A Tensor-Based Information Framework for Predicting the Stock Market

QING LI, Southwestern University of Finance and Economics, China
YUANZHU CHEN, Memorial University of Newfoundland, Canada
LI LING JIANG and PING LI, Southwestern University of Finance and Economics, China
HSINCHUN CHEN, University of Arizona, USA, and Tsinghua University, China

To study the influence of information on the behavior of stock markets, a common strategy in previous studies has been to concatenate the features of various information sources into one compound feature vector, a procedure that makes it more difficult to distinguish the effects of different information sources. We maintain that capturing the intrinsic relations among multiple information sources is important for predicting stock trends. The challenge lies in modeling the complex space of various sources and types of information and studying the effects of this information on stock market behavior. For this purpose, we introduce a tensor-based information framework to predict stock movements. Specifically, our framework models the complex investor information environment with tensors. A global dimensionality-reduction algorithm is used to capture the links among various information sources in a tensor, and a sequence of tensors is used to represent information gathered over time. Finally, a tensor-based predictive model to forecast stock movements, which is in essence a high-order tensor regression learning problem, is presented. Experiments performed on an entire year of data for China Securities Index stocks demonstrate that a trading system based on our framework outperforms the classic Top-N trading strategy and two state-of-the-art media-aware trading algorithms.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning; H.4.2 [**Information Systems Applications**]: Types of Systems

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Tensor, predictive model, stock, trading strategy, news, social media

---

This work has been supported by grants awarded to Dr. Qing Li from the National Natural Science Foundation of China (NSFC) (91218301,61170133,60803106,71401139), the Sichuan National Science Foundation for Distinguished Young Scholars (2013JQ0004), and the Fundamental Research Funds for the Central Universities (JBK151128, JBK120505). It also has been partially funded by grants awarded to Dr. Hsinchun Chen from the U.S. National Science Foundation (ACI-1443019, DUE-1303362, CMMI-1442116, SES-1314631) at the University of Arizona and the China National 1000-Talent Program at the Tsinghua University.
Authors' addresses: Q. Li, Collaborative Innovation Center for the Innovation and Regulation of Internet-based Finance, School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu, Sichuan, China; email: liq_t@swufe.edu.cn; Y. Z. Chen, Computer Science Department, Memorial University of Newfoundland, St. John's, NL, Canada; email: yzchen@mun.ca; L. L. Jiang, Alibaba Group, Zhejiang, China; email: liling.jiangliling@alibaba-inc.com; P. Li, International Exchange and Cooperation, Southwestern University of Finance and Economics, Chengdu, Sichuan, China; email: pingl@swufe.edu.cn; H. Chen, Management Information System Department, University of Arizona, Tucson, USA; email: hsinchun@email.arizona.edu. (Professor Chen is also a China National 1000-Talent Professor in the Department of Computer Science and Technology at the Tsinghua University of China.)

DOI: http://dx.doi.org/10.1145/2838731

ACM Reference Format:
Qing Li, Yuanzhu Chen, Li Ling Jiang, Ping Li, and Hsinchun Chen. 2016. A tensor-based information framework for predicting the stock market. ACM Trans. Inf. Syst. 34, 2, Article 11 (February 2016), 30 pages.
DOI: http://dx.doi.org/10.1145/2838731

1. INTRODUCTION

Stock movements are strongly affected by various highly interrelated sources and types of information that cover a wide range of topics including economics, politics, and psychology. The traditional Efficient Market Hypothesis (EMH) states that a stock price is always driven by "unemotional" investors to equal the firm's rational present value of expected future cash flows [Fama 1965]. New information related to markets may change investors' expectations and cause stock prices to move. In particular, stock investors are constantly updating their beliefs about the directions of markets as they receive new information, although they often disagree on the direction of movement. This disagreement leads to discrepancies between the actual price and the intrinsic value, with competing market participants causing a stock price to fluctuate around a stock's intrinsic value [Rechenthin and Street 2013] (i.e., new information influences asset prices intricately). However, empirical studies have demonstrated that prices do not strictly follow random walks [Lo and MacKinlay 1988]. Recent behavioral finance studies have attributed the nonrandomness of stock movements such as overreactions to unfavorable news to investors' cognitive and emotional biases [Long et al. 1990; Shleifer and Vishny 1997]. Although traditional finance and modern behavioral finance have opposing views regarding how information affects stock markets, both assume that new information affects stock movements.

Stock market information can be roughly categorized into *quantitative* market data and *qualitative* descriptions of financial standings. Quantitative analysis is a basic and powerful tool for both fundamental and technical stock analysts. Specifically, fundamental analysts generally attempt to study economic and business data to predict price trends. They believe that examining fundamentals such as the overall economy, industry conditions, financial conditions, and management capabilities provides insight into the future directions of stock prices. In contrast, technical analysts examine historical stock price trends in an attempt to predict future prices, with the belief that stock markets are cyclical and exhibit specific patterns and that these patterns repeat over time [Edwards et al. 2012].

Quantitative data, however, cannot entirely convey the limitless variety of financial standings of firms [Tetlock 2007]. Qualitative information, which is embedded in textual descriptions provided by news and social media, is actually complementary to quantitative data in enriching the knowledge of investors. This complementarity is particularly significant in the era of social media, in which the advent of Web 2.0 technologies has generated vibrant knowledge creation, sharing, and collaboration among investors. Stock information is updated rapidly and spreads with unprecedented speed, thereby providing first-hand information to investors in advance of formal statistical reports [Luo et al. 2013].

Changes in user engagement in social media including comments, ratings, and votes enable much more rapid exchanges of information and user interactions. This situation may lead to herd behavior in investing because investors' decisions tend to be influenced by the emotions of their peers. In practical computational investing, qualitative information provided by news articles and social media is quantified using Natural Language Processing (NLP) techniques. Typically, a news article is represented as a "bag of words" (term vector) to capture financial standings [Schumaker and Chen 2009b], and social media are processed using sentiment analysis to obtain social sentiments [Bollen et al. 2011].

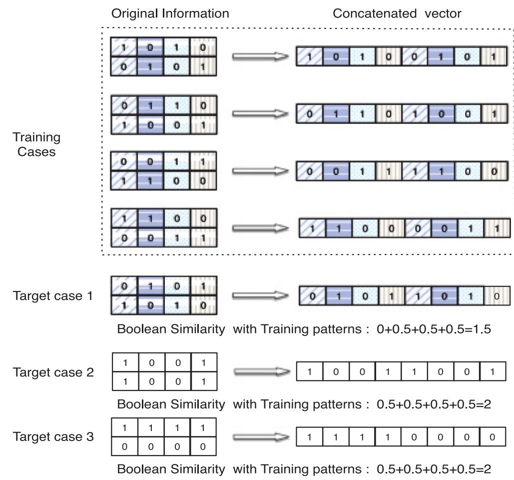


Fig. 1. Misclassification by the concatenated vector approach.

Essentially, stock information is multifaceted and interrelated. Event-specific information, firm-specific information, and sentiment information are the three primary sources of information (modes) that affect stock movements. Such complex information from various sources is also referred to as a *mosaic* information space [Francis et al. 1997], which implies strong interactions among the various information sources. Modeling the mosaic-like characteristic of investor information and studying the joint effect of multiple information sources on stock movements remain a challenge in computational analysis, and overcoming this challenge is critical for understanding the behavior of stock markets.

A common strategy in previous studies has been to concatenate the features of multiple information sources (or modes) into one compound feature vector, but as the dimension of the vector increases, solution strategies are subject to the “curse of dimensionality” [Bellman and Dreyfus 1962]. More importantly, in the mosaic approach, the various information sources are mingled and interact. With a concatenated vector representation, each information mode is assumed to be independent, and the contextual co-occurrence relations among the various information modes are reduced or possibly eliminated. Figure 1 shows the classification problem of mosaic information. In this study, a matrix is used to model a simplified investor information environment in which each row represents one information mode such as firm-specific, event-specific, or sentiment information. According to the mosaic information structure, feature patterns may exist either within an information mode (row) or between modes. More specifically, the matrix has the following three properties: (1) the sum of each column is 1, (2) the sum of each row is 2, and (3) the sum of all of the rows is 4. Theoretically speaking, target case *1* is more similar to target cases *2* and *3* because it possesses all of these patterns. However, if we remove the mosaic information structure and concatenate the features of the various information modes into one compound feature vector, target cases *2* and *3* may also seem similar to the training cases. This apparent similarity arises because the concatenated vector approach neglects the inherent links among the various information modes. Pattern (1) and pattern (2) vanish, and only pattern (3) is preserved. Such information losses cause misclassifications of the target cases. In addition to capturing these static interconnections among the various modes in one information matrix, it is important to identify and emphasize the dynamic connections among the various modes across a sequence of information matrices. For example, two news articles released at different times may be textually dissimilar, but both may contain favorable information regarding the same stock. Moreover, the corresponding firm-specific data at these two points in time may be similar and may indicate a good investment opportunity. In this scenario, the semantic similarities of different words can be enhanced by the similarities in the corresponding firm-specific data. Therefore, capturing, deducing, and reinforcing the dynamic connections among the various information modes are critical for predicting future stock trends.

For this purpose, we introduce a generic and scalable tensor-based computational framework to predict stock movements. Specifically, our framework models the complex information environment and its intrinsic links with tensors. A Global Dimensionality-Reduction (GDR) algorithm that captures the static and dynamic interconnections among the various information sources is presented. A tensor-based learning algorithm that predicts stock price trends in response to new information is then described. This framework provides a powerful tool for financial researchers to systematically study the joint effect of various information sources on stock movements.

In the remainder of this article, Section 2 briefly describes the related work of our research. Section 3 presents the design details of the proposed tensor-based information framework. Section 4 examines the effectiveness of the proposed approaches using a real stock data from the Chinese Stock Index (CSI) 100. This article is concluded with our findings and speculation on how the current work can be further improved in Section 5.

2. RELATED WORK

In this section, we evaluate the existing relevant research. We first review what types of information affect stock volatility and then discuss stock analysis models.

2.1. Information and Stock Volatility

A company's stock price reflects the value of the anticipated future profits of the company. Any information that is related to the firm's fundamentals or that affects investors' expectations may cause fluctuations in the stock price.

A number of studies using traditional finance approaches have examined the effects of quantifiable, firm-specific information such as firm size, cash flow, book-to-market equity, and historical data on stock price movements. For example, Dechow [1994] demonstrated that both accounting earnings and cash flows can be used to measure a firm's performance as reflected in stock returns. Jegadeesh and Titman [1993] found that stocks with higher returns in the previous 12 months tended to have higher future returns. Cheung and Ng [1992] observed that the relationship between stock price dynamics and firm size (market equity, or ME, which is the stock price multiplied by the number of shares) was stable, but the strength of the relationship appeared to change over time. Fama and French [1993] identified three risk factors for returns on stocks, the overall market, the firm size, and the book-to-market equity ratio (BE/ME), which is the ratio of the book value of the common equity to its market value.

Recent studies in behavioral finance have demonstrated that investment decisions are affected by the emotional impulses of investors [Li et al. 2014b]. In particular, the decisions of investors regarding stock prices are based less on fundamentals than traditional financial theories would suggest [Shleifer and Vishny 1997], and beliefs about future cash flows and investment risks are affected by emotions rather than by facts [Long et al. 1990]. Tetlock et al. [2008] demonstrated that negative sentiments expressed in news reports regarding Wall Street could be used to forecast decreases in firm earnings.

The importance of investor emotions has led to a burgeoning area of research into the effects of social media including news reports, discussion boards, Twitter posts ("tweets"), and micro-blogs on stock markets. For example, Frank and Antweiler [2004] extracted the positive ("bullish") and negative ("bearish") sentiments expressed in Yahoo! Finance postings and concluded that the effect of financial discussion boards on stocks was statistically significant. Schumaker and Chen [2009b] experimented with several textual news representation approaches including bags of words, noun phrases, proper nouns, and name entities to study the effect of breaking news on stock movements. Gilbert and Karahalios [2010] reported that increased levels of anxiety, worry, and fear produced downward pressure on the Standard and Poors (S&P) 500 index. Bollen et al. [2011] examined the use of public sentiment expressed via Twitter posts to forecast stock movements. Xu and Zhang [2013] found that Wikipedia enriched the information environment of investors and affected stock movements. Yu et al. [2013] demonstrated that social media had a stronger relationship to stock performance than did conventional media. Luo et al. [2013] studied the use of social media to predict firm equity values and found that blogs and consumer ratings were the most significant indicators of firm equity value. Zheludev et al. [2014] examined the relation between social sentiments and stock prices using mutual information analysis and Twitter posts. Li et al. [2014a, 2014b] measured public sentiment by observing the interactive behaviors of investors on social media and studied the effect of firm-specific news on stock prices and public sentiment.

Table I summarizes a representative sample of previous studies, indicating their foci, information sources, analysis models, and experimental approaches. The evidence suggests that stock markets are strongly affected by various types of highly interrelated information. However, few studies have examined the joint effects of these information sources on stocks or provided a generic and scalable information framework to systematically study the relationships between stock movements and new information.

2.2. Stock Analysis Models

There are two main approaches to the analysis of the effect of new information on stock movements.

Financial researchers typically apply linear regression models to analyze the causes of stock movements. For example, Fama and French [1993] applied a time-series regression model to study the effects of the overall market, firms, and BE/ME on stock prices. Tetlock et al. [2008] analyzed the relationship between news reports and stock prices using a linear regression model. Yu et al. [2013] studied the effects of social and conventional media on firm equity value using a revised version of the four-factor model proposed by Fama and French. These studies focused on linear correlations between information and stock prices.

Computer science offers alternative methods for analyzing the nonlinear relations between stock movements and new information through various artificial intelligence techniques. For example, Mittermayer and Knolmayer [2006] represented each news report as a term vector. Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) methods were applied to study the effect of news reports on stock movements. Schumaker and Chen [2009b] predicted future stock prices based on breaking news and historical stock prices using a Support Vector Regression (SVR) model. Lavrenko et al. [2000] proposed a Relevance Language Model (RLM) to associate stock price trends with news stories. Wang et al. [2012] studied the relation between news and stocks using a hybrid predictive model based on both ARIMA and SVR. Bollen et al. [2011] measured public sentiment expressed in Twitter posts and used this indicator to forecast stock price trends with Self-Organizing Fusion Neural Networks (SOFNNs). However, these studies all used vector-based approaches that

Table I. Comparison of Representative Studies

	Focus			Model			Experiment					
Literature	Target	Scale	Market	Quantified Input	Qualitative Input	Output	Method	Period	Data Size	Training & Test	Metrics	Cost
Fama and French [1993]	Stock return	Month	NYSE, Amex, NASDAQ	Economic data		Abnor. return	Regression model	1963-1991		No	Real value	No
Chan [2003]	Stock return	Month	NYSE, Amex, S&P500	Economic data	News Number	Abnor. return	Regression model	1980-2000	4200 stocks	No	Real value	No
Tetlock et al. [2008]	Stock return	Day	S&P500	Economic data	Word number in News	Abnor. return	Regression model	1980-2004	500 stocks	No	Real value	No
Wüthrich et al. [1998]	Index	Day	Five Major Indexes		News	Index	KNN, Regression model	12/06/1997-03/06/1997	No	Three months rolling	Trend	No
Lavrenko et al. [2000]	Stock Price	Hour	No		News	Stock price	Language model	10/15/1999-02/10/2000	127 stocks	90/40 days	Trend	No
Mittermayer and Knolmayer [2006]	Stock trend	Minute	S&P500		News	Trend	KNN, SVM	04/01/2002-12/31/2002	500 stocks	90%/10% stocks	Trend	No
Bollen et al. [2011]	Index	Day	NYSE	Past DJIA	Twitter Mood	Present DJIA	SOFNN	02/28/2008-12/19/2008		10/1 months	Real value	No
Wang et al. [2012]	Stock price	Quarter	No	Economic data	News	Stock price	ARIMA, SVR	1994-2010	6 stocks	20/5 stocks	Real value	No
Schumaker et al. [2012]	Stock price	Minute	S&P500	Current stock price	News	Future stock price	SVR	10/26/2005-11/28/2005	500 stocks	4/1 weeks	Real value, Trend	No
Yu et al. [2013]	Stock return	day	US	Economic data	social media	Future stock return	Regression	07/01/2011-09/30/2011	824 stocks	No	Real value	No
Li et al. [2014b]	Stock price	Minute, day, week	CSI100	Current stock price	News, public mood	Future stock price	SVR	01/01/2011-12/31/2011	100 stocks	9/3 months	Real value, Trend	No
this study	Stock price	Minute	CSI100	Economic data	news, public mood	Future stock price	Tensor-based Learning	01/01/2011-12/31/2011	100 stocks	9/3 months	Real value, Trend	No

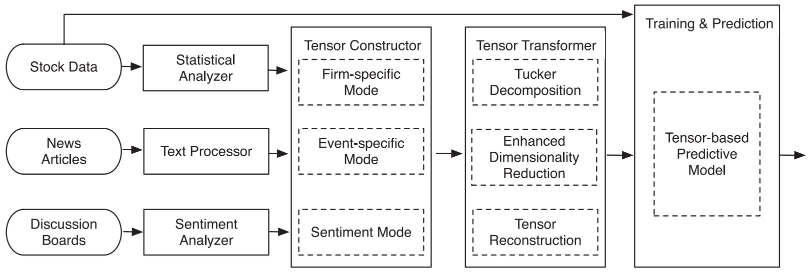


Fig. 2. System overview.

concatenated the features of multiple information sources into one compound feature vector. Such information linearization reduces or ignores the intrinsic associations among the various information sources. In this study, we represent the complex information that is available to investors using tensors, and we apply a tensor-based information framework to capture the nonlinear relation between new information and stock movements. This framework uses two innovative tensor-based approaches, global dimensionality reduction and tensor-based regression learning, to capture and emphasize the static and dynamic relations among several information sources and to study their joint effect on stock movements.

3. TENSOR-BASED INFORMATION FRAMEWORK

In this study, we propose a tensor-based information framework to investigate the effect of information on stock markets, particularly the joint effect of several information sources. Figure 2 presents an overview of the tensor-based information framework. The tensor constructor models the complex information environment that is available to investors in terms of three types of information (i.e., event-specific, firm-specific, and sentiment information). Event-specific information is obtained from daily news articles, and public sentiment with regard to investing is extracted from financial discussion boards. This information and firm-specific data are processed with a tensor transformer to remove noise and capture intrinsic associations among the three information modes. The transformed information tensors are used with a tensor-based model to predict future stock trends.

3.1. Investor Information Modeling

Various information factors that affect stock movements have been extensively studied in the past. Traditional finance focuses predominantly on the long-term effect of firm-specific factors [Cheung and Ng 1992; Fama and French 1993; Jegadeesh and Titman 1993; Dechow 1994], whereas modern behavioral finance concentrates on the short-term effects caused by public sentiment and current events [Frank and Antweiler 2004; Tetlock et al. 2008; Schumaker and Chen 2009b; Gilbert and Karahalios 2010; Bollen et al. 2011; Li et al. 2014b]. It is critical to model the multidimensional relations in this complex information environment to inform investors and to study the joint effect of various information sources on stock trends. In this study, we represent the investor information environment using three types of information, referred to as the firm-specific, event-specific, and sentiment modes.

- *Firm-specific mode:* In essence, a firm's stock price reflects its intrinsic value. We selected five attributes of a company to capture its future business value; each attribute has been shown to have some degree of predictive value in previous studies [Fama and French 1993; Li et al. 2014a, 2014b]. These attributes were the stock

Table II. Description of Notations

Symbol	Description
$\mathbf{x}$	a vector (boldface lower-case letter)
$\mathbf{X}$	a matrix (boldface captial letter)
$\mathcal{X}$	a tensor (script letter)
$I_1, \ldots, I_M$	the dimensionality of mode $1, \ldots, M$
$y_i\|_{i=1}^{M}$	a sequence of M numbers, i.e., $\{y_1, \ldots, y_M\}$
$\mathbf{U}_i\|_{i=1}^{N}, \mathcal{X}_i\|_{i=1}^{N}$	a sequence of N matrices or tensors
$\|\|\mathbf{X}\|\|$	the norm of a matrix $\mathbf{X}$

price, the trading volume, the turnover, the price-to-earnings (P/E) ratio, and the price-to-book (P/B) ratio.

- *Event-specific mode:* Previous studies have demonstrated that news articles play an important role in determining short-term stock movements [Tetlock et al. 2008; Schumaker and Chen 2009b; Li et al. 2014b]. With new information, stock investors are constantly updating their beliefs about the directions of stock prices. The influence of news has two aspects, fundamentals and emotions. News articles enrich investors' knowledge by conveying a more comprehensive view of a firm's financial standing than is provided by a firm's characteristics alone. The optimism or pessimism of news articles may affect the emotions of irrational investors. Therefore, we use news articles as event-specific information factors (Section 3.4). Specifically, each news article is represented using a term vector in which each entry is a weighted noun and sentiment word.
- *Sentiment mode:* With technological advancements that facilitate interactions among users, social media are becoming increasingly popular and provide an important platform to share opinions or feelings among investors. In actuality, investors may be irrational, tending to be influenced by emotions and their peers, which can lead to herd behavior in investments [Long et al. 1990; Shleifer and Vishny 1997]. Previous studies have presented an effective method for capturing social sentiment by tracking variations in the frequency of emotion-related words in social media [Bollen et al. 2011; Yu et al. 2013; Luo et al. 2013; Li et al. 2014b]. In this study, we tracked social sentiment regarding single stocks (stock mood) and the entire market (market mood). The details of the methodology for tracking the market and stock moods are presented in Section 3.4.

To preserve the multifaceted and interrelated characteristics of investor information, we modeled the information using a tensor representation. Investor information at time t is represented with a 3^{rd}-order tensor $\mathcal{X}_t$. Essentially, a tensor is a mathematical representation of a multi-dimensional array. A vector is a first-order tensor, and a matrix is a second-order tensor. More details on tensor algebra can be found in Kolda and Bader [2009]. In the following, we use $\mathbf{x}$ to denote a vector, $\mathbf{X}$ to denote a matrix, and $\mathcal{X}$ to denote a tensor. Table II presents the notation used in this study.

Figure 3 illustrates an example of a 3^{rd}-order tensor, $\mathcal{X}_t \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, representing the three-way relations of firm-specific, event-specific, and sentiment information at time t. The variables I_1, I_2, and I_3 are the numbers of firm-specific features, event-specific features, and sentiment features, respectively. The significance of the elements a_{i_1,i_2,i_3} of tensor $\mathcal{X}_t$ is as follows:

- $a_{i_1,1,1}$, $1 \leq i_1 \leq I_1$, denotes the firm-specific information feature values;
- $a_{2,i_2,2}$, $1 \leq i_2 \leq I_2$, denotes the event-specific information feature values;
- $a_{3,3,i_3}$, $1 \leq i_3 \leq I_3$, denotes the sentiment information feature values; and
- other elements are set to zeros originally.

Thus, investor information can be represented by a sequence of tensors rather than a sequence of concatenated vectors, as was used in previous studies. Each dimension of

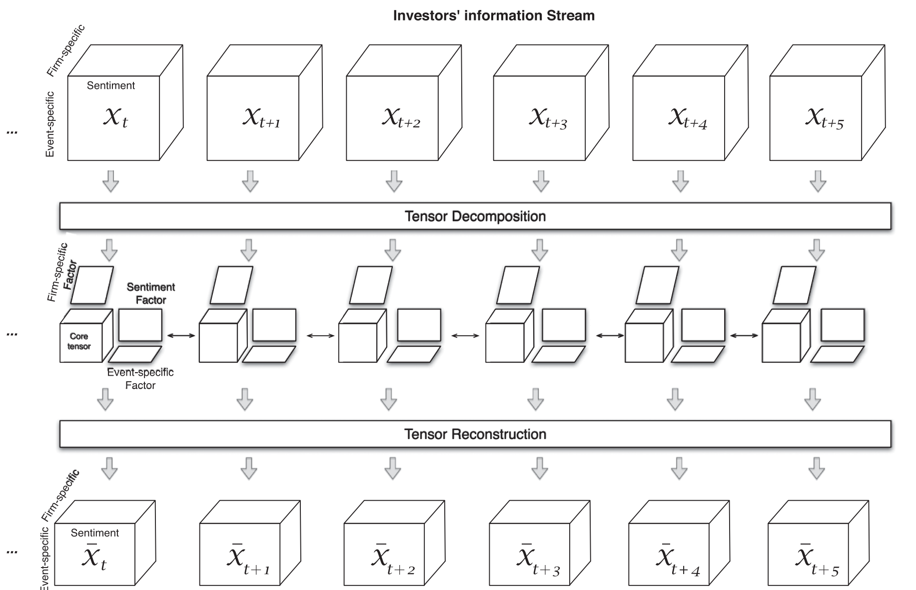


Fig. 3. Investor information environment via tensor representation. A snapshot of information stream at time t is 3^{rd}-mode data tensor $\mathcal{X}_t$. Tensor decomposition and reconstruction is applied to reduce the dimensionality and reinforce the intrinsic links of different information modes.

a tensor represents a subspace of one information mode, and the complementary mode subspaces are able to reflect the multidimensional relations present among them. The corresponding stock trend indicator, namely, the future stock price at time i, is denoted as $y_i|_{i=1}^{N}$.

3.2. Tensor Transformation

Given the tensor representation of investor information, a tensor decomposition and reconstruction technique is applied to reduce the dimensionality and derive the latent relations among the three information modes. The two most popular tensor decomposition methods are CP and Tucker decomposition [Kolda and Bader 2009]. However, both decomposition approaches focus on the geometric structure of a single tensor without considering the entire tensor sequence. Although both methods are able to identify the nonlinear structural information contained in multiple information modes in a tensor, they fail to capture the dynamic connections among multiple modes across a sequence of information tensors.

In this study, we introduce a GDR method that identifies the intrinsic connections among several information sources from the geometric structure of both a single tensor and a sequence of tensors. This identification is achieved by adjusting the factor matrices in the tensor subspace with two additional criteria. Specifically, for two tensors with similar stock prices, their decomposed factor matrices in each tensor subspace are adjusted to be similar. To avoid overadjusting, the variances of these adjusted factor matrices are maximized in each tensor subspace sequence. Figure 4 illustrates the GDR of a tensor sequence, and Table III gives the details of the proposed GDR algorithm.

This algorithm first decomposes a tensor $\mathcal{X}_i$ into $\mathcal{C}_i \times_1 \mathbf{U}_1^i \times_2 \mathbf{U}_2^i \times_3 \mathbf{U}_3^i$ using the Tucker decomposition [Kolda and Bader 2009]. Here, each factor matrix $\mathbf{U}_k^i$ describes one distinct facet of the investor information space (i.e., firm, event, or sentiment information). The core tensor $\mathcal{C}_i$ indicates the strengths of the relations among the three facets represented by a tensor $\mathcal{X}_i$.

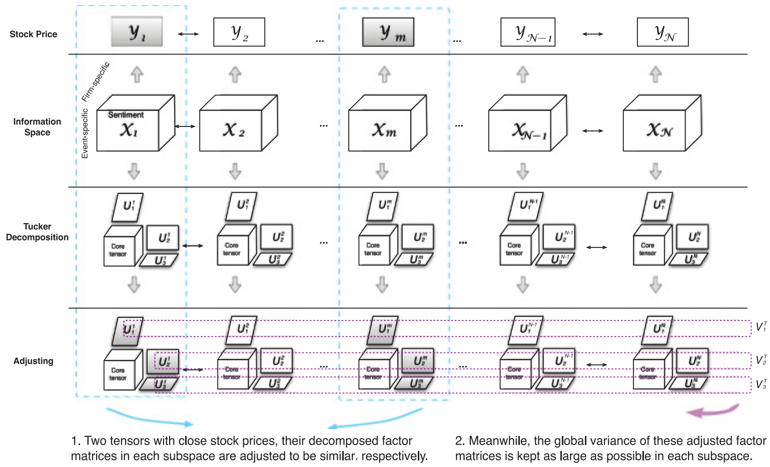


Fig. 4. GDR illustration.

Following Tucker decomposition, $\mathbf{U}_k^i|_{i=1}^N$ is further adjusted to preserve the global geometric structure of each information mode in the tensor sequence. For this purpose, we minimize the following objective function to obtain a correction factor $\mathbf{V}_k \in \mathbb{R}^{I_k \times J_k}$, where $J_k \leq I_k$, to adjust the original factor matrix sequence $\mathbf{U}_k^i|_{i=1}^N$.

$$\min_{\mathbf{V}_k} J(\mathbf{V}_k) = \frac{\sum_{i=1}^N \sum_{j=i}^N \left\| \mathbf{V}_k^T \mathbf{U}_k^i - \mathbf{V}_k^T \mathbf{U}_k^j \right\|^2 w_{i,j}}{\sum_{i=1}^N \left\| \mathbf{V}_k^T \mathbf{U}_k^i \right\|^2 d_{i,i}}. \tag{1}$$

In Equation (1), $d_{i,i}$ are elements of the matrix $\mathbf{D}$, a diagonal matrix in which the diagonal entries are column sums of $\mathbf{W}$ (i.e., $d_{i,i} = \sum_{m=1}^N w_{m,i}$), and $\mathbf{W}$ is a weighting matrix that captures the geometric structure of a tensor sequence $\mathcal{X}_i|_{i=1}^N$. Essentially, $\mathbf{W}$ is an upper triangular matrix in which $w_{i,j}|_{i=1,i\leq j}^N$ indicates the proximity of two tensors in the tensor sequence and $w_{i,j}|_{i=1,i>j}^N$ is zero. Here, the proximity of two tensors is weighted using

$$w_{i,j} = \begin{cases} 1, & \text{if } i \leq j \text{ and } |y_i - y_j|/y_j \leq 5\% \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In preliminary experiments, the performance was relatively robust when the threshold was between 5% and 10%. We set the threshold to 5% without loss of generality. In Equation (2), $||.||$ is the Frobenius norm of the tensor and is defined as

$$||A||^2 = \sum_{i=1}^m \sum_{i=j}^n |a_{i,j}|^2 = trace(AA^T). \tag{3}$$

In Tucker decomposition, the factor matrix $\mathbf{U}_k^i$ preserves only the intrinsic associations among the various information modes within the tensor $\mathcal{X}_i$. To capture the dynamic connections among the various modes over time, the objective function $J(\mathbf{V}_k)$ is optimized to determine a correction factor $\mathbf{V}_k$ to adjust $\mathbf{U}_k^i|_{i=1}^N$. The purpose of the objective function is to minimize the differences in the values of the k-model matrices, whose corresponding stock prices are similar, while attempting to retain the original properties of the matrices in mode k. Specifically, $||\mathbf{V}_k^T \mathbf{U}_k^i - \mathbf{V}_k^T \mathbf{U}_k^j||^2 w_{i,j}$ serves to correct $\mathbf{U}_k^i|_{i=1}^N$ by $\mathbf{V}_k$ to minimize the differences between matrices $\mathbf{U}_k^i$ and $\mathbf{U}_k^j$. The difference between $\mathbf{U}_k^i$ and $\mathbf{U}_k^j$ is measured in term of the variance between their corresponding stock price y_i and y_j as defined in Equation (2). To avoid overadjusting similar factor matrices in $\mathbf{U}_k^i|_{i=1}^N$, it is necessary to retain the original geometric structure of information mode k. That is, the variance of the factor matrices in each mode should be maximized. In general, the variance of a discrete random variable x is defined as

$$var(x) = \sum (x_i - \mu_x)^2 p_i, \qquad \mu = \sum x_i p_i, \tag{4}$$

where μ is the expected value of x, and p_i is the probability.

We assume that $\mathbf{V}_k^T \mathbf{U}_k$ is a random variable in the tensor subspace and that it has a mean of zero. The probability p_i can be estimated from the diagonal matrix $\mathbf{D}$ using spectral graph theory [Chung 1997]. Thus, the weighted variance of the factor matrices in mode k is estimated as

$$var\left(\mathbf{V}_k^T \mathbf{U}_k\right) = \sum_{i=1}^{N} \left\|\mathbf{V}_k^T \mathbf{U}_k^i\right\|^2 d_{i,i}. \tag{5}$$

Let $\mathbf{A}^i = \mathbf{V}_k^T \mathbf{U}_k^i$ be $\mathbf{A}^i$, so the objective function for information mode k can be rewritten as follows:

$$\begin{aligned}
J(\mathbf{V}) &= \frac{\sum_{i=1}^N \sum_{j=i}^N ||\mathbf{A}^i - \mathbf{A}^j||^2 w_{i,j}}{\sum_{i=1}^N ||\mathbf{A}^i||^2 d_{i,i}} \\
&= \frac{\sum_{i=1}^N \sum_{j=i}^N trace(\mathbf{A}^i - \mathbf{A}^j)(\mathbf{A}^i - \mathbf{A}^j)^T w_{i,j}}{\sum_i^N trace(\mathbf{A}\mathbf{A}^{iT}) d_{i,i}} \\
&= \frac{\sum_{i=1}^N \sum_{j=i}^N trace(\mathbf{A}^i\mathbf{A}^{iT} + \mathbf{A}^j\mathbf{A}^{jT} - \mathbf{A}^i\mathbf{A}^{jT} - \mathbf{A}^j\mathbf{A}^{iT}) w_{i,j}}{\sum_i^N trace(\mathbf{A}\mathbf{A}^{iT}) d_{i,i}} \\
&= \frac{trace\left(\mathbf{A}^1\mathbf{A}^{1T} \sum_{j=1}^N w_{1,j} + \mathbf{A}^2\mathbf{A}^{2T} \sum_{j=2}^N w_{2,j} \ldots + \mathbf{A}^N\mathbf{A}^{NT} \sum_{j=N}^N w_{N,j} - \sum_{i=1}^N \sum_{j=i}^N \mathbf{A}^i\mathbf{A}^{jT} w_{i,j}\right)}{trace\left(\sum_i^N \mathbf{A}\mathbf{A}^{iT} d_{i,i}\right)} \\
&= \frac{trace\left(\sum_{i=1}^N \mathbf{A}^i\mathbf{A}^{iT} d_{i,i} - \sum_{i=1}^N \sum_{j=i}^N \mathbf{A}^i\mathbf{A}^{jT} w_{i,j}\right)}{trace\left(\sum_i^N \mathbf{A}\mathbf{A}^{iT} d_{i,i}\right)}.
\end{aligned} \tag{6}$$

Let $\mathbf{D}_U = \sum_{i=1}^N d_{i,i}\mathbf{U}^i\mathbf{U}^{iT}$ and $\mathbf{W}_U = \sum_{i=1}^N \sum_{j=i}^N w_{i,j}\mathbf{U}^i\mathbf{U}^{jT}$. The objective function can then be rewritten as

$$\begin{aligned}
J(\mathbf{V}) &= \frac{trace\left(\sum_{i=1}^N \mathbf{V}^T\mathbf{U}^i\mathbf{U}^{iT}\mathbf{V} d_{i,i} - \sum_{i=1}^N \sum_{j=i}^N \mathbf{V}^T\mathbf{U}^i\mathbf{U}^{jT}\mathbf{V} w_{i,j}\right)}{trace\left(\sum_i^N \mathbf{V}^T\mathbf{U}^i\mathbf{U}^{iT}\mathbf{V} d_{i,i}\right)} \\
&= \frac{trace\left(\mathbf{V}^T\left(\sum_{i=1}^N d_{i,i}\mathbf{U}^i\mathbf{U}^{iT}\right)\mathbf{V} - \mathbf{V}^T\left(\sum_{i=1}^N \sum_{j=i}^N w_{i,j}\mathbf{U}^i\mathbf{U}^{jT}\right)\mathbf{V}\right)}{trace\left(\mathbf{V}^T\left(\sum_i^N d_{i,i}\mathbf{U}^i\mathbf{U}^{iT}\right)\mathbf{V}\right)} \\
&= \frac{trace\left(\mathbf{V}^T\mathbf{D}_U\mathbf{V} - \mathbf{V}^T\mathbf{W}_U\mathbf{V}\right)}{trace\left(\mathbf{V}^T\mathbf{D}_u\mathbf{V}\right)}.
\end{aligned} \tag{7}$$

Without constraints, there are several solutions to this optimization problem. Therefore, we add a normalization constraint, namely, $trace(\mathbf{V}^T\mathbf{D}_u\mathbf{V}) = 1$. The optimization problem is then

$$\begin{aligned}
&\min J(\mathbf{V}) = trace\left(\mathbf{V}^T\mathbf{D}_U\mathbf{V} - \mathbf{V}^T\mathbf{W}_U\mathbf{V}\right) \\
&\text{s.t. } trace\left(\mathbf{V}^T\mathbf{D}_u\mathbf{V}\right) = 1.
\end{aligned} \tag{8}$$

Table III. Algorithm: Global Dimensionality Reduction of a Tensor Stream

Input:	The training tensor stream $\mathcal{X}_i\|_{i=1}^N \in \mathbb{R}^{I_1\times I_2\times I_3}$ and the associated indicators $y_i\|_{i=1}^N \in \mathbb{R}$.
Output:	The mapped tensor steam $\bar{\mathcal{X}}_i\|_{i=1}^N \in \mathbb{R}^{J_1\times J_2\times J_3}$, where $J_k \le I_k$.
Step 1:	Calculate the weight matrix $\mathbf{W}$;
Step 2:	From $k = 1$ to 3
Step 3:	From $i = 1$ to N
Step 4:	Decompose the original tensor $\mathcal{X}_i$ into $\mathcal{C}_i \times_1 \mathbf{U}_1^i \times_2 \mathbf{U}_2^i \times_3 \mathbf{U}_3^i$ by Tucker decomposition;
Step 5:	End
Step 6:	$\mathbf{D}_{U_k} = \sum_{i=1}^N d_{i,i}\mathbf{U}_k^i\mathbf{U}_k^{iT}$;
Step 7:	$\mathbf{W}_{U_k} = \sum_{i=1}^N \sum_{j=i}^N w_{i,j}\mathbf{U}_k^i\mathbf{U}_k^{iT}$;
Step 8:	Obtain $\mathbf{V}_k$ by solving $(\mathbf{D}_{U_k} - \mathbf{W}_{U_k})\mathbf{V}_k = \lambda\mathbf{D}_{U_k}\mathbf{V}_k$;
Step 9:	End;
Step 10:	From $i = 1$ to N
Step 11:	$\bar{\mathcal{X}}_i = \mathcal{C}_i \times_1 (\mathbf{V}_1^T\mathbf{U}_1^i) \times_2 (\mathbf{V}_2^T\mathbf{U}_2^i) \times_3 (\mathbf{V}_3^T\mathbf{U}_3^i)$;
Step 12:	End.

To optimize $J(\mathbf{V})$, we construct a Lagrangian function L and optimize $J(\mathbf{V})$ by taking the partial derivative of L with respect to $\mathbf{V}$ [Bellman 1956].

$$L(\mathbf{V}) = trace\big(\mathbf{V}^T\mathbf{D}_U\mathbf{V} - \mathbf{V}^T\mathbf{W}_U\mathbf{V}\big) - \lambda\big(trace\big(\mathbf{V}^T\mathbf{D}_u\mathbf{V}\big) - 1\big). \tag{9}$$

Because $\mathbf{D}_U^T = \sum_{i=1}^N d_{i,i}(\mathbf{U}^i\mathbf{U}^{iT})^T = \sum_{i=1}^N d_{i,i}\mathbf{U}^i\mathbf{U}^{iT} = \mathbf{D}_U$, and $\mathbf{W}_U^T = \mathbf{W}_U$, it follows that

$$\begin{aligned}\frac{\mathrm{d}L(\mathbf{V})}{\mathrm{d}\mathbf{V}} &= \big(\mathbf{V}^T(\mathbf{D}_U - \mathbf{W}_U)\big)^T + (\mathbf{D}_U - \mathbf{W}_U)\mathbf{V} - \lambda\big(\big(\mathbf{V}^T\mathbf{D}_U\big)^T + \mathbf{D}_U\mathbf{V}\big)\\ &= \big((\mathbf{D}_U - \mathbf{W}_U)^T + \mathbf{D}_U - \mathbf{W}_U\big)\mathbf{V} - \lambda\big(\mathbf{D}_U^T - \mathbf{D}_U\big)\mathbf{V}\\ &= 2(\mathbf{D}_U - \mathbf{W}_U)\mathbf{V} - 2\lambda\mathbf{D}_U\mathbf{V}\\ &= 0.\end{aligned} \tag{10}$$

The correction matrix $\mathbf{V}$ can be calculated from

$$(\mathbf{D}_U - \mathbf{W}_U)\mathbf{V} = \lambda\mathbf{D}_U\mathbf{V}. \tag{11}$$

For information mode k, we calculate $\mathbf{V}_k$ using Equation (11). Thus, tensor $\mathcal{X}_i \in \mathbb{R}^{I_1\times I_2\times I_3}$ is mapped onto a lower-dimensional tensor $\bar{\mathcal{X}}_i \in \mathbb{R}^{J_1\times J_2\times J_3}$, where $\bar{\mathcal{X}}_i = \mathcal{C}_i \times_1 (\mathbf{V}_1^T\mathbf{U}_1^i)\times_2(\mathbf{V}_2^T\mathbf{U}_2^i)\times_3(\mathbf{V}_3^T\mathbf{U}_3^i)$. Table III presents a summary of the GDR algorithm. This algorithm not only identifies the latent relationships among the various information modes in a tensor but also reshapes the factor matrices by propagating and reinforcing their correlations over time.

3.3. Tensor-based Predictive Model

Predicting stock trends using new information is in essence a supervised learning problem. In our approach, the investor information space is modeled with a tensor sequence $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N\}$, and the corresponding stock trend indicator, namely, the future stock price, is denoted as y_i. Our goal is to determine the nonlinear relations between $\mathcal{X}_i|_{i=1}^N$ and $y_i|_{i=1}^N$. This is essentially a high-dimensional regression problem. In the remainder of this section, we first explain the proposed tensor regression learning algorithm using 2^{nd}-order tensors and then extend this algorithm to higher-order tensors. We begin by determining a 2^{nd}-order tensor mapping function $f(\mathcal{X})$ that differs by at most ε deviation from the target y_i (i.e., we will accept differences only if they are less than ε).

Definition 3.1 (*Second-Order-Tensor-Regression Learning Problem*). Given a set of training data $\{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \ldots, (\mathbf{X}_N, y_N)\}$, where the 2^{nd}-order tensor (matrix) $\mathbf{X}_t \in \mathbb{R}^{I_1 \times I_2}$ denotes the input patterns and $y_t \in \mathbb{R}$ is the output that is associated with $\mathbf{X}_t$, determine a 2^{nd}-order tensor mapping function $f(\mathbf{X}) = \mathbf{u}^T\mathbf{X}\mathbf{v} + b$ with $\mathbf{u} \in \mathbb{R}^{I_1}$, $\mathbf{v} \in \mathbb{R}^{I_2}$, and $b \in \mathbb{R}$ having an error no greater than ε from the targets for all training data and with the least complexity.

This definition is analogous to that of SVR models [Smola and Schölkopf 2004]. Essentially, SVR is a special case of our model in which the input data are 1^{st}-order tensors (vectors). The mapping function $f(\mathbf{X})$ can be rewritten using the inner product (i.e., $f(\mathbf{X}) = <\mathbf{X}, \mathbf{u}\mathbf{v}^T> + b$). Consequently, the model complexity can be measured by $||\mathbf{u}\mathbf{v}^T||$. The complexity should be minimal to allow generalization ability to new data [Smola and Schölkopf 2004]. Thus, this regression learning problem is transformed into a convex optimization problem:

$$\min_{\mathbf{u},\mathbf{v},b} J(\mathbf{u}, \mathbf{v}, b) = \frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2$$
$$\text{subject to} \begin{cases} y_i - <\mathbf{X}_i, \mathbf{u}\mathbf{v}^T> - b \leq \varepsilon, \\ <\mathbf{X}_i, \mathbf{u}\mathbf{v}^T> + b - y_i \leq \varepsilon. \end{cases} \tag{12}$$

It is assumed that a mapping function, $f(\mathbf{X})$, that approximates each pair $(\mathbf{X}_i, y_i)$ with an error less than ε exists. To improve generalizability, the model is permitted to tolerate some outliers that have a mapping error greater than ε. Therefore, similar to the "soft margin" loss function employed in SVM models [Cortes and Vapnik 1995], we introduce slack variables ξ_i, ξ_i^* to allow for the presence of some outliers in the training data and minimize the sum of the errors incurred by these outliers. The optimization problem can be expressed as

$$\min_{\mathbf{u},\mathbf{v},b,\xi_i,\xi_i^*} J(\mathbf{u}, \mathbf{v}, b, \xi_i, \xi_i^*) = \frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$\text{subject to} \begin{cases} y_i - <\mathbf{X}_i, \mathbf{u}\mathbf{v}^T> - b \leq \varepsilon + \xi_i, \\ <\mathbf{X}_i, \mathbf{u}\mathbf{v}^T> + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \ldots, N, \end{cases} \tag{13}$$

where C is a positive constant that is used to control the adjustment between the model complexity and the degree to which deviations larger than ε are tolerated. This optimization problem can be solved using an iterative algorithm [Cai et al. 2006]. Let $\mathbf{u} = (1, \ldots, 1)^T$, $\mathbf{x}_i = \mathbf{X}_i^T\mathbf{u}$, and $\beta_1 = ||\mathbf{u}||^2$. Then, $\mathbf{v}$ can be computed by solving the following optimization problem:

$$\min_{\mathbf{v},b,\xi_i,\xi_i^*} J(\mathbf{v}, b, \xi_i, \xi_i^*) = \frac{1}{2}\beta_1||\mathbf{v}||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$\text{subject to} \begin{cases} y_i - \mathbf{v}^T\mathbf{x}_i - b \leq \varepsilon + \xi_i, \\ \mathbf{v}^T\mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \ldots, N. \end{cases} \tag{14}$$

Table IV. Algorithm: Tensor-based Regression Learning

Input:	The training tensor stream $\mathcal{X}_i\|_{i=1}^N \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and the associated indicators $y_i\|_{i=1}^N \in \mathbb{R}$.
Output:	The parameters in tensor function $f(\mathcal{X}) = \mathcal{X} \times_1 \mathbf{W}_1 \times_2 \mathbf{W}_2 \times_3 \mathbf{W}_3 + b$, i.e., $\mathbf{W}_k\|_{k=1}^3 \in \mathbb{R}^{I_k}$, and b, and corresponding slack variables $\xi_i\|_{i=1}^N \in \mathbb{R}$, $\xi_i^*\|_{i=1}^N \in \mathbb{R}$.
Parameters:	User-specified parameters C, ε
Step 1:	Set $\mathbf{W}_k\|_{k=1}^3$ equal to random unit vectors in $\mathbb{R}^{I_k}$;
Step 2:	Do steps 3-7 iteratively until convergence;
Step 3:	From $m = 1$ to 3
Step 4:	Set $\beta_{k,k\neq m} = \|\|\mathbf{W}_k\|\|^2$, $\mathbf{x}_{i,1\leq i\leq N} = \mathcal{X}_i \prod_{1\leq k\leq 3}^{k\neq m} \times_k \mathbf{W}_k$;
Step 5:	Obtain $\mathbf{W}_m$ by optimizing $\min_{\mathbf{W}_k, b, \xi, \xi^*} J(\mathbf{W}_k, b, \xi, \xi^*) = \frac{1}{2} \prod_{1\leq k\leq 3}^{k\neq m} \beta_i \|\|\mathbf{W}_m\|\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$ **s.t.** $\begin{cases} y_i - \mathbf{W}_m^T \mathbf{x}_i - b \leq \varepsilon + \xi_i \\ \mathbf{W}_m^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \ldots, N \end{cases}$
Step 6:	End
Step 7:	Objective function convergence check.
Step 8:	End

Once $\mathbf{v}$ is obtained, let $\beta_2 = ||\mathbf{v}||^2$, and $\hat{\mathbf{x}}_i = \mathbf{X}_i \mathbf{v}$. The vector $\mathbf{u}$ can be obtained by solving the following optimization problem:

$$\min_{\mathbf{u}, b, \xi_i, \xi_i^*} J(\mathbf{u}, b, \xi_i, \xi_i^*) = \frac{1}{2} \beta_2 ||\mathbf{u}||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

$$\text{subject to} \begin{cases} y_i - \mathbf{u}^T \hat{\mathbf{x}}_i - b \leq \varepsilon + \xi_i, \\ \mathbf{u}^T \hat{\mathbf{x}}_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \ldots, N. \end{cases} \tag{15}$$

This iterative procedure to update $\mathbf{u}$ and $\mathbf{v}$ is performed until the objective function converges. Note that the optimization problems presented in Equations (14) and (15) are standard SVR problems. Any computational method for SVR can be used. As mentioned earlier, the GDR algorithm propagates and reinforces the correlations in each information mode across the timeframe while retaining the intrinsic relations among the various information modes in the tensor. In contrast, the iterative optimization procedure used in regression learning iteratively disseminates and strengthens the connections between each pair of information modes. These approaches facilitate our study of the joint effects of multiple information modes on stock prices. Details of the regression learning algorithm can be found in Appendix B.

The iterative optimization method and the regression learning algorithm for 2^{nd}-order tensors (matrices) can be directly extended to 3^{rd}-order or higher-order tensors. Table IV describes the generalized tensor-based regression learning algorithm.

Table V. Features

Information mode	Features
Firm	stock price, trading volume, turnover, P/E ratio, P/B ratio.
Event	noun, sentiment word.
Sentiment	optimistic mood of a stock (M_s^+), market optimistic mood (M^+), pessimistic mood of a stock (M_s^-), market pessimistic mood (M^-), stock sentiment intensity (I_s), market sentiment intensity (I).

3.4. Feature Extraction

The investor information is represented with tensors using three information modes: firm-specific, event-specific, and sentiment information. The features of each information mode used in this study are summarized in Table V. In this section, we first describe the features that reflect the intrinsic value of a firm. Next, we present a method for quantifying news articles to construct the event-specific information mode. In addition, we introduce a method to capture social sentiment from financial discussion boards to form the sentiment information mode.

Firm-specific features reflect the intrinsic value of a listed firm and can be obtained from a professional securities information database. We used the China Stock Market and Accounting Research (CSMAR) database. Based on previous studies [Fama and French 1993; Li et al. 2014a, 2014b], five firm-specific features were used to capture the future business value of a firm:

- *Stock price*: The price of a single share of the publicly traded shares of a company; the stock price is the highest price someone is willing to pay for the stock or the lowest price that it can be bought for.
- *Trading volume*: The number of shares or contracts that are traded in the capital markets during a given period of time; a higher volume for a stock is an indicator of higher liquidity.
- *Turnover*: The total value of stocks traded during a specific period of time; the higher the share turnover is, the more liquid the share of the company.
- *Price-to-earnings (P/E) ratio*: The valuation ratio of a company's current share price compared to its per-share earnings; generally, stocks with higher (or more certain) forecast earnings growth have a higher P/E, and those expected to have lower (or riskier) earnings growth have a lower P/E.
- *Price-to-book (P/B) ratio*: The stock's market value relative to its book value, calculated by dividing the current closing price of the stock by the latest quarter's book value per share; a lower or higher P/B ratio could indicate that the stock is under- or overvalued.

For the event-specific mode, we represent a news article as a term vector in which each entry is a weighted noun or a sentiment word in the article, as in Li et al. [2014b]. This textual representation is referred to as a *bag-of-words* model. These models have been widely used in NLP and information retrieval. Rather than using all of the words in an article, we use only selected nouns and sentiment words because the influence of news articles on stocks is determined by the information that is related to fundamentals and sentiments. We assume that nouns convey the state of a firm's fundamentals and that sentiment words represent the subjective orientation (i.e., tone) of a news article.

In this study, FNLP, a state-of-the-art lexical analysis system for the Chinese language,[1] was applied to extract proper nouns from news articles. Sentiment words were detected according to a finance-specific sentiment word list used by Li et al. [2014b]. After extracting nouns and sentiment words to represent an article as a term vector,

[1]FNLP is developed by Fudan University and accessible at http://code.google.com/p/fudannlp/.

the weight of each term, which indicates its topical importance, was measured using the standard TF/IDF weighting schema used in Baeza-Yates and Ribeiro-Neto [1999] and Li et al. [2010, 2014b].

For the sentiment mode, public moods were extracted from the financial postings in social media. Previous studies have captured public moods from user activities [Szabo and Huberman 2010] and online content [Mishne and De Rijke 2006; Li et al. 2014b]. In this study, we followed the approach proposed by Li et al. [2014b], which has been shown to be effective in capturing investors' moods from social media. In particular, two types of social sentiment were evaluated. One type was popular sentiment regarding single stocks (stock mood) and the other type was popular sentiment regarding the overall market (market mood). Each mood consisted of three features: optimism, pessimism, and sentiment intensity.

An optimistic or pessimistic mood regarding a stock or the market was calculated using the fraction of positive or negative words and adjusted according to the posting importance and publication date, similar to the approach used in Li et al. [2014b]. Specifically, an optimistic mood (M_s^+) for stock s is measured as

$$M_s^+ = \sum_{i=0}^{\tau} \sum_{j=0}^{K} \frac{P_{i,j}}{L_{i,j}} \times W_j \times T_i, \tag{16}$$

where $P_{i,j}$ denotes the number of positive words in an online posting j on the i^{th} day after the date of release, $L_{i,j}$ is the total number of words in posting j on the i^{th} day, W_j represents the importance of posting j, and T_i is the time factor, which reflects the diminishing influence of older postings on sentiments. The posting weight W_j was applied to differentiate the level of influence of the postings. Intuitively, influential postings are those whose contents are read or discussed by a large number of readers. Therefore, we calculated the weight from the number of users viewing a posting (i.e., clicks),

$$W_j = \frac{c_j}{\max_{t'} c_{t'}}, \tag{17}$$

where c_j is the number of clicks for posting j and $max_{t'} c_{t'}$ denotes the largest number of clicks on the day of publication of posting j. Because the influence on public sentiment wanes after several days [Tetlock et al. 2008], the time factor T_i was defined to adjust the degree of influence of a posting with the time elapsed since the date of publication, i.e.,

$$T_i = e^{-i/\beta}, \tag{18}$$

where β is a constant that adjusts the time attenuation scale and was set to 20 to simulate the attenuation for the average number of work days in a month. Consequently, the positive market mood (M^+) is a weighted sum of the optimistic stock mood for the stocks of the firms listed in the market:

$$M^+ = \sum_{i=0}^{k} M_s^+ \times \frac{n_i}{\max_{i'} n_{i'}}, \tag{19}$$

where n_i is the total number of posting for stock s and $max_{i'} n_{i'}$ denotes the largest number of postings regarding stock (i') in the market. A pessimistic mood (M_s^-) for stock s and a negative market mood (M^-) can be defined similarly.

The sentiment intensity is the difference between the optimistic mood and the pessimistic mood. The intensity can be measured as

$$I = \frac{M^{+} - M^{-}}{M^{+} + M^{-}}, \tag{20}$$

$$I_s = \frac{M_s^{+} - M_s^{-}}{M_s^{+} + M_s^{-}}, \tag{21}$$

where I denotes the sentiment divergence of the entire market and I_s is the sentiment divergence for stock s.

4. EXPERIMENTS

The goal of our experiments was to examine the effectiveness of the proposed tensor-based information framework, particularly the ability to capture the joint effect of multiple information sources on stock movements. This tool can improve our understanding of financial markets. Questions of particular interest include the following:

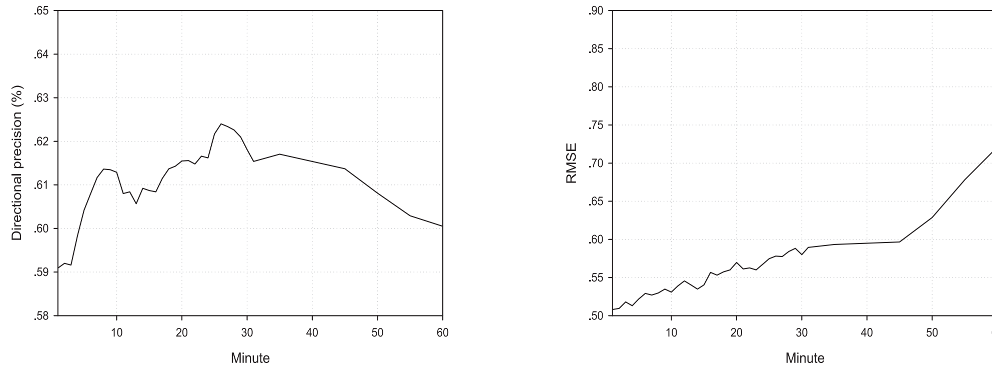
- What are the joint effect of multiple information sources on stock markets? Do multiple sources strengthen or neutralize each other's influence? Clear answers to these questions would provide an important guideline for developing an efficient media-aware trader.
- Typically, social sentiment represents popular attitudes regarding investments, and news articles reflect professional views. What is the combined effect of these potentially conflicting viewpoints? Does social sentiment reduce the effect of asymmetric information in stock markets? Answers to these questions would facilitate a better understanding of investors' behaviors, in particular herd behavior [Avery and Zemsky 1998].

We used data from stock markets in Mainland China. Because market makers are not permitted in Chinese markets, this study was able to evaluate stock movements in response to various types of market information free from the effects of trades by market makers.

4.1. Experimental Setup

In our experiments, we used stock market data provided by Li et al. [2014b]], which we extended with some additional information. In particular, our data consisted of three sets, each of which corresponded to one information mode.

- *Event-specific data set*: News reports can cause price fluctuations in relevant stocks. This dataset contained 124, 470 financial news articles related to the 100 companies listed in the CSI 100. These news articles were gathered from 72 Chinese financial websites during the period between January 1, 2011 and December 31, 2011. This dataset was constructed by querying the search engines Baidu and Google, specifying data ranges and websites to locate news articles with company names, abbreviations, or stock number IDs. The collected news articles were obtained from the 72 reputable Chinese financial websites including `finance.sina.com.cn`, `finance.caixin.com`, and `finance.people.com.cn`. We applied a bloom filter to detect and remove duplicate news articles [Jain et al. 2005] and retained the news articles with company name in the title, which reduced the number of retrieved articles [Tetlock et al. 2008].
- *Social sentiment dataset*: This dataset contained discussion threads from two premier financial discussion boards in China, `www.sina.com` and `www.eastmoney.com`. These two websites have more than 20 million visitors per day, who generate an enormous



(a) Directional accuracy (b) RMSE

Fig. 5. Predictive outlook window.

number of postings, votes, and user clicks. On both websites, each listed firm has its own discussion board. A crawler was applied to collect the postings related to each CSI 100-listed firm posted between January 1 and December 31, 2011.

- *Firm-specific dataset*: This dataset contained the financial statuses of CSI 100 companies between January 1, 2010, and December 31, 2011. The data included stock prices, trading volumes, turnovers, P/E ratios, and P/B ratios.

The data were divided into two sets, a training set and a testing set. The first 9 months of data were used for training the model, and the last 3 months of data were used for the model evaluation and investment experiments. Because the CSI 100 list is updated twice a year, the experiments included only the companies listed during the entire year of 2011. During the evaluation period, the CSI index decreased by 5.21%; specifically, 46.12% of stocks increased, 49.53% decreased, and 4.35% remained unchanged.

4.2. Metrics

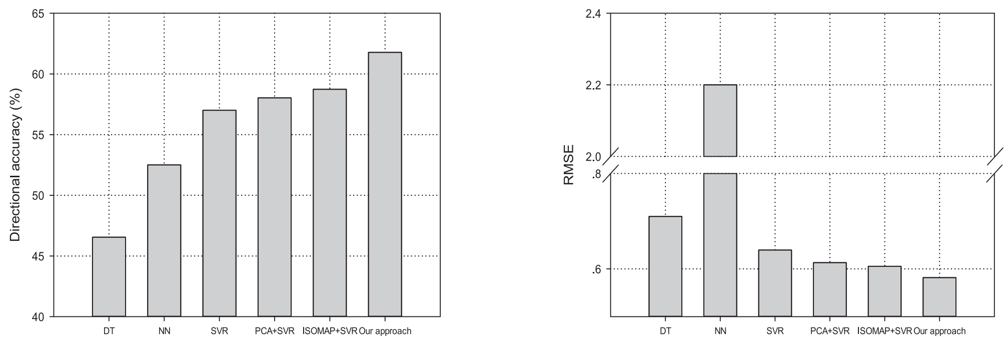
The directional accuracy (DA) and the error in the predicted value are two primary metrics used to evaluate the performance of trading strategies [Schumaker and Chen 2009b]. The DA measures the upward or downward change in the predicted stock price compared to the actual change in the stock price. Because predictions may be close in value but in the wrong direction, an error metric was used to evaluate the difference between the predicted and actual stock prices, the Root Mean Squared Errors (RMSE). The directional accuracy and the RMSE are defined as follows:

$$DA = \frac{S}{N}, \quad RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i - R_i)^2}{N}}, \tag{22}$$

where S is the number of predictions in which the predicted price and the actual stock price have the same movement direction. P_i is the predicted price at the i^{th} prediction, R_i is the actual stock price at the i^{th} prediction, and N is the total number of predictions.

4.3. Prediction Horizon

A theoretical 20-minute time window was used in previous studies. This window implies that the optimal horizon for predicting stock values is approximately 20 minutes following the introduction of new information [Gidofalvi 2001; Li et al. 2014a; Schumaker and Chen 2009a]. As shown in Figure 5, our results indicated that the best predictive performance was achieved at the 26th minute after the release of a news



(a) Directional accuracy (b) RMSE

Fig. 6. Comparison (vector vs. tensor).

item. This finding agrees with previous research, which observed that there is a lag between the time new information is introduced and the time at which the stock market corrects itself to a new equilibrium [LeBaron et al. 1999]. As in Li et al. [2014b], we determined the optimal horizon based on the directional accuracy rather than the RMSE. This choice was made because the investment performance (Section 4.8) relies heavily on the directional accuracy.

4.4. Effectiveness of the Tensor Model

The advantage of the proposed tensor-based framework over previous approaches using concatenated vectors is the ability to model multifaceted factors and their intrinsic relationships. To investigate the effectiveness of the proposed approach, we compared our tensor-based approach with the following vector-based approaches:

- DT: The CART decision tree algorithm[2] was directly applied to the original concatenated vector, which consisted of firm-specific, event-specific, and sentiment information features, to predict future stock prices.
- NN: A Back-Propagation (BP) neural network[3] was directly applied to the original concatenated vector, which consisted of firm-specific, event-specific, and sentiment information features, to predict future stock prices.
- SVR: SVR was directly applied to the original concatenated vector, which consisted of firm-specific, event-specific, and sentiment information features, to predict future stock prices.
- PCA+SVR: The linear dimensionality-reduction technique Principle Component Analysis (PCA) was applied to the original concatenated vector. SVR was used on this reduced-dimension vector to predict future stock prices.
- ISOMAP+SVR: The nonlinear dimensionality-reduction technique ISOMAP was applied to the original concatenated vector. Then, SVR was used on the reduced-dimension vector to predict future stock prices.

Figure 6 shows the performance of these methods in terms of the RMSE and the directional accuracy. Compared with DT and NN, SVR performed significantly better. Furthermore, the performance of PCA+SVR was slightly better than the baseline SVR approach because noise was partially removed by the PCA method. PCA is a statistical

[2]It is implemented using the Scikit-Learn package, accessible at http://scikit-learn.org/stable/.
[3]It is implemented using the Neuroph package for large-scale data, accessible at http://neuroph.sourceforge.net/.

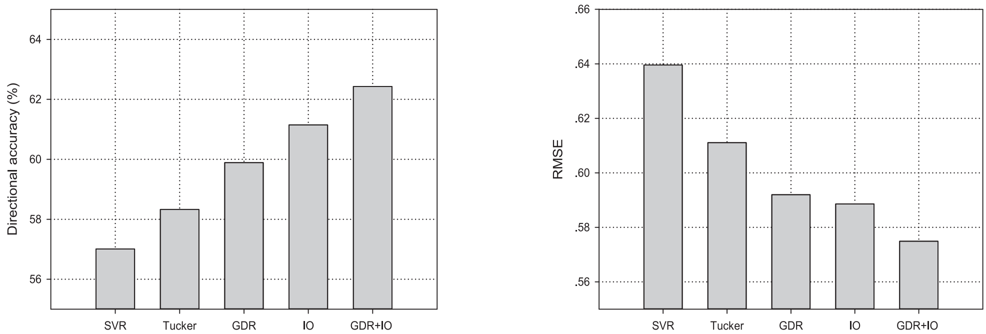
procedure that uses an orthogonal transformation to convert a set of observations of potentially related variables into a set of values of linearly unrelated variables. ISOMAP, which analyzes nonlinear information data, produced better results than PCA because the relation between stock movements and new information is nonlinear. It can be observed that our tensor-based approach outperformed the five vector-based approaches, with an improvement of 9.51% in the directional accuracy and 10.12% in the RMSE when compared with the classic SVR approach. The improved performance is the result of the tensor decomposition and the iterative optimization, which use the intrinsic relations among the various information sources. In addition, we performed t-tests using both directional accuracy and RMSE as the performance measure, and the p-values of these tests were all less than 0.05, so the results of the experiments were statistically significant.

4.5. GDR and IO Effectiveness

In this section, we explore the performance of GDR (Section 3.2) and Iterative Optimization (IO) (Section 3.3) in our tensor-based information framework. GDR organizes the complex information that is available to investors through a balanced consideration of two goals: (1) retaining the static relations among multiple information modes at time t and (2) propagating and strengthening the dynamic correlations within the *same* information mode across the timeframe. In contrast, the IO in the tensor-based regression learning procedure iteratively captures the interconnections between *different* information modes across the timeframe. We tested the following variants of the proposed approach to understand its characteristics:

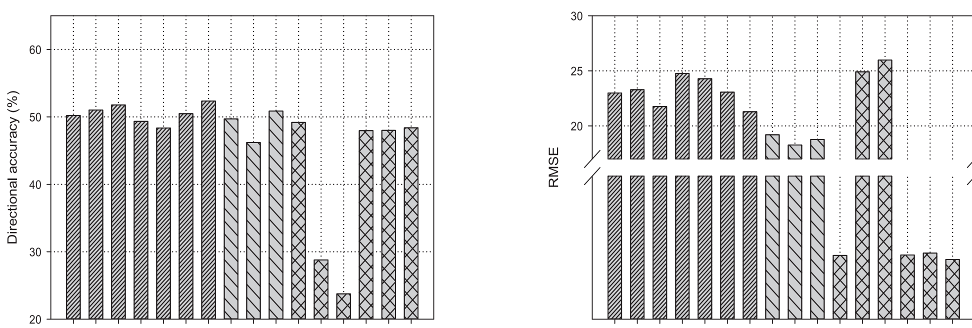
- SVR: SVR was directly applied to the original concatenated vector, which consisted of firm-specific, event-specific, and sentiment information features, to predict future stock prices.
- Tucker: The Tucker decomposition was first applied to reduce the dimension of the original tensor, and the entries of the new tensor were then concatenated into a compound feature vector. SVR was then applied to make predictions based on these concatenated vectors.
- GDR: The original tensors were first reshaped by the global dimensionality-reduction method, and the entries of the new tensor were then concatenated into a compound feature vector. SVR was then applied to make predictions based on these concatenated vectors.
- IO: The original tensor sequence was directly processed by the tensor-based learning model to make predictions without performing global dimensionality reduction.
- Our proposed tensor-based approach (GDR+IO).

Figure 7 shows that both GDR and IO enhanced the intrinsic associations among the various information sources. More importantly, both GDR and IO outperformed the traditional Tucker dimensionality-reduction technique. The combination of GDR and IO achieved the best performance. The p-values for the t-test were all less than the critical confidence value (0.05), indicating that the superior performance of the proposed approach was statistically significant. One reason is that the relations among the various investor information modes are interconnected throughout the timeframe, but the Tucker method captures only the relations at one point in time. Based on this series of experiments, we can conclude that capturing the intrinsic associations among multiple information sources is essential to discover the nonlinear relationships between stock movements and new information. That is, the failure to capture such associations results in the loss of important information and in the underestimation of the joint effects of multiple information sources on stock movements.



(a) Directional accuracy (b) RMSE

Fig. 7. GDR and IO effectiveness.



(a) Directional accuracy (b) RMSE

Fig. 8. Contribution of each information feature.

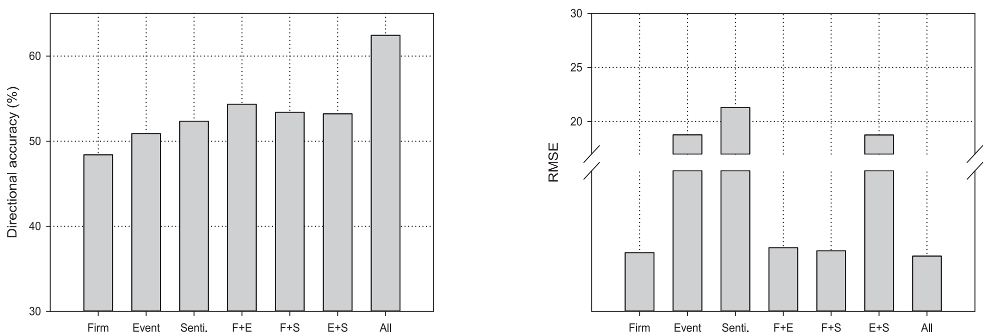
4.6. Level of Investor Information

In the "mosaic" concept of investor information, stock movements are affected by various types of information, each of which consists of various features [Francis et al. 1997]. To some extent, more information leads to a better understanding of financial markets. In contrast, in machine learning more dimensions can inhibit the performance of an algorithm (i.e., the "curse of dimensionality") [Bellman and Dreyfus 1962].

To better understand the role of each feature, we evaluated the individual contribution of each feature in the three information sources. Figure 8 shows the predictive performance of the various features. Each feature contributed to some degree to the prediction accuracy. Generally, all of the features except the trading volume and the turnovers were useful in predicting the stock price direction but not for predicting the future stock value. The P/B and P/E ratios, which are determined by the current stock price, showed the best accuracy in predicting the future stock value. When the features of one information mode were used together, the accuracy was better than that obtained with any individual feature.

To further understand the predictive performance using multiple information sources, we evaluated the joint contribution of the various information types. We considered the following combinations of information modes:

- Firm: Only firm-specific information was used, and the predictive model was SVR.
- Event: Only event-specific information was used, and the predictive model was SVR.



(a) Directional accuracy (b) RMSE

Fig. 9. Contribution of various information sources.

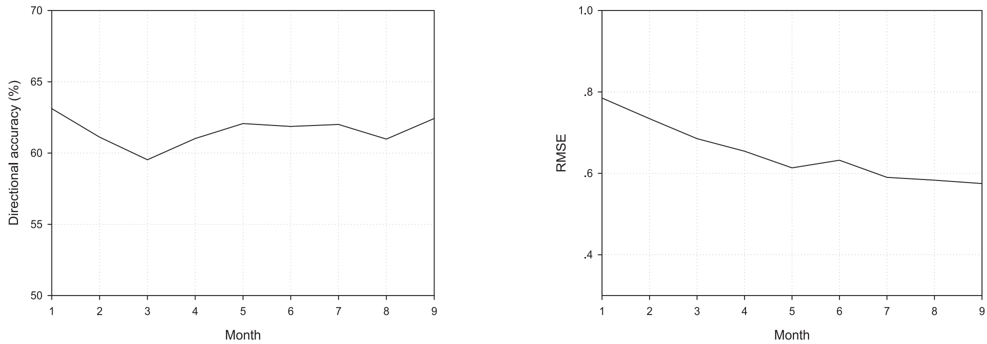
- Sentiment: Only sentiment information was used, and the predictive model was SVR.
- Firm+Event (F+E): Firm-specific and event-specific information were concatenated to form the input vector, and the predictive model was SVR.
- Firm+Sentiment (F+S): Firm-specific and sentiment information were concatenated to form the input vector, and the predictive model was SVR.
- Event+Sentiment (E+S): Event-specific and sentiment information were concatenated to form the input vetor, and the predictive model was SVR.
- All: All three types of information were concatenated into a tensor as the input, and the predictive model was the proposed tensor-based learning algorithm.

Figure 9 shows that the predictive performance increased as more types of information were used. The tensor-based representation that incorporates all three types of information achieved the best performance. A clear understanding of the joint effects can provide a guideline for creating efficient media-aware stock trading algorithms. As shown in Figure 9, the performance using only social sentiment was rather poor, although a statistically significant correlation between social sentiment and stock prices was demonstrated by Zheludev et al. [2014]. An alternative explanation is that extensive mutual information in statistical analysis may not adequately capture the subtle nonlinear relations between stock movements and new information. We observed that the predictability of social sentiment was strengthened along with firm-specific news information. Social sentiment typically reflects popular attitudes toward investments, and firm-specific news represents the insights of professionals regarding stock trends. In our study, these potentially conflicting viewpoints were complementary (i.e., social sentiment was not merely noise but rather valuable information about financial markets). This finding supports the notion that social media can reduce the effect of asymmetric information obtained from news articles on stock markets [Tetlock 2010].

4.7. Tensor-based Model Tuning

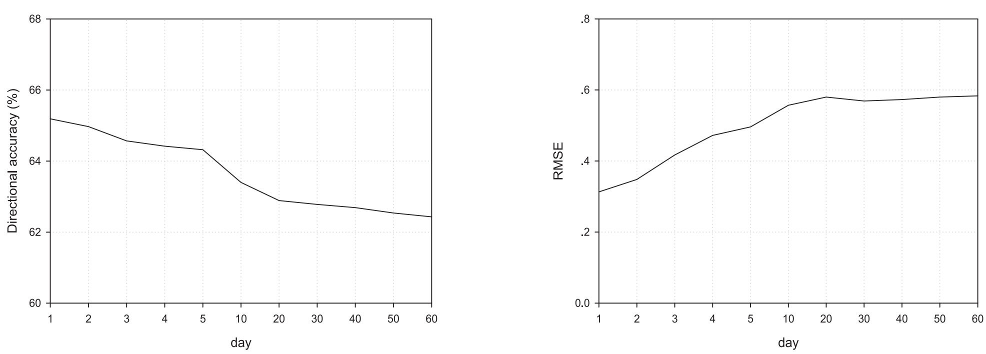
Experimental evaluations in previous studies typically relied on a static training strategy in which a model is obtained via a fixed division of training and testing data [Lavrenko et al. 2000; Mittermayer and Knolmayer 2006; Bollen et al. 2011; Wang et al. 2012; Schumaker et al. 2012; Li et al. 2014b]. In this section, we study the effect of various training strategies on the predictive of the model.

First, we investigated the effect of the amount of training data. Specifically, we first trained the tensor-based model with one month of data (September), and then we incrementally increased the training data by one month at a time in reverse-chronological order. As shown in Figure 10, with more training data, the predictive performance gradually increased in both the directional accuracy and the RMSE.



(a) Directional accuracy (b) RMSE

Fig. 10. Incremental training.



(a) Directional accuracy (b) RMSE

Fig. 11. Retraining.

Next, we studied the effect using a variable ΔT to govern the balance between the training frequency and the prediction performance. That is, we started with a bootstrap of T transaction days as the training data. In each iteration, we added ΔT days to the training data and evaluated the model's performance for the next ΔT transaction days. For example, for $\Delta T = 1$ the model is trained daily, whereas $\Delta T = 5$ specifies that the model is trained every week. Specifically, for a given ΔT ($\Delta T = 1, 2, 3, 4, 5, 10, 20, 30, 40, 50,$ and 60 days in our experiments), we used an iterative rolling test to create each data point shown in Figure 11(a) and (b). In the i^{th} iteration ($i = 1, 2, 3, \ldots$), we used $T + (i-1) \times \Delta T$ transaction days of training data and obtained predictions for days $T + (i-1) \times \Delta T + 1$, $T + (i-1) \times \Delta T + 2, \ldots, T + (i-1) \times \Delta T + \Delta T$. We averaged the directional accuracy and the RMSE for the entire series of a given ΔT. As can be observed in the figures, the system exhibited the best performance when it was trained every day (i.e., $\Delta T = 1$). The performance degraded as ΔT was increased. However, for $\Delta T \leq 5$, the performance degradation was moderate.

4.8. Investment Experiments

We designed and implemented a tensor-based stock information analyzer named TeSIA based on the proposed information framework. In this section, we compare the performance of TeSIA with that of the classic Top-N trading strategy and two state-of-the-art media-aware trading algorithms, AZFinText [Schumaker and Chen 2009b] and eMAQT [Li et al. 2014a].

Similar to a previous study [Li et al. 2014a], we chose RMB 10,000 (approximately USD 1,630) as the investment budget and compared the daily earnings of these approaches over a 3-month evaluation period, during which the CSI 100 index decreased by 5.21% (from 2,363 to 2,240). We assumed zero transaction costs as in previous studies [Chan 2003; Lavrenko et al. 2000; Schumaker et al. 2012; Wang et al. 2012; Li et al. 2014a, 2014b] because the transaction costs can be absorbed by increasing the volume of each transaction provided that the transaction remains profitable.

Top-N is a long-term strategy that assumes that if a combination of stocks has performed well in the past, the same combination will perform well in the near future. We invested in the 30 highest-performing stocks over the period between January 1 and September 30, 2011. The shares were purchased at the beginning of October 2011 and sold at the end of the 3-month evaluation period.

We compared the performance of the proposed TeSIA algorithm with two state-of-the-art media-aware trading algorithms, AZFinText [Schumaker and Chen 2009b] and eMAQT [Li et al. 2014a]. AZFinText applies an SVR model to capture the correlations between financial news and stock prices without considering the public mood. The eMAQT algorithm uses public mood in addition to firm-specific news to predict future stock trends. However, both approaches are based on concatenated feature vectors, which can reduce the intrinsic associations among multiple information modes that may exist in the complex environment of investors information. The proposed TeSIA algorithm is the first tensor-based media-aware trading algorithm that is capable of capturing the intrinsic relationships among several information sources and their joint effects on stock movements.

For these three media-aware algorithms, both selling short and buying long were allowed. Specifically, if a firm-specific news article was released, these algorithms forecasted the stock price for that firm 26 minutes in the future. In buying long, if the difference between the predicted future price and the current stock price was greater than 0.2% of the invested value,[4] then the stock was purchased immediately and sold it 26 minutes later. The investment gain was the spread (difference) between the sale and purchase prices. In selling short, if the spread between the predicted future price and the current stock price was less than 0.1% of the invested value, the stock was borrowed and sold immediately and purchased at the original price after 26 minutes. The investment gain was the stock price at the time the shares were borrowed minus the purchase price. To better understand the media-aware strategy, we computed the ideal investment income (i.e., the income assuming the algorithm was able to predict the future price with 100% accuracy at the time that a firm-specific news report was released). Figure 12 presents the daily return of this perfect strategy, which achieved a total return of RMB 2.8941×10^{14} after three months.

Figure 12 shows the daily investment incomes of these four methods over the 3-month assessment period. Notably, the Top-N method invoked a long-term investing strategy and traded only at the end of the assessment period. The daily income of the Top-N method reflects only the value of its portfolio on that day. For the three media-aware trading systems, the daily income was the sum of all transaction incomes earned on that day. Figure 12 shows that the Top-N method produced a small loss at the end of the 3-month assessment period, even with N set to the optimal value of 30. Compared with the decrease of 5.21% in the CSI 100 index and the 166.11% return of eMAQT, the proposed TeSIA algorithm yielded a remarkable return of 389.10% in three months. Note that this profit came from both buying long and selling short. The

[4]In Chinese stock markets, the long-selling transaction cost is about 0.2% of the invested value and 0.05% for a short-selling transaction.

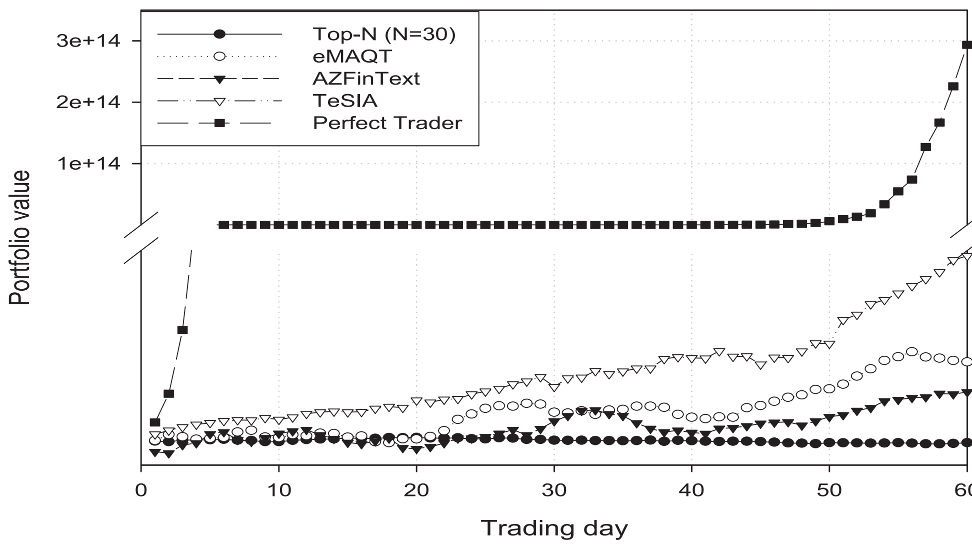


Fig. 12. Investment comparison.

profit from buying long was 150.27%. In contrast, the profit from selling short was 234.36%. Selling short was advantageous because the market trend was downward over the evaluation period.

A randomization test [Lavrenko et al. 2000; Li et al. 2014a] was conducted to determine if the investment income of the TeSIA algorithm was statistically significant. For this purpose, we first implemented a random method that bought or sold stocks randomly. Two constraints were set for the random trader: (1) each stock had the same probability of being traded as for the TeSIA method, and (2) each trading transaction occurred at the same time as for the TeSIA method. In 1,000 trials of the random trader, the investment incomes for 99.5% of the trials were less than that of the TeSIA algorithm (i.e., the performance of the TeSIA algorithm was statistically significant at the 1% level). In addition, a paired t-test showed that with directional accuracy as the performance measure, the TeSIA algorithm performed significantly better than the eMAQT and AZFinText methods at $p = 0.011$ and $p = 0.008$, respectively. This result shows the statistically significant improvement in the performance of the TeSIA algorithm over that of the state-of-the-art media-aware algorithms.

5. CONCLUSION AND FUTURE WORK

The stock market is strongly affected by various types of highly interrelated information, that interact in a complex fashion. Previous studies have typically concatenated the features of various information sources into one compound feature vector, but this step inevitably diminishes the interrelations among the various information sources. In this study, we introduced a tensor-based information framework for predicting stock movements in response to new information. This method is able to identify multi-faceted information factors and their intrinsic relations to stock movements. This capability is achieved through the use of two methods, GDR and IO in tensor-based learning. This combination captures the static and dynamic relations among multiple information sources. Experiments performed using data on CSI 100 stocks for one year demonstrated the importance of the joint effects of multiple information sources. This result supports the “mosaic” theory of investors information in which various types of information are interwoven and interact with each other, and stock markets are affected by the combined effect of these information sources [Francis et al. 1997]. The proposed framework provides a powerful methodology with which financial researchers can understand the behavior of stock markets.

In this study, the textual information for the event-specific mode was represented by isolated words. It would be interesting to investigate the interdependencies between word features to extract more meaningful information, as was done in Mikolov et al. [2013]. We assessed public opinion using financial discussion boards. Given the popularity of social media, it would be of great interest to study public opinion as expressed in various types of social media including micro-blogs, Wikipedia, and blogs. An investigation into the use of these new sources to model the comprehensive information space available to investors is imperative. This additional information presents a scalability challenge for the proposed approach. An alternative method is to solve the suboptimization problems (i.e., Equations (14) and (15)) in the iterative optimization procedure with parallel SVR processing techniques. Parallel SVR processing is a promising approach to addressing the scalability problem [Catanzaro et al. 2008]; however, its effectiveness for the proposed tensor-based information framework has yet to be explored.

Chinese markets impose a “T+1” transaction rule, which forbids purchasing and selling the same stock within one trading day. This constraint is incompatible with our presumption in the investment experiments that investors are able to trade the same stock within 26 minutes. In fact, we sought to investigate the effect of social media on stock movements, and the greatest influence appears to occur within 26 minutes after a news story is released. These volatile stock movements can be predicted in part by capturing the emotional impulses of investors.

More importantly, the tensor-based information framework is generalizable to other multidimensional learning problems in which the information modes of multidimensional data are interconnected and interact over time. Examples include detecting moving objects in video data [Cucchiara et al. 2003], context-aware mobile recommendations [Oku et al. 2006], and health care monitoring [Fleury et al. 2010]. We would be very interested in studying the effectiveness of the proposed tensor-based information framework in these applications.

APPENDICES

A. TENSOR

Tensors can be multiplied together. Here we only consider the tensor mode-m product, i.e., multiplying a tensor by a matrix in mode m.

Definition A.1 (*Mode-m Product*). The mode-m matrix product of a tensor $\mathcal{X} \in \mathbb{R}^{I_1,\ldots,I_N}$ with a matrix $\mathbf{U} \in \mathbb{R}^{J\times I_m}$ is denoted by $\mathcal{X} \times_m \mathbf{U}$ and results in a tensor of size $I_1 \times \cdots \times I_{m-1} \times J \times I_{m+1} \times \cdots \times I_N$. Let element $(i_1, i_2, \ldots, i_N)$ of tensor to be donoted as $\mathcal{X}_{i_1,i_2,\ldots,i_N}$. Elementwise, we have $(\mathcal{X} \times_m \mathbf{U})_{i_1 i_2 \ldots i_N} = \sum_{i_m=1}^{I_m} x_{i_1,i_2\ldots i_N} u_{j i_m}$. A tensor $\mathcal{X} \in \mathbb{R}^{I_1,\ldots,I_N}$ can multiply a sequence of matrices $\mathbf{U}_i|_{i=1}^{N} \in \mathbb{R}^{I_i\times R_i}$ as $\mathcal{X} \times_1 \mathbf{U}_1 \ldots \times_N \mathbf{U}_N$, which can be rewritten as $\mathcal{X} \prod_{i=1}^{N} \times_i \mathbf{U}_i$.

B. JUSTIFICATION OF THE PROPOSED ALGORITHM

The justification of the proposed tensor-based learning approach is provided in this section. First, we prove that $\mathbf{u}$ and $\mathbf{v}$ are dependent on each other, which leads to the iterative solution of the proposed approach. Second, we demonstrate the convergence of the iterative procedure.

B.1. Dependence in the Learning Algorithm

Definition B.1 (*Second-Order Tensor-Regression Learning Problem*). Given a set of training data $\{(\mathbf{X}_1, y_1), (\mathbf{X}_2, y_2), \ldots, (\mathbf{X}_N, y_N)\}$, where the 2^{nd}-order tensor (matrix) $\mathbf{X}_t \in \mathbb{R}^{I_1 \times I_2}$ denotes the input patterns and $y_t \in \mathbb{R}$ is the output that is associated with $\mathbf{X}_t$, determine a 2^{nd}-order tensor mapping function $f(\mathbf{X}) = \mathbf{u}^T \mathbf{X} \mathbf{v} + b$, where $\mathbf{u} \in \mathbb{R}^{I_1}$, $\mathbf{v} \in \mathbb{R}^{I_2}$, and $b \in \mathbb{R}$ that has at most ε deviation from the targets for all training data and simultaneously has as low a model complexity as possible.

The goal is to find a tensor mapping function $f(\mathbf{X})$ that has at most ε deviation from the targets for all training data. The slack variables ξ_i, ξ_i^* are introduced to allow for the presence of some outliers in training data and make the sum of the errors incurred by these outliers as small as possible. Thus, the learning problem is converted to:

$$\begin{aligned} &\min_{\mathbf{u},\mathbf{v},b,\xi_i,\xi_i^*} J(\mathbf{u}, \mathbf{v}, b, \xi_i, \xi_i^*) = \frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ &\text{subject to} \begin{cases} y_i - \mathbf{u}^T \mathbf{X}_i \mathbf{v} - b \leq \varepsilon + \xi_i \\ \mathbf{u}^T \mathbf{X}_i \mathbf{v} + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i, \geq 0 \quad i = 1, \ldots, N. \end{cases} \end{aligned} \tag{23}$$

The key idea is to construct a Lagrangian function from the objective function and the corresponding constraints by introducing a dual set of variables. Therefore, we proceed as follows:

$$\begin{aligned} L = &\frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \mathbf{u}^T\mathbf{X}_i\mathbf{v} + b\right) \\ &- \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \mathbf{u}^T\mathbf{X}_i\mathbf{v} - b\right) - \sum_{i=1}^{N}\eta_i\xi_i - \sum_{i=1}^{N}\eta_i^*\xi_i^*. \end{aligned} \tag{24}$$

Here, L is the Lagrangian and α_i, α_i^*, η_i, η_i^* are Lagrange multipliers. Note that

$$\begin{aligned} \frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 &= \frac{1}{2}trace(\mathbf{u}\mathbf{v}^T\mathbf{v}\mathbf{u}^T) \\ &= \frac{1}{2}(\mathbf{v}^T\mathbf{v})trace(\mathbf{u}\mathbf{u}^T) \\ &= \frac{1}{2}(\mathbf{v}^T\mathbf{v})(\mathbf{u}^T\mathbf{u}). \end{aligned} \tag{25}$$

Thus, L is rewritten as

$$\begin{aligned} L = &\frac{1}{2}(\mathbf{v}^T\mathbf{v})(\mathbf{u}^T\mathbf{u}) + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i\left(\varepsilon + \xi_i - y_i + \mathbf{u}^T\mathbf{X}_i\mathbf{v} + b\right) \\ &- \sum_{i=1}^{N}\alpha_i^*\left(\varepsilon + \xi_i^* + y_i - \mathbf{u}^T\mathbf{X}_i\mathbf{v} - b\right) - \sum_{i=1}^{N}\eta_i\xi_i - \sum_{i=1}^{N}\eta_i^*\xi_i^*. \end{aligned} \tag{26}$$

It follows from the saddle point condition that the partial derivatives of L with respect to the variables $\mathbf{u}$, $\mathbf{v}$, b, ξ_i, and ξ_i^*, must vanish to achieve optimality. This yields the following conditions:

$$\mathbf{u} = \frac{\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{X}_i\mathbf{v}}{\mathbf{v}^T\mathbf{v}}, \quad i = 1, \ldots, N. \tag{27}$$

$$\mathbf{v} = \frac{\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{u}^T \mathbf{X}_i}{\mathbf{u}^T \mathbf{u}}, \quad i = 1, \ldots, N. \tag{28}$$

$$\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0, \quad i = 1, \ldots, N. \tag{29}$$

$$C - \alpha_i - \eta_i = 0, \quad i = 1, \ldots, N. \tag{30}$$

$$C - \alpha_i^* - \eta_i^* = 0, \quad i = 1, \ldots, N. \tag{31}$$

As indicated by Equations (27) and (28), $\mathbf{u}$ and $\mathbf{v}$ are dependent on each other and cannot be solved independently. Therefore, we can apply an iterative approach to solve this problem, as explained in Section 3.3.

B.2. Convergence Proof

In this section, we provide a convergence proof of the iterative computational method described above. This justification is similar to the STM problem in Cai et al. [2006]. Specifically, we have the following theorem:

THEOREM B.1 (CONVERGENCE). *The iterative procedure to solve the optimization problems (14) and (15) will monotonically decrease the objective function (13), and hence the learning algorithm converges.*

PROOF. Define:

$$J(\mathbf{u}, \mathbf{v}) = \frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*).$$

This is solved by the iterative method. Specifically, it first fixes $\mathbf{u}_0$ with the initial value, and solves the optimization problem (14) with a traditional SVR approach to obtain $\mathbf{v}_0$. Likewise, fixing $\mathbf{v}_0$, and solving the optimization problem (15) with a traditional SVR approach to obtain $\mathbf{u}_1$. Since SVR is a convex optimization problem, the solution of SVR is globally optimum. That is, the solutions of Equations (14) and (15) are globally optimum. Therefore, it is a ground truth that

$$J(\mathbf{u}_0, \mathbf{v}_0) \geq J(\mathbf{u}_1, \mathbf{v}_0).$$

With the iterative procedure, we can obtain that

$$J(\mathbf{u}_0, \mathbf{v}_0) \geq J(\mathbf{u}_1, \mathbf{v}_0) \geq J(\mathbf{u}_1, \mathbf{v}_1) \geq J(\mathbf{u}_2, \mathbf{v}_1) \geq \cdots.$$

Since the lower bound of J is 0, the objective function J will be converged with sufficient iterations. □

REFERENCES

Christopher Avery and Peter Zemsky. 1998. Multidimensional uncertainty and herd behavior in financial markets. *American Economic Review* (1998), 724–748.

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. Modern information retrieval. *Addison Wesley Longman Publisher* (1999), 41–44.

Richard Bellman. 1956. Dynamic programming and lagrange multipliers. In *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 42. National Academy of Sciences, 767.

Richard Ernest Bellman and Stuart E. Dreyfus. 1962. *Applied Dynamic Programming*. Rand Corporation.

Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2, 1 (2011), 1–8.

Deng Cai, Xiaofei He, Ji-Rong Wen, Jiawei Han, and Wei-Ying Ma. 2006. Support tensor machines for text categorization. *Report No. UIUCDCS-R-2006-2714* (2006).

Bryan Catanzaro, Narayanan Sundaram, and Kurt Keutzer. 2008. Fast support vector machine training and classification on graphics processors. In *Proceedings of the 25th International Conference on Machine Learning*. ACM, 104–111.

Wesley S. Chan. 2003. Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics* 70, 2 (2003), 223–260.

Yin-Wong Cheung and Lilian K. Ng. 1992. Stock price dynamics and firm size: An empirical investigation. *Journal of Finance* 47, 5 (1992), 1985–1997.

Fan R. K. Chung. 1997. *Spectral Graph Theory*. Vol. 92. American Mathematical Society.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297.

Rita Cucchiara, Costantino Grana, Massimo Piccardi, and Andrea Prati. 2003. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 10 (2003), 1337–1342.

Patricia M. Dechow. 1994. Accounting earnings and cash flows as measures of firm performance: The role of accounting accruals. *Journal of Accounting and Economics* 18, 1 (1994), 3–42.

Robert D. Edwards, John Magee, and W. H. C. Bassetti. 2012. *Technical Analysis of Stock Trends*. CRC Press.

Eugene F. Fama. 1965. The behavior of stock-market prices. *Journal of Business* 38, 1 (1965), 34–105.

Eugene F. Fama and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 1 (1993), 3–56.

Anthony Fleury, Michel Vacher, and Norbert Noury. 2010. SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 274–283.

Jennifer Francis, J. Douglas Hanna, and Donna R. Philbrick. 1997. Management communications with securities analysts. *Journal of Accounting and Economics* 24, 3 (1997), 363–394.

Murray Z. Frank and Werner Antweiler. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59, 3 (2004), 1259–1294.

Gyozo Gidofalvi. 2001. *Using news articles to predict stock price movements*. Department of Computer Science and Engineering, University of California, San Diego (2001).

Eric Gilbert and Karrie Karahalios. 2010. Widespread worry and the stock market. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*. 1–8.

Navendu Jain, Mike Dahlin, and Renu Tewari. 2005. Using bloom filters to refine web search results. In *Proceedings of the 8th International Workshop on the Web and Databases (WebDB)*. ACM, 25–30.

Narasimhan Jegadeesh and Sheridan Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48, 1 (1993), 65–91.

Tamara G. Kolda and Brett W. Bader. 2009. Tensor decompositions and applications. *SIAM Review* 51, 3 (2009), 455–500.

Victor Lavrenko, Matt Schmill, Dawn Lawrie, Paul Ogilvie, David Jensen, and James Allan. 2000. Language models for financial news recommendation. In *Proceedings of the 9th International Conference on Information and Knowledge Management (CIKM)*. ACM, 389–396.

Blake LeBaron, W. Brian Arthur, and Richard Palmer. 1999. Time series properties of an artificial stock market. *Journal of Economic Dynamics & Control* 23, 9–10 (1999), 1487–1516.

Qing Li, Jia Wang, Yuanzhu Peter Chen, and Zhangxi Lin. 2010. User comments for news recommendation in forum-based social media. *Information Sciences* 180 (2010), 4929–4939.

Qing Li, Tiejun Wang, Qixu Gong, Yuanzhu Chen, Zhangxi Lin, and Sa-kwang Song. 2014a. Media-aware quantitative trading based on public web information. *Decision Support Systems* 61 (2014), 93–105.

Qing Li, TieJun Wang, Ping Li, Ling Liu, Qixu Gong, and Yuanzhu Chen. 2014b. The effect of news and public mood on stock movements. *Information Sciences* 278 (2014), 826–840.

Andrew W. Lo and Archie Craig MacKinlay. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1, 1 (1988), 41–66.

James Bradford De Long, Andrei Shleifer, Lawrence Henry Summers, and Robert James Waldmann. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98, 4 (1990), 703–738.

Xueming Luo, Jie Zhang, and Wenjing Duan. 2013. Social media and firm equity value. *Information Systems Research* 24, 1 (2013), 146–163.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *Workshop Proceedings of International Conference on Learning Representations (ICLR)*. https://sites.google.com/site/representationlearning2013/workshop-proceedings.

Gilad Mishne and Maarten De Rijke. 2006. Capturing global mood levels using blog posts. In *Proceedings of AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*. AAAI, 145–152.

Marc-Andre Mittermayer and Gerhard F. Knolmayer. 2006. Newscats: A news categorization and trading system. In *Proceedings of the 6th International Conference on Data Mining (ICDM)*. ACM, 1002–1007.

Kenta Oku, Shinsuke Nakajima, Jun Miyazaki, and Shunsuke Uemura. 2006. Context-aware SVM for context-dependent information recommendation. In *Proceedings of the 7th International Conference on Mobile Data Management*. IEEE, 109.

Michael Rechenthin and W. Nick Street. 2013. Using conditional probability to identify trends in intra-day high-frequency equity pricing. *Physica A: Statistical Mechanics and Its Applications* 392, 24 (2013), 6169–6188.

Robert P. Schumaker and Hsinchun Chen. 2009a. A quantitative stock prediction system based on financial news. *Information Processing & Management* 45, 5 (2009), 571–583.

Robert P. Schumaker and Hsinchun Chen. 2009b. Textual analysis of stock market prediction using breaking financial news: The AZFin text system. *ACM Transactions on Information Systems* 27, 2 (2009), 12:1–12:19.

Robert P. Schumaker, Yulei Zhang, Chun-Neng Huang, and Hsinchun Chen. 2012. Evaluating sentiment in financial news articles. *Decision Support Systems* 53, 3 (2012), 458–464.

Andrei Shleifer and Robert W. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52, 1 (1997), 35–55.

Alex J. Smola and Bernhard Schölkopf. 2004. A tutorial on support vector regression. *Statistics and Computing* 14, 3 (2004), 199–222.

Gabor Szabo and Bernardo A. Huberman. 2010. Predicting the popularity of online content. *Communications of the ACM* 53, 8 (2010), 80–88.

Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, 3 (2007), 1139–1168.

Paul C. Tetlock. 2010. Does public financial news resolve asymmetric information? *Review of Financial Studies* 23, 9 (2010), 3520–3557.

Paul C. Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63, 3 (2008), 1437–1467.

Baohua Wang, Hejiao Huang, and Xiaolong Wang. 2012. A novel text mining approach to financial time series forecasting. *Neurocomputing* 83 (2012), 136–145.

Beat Wüthrich, Vincent Cho, Steven Leung, D. Permunetilleke, K. Sankaran, J. Zhang, and W. Lam. 1998. Daily stock market forecast from textual web data. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2720–2725.

Sean Xin Xu and Xiaoquan Michael Zhang. 2013. Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly* 37, 4 (2013), 1043–1068.

Yang Yu, Wenjing Duan, and Qing Cao. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55 (2013), 919–926.

Ilya Zheludev, Robert Smith, and Tomaso Aste. 2014. When can social media lead financial markets? *Scientific Reports* 4 (2014).

Received December 2014; revised October 2015; accepted October 2015