Stock Trends Prediction Using the Feature Pattern Constructed with the Panel Data of Asian Stock Markets

Tsung-Nan Chou [1+]

[1] Department of Finance, Chaoyang University of Technology

Abstract. The artificial intelligence has been widely used in many different industries for many years. In particular, the deep learning approaches have profoundly improved the traditional banking services and generated a great impact on the financial industry. By using the artificial intelligence technology, the financial advisors will be able to explore various data sources, and the meaningful information can be discovered and analyzed more efficiently. In this study, both the conventional machine learning and current deep learning methods were applied to predict the trend direction of the weighted stock price index of Taiwan based on the panel data from associated Asian stock markets. A novel feature pattern constructed with the spatial and temporal correlation of the stock markets was used as the input to train neural network. The arrangement of two dimensional pattern for input data might facilitate the convolutional neural network to explore the embedded historical trend for stock index. Three different experiments were conducted to evaluate the prediction performance of models. The experimental results showed that the convolutional neural network built with 6 and 12 distinct kernels for convolution layers outperformed the other models, and achieved the highest prediction accuracy. Although the dimensionality of original input variables was increased, the result indicated the transformation of the input variables into a two dimensional feature pattern might be helpful for the deep learning approaches to learn the historical trend pattern.

Keywords: feature pattern, machine learning, convolutional neural network

1. Introduction

Since the world is currently in the era of big data, the amount of data volumes is expecting to grow rapidly and reach 44 zettabytes by the year 2020 according to the speculation from International Data Corporation (IDC). The significant increase of data volumes generated by the financial institutions augments the complexity of data processing and analytics, and frustrates financial advisors and managers to analyse useful information and provide diversified solutions for further decision making problems. Data mining techniques and machine learning algorithms can assist data managers to improve their problem-solving techniques by discovering meaningful patterns and trends in large amounts of data. As a result, data mining is not limited to any specific industry to use; almost every industry including the medical, biotechnology, manufacturing, finance and other industries are widely applying this technology to explore massive data and provide sophisticated analysis results for decision makers. However, the financial industry is witness to be the higher uncertainty and disruption among all industries since the global economy suffered banking catastrophe during the 2008 financial crisis [1]. Big data analysis and data mining techniques have been successfully applied to the fraud detection in finance for many years, and prevented from being lost billions dollars to the action of frauds if compared to the traditional detection methods which are time consuming and laborious. In addition, data mining can help investors to cultivate their talents to spot the invest opportunity and perform faster and more efficient decision-making in judging whether certain stocks or bonds are suitable for their intelligent investments.

[+] Corresponding author. Tel.: +886-4-23323000; fax: +886-4-23742333.
E-mail address: tnchou@cyut.edu.tw.

As most investors acquired their data from various public information sources, data exploration technology has become an important issue for investors to decide whether the data is valuable, and require the correlation and relevance analysis to bring important aspects of that data into focus. Both data mining and data analysis can help users to extract the necessary information than the human analysis and improve their investment performance. The objective of this study is to develop a novel structure of feature pattern based on the spatial and temporal correlation of the stock market data. The proposed feature pattern with two-dimensional formation will be used as a premise for the deep learning neural network to learn the historical trend of stock market index and predict their daily change. Three conventional machine learning algorithms will be evaluated and compared with that of deep learning approaches. Totally three different experiments were constructed to evaluate the prediction performance of models.

2. Research Methods and Experiment Design

In this paper, the integrated framework incorporates the special design of feature pattern to represent the mutual relationship between the Asian regional stock markets, and totally seven different classification models were created to predict the daily change of Taiwan weighted stock price index. The structure of the integrated framework is illustrated in figure 1. The implementation of the models includes two stages of processing. The first stage applies all of the 22 input variables collected from Asian stock markets and expend them to construct a two-dimensional feature pattern. The second stage requires to organize seven classification models into three different comparison experiments. The first experiment uses the original 22 variables as input data for three conventional machine learning models. The other two experiments use the derived features including 784 variables for training process with the modified neural network and convolutional neural network to predict the daily change of stock index. The major purpose of second stage is to examine the effectiveness of the two-dimensional feature pattern in modelling with the deep learning algorithms.
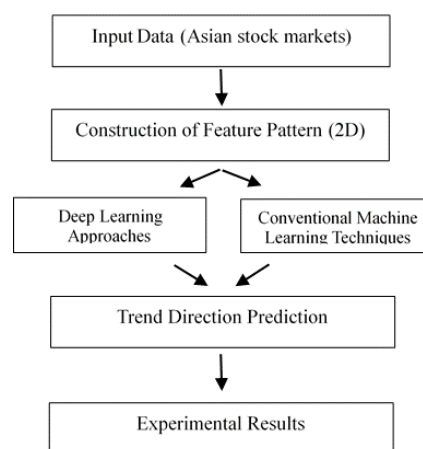

Fig. 1: Research framework.

2.1. Original features and machine learning techniques

To provide sufficient data for the evaluation of models, the input data was collected over the period of 2008 to 2015, and comprised the daily changes of 22 Asian stock markets. A total of 1750 cases were reviewed and subdivided into 1500 for the training dataset and another 250 for the testing dataset. The aggregation of these data is referred to as Original Features in this study. To deal with the data of different units and scale, both the training and testing datasets were normalized to a common scale prior to the classification. For the purpose of predicting the trend direction of stock market, the daily change of Taiwan stock market index was mapped to ten intervals prior the training process. Unlike the intraday change, the daily change calculates the difference between a given day's close of a stock index and the close of its previous trading day. Since the gain or loss of a trading day was divided into ten intervals, some intervals might span over both positive and negative values. Therefore, these intervals were regarded as highly fluctuation and might be difficult for the models to predict the trend direction of the market index.

Three conventional machine learning techniques [2] including Naive Bayes, Decision Tree, and Neural Network classifiers were employed to predict the daily change based on the Original Features. Naive Bayes method is a supervised learning algorithms based on Bayes theorem, and has worked very well in many research applications as it only requires a small amount of training data to estimate the required parameters. On the other hand, the C4.5 decision tree, improved from the original ID3 algorithm, can be used for handling training data with missing values and performs classification tasks with improved computational efficiency. Meanwhile, the Neural Network is incorporated as the third model, because its performance will be later compared with the deep learning algorithm which is the latest development in neural network. The hidden layer of neural network is defined as the average of (Number of independent variables + Number of dependent variables outcomes) in the first experiment. For speeding up the convergence of the training process, the learning rate and momentum are set to 0.5 and 0.2 respectively. The input variables are also normalized to improve the performance of network.

As high-dimensional data can increase the analytical difficulty for any models, many researches [3] use effective feature extraction and feature selection in order to reduces and eliminate the irrelevant variables for data analysis. For example, Principal Component Analysis (PCA) is a statistical method commonly used to evaluates a set of raw features and reduce the feature dimension with little losing accuracy. Although dimensionality reduction can reduce the complexity of the forecasting model, however, it might decrease the prediction accuracy by losing data integrity. On the other hand, employ feature engineering to explore and derive a new set of variables will increase the complexity of computation for analytical models. Fortunately, the fast development of AI hardware and deep learning techniques will benefit user to explore the high-dimensional data. The experimental data used in this study will be increased to 784 variables, and examine how the deep learning neural network deals with such a high-dimensional data analysis with more accuracy and efficiency.

2.2. Derived features and deep learning approaches

The original 22 variables were combined with their corresponding time series data based on specific length of time, and both the spatial and temporal data were arranged in a two-dimensional formation, which represented as a feature pattern for further training process of the deep learning neural network. It can be useful to create the two-dimensional feature pattern in our model, as the pattern can be easily scaled to different time frequency such as weekly, monthly, quarterly, and annual feature patterns for learning the historical trends of stock index. The monthly feature pattern is illustrated in the figure 2 and is referred as Derived Features in this study. Compare to the first group of models described in previous section which uses the original 22 variables, another two groups of models will use 784 variables based on the customized Derived Features. Unlike the basic neural network model used in the first group, the second group of models incorporated the same commonly used back propagation neural network, but the neurons of hidden layer were increased to 100 neurons for the comparison with the deep learning approaches later. The models in this group were slightly modified to apply with recently developed Dropout and Softmax functions mostly used in deep learning neural network [4].

The third group of models built with two different structures of the convolutional neural network [5-6], which developed from the traditional neural networks but incorporated more significant modifications on network structures and functions to learn abstract representations of the input data. The architecture of the convolutional neural network applied in this study includes six different layers. The feature pattern composed of 28x28 input variables were used in the first layer and worked as inputs to the network. The second layer was the first hidden layer of the network, and performed basic convolution operations over the input feature pattern. A 5x5 kernel was applied to filter the input pattern, and totally 6 distinct kernels were employed as filter mask in this convolutional layer to exploit the abstract features based on spatial and temporal correlation. The third layer was a max pooling layer, where a small 2x2 pooling region acted as a filter was applied to filter the output of the previous layer. As a result, the max value of that region was taken as the output of this layer and reduced the dimensionality of input pattern by eliminating non-maximal values. The forth layer was another convolution layer applying the same 5x5 kernel to filter the output of the third layer, instead of using another 12 distinct kernels. The fifth layer was a max pooling layer again which

identical to the previous one in third layer. Finally, the last layer was fully connected and correspond to a traditional neural network. The softmax function employed in this layer classified the outcomes into ten different intervals, and an associated dynamic threshold mechanism was designed to transform and interpret the intervals to the final three categories of trend direction for stock index. To explore the prediction performance of various convolutional neural networks, another network structure composed with 10 and 20 distinct kernels for each convolutional layer was also evaluated in this study.
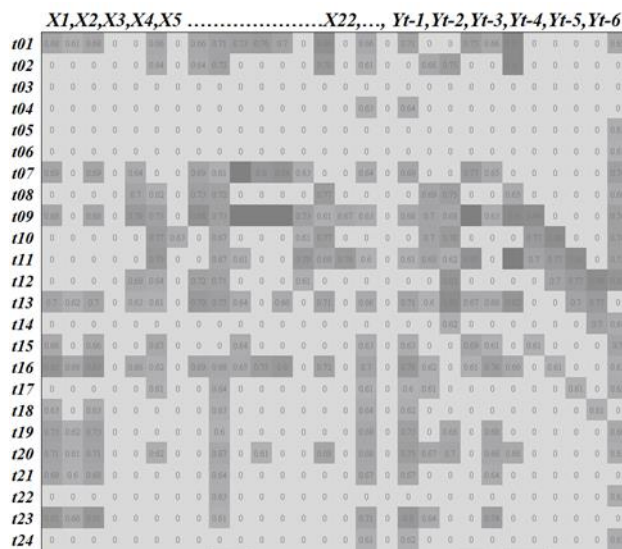

Fig. 2: Two-dimensional feature pattern.

3. Experimental Results

In this section, three different experiments were presented for all comparing models. The first experiment used the original 22 variables as inputs, and three conventional machine learning algorithms were evaluated for the prediction performance. As shown in Table 1, among all three models, the neural network model performed better than other models. However, the predictive accuracy for all models was less than 0.5, which were considered worse than the below-chance accuracy.

Table 1: Result for the first experiment.

Original Features	Accuracy	Sensitivity	Specificity	Precision	MCC	Kappa
Naive Bayes	0.448	0.367	0.682	0.363	0.063	0.195
Decision Tree	0.432	0.362	0.676	0.361	0.044	0.218
Neural Network	0.460	0.337	0.671	0.305	0.017	0.177

In the second experiment, the input features were reconstructed as a two-dimensional pattern, and the number of variables increased to 784 to account for the spatial and temporal relationship of variables. By referring to Table 2, the Sensitivity of NNW-Softmax model outperformed the other model. As a highlight, both models achieved higher Specificity, but performed lower Sensitivity. Normally, achieving an acceptable balance between Specificity and Sensitivity might not be applicable. If the losses for each trading day is a major issue concerning the short selling strategies to the investors, then Sensitivity will be a very important measurement. Based on the accuracy metric, the NNW-Dropout model performed better and reached the higher Accuracy at 0.640. Despite the fact that other metrics such as Sensitivity, Specificity and Precision were slightly lower than the NNW-Softmax model. Although the Accuracy was vulnerable to the predicted classes which were not equally represented in data set, Accuracy could still be used to assess the overall effectiveness of models. Basically, the Sensitivity and Precision metrics focused on the true positive cases rather than handling the true negative cases. Therefore, the Matthew's Correlation Coefficient (MCC) and Kappa statistic were also required to provide further comparison for models. A larger value for these metrics indicated a higher classification quality. In general, the MCC was regarded as a measure of the quality for the imbalanced data. The MCC was a correlation coefficient and its value ranged between -1 and $+1$. A coefficient of $+1$ represented a perfect prediction, while -1 indicated total disagreement between prediction

and observation. Particularly, a coefficient of zero noticed the result was similar to the random prediction. If a model simply predicts the more prevalent class and achieves a higher predictive accuracy, then the Kappa value can be used as an alternative measure for the validity of a model while its precision or sensitivity is high. The Kappa test of +1 means that the model is perfectly reliable. As shown in Table 2, the MCC Measure of the NNW-Softmax model was 0.330, and its Kappa test was 0.172, which was 1.8% slightly less than the NNW-Dropout model.

Table 2: Result for the second experiment.

Derived Features	Accuracy	Sensitivity	Specificity	Precision	MCC	Kappa
NNW-Dropout	0.640	0.494	0.801	0.494	0.301	0.190
NNW-Softmax	0.632	0.512	0.813	0.527	0.330	0.172

In Table 3, two convolutional neural networks based on different network structures were applied to extract features for the third experiment. The Specificity of CNN-B model was lower with 0.779, if compared to the CNN-A model. Although the Sensitivity of the CNN-A model achieved a higher value of 0.524, the CNN-B model performed better in Precision measure which was 0.574. Both models obtained the Kappa of 0.217 and 0.163 respectively, but the CNN-A model reached the higher MCC statistic at 0.331, which was 2% better than the other model. The CNN-A model achieved an accuracy of 0.652, and was considered as the most competent model among all models within all three experiments.

Table 3: Result for the third experiment.

Derived Features	Accuracy	Sensitivity	Specificity	Precision	MCC	Kappa
CNN-A	0.652	0.524	0.804	0.527	0.331	0.217
CNN-B	0.628	0.498	0.779	0.574	0.310	0.163

4. Conclusion

In this study, a novel feature pattern was constructed to represent the spatial and temporal correlation of the stock market data. Although this arrangement increased the number of variables from 22 to 784, the resulting two-dimensional pattern could enable the convolutional neural network to explore the embedded historical trend for any specified time interval. Totally seven classification models organized into three comparison groups were applied in experiments to evaluate the prediction performance. According to the above observation, the experimental results showed that the convolutional neural network built with 6 and 12 distinct kernels for each of two convolution layers outperformed the other models. The overall Accuracy of this CNN-A model was significantly improved to 0.652 with a Sensitivity of 0.524 and a Specificity of 0.804, respectively. Most of evaluation metrics, excluding the Specificity, indicated that CNN-A model achieved the best performance and could be used to predict the stock market trend.

Normally the traditional neural network encountered great difficulties in learning from high-dimensional input data. However, the result of the second experiment suggested the performance might be improved by partially using the functions developed from deep learning algorithms. As the dimensionality of original input data was significantly expanded to higher dimension, the experimental results also showed the two dimensional transformation of the input data might create a meaningful pattern for the deep learning approaches to learn the abstract features of the historical trend direction.

5. References

[1] Erkens, David H., Mingyi Hung, and Pedro Matos., "Corporate governance in the 2007–2008 financial crisis: Evidence from financial institutions worldwide," Journal of Corporate Finance 18(2), 389-411 (2012).

[2] Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J., [Data Mining: Practical machine learning tools and techniques], Morgan Kaufmann, (2016).

[3] Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A., "Feature extraction: foundations and applications," (207). Springer, (2008).

[4] Schmidhuber, Jürgen., "Deep learning in neural networks: An overview." Neural networks 61, 85-117 (2015).

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E., "Imagenet classification with deep convolutional neural networks," In Advances in neural information processing systems, 1097-1105 (2012).

[6] Vedaldi, Andrea, and Karel Lenc., "Matconvnet: Convolutional neural networks for matlab." Proceedings of the 23rd ACM international conference on Multimedia, ACM, (2015).