PACIS 2018 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

6-26-2018

Predicting Stock Price Movement Direction with Enterprise Knowledge Graph

Yuxin Zhang
Renmin University of China, ZYX1207214064@gmail.com

Kunlin Yang
Renmin University of China, kunliny@ruc.edu.cn

Wei Du
Renmin University of China, ahduwei@ruc.edu.cn

Wei Xu
Renmin University of China, weixu@ruc.edu.cn

Predicting Stock Price Movement Direction with Enterprise Knowledge Graph

Completed Research Paper

Yuxin Zhang
School of Information, Renmin University of China, Beijing, 100872, P.R. China
ZYX1207214064@gmail.com

Kunlin Yang
School of Information, Renmin University of China, Beijing, 100872, P.R. China
kunliny@ruc.edu.cn

Wei Du
School of Information, Renmin University of China, Beijing, 100872, P.R. China
ahduwei@ruc.edu.cn

Wei Xu
School of Information, and Smart City Research Center, Renmin University of China, Beijing, 100872, P.R. China
weixu@ruc.edu.cn

Abstract

Predicting stock price movement direction is a challenging task for financial investment. Previous researches focused on investigating the impacts of external factors (e.g., big events, economic influence and sentiments) in combination with the historical price to predict short-term stock price movement, while few researches leveraged the power of various relationships among enterprises. To bridge this gap, this research proposes power vector model and influence propagation model to mine the rich information in constructed Enterprise Knowledge Graph (EKG) for price movement prediction. In addition, Deep Neural Network (DNN) is introduced to train the model. The proposed model shows good prediction performance on the dataset of China top 500 enterprises.

Keywords: Enterprise Knowledge Graph, Stock Price Movement Direction Prediction, Influence Propagation, Deep Neural Network

Introduction

Enterprise Knowledge Graph (EKG) has been successfully applied in the financial field recently given its advantages in covering enterprise relationships. For instance, in 2016, EKG was utilized in investment consulting for securities companies (Ruan et al. 2016). In 2017, researchers studied EKG's application in mining investment relationships between enterprises and organizations based on the path-searching method (Hu et al. 2017). Here we explore a new EKG application in short-term stock price movement direction prediction.

It has been proved that part of enterprises' interaction will impact their business performance, like the cooperative and competitive relationships' influence according to research (Gnyawali et al.2001). On account of that the stock price can reflect the enterprise's business performance, we believe relevant enterprise will influence the focal enterprise's stock price.

In our model, we solve the problem of calculating and integrating influence among enterprises in a complex business network including diverse relationships. We utilize EKG analysis to calculate the influence of specific relationship from the relevant enterprise. Based on research (Adebiyi et al. 2014), we integrate influence from all relevant enterprises in combination with the historical stock price to predict the short-term price movement direction in the future. This research employs Influence Propagation Model to calculate influence between any given two enterprises in EKG, and in addition adopts DNN (Deep Neural Network) to integrate influence from several relevant enterprises.

In the following section 2, we will state our related work. In section 3, we will explain the model's framework in detail. In section 4, EKG's construction and representation will be stated. Section 5 present how to calculate the influence of relevant enterprises and utilize them to predict the future stock price movement direction. Section 6 provides empirical evaluation part. Conclusions are given at last.

Related Work

Stock Price Movement Direction Prediction

Several researches about stock price movement direction predictions focused on the external impacts on stock price movement, and part of them designed algorithms for utilizing external impacts to better predict future stock price movement. There are several types of external impacts, like the big events, economic influence and sentiments (Kohara et al. 1997; Park et al. 2013; Oh et al. 2013). Many researches adopted events to predict the stock price (Kohara et al. 1997; Ding et al. 2015). Sentiments are also popular factors among various relevant studies (Oh et al. 2013; Thien Hai et al. 2015; Deng et al. 2011). In 2015, researches proposed to use sentiment automatically extracted from social media to predict the future stock price movement (Thien Hai et al. 2015). Information from financial news or reports were also regarded as significant factors (Gidofalvi et al. 2001; Schumaker et al. 2009; Lin et al. 2011). In 2011, weighting scheme was utilized to integrate qualitative and quantitative factors from financial reports to predict the future stock movement direction (Lin et al.2011). In 2017, researchers utilized the deep learning framework to predict the news-oriented stock price trend (Hu et al. 2017).

Beyond that, researches also discussed the impacts of relevant enterprises' business performance. For instance, researchers proposed to predict the stock price movement direction based on a dynamic cooperative and competitive business network analysis (Zhang et al. 2016). They design Energy Cascade Model (ECM) to predict the stock price trend. However, the study only considered the influence of partners and rivals, while there are more complex relationships between enterprises, such as the controlling and controlled relationship between focal enterprise and its associated enterprises. Here associated enterprises refer to enterprises in which focal enterprise owns the significant portion of voting. It can be foreseen that focal enterprise's associated enterprises will also impact its stock price, on account of common profits shared by focal enterprise and its associated enterprises. Corporate Shareholders also influence the focal enterprise' business performance due to their participation in its governance. Because of Enterprise Knowledge Graph (EKG)'s wide coverage of different types of relevant enterprises, we utilize EKG analysis to calculate the influence of various relevant enterprises.

Influence Propagation in Graph

In our research, our influence calculation demands influence propagation pattern simulation in EKG. Previous researches have studied the influence propagation in the graph. In 2003, researchers constructed Independent Cascade (IC) model to simulate influence propagation in a social network (Kempe et al. 2013), however, here social network only contains one relationship between nodes and influence only travels in one type's link. EKG is actually a network with a single type of nodes and different types of links, and we should allow influence travel through different types of links. In the ECM (Zhang et al. 2016), influence can travel both in the cooperative link and competitive link, while this research utilized the sign of link's weight to differentiate cooperative link and competitive link. Because of four types of links' existence in our network, we can't simply employ the sign to tell the weights' type differences. In the Multi-relational Influence Propagation (MRIP) model (Yang et al. in 2012), the research achieved to propagate influence in a multi-relational network by calculating the correlation between different types of links. The correlations of different types of links actually are correlations of different sub-graphs, and they utilize conditional probabilities to represent the correlations. However, the conditional probabilities can't catch all topological features. In research (Dai et al. 2017), the belief vector method was proposed to better capture the similarities between sub-graphs' topological features. To combine all method's merits, in our influence calculation, we utilize the power vector method modified by belief vector method to replace conditional probability method in MRIP. Then we use MRIP method to calculate influence between every two nodes in EKG.

Model Framework

As depicted in Figure 1, there are four main steps in our model. The first step is to build the EKG. The second step is to calculate the influence value of relevant firms with influence propagation model and power vector method. The third step is to gain the stock information of enterprises in EKG and preprocess the data. After we gained the influence indicators' value, we need to integrate it with the stock price change. The last step is to convert the relevant firms' influence indicators to stock price influence indicators, and then combine them with some of the focal firm's basic historical stock performances as input to DNN for predicting the focal firm's short-term stock price movement direction.

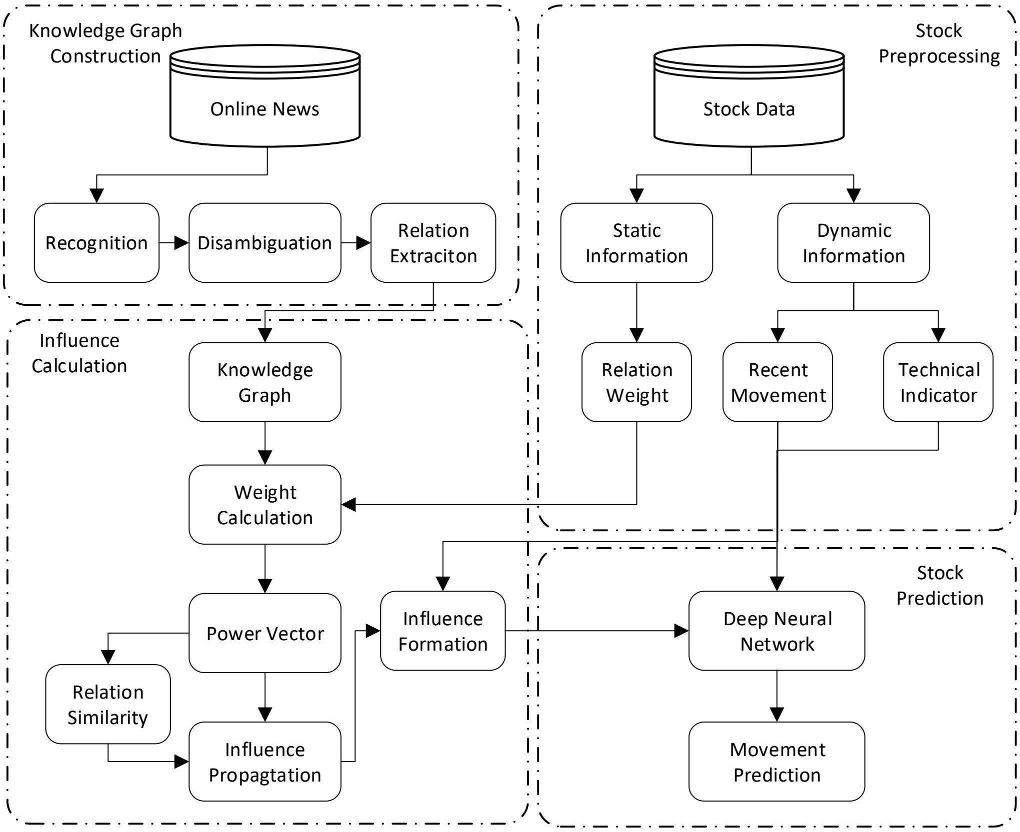

Figure 1. Model Framework

EKG Construction and Representation

EKG Construction

Our enterprise knowledge graph contains Top 500 enterprises in China and their partners, rivals, corporate shareholders and associated enterprises. Specifically, controlling and controlled links can be found in the accounting reports. Cooperation links can be identified from enterprises' financial news. Enterprises who have overlapped business scopes are treated as potential rivals.

We collect data from diverse sources, including web pages and non-textual data (e. g, PDF files). For instance, we extract data from XML files converted from PDF files, and we also extract web pages' textual data by webpage analysis.

Then we gain raw data containing the related enterprises' information and we employ Named Entities Recognition (NER) and Neural Relation Extraction (NRE) to process the documents in raw data. NER aims at recognizing enterprises in the documents. Furthermore, for disambiguation, if two recognized entities are actually the same enterprise, we replace them by the enterprise's whole name in Wikipedia. Then we attempt to link the recognized entities by NRE. Convolutional Neural Network (CNN) is adopted for relation extraction (Nguyen et al. 2015).

EKG Representation

The multi-relational network transformed by EKG can be represented by $G = (V, E)$, where V stands for the set of enterprise vertices, and E is the set of different types of relationship links connected between enterprises. In our EKG, $E = E_1 \cup E_2 \cup E_3 \cup E_4$, where E_1 is the cooperative link, E_2 is the competitive link, E_3 is controlled link, and E_4 is controlling link between enterprises. We denote the sub-graph as $G_r = (V, E_r), (r = 1,2,3,4)$. It can be noticed that G_1 and G_2 are undirected graphs, while G_3 and G_4 are directed graphs. We use $weight(i, j, r)$ to represent the weight of r-th type link between source node V_i and sink node V_j.

As for different relationships between enterprises, different ways can be used to assign weights for the links. Previous research has validated that earning per share determines PER (Price-to-Earning Ratio), and enterprise's business performance is closely related to earning per share (Zarowin et al). We consider PER as a key metric that reflects the enterprise's business performance. An acceptable statement is that if a partner performs and develops well in business, it will promote positive influence on the focal enterprise, and similarly if a rival has a good business performance, it will promote more threat on the focal enterprise. Thus, for both cooperative and competitive relationship, larger PER of relevant enterprises can lead to greater cooperative or competitive influence strength, then $\mathrm{LOG}(PE_h) * \mathrm{LOG}(PE_t)$ can be employed as cooperative and competitive links' weight between h and t nodes. For the controlling and controlled relationship, we may use controlling and controlled proportion. Then we normalize them respectively to [0,1] as the weight for the link of various types between different vertices.

Influence Calculation in EKG

Influence Propagation Pattern

Our model intends to calculate the influence of specific relationship between the focal enterprise and its relevant enterprises, such as the cooperative influence from enterprise's partner. Given that all enterprises may have indirect relationships with the focal enterprise, we regard all other enterprises in the EKG as focal enterprise's relevant enterprises, and a relevant enterprise can also be relevant in other relationships. For instance, as Figure 2 shows, A can be C's indirect partner. It is likely that the influence of cooperative relationship propagates through the cooperative edge between A and B, then the cooperative edge between B and C. Thus, A will have an influence on C. Notice that specific influence can also travel along other types of relationship links. A can be D's indirect partner too even if there is no path consisting of cooperative links between A and D. Part of cooperative influence travels through controlling link from B to D. And if the link between B and D is cooperative link, A may have more significant cooperative influence on D. Therefore, influence of specific relationship propagation should be differentiated between through its own relationship links and other types of relationship links. In the influence propagating process, the influence's significance will weaken with the increase of propagating distance. In the conclusion, influence propagation patterns should obey the following rules:

- The influence of specific relationship can travel through other types of relationship links. The significance's loss of traveling through other types of relationship links is determined by the interrelation rates between different types of links.
- The influence significance will weaken as it travels through more enterprise nodes
- The influence can travel in multiple paths to reach the final target enterprise node, once influence reaches the target enterprise node, the influence propagating process terminates.

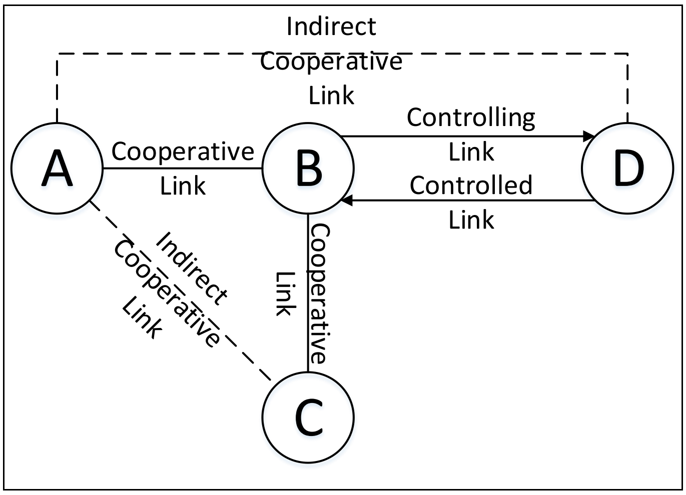

Figure 2. EKG Schema

In our influence propagation model, our target is to calculate the influence of specific relationship between two nodes in EKG. It can be denoted as $\mathrm{I(u, v, i)}$, which means the influence enterprise node u has on enterprise node v in i-th type relationship. In the modeling process of influence propagation in EKG, we should solve two main problems, one of which is how to calculate the interrelation rate between different types of relationship links, the other is how to model the influence propagation in a long distance. To solve these problems, we adopt two propagation processes, one is energy propagation for calculating the correlation between different relationships, and another is influence propagation for measuring the influence of indirectly related enterprises.

Correlation between Different Types of Links

In this section, we utilize a power vector method modified from belief vector model to measure the correlation between different types of relationship links, which is actually the similarity between different sub-graphs. The power vector method's prototype, belief vector method (Dai et al. 2017) was first proposed in 2017 as an effective method for relationship similarity estimation. To introduce the model, firstly we need to clarify some concepts. We create two main items in the model, one is power vector, the other is graph energy. In the power vector method, each node is assigned a power value. Power value is gained by energy accumulation, where energy is a virtual item propagating in the graph. Energy can be transmitted from each node to the neighbor nodes and modified recursively in the propagation. The energy value modification is iterative. This process will finally converge after a certain number of iterations.

In the energy propagation model, firstly we need to redefine the probability of energy transmitting from node v_i to one of its direct neighbor node v_j in G_r as based on original formula (Dai et al. 2017)

$$\psi(v_i, v_j, r) = \frac{weight(i,j,r)}{\sum_{k \epsilon N(i)} weight(i,k,r)} \tag{1}$$

where *N(i)* is the set of node *i*'s direct neighbor nodes.

The energy is passed iteratively between each node and its neighbor nodes, until convergence or a maximum iteration set by the user. Here we use $E_{ij}^{(r)}(v_j)$ to represent the energy from node v_i to node v_j in G_r. We let $E_{ij}^{(r)}(\mathrm{v_j})$ initial value be $\frac{1}{n+1}$, because we assume there are n vertices in each sub-graph. In the previous research (Dai et al. 2017), this value can be updated through energy propagation by the following rules:

$$E_{ji}^{(r)}(v_i) = \psi(v_j, v_i, r) \prod_{k \epsilon N(j) \backslash i} E_{kj}^{(r)}(v_j) \tag{2}$$

Here we modify it as the following rules, because it can better fit in our research:

$$E_{ji}^{(r)}(v_i) = \psi(v_j, v_i, r) \sum_{k \epsilon N(j) \backslash i} E_{kj}^{(r)}(v_j) \tag{3}$$

High power value is assigned to the nodes which have high energy value aggregation after convergence. A node's power value is proportional to the total energy value it receives. Due to that node v_i will easily has high energy value with many neighbor nodes, we define the power value of node v_i in the sub-graph G_r as:

$$P_{ri} = \frac{\sum_{j \epsilon N(i)} E_{ji}^{(r)}(v_i)}{\sum_{j \epsilon N(i)} \ weight(i,j,r)} \tag{4}$$

For sub-graph G_r consisting of r-th type of edge, we denote its power vector as $(P_{r1}, \ldots, P_{rn})$. The power vector can reflect sub-graph's both local and global topology. Any small difference can be spotted by comparing power matrices. Thus, by measuring the power matrices similarity, we can gain the sub-graphs similarity, and we let it be the links interrelation. Then we use Pearson coefficient to estimate similarity between two power matrices of sub-graph G_i and G_j, the similarity can be measured by the following equation:

$$\sigma(i,j) = \frac{n \sum_{k=1}^{n} (P_{ik} - P_i)(P_{jk} - P_j)}{\sum_{k=1}^{n} (P_{ik} - P_i)^2 \sum_{k=1}^{n} (P_{jk} - P_j)^2} \tag{5}$$

here $P_i = \frac{1}{n} \sum_{k=1}^{n} P_{ik}$

Influence Propagation Model

In this section, we will calculate the influence between any two given nodes in the graph, and the two nodes are not necessarily connected. First, we adopt the MRIP method (Yang et al. in 2012) to calculate the influence propagation pattern between two neighbor nodes, and we utilize the power vector method to replace the original link correlation measure method of MRIP. Second, we utilize the breadth-first search to propagate influence to any reachable node in the graph as many researches did before.

Here we intend to study the influence of *i*-th relationship the relevant enterprise v has on the focal enterprise *u*. Firstly we introduce influence value *I* to our model, and *I(v, u, i)* refers to the significance of the influence of *i*-th relationship that node v has on node u. We define the *I(v, v, i)=1*. Because a node's influence can propagate to its neighbors in a probability, the nodes with not influence value before can get part of influence energy from their neighbor nodes which have been assigned energy already. Thus, each node's energy can be calculated by influence propagation pattern between neighbor nodes. As long as we calculate the influence flow between two neighbor nodes, we can assign blank nodes with influence value. If we intend to calculate the *I(v, u, i)*, then the enterprise node *v* should be the influence propagation source, and the focal enterprise node *u* should be the target node. Once the target node is assigned an influence value, the influence propagation process terminates. The influence propagation will also terminate if the influence has no way to propagate in the graph.

After we fixed the source node, we can gain every reachable node's influence significance by influence propagation patterns between neighbors and breadth-first search procedure. In modeling the influence propagation patterns between neighbors, we need to define the probability of influence propagating through certain type link. We regard the process influence propagating to the neighbor nodes in the certain type of link the same way as energy transmitting in the power vector method section. Thus the probability that node *v* activates node *u* in *r*-th type of link is also $\psi(v, u, r)$. The difference between influence propagation and energy transmitting is that the influence of specific relationship can travel in other types of links, while energy only travels in a sub-graph at one time.

In the MRIP model(Yang et al. in 2012), the influence *I(v, u, i)* in *i*-th type link propagates between two neighbor nodes u_1 and u_2 is calculated as:

$$I(v,\ u_2, i) = I(v,\ u_1, i) * \beta * \psi(u_1,\ u_2, i)$$
$$+I(v,\ u_1, i) * \beta * \sum_{j \neq i} \big(\sigma(i,j) * \psi(u_1,\ u_2, j)\big) / (|N(u_1,\ u_2)|) \quad (6)$$

Here $\beta = 0.5$ is called katz factor, $I(v,\ u_1, i)$ means the source node (breadth-first search source node *v*) impacts on nodes u_1. $\sigma(i,j)$ means the similarity between link i and link j, and $|N(u_1,\ u_2)|$ is the number of link types between node v and node u. The existence of β is to punish the influence propagation of long distance and a single direction. It can be modified according to the real situations.

Preprocessing for Machine Learning Prediction

After we calculate every type of influence from every enterprise in the enterprise KG to the focal enterprise, we screen the Top K influence for each relation to reduce the indicators' number, where K is a hyper-parameter and related to relation type. However, here influence indicators do not refer directly to the influence on stock price. We need to transform it to the influence on stock price. We denote the focal enterprise as u, and we modify the I(v, u, r) as

$$I(v,u,r) = I(v,u,r) * SA_v \quad (7)$$

Here SA_v means the relevant enterprise *v*'s stock price movement amplitude ratio recently. We integrate the focal enterprise stock information including technical indicators KDJ index, historical stock price movement amplitude ratio with influence from the relevant firm as input to classifier model.

Deep Learning was adopted frequently in classification problem in recent years (Kussul et al. 2017; Qi et al. 2017; Esteva et al. 2017), and it was famous for learning complex properties well in previous experiments (Wang et al. 2017). It was proved to be effective in stock prediction field too. In 2015, Ding et al. discovered that deep learning was effective for event-driven stock prediction (Ding et al. 2015). In 2017, Huynh et al. employed DNN to predict stock price movements direction based on financial news information and stock historical price. (Huynh et al. 2017). And the research concluded that the DNN was a powerful framework for stock prediction. Thus, here we use DNN as stock price movement direction classification model.

Experimental Evaluation

Data Set Description

Our data source can be divided into two main parts, one for stock information and the other for EKG construction.

In terms of stock information, we export data from the wind terminal by using enterprises stock codes, including daily closing price and yearly price earning ratio (PE) in 2016. We label those days by comparing the closing price at time t to $t-1$, larger then assigned 1 else assigned 0, indicating movement goes up and down respectively. When it comes to bull market or bear market, although the focal enterprise's stock price is likely to move in the same direction as related enterprises, the phenomenon may be caused largely by the market effects. To avoid the market effects in causing the movements in the same direction, we exclude stock data in the period of the bull market or bear market. We treat the market as bull market or bear market when stock prices of more than 80% enterprises in our EKG are moving in the same direction. The statistics of labels after filtering are shown in Table 1.

Table 1. Statistics of Labels

Up	Down	Total
12960 (42.90%)	17520 (53.10%)	30210 (100%)

As for knowledge graph, we collect annual reports of Top 500 from 2014 to 2016 and crawled financial news data from Sina Finance, QQ Finance and Fenghuang Finance from June 2014 to Nov, 2016 with 54231 pieces of news after filtering. EKG is constructed based on latest news and accounting reports, and is assumed as constant for simplicity for the purpose of predicting stock price movement. We use the information of all the Top500 and some other enterprises closely related to Top500 to construct our knowledge graph, 2901 enterprises in total. Based on our data source, four relations are extracted for empirical study, which is cooperation, competition, controlling and controlled. We use PE for the weight of cooperation and competition relations and shareholding for the weight of controlling and controlled relations. The details of our knowledge graph are shown in Table 2.

Table 2. Statistics of Knowledge Graph

Relationship	#Rel	#Ent	Weight Calculation	Min	Max	Mean	Std
Cooperation	897027	2901	$\mathrm{LOG}(PE_h) * \mathrm{LOG}(PE_t)$	-41.24	66.04	8.67	10.51
Competition	38483	2901	$\mathrm{LOG}(PE_h) * \mathrm{LOG}(PE_t)$	-46.21	69.23	10.11	13.11
Controlling	2301	2901	$Shareholding_h$	0.02	84.64	18.47	21.70
Controlled	2301	2901	$Shareholding_t$	0.02	84.64	18.47	21.70

Where subscript h refers to the head of relationship link and t refers to the tail and function $LOG(x)$ is

$$LOG(x) = \begin{cases} -log(1-x), & x < 0 \\ log(x+1), & x \geq 0 \end{cases} \tag{8}$$

Performance Evaluation

We use three evaluations to measure our model predictive performance, namely accuracy (Acc), F-measure (FM) and Matthews Correlation Coefficient (MCC) (Ding et al. 2015; Mehra et al. 2014). Acc measures the overall accuracy rate of all the classes. However, it can be sensitive to unbalanced class distribution so we adopt another more robust metrics FM and MCC to reduce the effect of data skew. FM takes both precision and recalls into account while MCC take all confusion matrix into account and strikes lower deviation from data (Matthews 1975).

Given a confusion matrix:

$$\begin{bmatrix} True\ Positive\ (TP) & False\ Positive\ (FP) \\ False\ Negative\ (FN) & True\ Negative\ (TN) \end{bmatrix} \tag{9}$$

Our evaluations are calculated as follows:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{10}$$

$$FM = \frac{2TP}{2TP + FP + FN} \tag{11}$$

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{12}$$

Baselines and Proposed Model

We adopt three popular stock movement prediction methods as our baselines: Logistic Regression (LR), Support Vector Machine (SVM) and Random Forest (RF) (Ballings et al. 2015; Patel et al. 2015). In contrast to our baselines, DNN classifier is applied to predict the movements. As for the hyperparameters of our DNN, we totally use five layers, three layers of which are hidden layers, and each layer is followed by dropout methods to avoid overfitting. Nadam optimizer and mean squared error (MSE) loss are applied to our model and the learning rate is set to 0.001.

In order to focus on the major impact on focal enterprise, we set the number of influence of enterprises cooperation and competition relation to 10 and the that of controlling and controlled to 3, for the reason that in our knowledge graph, the relations of cooperation and competition are far more than controlling and controlled ones but these are hyper-parameters and flexible according to different graphs. In addition, we add the information of the past 3 days into the dataset, including a bool variable for the price movement and a float variable for price change ratio. KDJ Index is used as the technical indicator, the period of which is set to 9 days.

Results

We first apply our model to predict all the Top500 stock price movement to test the generalization and robustness of our model. For this part, we not only compare our model with those baselines, but also compare models trained with EKG information with those without it. All our experiments are performed with 10-folds cross-validations and we analyze the worst, mean and best results. The experimental results are shown in Table 3, 4 and 5.

Table 3. Acc Results of Top500 Movement Prediction

Model	Mean		Best		Worst	
	With EKG	Without EKG	With EKG	Without EKG	With EKG	Without EKG
LR	67.16%	59.05%	73.84%	66.99%	58.27%	44.37%
SVM	62.55%	61.16%	72.51%	71.19%	53.97%	52.98%
RF	73.48%	59.41%	82.11%	71.85%	68.87%	53.64%
DNN	**80.17%**	**67.17%**	**88.44%**	**75.16%**	**73.20%**	**55.62%**

Table 4. FM Results of Top500 Movement Prediction

Model	Mean		Best		Worst	
	With EKG	Without EKG	With EKG	Without EKG	With EKG	Without EKG
LR	0.55	0.40	0.66	0.59	0.44	0.23
SVM	0.41	0.41	0.71	0.70	0.30	0.29
RF	0.69	**0.51**	0.76	0.75	**0.64**	**0.44**
DNN	**0.71**	0.41	**0.79**	**0.75**	0.50	0.17

Table 5. MCC Results of Top500 Movement Prediction

Model	Mean		Best		Worst	
	With EKG	Without EKG	With EKG	Without EKG	With EKG	Without EKG
LR	0.31	0.13	0.47	0.31	0.15	-0.03
SVM	0.19	0.16	0.48	0.46	0.05	0.02
RF	0.46	0.16	0.59	0.42	**0.37**	0.05
DNN	**0.50**	**0.27**	**0.62**	**0.56**	0.35	**0.08**

Through comparison between dataset with EKG information and that without it, we can find that the EKG helps boost the performance significantly, which confirms that EKG is a useful tool for stock prediction. Given the same stock prediction model, being trained with EKG performs consistently better than being trained without it under every metric. These could be explained by the following two aspects. On one hand, the tradition technical indicators and history of the stock price cannot depict the features and dynamics of stocks accurately because their accuracies (around 60%) are just a little above pure guessing accuracy (50%). Without enough knowledge about the determinants of price, even the model knows the history, it fails to predict the future accurately. This contrast also demonstrates the importance of extracting suitable determinants instead of putting knowing facts into training blindly.

On the other hand, the EKG offers a lot of useful knowledge for stock prediction and the results also testify that relationship between firms would have an influence on each other's stock price. We are able to know the impact of other enterprises' effects on the focal firm with EKG and thus make an accurate judgment about the future movement.

Through comparison between DNN and other baselines, it can be drawn to the conclusion that DNN performs well when there exist complex relations and are able to learn underlying features automatically. This is mainly because DNN is better at analyzing high dimensions data quantitively and can extract the most representative features during the propagation between layers.

Conclusions

We demonstrated that relations between enterprises have impacts over each other's stock and by taking advantage of that impact, the performance of predicting stock price movement direction goes up significantly. We also prove that DNN is better at learning complex relation influence in EKG than other traditional machine learning models. The empirical experiments show features extracted from EKG can improve several machine learning model prediction performances, and DNN performs best among other machine learning models. Thus, We can conclude that the relevant enterprises do influence the focal enterprise's stock price and stock price movement direction can be better predicted with a combination of both relevant stock information and focal stock information.

Future work

In our future work, we plan to improve our EKG by adding more types of related enterprises, and design a dynamic EKG construction system which can update the information in EKG. Compared to the constant EKG in present paper, the dynamic EKG is expected to improve the prediction performance. More enterprises will be contained in our EKG and the EKG can be closer to the real business network. Then we will seek a stock price movement prediction's improvement by analysis of a more complicated and comprehensive EKG.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China (Grant No. 71301163, 71771212), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075, 15YJA630068), the People's Livelihood Investigation Project of Social Sciences Development Fund (201701602), Fundamental Research Funds for the Central Universities, and Research Funds of Renmin University of China (No. 15XNLQ08), Research Funds of Renmin University of China (No. 2018030009).

References

Adebiyi, A. A., Adewumi, A. O., and Ayo, C. K. 2014. "Comparison of ARIMA and artificial neural networks models for stock price prediction," *Journal of Applied Mathematics.*

Ballings, M., Van den Poel, D., Hespeels, N. and Gryp, R., 2015. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), pp.7046-7056.

Dai, C., Chen, L., Li, B., and Li, Y. 2017. "Link prediction in multi-relational networks based on relational similarity," *Information Sciences,* 394, pp.198-216.

Deng, S., Mitsubuchi, T., Shioda, K., Shimada, T., and Sakurai, A. 2011, December. "Combining technical analysis with sentiment analysis for stock price prediction," *In Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on IEEE.* pp. 800-807.

Ding, X., Zhang, Y., Liu, T., and Duan, J. 2015, July. Deep learning for event-driven stock prediction. *In Ijcai* ,pp. 2327-2333.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. 2017. “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature, 542(7639),* pp.115.

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

Gnyawali, D. R., and Ravi Madhavan. 2000. “Network structure and competitive dynamics: a structural embeddedness perspective,” Social Science Electronic Publishing.

Hu, X., Tang, X., and Tang, F. 2017, July. Analysis of Investment Relationships Between Companies and Organizations Based on Knowledge Graph. *In International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Springer,* Cham. pp. 208-218.

Hu, Z., Liu, W., Bian, J., Liu, X., and Liu, T. Y. 2017. “Listening to Chaotic Whispers: A Deep Learning Framework for News-oriented Stock Trend Prediction,” *arXiv preprint* arXiv:1712.02136.

Huynh, H. D., Dang, L. M., and Duong, D. 2017, December. “A New Model for Stock Price Movements Prediction Using Deep Neural Network,” *In Proceedings of the Eighth International Symposium on Information and Communication Technology*. *ACM*. pp. 57-62.

Kempe, David, Jon Kleinberg, and Éva Tardos. 2003. "Maximizing the spread of influence through a social network." *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM.*

Kohara, K., Ishikawa, T., Fukuhara, Y., and Nakamura, Y. 1997. “Stock price prediction using prior knowledge and neural networks,” *Intelligent systems in accounting, finance and management*, 6(1), pp.11-22.

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. 2017. “Deep learning classification of land cover and crop types using remote sensing data,” *IEEE Geoscience and Remote Sensing Letters*, 14(5), pp.778-782.

Lin, M. C., Lee, A. J., Kao, R. T., and Chen, K. T. 2011. “Stock price movement prediction using representative prototypes of financial reports,” A*CM Transactions on Management Information Systems (TMIS),* 2(3), pp.19.

Matthews, B. W. 1975. “Comparison of the predicted and observed secondary structure of T4 phage lysozyme,” *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), pp.442-451.

Mehra, A., Seidmann, A., and Mojumder, P. 2014. “Product Life - Cycle Management of Packaged Software,” *Production and Operations Management*, 23(3), pp.366-378.

Nguyen, T. H., and Grishman, R. 2015. “Relation extraction: Perspective from convolutional neural networks.” *In Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*. pp.39-48.

Nguyen, T. H., Shirai, K., and Velcin, J. 2015. “Sentiment analysis on social media for stock movement prediction,” *Expert Systems with Applications*, 42(24), pp.9603-9611.

Oh, C., and Sheng, O. 2011, December. “Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement”. *Icis*, pp. 1-19.

Park, K., and Shin, H. 2013. “Stock price prediction based on a complex interrelation network of economic factors,” Engineering Applications of Artificial Intelligence, 26(5-6), pp.1550-1561.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K., 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), pp.259-268.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. 2017. “Pointnet: Deep learning on point sets for 3d classification and segmentation,” P*roc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 1(2), pp.4.

Ruan, T., Xue, L., Wang, H., Hu, F., Zhao, L., and Ding, J. 2016, October. Building and Exploring an Enterprise Knowledge Graph for Investment Analysis. *In International Semantic Web Conference. Springer, Cham.* pp. 418-436.

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

Wang, L., Zhang, J., Liu, P., Choo, K. K. R., and Huang, F. 2017. "Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Computing*, 21(1), pp.213-221.

Yang, Y., Chawla, N., Sun, Y., and Hani, J. 2012, December. Predicting links in multi-relational and heterogeneous networks. *In Data Mining (ICDM), 2012 IEEE 12th International Conference on IEEE*. pp. 755-764.

Zarowin, 1990. What determines earnings-price ratios: revisited. *Journal of Accounting Auditing & Finance.*

Zhai, Y., Hsu, A., and Halgamuge, S. K. 2007, June. "Combining news and technical indicators in daily stock price trends prediction," In *International symposium on neural networks. Springer, Berlin, Heidelberg.* pp.1087-1096.

Zhang, W., Li, C., Ye, Y., Li, W., and Ngai, E. W. 2015. "Dynamic business network analysis for correlated stock price movement prediction," *IEEE Intelligent Systems*, 30(2), pp.26-33.