Proceedings of the 2015 International Conference on Machine Learning and Cybernetics, Guangzhou, 12-15 July, 2015

MINING GOLD IN SENIOR EXECUTIVES' POCKETS: AN ONLINE AUTOMATICALLY TRADING STRATEGY

CHAO MA[1], XUN LIANG[1]

[1]School of Information, Renmin University of China, Beijing 100872, China
E-MAIL: mc@ruc.edu.cn, xliang@ruc.edu.cn

Abstract:

Online financial news is an important part of financial big data. In this paper, we propose a model to promptly recognize valuable news about senior executives' behavior and an online automatically trading strategy based on the model. Our model consists of three phases. First, word segmentation and keyword extraction are employed to quantify the financial text. For a better efficiency and promptness, manifold learning is utilized to reduce the dimension of keyword vector. Second, the idea of financial event study is utilized to judge whether a specific type of news could produce significantly positive or negative return. Third, support vector machine is employed to recognize the specific financial news and associate the quantified text with the stock return. Experiments show that the recognition work performed excellently and the behavior of increasing shareholdings produces significant positive return. Our online automatically trading strategy based on the model obtained a return of 55.62%, outperforming three main benchmarks in the same period, 4.52%, 12.47% and -6.89% respectively.

Keywords:

Senior executives' behavior; Textual mining; Manifold learning; Support vector machine; Event study; Automatic trading

1. Introduction

In the era of big data, the instant and effective utilization of online financial text is a significant and challenging topic. Massive firm-specific news produced on the Internet is a universal but especially important type of online financial big data. Firstly, this kind of information is naturally connected with a specific listed company, which makes a straight-forward connection between the information and the stock price. Secondly, it always obviously influences on the investing value of corresponding stock. Thirdly, due to the property of the Internet, the transmission is large-extent and instant to some degree. However, the utilization of financial news on the Internet is tough. On one hand, the number of financial news produced online is huge. On the other hand, the type of financial news is various and complicated.

In the research of textual mining, scholars quantify the financial text as a vector and use data mining methods to find the relationship between the quantified text and corresponding stock price. Textual mining solves the problem of the huge amount but the explanation is not sufficient. Liang [1] regarded financial text as sentiment score based on a sentiment dictionary and associated the score with stock price using support vector regression. Groth [2] applied keyword vector to represent disclosures and predicted realized volatilities with data mining method including neutral network, naïve Bayesian, etc. Geva [3] used sentiment analysis tool to represent financial news with sentiment score, which is associated with the trend of corresponding stock price using neutral network and decision tree.

In the research of event study, scholars regard a specific kind of financial news as an event and the research focus on the stock price reaction to the event. Event study focus on specific event and the explanation is clear and concrete. However, compared with textual mining, there is a lack of intelligence and promptness. Muntermann [4] studied the reaction of stock price to specific disclosures (ad-hoc in Germany) and concluded that companies in smaller sizes are influenced more largely. Klößner [5] tested the volatilities within different windows around the published date of specific news and concluded that stock price overreacted to bad news. Groß-Klußmann [6] argued that stock price reacted differently to specific news with different sentiment indicators.

Because the listed companies are operated directly by their senior executives' decision, the news about senior executives' behavior always especially calls the attention of the market and scholars [7-8]. According to our research, in the mainland of China, the news reports three kinds of behavior of senior executives: senior executives resign (SER); senior executives increase their shareholdings (SEIS); senior executives decrease their shareholdings (SEDS).

In this paper, the idea of financial event study is employed to judge whether these three kinds of news could be applied to create profit. Textual mining methods are applied to associate the online financial news with the reaction pattern of

978-1-4673-7220-6/15/$31.00 ©2015 IEEE

corresponding stock price. Essentially, event study tells what is gold and textual mining finds them.

The remainder of the paper is organized as follows. Section 2 introduces the model including textual processing, data mining methods, and event study. Section 3 conducts experiments and constructs an online automatically trading strategy. Section 4 concludes the paper.

2. Research model

2.1. Textual processing

Online financial news is unstructured information. For an effective utilization of it, it is necessary to transfer the textual data to structured data. In this paper, word segmentation and keyword extraction are employed.

Obviously, the English words are single units, which are segmented naturally by space. However in Chinese all words are connected together. Word segmentation is necessary in the first phase of pre-processing, for the Chinese text does not have word boundaries. In the paper, we apply word segmentation software NLPIR (http://ictclas.nlpir.org/) to complete the pre-processing. After this phase, text is represented by word vector. Actually, the word vector could not represent a text appropriately because some words are redundant and the vector could not distinguish the text well. To solve this problem, the value of tf-idf = tf*idf is computed for every word [9-10], where tf = freq(w)/W, idf = log(L/(df(w)+1)), freq(w) denotes occurring frequency of word w, W denotes total occurring frequency of all words, df(w) denotes the number of documents containing w, and L denotes the total number of documents. Then the top D words with the highest values of tf-idf are regarded as the keywords of the text where D denotes the dimension of keyword vector.

2.2. Dimension reduction

Due to the fact that the dimension of keyword vector is high but in a specific text there are just several keywords occurred, the high-dimensional keyword vector is sparse, which greatly influences on the reaction promptness, efficiency and accuracy of our forecast models afterwards. In this paper, we employ manifold learning [11-12] to reduce the dimension. Essentially, manifold learning is a pre-processing layer added in the front of the input layer of the following forecast models, whose function is equal to the unsupervised learning in deep learning [13]. Here, the nonlinear mapping, from the high dimensional sparse vector to the low dimensional vector indicating the essential feature, is completed by manifold. Specifically, isomap algorithm [11] is applied to realize dimension reduction based on manifold learning. Given the keyword vector dataset $X=(x_1,x_2,\dots,x_n)\subset R^D$, where n denotes the size of X. The element in X can be obtained through a smooth embedding function f from a set Y in low-dimension space, namely, $x_i=f(y_i)$, where $Y=(y_1,y_2,\dots,y_n)\subset R^d, d\ll D, f\colon Y\to R^D$. Manifold learning outputs the nonlinear mapping $f^{-1}\colon R^D\to R^d$ and the resulted dimension is reduced.

The process of dimension reduction consists of following three steps [11]:

1) Construct neighborhood graph of financial news documents

Given $X=(x_1,x_2,\dots,x_n)\subset R^D$, we construct graph G containing all the points in X. Then we calculate the Euclidean distance du(i,j) between any two documents x_i and $x_j(i\neq j)$. If du(i,j) is less than the threshold ε or x_j is one of the K shortest points to x_i, they will be regarded being adjacent. In that condition, set graph G with edge E_{ij} and the weight of which is du(i,j).

2) Calculate shortest paths among financial news documents

If there exists edge E_{ij} between financial text x_i and $x_j (i\neq j)$, set the initial shortest distance dg(i,j) as du(i,j), or infinite otherwise. According to Dijkstra Algorithm [14], we update the shortest distance matrix, dg(i, j) = min { dg(i, j), dg(i, k) + dg(k, j)}, k = 1, 2, …, n, and set Dg = {dg^2(i, j)}.

3) Construct d-dimensional embedding

We apply MDS [11] in matrix Dg, and denote

$$S=\{s(i,j)\}=\left\{\delta(i,j)-\frac{1}{n}\right\},\ \text{where}\ \delta(i,j)=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases}$$

then we get
$$H=\frac{-SD_gS}{2} \tag{1}$$

Let $\lambda_1,\lambda_2,\dots,\lambda_d$ be the top *m* eigenvalues of H in descending order, and v_i be the according eigenvector of λ_i. We denote $U=(v_1,v_2,\dots,v_d)$. Then the d-dimensional embedding is

$$T=\mathrm{diag}\left(\lambda_1^{\frac{1}{2}},\lambda_2^{\frac{1}{2}},\dots,\lambda_d^{\frac{1}{2}}\right)U' \tag{2}$$

2.3. Forecast models

Through the processing before, every text can be represented by a low-dimensional vector. Next, we utilize SVM to recognize whether a financial text belongs to a specific type. The architecture of SVM is shown in Figure 1, where $x=(t_1,\dots,t_m)^T$ denotes the resulted low-dimensional vector processed by dimension reduction, $(\blacksquare)^T$ denotes the transpose of ■, and $y=\pm1$ denotes the label of x.

The SVM model is,

$$\max_{\alpha_1,\ldots,\alpha_l} = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i,x_j) + \sum_{i=1}^{l}\alpha_i$$
$$\text{s.t.} \sum_{i=1}^{l}\alpha_i y_i = 0$$
$$0 \le \alpha_i \le C,\ i=1,\ldots,l \quad (3)$$

where l denotes the number of data in the training set, $K(\blacksquare,\blacksquare)$ denotes kernel function, and α_i is the Lagrange multiplier.

The decision function is

$$\mathrm{Sgn}\left(\sum_{i=1}^{s}\alpha_i^* y_i K(x_i,x) + b^*\right) \quad (4)$$

where s denotes the number of support vectors, α_i^* denotes the optimal Lagrange multiplier, and b^* denotes the optimal offset.

There are three SVM classifiers for recognizing SER, SEIS and SEDS respectively.

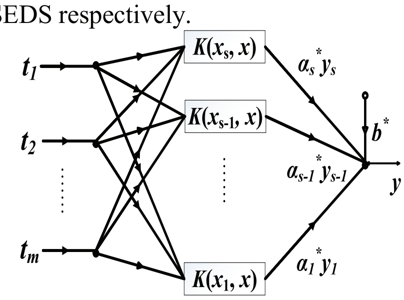

Figure 1. The svm architecture

2.4. Event study

Event study is employed to describe the reactive pattern of stock price. Furthermore, through event study, we could find the type which could create significantly positive or negative return, which is essential for deciding which type should be recognized. In normal event study, the event influencing stock price is determined on whether the cumulated abnormal return (cumulated AR, CAR) is significant in the corresponding time windows.

AR is defined as AR=R-E_R where R denotes the real return and E_R denotes the expected return. CAR is defined as,

$$\mathrm{CAR} = \sum_{i=1}^{w}\mathrm{AR}_i \quad (5)$$

where w denotes the length of the time window.

In normal event study, considering the possibility of insider dealing, the return is also calculated before the event day. In addition, ER is calculated for getting rid of the influence of market, industry, and size. In this paper, we aim at finding the news which could bring significantly positive return and design an automatically trading strategy so that actually we could simplify the normal event study. What we care about is just whether there is a positive return after specific financial news about senior executives' behavior, so in the paper E_R is set as zero and CAR is the cumulated return in the following 30 trading days.

3. Experiments analysis

3.1. Data

The data were collected from http://finance.sina.com.cn through our specially designed web crawlers. HTML pages were parsed and the content we need were extracted. Totally, we collected 20056 pieces of news from 2009 to 2014, containing 477 SER, 1327 SEIS, and 2532 SEDS.

3.2. Specific news recognition

After textual processing including word segmentations and keyword extraction, every text is represented by a high-dimensional keyword vector. In the paper, the dimension of keyword vector is set as 500. Due to improvement of the reaction promptness, accuracy and efficiency of our forecast models, manifold learning is utilized to reduce the dimension of the high-dimensional keyword vector. Residual error is the key indicator which reflects the outcomes of dimension reduction and the dimension with the lower residual error has the higher probability to be the essential dimension [11]. Figure 2 shows the residual errors with different embedded dimensions. It is clear that the change of the residual error does not decline obviously when the dimension is reduced to 15. The essential dimension of financial news should be within the range of 15 to 20.

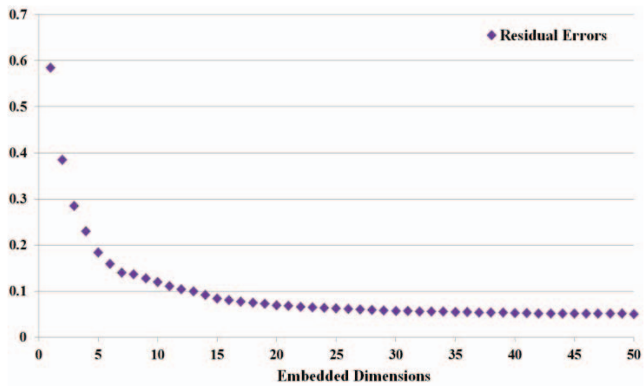

Figure 2. The residual errors in different embedded dimensions

After the previous process, the unstructured text can be transferred to a low-dimensional vector. Our forecast models based on SVM can recognize the category of each text automatically. To evaluate the forecast performance of the models, precision ratio (P) and recall ratio (R) are applied,

$$P=\frac{\text{\# of samples recognized accurately}}{\text{\# of samples recognized}}$$

$$R=\frac{\text{\# of samples recognized accurately}}{\text{\# of samples in corresponding category}}$$

There are many kernels to be used in SVM, such as polynomial (POLY) kernels and radial basis (RB) kernels. Nine kinds of SVM are applied: six POLY kernels $(ax_i^T x+1)^k$, with (a, k)=(0.01, 2), (a, k)=(0.01, 3), (a, k)=(0.1, 2), (a, k)=(0.1, 3), (a, k)=(1, 2), (a, k)=(1, 3) and three RB kernels $\exp\left(-a\|x_i-x\|^2\right)$ with a=0.01, a=0.1, a = 1. Table 1 presents the performance of different SVMs for different news categories. Higher precision means a higher probability that the recognized news is the one we want to recognize. Meanwhile, high recall means that the probability of missing this kind of news is low. From the result of Table 1, our forecast models performed excellently in terms of precision ratio.

Table 1. The prediction results of specific news category

Embedded Dimensions	Kernel Types	SER		SEIS		SEDS	
		P (%)	*R (%)*	*P (%)*	*R (%)*	*P (%)*	*R (%)*
15	POLY(a = 0.01, d =2)	**87.2±2.5**	65.6±4.7	**93.5±1.1**	67.1±1.8	**93.5±0.9**	81.4±1.8
	POLY(a = 0.01, d =3)	**99.7±0.5**	64.4±2.5	**99.4±0.3**	56.5±1.5	**98.4±0.7**	53.9±2.8
	POLY(a=0.1, d = 2)	**85.6±2.7**	85.1±3.0	**88.1±1.9**	81.3±3.0	**92.4±0.9**	88.1±1.2
	POLY(a=0.1, d = 3)	**91.3±2.0**	91.3±3.0	**90.6±1.4**	87.0±2.3	**93.0±0.9**	89.7±1.2
	POLY(a=1, d = 2)	**81.4±2.7**	82.6±3.0	**84.2±1.9**	82.4±2.7	**90.2±1.2**	87.5±1.5
	POLY(a=1, d = 3)	**87.9±3.4**	88.8±3.6	**85.2±1.9**	85.2±2.4	**88.0±1.3**	87.6±1.6
	RBF(a=0.01)	**94.2±1.9**	90.1±3.5	**92.4±1.9**	86.8±2.7	**93.5±0.8**	90.3±1.0
	RBF(a=0.1)	**91.0±2.7**	91.5±3.1	**90.2±1.4**	89.8±1.4	**92.8±0.7**	91.9±1.0
	RBF(a=1)	**84.5±8.4**	80.3±10.5	**77.0±1.6**	92.8±1.3	**85.3±1.3**	94.1±0.8
16	POLY(a = 0.01, d =2)	**87.4±2.3**	65.9±4.1	**93.8±1.0**	67.9±1.9	**93.7±0.9**	82.6±1.5
	POLY(a = 0.01, d =3)	**99.7±0.5**	65.3±2.8	**99.4±0.4**	56.6±1.4	**98.3±0.6**	54.3±2.9
	POLY(a=0.1, d = 2)	**84.9±2.2**	85.9±3.0	**88.2±1.8**	82.2±3.0	**92.3±1.0**	88.3±1.3
	POLY(a=0.1, d = 3)	**91.4±1.7**	91.6±3.0	**90.5±1.4**	87.3±2.0	**92.8±0.8**	89.8±1.1
	POLY(a=1, d = 2)	**81.1±3.0**	82.7±3.5	**84.6±1.9**	82.8±2.3	**89.9±1.2**	87.6±1.4
	POLY(a=1, d = 3)	**87.3±2.9**	89.1±3.1	**84.7±2.2**	84.6±2.5	**88.0±1.4**	87.3±1.7
	RBF(a=0.01)	**94.5±1.7**	90.0±3.1	**92.7±1.8**	86.4±2.4	**93.6±0.8**	90.1±0.9
	RBF(a=0.1)	**90.6±2.6**	91.6±3.4	**90.2±1.5**	89.2±1.6	**92.8±0.9**	91.7±1.1
	RBF(a=1)	**85.0±8.2**	79.6±10.7	**76.1±1.7**	93.0±1.2	**84.5±1.0**	94.3±0.8
17	POLY(a = 0.01, d =2)	**87.8±3.0**	65.6±4.4	**93.8±1.1**	67.7±2.0	**93.4±0.9**	82.9±1.3
	POLY(a = 0.01, d =3)	**99.7±0.6**	64.5±2.7	**99.4±0.3**	56.2±1.2	**98.3±0.7**	55.0±2.7
	POLY(a=0.1, d = 2)	**85.0±2.7**	85.7±3.3	**88.2±1.6**	82.6±2.8	**92.3±1.1**	88.1±1.3
	POLY(a=0.1, d = 3)	**91.0±2.0**	91.9±3.1	**90.5±1.4**	87.3±1.9	**92.8±0.9**	89.8±1.2
	POLY(a=1, d = 2)	**81.3±3.0**	82.3±3.4	**84.1±1.8**	82.0±2.8	**89.3±1.3**	87.5±1.4
	POLY(a=1, d = 3)	**87.7±3.3**	88.9±3.1	**84.7±2.2**	85.0±1.9	**88.1±1.5**	86.7±1.6
	RBF(a=0.01)	**94.2±1.8**	90.4±3.3	**92.4±1.7**	87.5±2.0	**93.7±0.9**	90.2±1.1
	RBF(a=0.1)	**90.8±2.8**	91.0±3.8	**90.5±1.5**	89.3±1.8	**92.6±0.9**	91.9±1.0
	RBF(a=1)	**83.4±8.9**	80.1±11.4	**75.7±1.7**	93.3±1.4	**84.2±1.2**	94.4±0.8
18	POLY(a = 0.01, d =2)	**87.0±2.7**	67.1±5.3	**93.8±1.0**	67.3±2.0	**93.4±0.8**	83.3±1.5
	POLY(a = 0.01, d =3)	**99.7±0.5**	65.6±3.1	**99.3±0.4**	56.6±1.3	**98.3±0.7**	55.6±2.6
	POLY(a=0.1, d = 2)	**85.0±2.6**	84.9±3.6	**88.3±1.5**	82.9±3.3	**92.2±0.9**	88.4±1.3
	POLY(a=0.1, d = 3)	**91.2±2.0**	91.6±2.9	**90.4±1.2**	87.3±2.0	**92.9±1.0**	89.9±1.1
	POLY(a=1, d = 2)	**80.9±3.2**	82.0±3.8	**84.5±1.8**	82.0±2.3	**89.0±1.1**	87.3±1.5
	POLY(a=1, d = 3)	**86.9±3.3**	90.4±3.2	**84.7±2.4**	85.1±2.0	**88.0±1.6**	87.1±1.6
	RBF(a=0.01)	**94.3±1.8**	90.6±3.5	**92.6±1.6**	87.5±1.8	**93.5±0.8**	90.7±0.9
	RBF(a=0.1)	**90.7±2.9**	91.1±3.5	**90.2±1.5**	89.8±1.6	**92.5±0.9**	91.9±0.9
	RBF(a=1)	**82.6±9.1**	81.3±11.5	**75.1±1.6**	93.4±1.2	**83.9±1.3**	94.3±0.8
19	POLY(a = 0.01, d =2)	**87.4±3.1**	67.3±5.9	**94.0±1.1**	67.9±2.0	**93.5±0.8**	83.2±1.5
	POLY(a = 0.01, d =3)	**99.7±0.5**	64.9±2.7	**99.3±0.4**	56.5±1.5	**98.3±0.7**	55.1±3.0
	POLY(a=0.1, d = 2)	**84.7±2.4**	84.2±4.0	**87.9±1.9**	83.4±2.8	**92.1±0.8**	88.7±1.2
	POLY(a=0.1, d = 3)	**91.0±2.4**	91.4±2.8	**90.4±1.5**	87.3±1.8	**92.8±1.0**	90.0±1.2
	POLY(a=1, d = 2)	**80.7±3.0**	82.6±3.7	**83.8±1.8**	82.5±2.2	**88.9±1.3**	87.5±1.5
	POLY(a=1, d = 3)	**87.3±2.9**	89.5±3.1	**84.4±2.2**	84.8±2.6	**87.9±1.4**	86.9±1.9
	RBF(a=0.01)	**94.1±1.9**	90.5±2.9	**92.8±1.7**	87.0±2.2	**93.3±0.8**	90.9±1.0
	RBF(a=0.1)	**89.6±3.0**	91.4±4.0	**90.5±1.5**	89.4±1.5	**92.3±0.9**	92.0±0.9
	RBF(a=1)	**80.1±9.7**	84.8±11.3	**74.6±1.5**	93.6±1.2	**83.2±1.4**	94.4±0.9
20	POLY(a = 0.01, d =2)	**86.9±2.9**	67.5±6.0	**94.0±1.1**	67.5±2.2	**93.3±0.8**	83.4±1.7
	POLY(a = 0.01, d =3)	**99.7±0.5**	64.7±2.9	**99.3±0.4**	56.8±1.7	**98.4±0.5**	55.6±3.2
	POLY(a=0.1, d = 2)	**84.4±2.3**	84.4±3.8	**88.1±1.8**	83.2±2.8	**92.1±1.0**	88.7±1.3
	POLY(a=0.1, d = 3)	**90.8±2.2**	91.7±3.2	**90.6±1.5**	87.2±1.9	**92.8±1.0**	89.9±1.1
	POLY(a=1, d = 2)	**80.8±3.0**	82.3±3.4	**83.7±2.0**	81.7±2.8	**88.8±1.4**	87.2±1.5
	POLY(a=1, d = 3)	**87.4±3.0**	90.4±2.9	**84.2±2.5**	85.0±2.1	**88.0±1.5**	87.5±1.4
	RBF(a=0.01)	**93.9±1.7**	90.9±2.8	**92.9±1.6**	87.0±2.3	**93.5±0.8**	91.0±1.0
	RBF(a=0.1)	**90.4±3.2**	90.1±4.1	**89.9±1.6**	89.6±1.7	**92.4±0.9**	91.8±0.8
	RBF(a=1)	**79.4±9.4**	85.0±11.5	**74.3±1.7**	93.2±1.2	**83.3±1.5**	94.4±0.7

For the demand of our automatically trading strategy afterwards, it is not necessary to recognize all the news so the forecast results are satisfied. Our models can be written in the memory of computers so that the specific news can be recognized rapidly. If the reaction of corresponding stock price is clear, then the decision will be completed automatically.

3.3. Specific news study

Our models could recognize specific news with an excellent precision ratio. However, not all the recognized news works. Financial event study is necessary to judge whether a specific kind of news is available. Our financial event study focuses on the cumulated return of 30 trading days after the news. Table 2 shows the results. Obviously, for SEIR, there is a significantly positive return, which is the basic of our trading strategy afterwards. Meanwhile, for SER and SEDS, the return is not significant. In terms of sentiment trend, SER and SEDS are relatively negative and SEIS is positive. Obviously, from our results, trading based on sentiment trend only is not convincing and sufficient. Financial event study is necessary.

Table 2. Results of event study

Types	Number	Average	Standard Deviation	T-statistic	P-value
SER	298	-0.196	11.838	-0.285	0.387
SEIS	770	0.954	11.259	***2.351******	***0.009***
SEDS	1485	0.455	14.189	1.235	0.108

Note: ***, ** and * mean significant at the significant levels of 1%, 5% and 10%, respectively.

3.4. Automatically trading strategy

To evaluate our model more objectively and put it into practice, we propose an automatically trading strategy. Based on the results of financial event study, SEIR could produce significantly positive return. Our automatically trading strategy recognizes SISH every day and long the corresponding stocks in the future one month. What we want to emphasize is that our event study result is based on the data before 2013 but the automatically trade strategy started at January of 2013 and ended at November of 2014, which confirms that the strategy does not utilize the information in the future. Shanghai and Shenzhen 300 stocks index (HS300), Shanghai stocks index (SHI) and Shenzhen stocks index (SZI) are universal benchmarks in the A share market [15], so we regard them as benchmarks also. Figure 3 illustrates the performance of our strategy and three benchmarks. Four portfolios start with the net value 1. After about 2 years, the return of our strategy is 55.62%, which excellently outperforms the three market benchmarks (HS300: 4.52%; SHI: 12.47%; SZI: -6.89%).

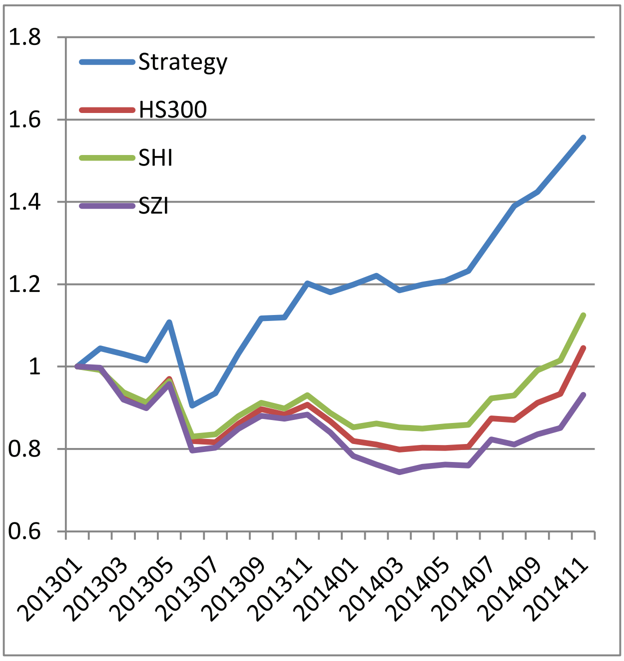

Figure 3. The performances of automatically trading strategy and three benchmarks

4. Conclusions

In the paper, we present a novel model to recognize the profitable news promptly and design an automatic trading strategy based on the model. Our model contains three parts. First, word segmentation and keyword extraction transferred the text to a keyword vector. To improve the promptness and accuracy, manifold learning is employed to reduce the dimension. Second, the idea of financial event study is applied to judge whether specific news could bring significantly positive or negative return, which is the basis of our trading strategy. Third, SVM realizes the recognition of specific news so as to associate the quantified text with a reactive pattern of stock price. Experiments show that the model could excellently recognize specific news in massive financial big data. The results of financial event study indicate that SEIS brings significant positive return. In the automatic trading strategy, SEIS is recognized promptly and holds 30 trading days. From January, 2013 to November 2014, the strategy obtained a return of 55.62%, and the benchmarks are 4.52%, 12.47%, and -6.89%, respectively.

Actually, we put forward an effective and efficient method to obtain positive return according to senior executives' behavior. The financial news contains various types besides the type about senior executives, such as profit report etc. In the further research, other types will be taken into consideration and the portfolio management will be refined.

Acknowledgements

The work was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (10XNI029), the Beijing NSF under grant number 4132067, and the NSF of China under grant number 71271211.

References

[1] X. Liang, R. C. Chen, Y. He, and Y. Chen, "Associating stock prices with web financial information time series based on support vector regression," Neurocomputing, Vol. 115, pp. 142-149, 2013.

[2] S. S. Groth, and J. Muntermann, "An intraday market risk management approach based on textual analysis," Decision Support Systems, Vol. 50 pp. 680-691, 2011.

[3] T. Geva, and J. Zahavi, "Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news," Decision Support Systems, Vol. 57, pp. 212-223, 2014.

[4] J. Muntermann, and A. Guettler, "Intraday stock price effects of ad hoc disclosures: the German case," Journal of International Financial Markets, Institutions and Money, Vol.17, pp. 1-24, 2007.

[5] S. Klößner, M. Becker, and R. Friedmann, "Modeling and measuring intraday overreaction of stock prices,"Journal of Banking & Finance, Vol. 36, pp. 1152-1163, 2012.

[6] A. Groß-Klußmann, and N. Hautsch, "When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions," Journal of Empirical Finance, Vol. 18, pp. 321-340, 2011.

[7] J. M. Salas, "Entrenchment, governance, and the stock price reaction to sudden executive deaths," Journal of banking & finance, Vol. 34, pp. 656-666, 2010.

[8] P. J. Brown, Z. Matolcsy, and P. Wells, "Group versus individual compensation schemes for senior executives and firm performance: Some evidence based on archival data," Journal of Contemporary Accounting & Economics, Vol. 10, pp. 100-114, 2014.

[9] S. M. Weiss, N. Indurkhya, T. Zhang, and F. Damerau, Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer Science & Business Media, 2010.

[10] M. W. Berry, J. Kogan, Text Mining: Applications and Theory, Wiley, 2010

[11] J. B. Tenenbaum, S. V. De, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," Science, Vol. 290, pp. 2319-2323, 2000.

[12] S. T. Roweis, and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, Vol. 290, pp. 2323-2326, 2000.

[13] G. E. Hinton, and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, Vol. 313, pp. 504-507, 2006.

[14] E. W. Dijkstra, "A note on two problems in connexion with graphs," Numerische Mathematik, Vol. 1, pp. 269-271, 1959.

[15] T. Zheng, and H. Zuo, "Range based regime switching stochastic volatility models with applications," Journal of Management Sciences in China, Vol. 16, pp. 82-94, 2013.