Applied Intelligence
https://doi.org/10.1007/s10489-021-02979-y

A multi-view time series model for share turnover prediction

Zhijin Wang[1] · Qiankun Su[1] · Guoqing Chao[2] · Bing Cai[1] · Yaohui Huang[1] · Yonggang Fu[1]

Accepted: 3 November 2021

Abstract

Share turnover is a key indicator for investing in the stock market, which represents how easy or difficult it is to trade a stock. Several techniques have been proposed to predict share turnover values. However, they are often inaccurate because they utilize single-view models that have an incomplete picture of the temporal dynamics. To address this issue, a multi-view time series model (MvT) is proposed to capture temporal dynamics using three views on two data groups. The temporal dynamics of target turnover data and exogenous turnover data are captured by a view generation component. The component generates three views in three different aspects. The predictions are then made by a view combination component and a full connected layer. Extensive experiments on two real stock datasets show the effectiveness and efficiency of the proposed MvT model, when compared with ten algorithms on four groups of stock data in terms of three metrics.

Keywords Share turnover · Prediction · Multi-view · Time series · Temporal dynamics

1 Introduction

Numerous stock investors attempt to seeking profit-maximization strategies to get promising gains. The share turnover, as an important quantitative index, catches much attention on supporting the decision-making for investors. The share turnover sums up the transactions of a stock within a trading day and reflects the stock activity degree on that day. Specifically, the high turnover value of a stock represents the recent high-frequency trading pattern, which leads to the high volatility of the stock price, and vice versa [31].

Despite the Efficient Market Hypothesis (EMH) indicates that the security exchange market is sufficiently efficient and impossible to predict stock movements, many time series methods have been applied to forecasting share turnover values and achieved good performance, such as mathematical methods [30] and deep learning methods [13]. It demonstrates that the market is less efficient than expected, and it is possible to make further predictions [7]. However, these single-view based methods suffer from the incomplete obtainment of temporal dynamics. According to phenomenons of the momentum effect and the reversal effect in stock market data [7], the upcoming values are potentially influenced by both the long and short-term trading patterns. Neglecting the local or global influence of share turnover

This article belongs to the Topical Collection: *Special Issue on Multi-view Learning*
Guest Editors: Guoqing Chao, Xingquan Zhu, Weiping Ding, Jinbo Bi and Shiliang Sun

✉ Qiankun Su
qiankun.su@jmu.edu.cn

Zhijin Wang
zhijinecnu@gmail.com

Guoqing Chao
guoqingchao@hit.edu.cn

Bing Cai
ccailearning@gmail.com

Yaohui Huang
yhhuang5212@gmail.com

Yonggang Fu
yonggangfu@jmu.edu.cn

1 Computer Engineering College, Jimei University, Yinjiang Road 185, Xiamen 361021, China

2 School of Computer Science and Technology, Harbin Institute of Technology, 2 West Culture Road, Weihai, 264209, People's Republic of China

observations will inevitably undermine the performance of prediction methods.

Moreover, trading patterns are also sensitively affected by many uncertain political-economic factors in the real world, such as government policies, corporate earnings performances, and related breaking news of enterprises. For these reasons, numerous methods assigned an independent view to extract the potential representation of exogenous data (such as search query data and other stock-related data), and then aggregate the target observations to generate predictions. These methods include but are not limited to PM [23], DGR [22] and DSAR [21]. Nevertheless, the above-mentioned multi-view methods cannot completely obtain temporal dynamics as well [29].

To address these issues, a multi-view time series model (MvT) is proposed to comprehensively capture the temporal dynamics in terms of target time series and exogenous time series. The MvT consists of a view generation component and a view combination component. The temporal dynamics of target turnover data and exogenous turnover data (i.e., Shanghai Connect) are captured by the view generation component, which generates three views to independently establish the association between each time point, period, and input variable. The predictions are made by the view combination component, which combines outputs from several views.

Traditional multi-view methods depend on the consensus and complementarity information among multiple distinct features to enhance the generalization performance [25–28]. However, for time series prediction problem, limited by the ability of time series data to generate other relevant features, the generated views should focus on exploring the inherent relationships of the sequences. Therefore, we proposed a novel view generation component to highlight the non-linear association of different aspects. Subsequently, the view combination component is utilized to fusion potential connections from each view and generate predictions.

In summary, the major contributions of this work are summarized as follows:

(1) Verified the superiority of integrating the Shanghai Connect turnover to improve the predictive performance in some other share turnovers.
(2) The proposed effective view generation component benefits well from exogenous turnover data.
(3) The MvT achieves better performance, not only on the target turnover time series, but also on the exogenous turnover time series.

The rest of this paper are organized as follows. Section 2 addressed this work with respect to multi-view learning. Section 3 formulates the multi-view time series prediction problem. Section 4 presents the proposed MvT. The experimental configurations for performance evaluation is given in Sections 5, and 6 shows the benefit of our scheme. Finally, Section 7 concludes this work.

2 Related work

This section recalls time series prediction methods to address this research, in terms of single-view algorithms and multi-source algorithms.

2.1 Single-view

Methods in this category develop a view to connect inputs and outputs. They are divided into statistical methods and deep learning methods.

Statistical methods apply statistical analysis techniques to exploit the different stochastic processes using a single-view. Autoregression (AR) and moving average (MA) are well-known linear statistical methods, autoregressive moving average (ARMA) and Auto-regressive integrated moving averages (ARIMA) are their variants. These methods have been introduced to predict the forex price [12], cryptocurrency price [8] and stock price [33]. However, these methods are suitable for stable time series. Difference processing is commonly utilized to correct unstable sequences, but the processing cannot cover overall series changes. Moreover, the non-linear patterns of observations are unable to be captured.

Learning methods are purposed to ameliorate the extraction of complicated non-linear relationships from time series. Vanilla Recurrent Neural Network (RNN) is one of the deep learning methods, which has the ability to produce temporal dynamics views and draw great attention in related domains [11]. CNNRNN [6], as its variant, using convolutional structures to extract local views and then feeding into the RNN component, which effectively alleviates the sequence length and highlights the key periods. However, the RNN structure is limited to consider long-term dependency. Hence, Long- and Short-Term Memory (LSTM) and Gated Recursive Network (GRU) are purposed to overcome the shortage [22], which has been applied for stock price forecasting [9], the stock market index [6] and financial volatility [5]. Recently, on the basis of LSTM, Encoder-Decoder (ED) architecture [4] remaps the inputs into further non-linear representation and has already been applied in predicting stock prices [10]. PM [23], MSL [20], attention mechanism [14] and transformer are also used to generate single-view. However, the local and global dynamical characteristics are ubiquitous of time series, using the single-view limits the correlation horizon of generations [32].

2.2 Multi-source

Methods in this category develop each view for a data source [15, 16]. Different sources may share a common view.

To be specific, DGR [22] captures the temporal dynamics from different time granularity sources and then consolidates dual granularity views to generate predictions. CARD [24] using decomposition techniques to obtain different relevant time series, and combine the two views from target and relevant time series to improve the performance accuracy. Similarly, DSAR [21] products the short-term patterns from multi-source data, and patterns are summarized in the combination stage.

Although these methods show the benefit for many realistic forecasting problems [1–3, 29], they were hard to achieve a precise prediction in more complex environments due to the various dynamical patterns of time series [18, 19].

Therefore, we integrated multiple direction views of each input time series and generated the predictions by learning a comprehensive observation. The observation effectively extracts the complex local correlation, global correlation and sequence correlation between inputs and prediction values. Besides, each view has an ability on different sources. Meanwhile, the linear representations of each input also are considered as a useful view to enhancing the robustness.

3 Problem definition

This section gives notations and problem definitions.

3.1 Notation

In this section, some definitions of time series prediction are given. For convenience, a list of symbols and their definitions are summarized in Table 1.

Time series Let symbol $\boldsymbol{Z} \in \mathbb{R}^{P\times 1}$ be the share turnover time series. Let symbol $\boldsymbol{E} \in \mathbb{R}^{P\times K}$ denote the Shanghai Connect turnover time series. P is the time steps of those time series. K is the factor number of Shanghai Connect.

Look-back window Let symbol T be the look-back window size. A look-back window observes the past turnover values in T time steps. For instance, $\boldsymbol{Z}_{1:T,1}$ denotes the observed values of $\boldsymbol{Z}$ within $[1, T]$ time steps.

Share turnover prediction problem Commonly, the prediction based on the past observed values is formulated as:

$$\hat{Z}_{T+1,1} \leftarrow F(\boldsymbol{Z}_{1:T,1}), \tag{1}$$

Table 1 Symbols and semantics

Symbol	Semantic
P	The time steps of a time series
T	Look-back window size
K	The number of exogenous factors
V	The number of views
δ	Bias
B, b	Batch size
$\boldsymbol{Z}$	The share turnover time series $\boldsymbol{Z} \in \mathbb{R}^{M\times 1}$
$Z_{m,1}$	The entry of $\boldsymbol{Z}$
$\boldsymbol{Z}_{1:T,1}$	The look-back window of $\boldsymbol{Z}$, $\boldsymbol{Z}_{1:T,1} \in \mathbb{R}^{T\times 1}$
$\boldsymbol{Z}^v_{1:T,1}$	The output of v-th view on $\boldsymbol{Z}_{1:T,1}$, $\boldsymbol{Z}^v_{1:T,1} \in \mathbb{R}^{T\times 1}$
$\boldsymbol{E}$	The Shanghai Connect time series, $\boldsymbol{E} \in \mathbb{R}^{M\times K}$
$E_{m,k}$	The entry of $\boldsymbol{E}$
$\boldsymbol{E}_{1:T,:}$	The look-back window of $\boldsymbol{E}$, $\boldsymbol{E}_{1:T,:} \in \mathbb{R}^{T\times K}$
$\boldsymbol{E}^v_{1:T,:}$	The output of v-th view on $\boldsymbol{E}_{1:T,:}$, $\boldsymbol{E}^v_{1:T,:} \in \mathbb{R}^{T\times K}$
$Z_{T+1,1}$	The real value in the upcoming day
$\hat{Z}_{T+1,1}$	The predicted value in the upcoming day
$F(\cdot)$	A mapping
$G_v(\cdot)$	The mapping of v-th view
$[;]$	Concatenation operation

where $\hat{Z}_{T+1,1} \in \mathbb{R}^{T\times 1}$ denotes the predicted value in the upcoming day. T is the look-back window size. $F(\cdot)$ is a mapping.

The prediction problem using exogenous data Share turnover trading patterns are sensitive to many uncertain political-economic factors, fusion exogenous data have the ability to reduce the effect of fluctuations.

The prediction based on both target and exogenous observed values is formulated as:

$$\hat{Z}_{T+1,1} \leftarrow F(\boldsymbol{Z}_{1:T,1}, \boldsymbol{E}_{1:T,:}), \tag{2}$$

where $\boldsymbol{E}_{1:T,:} \in \mathbb{R}^{T\times K}$ denotes the observations of Shanghai Connect turnover. T is the look-back window size.

3.2 Definition

In this section, the prediction problem via multi-view modeling is defined.

The prediction based on the past observed values using multiple views is formulated as:

$$\boldsymbol{Z}^v_{1:T,1} \coloneqq G_v(\boldsymbol{Z}_{1:T,1}), \tag{3}$$

$$\boldsymbol{E}^v_{1:T,:} \coloneqq G_v(\boldsymbol{E}_{1:T,:}), \tag{4}$$

where $G_v(\cdot)$ is the mapping of v-th view, a.k.a., view construction. $\boldsymbol{Z}^v \in \mathbb{R}^{B\times T\times 1}$ and $\boldsymbol{E}^v \in \mathbb{R}^{B\times T\times K}$ indicate the

v-th generated view of target observations and exogenous data respectively.

$$\hat{Z}_{T+1,1} := F(\{\boldsymbol{Z}^{v}_{1:T,1}\}, \{\boldsymbol{E}^{v}_{1:T,:}\}), \quad s.t., v \in [1, V], \tag{5}$$

where $F(\cdot)$ maps outputs of several view to the upcoming value, a.k.a., view combination, and $\hat{Z}_{T+1,1}$ is the prediction.

4 The Proposed MvT

This section illustrates the proposed multi-view time series model (MvT). The graphical scheme of MvT is drawn in Fig. 1.

Firstly, those time series are normalized, split and aggregated to generate trainable supervised data. Secondly, the target input data and exogenous input data are fed into three views to observe historical events from different viewpoints. Thirdly, the outputs from different views and input data are combined to generate the predictive values. Finally, all the predictive values are linearly sum up and de-normalized as prediction values.

4.1 Time series transformation

Min-Max normalization. To speed up the fitting process of the model, the Min-Max normalization is employed

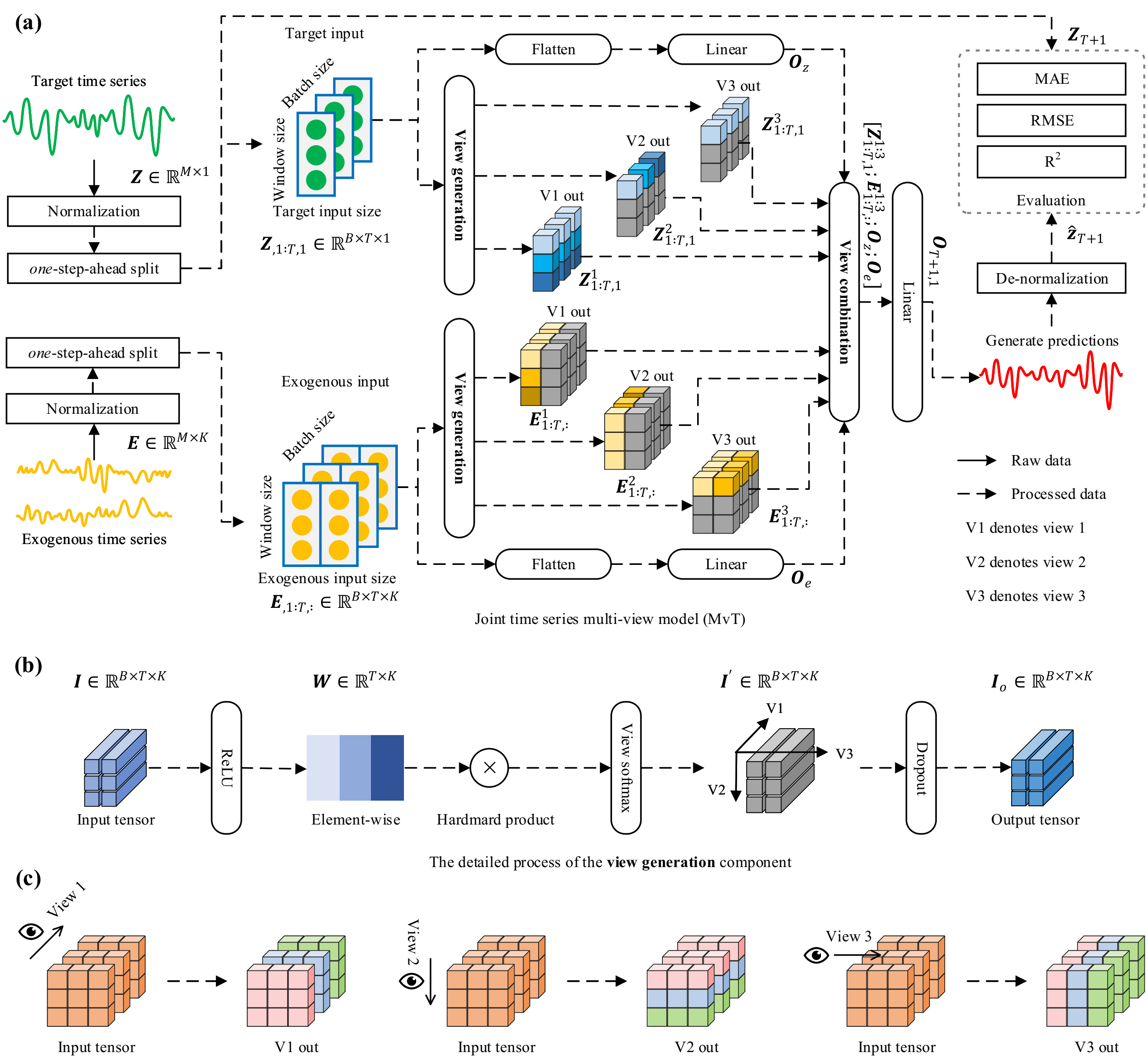


Fig. 1 The graphical diagram of the proposed multi-view time series model (MvT)

to compress the data into range [0, 1]. The normalization formula and its de-normalization formula are as follows:

$$\boldsymbol{d}' = \frac{\boldsymbol{d} - \min(\boldsymbol{d})}{\max(\boldsymbol{d}) - \min(\boldsymbol{d})}, \tag{6}$$

$$\boldsymbol{d} = \boldsymbol{d}' * (\max(\boldsymbol{d}) - \min(\boldsymbol{d})) + \min(\boldsymbol{d}), \tag{7}$$

where $\boldsymbol{d} \in \mathbb{R}^M$ represents a vector of all the observed samples, $\boldsymbol{d}' \in \mathbb{R}^M$ is the normalized data, $\max(\boldsymbol{d})$ is the maximum value of $\boldsymbol{d}$, and $\min(\boldsymbol{d})$ is the minimum value of $\boldsymbol{d}$. The de-normalization formula is applied for the outputs of models in the post-processing stage.

One-step-ahead split Due to the time series data can not be directly fed into a supervised model [22]. The one-step-ahead split technique is applied to these time series to generate supervised data.

Given the target time series $\boldsymbol{Z} \in \mathbb{R}^{M\times 1}$ with M consecutive time intervals, the exogenous time series $\boldsymbol{E} \in \mathbb{R}^{M\times K}$ with M consecutive time intervals and K factors, the one-step-ahead split is applied as below:

$$\begin{bmatrix} Z_{1,1} & \cdots & Z_{T,1} & \boldsymbol{E}_{1,:} & \cdots & \boldsymbol{E}_{T,:} \\ Z_{2,1} & \cdots & Z_{T+1,1} & \boldsymbol{E}_{2,:} & \cdots & \boldsymbol{E}_{T+1,:} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ Z_{M-T-1,1} & \cdots & Z_{M-1,1} & \boldsymbol{E}_{M-T-1,:} & \cdots & \boldsymbol{E}_{M-1,:} \end{bmatrix} \rightarrow \begin{bmatrix} Z_{T+1,1} \\ Z_{T+2,1} \\ \cdots \\ Z_{M,1} \end{bmatrix}, \tag{8}$$

where the left part is the inputs of a model, a.k.a, look-back window, the upper right part is the output of a model. For a lucid presentation, let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{(M-T-1)\times T\times(K+1)}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(M-T-1)\times 1}$ denote inputs and outputs, respectively. Several consecutive samples in $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}})$ are denoted by $(\boldsymbol{X} \in \mathbb{R}^{B\times T\times 1}, \boldsymbol{Y} \in \mathbb{R}^{B\times 1})$. $\boldsymbol{Z}_{T+1:T+B,1}$ is denoted by $\boldsymbol{Y}$, and $[\boldsymbol{Z}_{1:T,1}, \boldsymbol{Z}_{2:T+1,1}, \cdots, \boldsymbol{Z}_{B:T+B-1,1}]$ is denoted by $\boldsymbol{X}$.

4.2 View generation

As shown in Fig. 1a, the generated views $\boldsymbol{O}_z$ and $\boldsymbol{O}_e$ are introduced to further improve the robustness of MvT. These views directly extract the linear characteristics of the target and exogenous time series respectively, and then combine them in parallel to improve the accuracy of prediction. The processing can be formulated as follow:

$$\boldsymbol{O}_z = \sum_{t\in T} \boldsymbol{W}_z * \boldsymbol{Z} + \boldsymbol{\delta}_z, \tag{9}$$

$$\boldsymbol{O}_e = \sum_{t\in T} \boldsymbol{W}_e * \boldsymbol{E} + \boldsymbol{\delta}_e, \tag{10}$$

where $\boldsymbol{O}_z \in \mathbb{R}^{B\times T\times 1}$ is the linear view generated by observations, $\boldsymbol{O}_e \in \mathbb{R}^{B\times T\times 1}$ is the linear characteristics of exogenous data, $\boldsymbol{W}_z \in \mathbb{R}^{B\times T\times 1}$ and $\boldsymbol{W}_e \in \mathbb{R}^{B\times T\times 1}$ are the weight tensor of $\boldsymbol{Z}$ and $\boldsymbol{E}$ respectively, $\boldsymbol{\delta}_z \in \mathbb{R}^{B\times 1}$ and $\boldsymbol{\delta}_e \in \mathbb{R}^{B\times 1}$ are bias term.

Besides, the MvT designs a view generation component for both target time series and exogenous time series, as depicted in Fig. 1b. The view generation component using element-wise transformation makes a directional highlight for inputs. To alleviate overfitting and enhance the non-linearity of the network, the ReLU activation is integrated into a view unit as shown in Fig. 1c. The product view computation can be mathematically expressed as:

$$\boldsymbol{M} = \boldsymbol{W}_i * \mathrm{ReLU}(\boldsymbol{I}), \tag{11}$$

where $\boldsymbol{M}$ is element-wise weighted input matrix, $\boldsymbol{I} \in \mathbb{R}^{B\times T\times K}$ is the input tensor, $\mathrm{ReLU}(\cdot)$ is the rectified linear unit function, $\boldsymbol{W}_i \in \mathbb{R}^{B\times T\times K}$ is the weight tensor of $\boldsymbol{I}$, and * is the Hadamard product of $\boldsymbol{I}$ and $\boldsymbol{W}_i$.

To highlight the directional view generation in different aspects, the function of each view is described as follows:

View 1 observes each batch of inputs through the network. The batch size defines a period of consecutive training examples utilized in one iteration. The relationships between consecutive historical observations and the upcoming period are enhanced via highlight the critical period, the processing can be summarized as follows:

$$\boldsymbol{\mathcal{R}}^1{}_{b,t,k} = \frac{\exp(M_{b,:,:})}{\sum_b \exp(M_{b,:,:})}, \tag{12}$$

where $\boldsymbol{\mathcal{R}}^1 \in \mathbb{R}^{B\times T\times K}$ is generated by the transformation of view 1, $\boldsymbol{M} \in \mathbb{R}^{B\times T\times K}$ is element-wise weighted input matrix.

View 2 observes the relevant time steps from previous days of the look-back window to the next day. Those critical time steps are enhanced via the correlation to the output time step, the processing can be summarized as follows:

$$\boldsymbol{\mathcal{R}}^2{}_{b,t,k} = \frac{\exp(M_{:,t,:})}{\sum_t \exp(M_{:,t,:})}, \tag{13}$$

where $\boldsymbol{\mathcal{R}}^2 \in \mathbb{R}^{B\times T\times K}$ is generated by the transformation of view 2.

View 3 observes the relationships between input time series and target time series. Those critical time series are enhanced according to the contribution for output, the processing can be summarized as follows:

$$\boldsymbol{\mathcal{R}}^3{}_{b,t,k} = \frac{\exp(M_{:,:,k})}{\sum_k \exp(M_{:,:,k})}, \tag{14}$$

where $\boldsymbol{\mathcal{R}}^3 \in \mathbb{R}^{B\times T\times K}$ is generated by the transformation of view 3.

The generated views $\boldsymbol{Z}^{1:3}_{1:T,1}$ and $\boldsymbol{E}^{1:3}_{1:T,:}$ are produced by the above processing.

4.3 View combination and prediction

Finally, the generated views are combined through *View combination*, then, MvT predicts the upcoming values by a linear layer. The generate views from the above components are summarized as below:

$$\mathcal{G} := [\boldsymbol{Z}^{1:3}_{1:T,1}, \boldsymbol{E}^{1:3}_{1:T,:}, \boldsymbol{O}_z, \boldsymbol{O}_e], \tag{15}$$

where $\mathcal{G} \in \mathbb{R}^{B\times T\times(1+2+K)}$ is the concatenated outputs.

$$\boldsymbol{O}_{T+1,1} = \sum_{t\in T} \boldsymbol{W}_o * \mathcal{G} + \boldsymbol{\delta}_o, \tag{16}$$

where $\boldsymbol{O} \in \mathbb{R}^{B\times 1\times 1}$ is the output representation of the inputs $\mathcal{G}$, $\boldsymbol{W}_o$ is the weight tensor, and $\boldsymbol{\delta}_o \in \mathbb{R}^{B\times 1}$ is a bias term. After that, the predicted values $\boldsymbol{Y}$ are calculated by de-normalizing the $\boldsymbol{O}$ by Equation 7.

5 Experimental setup

This section gives data collection, performance metrics, benchmark methods and configurations.

5.1 Datasets

We examine the benefit of northbound trading to the prediction of share turnover. There are three stock exchanges in China, two of them (SSE, SESE) in mainland China and one in Hongkong. Shanghai-Hong Kong Stock Connect Program makes the Mainland China and Hong Kong markets are able to direct access to each other's stock market. Northbound trading refers to the trading of mainland-listed stocks from the Hong Kong Stock Exchange (HKEX)[1]. More specifically, the northbound trading is subdivided into Shanghai Connect and Shenzhen Connect, targeted for SSE and SESE respectively. There are released on November 17, 2014, and December 5, 2016 respectively.

By the day (January 31, 2021) we carry out the experiment, there are 1468 and 983 trading days for Shanghai Connect and Shenzhen Connect respectively. To have more training data, we select stocks from SSE. There are 568 stocks available on Shanghai Connect at the initial stage.

We select the two largest total market values at present among them, e.g., Ping An Insurance (Group) Company of China, Limited (stock code: 000001, shortened to *PingAn* in the following sections for the sake of conciseness) and Kweichow Moutai Company Limited (stock code: 600519, shorted to *Moutai*).

Considering the case of trading suspensions, we simply remove the records from Shanghai Connect during this period to keep the data aligned to the turnover of stocks. After this data preprocessing, both PingAn and Moutai coincidently have the same number of data points, 1442. Table 2 shows some basic statistics of the turnover of PingAn (denoted by S_1), Moutai (denoted by S_2), and Shanghai Connect (denoted by C) where SD is the abbreviation of standard deviation, and PCC denotes the Pearson correlation coefficient between S_1, S_2 and C. The value of Shanghai Connect on a certain day is the addition of the total value bought and sold on the given day. Clearly, there is a significant correlation between Shanghai Connect and PingAn, Moutai.

To present the time series of PingAn, Moutai, and Shanghai Connect in a more intuitive way, we plot the daily turnover of them during the period from November 17, 2014 to January 29, 2021, as shown in Fig. 2.

5.2 Performance metrics

The commonly used metrics are applied to measure the prediction performance. They are formulated below.

(1) Mean Absolute Error (MAE):

$$MAE = \frac{1}{|\Omega_{Test}|} \sum_{i\in\Omega_{Test}} |Z_{i,1} - \hat{Z}_{i,1}|, \tag{17}$$

(2) Mean Absolute Percentage Error (MAE):

$$RMSE = \sqrt{\frac{1}{|\Omega_{Test}|} \sum_{i\in\Omega_{Test}} (Z_{i,1} - \hat{Z}_{i,1})^2}, \tag{18}$$

(3) Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i\in\Omega_{Test}} (Z_{i,1} - \hat{Z}_{i,1})^2}{\sum_{i\in\Omega_{Test}} {Z_{i,1}}^2}. \tag{19}$$

where Ω_{Test} is the testing set.

For MAE and $RMSE$, the lower values have the better performance. For R^2, the higher values have the better performance.

5.3 Benchmarks

To evaluate our model objectively, the MvT is compared with the methods listed below.

(1) *Multiple linear regression* (MLR) describes how a response variable depends linearly on two or more explanatory variables.
(2) *Long short-term memory* (LSTM) is designed to address the problem of vanishing gradients and exploding gradients, and has a strong ability on sequential data.
(3) *Gated Recurrent Unit* (GRU) [4] is a variation from the LSTM, and has fewer parameters than LSTM. The memory ability is close to LSTM.

[1] Southbound trading refers to the trading of certain SEHK securities from mainland investors. It is likely named so because Hong Kong is located south of Shanghai and Shenzhen.

Table 2 The statistics of the turnover of PingAn, Moutai, and Shanghai Connect (Unit: 100M CNY)

Symbol	Min	Max	Medium	Mean	SD	PCC_C
S_1	4.35	679.41	40.53	53.69	47.55	0.2246***
S_2	1.68	169.01	21.54	27.31	22.35	0.8084***
C	6.89	921.64	90.72	149.89	149.05	—

*** indicates $p - value < 0.001$

(4) Encoder-Decoder (ED) [4] is a sequence-to-sequence framework, and is proposed for varying size of the inputs and outputs.
(5) CNNRNN [17] is a unified framework for multi-label image classification. It takes into consideration the dependencies among multi-labels.
(6) Dual-side autoregression (DSAR) [21] represents the search engine data and historical turnover values by considering dual sides autoregression results.

5.4 Configurations

We split the dataset into train and test data, the first 80% of data as training data and the remaining 20% as test data. All of the predictive methods are trained by the Adam optimizer, and the mean squared error (MSE) is selected as the loss function. For ED, CNNRNN, LSTM and GRU, the number of hidden neurons is set to {32, 64}. We adjust the window size to optimal for all methods from [1, 20], and the batch size is set to 32. The training epoch and learning rate are tuned to the optimal states of each method.

6 Predictive results and analyses

This section evaluates the proposed MvT performance in terms of MAE, $RMSE$ and R^2. These experiments aim to study the following issues:

(1) How parameter T affects the prediction performance?
(2) How does the combination of views affect prediction performance?
(3) Can MvT outperform the state-of-the-art methods?

6.1 Parameter *T*

In this section, we investigate the effects on look-back window size T on MvT. To eliminate the intervention of different group views, we set the parameter V to $\{1, 2, 3\}$, i.e., using all three views.

Figure 3a, b and c shows the evolution of our model accuracy over a series of input window size T from 1 to 20 in terms of MAE, $RMSE$ and R^2 for the stock PingAn. The results show that the best performance is achieved at

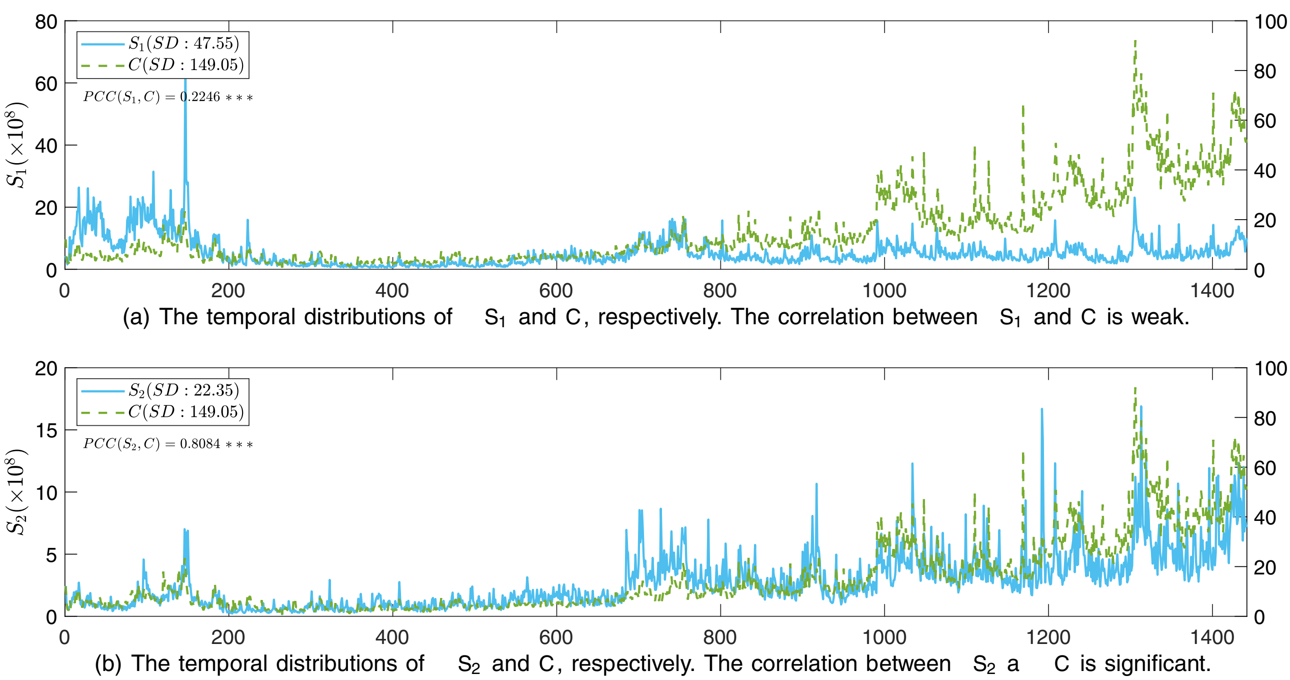
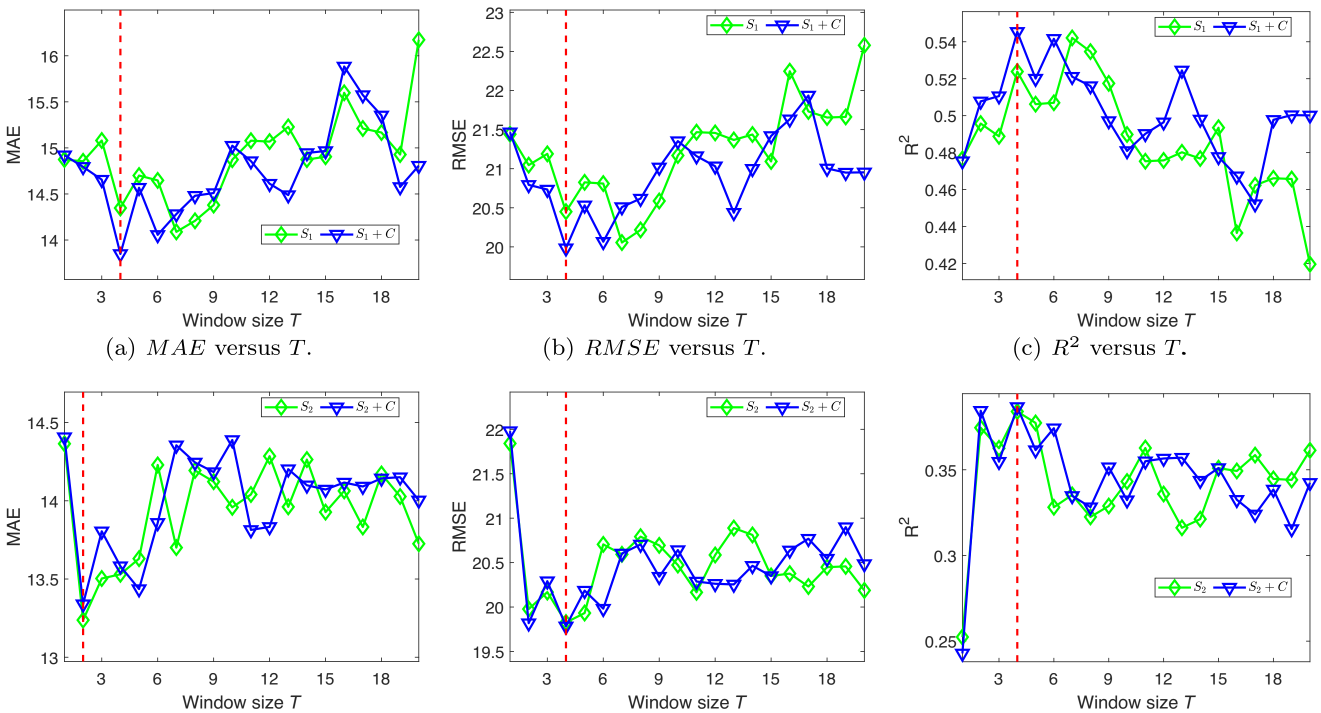


Fig. 2 The daily share turnover time series of PingAn (S_1), Moutai (S_2) and Shanghai Connect (C)



(a) MAE versus T. (b) $RMSE$ versus T. (c) R^2 versus T.

(d) MAE versus T. (e) $RMSE$ versus T. (f) R^2 versus T.

Fig. 3 The evolution of our model accuracy over a series of input window size T on PingAn in terms of MAE, $RMSE$ and R^2. Both V on S_2 and V on $S_2 + C$ is fixed at $\{1, 2, 3\}$. **a-c** stock 1. **d-f** stock 2. The optimal values are found around $T = 4$ as shown in red dashed lines

$T = 4$. In the beginning, the performance improves until $T = 4$ as T increases. After this, with the boosts of T, the performance decreases.

The potential reason behind this is that the local correlation is sensitive between historical observations and upcoming values. Redundant look-back windows reduce the stability of the model, and too short a look-back window cannot provide enough temporal information with the rapidly changing stock market. We carry out the same experiments for the stock Moutai, the results is plotted in Fig. 3d, e and f. Moutai has similar performance trends as PingAn. The best performance is achieved at $T = 4$ in terms of $RMSE$ and R^2. It is worth noticing that the optimal MAE performance found at $T = 2$, the difference is due to $RMSE$ enlarge the impact of the larger error and MAE represents the real error. More balanced predictions during peak period and trough period can be obtained by lower $RMSE$. Based on these empirical results, we fix T at 4 for the following experiments.

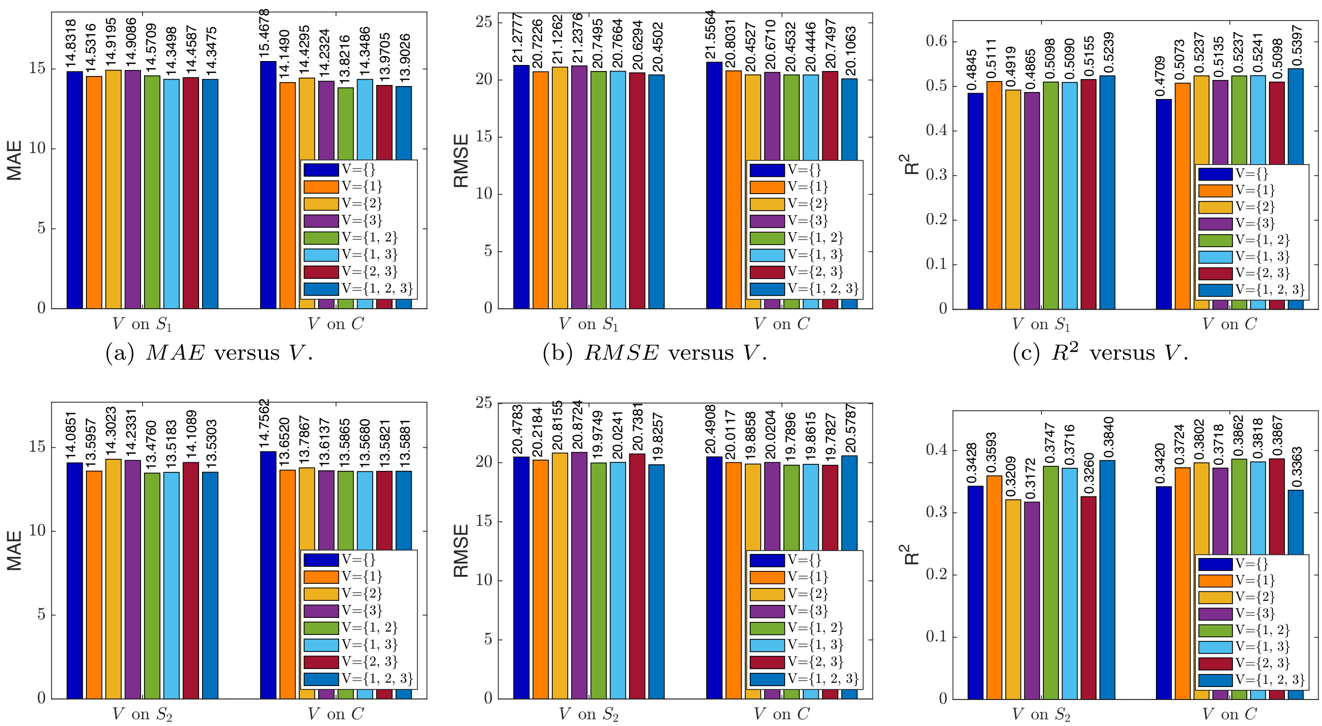
With the exogenous data, the method has more frequent fluctuations of performance compared to the other group without that. The potential reasons are: (1) there still has room to improve for this model; (2) the samples from Shanghai Connect is not enough; (3) the changes of exogenous time series delay feedback to investors.

6.2 View evaluation

In our model, the parameter V is optional and can be used in combinations arbitrarily. There are three optional views for the input data. Thus, we have six combinations of V in total.

To examine the effects of combinations on the performance of our model, we carry out the experiments on all combinations. The left part in Fig. 4a, b and c shows the performance of our model on all views for PingAn in terms of MAE, $RMSE$ and R^2. The best performance is achieved at the combination $\{1, 2, 3\}$. To investigate the benefits of the exogenous data, we first fix the view of the share turnover on the combination $\{1, 2, 3\}$ based on the above finding. After this, similar experiments are carried out on the share turnover with Shanghai Connect. The results are plotted on the right part in Fig. 4d, e and f. Not surprisingly, we achieve the best performance at the combination $\{1, 2, 3\}$. With the exogenous data, MAE and $RMSE$ decline by 3.1% and 1.7%, and R^2 improves 3%, respectively. Each of the views provides different non-linear relationships and temporal dynamical inputs, making effective use of them can generate more accurate predictions.

For the stock Moutai, we perform the same evaluation. The results are shown in Fig. 4d, e and f. Similarly, the best performance is achieved at the combination $\{1, 2, 3\}$



(a) MAE versus V. (b) $RMSE$ versus V. (c) R^2 versus V.

(d) MAE versus V. (e) $RMSE$ versus V. (f) R^2 versus V.

Fig. 4 The performance of our model on all combinations in terms of MAE, $RMSE$ and R^2 for PingAn with and without Shanghai Connect. The T is fixed at 4

for the share turnover without Shanghai Connect. However, with the exogenous data, the best performance obtains at the combination {2, 3}. One of the potential reasons behind this is the Pearson correlation coefficient of S_1 and C is less than that of S_2 and C. Due to the similar development trend of S_2 and C, especially in time steps representation, C unable to supply enough non-linear information for S_2.

6.3 Comparisons

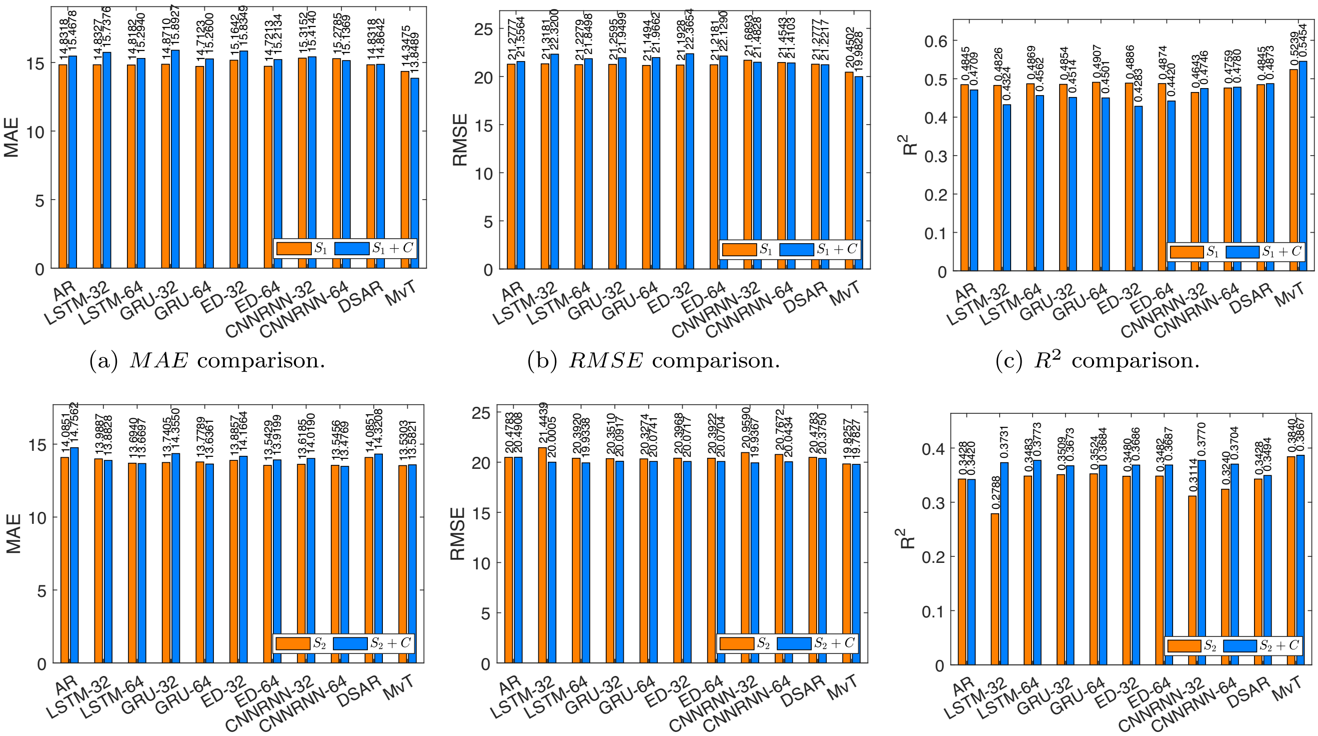
To provide unbiased and objective results, we repeat each experiment five times and average them up for all performance metrics.

We observe the results and summary the following key conclusions:

(1) The proposed MvT outperforms other benchmarks in terms of three metrics.
(2) The GRU-64 has the second best performance in terms of three metrics without exogenous data except in terms of MAE for the stock Moutai.
(3) The correlation of target and exogenous data is sensitive for RNN based methods.
(4) From the perspective of stock PingAn without exogenous data, CNNRNN-32 poorly performs than other methods.
(5) From the perspective of stock Moutai with exogenous data, AR has the worst performance than other methods.
(6) The AR based method poorly performs than other methods in stock Moutai.

Figure 5(a-c) show the performance of all models in terms of MAE, $RMSE$ and R^2 for the stock PingAn. The proposed MvT method outperforms the baseline strategies as expected. Without Shanghai Connect data, the best result of the benchmarks is achieved at GRU-64, the concise non-linear structure more appropriate for the small amount of samples. Compared with GRU-64, the improvements of our model reach 2.5%, 3.3% and 6.3% on MAE, $RMSE$ and R^2. In addition, due to insufficient extraction of CNN from observations, CNNRNN-32 poorly performs than other methods without exogenous time series. Not surprisingly, with the exogenous data, our previous model obtains the best performance. Compared with DSAR, our new model improves 6.8%, 5.8% and 11.9% on MAE, $RMSE$ and R^2 respectively. Furthermore, if we focus on our model, with Shanghai Connect, the improvements are still significant, a 3.5% decrease in MAE, a 2.3% decrease in $RMSE$, and a 4.1% increase in R^2.

Meanwhile, Fig. 5d-f show the performance of all models in terms of MAE, $RMSE$ and R^2 for the stock Moutai. Moreover, our model outperforms the baseline strategies as expected. Without Shanghai Connect data, compared with the best result of the baseline strategies, the improvements of our model reach 0.1%, 2.5% and 9.0%



(a) MAE comparison. (b) $RMSE$ comparison. (c) R^2 comparison.

(d) MAE comparison. (e) $RMSE$ comparison. (f) R^2 comparison.

Fig. 5 Comparison of all models on PingAn and Shanghai Connect data. $T = 4$, $V = \{1, 2, 3\}$

on *MAE*, *RMSE* and R^2. With the exogenous data, compared with the best result of the baseline strategies, the improvements of our model reach 0.4%, 1.1% and 2.5% on *MAE*, *RMSE* and R^2. The AR based methods, i.e. AR and DSAR, show relatively worse performance, which is unable to capture the non-linear dependencies between each time step. Weak correlation exogenous time series make more contributions to improving the prediction accuracy. A possible reason is that few similar patterns of time series are prone to boost the fitting errors, while weak correlation data increase complexity but enhance the robustness by various dynamical temporal patterns. In the stock Moutai, the performance with Shanghai Connect has a tiny improvement.

The potential reasons for such improvements brought are: (1) the benefits of the exogenous data, Shanghai Connect; (2) our method is adaptive to both past and future data; (3) our method effectively extracts the local and global dependencies from both internal and exogenous data.

7 Conclusion

This paper proposed a multi-view time series model (MvT) to predict the share turnover values in the following trading day. In order to provide comprehensive learning aspects, three views are generated to observe the temporal dynamics on PingAn, Moutai and Shanghai Connect turnover data. The predictions are figured out based on the combinations of three view outputs. Extensive experiments are carried out to verify the benefits of the proposed MvT.

In the future, the authors would like to apply the multi-view model on multi-horizon time series prediction problems.

Acknowledgments This work was supported in part by Jimei University (nos. ZP2021013, ZQ2018008 and ZP2020043), the Science project of Xiamen City (no. 3502Z20193048), the Education Department of Fujian Province (CN) (nos. JAT200277, JAT200232, JAT170327 and JAT200273), and the Natural Science Foundation of Fujian Province (CN) (nos. 2019J05099 and 2021J01859).

The authors would like to thank the editor and anonymous reviewers for their helpful comments in improving the manuscript quality.

Data Availability The data used to support the findings of this study are available from the corresponding author upon request.

Compliance with Ethical Standards

Conflict of interest None.

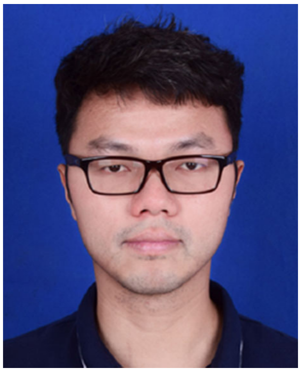
References

1. Chao G, Sun J, Lu J, Wang A-L, Langleben DD, Li C-S, Bi J (2019) Multi-view cluster analysis with incomplete data to understand treatment effects. Inf Sci 494(2):146–168

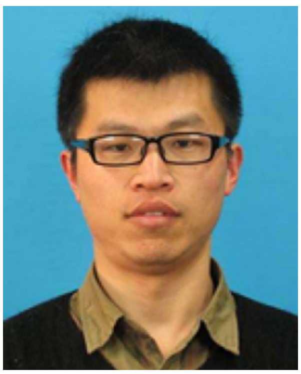
2. Chao G, Sun S (2016) Consensus and complementarity based maximum entropy discrimination for multi-view classification. Inf Sci 367(7):296–310
3. Chao G, Sun S (2018) Semi-supervised multi-view maximum entropy discrimination with expectation laplacian regularization. Inf Fusion 45(2):296–306
4. Cho K, van Merrienboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: Encoder-decoder approaches. In: Proceedings of 8th Workshop on Syntax, Semantics and Structure in Statistical Translation. Association for Computational Linguistics, Doha, pp 103–111
5. Kim HY, Won CH (2018) Forecasting the volatility of stock price index: a hybrid model integrating lstm with multiple garch-type models. Expert Syst Appl 103:25–37
6. Lee C-Y, Soo V-W (2017) Predict stock price with financial news based on recurrent convolutional neural networks. In: Proceedings of the Conference on Technologies and Applications of Artificial Intelligence. IEEE, Taipei, pp 160–165
7. Lin H, Zhou D, Liu W, Bian J (2021) Learning multiple stock trading patterns with temporal routing adaptor and optimal transport. In: Proceedings of the 27th International Conference on Knowledge Discovery and Data Mining. ACM, Virtual Event, pp 1017–1026
8. McNally S, Roche J, Caton S (2018) Predicting the price of bitcoin using machine learning. In: Proceedings of the 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing, Cambridge, pp 339–343
9. Minami S et al (2018) Predicting equity price with corporate action events using lstm-rnn. J Math Finance 8(01):58
10. Ozbayoglu AM, Gudelek MU, Sezer OB (2020) Deep learning for financial applications : a survey. Appl Soft Comput 93:106384
11. Selvin S, Vinayakumar R, Gopalakrishnan E, Menon VK, Soman K (2017) Stock price prediction using lstm, rnn and cnn-sliding window model. In: Proceedings of the international conference on advances in computing, communications and informatics. IEEE, Udupi, pp 1643–1647
12. Sermpinis G, Stasinakis C, Dunis C (2014) Stochastic and genetic neural network combinations in trading and hybrid time-varying leverage effects. J Int Financ Markets Inst Money 30:21–54
13. Sezer OB, Gudelek MU, Ozbayoglu AM (2020) Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. Appl Soft Comput 90:106181
14. Shih S, Sun F, Lee H (2019) Temporal pattern attention for multivariate time series forecasting. Mach Learn 108(8-9):1421–1441
15. Sun S, Chao G (2013) Multi-view maximum entropy discrimination. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence. AAAI Press, Beijing, pp 1706–1712
16. Sun S, Xie X, Dong C (2019) Multiview learning with generalized eigenvalue proximal support vector machines. IEEE Trans Cybern 49(2):688–697
17. Wang J, Yang Y, Mao J, Huang Z, Huang C, Xu W (2016) Cnn-rnn: a unified framework for multi-label image classification. CoRR arXiv:1604.04573
18. Wang Y, Gu J, Zhou Z, Wang Z (2015a) Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion. Knowl-Based Syst 88:12–23
19. Wang Y, Li J, Gu J, Zhou Z, Wang Z (2015b) Artificial neural networks for infectious diarrhea prediction using meteorological factors in shanghai (china). Appl Soft Comput 35:280–290
20. Wang Z, Cai B (2021) COVID-19 cases prediction in multiple areas via shapelet learning. Appl Intell 51:1–12
21. Wang Z, Huang Y, Cai B, Ma R, Wang Z (2021a) Stock turnover prediction using search engine data. J Circ Syst Comput 30:2150122:1–2150122:18
22. Wang Z, Huang Y, He B (2021B) Dual-grained representation for hand, foot, and mouth disease prediction within public health cyber-physical systems. Softw Practice Exper 51:2290–2305
23. Wang Z, Huang Y, He B, Luo T, Wang Y, Fu Y (2020) Short-term infectious diarrhea prediction using weather and search data in Xiamen, China. Sci Programm 2020:8814222:1–8814222:12
24. Wang Z, Wang Z, Lin Y, Liu J, Fu Y, Zhang P, Cai B (2021c) Prediction of HFMD cases by leveraging time series decomposition and local fusion. Wirel Commun Mob Comput 2021:5514743:1–5514743:10
25. Xie X (2018) Regularized multi-view least squares twin support vector machines. Appl Intell 49(2):688–697
26. Xie X, Sun S (2019) General multi-view learning with maximum entropy discrimination. Neurocomputing 332:184–192
27. Xie X, Sun S (2020a) General multi-view semi-supervised least squares support vector machines with multi-manifold regularization. Inf Fusion 62:63–72
28. Xie X, Sun S (2020b) Multi-view support vector machines with the consensus and complementarity information. IEEE Trans Knowl Data Eng 32(12):2401–2413
29. Xu C, Tao D, Xu C (2013) A survey on multi-view learning. CoRR arXiv:1304.5634
30. Zhang B, Chan JCC, Cross JL (2020) Stochastic volatility models with ARMA innovations: An application to G7 inflation forecasts. Int J Forecast 36:1318–1328
31. Zhang L, Aggarwal CC, Qi G (2017) Stock price prediction via discovering multi-frequency trading patterns. In: Proceedings of the 23rd International Conference on Knowledge Discovery and Data Mining. ACM, Halifax, pp 2141–2149
32. Zhao J, Xie X, Xu X, Sun S (2017) Multi-view learning overview: Recent progress and new challenges. Inf Fusion 38:43–54
33. Zhou X, Pan Z, Hu G, Tang S, Zhao C (2018) Stock market prediction on high-frequency data using generative adversarial nets. Math Probl Eng 2018:1–11

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Zhijin Wang received the Ph.D. degree from the Department of Computer Science and Technology, East China Normal University, Shanghai, China, in 2016. He is currently with the Computer Engineering College, Jimei University, Xiamen, China. His current research interests include recommendation system, data mining, and artificial intelligence in healthcare.



Qiankun Su received the Ph.D. degree from the INP-ENSEEIHT, University of Toulouse, France. Now, he is a lecturer at Computer Engineering College, Jimei University, Xiamen, P.R. China. His research interests include wireless network, network coding, and time series forecasting.



Guoqing Chao received the B.S. degree from Xinyang Normal University, Xinyang, China, in 2009. He got the Ph.D. degree with the Department of Computer Science and Technology, East China Normal University, Shanghai, China in 2015. After that, he was a PostDoc in University of Connecticut and Northwestern University, US and Singapore Management University, Singapore. Currently, he works with Harbin Institute of Technology at Weihai, China. His research interests include machine learning, data mining, medical informatics, bioinformatics and service computing.

Bing Cai is currently pursuing a B.S. degree in computer science in Jimei University, Xiamen, China. His research interests include artificial intelligence and time series forecasting.

Yaohui Huang received the B.S. degree in computer science from Chengyi University College, Jimei University, Xiamen, China, in 2020, and is currently pursuing the master's degree with Guangxi University for Nationalities, Nanning, China. His research interests include data mining, time series data processing and analysis, deep learning and data fusion.

Yonggang Fu received his bachelor and master degree from Xi'an Jiaotong University in Computational Mathematics in 1995 and 1998 respectively. He received his doctor degree in Computer Software and Theory in Shanghai Jiaotong University in 2005. Now he is a professor in Jimei University. His main research concentrated in deep learning, multimedia security, and artificial intelligence and so on.