Measuring Financial Crisis Index for Risk Warning through Analysis of Social Network

Zhongsheng Wang Kiyoaki Shirai
School of Advanced Science and Technology,
Japan Advanced Institute of Science and Technology
{wzswan1.0,kshirai}@jaist.ac.jp

1 Introduction

A financial crisis such as a drastic drop of stock prices may cause a lot of loss to many investors. At the same time, due to the globalization, a financial crisis in one country may influence to the whole world. As stock market is innate complex, dynamics and chaotic, management of financial risk has been proved to be a very difficult task.

In Socionomic Theory of Finance [7, 8], irrational speculation behaviors play an import role in a financial crisis. Nowadays, Social Networking Service (SNS), such as Twitter or Weibo, is widely used by many people. Those social networks provide us lots of information to monitor a financial market situation and can be used to predict a financial crisis.

The goal of this research is to propose a new financial index that measures an extent of a financial crisis. Microblog is used as a source of our new index. Hypotheses behind our financial crisis index are as follows.

- Not many investors focus on financial markets daily. When a bull market begins and stock price keeps going up, more beginners come into markets.
- Those people are intense to take irrational speculation behaviors. It may cause a panic and a bear market. Behaviors of beginners may make a financial risk more serious.
- When many novice investors come into the market, they posts messages about financial topics. Therefore, intensity of attention toward finance on microblog may has a positive correlation with a financial bubble level. That is, an amount of financial messages on SNS can be used as a financial crisis index.

In addition, a model to predict a movement(up or down) of stock prices is trained with past stock prices as well as our index derived from texts in microblog. Since historical prices is sequential data, Long Short-Term Memory (LSTM) has been used for prediction of a stock price.We train such an LSTM and measure its ability of prediction to evaluate effectiveness of our proposed financial crisis index.

2 Related work

Traditional previous researches on financial analysis only worked on history data and past stock prices. With a wide spread of Web forum and SNS, textual information obtained from them can be utilized for the market analysis. Nguyen et al. proposed a method based on sentiment analysis on social media to predict a movement of stock prices [4]. Jaramillo et al. proposed a method to predict a stock price using a history of prices as well as polarity of company reports and news [2]. Jianhong et al. applied a deep learning method on sentiment-aware stock market prediction [3]. Similar to these previous studies, we also use textual information for a stock market analysis. However, we mainly focus on not prediction of a stock value but financial risk aversion.

Most researchers and investors care about a good return in the financial market. But how to manage a risk is more import to avoid massive loss. Some of the previous work on the financial risk management were based on historical prices only, but several studies also tried to use textual analysis [5, 6, 10]. However, texts on SNS were not paid attention for risk management. In this paper, comments on Weibo is used as a source of textual information for management of a financial crisis.

3 Financial Attention Index

3.1 Definition

Financial Attention Index (FAI) is our proposed index that measures an extent of a financial risk. It is defined as Equation (1).

$$FAI \stackrel{def}{=} \frac{\begin{array}{l}\textit{number of financial related}\\ \textit{comments on SNS}\end{array}}{\textit{total number of comments on SNS}} \tag{1}$$

As discussed in Section 1, we suppose that financial topics are to be more mentioned and discussed on SNS when many novice investors participate in the market. Such a situation might cause the market unstable. Therefore, FAI is supposed to be positively correlated to a financial risk. Although a financial crisis can be measured by various points of view, FAI can be a financial crisis index from one perspective.

Comments on SNS are classified whether they are related to financial topics or not. Weibo is chosen as SNS in this study. Then the number of financial comments and total comments are counted to get FAI. A classifier of financial related comments is not trained from Weibo comments. Instead, a labeled data of news articles is used for training. In our research, FAI is calculated for every week; the number of comments posted in a period of one week is counted to get FAI.

3.2 Calculation of FAI

3.2.1 Type of classifier

Two kinds of financial classifiers are trained. One is the two-way classifier, which classifies a comment on Weibo as a financial or non-financial related comment. The other is the three-way classifier. It classifies a comment on Weibo into three classes: (1) stock-related, which is a comment related to the stock market, (2) financial-related, which is a comment related to financial markets such as future market, bond market and so on, but not related to the stock market, and (3) other, which is a comment not related to financial topics. Since we mainly focus on analysis of stock prices for detection of a financial crisis, we distinguish topics about stocks with other financial related topics. When the three-way classifier is used for calculation of FAI, both stock-related and financial-related comments are treated as the financial comments.

3.2.2 Data collection

Two kinds of data are collected to calculate FAI.

Weibo dataset

Comments on Weibo posted from 2013-7-1 to 2016-12-5 are crawled. This period contains stable, bull, and bear markets as reported later. The number of collected comments is 2,104,746. They are collected with their posting time to calculate FAI for each period of one week.

News dataset

News articles are collected from two sources: the Tencent news website[1] and the text collection of THUC Project developed by Tsinghua University [9]. In this dataset, each document is annotated with its topic (stock related, financial related and other). Table 1 shows the number of documents in this dataset.

The news dataset is used to train both the two-way and three-way financial classifiers. When the two-way classifier is trained, news articles of the stock and financial classes are treated as the financial related documents.

[1] http://news.qq.com/, http://finance.qq.com/

Table 1: News dataset

Class	Website	THUC Project
Financial related	800,000	5,000,000
Stock related	5,000	200,000,000
Other	165,000	500,000,000

3.2.3 Training of classifier

Support Vector Machine (SVM) is used for training our financial classifier. A linear function is chosen as the kernel function of SVM. The tools of gensim, sklearn, scipy and jieba are used for training SVM.

Bag-of-words are used as features for training SVM. Function words are removed by a preprocessing and all content words in a document are extracted as the features. The weight of each feature is set as the TF-IDF score.

We found that the number of the features was high, i.e. nearly 200,000. Therefore, we apply Latent Semantic Analysis (LSA) to reduce the feature space. LSA is a technique to reduce a size of a matrix of words by documents using Singular Value Decomposition (SVD). In this study, the number of the features is reduced to 50.

4 Stock price movement prediction with FAI

FAI is used for prediction of a stock index to evaluate its ability for risk management. LSTM [11] is chosen for training a prediction model, since it is well used for prediction of time series. In our model, input of LSTM is a time sequence of either the stock index, a difference of the stock index, or FAI. The difference of the stock index is defined as a change of the stock index between the current and previous periods. We also train a model where these three kinds of values are concatenated as a vector and passed to the input layer of LSTM. The output of LSTM is a stock index of the next period. We define a period of LSTM as one transaction day. Note that FAI is calculated for each week. If FAI is used for LSTM, the same value is entered during days in a week. Our LSTM structure consists of one input layer, two LSTM layers and one output layer. The input and output layers consist on one neuron node. The first and second LSTM layers contain 5 and 100 nodes respectively.

LSTM is learned through training by python deep learning library Keras [1]. The activate function in LSTM units is ‘linear’. The model is trained by the rmsprop method with 1 examples in a batch, with categorical cross entropy as the objective loss function. A validation fraction is set as 0.1%. A learning rate is set as 0.001. All initial weights are set to be small positive constant values. To prevent overfitting, a dropout is set at 20% and an L2 regularization constraint is set as 0.01.

5 Evaluation

5.1 Classification of financial comment

The classifier to judge whether a comment is related to financial topics takes an import role in FAI. First, the financial classifier is empirically evaluated. Ten-fold cross validation on our news dataset is carried out. The performance of the classifier is measured by the accuracy, which is a ratio of the number of correctly classified comments to the total number of comments.

The accuracy of the two-way and three-way classifier is 83.35% and 86.18% respectively[2]. The performance of the financial classifiers is satisfying. We found that the three-way classifier outperformed the two-way classifier. So the three-way classifier is used in our next experiments.

5.2 Prediction of stock index movement

5.2.1 Experimental setting

Our LSTM-based stock index prediction model is evaluated for its ability to predict not the stock index but a movement of the stock index. More precisely, we consider a task to classify a movement of the stock index in one week into one of the following three classes.

Up : A case where the stock index goes up by T or more as $P_e - P_s > T$, where P_s and P_e is the stock index at the start and end of the period.

Keep : A case where the stock index does not change drastically as $|P_e - P_s| \leq T$.

Down : A case where the stock index goes down by T or more as $P_e - P_s < -T$.

We set T as 0.02 in this experiment.

Since our FAI is derived from comments of Weibo that is Chinese microblogging service, SSE Composite Index (SCI) is chosen as the stock index in this experiment. SCI is computed from prices of stocks of Chinese companies. It is a tool widely used by investors to describe the market. The SCI values are obtained from the finance sina website[3]. The dataset contains the SCI values of 679 trading days during 2013-10-28 to 2016-8-2, which is almost the same period where the comments on Weibo are downloaded.

One hundred and twenty continuous weeks in our stock index dataset are chosen as test periods. For each test period, all past values are used as the training data.

[2]Recall that the comments are classified as either "financial-related" or "other" even when the three-way classifier is used; both stock-related and financial-related comments are regarded as financial-related comments.

[3]http://finance.sina.com.cn/

5.2.2 Evaluation criteria

Two evaluation criteria are used in this experiment.

Accuracy

It is a proportion of the number of the test periods for which the predicted movement class agrees with the true class.

No Lost Accuracy

The main goal of this research is to help investors not to get a good return but to avoid financial risks. It is more important to predict the movement "Down" in order not to cause loss. Therefore, we introduce another criterion called "No Lost Accuracy". It is the accuracy of the stock movement prediction task where the test periods are classified as either "Down" or not. That is, "Up" and "Keep" classes are merged into one class "Not-Down".

5.2.3 Result and discussion

Table 2 shows the Accuracy (A) and No Lost Accuracy (NLA) of four different prediction models. "SI", "DI", "FAI" stand for the model using only the stock index, the difference of the stock index, and FAI, respectively. "All" stands for the model using three values.

Table 2: Result of stock movement prediction

Model	SI	DI	FAI	All
A	31.1%	42.9%	41.2%	35.3%
NLA	47.9%	62.2%	54.6%	57.1%

"FAI" is better than "SI" in terms of the Accuracy and No Lost Accuracy, indicating that FAI is effective to predict the movement of the stock index. When combining FAI with other indexes, the No Lost Accuracy is also improved. However, "FAI" does not outperform "DI". It is found that the difference of the stock index is a strong indicator of the movement of the stock market.

Figure 1 shows a change of SSE Composite Index in our dataset. Both a bull market and bear market are found in this graph. To evaluate the performance of the prediction models in different situations (bull market, bear market etc.), we divide the test periods into 12 terms, where each term consists of 10 test periods (10 weeks), and measure an average of Accuracy and No Lost Accuracy on each term. Results are shown in Table 3. The tables include the situation, which is graphically shown in Figure 1, for each term.

The results of the experiment show a tendency that the model "FAI" achieves better performance than the others in the situations of both a bull and bear market. Especially, it is good on a front bull situation (T5). These results indicate that FAI is effective

Figure 1: Change of SSE Composite Index

to predict a drastic movement of the market. On the other hand, the model using the difference of the stock index (DI) works well for the stable situations. The model using the stock index (SI) is the worst in our experiment.

Table 3: Prediction for different situation

(a) Accuracy

Term	**Situation**	**SI**	**DI**	**FAI**	**All**
T1	stable	50%	60%	40%	40%
T2	stable	10%	60%	50%	40%
T3	stable	30%	40%	40%	20%
T4	prim bull	30%	70%	10%	50%
T5	front bull	40%	30%	90%	50%
T6	mid bull	40%	30%	40%	40%
T7	late bull	20%	50%	40%	30%
T8	front bear	20%	30%	40%	0%
T9	mid bear	10%	40%	30%	50%
T10	late bear	40%	20%	50%	10%
T11	stable	30%	40%	40%	50%
T12	stable	56%	44%	22%	44%

(b) No Lost Accuracy

Term	**Situation**	**SI**	**DI**	**FAI**	**All**
T1	stable	70%	60%	50%	70%
T2	stable	10%	80%	60%	70%
T3	stable	70%	60%	50%	20%
T4	prim bull	50%	90%	40%	50%
T5	front bull	60%	50%	100%	80%
T6	mid bull	50%	50%	50%	70%
T7	late bull	30%	60%	40%	50%
T8	front bear	40%	50%	50%	50%
T9	mid bear	30%	60%	70%	60%
T10	late bear	50%	50%	50%	40%
T11	stable	50%	70%	50%	70%
T12	stable	67%	67%	44%	56%

6 Conclusion

As a novel index to measure an extent of a financial risk, this paper proposed the Financial Attention Index (FAI) that calculates the proportion of the financial related comments on SNS. We empirically proved the effectiveness of the FAI in the task of stock movement prediction. In future, we will investigate a better design of a structure of LSTM to accept multiple inputs, i.e. the stock index, the difference of the index, and FAI.

References

[1] François Chollet. Keras. *Git-hub*, 2015.

[2] Johana Jaramillo, Juan David Velasquez, and Carlos Jaime Franco. Research in financial time series forecasting with SVM: Contributions from literature. *IEEE Latin America Transactions*, 15(1):145–153, Jan 2017.

[3] Jiahong Li, Hui Bu, and Junjie Wu. Sentiment-aware stock market prediction: A deep learning method. In *2017 International Conference on Service Systems and Service Management*, pages 1–6, June 2017.

[4] Thien Hai Nguyen, Kiyoaki Shirai, and Julien Velcin. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.*, 42(24):9603–9611, December 2015.

[5] Michael Niemira and Thomas Saaty. An analytic network process model for financial-crisis forecasting. *International Journal of Forecasting*, 20:573–587, 2004.

[6] Kyong Joo Oh, Tae Yoon Kim, Hyoung Yong Lee, and Hakbae Lee. *Using Neural Networks to Support Early Warning System for Financial Crisis Forecasting*. 2005.

[7] Robert R. Prechter Jr, Deepak Goel, Wayne D. Parker, and Matthew Lampert. Social mood, stock market performance, and U.S. presidential elections: A socionomic perspective on voting results. *SAGE Open*, 2(4):2158244012459194, 2012.

[8] Robert R. Prechter Jr and Wayne D. Parker. The financial/economic dichotomy in social behavioral dynamics: The socionomic perspective. *Journal of Behavioral Finance*, 8(2):84–108, 2007.

[9] Maosong Sun, Zhipeng Guo, Yu Zhao, Yabin Zheng, Xiance Si, and Zhiyuan Liu. Thu Chinese text classification, 2006.

[10] Roman Trusov, Alexey Natekin, Pavel Kalaidin, Sergey Ovcharenko, Alois Knoll, and Aida Fazylova. Multi-representation approach to text regression of financial risks. In *2015 Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference (AINL-ISMW FRUCT)*, pages 110–117, 2015.

[11] Ruoxuan Xiong, Eric Nichols, and Yuan Shen. Deep learning stock volatilities with google domestic trends. 2015.