Deep Learning for Stock Prediction Using Numerical and Textual Information

Ryo Akita
Graduate School of System Informatics, Kobe University
akita@ai.cs.kobe-u.ac.jp

Akira Yoshihara
Graduate School of System Informatics, Kobe University
yoshihara@ai.cs.kobe-u.ac.jp

Takashi Matsubara
Graduate School of System Informatics, Kobe University
matsubara@phoenix.kobe-u.ac.jp

Kuniaki Uehara
Graduate School of System Informatics, Kobe University
uehara@kobe-u.ac.jp

Abstract—**This paper proposes a novel application of deep learning models, Paragraph Vector, and Long Short-Term Memory (LSTM), to financial time series forecasting. Investors make decisions according to various factors, including consumer price index, price-earnings ratio, and miscellaneous events reported in newspapers. In order to assist their decisions in a timely manner, many automatic ways to analyze those information have been proposed in the last decade. However, many of them used either numerical or textual information, but not both for a single company. In this paper, we propose an approach that converts newspaper articles into their distributed representations via Paragraph Vector and models the temporal effects of past events on opening prices about multiple companies with LSTM. The performance of the proposed approach is demonstrated on real-world data of fifty companies listed on Tokyo Stock Exchange.**

1. Introduction

Stock prices react to events related to business performances or overseas markets. Investors may judge on the basis of technical analysis, such as a company's charts, market indices, and on textual information such as newspapers or news-carrying microblogs. It is however difficult for investors to analyze and predict market trends according to all of these information. Many Artificial Intelligence (AI) approaches have been investigated to predict those trends automatically [1], [2], [3]. For example, investment simulation analysis with artificial markets [4], [5] or stock trend analysis with lexical cohesion based metric of sentiment polarity of financial news [6].

However, these works encounter four issues. First, many of them which use textual information represent the information as Bag-of-Words [2], [7] despite that the BoW model cannot capture the some of useful information for our purpose since this model cannot consider any interpretation of linguistic patterns or aspects such as word order, synonyms, co-references, and pronoun resolution. Next, previous works often used either textual or numerical information for market analysis, while the investors make decisions on various information [8]. By using both information, a model can capture more complex relationships between them and the stock price. The third problem is, many previous works have considered only one company on training model, while the stock prices between companies should be correlational [9]. Finally, the previous works which use the textual information did not consider stock prices as time series [10]. This is because of the difficulty to make rules how textual information influences time series.

This study focuses on these four issues. We apply Paragraph Vector [11] to obtain a continuous distributed representation of each news article. We use the distributed representations and the diary open prices of 10 companies to predict their close price by regression analysis. As the predictive model, we employ the Long Short-Term Memory (LSTM) model [12], [13] for dealing with the influence of time series. Experiments of stock market simulation on financial news datasets demonstrate that our model effectively overcomes these issues.

2. Related Works

2.1. Stock price prediction depending on textual information

There has been a number of efforts for predicting the trend of stock prices on the basis of textual information [14], [15], [16], [17], [18]. For example, Lavrenko et al. [7] combined the trends of stock prices and financial news article, and predicted the trends using the content of news articles before the trends actually appear. The predicted trends in their simulation were able to make profit.

Schumaker and Chen [8] compared several different textual representations such as Bag-of-Words, Noun Phrases, and Named Entities for stock price prediction. They showed that Bag-of-Words is not sufficient, while support vector machines (SVM) with proper noun features is superiority of predicting the trend of stock prices. Hagenau et al. [10] predicted the difference of open and close prices of a stock with the data from DGAP and EuroAdhoc which are corporate announcements of Germany and UK, respectively. They used bigram, a sequence of two adjacent words and

978-1-5090-0806-3/16/$31.00 copyright 2016 IEEE
ICIS 2016, June 26-29, 2016, Okayama, Japan

2-word combination as features, and conducted feature selection according to χ^2-statistic with respect to each brand of stock prices. Their experiment showed that Chi-Square based feature selection method allowed lifting classification accuracy and reduced overfitting.

However, these works share a common problem; the efficiency in dealing with large scale textual information. Ding et al. [9] employed Open IE（Information Extraction） techniques to extract the actor and object of events from title of news articles, and predicted the S&P 500 index. They used the deep neural network model as classifier and achieved better performance than SVM in their experiments.

2.2. Discussion about issues on previous works

The following features are considered to be useful for predicting daily stock price,

Textural information should be represented as a fixed-length vector without losing semantics of the words. Lavrenko et al. [7] employed Bag-of-Words based document representation. However, the method is not capable of preserving word order or semantics in the original document.

A model should deal with time series data since stock price data is also time series. Schumaker et al. [8] and Ding et al. [9] employed SVM and deep neural network, respectively, although these models consider the influence of past events to have a fixed duration. While many events have a similar duration of effect, some words, such as "financial crisis", will have long-lasting influences. To capture such influences, a model should take time series data into consideration. Many of previous works used only one of textual, numerical, or image information for stock price prediction, and their model was trained with consideration about a single company. Nevertheless, it is desirable for the prediction model to consider multimodal information and multiple companies at the same time since investors make decisions depending on various factors such as relationships between companies.

3. Proposed Approach

Fig.1 shows an overview of our approach. Our approach predicts 10 company's closing stock prices by regression from textual and numerical information by using LSTM, which can memorize the previous timesteps due to its architecture. We use multiple companies to learn the correlations between companies. For example, an event like "Nissan recalls..." might make Nissan's stock price decrease while making the stock price of Toyota (another company in the same industry) to increase at the same time. We decide that the number of predicting companies to be 10 due to computation time constraint.

3.1. Representation of textual information

We employ the technique of Paragraph Vector in order to obtain the distributed representation by mapping variable-length pieces of texts to a fixed-length vector. Paragraph

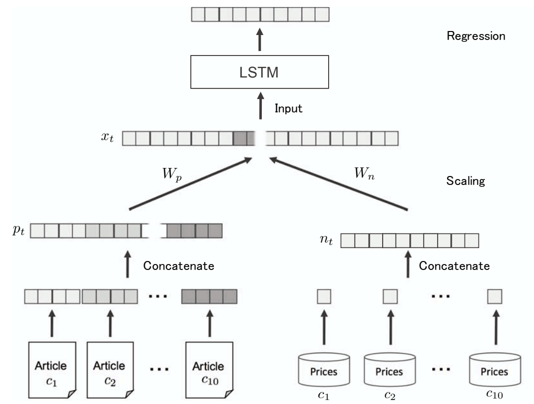

Figure 1. A graphical structure of the proposed method.

Vector can be classified into two categories, Distributed Memory Model of Paragraph Vectors (PV-DM) and Distributed Bag of Words version of Paragraph Vector (PV-DBOW). The main structure of these two Paragraph Vector methods is to predict word(s) in a context (a sequence of training words). The proposer of Paragraph Vector, Le et al., recommended [11] the combination with these models and we use both at the same time.

Every article is represented as a vector of fixed-length d_a by using two kind methods of Paragraph Vector as mentioned above, and then we concatenate the distributed representations of articles concerning 10 companies to make the group article vector $\boldsymbol{p_t}$. Hence, the vector $\boldsymbol{p_t}$ has 10 times the dimensions of a vector of each article.

However, this method assumes that there is always one article about every company published at every timestep, which is rare case in reality. To deal with this, we first fix the position of each company in the vector, and if there are no article about a certain company in a timestep, we insert zero vector instead. In the case that multiple articles about a company are published at a single timestep, we average these article's distributed representations. For example, we assume that each companies denoted as $\{c_1, c_2, \ldots, c_{10}\}$, and if there are no articles about company c_2 and two articles about c_{10}, the vector of article group at timestep t is

$$\boldsymbol{p_t} = \left\{ \boldsymbol{a^t_{c_1}}, \boldsymbol{0}, \ldots, \frac{\boldsymbol{a^t_{c_{10,1}}} + \boldsymbol{a^t_{c_{10,2}}}}{2} \right\}, \tag{1}$$

where $\boldsymbol{a^t_{c_n}}$ is the article distributed representation about company c_n at timestep t.

3.2. Representation of numerical information

We denote our numerical information (stock prices) of all 10 companies at timestep t as vector $\boldsymbol{n_t}$. Since the scale of price depends on companies, we normalize stock price.

We normalize the price for a given stock to be within the range of "[-1, 1]", by the following fomula:

$$value^t_{c_n} = \frac{2 * price^t_{c_n} - (max_{c_n} + min_{c_n})}{max_{c_n} - min_{c_n}}, \tag{2}$$

where $price_{c_n}^t$ is the stock price of company c_n at day t, and max_{c_n} and min_{c_n} are the maximum and minimum stock price of c_n during the priod of training dataset respectively.

3.3. Concatenation of textual and numerical information

The input vector $\boldsymbol{x_t}$ of LSTM is obtained from the combination of article group vector $\boldsymbol{p_t}$ and stock prices vector $\boldsymbol{n_t}$. The two vector $\boldsymbol{p_t}$ and $\boldsymbol{n_t}$ might not be concatenate directly because there is a great difference between the number of their dimensions. This is likely to cause that LSTM will be influenced more heavily by the textual information, which has much more dimensions than stock prices. This imbalance could hinder the accuracy of predictions.

To solve this problem, we scale the size of these vectors in order to get each of them having the amount of same dimensions. In other words, we extend/reduce the dimensions of the vectors $\boldsymbol{p_t}$ and $\boldsymbol{n_t}$ to be the half of dimensions d_x of the input vector $\boldsymbol{x_t}$. We utilize a neural network as the method of scaling. By using the neural network, we transform the vector into any dimensions. We also expect our neural network to learn information important for stock price prediction during training. Input vector $\boldsymbol{x_t}$ is made by concatenating $\boldsymbol{P_t}$ and $\boldsymbol{N_t}$, which are shrunken $\boldsymbol{p_t}$ and extended $\boldsymbol{n_t}$,respectively, where $\boldsymbol{P_t}$, $\boldsymbol{N_t}$ and $\boldsymbol{x_t}$ are computed as follows:

$$\boldsymbol{P_t} = W_p \boldsymbol{p_t} + \boldsymbol{b_p}, \tag{3}$$

$$\boldsymbol{N_t} = W_n \boldsymbol{n_t} + \boldsymbol{b_n}, \tag{4}$$

$$\boldsymbol{x_t} = \begin{bmatrix} P_t & N_t \end{bmatrix}, \tag{5}$$

where the matrices $W_p \in \mathbb{R}^{d_x \times d_p}$ and $W_n \in \mathbb{R}^{d_x \times d_n}$ represent the weights, and $\boldsymbol{b_p}$ and $\boldsymbol{b_n}$ represent their biases.

4. Evaluation

In this section, we examined the validity of the approach with real-world newspaper articles and stock price data. We checked qualitatively whether the vector represented by Paragraph Vector was able to express events shown in the articles of the article event, and summarized the results of market simulation of our model.

4.1. Experimental settings

We used the morning edition of the Nikkei newspaper published from 2001 to 2008 for our experiments, with the news from year 2001 to 2006 as the training data, 2007 as validation data, and 2008 as test data. The target 10 companies were chosen from Nikkei 225 and the same industries. We chose the 10 companies that most frequently appeared in the news articles for the entire period.

To acquire fixed-length vector representation for articles by Paragraph Vector, we first parsed the articles into words to get the vocabulary through a morphological analysis. We used MeCab [19] as the morphological analyzer, and added Wikipedia entries and Nihon Keizai Shinbun's keywords to the dictionary of MeCab to deal with possible proper nouns. Moreover, Ding et al. [9] suggest that news titles alone are able to provide sufficent information to represent news articles and more helpful for prediction compared to the article's contents. Hence, our experiments also used distributed representation of only news titles.

TABLE 1. TOP 10 HEADLINES AND THEIR COSINE SIMILARITIES WITH RESPECT TO THE OBJECT HEADLINE.

object	トヨタ 系 部品 2 社 合併 発表 。 Announcement of consolidation of two suppliers affiliated with Toyota.
similarity	**headline**
0.794798	日産系部品 2 社の合併承認。 Approval of consolidation of two suppulier affiliated with Nissan.
0.681050	日興系ネット証券合併発表。 Announcement of consolidation of stock company affiliated with Nikko Securities
0.648905	三菱商事系関西電炉 2 社合併、来年 4 月メド合意。 Two electric steel comp. affiliated with Mitsubishi in the Kansai region agree about the consolidation.
0.632631	日産、グループ金融 4 社統合発表。 Nissan announces the merger of affiliated four financial companies.
0.626224	神奈川県内ディーラー再編 ——トヨタ系、高級車チャネル、日産系、販社 2 社合併。 Dealers in Kanagawa reorganize —— consolidation of deluxe car channel affiliated Toyota and sales company affiliated Nissan.
0.593248	三菱自、部品生産移管——トヨタ系にも。 Mitsubishi moters conducts car production transfer —— also influenced on Toyota.
0.590510	ハイブリッド部品 10 万台分、トヨタ、日産に供給——提携発表。 Toyota supplies parts of 100 thousand hybrid cars with Nissan —— announcement about cooperation.
0.587197	富士スピードウェイ、トヨタ、買収交渉。 Toyota negotiates acquisition of Fuji International Speedway.
0.580533	日産、日産ディーゼル、トラック共同開発発表。 Nissan and Nissan diesel motor announce about co-development of truck.
0.579858	GM・富士重、資本提携を発表——スズキ・富士重も交渉。 GM and Fuji heavy industries announce about capital alliance —— Suzuki and Fuji Heavy Industries are also in negotiation.

4.2. Acquiring distributed representation

As mentioned in Section 3.1, we converted articles into their distributions, and combined those into article groups. We evaluated the validity of these approaches, the results are shown below.

4.2.1. Distributed representation of articles. In terms of distributed representation, words and sentences which have similar meaning are mapped into similar positions in the vector space. Hence, cosine similarity of similar meaning articles are high. We calculated the cosine similarity between a vector of certain article with respect to others. We took up the top 10 articles that had the highest cosine similarity to evaluate whether these articles did indeed have similar meanings.

Table 1 shows that top 10 articles and their cosine similarities in regard to the article "トヨタ系部品 2 社合併発表。(Announcement of consolidation of two suppliers affiliated with Toyota.)". Name of auto manufacturer such as "Toyota" and "Nissan" appears in the almost all of the top 10 news title. Moreover, many of news titles have "consolidation" and similar words like “merger” or “cooperation”. These 10 news titles do indeed have similar meanings when compared to the object. It appears that Paragraph Vector is capable of capturing the article’s meanings as indeed.

4.2.2. Distributed representation of article groups. We evaluated the distributed representation of article groups by combining each article representations whether considered the meaning with cosine similarity, same as for articles. However, owing to limitations of space, we present only the group with the highest cosine similarity. Note that, for the purpose of this evaluation, we combined articles by their dates, and thus expected to find the day that have the most similar articles, if our approach did indeed work.

Table 2 shows two article groups. The upper part shows the object group and lower part shows the group which has the highest similarity score. We concluded that these articles were similar by comparing the contents of articles among each like “vise president, company funeral” and “president, farewell party” at the article V and G. Further, the article VI and A have similar since both of them had a negative affect on Fujitsu.

Cosine similarity between these groups was about 0.33. Compared to previous distributed representation of article evaluation, the overall score was smaller. This was however natural since there were few article groups having similar meaning to a certain group. The closer to 1 cosine similarity gets, the more two vectors are getting towards perfect matching. In a article group, perfect matching would imply every single article about the 10 companies is reporting the same events on the same day; this is extremely unlikely to occur. Thus, the cosine similarity 0.33 was sufficient to justify our method of combining articles into groups as valid.

TABLE 2. THE GROUP OF ARTICLES HAVING THE HIGHEST COSINE SIMILARITY WITH RESPECT TO THE OBJECT GROUP WHICH IS UPPER.

I.	数表 業績予想 修正 ・ 配当 異動 。 Table of earnings revision and change of dividend policies.
II.	きょう の 決算 発表 。 Today’s announcement of financial statements.
III.	日韓 の 電子商取引 市場 、 参加 企業 の 募集開始 —— 三菱総研 など 、 来年 、 実証実験 。 Japanese and Korean of e-commerce market starts to invite participant company Mitsubishi Research Institute etc. conduct demonstration experiment in the next year.
IV.	利益 確定 売り で 続落 、 値下がり 銘柄 1000 超す Profit booking makes futher losses, the number of declining stocks exceeds 1000.
V.	**(株式) 故 森永 範 興 氏 (元 NTT ドコモ 副社長) の 社葬 。** The company funeral of Norioki Morinaga (the former vise president of NTT docomo).
VI.	9 月 中間決算 本格 化 、 業績 選別 色 強まる —— 東芝 ・ 富士通 など 下落 。 Interim results of September become serious, and the ability of judge performance is required —— the price of Toshiba and Fujitsu were declined.
VII.	産業空洞化 対策 、 経営者 ら で 会議 、 経産省 が 設立 。 A measure of hollowing out of industry was discussed by managers and established by ministry of economy, trade and industry.
A.	**富士通 、 3000 人 を 異動 ・ 削減 —— 来年 度 、 黒字 目指す 。** Fujitsu made change and fired 3000 people —— go toward the black in the next year.
B.	アル チザ 、 1 月 中間 、 13 % 減益 、 6 億円前後 —— 現行 の 携帯 サービス 向け 不振 。 Artiza’s January interim results decrease 13% in profit, six hundred million yen —— current service for mobile phone is dull.
C.	データ から (3) 複雑 な 金融 税制 —— 個人 投資 妨げる 恐れ From data (3), a complex financial tax system —— it is in danger of preventing individual investor.
D.	(税 を ただす)「 日本 買い 」 期待 膨らむ —— 時価総額 上位 銘柄 が 堅調 (Correct tax) ”Buy Japanese” to be full of hope —— brands which are higher lank of market capitalization are steady.
E.	(株式) 長 時間 録音 の IC レコーダー —— 東芝 A IC recorder which is able to record long time —— Toshiba.
F.	(ニューフェース) 三菱商 (下) リスク に 負け ない 組織 へ —— 事業 資産 など 一元管理 (New face) Mistubishi aim not to lose risk —— unitarily manage the business property.
G.	**(会社 研究) 故 田部文一郎 氏 (元 三菱商事 社長) の お別れ会 。** (Company research) A farewell party of Bunichiro Tanabe (the form president of Mitsubishi).

4.3. Market simulation

In this section, we evaluated the results of a market simulation on our approach with parts relaced with other baseline approaches.

4.3.1. Strategy on market simulation. Our approach receives the opening prices of day t and vectors representing news titles about 10 companies, and predicts the close prices by regression. We simulated real-world stock trading by the strategy proposed by Lavrenko et al. [7], which was given by

$$r_{c_n}(t) = \frac{closing^{pred}_{c_n}(t) - opening^{true}_{c_n}(t)}{opening^{true}_{c_n}(t)}, \qquad (6)$$

$$gain_{c_n}(t) = \begin{cases} buy \rightarrow sell & (r_{c_n}(t) > 0), \\ sell \rightarrow buy & (r_{c_n}(t) < 0), \end{cases} \qquad (7)$$

where $buy \rightarrow sell$ denotes a transaction purchasing stocks at the opening price. During the holding time, if a stock is able to make profits of 2% or more, the stock is sold instantly. Otherwise, the stock is sold at the closing price. We used the same strategy for short selling. $sell \rightarrow buy$ denotes a transaction selling short at the opening price. If a stock price decrease 2% or more at the opening price, the stock is bought instantly. Otherwise, the stock is bought at the closing price. The upper limit of a purchase was 1,000,000 Japanese yen per company. We made our transaction costs as close to 1,000,000 Japanese yen as possible depending on the share unit number[1].

4.3.2. Parameters of our approach. The parameters of our approach selected using validation data are as follows.

Paragraph Vector. The sliding window size (the length of words considered in Paragraph Vector) of PV-DM and PV-DBOW was 9. In each model, the learned vector representations had 100 dimensions. Hence the dimensions of an article’s distributed representation d_a which was the concatenation of two vectors was 200, and the number of dimensions of article group d_p was 2000.

Input data. We reduced the dimensions of article group vector $\boldsymbol{p_t}$ from 2000 to 500, and extended the dimensions of opening price vector $\boldsymbol{n_t}$ from 10 to 500. Thus making the input vector $\boldsymbol{x_t}$ to have 1000 dimensions, and the target industries were “Transportation Equipment”, “Pharmaceutical”, “Machinery”, “Electronics” and “Wholesale Trade”. The reason we chose them was that there were the

1. The unit of the number of purchasing/selling stock.

TABLE 3. COMPARISON OF TRADE GAINS WITH INPUT VARIATIONS ON FIVE INDUSTRIES.

(Ten Thousands of Yen)

Industry/model	Num	BoW+Num	PV+Num (ours)
Transportation Equipment	272.15	433.33	**548.02**
Pharmaceutical	148.14	214.80	**263.6**
Machinery	−25.06	−214.46	**122.75**
Electronics	73.72	54.69	**87.46**
Wholesale Trade	**255.38**	−58.28	190.07
Overall	724.33	430.08	**1211.90**

TABLE 4. COMPARISON OF TRADE GAINS WITH LEARNER VARIATIONS ON 5 INDUSTRIES.

(Ten Thousands of Yen)

Industry/model	SVR	MLP	Simple-RNN	LSTM (ours)
Transportation Equipment	−21.34	−38.33	211.32	**548.02**
Pharmaceutical	48.67	19.56	70.01	**263.6**
Machinery	−154.72	−341.41	−321.29	**122.75**
Electronics	25.91	−204.97	**129.55**	87.46
Wholesale Trade	55.02	4.01	172.95	**190.07**
Overall	−46.96	−561.14	262.54	**1211.90**

only industies that had more than 10 companies in Nikkei 225.

LSTM. The size of each minibatch was 30, LSTM had one layer and were unrolled 20 steps, and on the basis of Wojciech et al. [20], we applyed 50% dropout on the non-recurrent connections. We trained the LSTM for 50 epochs with Adam [21].

4.3.3. Results. We evaluated the effectiveness of our approach with the following experiments:

Effectiveness of Paragraph Vector. We compared with the two baseline methods by replacing input in order to evaluate the effectiveness of Paragraph Vector. The first baseline used only the numerical data as input to investigate the effectiveness of textual information. The input vector $\boldsymbol{x_t}$ of this method had 500 dimensions and no dropout. Second baseline employed the vector represented by BoW instead of Paragraph Vector as textual information. We applyed feature selection method on the basis of the word's frequency of occurrence to improve prediction accuracies. All words were sorted in descending order of their occurrence, and words between 30% and 90% were selected as the vocablary. The number of words in our vocablaries is 9,153 / 7,891 / 10,963 / 9,916 / 4,334 of the industries from Section 4.3.2 respectively.

The experimental results are shown in Table 3. Our approach was able to make the more profits from all industries compared to using BoW (BoW+Num). The overall profit of our approach was 1211.90, much higher than that 430.08 of BoW+Num. Comparing with the method that used only numerical information as inputs (Num), our approach got more profits in the four out of five industries and the total was also higher by about 490 points. From the above results, we concluded that it was indeed effective for stock price prediction to employ distributed representations by utilizing Paragraph Vector.

TABLE 5. WITH VARIATIONS IN NUMBER OF THE OBJECT COMPANIES, COMPARISON OF TRADE GAINS ON 5 INDUSTRIES.

(Ten Thousands of Yen)

Industry/model	one-Co.	same-Ind. (ours)
Transportation Equipment	−310.67	**548.02**
Pharmaceutical	170.5	**263.6**
Machinery	−588.83	**122.75**
Electronics	−297.42	**87.46**
Wholesale Trade	−158.01	**190.07**
Overall	−1184.43	**1211.90**

TABLE 6. COMPARISON OF TRADE GAINS ON 10 BRANDS.

(Ten Thousands of Yen)

brand code/model	one-Co.	diff-Ind.	same-Ind.(ours)
7203	23.39	−6.23	**37.11**
7201	−47.80	−47.77	**−27.26**
4502	−4.29	9.97	**12.85**
4452	**57.23**	15.93	28.55
6301	29.19	−38.22	**63.11**
6367	−5.88	24.02	**30.38**
6758	−45.55	−51.14	**−13.3**
6501	4.34	−28.80	**62.0**
8058	−26.21	43.06	**50.93**
8031	−3.43	**64.90**	39.82
Overall	−19.01	62.14	**284.19**

Effectiveness of LSTM. In the next experiment, we evaluated the performance of LSTM by replacing it with three baseline classifiers. The first baseline classifier was Multi Layer Perceptron (MLP). The MLP we used have two layers with 500 and 250 units. Second baseline classifier was Support Vector Regression with Radial Basis Function (RBF) as its kernel and parameters γ and C were chosen by grid search for each brands. The searching ranges for γ and C were $[10^{-4}, 10^{0}]$ and $[10^{0}, 10^{4}]$, respectively. We verified the necessity of considering time series data in comparison with MLP and SVR which do not consider. The other classifier was RNN (Simple-RNN) to evaluate that LSTM cells was able to capture the time series influences which were caused by news events.

The results are reported in Table 4. As we can see, the profits of our model over all industries were higher than the MLP. In comparison with SVR, our model received higher profits in four out of five industries. Moreover, RNN and LSTM were classifiers that received any profits at all, which indicated a model considering time series data was essential for stock price prediction. LSTM had significantly outperformed Simple-RNN in spite of both model were capable of considering time series data. The difference of these networks is that the dynamics is either deterministic or nondeterministic transactions from previous to current hidden states. Since LSTM, being nondeterministic transactions got higher profits, it was assumed that LSTM was able to capture the fluctuating time series changes well.

Effectiveness of considering multiple companies in the same industry. Finally, we performed two experiments to evaluate the effectiveness of considering multiple companies within the same industry. In the first experiment, we compared our result with a baseline method where learned and predicted only one brand without the context of others. This was done for each of the 10 comanies to evaluate the effectiveness of considering multiple companies. Table 5 shows the experimental results of our model considering a separated single company (one-Co.) and that considering 10 companies simultaneously (same-Ind.). By examining the results in table 5, we found that both individual industries and the total amount of earnings of our model were higher than one-Co.

In the second experiment, we chose two companies from each of the five industries, and evaluated our method against two other baseline methods. This experiment was intended to evaluate the effectiveness of considering the same industry. Table 6 shows the experimental results where our model, one-Co. and the model of considering 10 companies that belongs to different industies together (diff-Ind.). We compared with the results of the target brands of diff-Ind., since could not compare among industries. By examining the results in table 6, the total earnings of diff-Ind. was higher than that of one-Co.. Hence, it is effective for predicting stock prices to consider multiple companies together. Table 6 also shows that our model got higher earnings in the nine brands compared with diff-Ind. because our approach captured correlations between companies in the same industry.

5. Conclusion

This paper proposed an approach to predict stock prices by employing distributed representations of news articles and considering the correlations between multiple companies within the same industry. In our approach, a recurrent network captured the changes of time series influence on stock price.

To evaluate the effectiveness of the approach, we conducted experiments on market simulation. Experimental results showed that distributed representations of textual information are better than the numerical-data-only methods and Bag-of-Words based methods, LSTM was capable of capturing time series influence of input data than other models, and considering the companies in the same industry was effective for stock price prediction.

For future work, we would like to incoorporate more technical indices such as the moving average (MA) and the moving average convergence divergence (MACD) for better profit-making capabilities.

The authors would thank to Associate Professor Kazuhiro Seki of Konan University for valuable discussions. This work was partially supported by the JSPS KAKENHI (16K12487 and 26280040).

References

[1] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1, pp. 307–319, 2003.

[2] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Language models for financial news recommendation," in *Proceedings of the 9th International Conference on Information and Knowledge Management (CIKM-00)*, 2000, pp. 389–396.

[3] F. E. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, pp. 309–317, 2001.

[4] B. Bick, H. Kraft, and C. Munk, "Solving constrained consumption-investment problems by simulation of artificial market strategies," *Management Science*, vol. 59, no. 2, pp. 485–503, 2013.

[5] B. Adrian and L. Nathan, "An introduction to artificial prediction markets for classification," *Journal of Machine Learning Research*, vol. 13, pp. 2177–2204, 2012.

[6] A. Devitt and K. Ahmad, "Sentiment polarity identification in financial news: A cohesion-based approach," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL-07)*, June 2007, pp. 984–991.

[7] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, "Mining of concurrent text and time series," in *Proceedings of the KDD-2000 Workshop on Text Mining*, 2000, pp. 37–44.

[8] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The AZFin text system," *ACM Transactions on Information Systems (TOIS-09)*, vol. 27, no. 2, pp. 12:1–12:19, 2009.

[9] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Using structured events to predict stock price movement: An empirical investigation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*. Association for Computational Linguistics, 2014, pp. 1415–1425.

[10] M. Hagenau, M. Liebmann, M. Hedwig, and D. Neumann, "Automated news reading: Stock price prediction based on financial news using context-specific features," in *System Sciences, 2012. Proceedings of the 45th Annual Hawaii International Conference on System Sciences (HICSS-12)*. IEEE, 2012, pp. 1040–1049.

[11] M. Robert, R. Bharath, S. Mohammad, K. Gert, and P. Vijay, "Understanding protein dynamics with l1-regularized reversible hidden markov models," in *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 2014, pp. 1197–1205. [Online]. Available: http://jmlr.org/proceedings/papers/v32/mcgibbon14.pdf

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with lstm," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

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

[15] K. Izumi, T. Goto, and T. Matsui, "Trading tests of long-term market forecast by text mining," in *Proceedings of the tenth IEEE International Conference on Data Mining Workshops (ICDM-10)*, 2010, pp. 935–942.

[16] B. Johan and M. Huina, "Twitter mood as a stock market predictor," *IEEE Computer*, vol. 44, no. 10, pp. 91–94, 2011.

[17] M. andr Mittermayer, "Forecasting intraday stock price trends with text mining techniques," in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS-04)*, 2004.

[18] S. Jianfeng, M. Arjun, L. Bing, J. P. Sinno, L. Qing, and L. Huayi, "Exploiting social relations and sentiment for stock prediction," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*, 2014, pp. 1139–1145.

[19] T. Kudo, "Mecab: Yet another part-of-speech and morphological analyzer," *http://mecab.sourceforge.net/*, 2005.

[20] Z. Wojciech, S. Ilya, and V. Oriol, "Recurrent neural network regularization," *CoRR*, vol. abs/1409.2329, 2014. [Online]. Available: http://arxiv.org/abs/1409.2329

[21] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980