Knowledge-Based Systems 262 (2023) 110262

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Essential tensor learning for multimodal information-driven stock movement prediction

Jun Wang [a], Yexun Hu [a], Tai-Xiang Jiang [a,*], Jinghua Tan [b,*], Qing Li [a]

[a] School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] College of Economics, Sichuan Agricultural University, Sichuan Agricultural University, Chengdu 611130, China

ARTICLE INFO

Article history:
Received 28 June 2021
Received in revised form 25 December 2022
Accepted 2 January 2023
Available online 6 January 2023

Keywords:
Tensor robust principal component analysis
Stock movements
Multimodal and multitemporal

ABSTRACT

In the literature, an increasing amount of information from various sources related to the stock market is being considered for stock movement prediction. However, previous studies usually modeled market information as a vector, failing to effectively utilize the inner structure in terms of multimodal and multitemporal characteristics. Moreover, the release, dissemination, and absorption of information causing spillover effects from stocks related to the target stock should not be neglected in today's information society. Thus, this study proposes a general tensor representation and fusion framework to capture the intrinsic interactions of multimodal and multitemporal stock market information based on the invariant correlations among stocks within a short period. Specifically, we construct a general correlation matrix to represent the correlation between the stocks with respect to a given mode of information for a single day. Then, for a short period, with multimodal information, the matrices are concatenated into a tensor, which is highly inner correlated. A tensor robust principal component analysis (TRPCA) model is then employed to fuse the multimodal and multitemporal information, adaptively infer essential interactions, and faithfully enhance the inner correlation of the constructed tensor. Experiments on real datasets show that the proposed tensor representation and fusion framework can efficiently improve the performance of stock movement prediction. The performance of the investment simulation further illustrates the superiority of the proposed method in terms of the return rate (26.73%) for a full year.

1. Introduction

The stock market is a place for investors to realize financial asset allocation, and its healthy and stable development is an important guarantee of economic growth. With the rise of Internet media, the emergence of massive market information has led to constant changes in investors' expectations of stock prices, and the resulting risk of abnormal stock price fluctuations has attracted considerable attention from academia and industry. In fact, both the "efficient market hypothesis" and "irrational investor theory" collaborated to verify the close correlation between stock market volatility and media information release, dissemination, and absorption [1,2]. Therefore, how to accurately quantify the effects of media information has become an important challenge in the field of stock market volatility research.

In previous studies, scholars have identified many market information factors that influence stock price volatility, including trading data, news, social media, public sentiment, and search behavior [3]. Therefore, scholars have attempted to analyze stock market volatility using fused market information. For example, [4] synthesized stock information and *Wall Street Journal* content and found that news sentiment combined with stock fundamentals can effectively portray stock market volatility trends. [5] also found that multidimensional data have better predictive performance. These studies show that the stock market is influenced by multidimensional market information, and the fusion analysis of different types of market information on stock volatility has received much attention.

However, fusing multimodal market information accurately and efficiently remains challenging. In recent years, as a generalization of the matrix, the tensor has become an important tool for better modeling the relationships of multisource or multimodal data, and has been applied to a wide range of real-world problems, such as image processing, earthquake prediction, personal web search, and higher-order web link analysis. For the stock market, the tensor format has been shown to well capture the interaction among different pieces of market information [6,7]. Furthermore, considering that most previous studies have been trained on single stocks, severing the correlation between

* Corresponding authors.
E-mail addresses: wangjun1987@swufe.edu.cn (J. Wang), 220081202001@smail.swufe.edu.cn (Y. Hu), taixiangjiang@gmail.com (T.-X. Jiang), 595915575@qq.com (J. Tan), liqing@swufe.edu.cn (Q. Li).

https://doi.org/10.1016/j.knosys.2023.110262
0950-7051/

stocks and the stock market environment, scholars are beginning to take the stock comovement relationship into account in their characteristic representations by decomposing and reconstructing the tensor [6]. For example, [8] used coupled stock similarity to construct the stock-correlation matrix and incorporate the correlation into the stock data tensor representation through a tensor reconstruction algorithm to achieve better prediction results. [9] constructed a stock association network based on media news data and characterized a single stock in the tensor manner.

In fact, because the amount and speed of information dissemination in today's stock markets have increased dramatically, the effects of stock comovements are reflected at a faster rate, bringing new challenges and opportunities for stock market forecasting. Unfortunately, most previous studies treat stocks in the market as isolated, which makes it difficult to capture the underlying mechanisms of stock comovements. Therefore, instead of treating stocks independently, we model the correlation of stocks, with respect to different information sources in the market, under a unified tensor representation scheme. Then, a tensor robust principal component analysis (TRPCA) model is further established to adaptively fuse the multimodal and multitemporal data by employing the homogeneity of those correlations. The fusion result captures the intrinsic correlations and interactions among different stocks from the multimodal and multitemporal market data. Finally, an attention-based long short-term memory (LSTM) classifier is trained based on the fusion results and used for stock market forecasting. The contributions of this study are as follows.

- We propose a novel stock-correlation representation scheme, in which the multimodal and multitemporal market information is well structured in the tensor format. Then, to fully exploit the essential inner correlations, we employ a tensor robust principal component analysis (TRPCA) model to further organically fuse multimodal and multitemporal market information.
- Based on the results from our general tensor representation and fusion framework, a long short-term memory (LSTM) with the attention mechanism is tailored to predict stock movements. Experiments are conducted on one full year of data on the Chinese securities market and one and a half years of data on S&P 500 firms. Our method, which consists of the tensor representation, TRPCA fusion, and the attention-based LSTM, achieves better performances compared with state-of-the-art methods.

2. Related work

In this section, we first review the relevant literature on stock movement prediction and tensor representation. Subsequently, some basic concepts are provided on which we established our tensor framework.

2.1. Media-aware stock movements

In the digital age, stock price fluctuations depend on the combination of numerous pieces of market information. In previous studies, owing to the simple structure of the market information environment, scholars focused on refining the main influential factors of stock market operation [10–12], such as exploring the effect of different market factors on stock volatility one by one from the perspectives of economic indicators, market environment, policy changes, investor sentiment, etc. That is, they dismantled and analyzed the causal links between various market factors and stock market volatility. However, with the increasing complexity of the stock market information environment, the analysis of market factors based on a single dimension is too one-sided, and market-influencing factors are no longer a simple numerical value to be characterized. As a result, scholars have begun to note that the fusion analysis behind various types of market information is the key to resolving stock price volatility. In this context, the study of various types of market information fusion related to stock volatility has become mainstream in academia.

In the study of market information fusion, scholars have broadly divided market information into two categories, i.e., fundamental information and media information, and based on the combination of these two types of market information, they have explored the effect of market information fusion. In the process of modeling the fusion of fundamental and media information, the market information fusion process aims to capture the combined effect of multiple pieces of market information on stock volatility. How to preserve as much as possible the interaction between different types of market information while fusing them is the key to the fusion process. The common strategy of fusion methods, however, is to simply concatenate numerical economic indicators and textual vector-based media into a super compound vector, which inevitably ignores the interactions among different information types, resulting in the loss of some key information [13,14]. To capture the intrinsic associations among different information sources, some researchers have applied tensor theory to model the complicated market information space to gain a better understanding of stock market movements [6,15]. Specifically, those studies represent one type of information as a tensor mode, and the core tensor is applied to record the links among different information sources. The core tensor is constructed based on the input feature space of market information via tensor decomposition and reconstruction, which can be considered a static fusion method.

Notably, the stock market has two important characteristics: time series and comovement. First, the stock trend itself is a time series. Since stock prices are not completely controlled by random factors, the nature of the time series can objectively describe the pattern of historical data changes over time, which is useful for predicting future stock price trends [6]. The static fusion analysis approach disconnects the coherence of market information in time and constructs the links of different sources based merely on their physical structures rather than their natural interactions, resulting in stock price forecasts that are much less effective. Second, modern finance research has pointed out that the relationship among listed firms causes momentum spillover effects, i.e., the stock returns of relevant listed companies help in the prediction of the stock trend of the target firm. These momentum spillover effects among related stocks are affected by a variety of interfirm linkages or similarities. Restricting interfirm relatedness to a particular type of firm relation makes it difficult to capture the entirety and nature of the momentum spillover effects among related firms, which inevitably causes errors in stock predictions.

To summarize, in contrast to most existing studies, which consider only market information fusion in the representation stage, our method fuses multimodal and multitemporal data by employing the homogeneity of those correlations to capture the intrinsic correlations and interactions among different stocks. The proposed approach of market information fusion aims to achieve better stock predictive power, which has rarely been covered in previous studies.

2.2. Tensor

The tensor is a higher-order generalization of the vectors and matrices. It is able to structurally characterize data from different sources, such as hyperspectral images [16,17], magnetic resonance imaging (MRI) data [18], video data [19], high-order web links [20], personalized web search data [21], and seismic data [22]. By representing the data in a tensor-based manner, the inner structure of the data, which are always inherently multidimensional, would naturally be preserved. In general, those data in the tensor format are always inner correlated, resulting in low rankness. This contributes to mining the high-order correlations for subsequent applications, such as action classification, multiview subspace spectral clustering, face recognition, and multivariate spatiotemporal analysis [23,24].

Since a tensor is able to fuse multidimensional data effectively and can also reduce the noise of fusion via multiple decompositions, it has received extensive scholarly attention in media-aware stock movements. For example, Li et al. [6] applied tensors to stocks by exogenously fusing multidimensional information on historical transaction data, media sentiment, and company business conditions into a tensor to investigate the effect of multidimensional features on stock price changes. Wang et al. [25] proposed a novel multimodal tensor fusion network (MTFN) to achieve significant matching performance between images and text at acceptable model complexity.

The reason that tensors can capture the inner interactions of multiple points of market data is that decomposition methods for tensors can reduce the multidimensional data to achieve the effect of data fusion. Common tensor decomposition methods include the CANDECOMP/PARAFAC (CP) decomposition [26,27], the Tucker decomposition [28,29], and tensor singular value decomposition (t-SVD) [30–32]. The main purpose of tensor decomposition is to reduce the rank of the tensor to obtain a representation of the core data in the tensor. By decomposing the constructed tensor, we can obtain the multisource fusion information containing the best representation.

In particular, the ranks of the tensors are still not unique, and there are four mainstream notions of the tensor rank, i.e., the CP-rank [33] based on CP decomposition, the Tucker-rank corresponding to the Tucker decomposition [28], the tensor train (TT)-rank derived from the TT decomposition [34], and the newly emerged tubal-rank defined with t-SVD [31]. In this work, we mainly focus on the t-SVD framework, since the tensor–tensor product is first defined within it, avoiding the loss of information inherent in matricization or flattening [31]. Although the tensor rank minimization problems are always NP-hard, minimizing the tensor nuclear norm (TNN), which is a surrogate of the tubal-rank and easy to optimize, has shown its effectiveness in enhancing the low-rankness [35,36].

2.3. Tensor basics and preliminaries

Throughout this study, we use lowercase letters, e.g., x, boldface lowercase letters, e.g., $\mathbf{x}$, boldface uppercase letters, e.g., $\mathbf{X}$, and boldface calligraphic letters, e.g., $\boldsymbol{\mathcal{X}}$, to denote scalars, vectors, matrices, and tensors, respectively. Here, $\boldsymbol{\mathcal{X}}_{ijk}$ denotes the (i,j,k)-th element of a given third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times n_2\times n_3}$. The tensor Frobenius norm of a third-order tensor $\boldsymbol{\mathcal{X}}$ is defined as $\|\boldsymbol{\mathcal{X}}\|_F := \sqrt{\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{X}}\rangle} = \sqrt{\sum_{ijk} \boldsymbol{\mathcal{X}}_{ijk}^2}$.

Next, we list the basic definitions and a theorem of the t-SVD framework and the tensor nuclear norm (TNN). Those contents were derived in [32,37]; we restate them here for the convenience of the reader.

Definition 1 (*Tensor Conjugate Transpose*)**.** The conjugate transpose of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_2\times n_1\times n_3}$ is the tensor $\boldsymbol{\mathcal{A}}^{\mathrm{H}} \in \mathbb{C}^{n_1\times n_2\times n_3}$ obtained by conjugate transposing each of the frontal slices and then reversing the order of the transposed frontal slices 2 through n_3, that is, $\left(\boldsymbol{\mathcal{A}}^{\mathrm{H}}\right)^{(1)} = \left(\boldsymbol{\mathcal{A}}^{(1)}\right)^{\mathrm{H}}$ and $\left(\boldsymbol{\mathcal{A}}^{\mathrm{H}}\right)^{(i)} = \left(\boldsymbol{\mathcal{A}}^{(n_3+2-i)}\right)^{\mathrm{H}}$ ($i = 2,\ldots,n_3$).

Definition 2 (*t-prod*)**.** The tensor–tensor product (t-prod) $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}}$ of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n_2\times n_4\times n_3}$ is a tensor of size $n_1 \times n_4 \times n_3$, where the (i,j)-th tube $\mathbf{c}_{ij:}$ is given by

$$\mathbf{c}_{ij:} = \boldsymbol{\mathcal{C}}(i,j,:) = \sum_{k=1}^{n_2} \boldsymbol{\mathcal{A}}(i,k,:) * \boldsymbol{\mathcal{B}}(k,j,:) \tag{1}$$

where $*$ denotes the circular convolution between two tubes of the same size.

Definition 3 (*Special Tensors*)**.** The **identity** tensor $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n_1\times n_1\times n_3}$ is a tensor whose first frontal slice is the $n_1 \times n_1$ identity matrix and whose other frontal slices are all zeros. A tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{C}^{n_1\times n_1\times n_3}$ is **orthogonal** if it satisfies $\boldsymbol{\mathcal{Q}}^{\mathrm{H}} * \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} * \boldsymbol{\mathcal{Q}}^{\mathrm{H}} = \boldsymbol{\mathcal{I}}$. A tensor $\boldsymbol{\mathcal{A}}$ is called **f-diagonal** if each frontal slice $\boldsymbol{\mathcal{A}}^{(i)}$ is a diagonal matrix.

Theorem 1 (*t-SVD*)**.** *For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the t-SVD of $\boldsymbol{\mathcal{A}}$ is given by*

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^H \tag{2}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal tensors and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is an f-diagonal tensor.

Definition 4 (*Tensor Tubal-Rank*)**.** The tubal-rank of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, denoted by $\mathrm{rank}_t(\boldsymbol{\mathcal{A}})$, is defined as the number of nonzero singular tubes of $\boldsymbol{\mathcal{S}}$, where $\boldsymbol{\mathcal{S}}$ comes from the t-SVD of $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^\top$. That is, $\mathrm{rank}_t(\boldsymbol{\mathcal{A}}) = \#\{i : \boldsymbol{\mathcal{S}}(i,:,:) \neq 0\}$.

Definition 5 (*Block Diagonal Operation*)**.** The block diagonal operation of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is given by

$$\mathtt{bdiag}(\boldsymbol{\mathcal{A}}) \triangleq \begin{bmatrix} \boldsymbol{\mathcal{A}}^{(1)} & & & \\ & \boldsymbol{\mathcal{A}}^{(2)} & & \\ & & \ddots & \\ & & & \boldsymbol{\mathcal{A}}^{(n_3)} \end{bmatrix}, \tag{3}$$

where $\mathtt{bdiag}(\boldsymbol{\mathcal{A}}) \in \mathbb{C}^{n_1 n_3\times n_2 n_3}$.

Definition 6 (*Tensor Nuclear Norm (TNN)*)**.** The TNN of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, denoted by $\|\boldsymbol{\mathcal{A}}\|_{\mathrm{TNN}}$, is defined as

$$\|\boldsymbol{\mathcal{A}}\|_{\mathrm{TNN}} \triangleq \|\mathtt{bdiag}(\boldsymbol{\mathcal{Z}})\|_*, \tag{4}$$

where $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is the Fourier transformed (along the third mode) tensor of $\boldsymbol{\mathcal{A}}$. The TNN can be computed by summing the matrix nuclear norms of the frontal slices of $\boldsymbol{\mathcal{Z}}$. That is, $\|\boldsymbol{\mathcal{A}}\|_{\mathrm{TNN}} = \sum_{i=1}^{n_3} \|\boldsymbol{\mathcal{Z}}(:,:,i)\|_*$.

3. Proposed essential tensor learning framework

As aforementioned, the stock market is affected by many factors from various sources. In the early literature, researchers usually constructed a feature vector with the direct concatenation of the multimodal and heterogeneous data. This will unavoidably destroy the intrinsic structure of these data, causing the loss of abundant important information [13]. In addition, stocks in the market are not isolated. Based on the momentum spillover effect, the price of one stock will also be affected by the correlated companies. In this paper, we propose an essential tensor learning framework for multimodal and multitemporal market information, which consists of three modules, i.e., tensor representation, fusion stage, and prediction. The tensor representation and fusion stage could capture the inner correlation between different stocks from multimodal factors. Subsequently, these fusion results are utilized to track stock movements through the prediction module. Fig. 1 shows an overview of the proposed approach.

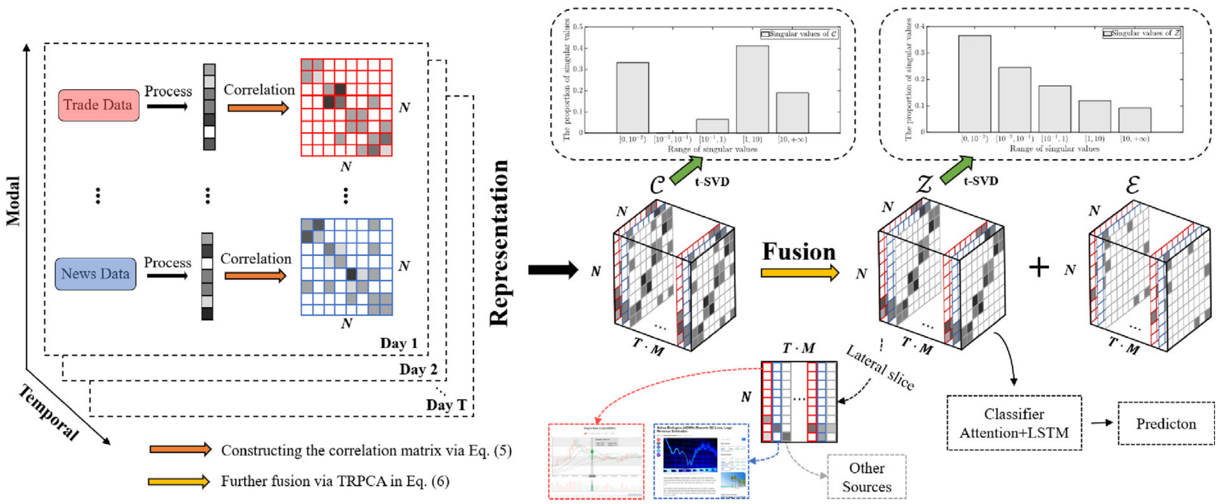


Fig. 1. Flowchart of the proposed tensor representation and fusion framework for stock prediction.

3.1. Tensor-based representation for multimodal and multitemporal data

In this section, we propose unifying multimodal and multitemporal market information factors in the tensor format, while maintaining their inner structure.

Suppose there are N stocks in the market, and for each of them, we have M types of market information, e.g., historical transactions, media sentiment, and company business conditions, that cause stock market volatility. At the tth ($t = 1, 2, \ldots, T$) day, for the ith company ($i = 1, 2, \ldots, N$), its jth ($j = 1, 2, \ldots, M$) type of the market information can be written as a feature vector as

$$\mathbf{x}_i^{j(t)} \in \mathbb{R}^{d(j)\times 1},$$

where $d(j)$ indicates the dimension of the features of the jth type of the market information. For instance, the historical transaction data contain the highest price, lowest price, opening price, closing price, turnover, trading volume, and P/B and P/E ratios, composing an eight-dimensional feature vector.

The jth type of market information for all stocks in the market can be formulated as a matrix:

$$\mathbf{X}^{j(t)} = \left[\mathbf{x}_1^{j(t)}, \mathbf{x}_2^{j(t)}, \ldots, \mathbf{x}_N^{j(t)}\right]^\top \in \mathbb{R}^{N\times d(j)}.$$

We can see that $X^{j(t)}$s ($j = 1, 2, \ldots, M$ and $t = 1, 2, \ldots, T$) constitute multimodal, multitemporal, and heterogeneous factors. It is not easy to directly fuse them without destroying their inner structures. Fortunately, the inner correlation between different stocks is expected to be homogeneous. Thus, for the jth type of market information, we construct the correlation matrix of different stocks as

$$\mathbf{C}^{j(t)} = \gamma^{j(t)}\mathbf{X}^{j(t)}\mathbf{X}^{j(t)\top}\mathbf{S}^{j(t)} \in \mathbb{R}^{N\times N}, \tag{5}$$

where $\mathbf{X}^{j(t)} \in \mathbb{R}^{N\times N}$ is the similarity matrix of the jth type of market information on the tth day and $\gamma^{j(t)}$ is a nonnegative weighting parameter[1] related to the scale of $x_i^{j(t)}$. Specifically, with scale information, the use of nonnegative weighting parameter $\gamma^{j(t)}$ can further preserve the heterogeneous characteristics of each stock while preserving the homogeneous correlation of stocks. In this study, we set $\gamma^{j(t)} = \langle \sum_{i=1}^N \mathbf{x}_i^{j(t)}, \sum_{i=1}^N \mathbf{x}_i^{j(t)} \rangle$, and it works well. Here, $\mathbf{S}^{j(t)} \in \mathbb{R}^{N\times N}$ is a diagonal matrix, defined as follows:

$$\mathbf{S}^{j(t)} = \begin{bmatrix} \sum_{i=k}^N s_{k,1} & 0 & \cdots & 0 \\ 0 & \sum_{i=k}^N s_{k,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sum_{i=k}^N s_{k,N} \end{bmatrix},$$

where $s_{k_1,k_2} = \exp(\frac{\|\mathbf{x}_{k_1}^{j(t)} - \mathbf{x}_{k_2}^{j(t)}\|_2^2}{\eta^2})$ and $k_1, k_2 = 1, 2, \ldots, N$. The role of $\mathbf{S}^{j(t)}$ in Eq. (5) is to further enhance the similarity among stocks, which are originally close to each other, in the correlation matrix.

Finally, the correlation matrices are stacked together as a tensor $\mathcal{C} \in \mathbb{R}^{N\times N\times MT}$. We can see that each N-by-N slice of $\mathcal{C}$ represents the correlation of N stocks in the market with respect to a single day and a given factor. Moreover, as shown in Fig. 1, a lateral slice reflects the correlation of a given stock with other stocks in T days and with M types of market information. The columns of this lateral slice are expected to be relevant, which allows for further fusion in the subsequent section.

3.2. TRPCA for multimodal and multitemporal data fusion

The tensor $\mathcal{C} \in \mathbb{R}^{N\times N\times MT}$ represents M types of market information, with respect to N stocks in T days. To adaptively infer the homogeneity and reject the bias caused by outliers, which are unavoidable in the market information, we employ the TRPCA [36] model:

$$\begin{aligned} \min_{\mathcal{Z},\mathcal{E}} \quad & \|\mathcal{Z}\|_{\text{TNN}} + \lambda\|\mathcal{E}\|_1 \\ \text{s.t.} \quad & \mathcal{Z} + \mathcal{E} = \mathcal{C}, \end{aligned} \tag{6}$$

where $\mathcal{Z}, \mathcal{E} \in \mathbb{R}^{N\times N\times MT}$ are the homogeneous part and the sparse error part, respectively, and λ is a nonnegative parameter. Here, $\|\mathcal{Z}\|_{\text{TNN}}$ is the TNN of $\mathcal{Z}$. Minimizing the TNN enhances the low-rankness of $\mathcal{Z}$, which enforces the essential structure of those N-by-N slices of $\mathcal{C}$ fused organically in $\mathcal{Z}$. $\|\mathcal{E}\|_1$ denotes the ℓ_1 norm of $\mathcal{E}$, i.e., the sum of the absolute values of the entries in $\mathcal{E}$, the minimization of which enhances the sparsity of $\mathcal{E}$.

[1] We remark here that multiplying by a nonnegative weighting parameter will not break the homogeneity of those correlation matrices and that retaining the scale information will be helpful for subsequent prediction.

We adopt the alternating direction method of multipliers (ADMM) to minimize Eq. (6). First, the augment Lagrangian function of Eq. (6) is

$$\begin{aligned} &L_\beta(\mathcal{Z},\mathcal{E},\mathcal{M})\\ =&\|\mathcal{L}\|_* + \lambda\|\mathcal{E}\|_1 + \langle\mathcal{M}, \mathcal{C}-\mathcal{L}-\mathcal{E}\rangle + \frac{\beta}{2}\|\mathcal{C}-\mathcal{L}-\mathcal{E}\|_F^2 \\ =&\|\mathcal{L}\|_* + \lambda\|\mathcal{E}\|_1 + \frac{\beta}{2}\|\mathcal{C}-\mathcal{L}-\mathcal{E}+\frac{\mathcal{M}}{\beta}\|_F^2 + \text{const.}, \end{aligned} \tag{7}$$

where β denotes the Lagrange parameter, $\mathcal{M}\in\mathbb{R}^{N\times N\times MT}$ denotes the Lagrangian multiplier, and const. $= -\frac{\beta}{2}\|\frac{\mathcal{M}}{\beta}\|_F^2$ is constant with respect to $\mathcal{Z}$ and $\mathcal{E}$.

Then, we alternately update $(\mathcal{Z},\mathcal{E},\mathcal{M})$ as

$$\begin{cases} \mathcal{Z}^{k+1} = \mathcal{U} * \mathcal{S}\text{hrink}_{\frac{1}{\beta}}(\mathcal{S}) * \mathcal{V}^\top, \\ \mathcal{E}^{k+1} = \mathcal{S}\text{hrink}_{\frac{\lambda}{\beta}}\left(\mathcal{C}-\mathcal{Z}^{k+1}+\frac{\mathcal{M}^k}{\beta}\right), \\ \mathcal{M}^{k+1} = \mathcal{M}^k + \beta(\mathcal{C}-\mathcal{Z}^{k+1}-\mathcal{E}^{k+1}), \end{cases} \tag{8}$$

where $\mathcal{U}*\mathcal{S}*\mathcal{V}^\top$ denotes the t-SVD of $\mathcal{C}-\mathcal{E}^k-\frac{\mathcal{M}^k}{\beta}$ and the tensor **soft-thresholding operator** $\mathcal{S}\text{hrink}_v(\cdot)$ indicates that

$$\mathcal{S}\text{hrink}_v(\mathcal{A}) = \overline{\mathcal{A}}$$

with

$$\overline{\mathcal{A}}_{i_1 i_2\cdots i_N} = \begin{cases} \mathcal{A}_{i_1 i_2\cdots i_N} - v, & \mathcal{A}_{i_1 i_2\cdots i_N} > v, \\ 0, & \text{otherwise.} \end{cases}$$

From the histogram of the singular values[2] of $\mathcal{C}$ and $\mathcal{Z}$ after the t-SVD in Fig. 1, we can see that the singular values in $\mathcal{Z}$ are sparser than those in $\mathcal{C}$. Thus, the correlations among stocks are believed to be more well-organized in the fusion result $\mathcal{Z}$. Next, the fusion result $\mathcal{Z}$ is adopted to train the classifier for prediction.

3.3. Attention-based LSTM

The prediction of stock trends is a time-series problem since the stock movements are also determined by historical market information in addition to the current condition. To capture the time-series relationship between fused features and stock movements, the LSTM model is utilized to make predictions in our proposed framework. The LSTM model is a variant of the recurrent neural network (RNN). The RNN is a deep network architecture in which the connections between hidden units form a directed cycle and the previous information on hidden states can be kept with this feedback loop mechanism. Therefore, RNNs are preferred for problems where the system needs to store and update the context information for long-term dependencies. However, because of the vanishing and exploding gradient problems, the gradient becomes too small or too large, which makes it difficult to optimize the long-term mechanism of RNNs. To handle long-term dependencies, the LSTM model is proposed to address the optimization process with a gate structure [38].

Specifically, the LSTM unit comprises forget, input, and output gates. Once fused market information $\mathbf{z}_t$ and the cell memory $\mathbf{c}_{t-1}$ are obtained, forget gate f_t is used to allow for useless information in $\mathbf{c}_{t-1}$ to be discarded with respect to $\mathbf{z}_t$. Note that, $\mathbf{z}_t$ is the vectorized vector of $\mathcal{Z}$ in Eq. (6). The input gate $\mathbf{i}_t$ is utilized to control how much current market information should be absorbed into the candidate cell memory flow $\widetilde{\mathbf{c}_t}$. In this design, the previous cell memory $\mathbf{c}_{t-1}$ can be updated with the current information and evolves into the current memory $\mathbf{c}_t$. Then, $\widetilde{\mathbf{c}_t}$, $\mathbf{f}_t$, $\mathbf{i}_t$, and $\mathbf{c}_t$ are calculated as follows:

$$\widetilde{\mathbf{c}_t} = \tanh(\mathbf{W}_c\mathbf{z}_t + \mathbf{U}_c\mathbf{h}_{t-1} + \mathbf{b}_c) \tag{9}$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f\mathbf{z}_t + \mathbf{U}_f\mathbf{h}_{t-1} + \mathbf{b}_f) \tag{10}$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i\mathbf{z}_t + \mathbf{U}_i\mathbf{h}_{t-1} + \mathbf{b}_i) \tag{11}$$

$$\mathbf{c}_t = f_t * \mathbf{c}_{t-1} + \mathbf{i}_t * \widetilde{\mathbf{c}_t}, \tag{12}$$

where $\mathbf{z}_t$ represents the current market information, $\mathbf{h}_{t-1}$ denotes the previous hidden states, and $\mathbf{c}_{t-1}$ denotes the previous cell memories. Here, $\{\mathbf{W}_c, \mathbf{U}_c, \mathbf{b}_c\}$, $\{\mathbf{W}_f, \mathbf{U}_f, \mathbf{b}_f\}$, and $\{\mathbf{W}_i, \mathbf{U}_i, \mathbf{b}_i\}$ are the network parameters of the candidate memory and the forget and input gates, respectively. Two classic activation functions, sigmoid and tanh, are also adopted. Finally, the cell memory $\mathbf{c}_t$ is able to capture the valuable market patterns hidden in both the previous and current periods. The output gate $\mathbf{o}_t$ is further utilized to process the current cell memory $\mathbf{c}_t$ and the market information to obtain the output $\mathbf{h}_t$. Specifically,

$$\mathbf{o}_t = \sigma(\mathbf{W}_o\mathbf{z}_t + \mathbf{U}_o\mathbf{h}_{t-1} + \mathbf{b}_o) \tag{13}$$

$$\mathbf{h}_t = \mathbf{o}_t \cdot \tanh(\mathbf{c}_t). \tag{14}$$

By using the proposed architecture, we are able to capture time-series rules for the market information. In addition, by considering the different levels of influence of the market information in different periods, it is critical to identify important information [39]. Inspired by the attention mechanism of the human brain [40,41], which can use limited attention resources to select a small amount of critical information from a large amount of data, we attempt to use the attention mechanism to further distinguish the effects of market information across time in the LSTM model. In particular, assume that $\mathbf{H}$ is the cell state vector $[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_t]$ produced by the LSTM layer. The output $\mathbf{c}_t$ calculated via the attention layer with $\mathbf{H}$ is described as

$$\mathbf{c}_t = \sum_{i=1}^{t} a_{t,i}\mathbf{h}_i, \tag{15}$$

where

$$a_{t,i} = \frac{\exp(e_i)}{\sum_{j=1}^{t}\exp(e_j)}, \quad e_j = \mathbf{w}_u\left(\tanh(\mathbf{w}_a\mathbf{h}_j + \mathbf{b}_a)\right)^\top, \tag{16}$$

e is the score function, and $\mathbf{w}_u$, $\mathbf{w}_a$, and $\mathbf{b}_a$ are the variables in the attention layer. The training details of this classifier are presented in the subsequent section.

4. Experimental results

In this section, we conduct experiments to test the effectiveness of the proposed tensor representation and fusion framework for predicting information-driven stock movements. After giving the implementation details of the experiments, we first examine the performance of the proposed classifier by comparing it with state-of-the-art methods. Next, discussions together with the ablation study are reported. Finally, a simulation investment is conducted to evaluate the proposed approach.

Dataset. In this study, we extend the stock data provided by [13] with the news information crawled by our focus-topic crawler. Specifically, our dataset focuses on the data of companies in CSI 100 and covers January 1 to December 31, 2015. Thirty-six companies ($N = 36$) with completed fundamental data are selected. The fundamental data consist of the volume, turnover rate, P/E

2 Here, singular values refer to the diagonal entries in each slice of the f-diagonal tensor in Eq. (2) after the t-SVD.

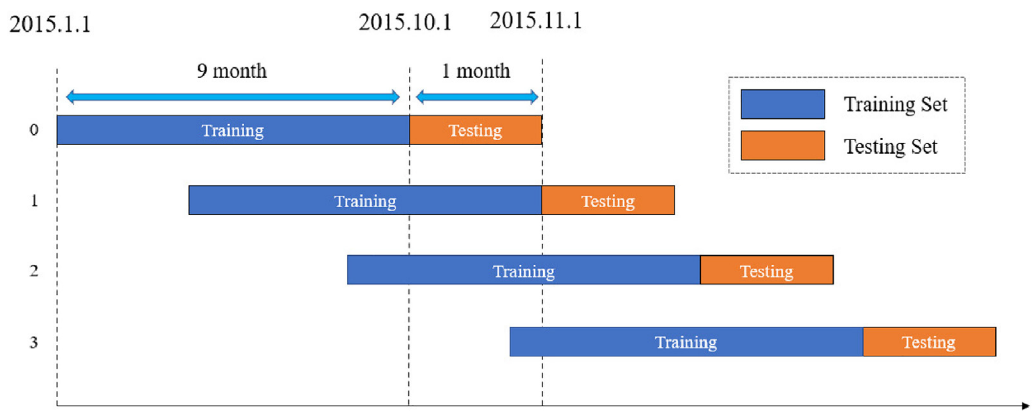


Fig. 2. Example of the rolling training-test data splits.

ratio, P/B ratio, and opening, closing, highest, and lowest prices, constituting 8-length feature vectors ($d_1 = 8$). The news in our dataset is crawled from East Money,[3] which has almost 10 million views per day. By preprocessing via the sentiment qualification technique [9], we obtain the sentiment features, including the positive and negative scores, of length 3 ($d_2 = 3$). Therefore, we have two modes of market information (i.e., $M = 2$) in our framework. To ensure the robustness of the experiments, we further carry out a series of experiments on the actual market data of S&P 500 firms. This dataset contains the transactional records and relevant news articles of the period from February 8, 2011, to November 18, 2013. Daily transaction data are obtained from Wharton Research Data Services (WRDS).[4] The media corpus is generously provided by [42], which contains textual financial news from Reuters and Bloomberg. Every successive 10 months of data are set as one period, in which the first 9 months are for training and the remaining 1 month is for testing (see Fig. 2).

Evaluation metrics. Stock movement prediction is a binary classification problem. Following the studies of [9,43], we use the accuracy (ACC) and Matthews correlation coefficient (MCC) to evaluate the effectiveness of all the approaches. Given the confusion matrix, which contains the number of samples classified as true positive (TP), false positive (FP), true negative (TN), and false negative (FN), the formulas for the ACC and MCC are as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \tag{17}$$

and

$$\mathrm{MCC} = \frac{\mathrm{TP}\cdot\mathrm{TN}-\mathrm{FP}\cdot\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}. \tag{18}$$

Higher ACC and MCC values indicate better prediction results.

Experimental settings. All experiments are implemented using Python modules Scikit-learn and Keras. The optimal parameters, including η (in Eq. (5)), λ (in Eq. (6)), and T (in Eq. (6)) are set by grid search. Specifically, we test our method under different hyperparameters, with η varying from 0.2 to 0.6 with a step of 0.05, λ varying from 10^{-2} to 10^{2} with a step of 0.5 in terms of its index, and T varying between 1 and 20 with a step of 1. Figs. 3 and 4 present the predicted results in terms of the evaluation metrics. From Fig. 3, we can see that the prediction accuracy is best when $T = 12$. This result supports the previous findings that the historical pattern is crucial for tracking stock movements [6,7]. Moreover, the ACC value is best when $\eta = 0.5$ and $\lambda = 10^{1/2}$, as shown in Fig. 4. Specifically, we present the results of parameters η and λ in terms of $T = 1$, $T = 6$, and $T = 12$ (the best). Thus, we set $\eta = 0.5$, $\lambda = 10^{1/2}$, and $T = 12$ throughout all the experiments. When training the classifier, we adopt the Adam optimizer [44] and set the learning rate to 0.001. In the attention neural network, the batch size is set as 8 with epoch 100, and the number of neural nodes is selected from $\{16, 32, 64, 128\}$. To prevent overfitting, dropout [45] with a rate of 0.5 is applied. Here, we report results for the parameters on the CSI 100 dataset. Similar parameter results can be found on the S&P 500 dataset. Note that, to ensure the robustness of classifier evaluation, we trained each classifier in our experiments 10 times with different initializations as suggested by [46]. The average of selecting runs in the testing set is reported to eliminate the fluctuations caused by random initializations.

4.1. Comparisons with state-of-the-art methods

In this section, we compare the proposed method with five state-of-the-art methods that consider multimodal market information for predictions. These are as follows.

- TeSIA [6] is a tensor-based learning approach that fuses firm-mode, event-mode, and sentiment-mode data to make stock movement predictions.
- HAN-SPL [47] addresses the influence of news on stock prediction by capturing sequential content dependency and diverse influence and applying the self-paced learning mechanism.
- CMT [7] fuses heterogeneous data into a tensor and captures the intrinsic relations among the events and the investors' sentiments with quantitative features and correlation matrices.
- MFN [43] learns the representation of each review by using a convolutional neural network (CNN) and integrating multiview textual features and extended knowledge.
- Multi-GCGRU [48] utilizes a graph convolutional network and gated recurrent units to incorporate the cross effect from related stocks to make predictions.

We compare our approach with the baselines on stock prediction groups using the same dataset. Table 1 presents details of the comparison of our experimental results. It shows that in terms of ACC and MCC metrics, the proposed method outperforms other machine learning models, with the best average ACC of 0.5628 and MCC of 0.1316, respectively, representing an improvement in the ACC of at least 1.1%. For the S&P 500 dataset, the proposed method achieves the best performance, with an average ACC of 0.5212 and MCC of 0.0415, representing an improvement in

3 https://www.eastmoney.com/
4 https://wrds-www.wharton.upenn.edu

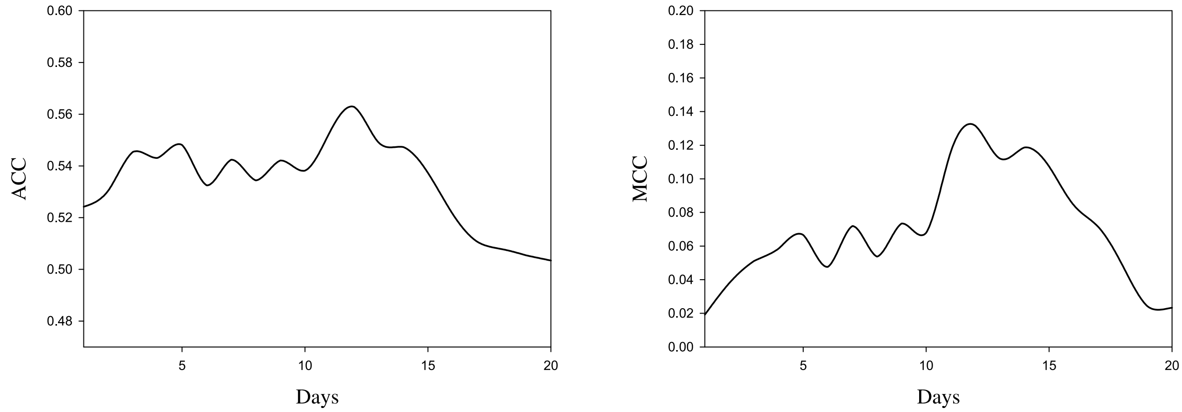
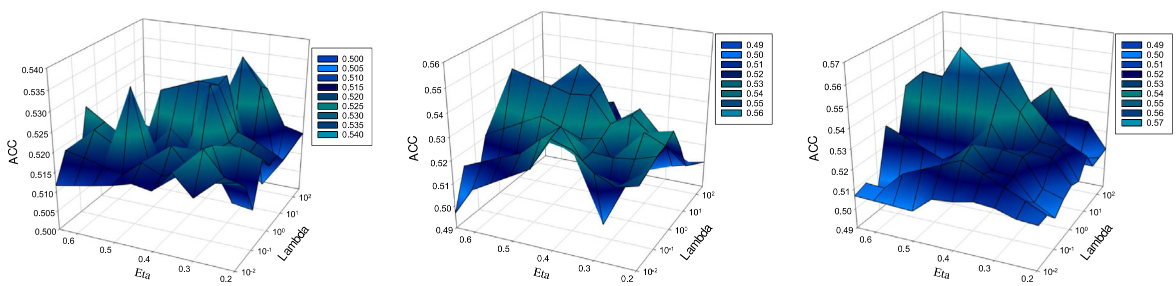


Fig. 3. ACC (left) and MCC (right) values for different days.



Parameter $T = 1$ Parameter $T = 6$ Parameter $T = 12$

Fig. 4. ACC values for different values of λ and η.

Table 1
Quantitative metrics of the prediction results of different methods.

Classifier	CSI 100		S&P 500	
	ACC	MCC	ACC	MCC
TeSIA [6]	0.5278	0.0551	0.5056	−0.0163
HAN-SPL [47]	0.5302	0.0617	0.5186	0.0102
CMT [7]	0.5437	0.0899	0.5199	0.0078
MFN [43]	0.5567	0.1094	0.5169	0.0401
Multi-GCGRU [48]	0.5349	0.0665	0.5172	0.0329
Ours	**0.5628**	**0.1316**	**0.5212**	**0.0415**

the ACC of at least 1.56%. Clearly, our framework is effective and meets the top level of the current research. These findings further prove the effectiveness of the tensor-based multimodality fusion mechanism with low-rank learning. In addition, the p-values for the t-tests are all less than the critical confidence value (0.05), indicating that the superior performance of the proposed approach is statistically significant.

4.2. Ablation study

Our framework consists of three dominant pipelines: the representation part, the fusion stage, and the classifier. In this section, we present the results of ablation experiments that are conducted to analyze these three elements of the proposed framework. Then, we further explore the scalability of the proposed framework. Specifically,

4.2.1. Effectiveness of tensor representation

Within the representation part, the employment of multimodal and multitemporal information plays a key role. Therefore, we remove the multimodal part or the multitemporal part, respectively. That is, we use only the trade data ($M = 1$) to construct the correlation matrix or consider only the multimodal data for one day ($T = 1$). In the first block of Table 2, we present the ACC and MCC values of the prediction results obtained by the degraded versions of our method mentioned previously. It can be seen that when our method lacks the multimodal structure, the ACC decreases by 0.0247 and the MCC drops by 0.0541. This result shows that the utilization of multimodalities is helpful for stock movement prediction. The ACC and MCC values decrease by 0.0386 and 0.1124, respectively, when the multitemporal information is not considered. This outcome shows that historical information in the market is important.

4.2.2. Effectiveness of fusion stage

As mentioned in Section 3.2, a fusion stage is proposed to further capture the essential structure of the multimodal and multitemporal data. To evaluate its effectiveness, we test the prediction performance by setting the representation result $\mathcal{C}$ as the training data instead of the fusion result $\mathcal{Z}$, namely, omitting the fusion stage. In the second block of Table 2, we can observe a sharp decline in both ACC and MCC (only 0.5177 and 0.0041, respectively), when the fusion stage of performing the optimized TRPCA model is omitted. This finding proves that the fusion stage of the proposed TRPCA is able to effectively capture the intrinsic structure of the multimodal and multitemporal data.

4.2.3. Effectiveness of the proposed classifier

In this section, we first present a series of experiments conducted to analyze the attention mechanism in the proposed classifier. Specifically, we compare it with two classic attention mechanisms (attention_dot and attention_general), which differ from ours in their score function [?]. In the third block of Table 2, we can observe there is no significant gap between classifiers with different attention mechanisms in terms of both ACC and MCC, although the best performance is achieved by our proposed

Table 2
Ablation comparison.

Element	Classifier	CSI		S&P	
		ACC	MCC	ACC	MCC
Tensor Representation	Ours w/o the multimodal structure	0.5381	0.0775	0.5160	0.0037
	Ours w/o the multitemporal structure	0.5242	0.0192	0.5051	−0.0637
Fusion Stage	Ours w/o the fusion stage	0.5177	0.0041	0.4984	−0.0784
Attention Mechanism	attention (*dot*)	0.5619	0.1301	0.5203	0.0394
	attention (*general*)	0.5602	0.1284	0.5182	0.0381
	Ours	**0.5628**	**0.1316**	**0.5212**	**0.0415**

Table 3
Stock movements prediction results of several different classifiers.

Classifier	CSI 100		S&P 500	
	ACC	MCC	ACC	MCC
BPNN	0.5077	0.0086	0.5096	−0.0123
CNN	0.5255	0.0469	0.5132	0.0272
LSTM	0.5288	0.0741	0.5077	0.0086
The proposed classifier	**0.5628**	**0.1316**	**0.5212**	**0.0415**
BPNN-v	0.5019	0.0049	0.5043	−0.0587
CNN-v	0.5184	0.0299	0.5060	−0.0355
LSTM-v	0.5248	0.0343	0.5097	−0.0394
The proposed classifier-v	0.5192	0.0133	0.5200	0.0175

Table 4
Stock movements prediction results with different input dimensions.

Classifier	CSI 100		S&P 500	
	ACC	MCC	ACC	MCC
Ours+media ($M = 3$)	0.5631	0.1335	0.5218	0.0427
Ours+tech ($M = 3$)	0.5636	0.1349	0.5224	0.0449
Ours+media & tech ($M = 4$)	0.5648	0.1367	0.5241	0.0472
Ours ($M = 2$)	0.5628	0.1316	0.5212	0.0415

method. This suggests that distinguishing market information at different periods via an attention mechanism can improve forecasting performance, but how this distinction (via different attention mechanisms) is achieved does not have a significant effect on our classifier. A good explanation is that the representations of market information are sufficient for the prediction task, which consolidates the effectiveness of the proposed framework, especially the TRPCA model for essential tensor learning.

In addition, from the perspective of the classifier models, we further examine the effectiveness of the proposed attention-based LSTM classifier and three selected classification techniques: the back-propagation neural network (BPNN), the CNN, and a plain LSTM classifier. Table 3 presents the quantitative metrics of the results for the different classifiers. We also train these classifiers with vectors, which are directly concatenated by the features introduced in Section 4. In Table 3, classifiers with the suffix "-v" denote that the training data are feature vectors.

From Table 3, we can see that the performance of the proposed attention-based LSTM trained using our fusion result is significantly superior to that of the others in terms of both the ACC and MCC metrics. We can also see that the performances of the classifiers trained with vectors are limited. This result reveals that the proposed tensor representation and fusion framework can efficiently fuse multimodal and multitemporal information, with the inner structure being well preserved. In addition, the performance of the proposed attention-based LSTM is significantly worse than that when trained using our fusion result. On average, "LSTM-v" is better than our attention-based LSTM. This interesting phenomenon shows that the attention mechanism is more suitable for tensor data generated by our tensor representation and fusion framework, being able to further identify useful information based on the obtained fused features. For the S&P 500 dataset, the proposed method again achieves the best performance, with improvements of at least 1.56% and 52.57% in terms of the ACC and MCC, respectively.

4.2.4. Scalability of the proposed framework

In addition, to further investigate the scalability of the multimodal design, we conduct a prediction task based on higher dimensional market information. Specifically, previous studies have demonstrated that technical factors and social media can shape stock movements [6]. We evaluated the multimodal framework of this study by incorporating technical factors and social media into our dataset. From the perspective of textual information, we incorporate the social media information to extend the dimension of the input (ours+media). From the perspective of numerical data, we incorporate the technical indicators to extend the dimension (ours+tech). These two methods extend the bimodal input to a trimodal input (i.e., $M = 3$). Moreover, we further consider both social media and technical indicators to evaluate the proposed framework ($M = 4$). Note that, similar to the news, textual media information is pre-processed by quantifying sentiment.

Table 4 indicates that the proposed method is a scalable framework for multimodal inputs and its predictive power can be enhanced by incorporating more useful information. It also can be observed that the performance improvement is somewhat limited. A good explanation is that fundamental data and financial news are the two main types of information that influence market volatility in multimodal market data, which is consistent with the previous findings of [9,46].

4.3. Investment simulation

To better evaluate our model, we conduct experiments to simulate stock investments using data from mid-September to December 2015. At the same time, we compare our method with the five state-of-the-art methods mentioned in Section 4.1 on real investments. In each investment, we set the initial capital to RMB 100,000 (approximately USD 14,390) and compare the cumulative daily returns on the basis of continuous investment, during which time the CSI 100 index rises by approximately 10%. In the simulation, we ignore transaction fees and sell and buy during each trading day. Stocks for investment options are selected based on our framework's ranking of probability predictions of the daily ups and downs of stocks. When the remaining funds are sufficient, we maximize the purchase of stocks that are at the top of the forecast for future ups and downs. Fig. 5 plots the income over time. Our method (the red line) achieves the best return of RMB 127,931 (p-values$<$ 0.05). The return rate of our method is 27.93% higher than those of MFN (18.34%), HAN-SPL (14.13%), CMT (17.50%), Multi-GCGRU (17.87%), and TeSIA (9.01%). From Fig. 5, we can also see that the performance of Multi-GCGRU, whose final return rate is the second best, fluctuates, whereas our method ranks first most of the time.

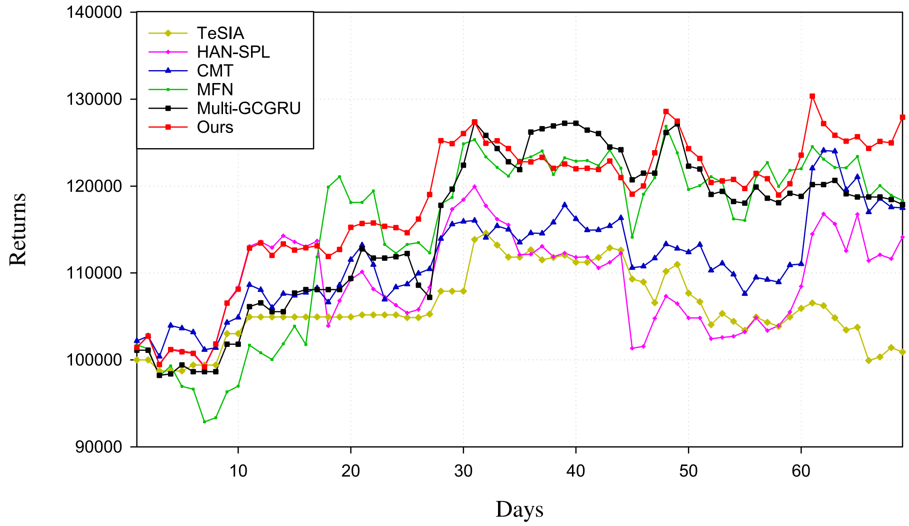


Fig. 5. Investment simulations.

5. Conclusions

This study has investigated stock movement prediction from multiple sources of market-related data. First, to exploit the inner correlation among different stocks in the market, a general tensor-based correlation representation strategy is proposed. In this way, the market-related data, which are always multimodal and multitemporal, can be employed in a unified tensor structure. Considering the homogeneity of the correlation among different companies with respect to different types of market-related data in a short time, we optimize the TRPCA model to further enhance the inner correlations within the tensor data. The essential structure of those multimodal and multitemporal data is believed to be organically fused in the low-rank output of the TRPCA model, whereas some outliers are rejected and stored in the sparse component. Then, we design an LSTM classifier with the self-attention mechanism and train it on the fusion result of the TRPCA model. Abundant numerical experiments have been conducted on the fundamental data and news data, revealing the superiority of the proposed framework. Moreover, investment simulations show that our method could achieve the best return rate (26.73%) for a full year among all the state-of-the-art methods. In the future, we could extend our method by unrolling the optimization algorithm into a deep neural network and training it in an end-to-end manner.

CRediT authorship contribution statement

Jun Wang: Conceptualization, Writing – original draft. **Yexun Hu:** Writing – review & editing. **Tai-Xiang Jiang:** Supervision, Methodology, Conceptualization, Writing – review & editing. **Jinghua Tan:** Methodology, Software, Data curation. **Qing Li:** Visualization, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC), under Grants 12001446, 12171072 71671141, 71873108, and 62072379, and the Natural Science Foundation of Sichuan, China (NSFSC), under Grants 2022NSFSC1798, 2022NSFSC1798, and 23NSFSC3330, the Jiaozi Institute of Financial Technology Innovation, Southwest University of Finance and Economics, under Grant CGZH20210204, and the Research Program of Science and Technology at Universities of Inner Mongolia Autonomous Region, under Grant 2021GG0164

References

[1] E.F. Fama, The behavior of stock market prices, J. Bus. 38 (1) (1965) 34–105.
[2] J.B. De Long, A. Shleifer, L.H. Summers, R.J. Waldmann, Noise trader risk in financial markets, J. Polit. Econ. 98 (4) (1990) 703–738.
[3] P.C. Tetlock, M. Saar-Tsechansky, S. Macskassy, More than words: Quantifying language to measure firms' fundamentals, J. Finance 63 (3) (2008) 1437–1467.
[4] R.P. Schumaker, H. Chen, A quantitative stock prediction system based on financial news, Inf. Process. Manage. 45 (5) (2009) 571–583.
[5] B. Weng, M.A. Ahmed, F.M. Megahed, Stock market one-day ahead movement prediction using disparate data sources, Expert Syst. Appl. 79 (2017) 153–163.
[6] Q. Li, Y. Chen, L.L. Jiang, P. Li, H. Chen, A tensor-based information framework for predicting the stock market, ACM Trans. Inf. Syst. (TOIS) 34 (2) (2016) 11.
[7] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, S.Y. Philip, Improving stock market prediction via heterogeneous information fusion, Knowl.-Based Syst. 143 (2018) 236–247.
[8] X. Zhang, A nonconvex relaxation approach to low-rank tensor completion, IEEE Trans. Neural Netw. Learn. Syst. 30 (6) (2018) 1659–1671.
[9] Q. Li, J. Tan, J. Wang, H.C. Chen, A multimodal event-driven LSTM model for stock prediction using online news, IEEE Trans. Knowl. Data Eng. (99) (2020) 1.
[10] C.W. Calomiris, H. Mamaysky, How news and its context drive risk and returns around the world, J. Financ. Econ. 133 (2) (2019) 299–336.
[11] U. Alia, D. Hirshleifer, Shared analyst coverage: Unifying momentum spillover effects, J. Financ. Econ. 136 (3) (2020) 649–675.
[12] W. Long, Z. Lu, L. Cui, Deep learning-based feature engineering for stock price movement prediction, Knowl.-Based Syst. 164 (JAN.15) (2019) 163–173.
[13] Q. Li, L. Jiang, P. Li, H. Chen, Tensor-based learning for predicting stock movements, in: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015, pp. 1784–1790.
[14] H. Wang, S. Lu, J. Zhao, Aggregating multiple types of complex data in stock market prediction: A model-independent framework, Knowl.-Based Syst. 164 (2019) 193–204.
[15] X. Ding, Y. Zhang, T. Liu, J. Duan, Deep learning for event-driven stock prediction, in: International Joint Conferences on Artificial Intelligence, IJCAI, 2015, pp. 2327–2333.

[16] T.-X. Jiang, L. Zhuang, T.-Z. Huang, X.-L. Zhao, J.M. Bioucas-Dias, Adaptive hyperspectral mixed noise removal, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13.
[17] L. Zhuang, M.K. Ng, X. Fu, J.M. Bioucas-Dias, Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–15.
[18] T.-X. Jiang, M.K. Ng, X.-L. Zhao, T.-Z. Huang, Framelet representation of tensor nuclear norm for third-order tensor completion, IEEE Trans. Image Process. 29 (2020) 7233–7244.
[19] T.-X. Jiang, M.K. Ng, J. Pan, G.-J. Song, Nonnegative low rank tensor approximations with multidimensional image applications, Numer. Math. (2022) http://dx.doi.org/10.1007/s00211-022-01328-6.
[20] T.G. Kolda, B.W. Bader, J.P. Kenny, Higher-order web link analysis using multilinear algebra, in: Fifth IEEE International Conference on Data Mining, ICDM'05, IEEE, 2005, pp. 8–pp.
[21] P. Symeonidis, A. Zioupos, Matrix and tensor decomposition in recommender systems, in: RecSys, 2016, pp. 429–430.
[22] N. Kreimer, M.D. Sacchi, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation, Geophysics 77 (3) (2012) 113–122.
[23] C. Jia, G. Zhong, Y. Fu, Low-rank tensor learning with discriminant analysis for action classification and image recovery, in: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014, pp. 1228–1234.
[24] C. Zhang, H. Fu, S. Liu, G. Liu, X. Cao, Low-rank tensor constrained multiview subspace clustering, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1582–1590.
[25] T. Wang, X. Xu, Y. Yang, A. Hanjalic, H.T. Shen, J. Song, Matching images and text with multi-modal tensor fusion and re-ranking, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 12–20.
[26] J.D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition, Psychometrika 35 (3) (1970) 283–319.
[27] J. Douglas Carroll, S. Pruzansky, J.B. Kruskal, CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters, Psychometrika 45 (1) (1980) 3–24.
[28] L.R. Tucker, Some mathematical notes on three-mode factor analysis, Psychometrika 31 (3) (1966) 279–311.
[29] I.V. Oseledets, D. Savostianov, E.E. Tyrtyshnikov, Tucker dimensionality reduction of three-dimensional arrays in linear time, SIAM J. Matrix Anal. Appl. 30 (3) (2008) 939–956.
[30] K. Braman, Third-order tensors as linear operators on a space of matrices, Linear Algebra Appl. 433 (7) (2010) 1241–1253.
[31] M.E. Kilmer, C.D. Martin, Factorization strategies for third-order tensors, Linear Algebra Appl. 435 (3) (2011) 641–658.
[32] M.E. Kilmer, K. Braman, N. Hao, R.C. Hoover, Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging, SIAM J. Matrix Anal. Appl. 34 (1) (2013) 148–172.
[33] H.A. Kiers, Towards a standardized notation and terminology in multiway analysis, J. Chemometrics 14 (3) (2000) 105–122.
[34] I.V. Oseledets, Tensor-train decomposition, SIAM J. Sci. Comput. 33 (5) (2011) 2295–2317.
[35] C.J. Hillar, H. Lim, Most tensor problems are NP-hard, J. ACM 60 (6) (2013) 1–39.
[36] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, S. Yan, Tensor robust principal component analysis with a new tensor nuclear norm, IEEE Trans. Pattern Anal. Mach. Intell. 42 (4) (2019) 925–938.
[37] Z. Zhang, G. Ely, S. Aeron, N. Hao, M. Kilmer, Novel methods for multilinear data completion and de-noising based on tensor-SVD, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2014, pp. 3842–3849.
[38] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[39] J. Li, B. Yang, H. Li, Y. Wang, C. Qi, Y. Liu, DTDR–ALSTM: Extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models, Knowl.-Based Syst. 211 (2021) 106508.
[40] D. Bahdanau, K.H. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, in: The 3rd International Conference on Learning Representations, ICLR 2015, 2015.
[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, pp. 5998–6008.
[42] J. Duan, Y. Zhang, X. Ding, C.-Y. Chang, T. Liu, Learning target-specific representations of financial news documents for cumulative abnormal return prediction, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 2823–2833.
[43] H. Wang, T. Wang, Y. Li, Incorporating expert-based investment opinion signals in stock prediction: A deep learning framework, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2020, pp. 971–978.
[44] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations, 2015.
[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A aimple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[46] F. Feng, X. He, X. Wang, C. Luo, Y. Liu, T.-S. Chua, Temporal relational ranking for stock prediction, ACM Trans. Inf. Syst. 2 (37) (2019) 1–30.
[47] Z. Hu, W. Liu, J. Bian, X. Liu, T.Y. Liu, Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction, in: WSDM' 18, 2017.
[48] J. Ye, J. Zhao, K. Ye, C. Xu, Multi-graph convolutional network for relationship-driven stock movement prediction, Papers (2020).