A Multimodal Event-driven LSTM Model for Stock Prediction Using Online News

Journal:	*Transactions on Knowledge and Data Engineering*
Manuscript ID	TKDE-2019-05-0439.R1
Manuscript Type:	Regular
Keywords:	Stock Prediction, Tensor, Multimodality, Deep Learning, LSTM

A Multimodal Event-driven LSTM Model for Stock Prediction Using Online News

Qing Li, *Member, IEEE,* Jinghua Tan, *Member, IEEE,* Jun Wang, Hsinchun Chen, *Fellow, IEEE*

Abstract—In finance, it is believed that market information, namely, fundamentals and news information, affects stock movements. Such media-aware stock movements essentially comprise a multimodal problem. Two unique challenges arise in processing these multimodal data. First, information from one data mode will interact with information from other data modes. A common strategy is to concatenate various data modes into one compound vector; however, this strategy ignores the interactions among different modes. The second challenge is the heterogeneity of the data in terms of sampling time. Specifically, fundamental data consist of continuous values sampled at fixed time intervals, whereas news information emerges randomly. This heterogeneity can cause valuable information to be partially missing or can distort the feature spaces. In addition, the study of media-aware stock movements in previous work has focused on the one-to-one problem, in which it is assumed that news affects only the performance of the stocks mentioned in the reports. However, news articles also impact related stocks and cause stock co-movements. In this article, we propose a tensor-based event-driven LSTM model to address these challenges. Experiments performed on the China securities market demonstrate the superiority of the proposed approach over state-of-the-art algorithms, including AZFinText, eMAQT, and TeSIA.

Index Terms—Stock Prediction, Tensor, Multimodality, Deep Learning, LSTM.

1 Introduction

A company's stock price reflects investor perception of its ability to earn and grow profits in the future. The traditional efficient market hypothesis (EMH) states that the price of a stock is always driven by 'unemotional' investors [1, 2]. New information related to markets will change investors' expectations about the markets and cause stock prices to move [3]. On the other hand, in behavioral finance studies, stock movements are attributed to investors' cognitive and emotional biases [4]. Although the two theories are based on different views regarding how information shapes stock movements, both agree that the volatility of stock markets stems from the release, dissemination and absorption of information [5].

In previous studies, scholars have found that stock movements are affected by various sources of information, including transaction data, news, social media, and search behavior [6, 7, 8]. Some researchers have taken a further step by examining the joint effects of various types of information, which has proven helpful in capturing stock movements [9, 10, 11]. Essentially, stock markets are affected by multiple information sources, which can be roughly categorized into two subgroups: fundamentals (e.g., turnover, opening prices, and trading volumes) and financial news [12]. Thus, the problem of modeling stock movements is essentially a multimodal learning problem.

The first challenge lies in identifying the joint effects of fundamental data and news information on stock markets. The traditional strategy is to concatenate these information into a compound vector and utilize various learning models, including support vector machines (SVMs), decision trees (DTs), and artificial neural networks (ANNs), to make predictions [13, 14, 15]. However, such vector-based models may ignore the inherent links among multiple sources of information and thus fail to capture their interconnections. To overcome this challenge, some scholars have modeled multidimensional information with tensors to achieve better performance [16, 17].

Another important issue facing multimodal models is the heterogeneity of the sampling times among different modes. For stock markets in particular, the fundamental data are characterized by continuous values sampled at equal time intervals (i.e., one day). By contrast, news information consists of discrete values sampled at nonequal time intervals because of the randomness of the occurrence of news events. A good example is presented in Figure 1. This figure shows news articles about stock "000001" published between January 1, 2015, and April 1, 2015. The occurrence of these news events is irregularly distributed, with varying intervals ranging from days to weeks or even months, while the fundamental information is represented by daily continuous data. The problem of how to fuse these two types of data for solving a supervised learning problem has yet to be explored.

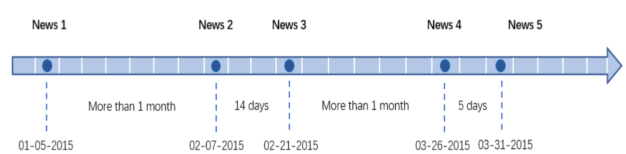


Fig. 1: An example of news events between Jan 1, 2015, and April 1, 2015, with varying time intervals.

- *Qing Li, Jinghua Tan and Jun Wang are with the Financial Intelligence & Financial Engineering Key Lab of Sichuan Province and the School of Economic Information Engineering, Southwestern University of Finance and Economics, China.*
  Email address: liq_t@swufe.edu.cn (Qing Li)
- *Hsinchun Chen is with the Department of Management Information Systems, University of Arizona, USA and Tsinghua University, China.*

In previous studies, this problem has typically been solved by using only a portion of the available data; that is, only data sampled at the times of news events are retained for further analysis. For instance, Schumaker and Chen utilized transaction data within 20 minutes following the release of breaking news to study media-aware stock movements [18]. There were two data dimensions in this study: the continuous transaction data and the discrete news article data. Only part of the transaction data, i.e., the data that coincided with news publications, were utilized, while other transaction records without corresponding news reports were discarded. One alternative solution is to obtain a sparse feature space by filling the missing values in the news mode with Null. However, such sparsity in the news dimension will distort the entire feature space.

In addition, previous studies on media-aware stock movements have focused on only the one-to-one problem, in which news articles about a company are assumed to affect only that company, without considering the indirect effects on related companies. However, the fluctuation of one company is also affected by its related companies. For instance, a news report on the alternative energy supply on November $14th$, 2017, applied downward pressure on the PetroChina (601857) stock, resulting in a decrease of 1.35%. By contrast, due to its savings on transportation costs, Air China (601111) saw its stock increase by 15.35%. Incorporating such correlations among relevant companies to quantify media-aware stock movements is of great interest.

To address the above challenges, we propose a multimodal event-driven long short-term memory (LSTM) model with several unique contributions, as follows.

- We first represent the complicated market information space with tensors to preserve the interconnections among different information modalities.
- We then propose an event-driven LSTM model to address the heterogeneity of the sampling times in different modes. This is achieved by controlling the memory in the neural network so as to fuse the continuous data sampled at equal intervals (fundamental data) with the discrete values sampled at nonequal intervals (news).
- We also consider the indirect influence of related companies on media-aware stock movements by constructing a media-based enterprise network to reshape the market information space represented by tensors.
- Experiments performed on one full year of data on the China securities market demonstrate the superiority of the proposed approach over state-of-the-art algorithms, including AZFinText, eMAQT, and TeSIA. Relative to these algorithms, the proposed approach achieves a performance improvement of at least 22.8%.

2 Related work

In this section, we review the relevant literature from three perspectives: the influence of information on stock volatility, stock comovements and the approaches for quantifying such media-aware movements.

2.1 Information and Stock Volatility

The price of a stock reflects investors' expectations regarding a company's future cash flows. Investors may change their expectations as they receive new information, resulting in stock fluctuations. Stock market information can be roughly categorized into three subgroups: fundamental data, media information and a combination of the two [12].

- **Fundamental information:** A number of studies in traditional finance have examined the effects of fundamental information. Haugen and Baker showed that cash flows can provide additional information content for better understanding stock markets [22]. Fama and French found that a stock's performance is determined mainly by three risk factors: the overall market, the firm size, and the book-to-market equity ratio (BE/ME) [23]. Jegadeesh and Titman observed that stocks with higher returns in the previous 12 months tended to have higher future returns [24].
- **Media information:** The pilot research on media-aware stock movements can be traced back to work on the influence of financial reports on stocks [25]. Later, researchers observed the influence of online media on stock fluctuations [15, 21]. In particular, investors' decisions can be influenced by the opinions of others as expressed via online media, which may result in herd behavior in investment. For example, Schumaker and Chen experimented with several textual news representation approaches to study the effects of breaking news on stock movements [10]. Bollen, Mao and Zeng found that the collective mood states derived from 10 million tweets were correlated with the index of the Dow Jones Industrial Average (DJIA) using a self-organizing fuzzy neural network (SOFNN) model [14]. These works have prompted the birth of media-aware hedge funds, including Derwent Capital Markets, DCM Capital and Cayman Atlantic.
- **Combination:** Many studies have shown that both fundamental and media information can shape stock movements. However, difficulties arise in modeling these two types of information [26, 27]. One common strategy is to concatenate these information into a compound vector, thus treating each value as an independent variable. For example, Tetlock measured the positive (negative) sentiment polarity of an article and applied a linear regression model to capture stock returns [12]. Mittermayer and Knolmayer represented market information spaces with vectors and applied SVM and KNN models to study the impacts of news on stock markets [15]. However, such vector-based methods dilute or even ignore the intrinsic associations among various information sources. Alternatively, researchers have found that tensor representations are able to capture the interconnections among various modes of market information, thus providing a better understanding of stock movements. Li et al. was the first to apply tensor theory to model the complicated market information space and show that such a representation is able to capture the joint effects of different information sources [16].

TABLE 1: Representative research on the influence exerted by news articles on stock markets.

Category	Reference	Model	Focus		Experiment		
			Information source	Scale	Response	Predictor	Period
Statistical models and regression models	[7]	Statistical model	Wikipedia	Week	Index	Number of page views	12/10/2007-04/30/2012
	[8]	Mutual information	Twitter	Hour	Price	Message volume	11/12/2012-12/03/2013
	[12]	Linear model	DJNS, WSJ	Day	Return	Number of emotion words	1980-2004
Classical ML-based models	[15]	KNN, SVM	PR Newswire	Minute	Stock trend	News content	12/06/1997-03/06/1997
	[19]	KNN	Yahoo	Day	Index	News content	04/01/2002-12/31/2002
Deep learning models	[20]	Neural network	Reuters, Bloomberg	Week	Return, volatility	Sentiment	01/2003-02/2014
	[21]	LSTM	Microblogs	Hour	Stock trend	Social media content	2015

A tensor representation provides a promising solution for retaining the interactions among different information modes for multimodal learning problems. However, the previous work [16] relied on a framework based on support tensor regression, which requires iterative estimation of the parameters for each mode until the objective function converges. Such iterative calculations are time consuming and constitute a bottleneck for the parallelization of processing. Moreover, most machine learning algorithms can utilize only vectorized input features. It is quite challenging to apply a tensor representation in combination with machine learning approaches, especially deep learning networks.

2.2 Stock Comovements

As discussed in regard to the above examples of PetroChina and Air China, it is of great interest to consider the influence of related companies on stock movements. The challenge here lies in how to identify related companies. Related to this challenge is the work on stock comovements.

There are two mainstream methods of studying stock comovements: from the perspective of fundamentals and from the perspective of investor behavior. Traditional financial researchers have attributed stock comovments to the fundamental characteristics of the listed companies [28, 29, 30]. For instance, Pindyck and Rotemberg discovered that company size and the degree of institutional ownership influence stock comovements [28]. Preis et al. reported that stock correlations are reflected by normalized DJIA index returns on various time scales [29]. Aghabozorgi and Teh attributed stock comovements to historical transaction prices [30]. In contrast, in modern behavioral finance, it is believed that irrational behaviors of investors cause the comovements of related stocks. For example, Rashes found a highly abnormal positive correlation between two companies with similar names but nothing else in common, caused by the irrational feelings of investors [31].

With the advancement of the Web 2.0 era, the influence of online information on stocks has become salient [9]. The influence of online media involves two aspects: fundamentals and emotions. Web media enrich investors' knowledge by conveying a more comprehensive view of a firm's financial standing. In addition, Web media provide a platform for expressing the options of experts and the public moods of investors, which inevitably affect investor behavior and can even elicit herd behavior [9]. Essentially, Web media act as a sort of mirror reflecting the fundamentals of listed companies and affecting investor behavior to some degree. In this study, we build an innovative media-based enterprise network to identify related companies in terms of their media performance.

2.3 Stock Analysis Models

Once fundamental data and news information have been obtained in a machine-friendly form, various types of analysis models can be applied to study media-aware stock movements. There are three mainstream classes of such models: statistical models (originating from statistics), regression models (originating from econometrics) and machine learning models (originating from computer science). Table 1 summarizes the related work in terms of these classes of models.

Statistical models emphasize the correlations between a single feature and stock markets [7, 8]. For example, Moat et al. applied the Wilcoxon test to identify the linkage between a company's browsing frequency on Wikipedia and its stock fluctuations. Econometric models focus on the causal relationships between specific features and market movements [12, 32, 33]. For example, Huang et al. applied logistic regression models and found that abnormal optimism in a company's earnings report exerted a drag on its stock performance [34]. However, both statistical models and econometric models often have difficulty preserving the interconnections among multiple data sources and thus fail to capture their joint effects on stock performance. Thus, computer scientists have taken the further step of utilizing machine learning algorithms to capture such complex nonlinear relationships.

The problem of modeling media-aware stock movements is essentially a binary classification problem. Given the ability to take high-dimensional data as input, many machine learning algorithms, including SVMs, Bayesian classifiers and DT methods, have been applied to solve this problem [13, 15, 35]. For example, early research can be traced back to the work of Wuthrich et al., who forecast the daily trends of five major stock market indexes using a neural network and the KNN algorithm [19]. Later, Schumaker and Chen estimated a discrete stock price 20 minutes after

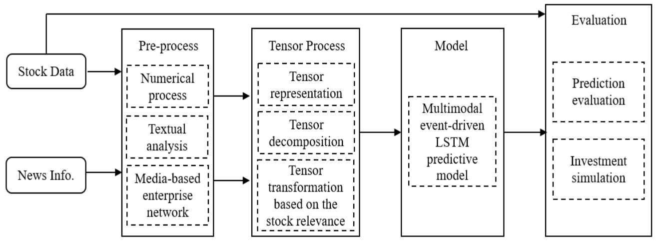


Fig. 2: System framework.

the release of a related news article using support vector regression (SVR) [18].

With the great success of deep learning in various fields, including text processing [36], image recognition [37], and speech recognition [38], some researchers have begun to explore the power of deep learning for capturing media-aware stock movements [39]. For example, Ding et al. proposed a deep learning method to model both the short-term and long-term influences on stock price movements and found that the performance of a deep neural network (DNN) was better than that of an SVM [39]. Huang et al. applied a convolutional neural network (CNN) to explore the impact of public sentiment, as extracted from tweets, on stock markets [40]. Inspired by the application of recurrent neural networks (RNNs) to time-series problems, LSTM models have been widely applied to study media-aware stock movements [21, 41]. However, these approaches simplify the market information space by adopting a vector representation, which ignores the interconnections among different information modes. In addition, the standard LSTM technique fails to address the heterogeneity of the sampling times among different market information modes. In this article, we model the market information with tensors and apply an event-driven mechanism to capture the interconnections and balance the heterogeneity of the sampling times between the different information modes [16].

3 Model Architecture

Stock markets are influenced by various information sources, including fundamental data and media information. A common strategy in previous studies has been to concatenate information from these heterogeneous data sources into a compound vector. However, these vector-based models treat different information sources as independent features, ignoring the inherent links between them and thus failing to properly capture stock movements [16]. In addition, fundamental data are continuous values sampled at equal time intervals, whereas media information emerges randomly. This heterogeneity results in valuable information being partially missing. Therefore, we use tensors to represent market information to preserve the multifaceted and interrelated nature of the data. On this basis, a multimodal event-driven LSTM model is proposed to capture the nonlinear relations between market information and stock movements. Figure 2 shows an overview of the proposed approach.

3.1 Tensor Representation

A tensor is a mathematical representation of a multidimensional array. Specifically, an N-way or N^{th}-order tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system. Essentially, a first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three or higher are called higher-order tensors. Figure 3 illustrates an example of a second-order tensor sequence for one stock. Additional details on tensor algebra can be found in [42].

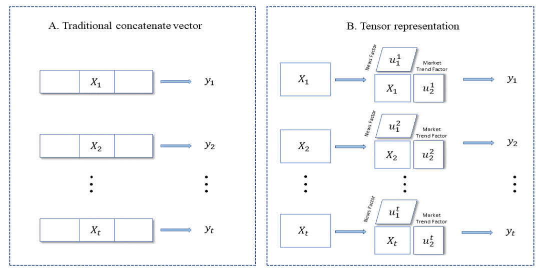


Fig. 3: Market information represented by a second-order tensor sequence. The tensor representation is used to reinforce the intrinsic links among multiple information sources.

In this study, stock market information is categorized into two subgroups, namely, fundamental data and media information, as described below.

$Fundamental\ data$: The price of a stock is a reflection of a firm's intrinsic value. Eight firm attributes are selected to capture the future business value of a firm, and each attribute has been shown to have some degree of predictive value [23, 43]. These attributes are the following: highest price, lowest price, opening price, closing price, turnover, trading volume, P/B and P/E ratio. More detailed explanation can be found in Table 2.

TABLE 2: Definition of the stock attributes.

Predictor	Explanations
Opening/ Closing price	The first (final) price at which a security is traded on a given trading day.
Highest/ Lowest price	The highest (lowest) price at which a security is traded on a given trading day.
Volume of trade	The total quantity of shares or contracts traded for a specific security.
Turnover	Turnover calculates how quickly a firm collects cash from accounts receivable or how fast the firm sells its inventory.
P/E ratio	Price-Earnings Ratio is for valuing a company that measures its current share price relative to its per-share earnings. It can be calculated as: Market Value per Share / Earning per Share.
P/B ratio	Price-to-Book Ratio is used to compare a stock's market value to its book value. It can be calculated as: Market Price per Share/Book Value per Share.

$Media\ data$: In modern behavioral finance, it is believed that investors are irrational, tending to be influenced by experts' opinions as expressed in the media. To capture media sentiment, we extract the following characteristics: positive media sentiment (P_t^+), negative media sentiment (P_t^-) and media sentiment divergence (D_t). These characteristics are calculated as follows:

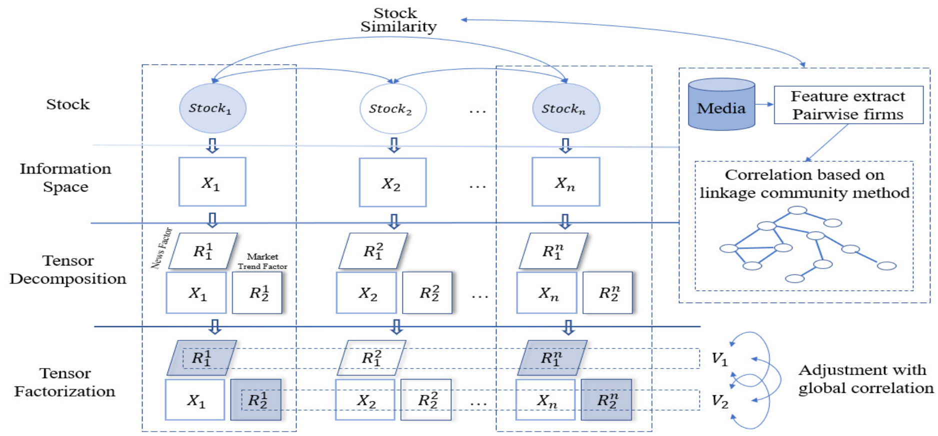


Fig. 4: Illustration of tensor transformations.

$$P_t^+ = \frac{N_t^+}{N_t^+ + N_t^-}, P_t^- = \frac{N_t^-}{N_t^+ + N_t^-}, D_t = \frac{N_t^+ - N_t^-}{N_t^+ + N_t^-}, \tag{1}$$

where N_t^+ (N_t^-) is the sum of the frequency of each positive (negative) sentiment word found in the media on the t^{th} day. D_t denotes the sentiment divergence on the t^{th} day. Previous studies have relied on a general emotion word dictionary to capture media sentiment. However, 73.8% of the negative sentiment words in this general sentiment dictionary no longer express negative emotional meanings in the financial field [9]. For instance, the word "bear" originally referred to an ursine animal but indicates poor earnings returns in the financial domain, e.g., "a bear stock". Therefore, we resort to a finance-oriented sentiment dictionary created in our previous study [16].

After obtaining fundamental information and media sentiment data, we construct a second-order tensor $X_t \in R^{I_1 \times I_2}$ to represent the market information at time t. The variables I_1 and I_2 represent the numbers of features in the fundamental data and media data, respectively. In this way, the interconnections among multiple sources of information can be preserved. The significance of the elements a_{i_1,i_2} of the tensor X_t is defined as follows:

- $a_{i_1,1}$, $1 < i_1 \leq I_1$, denotes the value of the i_1^{th} fundamental information feature.
- a_{2,i_2}, $1 < i_2 \leq I_2$, denotes the value of the i_2^{th} media sentiment information feature.
- All other elements are initially set to zero.

Unlike in traditional vector-based methods, this second-order tensor is able to capture the correlations characterizing market information in complementary subspaces.

3.2 Tensor Decomposition and Reconstruction

In this study, we introduce a unique tensor framework to allow the intrinsic connections between two different information sources to be identified from the geometric structure of the tensor X. Such identification is achieved through tensor transformations, namely, Tucker decomposition and tensor reconstruction. Tucker decomposition is applied to decompose the tensor X into $C \times_1 R_1 \times_2 R_2$ [42]. Here, each factor matrix R_k $(k = 1, 2)$ describes one distinct facet of the information space of the stock market (i.e., fundamental information and media information), and the core tensor C reflects the strength of the relations between these two facets. Thus, the decomposition captures the intrinsic associations and interactions within the tensor X.

After Tucker decomposition, the R_k are further adjusted to preserve the stock connections based on a stock correlation matrix, which is constructed on the basis of stock comovements. The intuition here is that if two stocks are highly correlated, then news articles about one stock are likely to effect a similar shock to the other stock.

Figure 4 details the tensor transformation process. For this purpose, we minimize the following Lagrangian objective function to obtain correction factors V_k $(k = 1, 2)$ with which to adjust the original factor matrix sequences $R_k^i|_{i=1}^N$:

$$\min_{V_{k,k=1,2}} L(V_k) = \frac{\lambda}{2} \sum_{i=1}^{N} \left\| X_t^i - C \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i) \right\|^2$$

$$+ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} \left\| V_1^T R_1^i - V_1^T R_1^j \right\|^2 s_{i,j}$$

$$+ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=i}^{N} \left\| V_2^T R_2^i - V_2^T R_2^j \right\|^2 s_{i,j}. \tag{2}$$

Here, $\left\| X - C \times_1 (V_1^T R) \times_2 (V_2^T R) \right\|^2$ is used as a normalization constraint to avoid overfitting and to control the adjusted tensor decomposition to be close to the real values, whereas $\left\| V_k^T R_k^i - V_k^T R_k^j \right\|^2$ serves to correct R_k^i by V_k to minimize the differences among stocks with higher correlations $s_{i,j}$. The purpose of the adjustment factor $s_{i,j}$ is to minimize the differences in the values of each mode for related stocks. The method of calculating the $s_{i,j}$ is described in Section 3.3.

To solve this Lagrangian function L and optimize the objective function, we apply an iterative algorithm to update the entries in V via gradient descent [27]. The gradient for each variable is derived as follows:

$$\nabla_{v_1} L = \lambda \sum_{i=1}^{N} (C \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i) R_1^i) + (D_{R_1} - S_{R_1}) V_1$$
$$\nabla_{v_2} L = \lambda \sum_{i=1}^{N} (C \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i) R_2^i) + (D_{R_2} - S_{R_2}) V_2. \tag{3}$$

Here, $D_{R_k} = \sum_{i=1}^{N} (R_k^i R_k^{iT}) d_{i,j}$, where $d_{i,i}$ are the diagonal entries and are column sums of the correlation matrix $S(i.e., d_{i,i} = \sum_{m=1}^{N} s_{m,i})$, while S_{R_k} is calculated as $\sum_{i=1}^{N} \sum_{j=i}^{N} (R_k^i R_k^{iT}) s_{i,j}$. The details of the partial derivative can be found in [16]. The iterative procedure for updating V_1 and V_2 is performed until the objective function converges. Thus, the reconstructed tensor is defined as $\widetilde{X} = C \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i)$. The algorithm for the learning process is detailed in Algorithm 1.

Algorithm 1 Iterative machine learning approach for tensor transformation

Input: The tensor stream X, the global correlation matrix S, and the error threshold ε .
Output: The mapped tensor stream $\widetilde{X} = C^i \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i)$

From $i = 1$ to N (N is the total number of stocks)
  Decompose the tensor X^i into $C^i \times_1 R_1^i \times_2 R_2^i$
End
Set η as the step size for gradient descent
While $(Loss^n - Loss^{n-1} > \varepsilon)$
  Get $\nabla_{v_1} L$, $\nabla_{v_2} L$
  $v_1^{n+1} = v_1^n - \eta \nabla_{v_1} L$
  $v_2^{n+1} = v_2^n - \eta \nabla_{v_2} L$
  $n = n + 1$
End while
From $i = 1$ to N
  $\widetilde{X} = C \times_1 (V_1^T R_1^i) \times_2 (V_2^T R_2^i)$
End

3.3 Stock Relatedness

Stock movements are influenced by fluctuations of related stocks. There are two ways to define the concept of 'related' stocks [44]. In traditional finance, related stocks are determined by their fundamental characteristics [28, 29]. From this perspective, the correlations $s_{i,j}$ in Equation (2) can be calculated as follows:

$$s_{i,j} = \frac{E((x_i - u_{x_i})(y_j - u_{y_j}))}{\sigma_{x_i} \sigma_{y_j}}, \tag{4}$$

where the x_i are the fundamental features of firm i and the y_j are the fundamental features of firm j. Essentially, the $s_{i,j}$ are the Pearson correlation coefficients applied to calculate the correlations between the two firms. u is the mean value, and σ is the standard error.

In modern behavioral finance, stock comovements are considered to be affected by the collective opinions of irrational investors [31]. To some extent, financial news articles provide summaries of both firm fundamentals and investor opinions. In other words, news articles enrich investors' knowledge by conveying a more comprehensive view of a firm's financial standing than is provided by a firm's price alone. The optimism and pessimism characterizing news articles may affect the emotions of irrational investors.

In this study, we construct a media-based enterprise network to identify the stocks related to a target firm. In this network, each node represents a listed firm, and an edge between two nodes represents the news co-exposure of the two corresponding firms. Specifically, if two firms are mentioned in the same news articles, there is a link between them. The edge between two firms is weighted by the total number of news articles mentioning both firms. Note that we disregard news articles mentioning more than five firms consecutively because such news items usually do not convey useful information [27].

In this enterprise network, we can simply treat firms with direct connections to the target firm i as related firms. However, this approach ignores the transitive effect. Specifically, as shown in Figure 5, a news article related to firm A may affect firms B, C and D since they are directly linked to firm A. Moreover, firm E could be affected as well. This is because the influence of firm A could be passed through firms B and C to firm E. Therefore, we take the further step of adopting the community linkage method to bridge relevant firms without direct linkages [45].

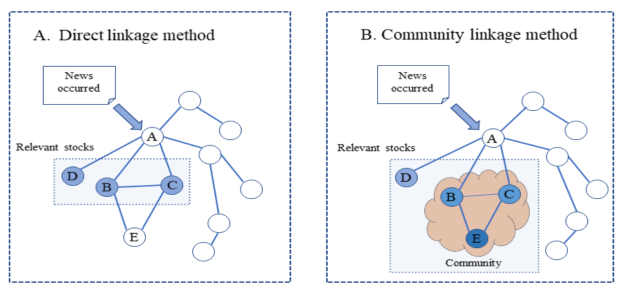


Fig. 5: Simple pairwise correlations and the community linkage method.

In particular, we first calculate the relationship between firms i and j as follows:

$$\overline{s}_{i,j} = \frac{\sum(N_i \bigcap N_j)}{\sum(N_i \bigcup N_j)}, \tag{5}$$

where N_i is the sum of the correlations of Node i and N_j is the sum of the correlations of Node j. A larger $\overline{s}_{i,j}$ indicates a higher correlation. The final relation matrix $S \in R^{N \times N}$ is a Boolean matrix, and each of its entries $s_{i,j}$ is defined as follows:

$$s_{i,j} = \begin{cases} 1 & \text{if } i \leq j \text{ and } \bar{s}_{i,j} \geq \theta \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where θ is a threshold. Therefore, the related firms are the firms with an entry value of 1 with respect to the target firm. This information is utilized in Equation (2) to bridge and reinforce the interactions among related firms.

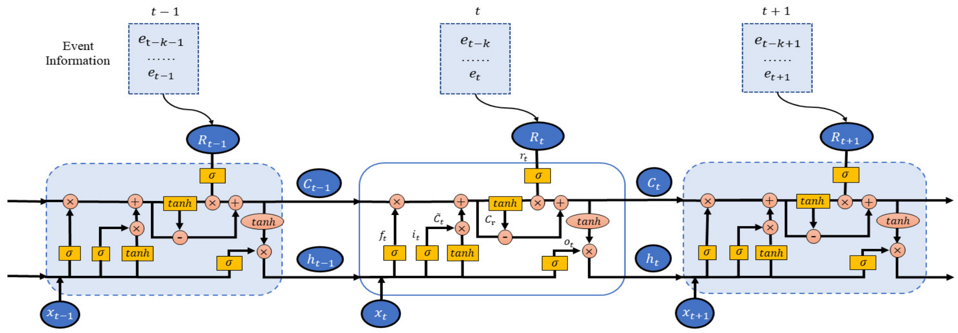


Fig. 6: Illustration of the proposed event-driven long short-term memory unit and its application for analyzing stock market information. At the t^{th} day, it considers the event effect of the previous k days.

3.4 Multimodal Event-driven LSTM Model

A long short-term memory (LSTM) model is a variant of a recurrent neural network (RNN) that is able to handle long-term dependencies by means of a gate mechanism [46]. Such a model is designed to handle time sequence data collected at equal intervals, such as daily transaction data [21]. However, predicting media-driven stock movements is essentially a multimodal problem, in which each mode has unique characteristics. Specifically, the information space for stock markets consists of both fundamental information and media information. The fundamental data are daily transaction records sampled at equal intervals. In contrast, the data for the media mode consist of discrete values sampled at nonequal time intervals due to the random time distribution of news releases. This randomness can lead to failure of the long-term dependency mechanism of an LSTM model [47]. Specifically, if two similar news articles are released with a sufficiently large time gap, the LSTM model may forget the knowledge learned from the first news article before processing the information from the second. To solve this problem, we propose a novel event-driven LSTM model by extending an LSTM model to include an event-driven memory mechanism. In addition, we adopt a tensor representation to capture the interactions among different modes of the multimodal data.

3.4.1 *Event-driven LSTM Model*

Market information takes the form of multimodal data with a continuous fundamentals mode and a discrete news mode. To solve the stock movements prediction problem given data collected at nonuniformly distributed time intervals, we propose an event-driven LSTM model, in which a triggering strategy is applied to reinforce the event-based information obtained in previous stages. Figure 6 shows the details of the proposed event-driven LSTM model.

As seen in Figure 6, all market information at time t is represented by a tensor X_t. The event information E_t is represented by a vector $< e_0, e_1, \ldots, e_t >$, where e_t is the total number of news articles at time t. There are two information flows recording the learned knowledge in the network. Specifically, the cell memory C_t records the event-based knowledge learned from previous stages, and the output H_t records the patterns learned from previous market information at time t. A forget gate F_t is utilized to control how much information should be retained or forgotten from the cell memory C_{t-1} at time t. In other words, F_t allows information in C_{t-1} that is useless with respect to H_{t-1} and X_t to be discarded. An input gate I_t is used to control how much current information should be absorbed into the event knowledge flow C. In this structure, the temporary memory $\widetilde{C}_t$ stores the knowledge learned from both the current market information X_t and the information H_{t-1} from the previous stage via the mapping function in the neural network. Therefore, $\hat{C}_t$ is able to capture the valuable rules and patterns hidden in both the previous and current time periods. $\widetilde{C}_t$, F_t, I_t and $\hat{C}_t$ are calculated as follows:

$$\widetilde{C}_t = tanh(W_c * X_t + U_c * H_{t-1} + V_c * E_t + B_c) \quad (7)$$

$$F_t = \sigma(W_f * X_t + U_f * H_{t-1} + V_f * E_t + B_f) \quad (8)$$

$$I_t = \sigma(W_i * X_t + U_i * H_{t-1} + V_i * E_t + B_i) \quad (9)$$

$$\hat{C}_t = f_t \circ C_{t-1} + I_t \circ \widetilde{C}_t, \quad (10)$$

where $\{W_c, U_c, V_c\}$ are the parameters of the temporary cell, with B_c being the corresponding bias; $\{W_f, U_f, V_f\}$ are the parameters of the forget gate, with B_f being the corresponding bias; $\{W_i, U_i, V_i\}$ are the parameters of the input gate, with B_i being the corresponding bias; '$\circ$' denotes the Hadamard product; and '$*$'denotes the convolution operator. The convolution operator '$*$', which is used to process the tensor-based market information, is explained in Section 3.4.2.

To address events occurring at nonequal time intervals, E_t is used to control what type of market information should be utilized on the basis of event occurrence at time t. Thus, we obtain an event control factor r_t via a nonincreasing mapping function, as follows:

$$r_t = \sigma(V_r * E_t + B_r) \quad (11)$$

If the market information is strongly related to current events, r_t tends to be large. In this case, more event-related information C_r will be retained in the cell memory. Specifically, the event-related memory C_r can be extracted via a tanh function, and the cell memory C_t at time t is determined by Equation (13):

$$C_r = tanh(\hat{C}_t) \tag{12}$$

$$C_t = \hat{C}_t + (C_r \circ r_t - C_r) \tag{13}$$

Finally, the cell memory C_t and the market information X_t together are passed through the output gate O_t to obtain the output H_t. Specifically,

$$O_t = \sigma(W_o * X_t + U_o * H_{t-1} + V_o * E_t + B_o) \tag{14}$$

$$H_t = O_t \circ tanh(C_t) \tag{15}$$

With this proposed architecture, we are able to address multimodal data with heterogeneous sampling intervals, specifically, data for which some data modes are sampled at equal intervals and other modes are sampled at nonequal intervals. The pseudocode for this proposed algorithm is presented in Algorithm 2.

Algorithm 2 Event-driven long short-term memory model

Input: The training tensor stream $X_t^i|_{i=1}^N$ and the associated stock trends $y_t^i|_{i=1}^N$.
Output: Trained proposed model for stock predictions.
1: **For** time step t = 1 to T **Do**
2: Obtain candidate cell state $\widetilde{C}_t$ from the input at time t and the output at time $t-1 (t > 1)$ by Eq. 7.
3: Process information in time t through forget gate F_t and input gate I_t by Eq. 8, 9.
4: Obtain cell memory $\hat{C}_t$ at time t by Eq. 10.
5: Update cell memory $\hat{C}_t$ through the event-driven mechanism by Eq. 11- 13 and obtain the new cell state C_t at time t.
6: Obtain the output H_t through output gate O_t by Eq. 14, 15.
7: **End for**

3.4.2 *Tensor-based Convolution Operation:*

In Figure 6, the market information X_t is fed into the network along with the learned knowledge H_{t-1} from the previous stage $t-1$ for further analysis at time t. However, X_t and H_{t-1} are represented by tensors, which cannot be concatenated into a super compound vector, as is the case in a traditional LSTM model.

To merge, multiply, and sum this knowledge and information represented by tensors in the proposed network, as shown in Figure 6, we apply the convolution operations of the ConvLSTM model to process the tensors, as done in [48]. By virtue of the advantages of local connections and weight sharing possessed by these convolution operations, it becomes possible to capture the interactions among different information sources modeled as different tensor subspaces. Essentially, ConvLSTM provides the unique feature of temporally propagating interconnections through each ConvLSTM state. This makes it possible for us to process time-series data represented as tensors.

The convolution operation in Equations (7) to (14) is defined as $'*'$ and makes it possible to process tensors instead of vectors in the proposed network. Therefore, the interrelations among different sources of market information $X^t \in R^{I_1 \times I_2}$ can be captured for further analysis.

4 Experimental Evaluation

To gauge the effectiveness of the proposed approach for predicting media-aware stock movements, we conducted a series of experiments using actual market transaction data from January 1, 2015, to December 31, 2015. The source code and dataset are accessible on GitHub[1].

4.1 Experimental Data

In our experiments, we extended the CSI 100 stock data provided by Li et al. [9] with additional financial news articles crawled by our focus-topic crawler, as follows,

- Fundamental data: This dataset contains the financial statuses of 100 companies listed on the China Securities Index (CSI 100) between January 1, 2015, and December 31, 2015. Since the companies on the CSI 100 list are updated every six months, we selected 91 companies that remained on the list for the entire period.
- Media data: The release of important news information affects investors' expectations concerning a company's future, resulting in stock market fluctuations. We selected 45,021 news data points related to the 91 selected companies listed on the CSI 100 from www.eastmoney.com, which is one of the largest financial information websites in China.

Table 3 summarizes the descriptive statistics of the main variables of these two data sources. From Table 3(a), it can be observed that the variation in volume is substantial. Conversely, the variations in stock prices (opening price, closing price, highest price and lowest price) are much smaller. The standard derivations of these prices are all around 20. This is understandable because the highest price of any stock is 116, while volume can reach a considerable quantity of shares (51,354,670). Furthermore, as indicated earlier, the stock prices wander around a firm's intrinsic value and thus have relatively low volatility.

Table 3(b) shows the statistical information of our media data set. There are 34,656 positive and 10,365 negative news articles. The number of positive news articles is approximately 3.5 times the number of negative news articles. This is consistent with the findings that news about the stock markets tends to present positive sentiment towards a firm [49]. The shortest positive news consists of 18 words, while the shortest negative news is 16 words. The longest news article reaches 20,936 words. At the company level, the firm with the least news coverage has only 74 news reports, while the firm with the most public attention has 5,703 news articles.

In our experiments, the data of the first 9 months are utilized to train the model, and the last 3 months of data are used for evaluation. Note that the evaluation is conducted in a rolling window fashion. In particular, the basic model is trained with the first 9 months of data. When evaluating the t^{th} day in the test period, the input data prior to the t^{th}

1. https://github.com/Tanny16/Tensor-based-eLSTM.

day are continuously added to update the training model, ensuring the latest market information is included.

TABLE 3: Descriptive statistics of features.

(a) Descriptive statistics of fundamental data

Variables	Min	Mean	Max	Std	25%	75%
Opening price	2.200	26.602	110.710	20.479	12.410	32.400
Closing price	2.220	26.633	112.300	20.517	12.438	32.450
Highest price	2.290	27.118	116.000	20.887	12.715	32.990
Lowest price	2.200	26.150	110.710	20.134	12.150	31.818
Turnover	0.009	1.685	62.248	2.413	0.439	1.916
Volume	1,018	948,717	51,354,670	2,177,324	138,306	871,434
P/B ratio	0.817	3.367	14.889	1.911	1.920	4.420
P/E ratio	6.642	40.428	477.466	61.221	12.113	36.977

(b) Descriptive statistics of media data

Category		N_n	News Length			Company News Count		
			Min	Avg	Max	Low	Avg	Top
News	P	34,656	18	1,201	20,936	63	1,264	5,512
	N	10,365	16	728	8,758	9	42	210
Total		45,021	16	1,137	20,936	74	1,418	5,703

4.2 Evaluation Settings

In this study, the directional accuracy (DA) and the Matthews correlation coefficient (MCC) were selected as evaluation metrics to measure the system performance [16, 27, 39, 50]. The DA is the most popular metric for stock classification tasks. This metric measures the upward or downward differences in the predicted trends compared to the actual changes in stock prices.

However, the DA tends to exhibit bias if the classes are of very different sizes. Suppose that there are 100 samples, among which 98% are positive samples and the remainder are negative samples. If a classifier judges all samples to be positive samples, the DA achieved is 98%. However, this classifier fails to recognize the negative samples even though it has a high DA. Therefore, in this study, we also adopted the MCC metric avoid such bias caused by skewed data. For both metrics, a larger value indicates better performance. The two metrics are defined as follows:

$$DA = \frac{n}{N} \tag{16}$$

$$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}}, \tag{17}$$

where n is the number of predictions for which the predicted trend and the actual stock trend show the same direction of stock movements and N is the total number of predictions over several days. $tp(tn)$ is the number of positive (negative) samples correctly classified as positive (negative) ones. $fp(fn)$ is the number of negative (positive) samples falsely classified as positive (negative) ones.

In addition, considering the financial scenario, the Sharpe ratio is applied to measure the investment robustness of the proposed multimodal event-driven LSTM framework. The Sharpe ratio helps investors compare investments in terms of both risks and returns [49]. Specifically, the Sharpe ratio describes how much excess return the investors receive for the extra volatility they endure for holding a riskier asset. The higher the ratio is, the greater the investment return relative to the amount of risk taken, and thus, the better the investment. It can be defined as,

$$S = \frac{r_m - r_f}{\sigma}, \tag{18}$$

where r_m is the mean return of an investment, r_f is the risk-free rate represented by the bank rate, and σ is the standard deviation of the investment return/volatility. Thus, the Sharpe ratio S measures the excess return per unit of risk. It allows an investor to better isolate the profits associated with risk-taking activities.

We take the further step of evaluating the k-day-ahead influence of media information. In particular, there are several k-day-ahead stock trends that can be of interest to investors. We carry out three tracks with different trends, as suggested by [11]. In Table 4, target 1 compares the opening stock price on day $i+k$ with the opening price on day i. Target 2 follows the same logic but based on the closing price instead of the opening price. For target 3, we compare the closing price on day $i+k$ with the opening price on day i. Although the targets of these three tracks are different, their inputs, which consist of the fundamental and media features from the previous k days, are the same. It can be defined as $< f_t^1, f_t^2, ..., f_t^m, ..., f_{t+k-1}^1, f_{t+k-1}^2, ..., f_{t+k-1}^m >$, where m is the number of features.

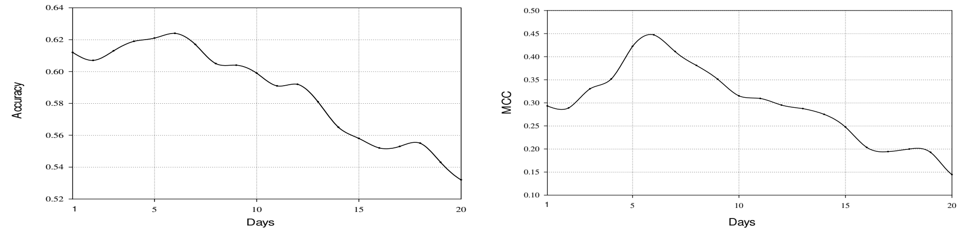
TABLE 4: Three k-days-ahead targets.

Tracks	Targets formula
Target 1	$price_{i+k}^{open} - price_{i+k-1}^{open}$
Target 2	$price_{i+k}^{close} - price_{i+k-1}^{close}$
Target 3	$price_{i+k}^{close} - price_{i+k-1}^{open}$

4.3 Model Parameters

Essentially, the proposed algorithm is an extension of the traditional LSTM approach. The basic framework of our algorithm was implemented using Keras and TensorFlow. We tuned the parameters to achieve the optimal performance of the proposed method. In our preliminary study, we found an optimal θ of 0.8 in Equation (6). We also found that among several classical activation functions, including the sigmoid, tanh, and exponential linear unit (ELU), the rectified linear unit (ReLU) function achieved the best performance. We utilized the Adam optimizer because it allows the learning rate to be set automatically based on the update history of the model weights.

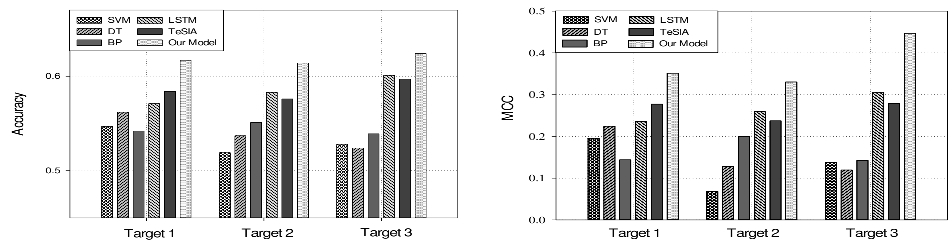
In addition, we carry out a series of experiments with different k days ahead to investigate the optimal period of influence of the media on stock movements. Figure 7 shows that stock markets respond to the news immediately, and the impact of the media on stock markets lasts for a few days. Such influence is able to last up to 6 days and attenuates thereafter in terms of both the DA and MCC in our experimental setting. This finding supports the findings of previous studies regarding the short-term effect of media-aware stock movements [9, 18]. Note that, since k indicates the future time after releasing a news article, the range of k starts from 1 as shown in the x-axis of Figure 7.



(a) Directional accuracy

(b) Matthews correlation coefficient

Fig. 7: DA (a) and MCC (b) over different day intervals for the proposed multimodal event-driven LSTM model.



(a) Directional accuracy

(b) Matthews correlation coefficient

Fig. 8: Prediction results for the directions of stock movements.

4.4 Comparison

To gauge the overall performance of the proposed approach, we compared it with several classic methods, including SVM, DT, backpropagation (BP) neural network, and LSTM models. The baselines are described as follows:

- SVM: We directly concatenated the fundamental and media information to form a super compound vector, and used this vector as the input to the SVM model.
- DT: The DT approach is an effective modeling method for stock forecasting. Therefore, we applied it as one of our benchmarks. The concatenated compound vector was directly fed into the DT model for predictions.
- BP: We concatenated the fundamental and media information into a compound vector, which was fed into the BP neural network to generate predictions.
- LSTM: LSTM networks can achieve excellent performance on time-series data. Here, we applied an LSTM model to capture the time dependency of stock data. The concatenated compound vector was used as the input to the LSTM model.
- TeSIA: TeSIA is one of the state-of-the-art methods for forecasting media-aware stock movements [16]. In TeSIA, the market information is modeled with tensors to capture the interconnections among different information modes. Here, we modeled the fundamental and media information as tensors instead of using vectors as input.
- Our model: For our proposed multimodal event-driven LSTM model, the fundamental and media information are modeled as tensors to be used as the input to the model.

Thus, we compared the proposed method to five classic approaches (SVM, DT, BP, LSTM, and TeSIA) with three different targets representing different predicted outcomes. Table 5 and Figure 8 present the details of our experimental results.

TABLE 5: Prediction results for the directions of stock movements.

Model	Target 1		Target 2		Target 3	
	DA	MCC	DA	MCC	DA	MCC
SVM	0.547	0.1956	0.519	0.0679	0.528	0.1374
DT	0.562	0.2244	0.537	0.1277	0.524	0.1195
BP	0.542	0.1438	0.551	0.1997	0.539	0.1422
LSTM	0.571	0.2354	0.583	0.2594	0.601	0.3058
TeSIA	0.584	0.2775	0.576	0.2371	0.597	0.2789
Our model	0.617	0.3516	0.614	0.3304	**0.624**	**0.4472**

In terms of both the DA and MCC metrics, the SVM and DT models achieved the best performance for target 1 among the three considered targets. The BP model achieved its best performance for target 2, whereas LSTM, TeSIA and the proposed approach achieved their best performance for target 3. Among the baseline models, the models that achieved their best performance for target 3 also achieved better performance overall compared with the traditional SVM, DT and BP models. A good explanation for this behavior is that the memory dependency in the LSTM network enhances the effectiveness of LSTM-based methods on time-series data. Furthermore, the result of experiments performed on the entire year of data on the China securities market demonstrate the superiority of the proposed approach over even these superior baselines, with performance enhancements of at least 2.3% and 14.1% in terms of the DA and MCC, respectively. 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 was statistically significant.

4.5 Effectiveness of the Proposed Approach

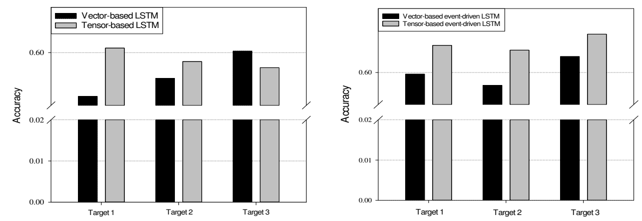
The modeling of media-aware stock movements is essentially a multimodal problem. Two unique challenges arise in processing such multimodal data. First, the information from one data mode interacts with information from other data modes. One common strategy for addressing this challenge is to concatenate information from various data modes into one compound vector, thus simply ignoring the interactions among the different modes. The second challenge is the heterogeneity of the data in terms of sampling time. Specifically, fundamental stock data consist of continuous values sampled at fixed time intervals, whereas news information emerges randomly. This heterogeneity can lead to some valuable information being missing or can even distort the feature spaces. In addition, previous studies on media-aware stock movements have focused on the one-to-one problem, in which news is assumed to affect the performance of only the stocks mentioned in the news report. However, news can also impact related stocks and cause stock comovements. In this section, we examine the effectiveness of the unique features of the proposed framework in addressing these three issues.

4.5.1 *Effectiveness of the Tensor Representation*

As mentioned above, both fundamental and media information can shape stock movements. Predicting media-aware stock movements is essentially a multimodal problem. Difficulties arise in attempting to model the two relevant types of information without ignoring the interactions among different data types. One common strategy is to concatenate multiple types of information into one compound vector, thus inevitably diluting or even ignoring the intrinsic associations between the two information sources. In this study, we modeled market information in the form of tensors to retain the interactions between different modes when addressing the multimodal learning problem. To investigate the effectiveness of this tensor representation, we first compared the results of a vector-based LSTM model with those of a tensor-based LSTM model and then explored the differences between the results of a vector-based event-driven LSTM model and the proposed tensor-based event-driven LSTM model, in which an event-driven mechanism is adopted to account for data that are heterogeneous in terms of their sampling times for the multimodal learning problems.

Figure 9 compares vector-based models with tensor-based models to show the superiority of the tensor representation. Specifically, the tensor-based LSTM achieves better performance than the vector-based LSTM in terms of target 1 and 2 in Figure 9(a). Note, however, that the vector-based LSTM model performs slightly better than the tensor-based LSTM model for target 3. In addition, Figure 9(b) shows that the proposed tensor-based event-driven LSTM model outperforms the vector-based event-driven LSTM model for all targets. These results prove that the interactions among different information modes affect stock movements and that the tensor representation can efficiently preserve such connections compared with the vector representation.

One possible explanation of the failure of the tensor representation in the LSTM model for target 3 is that this



(a) LSTM models (b) Event-driven models

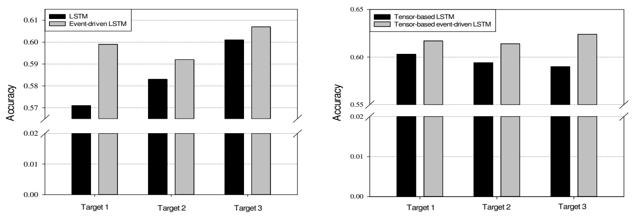
Fig. 9: DA results for (a) LSTM models and (b) Event-driven models using vector and tensor representations.

target utilizes both the opening price and the closing price, and using both prices allows additional information to be considered, thus overcoming the limited ability of the vector representation in the traditional LSTM model to capture interaction information. This observation also supports the finding that nontransactional time information can be reflected in the opening and closing prices [51]. In contrast, the success of the tensor representation in the event-driven LSTM model with respect to target 3 can be attributed to the event-driven mechanism in the event-driven LSTM model, which captures even more valuable interaction information, thus overwhelming the benefit gained from the additional information absorbed by considering both the opening and closing prices. A more comprehensive gauge of the effectiveness of the event-driven mechanism in accounting for the heterogeneity of the data in terms of sampling time for multimodal learning is presented in the next section.

4.5.2 *Effectiveness of the Event-driven Mechanism*

To capture media-aware stock movements, accounting for the heterogeneity of the sampling times between the two information modes is a critical issue for this multimodal learning problem. Specifically, the fundamental data consist of continuous values sampled at equal time intervals, i.e., one day, whereas the news information consists of discrete values sampled at nonequal time intervals because of the randomness of news releases. This randomness leads to failure of the long-term dependency mechanism of the traditional LSTM model [47]. Specifically, if two similar news articles are released with a sufficiently large time gap, the LSTM model may forget the knowledge learned from the first news article before processing the information from the second. In this study, we have proposed an event-driven memory mechanism to solve this problem of heterogeneous data sampling for multimodal learning.

Figure 10 presents the effectiveness of the proposed event-driven mechanism in both the vector-based and tensor-based LSTM models. Figure 10(a) shows that the vector-based event-driven LSTM model performs better than the traditional LSTM model for all targets. Similarly, Figure 10(b) shows that the tensor-based event-driven LSTM model outperforms the tensor-based LSTM model for all targets. These findings confirm that the event-driven mechanism allows the event-driven LSTM model to better find the rules and patterns characterizing stock markets given random news event occurrences. In previous studies, this problem has typically been solved by using only a portion of the data; that is, only the data sampled at the time of a news event are retained for further analysis. However, the failure to address the sampling heterogeneity leads to a loss of important patterns, inevitably causing historical information to be undervalued. By contrast, the event-driven memory mechanism proposed in this study provides a promising method of addressing the problem presented by the sampling heterogeneity among different data sources in multimodal learning.



(a) Vector-based methods (b) Tensor-based methods

Fig. 10: DA results for (a) vector-based and (b) tensor-based LSTM using event-driven LSTM models.

4.5.3 *Effectiveness of Stock Relatedness*

In this section, we evaluate the performance enhancement achieved by considering the influence of related stocks on a target stock. Previous studies on media-aware stock movements have focused only on the one-to-one problem, without considering the impact of related stocks. The main challenge for considering related stocks is how to define the relevant relationships among stocks. In this study, we have built a media-based enterprise network under the assumption that the co-occurrence of stocks in a news article reflects their relatedness to some degree. Such relevant influences are absorbed via the tensor decomposition and reconstruction when modeling the market information (Section 3.3).

As mentioned above, there are two ways to identify stocks related to a target stock. One approach (implemented in the tensor-based event-driven LSTM_dl model) is to treat only firms with direct links to the target firm in the media-based enterprise network as related firms. The other approach (implemented in the tensor-based event-driven LSTM_lc) is to treat all firms in the same link communities as related firms, thus considering the transitive effect. Figure 11 presents the performance of the tensor-based event-driven LSTM_none, tensor-based event-driven LSTM_dl, and tensor-based event-driven LSTM_lc models (our approach). The approaches that consider the influence of related stocks (tensor-based event-driven LSTM_dl and tensor-based event-driven LSTM_lc) outperform the tensor-based event-driven LSTM_none model, which ignores stock comovements. This finding confirms that a target firm is affected by its related firms and that media coverage is an effective way to measure such relatedness. In addition, the tensor-based event-driven LSTM_lc model, which considers the transitive effect when identifying related stocks in the media-based enterprise network, outperforms the tensor-based event-driven LSTM_dl model, which identifies related stocks on the basis of only direct linkages in the network. This finding indicates that the way in which relatedness is defined has a critical effect on the prediction performance and that considering link communities is a promising approach to defining relatedness.

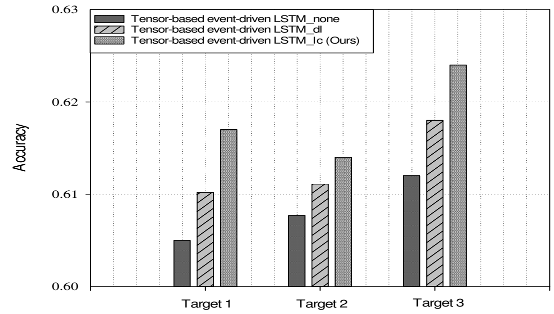


Fig. 11: DA results for tensor-based event-driven LSTM models considering stock relatedness.

4.6 Investment Simulation

In this section, we describe the design and implementation of a tensor-based stock market information analyzer based on the proposed multimodal event-driven LSTM framework. We compare the performance of our proposed analyzer with the performances of three state-of-the-art trading algorithms, namely, eMAQT [9], TeSIA [16] and AZFinText [18], as well as the classic top-N trading strategy. Top-N is a long-term strategy based on the assumption that if a certain combination of stocks has performed well in the past, then the same combination will perform well in the near future. We invested in the N highest-performing stocks over the period between October 1 and December 31, 2015.

We chose RMB 10,000 as the investment budget, and we further assumed zero transaction costs, as in previous studies [16]. Finally, we compared the daily earnings of the five approaches over three months, during which time the CSI 100 index decreased by 5.21% (from $2,363$ to $2,240$).

In this simulation, both selling short and buying long were allowed. Specifically, when a firm-specific news article was released, these algorithms were used to forecast the future stock price for that firm. For buying long, if the trend of the predicted future price over the current stock price was a rising signal, then the stock was purchased immediately and sold 6 days later. The investment gain was calculated as the spread between the sale and purchase prices. For selling short, if the trend of the predicted future price over the current stock price was a falling signal, then the stock was borrowed, sold immediately and purchased at the original price after 6 days. The investment gain was calculated as the stock price at the time when the shares were borrowed minus the purchase price. The horizon of 6 days was set in accordance with our optimal parameters in Section 4.3. Figure 12 presents the accumulated daily return (x-axis) of these five methods over the 3-month assessment period (y-axis) along with the CSI 100 index in the same period. Notably, the top-N method relied on a long-term investing strategy, with trading only at the end of the assessment period. Thus, the daily income of the top-N method reflects only the value of its portfolio on that day. For the other media-aware trading systems, the daily income is the sum of all transaction incomes earned on that day.

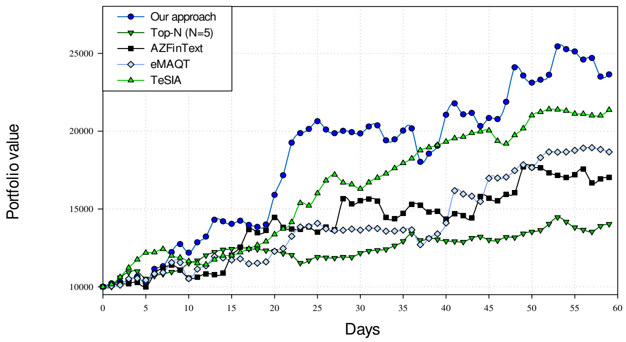


Fig. 12: Investment simulation comparison.

Figure 12 shows that among all baselines, TeSIA achieved the best return of 113.62% (p-values<0.05), and the top-N (N=5) method achieved the lowest return of 40.32% at the end of the 3-month assessment period. Among all systems, our proposed approach achieved superior performance, yielding a remarkable return of 136.42% (p-values<0.05). This result illustrates the superiority of the proposed framework, including the mechanism of tensor representation, event-driven and stock relatedness. Note that this profit came from both buying long and selling short. The profit from buying long was 62.35%, whereas the profit from selling short was 74.07%. Selling short was more advantageous because the predominant market trend was downward over the evaluation period.

Specifically, AZFinText algorithm, the first framework to incorporate both stock data and news data for predictions, achieved a return of 70.30% in our experiment. By considering finance-oriented sentiment, eMAQT achieved better performance, with an enhancement of 16.38%, compared with AZFinText. In addition, the tensor-based methods (TeSIA and our method) achieved superior returns with an enhancement of 26.94% and 49.74%, respectively, compared with the optimal vector-based method (eMAQT). The results further demonstrate the effectiveness of the tensor representation.

Recall that a higher Sharpe ratio means a greater return relative to the amount of risk taken. Table 6 presents the Sharpe ratio of four baselines and our proposed method in the 3-month assessment period. These results further indicate the superiority of the proposed method, especially in terms of the trade-off between the return and the risk.

TABLE 6: The Sharpe ratio for the different algorithms.

Algorithm	Sharpe ratio
Top-N	1.195%
AZFinText	1.023%
eMAQT	1.255%
TeSIA	1.307%
Our method	1.418%

5 Conclusions and Future Work

Stock markets are strongly affected by various types of highly interrelated information. In both traditional finance and behavioral finance, it is believed that market information, especially fundamental information and news report information, shapes stock movements. Predicting future stock trends based on market information is essentially a multimodal data problem. Multimodal data consist of several modes, each corresponding to a group of similar data sharing the same attributes. In this study, market information data are considered to consist of two modes: fundamentals and news. Two unique challenges arise in processing these multimodal data. The first challenge is that the information from one data mode interacts with information from other data modes, violating the assumption of feature independence that is adopted in traditional supervised learning. A common strategy in previous studies has been to concatenate the information from various data modes into a compound vector, thereby ignoring the interactions among different modes. By contrast, in this study, we proposed a tensor representation approach for modeling multimodal market information. This method is able to preserve the interrelations between fundamental and news information and to capture their joint effects. The second challenge is the sampling heterogeneity of the different data modes. For market information, fundamental data are continuous values sampled at fixed time intervals, whereas news information emerges randomly. This heterogeneity can result in a partial loss of valuable information and can even distort the feature space. In this study, we proposed an event-driven memory mechanism to address the sampling heterogeneity among different data sources for multimodal learning. Experiments performed on an entire year of data from the China securities market demonstrate the superiority of the proposed approach over state-of-the-art algorithms, including AZFinText, eMAQT, and TeSIA, and our method achieved a return of 136.42% in an investment simulation.

In this study, we focused on media-aware stock movements. However, the proposed tensor-based event-driven LSTM framework can be generalizable to many other multimodal learning problems in which the information space consists of several interacting data modes with sampling heterogeneity. For instance, in health care monitoring, both daily monitoring indicators and random sickness records are applied to detect health abnormalities [47]. Another good example is the prediction of crop growth in agriculture based on daily growth indicators and uncertain conditions, including rainfall, wind and disasters [48]. However, the effectiveness of the proposed method in related fields has yet to be explored.

6 Acknowledgments

This work has been supported by the National Natural Science Foundation of China (NSFC) (71671141, 71873108), Fundamental Research Funds for the Central Universities (JBK 171113, JBK 170505, JBK 1806003), Sichuan Province Science and Technology Department (2019YJ0250), and the Financial Innovation Center of the Southwestern University of Finance and Economics. It also has been partially funded by grants awarded to Dr. Hsinchun Chen from the U.S. National Science Foundation (ACI-1443019, CMMI-1442116) at the University of Arizona and the China National 1000-Talent Program at Tsinghua University.

References

[1] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, "Noise trader risk in financial markets," *Journal of Political Economy*, vol. 98, no. 4, pp. 703–738, 1990.
[2] E. F. Fama, "The behavior of stock-market prices," *The journal of Business*, vol. 38, no. 1, pp. 34–105, 1965.
[3] M. Rechenthin and W. N. Street, "Using conditional probability to identify trends in intra-day high-frequency equity pricing," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 24, pp. 6169–6188, 2013.
[4] A. Shleifer and R. W. Vishny, "A survey of corporate governance," *Journal of Finance*, vol. 52, no. 2, pp. 737–783, 1997.
[5] Q. Li, C. Yan, J. Wang, Y. Chen, and H. C. Chen, "Web media and stock markets : A survey and future directions from a big data perspective," *IEEE Transactions on Knowledge & Data Engineering*, vol. 30, no. 2, pp. 381–399, 2018.
[6] W. Nuij, V. Milea, F. Hogenboom, F. Frasincar, and U. Kaymak, "An automated framework for incorporating news into stock trading strategies," *IEEE Transactions on Knowledge & Data Engineering*, vol. 2, no. 11, pp. 823–835, 2014.
[7] H. S. Moat, C. Curme, A. Avakian, D. Y. Kenett, H. E. Stanley, and T. Preis, "Quantifying wikipedia usage patterns before stock market moves," *Scientific Reports*, vol. 3, p. 1801, 2013.
[8] I. Zheludev, R. Smith, and T. Aste, "When can social media lead financial markets?" *Scientific Reports*, vol. 4, p. 4213, 2014.
[9] Q. Li, T. Wang, P. Li, L. Liu, Q. Gong, and Y. Chen, "The effect of news and public mood on stock movements," *Information Sciences*, vol. 278, pp. 826–840, 2014.
[10] R. P. Schumaker and H. Chen, "A quantitative stock prediction system based on financial news," *Information Processing & Management*, vol. 45, no. 5, pp. 571–583, 2009.
[11] B. Weng, M. A. Ahmed, and F. M. Megahed, "Stock market one-day ahead movement prediction using disparate data sources," *Expert Systems with Applications*, vol. 79, pp. 153–163, 2017.
[12] P. C. Tetlock, M. Saar-Tsechansky, and S. Macskassy, "More than words: Quantifying language to measure firms' fundamentals," *Journal of Finance*, vol. 63, no. 3, pp. 1437–1467, 2008.
[13] Y. F. Wang, "On-demand forecasting of stock prices using a real-time predictor," *IEEE Transactions on Knowledge & Data Engineering*, vol. 15, no. 4, pp. 1033–1037, 2003.
[14] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
[15] M.-A. Mittermayer and G. F. Knolmayer, "Newscats: A news categorization and trading system," in *The 6th International Conference on Data Mining (ICDM)*. IEEE, 2006, pp. 1002–1007.
[16] Q. Li, Y. Chen, L. L. Jiang, P. Li, and H. Chen, "A tensor-based information framework for predicting the stock market," *ACM Transactions on Information Systems (TOIS)*, vol. 34, no. 2, p. 11, 2016.
[17] Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Transactions on Signal Processing*, vol. PP, no. 99, pp. 3675–3687, 2016.
[18] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The azfin text system," *ACM Transactions on Information Systems (TOIS)*, vol. 27, no. 2, pp. 1–19, 2009.
[19] B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, and J. Zhang, "Daily stock market forecast from textual web data," in *IEEE International Conference on Systems, Man, and Cybernetics*, 1998, pp. 2720–2725.
[20] S. Qiang, A. Liu, and S. Y. Yang, "Stock portfolio selection using learning-to-rank algorithms with news sentiment," *Neurocomputing*, vol. 264, pp. 20–28, 2017.
[21] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, "Predicting stock price returns using microblog sentiment for chinese stock market," in *The 3rd International Conference on Big Data Computing and Communications (BIGCOM)*. IEEE, 2017, pp. 87–96.
[22] R. A. Haugen and N. L. Baker, "Commonality in the determinants of expected stock returns," *Journal of Financial Economics*, vol. 41, no. 3, pp. 401–439, 1996.
[23] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, vol. 33, no. 1, pp. 3–56, 1993.
[24] N. Jegadeesh and S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency," *Journal of Finance*, vol. 48, no. 1, pp. 65–91, 1993.
[25] W. DeBondt and R. Thaler, "Does the stock market overreact?" *Journal of Finance*, vol. 60, no. 3, pp. 793–805, 1985.
[26] L. Yu, H. Lunchao, and T. Ling, "Stock selection with a novel sigmoid-based mixed discrete-continuous differential evolution algorithm," *IEEE Transactions on Knowledge & Data Engineering*, vol. 28, no. 7, pp. 1891–1904, 2016.
[27] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and S. Y. Philip, "Improving stock market prediction via heterogeneous information fusion," *Knowledge-Based Systems*, vol. 143, pp. 236–247, 2018.
[28] R. S. Pindyck and J. J. Rotemberg, "The comovement of stock prices," *The Quarterly Journal of Economics*, vol. 108, no. 4, pp. 1073–1104, 1993.
[29] T. Preis, D. Y. Kenett, H. E. Stanley, D. Helbing, and E. Ben-Jacob, "Quantifying the behavior of stock correlations under market stress," *Scientific Reports*, vol. 2, p. 752, 2012.
[30] S. Aghabozorgi and Y. W. Teh, "Stock market comovement assessment using a three-phase clustering method," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1301–1314, 2014.
[31] M. S. Rashes, "Massively confused investors making conspicuously ignorant choices (mci–mcic)," *Journal of Finance*, vol. 56, no. 5, pp. 1911–1927, 2001.
[32] W. Antweiler and M. Z. Frank, "Is all that talk just noise? the information content of internet stock message boards," *Journal of Finance*, vol. 59, no. 3, pp. 1259–

1294, 2004.

[33] X. Luo, J. Zhang, and W. Duan, “Social media and firm equity value,” *Information Systems Research*, vol. 24, no. 1, pp. 146–163, 2013.

[34] Z. Y. Huang X, Teoh S H, “Tone management,” *The Accounting Review*, vol. 89, no. 3, pp. 1083–1113, 2013.

[35] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification,” *IEEE Transactions on Systems Man & Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[36] G. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[37] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[38] A. R. Mohamed, G. E. Dahl, and G. Hinton, “Acoustic modeling using deep belief networks,” *IEEE Transactions on Audio Speech & Language Processing*, vol. 20, no. 1, pp. 14–22, 2011.

[39] X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep learning for event-driven stock prediction.” in *International Joint Conferences on Artificial Intelligence (IJCAI)*, 2015, pp. 2327–2333.

[40] Y. Huang, K. Huang, Y. Wang, H. Zhang, J. Guan, and S. Zhou, “Exploiting twitter moods to boost financial trend prediction based on deep network models,” in *International Conference on Intelligent Computing*. Springer, 2016, pp. 449–460.

[41] L. Troiano, E. M. Villa, and V. Loia, “Replicating a trading strategy by means of lstm for financial industry applications,” *IEEE transactions on industrial informatics*, vol. 14, no. 7, pp. 3226–3234, 2018.

[42] T. Kolda and B. Bader, “Tensor decompositions and applications,” *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[43] R. D. McLean and J. Pontiff, “Does academic research destroy stock return predictability?” *Journal of Finance*, vol. 71, no. 1, pp. 5–32, 2016.

[44] L. Ling, W. Jing, L. Ping, and Q. Li, “A social-media-based approach to predicting stock comovement,” *Expert Systems with Applications*, vol. 42, no. 8, pp. 3893–3901, 2015.

[45] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, “Link communities reveal multiscale complexity in networks,” *Nature*, vol. 466, no. 7307, p. 761, 2010.

[46] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[47] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou, “Patient subtyping via time-aware lstm networks,” *SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65–74, 2017.

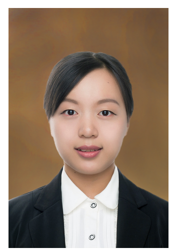
[48] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” in *Advances in Neural Information Processing Systems*, 2015, pp. 802–810.

[49] S. A. Sharpe, “Financial market imperfections, firm leverage, and the cyclicality of employment,” *The American Economic Review*, vol. 84, no. 4, pp. 1060–1074, 1994.

[50] S. Lei, Z. Teng, W. Le, Z. Yue, and A. Binder, “Deepclue: Visual interpretation of text-based deep stock prediction,” *IEEE Transactions on Knowledge & Data Engineering*, vol. PP, no. 99, pp. 1–1, 2018.

[51] K. Tissaoui, “The intraday pattern of trading activity, return volatility and liquidity: Evidence from the emerging tunisian stock exchange,” *International Journal of Economics and Finance*, vol. 4, no. 5, p. 156, 2012.

Qing Li is a professor at Southwestern University of Finance and Economics, China. Prior to taking that post, he was a postdoctoral researcher at Arizona State University and at the Information & Communications University of Korea. Li’s research interests lie primarily in intelligent information processing and business intelligence. He has served on the editorial boards of Electronic Commerce Research and Applications, the Journal of Database Management, and the Journal of Global Information Management as well as the program committees of various international conferences, including PACIS, SIGIR, and CIKM. He received his Ph.D. from Kumoh National Institute of Technology in February of 2005 and his M.S. and B.S. degrees from Harbin Engineering University, China.



Jinghua Tan is a Ph.D. student at Southwestern University of Finance and Economics, China. She obtained her bachelor’s degree from Southwestern University of Finance and Economics in 2016. She was a visiting scholar at Memorial University of Newfoundland at St. John’s, Canada in 2019. Her research primarily focuses on data mining and finance intelligence.

Jun Wang is an associate professor at Southwestern University of Finance and Economics, China. Prior to taking that post, he was a researcher at Memorial University of Newfoundland at St. John’s, Canada. He was awarded the National Scholarship in 2017. His research interests lie primarily in social media, social network analysis, financial analysis and business intelligence.

Hsinchun Chen is a Fellow of IEEE, ACM, and AAAS. He is a University of Arizona Regents Professor and the Thomas R. Brown Chair in Management and Technology in the Department of Management Information Systems (MIS) in the School of Management of the University of Arizona. His research primarily focuses on data/web/text mining and knowledge management techniques. He is Editor-in-Chief of ACM Transactions on Management Information Systems, Editor-in-Chief of the Springer Security Informatics journal, and Associate Editor-in-Chief of IEEE Intelligent Systems. He has served on the following editorial boards: ACM Transactions on Information Systems; IEEE Transactions on Systems, Man, and Cybernetics; Decision Support Systems; Journal of the American Society for Information Science and Technology; International Journal of Digital Libraries; International Journal of Electronic Business; Journal of Information Technology and Politics; and Encyclopedia of Library and Information Sciences. He received his B.S. degree from the National Chiao-Tung University in Taiwan, his MBA degree from SUNY Buffalo, and his Ph.D. degree in Information Systems from New York University.