Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Tensor-Based Learning for Predicting Stock Movements

Qing Li and **LiLing Jiang** and **Ping Li**
Computer Science Department
Southwestern University of Finance and Economics
555 Liutai Ave., Chengdu, 611130, China

Hsinchun Chen
Management Information Systems
Department, University of Arizona
1130 E. Helen St. Tucson, Arizona, 85721-0108, USA

Abstract

Stock movements are essentially driven by new information. Market data, financial news, and social sentiment are believed to have impacts on stock markets. To study the correlation between information and stock movements, previous works typically concatenate the features of different information sources into one super feature vector. However, such concatenated vector approaches treat each information source separately and ignore their interactions. In this article, we model the multi-faceted investors' information and their intrinsic links with tensors. To identify the nonlinear patterns between stock movements and new information, we propose a supervised tensor regression learning approach to investigate the joint impact of different information sources on stock markets. Experiments on CSI 100 stocks in the year 2011 show that our approach outperforms the state-of-the-art trading strategies.

Introduction

Essentially, stock movements are information-driven activities in which new information affects the beliefs of investors and causes fluctuations of stock prices. Traditional finance believes that stock prices are affected by new information randomly. In particular, a stock price is always driven by "unemotional" investors to equal the firm's rational present value of expected future cash flows (Fama 1965). Stock investors are constantly updating their beliefs on the future business value, although they will disagree on the the direction of the company's business value with new information. This will lead to a discrepancy between the actual price and the intrinsic price which causes the stock to wander randomly around its intrinsic value. However, in real financial markets, stock investors are emotional. Empirical studies have shown that stock prices do not completely follow random walks (Lo and MacKinlay 1988). Modern behavioral finance studies attribute non-randomness stock movements to investors' cognitive and emotional biases (DeLong et al. 1990; Nofsinger 2005; Shleifer and Vishny 1997). In spite of the fact that traditional finance and modern behavioral finance are conflicted in the way of how information affects stock markets, they believe that information shapes stock movements.

Financial information can be roughly categorized into *quantified* data and *qualitative* descriptions of firms. Stock analysts, whether the technician or the fundamentalist, rely heavily on the quantified information. In traditional finance, a number of literature examines the effect of quantified market data like firm size, cash flow, book-to-market equity, past return, on stock movements. For example, Dechow (1994) shows that accounting earnings and cash flows help measure the firm performance, as reflected in stock returns. Jegadeesh and Titman (1993) find that stocks with higher returns in the previous twelve months tend to have higher future returns. Chen and Ng (1992) document that there is a stable relations between stock price dynamics and firm size (ME, stock price times number of shares) but the strengths of the relationships change over time. Fama and French (1993) identify three risk factors in the returns on stocks, i.e., overall market, firm size, book-to-market equity (BE/ME, the ratio of the book value of common equity to its market value).

However, the quantified data cannot entirely convey the limitless variety of firms' financial standings. Qualitative information, hidden in the textual descriptions of conventional news and social media, is complementary to quantified data to enrich investors' information environment, especially in social media era. Social media including blogs, tweets/micro-blogs, and discussion boards is updated rapidly and spreads virally at an unprecedented speed, providing first-hand information to investors ahead of formal statistical reports (Luo, Zhang, and Duan 2013). Meanwhile, the adaption of user engagement in social media, such as comments, ratings, votes, and so forth, enables vibrant information creation, sharing, and collaboration among investors. With such rapid information influx and user interactions, decisions of investors tend to be influenced by the emotion of peers and the public. It may well lead to a herd behavior in investment. This is evidenced by the recent behavior finance studies. For instance, Frank and Antweiler (2004) extract the bullish and bearish sentiments of Yahoo! Finance postings, concluding that the effect of financial discussion boards on stocks is statistically significant. Gilbert and Karahalios (2010) report that an increase of anxiety, worry, and fear emotions produces downward pressure on the S&P 500 index. Bollen, Pepe, and Mao (2011) capture the public mood from tweets to forecast stock movements. Yu, Duan, and Cao (2013) show that social media has a stronger relationship with firm stock market performance than conventional media. Luo, Zhang, and Duan (2013) study the predictive relationship between social media and firm equity value, and find that Web blogs and consumer ratings are the two most significant leading indicators of firm equity value in social media. Li et al. (2014a; 2014b) propose a media-aware quantitative trader capturing public mood from interactive behaviors of investors in social media, and study the impact of firm-specific news sentiment on stocks along with this public mood.

Essentially, the information related with markets is multi-faceted and multi-relational. The primary sources (modes) include event-specific, firm-specific, and sentiment information. Such complex information, referred to as a *mosaic* notion of investors' information environment (Francis, Douglas Hanna, and Philbrick 1997), implies the joint influence of different information sources on stock movements. The way to explore such joint impact remains a great challenge in computational investing, which is critical to understand the invisible hand of stock markets. The challenge lies in negotiating the "semantic gap" while mingling these low-level features with the high-level concepts.

The common strategy in previous studies (Lavrenko et al. 2000; Li et al. 2014a; Mittermayer and Knolmayer 2006; Schumaker and Chen 2008; 2009b; Schumaker et al. 2012; Wang, Huang, and Wang 2011; Xu and Zhang 2013; Yu, Duan, and Cao 2013) is to concatenate features of different information sources into one super feature vector, whose high dimensionalities always cause the problem of "curse of dimensionality" (Bellman and Dreyfus 1962). More importantly, from "mosaic" perspective, different information sources are interlaced and interacted to construct the complex investors' information space. With the concatenated vector representation, each vector element is assumed to be independently, and the contextual coocurrence relations between different information sources are somehow weakened, even ignored. For example, two positive news articles about a stock may be textually dissimilar, since nature language is rich and diverse. Contrastingly, the quantified statistics of profits, sales, debt levels, and dividends showing a good investing chance may strengthen the semantic similarities of different words in these two articles. It is necessary to propagate and reinforce these contextual coocurrence relations among different information sources to capture the nonlinear patterns between stock movements and new information.

In this article, we employ the algebra of higher-order tensor to model the multi-faceted investors' information and their intrinsic links. It provides a generalizable and scalable framework to analyze the complex investors' information on stock movements, in which the multi-faceted factors are complementary to each other. To identify the nonlinear patterns between stock movements and new information, we propose a supervised tensor regression learning approach to investigate the joint impact of different information sources on stock markets. It provides a powerful methodology for financial researchers to explore the impact of qualitative information and quantified data on stock movements.

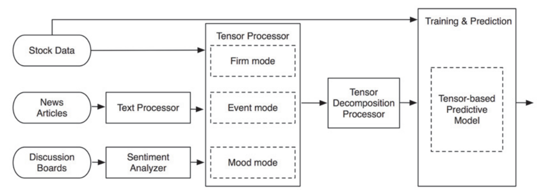

Figure 1: Design Scheme.

System Framework

In this study, we implement a tensor-based stock information analyzer, dubbed TeSlA, to systematically study the information impact on stocks. The framework of TeSlA is sketched in Figure 1. It first represents three types of information sources as tensors. Tucker decomposition is then applied to remove noise and capture intrinsic links of different modes in these tensors. These transformed information tensors are feed into the predictive model for predicting future stock movements.

Investors' Information Modeling

There are various of information factors on stock movements which have been studied extensively before. Traditional finance mainly focuses on the long-term impact of firm-specific factors (Cheung and Ng 1992; Dechow 1994; Fama and French 1993; Jegadeesh and Titman 1993), whereas modern behavioral finance is interested in the short-term shock of sentimental factors and event-specific factors (Bollen, Pepe, and Mao 2011; Frank and Antweiler 2004; Gilbert and Karahalios 2010; Schumaker and Chen 2009b; Tetlock, Saar-Tsechansky, and Macskassy 2008). It is critical to model the complex investors' information space of different information sources, and study their joint impact on stocks. In this study, we construct investors' information space in terms of three different types of information sources, i.e., firm, event, and sentiment. In particular,

- *Firm-specific Mode:* The stock price reflects the intrinsic value of a firm. Investors generally have higher expectations on healthy companies. Here, we select six key characteristics of a company to capture its future business value, each of which shows a predictive ability to some degree in previous literatures (Fama and French 1993; Li et al. 2014a; 2014b). That is, stock price, trading volume, turnover, price-to-earnings (P/E) ratio, price-to-book (P/B) ratio, and industry category.
- *Event-specific Mode:* Stock investors are constantly updating their beliefs about the direction of the market with new information, which leads to fluctuations of stock prices. Recent studies show that news articles play an important role in short-term stock movements (Fama and French 1993; Li et al. 2014a). Therefore, we utilize news articles as the event-specific information factor. Specifically, each news article is represented as a term vector,

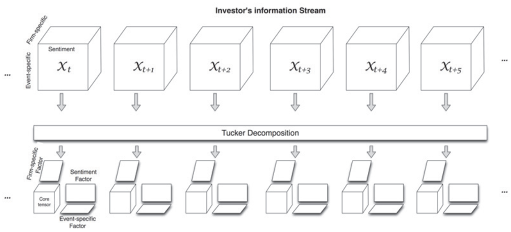

Figure 2: Investors' Information via Tensor Representation. A snapshot of information stream at time t is 3rd-mode data tensor $\mathcal{X}_t$ which can be summarized by a core tensor and three factor matrices via Tucker decomposition.

where each entry is a weighted noun and sentiment word extracted from the article (Li et al. 2014b).

- *Sentiment Mode:* With the popularity of social media, it provides an important platform to share opinions or feelings among investors. In real financial market, irrational investors tend to be influenced by peers, most likely leading to a herd behavior in investing (DeLong et al. 1990). Previous studies put forward an effective way to capture social sentiment by tracking the changes of emotion words in social media (Bollen, Pepe, and Mao 2011). Here, we capture social sentiment in terms of the positive and negative mood of investors as the way proposed by Li et al. (2014b).

The multiple modes of investors' information are complementary in essence, which could interact on each other. In this article, the investors' information environment is modeled in the form of tensor streams, a snapshot of investors' information at time t is represented as order-3 tensor $\mathcal{X}_t$. Essentially, a tensor is a mathematical representation of a multi-way array. A first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three is called 3rd-order tensors. More details of the tensor algebra can be found in Kolda and Bader (2009). Note that we use **x** to denote a vector, **X** denote a matrix, and $\mathcal{X}$ a tensor.

Figure 2 illustrates an example 3rd-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, sentiment information at time t. Here, I_1, I_2, and I_3 are the dimensions of firm-specific features, event-specific features, and sentiment features, respectively. The element values a_{i_1,i_2,i_3} of each information snapshot at time t are defined as:

- $a_{i_1,1,1}$ denotes features of firm-specific information;
- $a_{2,i_2,2}$ denotes features of event-specific information;
- $a_{3,3,i_3}$ denotes features of sentiment information;
- other elements are set to zeros originally.

Thus, investors' information can be represented by a tensor stream instead of a vector stream in traditional approaches. Each order of a tensor represents a subspace of one information mode.

Tensor Decomposition & Reconstruction

Once represent investors' information with tensors, a decomposition technique is applied to derive latent relationships between different information modes. CP and Tucker decomposition are the two most popular tensor decomposition methods (Kolda and Bader 2009). In this article, we apply Tucker decomposition to derive latent relationships inherent in a tensor. Essentially, Tucker decomposition is a form of higher-order PCA. It decomposes a tensor into a core tensor multiplied by a matrix along each mode.

Definition 1 (Tucker Decomposition) *Tucker decomposition of* $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_M}$ *yields a* core tensor $\mathcal{C}$ *of specified size* $R_1 \times \cdots \times R_M$ *and* factor matrices $\boldsymbol{U}_m|_{m=1}^{N} \in \mathbb{R}^{I_m \times R_m}$ *such that*

$$\mathcal{X} \approx \mathcal{C} \prod_{m=1}^{M} \times_m \boldsymbol{U}_m, \tag{1}$$

i.e., the reconstruction error $e = ||\mathcal{X} - \mathcal{C} \prod_{m=1}^{M} \times_m \mathbf{U}_m^T||$ is minimized. Here, $\times_m$ denotes the *mode-m product*.

Figure 2 depicts the Tucker decomposition of the third-order tensor. The third-order tensor $\mathcal{X}$ is decomposed as $\mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$. Here, factor matrices, $\mathbf{U}_1$, $\mathbf{U}_2$, and $\mathbf{U}_3$, describe one distinct facet of the investors' information, i.e., firm, event, and sentiment, respectively. The core tensor, $\mathcal{C}$, indicates the strength of relationships among three facets. Each information mode is represented by an "order" in tensor, which considers the difference of scale. During the tensor decomposition, the hidden compensations and interactions between different modes are emphasized and strengthened. It overcomes the weakness of the concatenated feature vectors in previous studies which ignores not only the scale difference of different modes but also their correlations and interactions.

After decomposition, we reconstruct a new tensor, $\hat{\mathcal{X}}$, which is able to reveal the latent information relationship of firms, events, sentiment in the form of new entries. The reconstruction transformation is equivalent to the tensor operation $\hat{\mathcal{X}} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3$ with the computational cost $O(N)$ where $N = I_1 + I_2 + I_3$. Let A and $\hat{A}$ be the set of tenor entires in $\mathcal{X}$ and $\hat{\mathcal{X}}$, respectively. The reconstructed tensor $\hat{\mathcal{X}}$ consists of a set of triplets $\hat{a}(i,j,k) \in \hat{A}$, where $A \subset \hat{A}$. Each $\hat{a}(i,j,k)$ indicates the intrinsic relations among the information related with firm, event, and sentiment. Figure 3 shows a simplified example of this transformation, which obtains the enhanced knowledge to identify the interactions and correlations between different information modes. The original tensor $\mathcal{X} \in \mathbb{R}^{3 \times 3 \times 5}$ is decomposed into one core tensor $\mathcal{C} \in \mathbb{R}^{2 \times 2 \times 2}$, and three factor matrices of $\mathbf{U}_1 \in \mathbb{R}^{3 \times 2}$, $\mathbf{U}_2 \in \mathbb{R}^{3 \times 2}$, and $\mathbf{U}_3 \in \mathbb{R}^{5 \times 2}$, when we choose 2 as the reduction size. The reconstructed tensor $\hat{\mathcal{X}}$ is derived by multiplying the core tensor and three factor matrices. It can be observed that tensor decomposition and reconstruction has updated the value for each existing entries indicating its importance and identify some new entries showing the latent relationships.

Tensor-based Supervised Learning

Predicting stock movements is essentially a supervised learning problem. In this study, investors' information is

Original Tensor $\mathcal{X}$

Firm	Event	Sentiment	Value
1	1	1	1
2	1	1	1
3	1	1	1
2	1	2	1
2	2	2	1
2	3	2	1
3	3	1	1
3	3	2	1
3	3	3	1
3	3	4	1
3	3	5	1

Decomposition

Reconstruction

Recontructed Tensor $\hat{\mathcal{X}}$

Firm	Event	Sentiment	Value
1	1	1	0.28
1	1	2	0.18
1	3	2	0.33
2	1	1	0.83
2	1	2	0.83
2	1	3	0.38
2	1	4	0.38
2	1	5	0.38
2	2	1	0.18
2	2	2	0.28
2	3	2	0.86
2	3	3	0.15
2	3	4	0.15
2	3	5	0.15
3	1	1	0.86
3	2	1	0.32
3	3	1	1.24
3	3	2	1.24
3	3	3	0.57
3	3	4	0.57
3	3	5	0.57

Figure 3: Example of Tensor Transformation.

modeled with a tensor stream $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_N\}$, and the corresponding stock trend indicator such as stock earning, index, and price at time i is denoted as y_i. Our goal is to find the hidden patterns between $\mathcal{X}_i$ and y_i. It can be converted to a high-dimensional regression problem. In particular, find a tensor function $f(\mathcal{X})$ that has at most ε deviation from the actually obtained targets y_i for all the training data. That is, we do not care about errors as long as they are less than ε, but will not accept any deviation larger than this. This definition is analogous to the support vector regression (SVR) (Smola and Schölkopf 2004). Essentially, SVR is a special case of our study in which the input data is 1st-order tensor (vector). In following part, we first explain the proposed supervised learning algorithm in the form of 2nd-order tensor, and extended it to higher-order tensor.

Definition 2 (Learning problem) *Given a set of training data $\{(\boldsymbol{X}_1, y_1), (\boldsymbol{X}_2, y_2), \ldots, (\boldsymbol{X}_N, y_N)\}$, where 2nd-order tensor (matrix) $\boldsymbol{X}_t \in \mathbb{R}^{I_1 \times I_2}$ denotes the input, and $y_t \in \mathbb{R}$ is the output associated with $\boldsymbol{X}_t$, find a 2nd-order tensor mapping function $f(\boldsymbol{X}) = \boldsymbol{u}^T \boldsymbol{X} \boldsymbol{v} + b$, where $\boldsymbol{u} \in \mathbb{R}^{I_1}$, $\boldsymbol{v} \in \mathbb{R}^{I_2}$, and $b \in \mathbb{R}$, that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time the complexity of the model is as low as possible.*

Here, the model complexity is measured by $||\mathbf{u}\mathbf{v}^T||$ (Smola and Schölkopf 2004). Instead of limiting the function $f(\mathbf{X})$ actually exists that approximates all pairs $(\mathbf{X}_i, y_i)$ with ε deviation, we introduce slack variables ξ_i, ξ_i^* to allow mapping errors. Therefore, we can write this problem as a convex optimization problem:

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

where C is a positive constant parameter used to control the tradeoff between the model complexity and the amount up to which deviations larger than ε are tolerated. To solve this optimization, the key idea is to construct a Lagrange function from the objective function and the constraints by introducing a dual set of variables. Therefore, we proceed as follows:

$$\begin{aligned} L = &\frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\eta_i\xi_i - \sum_{i=1}^{N}\eta_i^*\xi_i^* \\ &- \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i + \mathbf{u}^T\mathbf{X}_i\mathbf{v} + b) \\ &- \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{u}^T\mathbf{X}_i\mathbf{v}b). \end{aligned} \tag{3}$$

Here, L is the Lagrangian and α_i, α_i^*, η_i, η_i^* are Lagrange multipliers. Note that: $\frac{1}{2}||\mathbf{u}\mathbf{v}^T||^2 = \frac{1}{2}(\mathbf{v}^T\mathbf{v})(\mathbf{u}^T\mathbf{u})$. Thus, L is rewritten as

$$\begin{aligned} L = &\frac{1}{2}(\mathbf{v}^T\mathbf{v})(\mathbf{u}^T\mathbf{u}) + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\eta_i\xi_i \\ &- \sum_{i=1}^{N}\eta_i^*\xi_i^* - \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{u}^T\mathbf{X}_i\mathbf{v} - b) \\ &- \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i + \mathbf{u}^T\mathbf{X}_i\mathbf{v} + b). \end{aligned} \tag{4}$$

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, ξ_i^*) have to vanish for optimality. This gives the conditions:

$$\mathbf{u} = \frac{\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{X}_i\mathbf{v}}{\mathbf{v}^T\mathbf{v}} \quad i = 1, \cdots, N. \tag{5}$$

$$\mathbf{v} = \frac{\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{u}^T\mathbf{X}_i}{\mathbf{u}^T\mathbf{u}} \quad i = 1, \cdots, N. \tag{6}$$

As shown in Equations (5) and (6), $\mathbf{u}$ and $\mathbf{v}$ are dependent on each other, and can not be solved independently. Therefore, we can apply the iterative approach to solve this problem. In particular, first, let $\mathbf{u} = (1, \ldots, 1)^T$, $\mathbf{x}_i = \mathbf{X}_i^T\mathbf{u}$, and $\beta_1 = ||\mathbf{u}||^2$, $\mathbf{v}$ can be computed by solving the following optimization problem:

$$\min_{\mathbf{v},b,\xi_i,\xi_i^*} J(\mathbf{v}, b, \xi_i, \xi_i^*) = \frac{1}{2}\beta_1||\mathbf{v}||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$\text{subject to} \begin{cases} y_i - \mathbf{v}^T\mathbf{x}_i - b \leq \varepsilon + \xi_i, \\ \mathbf{v}^T\mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \cdots, N. \end{cases} \tag{7}$$

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

$$\min_{\mathbf{u},b,\xi_i,\xi_i^*} J(\mathbf{u}, b, \xi_i, \xi_i^*) = \frac{1}{2}\beta_2||\mathbf{u}||^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$\text{subject to} \begin{cases} y_i - \mathbf{u}^T\hat{\mathbf{x}}_i - b \leq \varepsilon + \xi_i, \\ \mathbf{u}^T\hat{\mathbf{x}}_i + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i^*, \xi_i \geq 0, \quad i = 1, \cdots, N. \end{cases} \tag{8}$$

Table 1: Tensor-based Learning Algorithm

Input:	Tensor stream $\mathcal{X}_i\|_{i=1}^N \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ 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 $\xi_i\|_{i=1}^N \in \mathbb{R}$, $\xi_i^*\|_{i=1}^N \in \mathbb{R}$.
Step 1:	Set $\mathbf{W}_k\|_{k=1}^3$ equal to random unit vectors;
Step 2:	Do steps 3-7 iteratively until convergence;
Step 3:	From $m = 1$ to 3
Step 4:	Set $\beta_{k,k \neq m} = \|\|\mathbf{W}_k\|\|^2$, $\mathbf{x}_{i,1 \le i \le N} = \mathcal{X}_i \prod_{1 \le k \le 3}^{k \neq m} \times_k \mathbf{W}_k$;
Step 5:	Obtain $\mathbf{W}_m$ by optimizing $\min_{\mathbf{W}_k, b, \xi, \xi^*} J(\mathbf{W}_k, b, \xi, \xi^*) =$ $\frac{1}{2} \prod_{1 \le k \le 3}^{k \neq m} \beta_i \|\|\mathbf{W}_m\|\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$ $\text{s.t.} \begin{cases} y_i - \mathbf{W}_m^T \mathbf{x}_i - b \le \varepsilon + \xi_i \\ \mathbf{W}_m^T \mathbf{x}_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i \ge 0, \quad i = 1, \cdots, N \end{cases}$
Step 6:	End
Step 7:	Convergence Checking
Step 8:	End

Note that optimization problem (7) and (8) can be solved by a standard SVR algorithm. Any computational method for SVR can also be used here. This iterative procedure to update **u** and **v** is carried on until the objective function converges.

While Tucker decomposition, the relations of different information mode (firm, event, sentiment) of a tensor are reinforced (vertical compensation). During the iterative optimization, the intrinsic links of different modes are strengthened across the timeline (horizontal interaction). This allows us to study the joint impact of different information modes on stock movements. With the solution of iterative optimization, the learning algorithm in the form of 2nd-order tensor can be straightly extended to 3rd-order tensor or even the higher order tensor. The generalized tensor-based regression learning algorithm is given in Table 1.

Experimental Evaluation

The ultimate goal of this study is to examine the effectiveness of the proposed tensor-based learning approach to capture the joint impact of different information modes on stock movements. In our experiments, we use the stock data generously provided by Li et al. (2014b). It consists of three parts:

- *Financial News*: It contains $124,470$ financial news articles related with 100 companies listed in China Securities Index (CSI 100).
- *Discussion Board*: It contains the discussion threads of CSI100 companies during *Jan.1, 2011* and *Dec.31, 2011* from two premier financial discussion boards in China, i.e., Sina.com and EastMoney.com.

Table 2: Comparison (Vector vs. Tensor)

Method	RMSE	Direction
SVR	0.6396	57.01%
PCA+SVR	0.6132	58.03%
ISOMAP+SVR	0.6054	58.74%
Our tensor-based approach	0.5818	61.78%

- *Stock Data*: It contains the high-frequency financial data during *Jan.1, 2011* and *Dec.31, 2011*. It provides intraday transaction information including price, volume and time in the second-level.

In our experiments, we used the data from the first 9 months of 2011 as a training corpus and the last 3 months of 2011 for testing. We removed several companies from the available 100 companies due to inconsistencies and abnormalities. In our testing period, the upward trend was 46.12%, the downward trend was 49.53%, and the remaining percentage was still. The standard deviation of the stock prices in this testing period was 27.12. Here, closeness and directional accuracy are used as the evaluation metrics. Directional accuracy measures the percentage of price forecast with right directions in the total forecast. Root Mean Squared Error (RMSE) between the predicted value and the real stock price is used as the closeness metric.

Time Window of Prediction

There exists a time window to foreseen the direction of a stock with the release of new information (Chan 2003; Gidofalvi 2001). A "20-minute" theory shows that an optimal outlook time window to sense stock movement is approximately 20 minutes after introducing new information (Gidofalvi 2001; Li et al. 2014a; Schumaker and Chen 2009a). In our study, we also observe the "20-minute" phenomena, and find that the best predictive performance is achieved while predicting +26-minute future prices after news releases. This finding indeed agrees with the previous research that reported the existence of lag time between information introduction and stock market correction to equilibrium (LeBaron, Arthur, and Palmer 1999).

Joint Impact of Investors' Information

Comparing with previous concatenated vector approaches, the advantage of the proposed tensor-based learning algorithm is able to model the multi-faceted factors and their intrinsic links of the complex investors' information. To investigate the effect of the proposed approach, we compare our tensor-based approach with the following vector-based approaches:

- SVR: SVR is directly applied to the original concatenated vector which consists of firm-specific, event-specific, and sentiment information features.
- PCA+SVR: PCA is firstly applied to the original concatenated vector to reduce the vector dimension, and then SVR is performed on the dimension-reduced vector.

Table 3: Vertical & Horizontal Compensation

Method	RMSE	Direction
Without iterations	0.6111	58.33%
Without Tucker	0.5886	61.15%
Our tensor-based approach	0.5818	61.78%

- ISOMAP+SVR: ISOMAP is firstly applied to the original concatenated vector to reduce the vector dimension, and then SVR is performed on the dimension-reduced vector.

Table 2 shows the prediction results of these methods in terms of RMSE and directional accuracy. The performance of PCA+SVR is a little better than the classic SVR approach, since a certain amount of noise has been removed by PCA. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. Comparing with PCA, ISOMAP targeted for nonlinear information data gains better result than PCA in our test. It can be observed that our tensor-based approach outperforms the other three vector-based approaches, with a directional accuracy improvement of 8.37% and RMSE improvement of 9.04% than the classic SVR approach. Such gains come from the tensor decomposition and iterative optimization, which utilize the intrinsic links of different information sources for predicting.

Vertical & Horizontal Compensation

To well understand how information of different modes interacts, we explore the power of the vertical compensation and the horizontal interaction in our proposed approach. Specifically, Tucker decomposition and reconstruction vertically capture the relations of different information modes at time t. Whereas, the iterative optimization horizontally diffuses the interconnections of different modes across the timeline. Here, we study the variants of the proposed approach to understand the inner functions:

- Without iterations: the original tensor is decomposed and reconstructed as a relation-reinforced tensor. Entries of this new tensor are concatenated into a vector. SVR is applied for making predictions.
- Without Tucker: the original tensor steam is directly feed into the tensor-based learning algorithm for prediction without Tucker decomposition.

Table 3 shows that both vertical and horizontal compensation procedures contribute to the relation enhancement of different information modes. The horizontal compensation plays a little more important role than the vertical compensation. That is, the relations of different investors' information modes are interlaced and mingled more efficiently across the timeline than at a static time point.

Investment Experiments

In this study, we design and implement a tensor-based stock information analyzer, TeSIA. We compare TeSIA with two classic trading strategies, i.e., Top-N and simple moving average (SMA) (James 1968), and one state-of-the-art media-aware trader, AZFinText (Schumaker and Chen 2009b). We set RMB10,000 (approximately USD1,630) as the investment budget and compare the daily earnings of these approaches in our 3-month evaluation period, during which the CSI Index was down by 5.21% from 2,363 to 2,240.

In our experiments, even with the optimal top-30 combination, a small loss is still experienced for Top-N approach. Different from the long-term strategy Top-N, SMA focuses on short-term transactions. The SMA strategy is triggered when an actual market stock price crosses through the daily moving average of the same stock. There is no positive earnings with SMA within the 3-month assessment time. AZFinText (Schumaker and Chen 2009b) is a media-aware trading system as TeSIA. It applies SVR model to capture the correlation between financial news and stock prices. Comparing the change of -5.21% in CSI100 and the 103.23% return in AZFinText, the proposed TeSIA yields a remarkable return of 235.20% in three months.

Conclusion and Future Work

Fama's "Efficient Market Hypothesis" (Fama 1965) reveals that new information shapes stock markets, and paves the way for his Nobel Prize in 2013. Following Fama's approach, linear regression models are generally adopted to examine the correlation between stocks and information. With the advent of natural language processing and machine learning techniques, it allows us to investigate the nonlinear patterns between information and stock movements. A common strategy in these approaches is to concatenate the features of different information sources into one super vector, which breaks the intrinsic links between different information sources. This work is a pilot study to model the multi-faceted investors' information and their intrinsic links to explore their joint impacts on stocks. The proposed tensor-based modeling and learning approaches are generalizable and scalable to incorporate any new information source. It provides a powerful methodology for financial researchers to understand the "invisible hand " of stock markets.

The investment experiment on CSI 100 stocks shows a promising earning return of the proposed approach in the year of 2011. It is quite interesting to explore the predictive power of TeSIA in other time windows to check its robustness. In addition, as the popularity of social media, the predictability of various kinds of social media including micro-blogs, wikipedias, and blogs are of great necessity to be investigated. With these extra information sources, it also brings a scalability challenge to the proposed approach. The paralleled SVR processing is a promising way to deal with the scalability problem (Catanzaro, Sundaram, and Keutzer 2008). However, its effectiveness with the proposed tensor-based framework is yet to be explored in our feature work.

Acknowledgments

This work has been supported by the National Natural Science Foundation of China (NSFC) (60803106, 61170133), and the Program for Sichuan National Science Foundation for Distinguished Young Scholars (2013JQ0004).

References

Bellman, R. E., and Dreyfus, S. E. 1962. *Applied dynamic programming*. Rand Corporation.

Bollen, J.; Pepe, A.; and Mao, H. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.

Catanzaro, B.; Sundaram, N.; and Keutzer, K. 2008. Fast support vector machine training and classification on graphics processors. In *Proceedings of the 25th international conference on Machine learning*, 104–111. ACM.

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

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

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

DeLong, J. B.; Shleifer, A.; Summers, L. H.; and Waldmann, R. J. 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98(4):703–738.

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

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

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

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

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

Gilbert, E., and Karahalios, K. 2010. Widespread worry and the stock market. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 59–65.

James, F. E. 1968. Monthly moving averages - an effective investment tool? *Journal of Financial and Quantitative Analysis* 3(3):315–326.

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

Kolda, T. G., and Bader, B. W. 2009. Tensor decompositions and applications. *SIAM review* 51(3):455–500.

Lavrenko, V.; Schmill, M.; Lawrie, D.; Ogilvie, P.; Jensen, D.; and Allan, J. 2000. Language models for financial news recommendation. *Proceedings of the 9th International conference on Information and Knowledge Management (CIKM)* 389–396.

LeBaron, B.; Arthur, W. B.; and Palmer, R. 1999. Time series properties of an artificial stock market. *Journal of Economic Dynamics & Control* 23(9-10):1487–1516.

Li, Q.; Wang, T.; Gong, Q.; Chen, Y.; Lin, Z.; and Song, S.-k. 2014a. Media-aware quantitative trading based on public web information. *Decision Support Systems* 61:93–105.

Li, Q.; Wang, T.; Li, P.; Liu, L.; Gong, Q.; and Chen, Y. 2014b. The effect of news and public mood on stock movements. *Information Sciences* 278:826 – 840.

Lo, A. W., and MacKinlay, A. C. 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies* 1(1):41–66.

Luo, X.; Zhang, J.; and Duan, W. 2013. Social media and firm equity value. *Information Systems Research* 24(1):146–163.

Mittermayer, M. A., and Knolmayer, G. F. 2006. Newscats: A news categorization and trading system. In *Proceedings of the 6th International Conference on data mining (ICDM)*, 1002–1007. IEEE.

Nofsinger, J. R. 2005. Social mood and financial economics. *Journal of Behavioral Finance* 6(3):144–160.

Schumaker, R. P., and Chen, H. 2008. Evaluating a news-aware quantitative trader: The effect of momentum and contrarian stock selection strategies. *Journal of the American Society for Information Science and technology* 59(2):247–255.

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

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

Schumaker, R. P.; Zhang, Y. L.; Huang, C. N.; and Chen, H. 2012. Evaluating sentiment in financial news articles. *Decision Support Systems* 53(3):458–464.

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

Smola, A. J., and Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and computing* 14(3):199–222.

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

Wang, B.; Huang, H.; and Wang, X. 2011. A novel text mining approach to financial time series forecasting. *Neurocomputing* 83(2012):136–145.

Xu, S. X., and Zhang, X. M. 2013. Impact of wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly* 37(4):1043–1068.

Yu, Y.; Duan, W.; and Cao, Q. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55:919–926.