalanobis distance:

$$A_{shared}(F_{shared,test}^i) = \sqrt{(F_{shared,test}^i - \mu_{shared})^T \Sigma_{shared}^{-1} (F_{shared,test}^i - \mu_{shared})} \quad (4)$$

The individuality based on unique features is calculated by computing the cosine similarity between the unique features and common features of the test individual sample $F_{shared,test}$

$$A_{unique}(F_{shared,test}^i) = - \frac{\langle F_{shared,test}^i, \mu_{shared} \rangle}{\|F_{shared,test}^i\|^2 \|\mu_{shared}\|^2}. \quad (5)$$

Calculating the comprehensive uniqueness score: By combining the uniqueness of common features and unique features, the total uniqueness score is obtained:

$$A(F_{shared,test}^i) = \lambda A_{shared}(F_{shared,test}^i) + (1 - \lambda) A_{unique}(F_{shared,test}^i), \quad (6)$$

where $\lambda \in (0,1)$ is the weight parameter between the two.

Algorithm 1 Extract common-Unique Multivariate Shapelets

Require: Raw data $X \in \mathbb{R}^{N \times C \times L}$, Shapelet configuration $\{(L_r, K_r)\}$, Augmentation library A, Hyperparameters: τ (temperature), λ (alignment weight), λ_s (ortho-reg weight)

Ensure: Learned shapelets: FilteredShapelets

- 1: $X_win \leftarrow \text{SlidingWindow}(X, w)$
 $X_norm \leftarrow \text{MinMaxScale}(X_norm)$
 - 2: $S_r \sim U(-1, 1)$ for $r = 1$ to R // Initialize R shapelets
 - 3: for epoch = 1 to Epoch do
for batch X_b in X_norm do
 $X', X'' \leftarrow \text{RandomAugment}(X_b, A), \text{RandomAugment}(X_b, A)$
//Augmented views
 $Z', Z'' \leftarrow \emptyset$
For $r = 1$ to R do
 $z'_r \leftarrow [\min_t \|X'[t:t+L_r] - s\|_2 \text{ for } s \in S_r]$
 $z''_r \leftarrow [\min_t \|X''[t:t+L_r] - s\|_2 \text{ for } s \in S_r]$
end for
 $L_c \leftarrow \text{InfoNCE}(\text{concat}(Z'), \text{concat}(Z''), \tau)$ // Coarse-grained
 $L_f \leftarrow \sum_r \text{InfoNCE}(z'_r, z''_r, \tau)$ // Fine-grained
 $L_a \leftarrow \text{AlignmentLoss}(Z', Z'') + \lambda_s \text{OrthoReg}()$
Update $\theta \leftarrow \text{SGD}(\theta, L_c + L_f + \lambda L_a)$ //Update parameters
end for
end for
 - 4: for x_test in X_test do
shapelets, indices $\leftarrow f_\theta(x_test)$ // Trained encoder f_θ
FilteredShapelets $\leftarrow \text{Filter1}(\text{shapelets}, \text{indices})$ // The most repeated shapelets
FilteredShapelets' $\leftarrow \text{Filter2}(\text{FilteredShapelets}, \text{indices})$ // Max distance filter
SaveShapeletPatterns(FilteredShapelets', "output")
end for
-

3.2 Fuzzy Temporal Sequence Similarity Recognition Model

The objective of the event sequence similarity recognition phase is to rapidly and accurately retrieve sequences from historical data that are highly similar to the target event, for use in subsequent predictions. This paper adopts a fuzzy temporal sequence similarity recognition method, with the fuzzy temporal event similarity measurement function as its core.

Based on the fuzzy temporal event[23] setting, $(A_l, s_{ij}, d_i(t_k), v(t_k))$ represents any fuzzy temporal event in space, describing that the object A_l has state s_{ij} in property S_i at temporal factor $v(t_k)$, with a value of $d_i(t_k)$. The state of the event is determined by the fuzzy membership value $\theta_{ij}(d_i(t_k))$ of attribute $d_i(t_k)$ to attribute state s_{ij} , which is the event function $E(A_l, s_{ij}, d_i(t_k), v(t_k))$. When an event specifically belongs to a certain attribute state, the fuzzy temporal event can be described using the event's membership state or state code: $\tilde{E}(A_l, \bar{S}_i(\alpha_i), v(t_k)) = [s_{ij}or0]_{1 \times j} = [ijor0]_{1 \times j}$.

For example, if a certain stock has an opening price increase of +3% on November 11, 2021, and it's uncertain whether this belongs to a *small price increase*, then the event function $E(\text{stock}, \text{smallpriceincrease}, +3\%, 2021 - 11 - 11)$ can be calculated for the fuzzy membership degree $\theta_{\text{smallpriceincrease}}(+3\%)$ to the attribute state of *small opening price increase* (coded as 12).

If the value of $\theta_{12}(+3\%)$ is greater than the fuzzy threshold α_i (where $\alpha_i \in [0,1]$), then $E(\text{stock}, \text{smallpriceincrease}, +3\%, 2021 - 11 - 11)$ is represented by 12, indicating the event: the stock experienced a *small opening price increase* on 2021-11-11.

Let the input target fuzzy temporal event sequence be M , $M = [\tilde{E}_m]_{1 \times m}$ ($1 \leq m \leq g$), where g is a constant representing the maximum number of fuzzy temporal events contained in the target sequence. To facilitate distinction from ordinary fuzzy temporal events, the target fuzzy temporal event sequence is represented as $\tilde{E}(A_l, \bar{S}_i(\alpha_i), v(t_k)) = [\tilde{s}_{ij}or0]_{1 \times g}$.

It is established that the similarity between fuzzy temporal events can be approximately measured by the distance between these fuzzy temporal events.

Definition 3.1 The similarity between different states of the same attribute can be approximately measured by the distance between states, denoted as D , with the calculation formula as follows:

$$D(s_{ij_1}, s_{ij_2}) = \sqrt{(s_{ij_1} - s_{ij_2})^2} = \sqrt{(10j_1 - 10j_2)^2} \quad (7)$$

Definition 3.2 The difference between any two fuzzy temporal events, or the distance between them, can be represented by the sum of the distances between the corresponding attribute states contained in the events, denoted as $Dist$. The distance calculation between any event and the target event can be expressed as:

$$Dist(E_n, \tilde{E}_m) = |E_n - \tilde{E}_m| = \left[\sum_{i=1}^p (s_{ij} - \tilde{s}_{ij})^2 \right]^{\frac{1}{2}} \quad (8)$$

Definition 3.3 Let the maximum possible event distance be denoted as D_{\max} . The matching similarity between any two fuzzy temporal events can be represented by the ratio of the difference between their event distance and the maximum event distance,

denoted as R . The calculation of the matching similarity between any event and the target event is expressed as:

$$R(E_n, \tilde{E}_m) = 1 - \frac{Dist(E_n, \tilde{E}_m)}{D_{max}} \quad (9)$$

Definition 3.4 Set the minimum fuzzy similarity threshold as η . When $R \geq \eta$, the two fuzzy temporal events can be considered similar.

Algorithm 2 Fuzzy Time Series Similarity Recognition Algorithm

Require: Fuzzy time series event list ori_data, settings mask_data, error_point_limit

Ensure: Similar event index list temp_index, save_dict

- 1: j_index_list = Get a sublist of ori_data's index, excluding the last (mask_data.shape[0]-1) elements
 save_dict = {'iter': [], 'index': []}
 bias_count = 0
 - 2: for i_index in mask_data.index:
 a_data = mask_data.loc[i_index, 'case']
 temp_index = [
 j_index for j_index in j_index_list
 if sum(1 for k in range(len(mask_data))
 if (a_data[k] or ori_data.loc[j_index + bias_count, 'case'][k])
 and not (a_data[k] and ori_data.loc[j_index + bias_count, 'case'][k])
 <= error_point_limit)
 save_dict['iter'].append(bias_count + 1)
 save_dict['index'].append(temp_index)
 j_index_list = temp_index
 bias_count += 1
 - 3: Replace NaN in data with 30
 count_num = []
 while True:
 mean_point = np.nanmean(data[:, 1:], axis=0)
 distance_value = np.sqrt(np.sum(np.square(data[:, 1:] - mean_point), axis=1))
 count_num.append((mean_point, np.median(distance_value),
 np.mean(distance_value), np.max(distance_value)))
 if np.max(distance_value) < distance_limit:
 break
 - 4 data = np.delete(data, np.argmax(distance_value), axis=0)
-

3.3 Sequence Clustering Prediction and Multi-step Prediction Model

The prediction module constructs a statistical analysis framework based on historical similarity, utilizing high-quality historically similar subsequences from previous steps to predict future trends of the target sequence. This module employs a similarity clustering algorithm to analyze historical data of event sequences, identifying the most frequently occurring event categories following the target event sequence M , and determining them as the most likely subsequent event types. This method provides both single-step event prediction and multi-step event sequence prediction, offering comprehensive temporal development references for securities trading decisions. Specifically, it includes:

1) Event Clustering Based on Similarity

The similarity between fuzzy temporal events can be approximately measured by the distance between them. For all historical records matching the target event sequence $M = \{E_n, E_{n+1}, \dots, E_{n+g}\}$, we extract the set of events $\{E_{n+g+1}\}$:

$$\{E_{n+g+1}\} = E_{n+g+1}^{(1)}, E_{n+g+1}^{(2)}, \dots, E_{n+g+1}^{(o)}$$

Where any immediate subsequent event $E_{n+g+1}^{(o)}$ has the attribute state vector $S_i = [s_{i1}, s_{i2}, \dots, s_{ip}]$, where p is the dimension of the state.

Calculate the mean of each attribute for all events in $\{E_{n+g+1}\}$, denoted as $\bar{S}_i = \bar{s}_{ij} = \frac{1}{q} \sum_{j=1}^q s_{ij}$. Combine all attribute means to form the mean fuzzy temporal event $\bar{E}_1 = [\bar{s}_{ij}]_{1 \times p}$. Calculate the similarity between each event $E_{n+g+1}^{(i)} \in \{E_{n+g+1}\}$ and the fuzzy mean event \bar{E}_1 :

$$R(E_{n+g+1}, \bar{E}_1) = 1 - \frac{Dist(E_{n+g+1}, \bar{E}_1)}{D_{max}}, \quad (10)$$

Perform hierarchical clustering on all $R(E_{n+g+1}, \bar{E}_1)$, extract cluster events based on the distance threshold, and retain the new event set.

Recalculate the fuzzy mean event $\bar{E}_2 = [\bar{s}_{ij}]_{1 \times p}$ for the filtered event set, and repeat the above steps until the minimum similarity of the remaining events meets the constraint condition: $R_{min} \geq \eta$, ending the iterative update of the fuzzy mean event. The final remaining event set is:

$$\{E_{n+g+1}^{final}\} = \{E_{n+g+1}^{(o)} \mid R(E_{n+g+1}^{(o)}, \bar{E}_k) \geq \eta\}, \quad (11)$$

These events are the most likely immediate subsequent event types after the occurrence of the target event sequence M .

The occurrence probability of the immediate subsequent event E_{n+g+1}^{final} can be obtained through the ratio of the number of events after clustering to the number of original events. For example, $P(E_{n+g+t})$ represents the occurrence probability of the predicted event at step t :

$$P(E_{n+g+t}) = \frac{|\{E_{n+g+1}^{final}\}|}{|E_{n+g+t}|}, \quad (12)$$

Therefore, $P(E_{n+g+t})$ is also referred to as the prediction accuracy.

2) Multi-step prediction

If the prediction of multiple subsequent event sequences following the target event sequence M is required, then based on each step of prediction, recursive clustering

prediction of subsequent events is performed. For example, E_{n+g+2} must originate from the immediate subsequent event of E_{n+g+1} , satisfying: $\{E_{n+g+2}\} = \{E_{n+g+2} | \exists E_{n+g+1} \in \{E_{n+g+1}\}_{final}\}$. Similarly, for multi-step prediction, the prediction accuracy of the subsequent step also needs to be multiplied by the occurrence probability of the previous event, referred to as the posterior prediction accuracy:

$$P(E_{n+g+t}) = \prod_{t=1}^k P(E_{n+g+t}) \quad (13)$$

By analogy, all events occurring after the target event sequence M can be predicted. However, it is recommended that the length k of the multi-step predicted sequence should not exceed the length g of the target event sequence M itself.

Algorithm 3 Fuzzy Time Series Similarity Prediction Mining Algorithm

Require: match_results, reference_data

Ensure: stats_predict

- 1: Load match_results into result_data
 - 2: Convert result_data to results_table (rows: iterations, columns: indices)
 - 3: For iter_i in results_table.index:
 - Initialize temp_list as an empty list.
 - Print debug: "Processing iteration {iter_i}".
 - 4: For col_j in results_table.columns:
 - cell_data \leftarrow results_table[col_j][iter_i].
 - If cell_data is NaN or empty, continue to the next column.
 - If cell_data is a list/array:
 - Append all elements of cell_data to temp_list.
 - Else:
 - Append cell_data to temp_list.
 - 5: If temp_list is empty:
 - Print debug: "No data for iteration {iter_i}".
 - Continue to the next iteration.
 - 6: temp_array \leftarrow Convert temp_list to NumPy array.
 unique_elements \leftarrow Unique elements in temp_array.
 stats_dict \leftarrow Compute counts of each element in unique_elements.
 - 7: stats_predict \leftarrow Convert stats_dict to DataFrame.
 - 8: Return stats_predict
-

4 Experiment

4.1 Individual Stock Trading Event Feature Subsequence Mining

This study examines the mid-to-high-end liquor sector stock group, focusing on six representative stocks with price ranges between 100-200 yuan: Shanxi Fenjiu (600809.SH), Luzhou Laojiao (000568.SZ), Gujing Gongjiu (000596.SZ), Jiugui Liquor (000799.SZ), Jinshiyuan (603369.SH), and Yanghe Shares (002304.SZ). The research utilizes minute-level trading data for these stocks from 2010-2019, comprising six dimensions: trading time, opening price (OP), highest price (HP), lowest price (LP),

closing price (CP), and trading volume (TV). All trading data was sourced from China's major securities exchanges and acquired through the open-source Python financial data interface package Tushare.

According to the analysis of the relationship between temporal types and prediction accuracy shown in Fig. 2, this research selected a 5-minute level temporal granularity for the experiment. Accordingly, all minute-level trading data was preprocessed into 5-min temporal data. The model's key parameters include: a sliding window (window_size) of 12, minimum Shapelets length of 3, and shapelet_lengths = [3,4,5,6,7,8,9] dynamically set according to window_size. To enhance model robustness, a mixed distance measurement (dist_measure) method incorporating three approaches was implemented for similarity calculation.

This experiment utilized 10-year trading data from the first five stocks as the training set for the multivariate time series feature subsequence recognition model to extract common group features and construct trading patterns for the mid-to-high-end liquor sector. Yanghe's 10-year trading data served as the test set to identify unique subsequence features of this individual stock.

Results indicate that the optimal Shapelets length for Yanghe is 5, with the five highest-scoring multivariate time series subsequences at this length shown in the table 1. Considering trading session continuity, the trading sequence from 10:00-10:20 on June 9, 2010, was identified as Yanghe's unique feature sequence relative to the mid-to-high-end liquor sector group. This subsequence pattern exhibited distinct specificity in Yanghe's trading data, differing significantly from trading patterns of other stocks in the sector.

Table 1. Characteristic subsequences in individual stock trading event

	date	start time	end time	window size
1	2010/6/9	10:00	10:20	5
2	2012/3/16	11:10	13:00	5
3	2012/4/20	14:10	14:30	5
4	2012/5/15	9:40	10:00	5
5	2012/5/16	9:35	9:55	5

4.2 Fuzzy Temporal Sequence Similarity Recognition Mining

4.2.1 Data Preparation

The minute-level trading data of Yanghe over 10 years was converted into temporal format. By calculating the ratio of difference between the current attribute value and the previous temporal value, the state values of event attributes were obtained. These state values effectively describe the trend states of opening price(OP), highest price(HP), lowest price(LP), closing price(CP), and trading volume(TV). The selected 1-min, 3-min, 5-min, 10-min, and 15-min temporal types were used as event temporal

types for temporal event conversion. Table 2 shows the state values of various attributes for the selected unique trading feature events under the 5-min temporal type.

Table 2. Partial Trading Events of Yanghe Under 5-Min Temporal Type

Time	Opening Price	Highest Price	Lowest Price	Closing Price	Trading Volume	Fuzzy Temporal Event (case)
2010/6/9 10:00	0.0069	-0.0001	-0.0004	0.0042	-0.1042	[14, 15, 24, 34, 35, 44, 52]
2010/6/9 10:05	-0.0024	0.0096	0.0118	0.0003	1.0000	[12, 25, 34, 35, 45, 55]
2010/6/9 10:10	0.0124	0.0049	0.0028	0.0113	0.1030	[15, 24, 25, 35, 44, 45, 54]
2010/6/9 10:15	0.0046	0.0058	0.0055	0.0050	-0.0308	[14, 15, 24, 25, 34, 35, 44, 45, 52]
2010/6/9 10:20	0.0056	-0.0056	0.0000	0.0006	-0.0768	[14, 15, 34, 35, 41, 42, 52]

The temporal event attribute states were discretized and labeled using fuzzy set theory and clustering methods. The Goodness of Variance Fit (GVF) method applied to preprocessed data determined an optimal classification of 5 categories for each attribute. Price and volume attribute states were discretized into five fuzzy states: "Large Decrease ($\downarrow L$)", "Small Decrease ($\downarrow S$)", "Neutral (M)", "Small Increase ($\uparrow S$)", and "Large Increase ($\uparrow L$)". The trapezoidal function was selected as the membership function based on common adaptation scenarios. Using $K=5$, the K-means algorithm identified cluster centers for attribute states, with the trapezoid's upper base formed by extending the cluster center by 10% on both sides, while the lower base corresponded to adjacent cluster center points. Table 2 presents the final fuzzy attribute state domain value divisions, corresponding intervals, and symbols, where symbol j in state S_{ij} represents attributes (S_i): opening price (1), highest price (2), lowest price (3), closing price (4), and trading volume (5).

Table 3. Correspondence table of attribute state domain value intervals and symbols

Status	Price		Status	Trading Volume	
	Range	Symbol(j)		Range	Symbol(j)
P \downarrow L	[-0.1, -0.0011]	1	TV \downarrow L	[- ∞ , -0.1782]	1
P \downarrow S	[-0.0051, 0]	2	TV \downarrow S	[-0.5297, 0]	2
PM	0	3	TVM	0	3
P \uparrow S	[0, 0.0051]	4	TV \uparrow S	[0, 0.5297]	4
P \uparrow L	[0.0011, 0.1]	5	TV \uparrow L	[0.1782, ∞]	5

This experiment converted original stock trading data into fuzzy temporal event symbols to generate standardized event symbol sequences. Specific time period sequences were extracted as target event sequences, as displayed in the fuzzy temporal event column in Table 3. Utilizing Yanghe's 10-year historical data, the fuzzy temporal sequence similarity recognition model was applied for similar matching of target event sequences and subsequent temporal event prediction. Experimental parameters included a fuzzy threshold of 0.5, maximum allowable error points of 2.5, and maximum distance limit of 30. These parameter controls facilitated accurate sequence matching and prediction.

4.2.2 Experimental Results and Analysis

(1) Recognizability of Similar Fuzzy Temporal Event Sequences

The experiment identified 8,634 fuzzy temporal event sequence sets similar to the target sequence events. Table 4 presents the 5 sequences with the highest and lowest matching similarity. Even with complex target sequence events, the algorithm-identified similar sequences maintained consistency with the target sequence events' rise and fall patterns across all attributes. In the lowest similarity cases, inconsistencies in rise and fall patterns occurred in at most one attribute. The similarity values in Table 5 demonstrate that not only were individual events highly consistent with the target events, but the fuzzy sequence sets composed of four events also highly matched the target sequence sets, with minimum overall similarity exceeding 0.6. These results confirm the effectiveness of the fuzzy temporal event sequence similarity recognition algorithm for stock event pattern recognition, validating its capability to identify similar stock patterns.

Table 4. Partial table of sequences with highest and lowest similarity to the target sequence

Index	Time	Event1	Event2	Event3	Event4	Event5
Target	2010/6/9 10:00	[14, 15, 24, 34, 35, 44, 52]	[12, 25, 34, 35, 45, 55]	[15, 24, 25, 35, 44, 45, 54]	[14, 15, 24, 25, 34, 35, 44, 45, 52]	[14, 15, 34, 35, 41, 42, 52]
1	2019/4/4 13:10	[14, 15, 24, 34, 35, 44, 45, 52]	[12, 25, 34, 35, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 54]	[14, 15, 24, 25, 34, 35, 44, 45, 54, 55]	[14, 15, 34, 35, 41, 42, 51, 52]
2	2017/10/17 10:30	[14, 15, 24, 34, 35, 44, 52]	[12, 24, 25, 34, 35, 44, 45, 54, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 54, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 52]	[14, 15, 34, 35, 44, 51, 52]
3	2011/6/28 13:40	[14, 15, 24, 34, 35, 44, 52]	[14, 24, 25, 34, 44, 45, 55]	[14, 15, 24, 25, 34, 35, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 52]	[14, 15, 34, 35, 42, 51]
4	2015/9/15 14:40	[14, 15, 24, 25, 34, 44, 52]	[12, 24, 34, 44, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 51, 52]	[15, 22, 34, 35, 41, 42, 52]
5	2018/6/12 10:20	[14, 15, 24, 25, 35, 44, 52]	[12, 24, 25, 44, 45, 54, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 52]	[14, 15, 34, 42, 51]
1	2013/7/30 14:15	[12, 24, 25, 34, 35, 44, 45, 54, 55]	[14, 15, 24, 25, 34, 35, 42, 55]	[14, 15, 21, 22, 34, 35, 44, 51]	[12, 22, 34, 44, 45, 51]	[14, 15, 34, 44, 45, 55]
2	2012/1/5 14:55	[11, 12, 24, 44, 51]	[14, 15, 22, 34, 35, 44, 45, 55]	[14, 15, 22, 32, 44, 55]	[11, 12, 24, 34, 35, 41, 42, 52]	[15, 24, 25, 34, 35, 45, 51]
3	2010/3/5 14:55	[12, 22, 34, 35, 44, 51, 52]	[14, 15, 25, 34, 35, 45, 51, 52]	[15, 34, 35, 41, 42, 54, 55]	[11, 12, 21, 22, 34, 44, 51, 52]	[14, 34, 44, 45, 55]
4	2017/10/11 14:35	[14, 15, 24, 25, 34, 35, 44, 45, 55]	[14, 15, 24, 34, 35, 44, 45, 51]	[14, 15, 24, 31, 32, 42, 54]	[12, 24, 25, 34, 35, 44, 45, 55]	[14, 15, 24, 25, 34, 35, 44, 45, 51, 52]
5	2011/5/13 14:15	[11, 12, 25, 34, 35, 44, 45, 55]	[15, 24, 34, 35, 44, 45, 51]	[11, 12, 22, 34, 35, 44, 45, 55]	[14, 15, 24, 25, 44, 45, 52]	[14, 15, 25, 34, 35, 45, 51, 52]

Table 5. Similarity values of partially matched sequences

Index	Time	Event1 Similarity	Event2 Similarity	Event3 Similarity	Event4 Similarity	Event5 Similarity	Sequencev Similarity
Target Sequence	2010/2/3 9:50	1.00	1.00	1.00	1.00	1.00	1.00
1	2010/12/20 9:50	0.95	1.00	0.93	0.75	0.95	0.92
2	2015/8/13 11:05	1.00	0.91	0.91	1.00	0.75	0.91
3	2013/10/23 11:00	1.00	0.78	0.87	1.00	0.89	0.91
4	2014/9/16 10:10	0.93	0.88	0.88	0.95	0.89	0.90
5	2012/11/12 13:15	0.93	0.83	0.88	1.00	0.88	0.90
1	2014/2/19 10:45	0.64	0.61	0.57	0.63	0.57	0.60
2	2019/3/20 10:30	0.65	0.61	0.59	0.57	0.60	0.60
3	2018/7/2 13:50	0.68	0.57	0.66	0.57	0.57	0.61
4	2011/12/15 10:10	0.69	0.52	0.56	0.61	0.66	0.61
5	2013/7/10 10:15	0.56	0.49	0.56	0.85	0.59	0.61

(2) Predictability of Subsequent Events in Fuzzy Temporal Event Sequences

Empirical evidence confirms the effectiveness of fuzzy temporal sequence similarity algorithms for forecasting stock trading events. This predictive methodology, validated through application to Yanghe data, reveals valuable patterns that enhance stock event prediction capabilities.

Firstly, prediction accuracy correlates significantly with the temporal granularity of fuzzy temporal events. Fig. 2 demonstrates an inverse relationship: as temporal resolution increases from 1-min to 15-min temporal, predictive accuracy diminishes substantially. Generally, finer temporal granularity produces superior predictive precision.

Although this pattern persists across all temporal classifications, the 5-min temporal resolution demonstrates exceptional performance in predicting subsequent events within target sequences, surpassing even finer temporal resolutions such as the 1-min temporal. Based on these observations with Yanghe data, we conclude that employing a 5-min temporal resolution for fuzzy temporal data mining optimizes latent information extraction. Consequently, when analyzing diverse subjects, researchers must identify the optimal temporal resolution, as this parameter fundamentally influences both similarity mining efficacy and predictive outcomes. Notably, predictive accuracy for immediately subsequent fuzzy temporal events deteriorates rapidly as temporal intervals extend.

Secondly, target sequence event selection significantly influences prediction outcomes and accuracy. Fig. 3 illustrates that when examining target sequences with continuous downward price trends—specifically (41,41,41) and (41,41,41,41)—the four consecutive price decreases display a more pronounced downward trajectory than three consecutive decreases. Interestingly, prediction accuracy for events following the

four-decline sequence is markedly lower than that of the three-decline sequence, and inferior to the four-element target sequence (41,41,41,45).

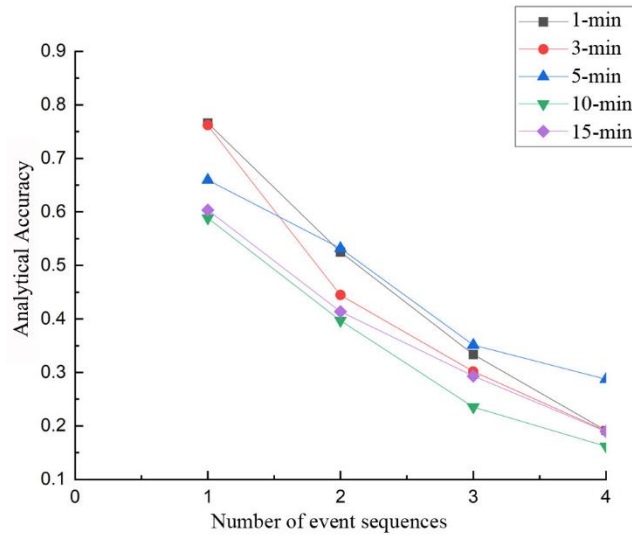


Fig. 2. Relationship between Different Temporal Types and Prediction Accuracy.

This phenomenon suggests that for Yanghe, increased uniformity and prominence in early trading patterns correlates with decreased predictability of subsequent trading behaviors. Conversely, sequences featuring trend reversals, such as (41,41,41,45)—characterized by consistent initial trends followed by an abrupt reversal—constitute distinctive formations where the terminal reversal carries significant directional implications, yielding superior accuracy in subsequent event prediction.

Therefore, when implementing this algorithm to identify similar patterns and forecast subsequent events, the careful selection of initial target sequences becomes paramount, as it directly impacts both the efficacy of fuzzy temporal sequence similarity mining and the reliability of predictive outcomes.

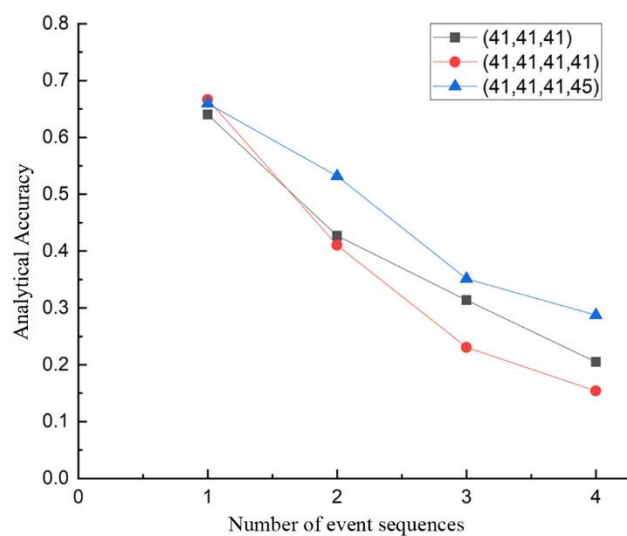


Fig. 3. Relationship between Different Target Sequence Events and Prediction Accuracy.

(3) Practical Application of Fuzzy Temporal Event Sequence Similarity Recognition and Prediction

Under the 5-min temporal type, we predicted four consecutive event sequences following the target sequence set. During the prediction process, the occurrence of attribute states in any subsequent event must be predicated on the occurrence of the immediately preceding event. The posterior predictions and accuracy of attribute states for each event are shown in Table 6.

Table 6. Partial List of Posterior Prediction Accuracy for Fuzzy States of Predicted Events

Fuzzy State	Event1	Event2	Event3	Event4
12	0.03	0.08	0.10	0.12
14	0.54	0.68	0.50	0.58
15	0.48	0.58	0.32	0.33
21	0.00	0.03	0.04	0.06
22	0.00	0.10	0.16	0.15
24	0.56	0.50	0.42	0.45
25	0.37	0.31	0.30	0.27
32	0.02	0.06	0.02	0.03
34	0.52	0.65	0.48	0.70
35	0.54	0.52	0.32	0.42
41	0.02	0.02	0.00	0.00
42	0.11	0.24	0.02	0.15
44	0.50	0.53	0.56	0.67
45	0.37	0.34	0.40	0.42
51	0.32	0.26	0.24	0.39
52	0.45	0.35	0.34	0.39
54	0.11	0.29	0.20	0.21
55	0.06	0.27	0.16	0.24

In Yanghe stock trading, once the target event sequence occurs, the first trading event immediately following it (i.e., subsequent event 1) is likely to exhibit concurrent increases in both volume and price. In subsequent event 1, the posterior prediction accuracies for small and large increases in the lowest price are 0.52 and 0.54, respectively. The opening price, highest price, and closing price are also predicted to rise, though with a higher probability of small increases.

Based on the actual occurrence of subsequent event 1 predictions, subsequent event 2 shows posterior prediction accuracies greater than 0.6 and 0.5 for small and large increases in opening price and lowest price, respectively. Combined with posterior prediction accuracies for small increases in the highest and lowest prices significantly exceeding 0.5, this indicates that subsequent event 2 demonstrates a more pronounced overall upward trend in both volume and price. Similarly, the attribute states predicted in subsequent events 3 and 4 have different emphases. In event 3, the posterior prediction accuracies for various attribute states are generally low, making the upward trend prediction unclear. However, in event 4, the posterior prediction accuracy for a small increase in the lowest price reaches 0.7, indicating a clear upward trend.

These predictions of immediate subsequent events for target sequences, which systematically base posterior predictions on prior validations, enable traders to develop

more precise and effective trading strategies. For Yanghe, our experimental subject, once a target event is identified as partially or fully occurring—and considering the high probability of future upward trends—traders can immediately initiate position establishment through stock acquisition. If subsequent event 1 occurs as predicted, additional position accumulation is warranted. Similarly, if subsequent event 2 materializes according to predictions, continued position building remains advisable. Conversely, if subsequent event 2 deviates from predictions, immediate position reduction becomes prudent to secure partial profits. This methodical process continues through subsequent event 4, at which point complete position liquidation should be considered. Through rigorous similarity matching based on the initial four event sets and their prior occurrences, combined with sequential posterior predictions for the subsequent four event sets, traders can implement flexible, incremental position management. This approach significantly mitigates risks associated with non-algorithmic trading decisions and facilitates scientifically sound profit capture from price differentials.

4.3 Contrast Experiment

Dynamic Time Warping (DTW), as the preferred distance measurement method in time series data mining, has been proven to possess the most extensive applicability and highest accuracy in sequence similarity matching tasks. Its multidimensional extension—Multivariate Dynamic Time Warping (MDTW)—has likewise gained widespread application[5]. This study designed rigorous comparative experiments, simultaneously applying target sequences and test data to both the standard MDTW algorithm and our proposed FS-FTSP algorithm. The input data expanded preceding and subsequent trading events based on selected 5-sequence multidimensional temporal series, forming five target sequences of different lengths: 2010/6/9 10:00-10:10 (3 sequences); 2010/6/9 10:00-10:20 (5 sequences); 2010/6/9 10:00-10:30 (7 sequences); 2010/6/9 10:00-10:45 (10 sequences); and 2010/6/9 10:00-11:05 (14 sequences).

As multivariate time series length increased, the FS-FTSP algorithm proposed in this paper demonstrated significant efficiency advantages. As illustrated in the Fig. 4, FS-FTSP execution time exhibited a gradual declining trend, decreasing from 58.81 seconds for 3 sequences to 39.33 seconds for 14 sequences, indicating that computational efficiency actually improved with increasing sequence length. In stark contrast, traditional MDTW algorithm execution time rose dramatically, surging from 67.74 seconds for 3 sequences to 224.82 seconds for 14 sequences. FS-FTSP substantially outperformed MDTW across all sequence lengths; even in the worst-case scenario, its runtime remained lower than MDTW's best performance. Computational complexity was virtually unaffected by increasing sequence length, and efficiency actually improved when processing larger-scale sequences.

Regarding mining results, as shown in the Fig. 5, our algorithm demonstrates exceptional advantages in multivariate temporal similarity recognition. As temporal sequence length increases, FS-FTSP consistently maintains powerful pattern discovery capabilities: with 3-sequence, 5-sequence, and 7-sequence time series, FS-FTSP

identifies patterns at volumes exceeding MDTW by more than 7-fold to several hundred-fold, showing clear superiority. Particularly noteworthy is that when sequence length reaches 10 and 14, FS-FTSP can still identify 300 and 752 meaningful patterns respectively, while the MDTW algorithm almost completely fails in these higher-complexity scenarios, identifying only a single pattern or none at all. This comparison thoroughly demonstrates FS-FTSP's excellent adaptability and stability when processing long sequences and complex patterns.

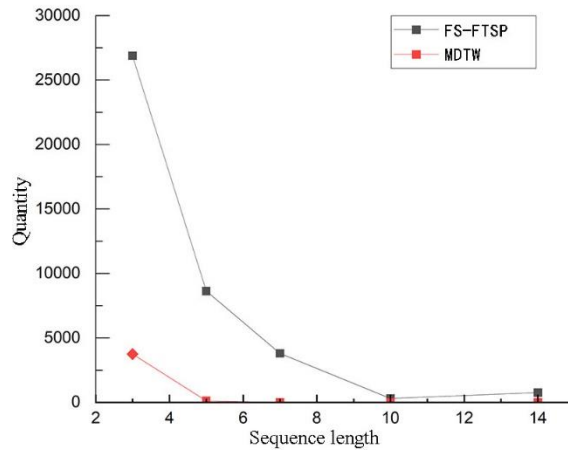


Fig. 4. Comparison of Similarity Matching Mining Quantity between Two Algorithms

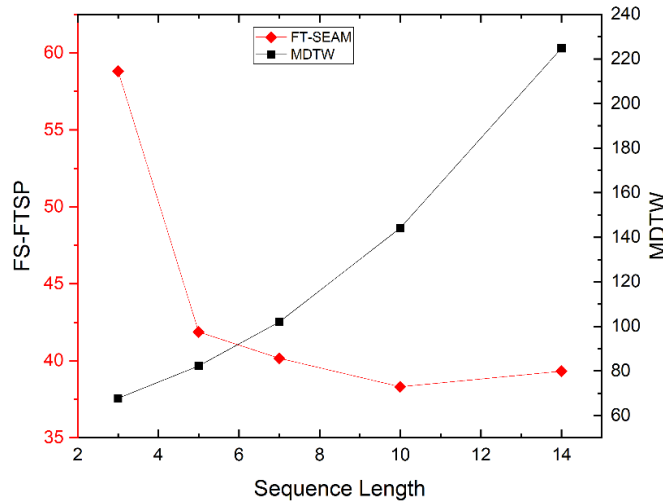


Fig. 5. Comparison of Similarity Matching Mining between Two Algorithms

Despite the promising results, this study exhibits several methodological limitations. First, the interdependencies and correlations among different stocks remain insufficiently incorporated into the analytical framework. Second, the current model architecture necessitates manual definition of fuzzy states and membership functions, thereby lacking autonomous learning capabilities essential for adaptive processing. Third, the absence of causal relationship modeling between variables potentially generates spurious pattern identification, compromising analytical validity.

5 Conclusions

This investigation presents a securities trading prediction framework (FS-FTSP) predicated on feature subsequence extraction and fuzzy temporal sequence similarity recognition, addressing the high-dimensionality, non-linearity, and non-stationarity characteristics inherent in securities market data. The framework systematically integrates a Self-supervised Unsupervised Model for common features (SUMS), a fuzzy temporal sequence similarity recognition algorithm, and a clustering-based multi-step prediction model, providing a novel solution for securities trading prediction.

Regarding theoretical contributions, this investigation achieves innovative breakthroughs along three critical dimensions. First, we have constructed a self-supervised special event subsequence extraction mechanism based on contrastive learning that achieves multi-level characterization of complex trading patterns through a separation strategy for common and unique features. Second, we have designed a temporal sequence similarity recognition algorithm integrating fuzzy set theory that effectively addresses the inherent uncertainty and fuzziness problems in financial time series. Third, we have developed a clustering prediction model based on historical similarity that supports recursive prediction from single-step to multi-step, enhancing both the accuracy and practicality of predictions.

Empirical analyses demonstrate that our methodology achieves exceptional performance in predicting both mid-to-high-end liquor sector stock groups and individual Yanghe stock. Compared with traditional MDTW algorithms, FS-FTSP exhibits significant advantages in computational efficiency, with execution time decreasing rather than increasing as sequence length grows. Regarding pattern recognition capability, our method maintains robust pattern discovery ability even when processing extended sequences, whereas MDTW algorithms become virtually ineffective in high-complexity scenarios. The investigation further reveals that prediction accuracy is intimately associated with temporal granularity selection, with 5-min temporal type exhibiting optimal performance in our research scenario, providing critical guidance for parameter selection in practical applications.

The theoretical significance of this investigation lies in advancing the integration of multivariate time series analysis techniques with fuzzy theory, providing a novel methodological framework for addressing complexity in financial time series. Its practical value is manifested in delivering interpretable, highly efficient prediction tools for quantitative investment decisions, contributing to enhanced scientific rigor in investment decision-making. Future investigations may further expand by incorporating causal inference mechanisms, developing adaptive fuzzy state definition methods, and considering cross-market interactions, thereby constructing a more comprehensive financial time series prediction system.

References

- [1] Grabocka, J.; Schilling, N.; Wistuba, M.; Schmidt-Thieme, L. Learning Time-Series Shapelets. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining; 2014; pp 392–401.
- [2] Li, G.; Choi, B.; Xu, J.; Bhowmick, S. S.; Chun, K.-P.; Wong, G. L.-H. Efficient Shapelet Discovery for Time Series Classification. *IEEE transactions on knowledge and data engineering* 2020, 34 (3), 1149–1163.
- [3] Hills, J.; Lines, J.; Baranauskas, E.; Mapp, J.; Bagnall, A. Classification of Time Series by Shapelet Transformation. *Data mining and knowledge discovery* 2014, 28, 851–881.
- [4] Yadav, M.; Alam, M. A. Dynamic Time Warping (Dtw) Algorithm in Speech: A Review. *International Journal of Research in Electronics and Computer Engineering* 2018, 6 (1), 524–528.
- [5] Moser, U.; Schramm, D. Multivariate Dynamic Time Warping in Automotive Applications: A Review. *Intelligent Data Analysis* 2019, 23 (3), 535–553.
- [6] Sun, B.; Guo, H.; Karimi, H. R.; Ge, Y.; Xiong, S. Prediction of Stock Index Futures Prices Based on Fuzzy Sets and Multivariate Fuzzy Time Series. *Neurocomputing* 2015, 151, 1528–1536.
- [7] Behbood, V.; Lu, J.; Zhang, G. Fuzzy Refinement Domain Adaptation for Long Term Prediction in Banking Ecosystem. *IEEE Transactions on Industrial Informatics* 2013, 10 (2), 1637–1646.
- [8] Li, J.; Ren, W.; Han, M. Mutual Information Variational Autoencoders and Its Application to Feature Extraction of Multivariate Time Series. *International Journal of Pattern Recognition and Artificial Intelligence* 2022, 36 (06), 2255005.
- [9] Bianchi, F. M.; Livi, L.; Mikalsen, K. Ø.; Kampffmeyer, M.; Jenssen, R. Learning Representations of Multivariate Time Series with Missing Data. *Pattern Recognition* 2019, 96, 106973.
- [10] Liu, F.; Zhang, L.; Yang, J.; Wu, W. Adaptive Fuzzy Population Coding Method for Spiking Neural Networks. *International Journal of Fuzzy Systems* 2023, 25 (2), 670–683.
- [11] Yoon, H.; Yang, K.; Shahabi, C. Feature Subset Selection and Feature Ranking for Multivariate Time Series. *IEEE transactions on knowledge and data engineering* 2005, 17 (9), 1186–1198.
- [12] Huang, W.; Yue, B.; Chi, Q.; Liang, J. Integrating Data-Driven Segmentation, Local Feature Extraction and Fisher Kernel Encoding to Improve Time Series Classification. *Neural processing letters* 2019, 49, 43–66.
- [13] Ye, L.; Keogh, E. Time Series Shapelets: A Novel Technique That Allows Accurate, Interpretable and Fast Classification. *Data mining and knowledge discovery* 2011, 22, 149–182.
- [14] Liang, Z.; Zhang, J.; Liang, C.; Wang, H.; Liang, Z.; Pan, L. A Shapelet-Based Framework for Unsupervised Multivariate Time Series Representation Learning. *arXiv preprint arXiv:2305.18888* 2023.

-
- [15] Cai, B.; Huang, G.; Yang, S.; Xiang, Y.; Chi, C.-H. Se-Shapelets: Semi-Supervised Clustering of Time Series Using Representative Shapelets. *Expert Systems with Applications* 2024, 240, 122584.
- [16] Rak, R.; Drozd, S.; Kwapien, J.; Oswiecimka, P. Detecting Subtle Effects of Persistence in the Stock Market Dynamics. *arXiv preprint physics/0504158* 2005.
- [17] Oświęcimka, P.; Drożdż, S.; Kwapien, J.; Górski, A. Fractals, Log-Periodicity and Financial Crashes. *Acta Physica Polonica A* 2010, 117 (4), 637–639.
- [18] Berndt, D. J.; Clifford, J. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the 3rd international conference on knowledge discovery and data mining; 1994*; pp 359–370.
- [19] Li, S.; Wu, J.; Jiang, X.; Xu, K. Chart GCN: Learning Chart Information with a Graph Convolutional Network for Stock Movement Prediction. *Knowledge-based systems* 2022, 248, 108842.
- [20] Dorr, D. H.; Denton, A. M. Establishing Relationships among Patterns in Stock Market Data. *Data & Knowledge Engineering* 2009, 68 (3), 318–337.
- [21] Bitencourt, H. V.; Orang, O.; Silva, P. C.; Guimarães, F. G.; others. A Multi-Step Multivariate Fuzzy-Based Time Series Forecasting on Internet of Things Data. *IEEE Internet of Things Journal* 2025.
- [22] Papageorgiou, K. I.; Poczeta, K.; Papageorgiou, E.; Gerogiannis, V. C.; Stamoulis, G. Exploring an Ensemble of Methods That Combines Fuzzy Cognitive Maps and Neural Networks in Solving the Time Series Prediction Problem of Gas Consumption in Greece. *Algorithms* 2019, 12 (11), 235.
- [23] Zhu, A.; Meng, Z.; Shen, R. Research on Fuzzy Temporal Event Association Mining Model and Algorithm. *Axioms* 2023, 12 (2), 117.