

# Rank Prediction for Portfolio Management Using Artificial Neural Networks

Jiyeon Bae, Ghudae Sim, Hyungbin Yun, Junhee Seok\*  
The school of Electrical Engineering  
Korea University  
Seoul, Korea  
{jiyeon\_b, ghudaesim, hb\_yun, jseok14}@korea.ac.kr

**Abstract**— The rank of equities is often used to determine the investment portfolio instead of prices because ranking is in general believed to be robust. In this paper, we propose a rank prediction method for portfolio management using ANN. While an ANN requires a large dataset to train the model, the sample size is usually insufficient in stock market data. Therefore, the proposed method uses data augmentation and an ensemble ANN model. In the simulation study, the proposed method shows 13 percentage of performance improvement from the other methods to predict the profit rank of equities in South-East Asian market.

**Keywords**—artificial neural network; stock market prediction; portfolio management.

## I. INTRODUCTION

Artificial neural networks (ANN) have been applied in a variety of areas including finance, medicine, image recognition and speech recognition. Especially, in finance, ANN have been used for bankruptcy prediction, credit risk assessment, stock market prediction [1,2].

For the stock market prediction, Guresen [3] evaluated the performance of ANN models for forecasting real exchange daily rate values of NASDAQ stock exchange index. Guresen compared ANN architectures, i.e. multi-layer perceptron (MLP), dynamic artificial neural network (DAN2) and the hybrid neural networks, for the prediction of daily rate values. Guresen demonstrated that MLP shows good performance than the other models; however, the difference between Mean Square Error (MSE), Mean Absolute Deviate (MAD) of each model is not significant.

Yakup Kara [4] proposed a prediction method for direction of stock price index movement using ANN and support vector machines (SVM). that the ANN model outperforms the SVM model in general.

Cui [5] employed data augmentation method to address limitation of training data size in deep neural network acoustic modeling for large vocabulary continuous speech recognition.

By using a modified version of a previously proposed approach using vocal tract length perturbation (VTLP) and a novel data augmentation approach based on stochastic feature mapping (SFM) in a speaker adaptive feature space, Cui showed that the proposed method can improve recognition performance by using both cross-entropy (CE) and state-level minimum Bayes risk (sMBR) as the cost function of ANN models.

Hellstrom [6] proposed rank measure method taking into account a large number of securities and grades them according to the relative returns. Hellstrom showed that rank measure, besides being more related to a real trading situation, can be more predictable than the individual returns. As the rank prediction is easier than numeric prediction, in field of finance, to be specific, quantitative analysis, for overcoming existing algorithm, it is expected to be a useful method to apply results of rank prediction to predict profit.

In this paper, we propose the rank prediction for portfolio management using ANN. The object of the rank prediction is to forecast the rank of equity profit (earning rates) after 80 days, which can be used for an assistant system for the portfolio optimization of an 80-day rebalancing fund.

In the proposed method, we use an ensemble approach for ANN. In the simulation study, we use 100 ensemble ANN models to compensate the limitation of single ANN model such as the initialization of weight and randomness of stochastic gradient decent algorithm. In addition, the proposed method conducts the data augmentation method due to the insufficient sample size in training data. In the stock market data, the number of sample is usually insufficient because each day corresponds to a sample. For example, collected data for ten years correspond to only 3,650 samples, which are, in general, insufficient to train ANN models.

We evaluated the performance of the proposed method by applying South-East Asian market data. We compared the performance of the proposed method and other methods by evaluating accuracy of binary classification where 50% of high-rank equities are considered as the classification target. The proposed method is beneficial for portfolio management for periodically rebalanced funds. Since the rank is an intuitive

---

This work was supported by the National Research Foundation of Korea grant (NRF-2017R1C1B2002850) and a grant from Mirae Asset Global Investments Co. Ltd. Correspondence should be addressed to [jseok14@korea.ac.kr](mailto:jseok14@korea.ac.kr)

output, a fund manager can easily understand the value and optimize the portfolio based on the predicted rank.

## II. DATA & METHOD

The proposed system is applied to South-East Asian market data. All data used this study were offered from Mirae Asset Global Investments Co. Ltd. Offered data including macro data and profit data come from January 1, 2001 till December 26, 2016 (5,841 days). We consequentially employed macro data and profit data during 2049 days for using without missing data. Macro data used as input data are composed of currency index, volatility index and international stock market index from January 16, 2009 till November 24, 2016. The profit dataset that have been used as target data in this study comes from five countries on South-East Asian market, which are Singapore, Malaysia, Philippines, Thailand, and Indonesia, from January 16, 2009 till November 24, 2016. In addition, the profit dataset includes 79 equities as variables, which belong to real estate, industrials, energy, material and information technology. Before using macro data and profit data, we performed to extract overlapped date values between macro data and profit data, which as a result have 2,049 days so the macro data have 2,049 x 114 dimension of matrix as the input data for ANN.

Since the sample size is insufficient to train ANN models, a data augmentation algorithm is used. We generated sample data between each gap of two raw samples. The value of generated sample is as follows:

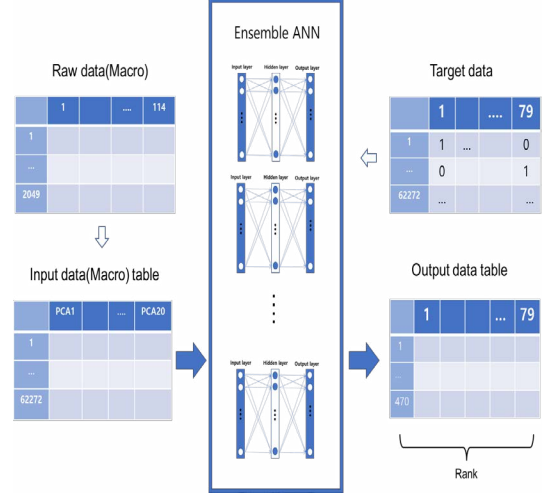
$$x_{k,t} = x_t + k \frac{(x_{t+1} - x_t)}{n} \text{ where } n = 50, k = 1, 2, \dots, 50 \quad (1)$$

In the equations,  $x_t$  is the input value at specific date  $t$ .  $x_{k,t}$  are the new samples between  $x_{t+1}$  and  $x_t$ . As a result, we can produce  $k$  samples between raw samples with this algorithm. To compress the training dataset, principal component analysis algorithm (PCA) is used. The training dataset is compressed to having 20 variables. After these data processing, the dimension of input dataset is converted to 62,272 x 20. We also applied the method to the profit data as target data in order to make number of data of each dataset be same. We processed the target data to be 0 or 1, which represent whether the profit of each equity is below than zero or not, respectively.

In the proposed rank prediction method, we use ensemble ANN models to compensate the limitation of single ANN model. We used input dataset by separating into two parts, training set and test set. The training set are composed of the data from January 16, 2009 till September 17, 2014 and test set are composed of the data from January 17 till August 5, 2016. Figure 1 shows the architecture of the proposed method. The ensemble ANN model consists of 100 different ANN models where each model uses different training days and initial weights. We trained each ensemble model constructed 1 hidden layer with 20 nodes with 20 input variables. Each ANN model is constructed with the architecture as follows:

$$\hat{Y} = \text{softmax}(W * X_{i \in S} + B) \text{ where } S \sim U(d_1, d_{2049}) \quad (2)$$

Here  $X_{i \in S}$  is macro data matrix. Sampling is performed following uniform distribution from day 1 to day 2049, which is applied each ensemble ANN model. Softmax function is applied to hidden layer as activation function, which as a result results in  $\hat{Y}$ .  $\hat{Y}$  can be interpreted as the probability.



**Figure 1.** The architecture of proposed method. One hundred ensemble ANN models are used. Macro dataset are processed by PCA algorithm to shrink the dimension. Binary target data (up, down) are used.

## III. RESULTS

In this section, the proposed method were applied to South-East Asian market data. While the raw macro data had 2,049 x 114 dimension of matrix as the input data for ANN, we conducted PCA and data augmentation algorithm in order to reduce the number of variables and increase the sample size, as we described in the previous section. After the data processing, we obtained 62,272 x 20 dimension of input dataset. And we also applied the method to the target data. With such an architecture, the output values can be interpreted as the probability of increment for each equity. Therefore, the outputs were ranked with these values, which corresponded to the predicted rank of the system.

For the validation of the results, we used Receiver Operation Characteristic (ROC) curve. ROC curve is a graphical plot which can evaluate binary classifier. The Area Under Curve (AUC) value, i.e. surface area under the ROC curve, can represents the accuracy of the model. We defined true positive like that binary classification where 50% of high-rank equities are considered as the classification target. We demonstrated that the ensemble model can compensate the limitation of single model. We tested the ensemble model with ten ANN and the ensemble model with 100 ANN. The results are shown in Figure 2. As shown in the figure, the ensemble model with 100 ANN outperform the ensemble model with ten ANN. It shows 40 percent of performance improvement of the ensemble model with 100 ANN. The performance can be enhanced when more ANN are used. Therefore, we can conclude that the ensemble model can enhance the prediction performance. To evaluate

effectiveness of the proposed system, we compared the performance of the system with that of linear regression and random forest. In this comparison, we applied equally 20 input variables obtained through PCA to ANN, linear regression and random forest model. The results are shown in Figure 3. As shown in Figure 3, the proposed system significantly shows 13 percentage of performance improvement from the other methods the other methods.

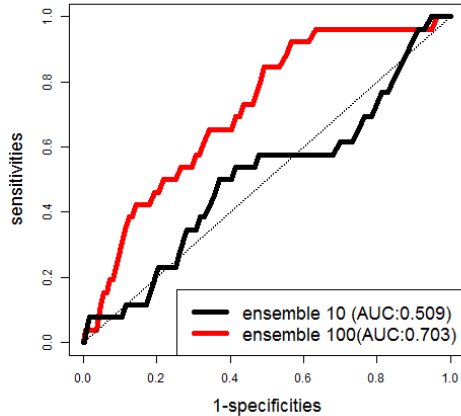


Figure 2. ROC curves with respect to the number of ANN models.

#### IV. CONCLUSION

In this work, we proposed a rank prediction system for portfolio management using ANN. We demonstrated the usefulness of the method by applying the South-East Asian market data to the proposed system. In the simulation study, the proposed system outperforms predictions with random forest and linear regression. The proposed method is expected to be a

support or advice system for the portfolio optimization of periodically rebalanced funds, by prediction the rank of equities. For the future work, we will apply not only macro data used this

study but also index data as input dataset to our rank prediction system for performance improvement and optimize ANNs structure for using a large number of input dataset and ensemble.

#### REFERENCES

- [1] Chen, An-Sing, Mark T. Leung, and Hazem Daouk. "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index." *Computers & Operations Research* 30.6 (2003): 901-923.
- [2] Yu, Lean, Shouyang Wang, and Kin Keung Lai. "Credit risk assessment with a multistage neural network ensemble learning approach." *Expert systems with applications* 34.2 (2008): 1434-1444.
- [3] Guresen, Erkam, Gulgun Kayakutlu, and Tugrul U. Daim. "Using artificial neural network models in stock market index prediction." *Expert Systems with Applications* 38.8 (2011): 10389-10397.
- [4] Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange." *Expert systems with Applications* 38.5 (2011): 5311-5319.
- [5] Cui, Xiaodong, Vaibhava Goel, and Brian Kingsbury. "Data augmentation for deep neural network acoustic modeling." *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)* 23.9 (2015): 1469-1477.
- [6] Hellstrom, Thomas. "Predicting a rank measure for stock returns." *Theory Stochastic Process* 6.22 (2000): 64-83.

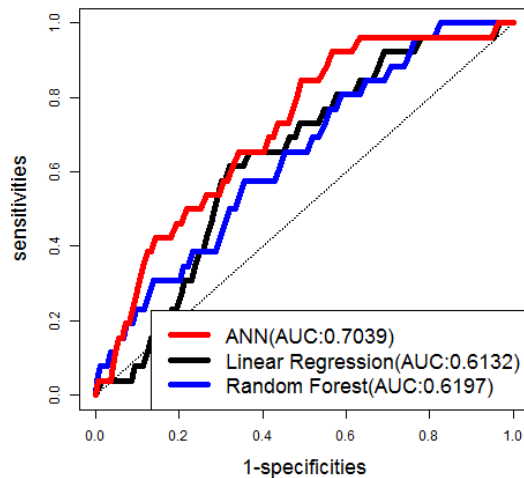


Figure 3. ROC curves of the proposed system(ANN), linear regression and random forest.