ISSN 2278-3091
Volume 9, No.3, May - June 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse50932020.pdf
https://doi.org/10.30534/ijatcse/2020/50932020

NEWFM-Based Feature Extraction for KOSPI Forecasting

Sang-Hong Lee
Dept. of Computer Science & Engineering, Anyang University, Republic of Korea, shleedosa@gmail.com

ABSTRACT

This paper proposes to make a KOSPI forecasting after one day, using minimum features selected by the distributed non-overlap area measurement method (NAMM) based on the neural network with weighted fuzzy membership functions (NEWFM). NEWFM uses the $CPC_{n,m}$ (Current Price Change of n days for n-1 to past n-m) of past 32 days of the KOSPI to forecast the rise and fall of the KOSPI in one day. Of the 38 coefficients by the wavelet transforms that were converted from $CPC_{n,m}$ and recently 32 days $CPC_{n,m}$ as feature, 5 minimum features selected by applying the NAMM. Using the proposed method, the results of using the experimental group from 1991 to 1998 showed an average forecasting rate of 67.62%.

Key words : Fuzzy Neural Networks, KOSPI, NEWFM, Wavelet Transforms.

1. INTRODUCTION

Fuzzy Neural Network (FNN), an adaptive decision support tool that combines neural network and fuzzy set theory for pattern classification, diagnosis, and forecasting, has been proposed [1][2][3][4]. FNNs of various structures were presented in addition to algorithms for learning, adaptation, and rule extraction for the extraction of knowledge from a given set of learning data [5][6][7]. Using these artificial intelligence systems, various applications for financial forecasting are being made [8]. Nonlinear time series forecasting methods, stock transactions using profit-and-loss determination systems, and forecasts of a turning point are being studied based on historical statistical indicators [8][9][10].

This paper proposes a plan to extract fuzzy rules for forecasting KOSPI based on the neural network with weighted fuzzy membership functions (NEWFM) and to minimize features by using the distributed non-overlap area measurement method (NAMM) [11][12]. NEWFM uses the bounded sum of fuzzy theory for each input to generate weighted fuzzy functions for classification. In addition, the minimum features are extracted while removing low-importance features using the NAMM. A simple fuzzy rule can be created by weighted fuzzy functions for such simplified minimum features. It can also give more effective forecasting or classification results by eliminating features that reduce the efficiency of forecasting or classification.

The experimental group and results of a genetic algorithm were used to compare the forecasting performance of the KOSPI using NEWFM [13]. A genetic algorithm was proposed to eliminate noisy data to improve KOSPI forecasting performance. A genetic algorithm showed an average one-day forecasting rate of 65.47% using the neural network after about 30% to 40% data of the entire experimental group were removed by instance selection.

Using the NEWFM-based the NAMM, this paper selects the five features with the highest importance for the classification of KOSPI rise and fall after one day, and presents the learned weighted fuzzy belonging function for these feature. The experimental groups from 1991 to 1998 were used to verify that the five selected features could be applied to various experimental groups. The results showed an average forecasting rate of 67.62%, with 72.41%, 62.07%, 62.72%, 61.02%, 65.52%, 79.31%, 67.24% and 70.69%, respectively.

2. WAVELET TRANSFORMS FOR PREPROCESSING

The Wavelet transforms complements the shortcomings of Fourier analysis, which gives information on global frequency characteristics by analyzing frequency characteristics at certain local points in time in signal processing [14]. Non-continuous wavelet transform separates the time-frequency signal into non-continuous signals on various scales. Figure 1 shows the filter bank for the implementation of non-continuous wavelet transforms. The g(n), called detail, is the high-pass filters (FIR) high-pass filters coefficient, and h(n) is the FIR low-pass filters coefficient. Repeat conversion at the next scale level with the h(n) signal whose length has been halved past each filter. The coefficient extracted by the conversion of the wavelet transform is the similarity to the mother wavelet, which represents the frequency signal over time given by the scale. In Figure 1, di and ai mean the details and approximation coefficients for each scale level i.

In this paper, $CPC_{n,m}$ (Current Price Position) is proposed as a technical indicator for forecasting after one day. $CPC_{n,m}$ is an indicator of where the closing price at the base date n is for the moving average from n-1 to n-m. In the following expression, C_n is the closing price of the base date of n days, and $MA_{n-1,n-m}$ refers to the moving average from n-1 to n-m.

$$CPP_{n,m} = ((C_n - MA_{n-1,n-m}) / MA_{n-1,n-m}) \times 100 \quad (1)$$

For the extraction of characteristic inputs using this indicator, 32 $CPC_{n,5}$ from $CPC_{n,5}$ to $CPC_{n-31,5}$, 31 days before the base date, a two-minute, non-continuous Haar wavelet transform with a scale level of 5 was performed as shown in Figure 1 to generate 38 coefficients. They consist of 16 numbers of d1, 8 numbers of d2, 4 numbers of a3 and numbers of d3, 2 numbers of each in a4 and d4, and 1 number of coefficient each in a5 and d5. A total of 39 features, including $CPC_{n,5}$ values, are initial features. In order to extract the minimum features, the NEWFM-based the distributed non-overlap area measurement method was used to extract the following five features with the highest importance among the 39 initial features.

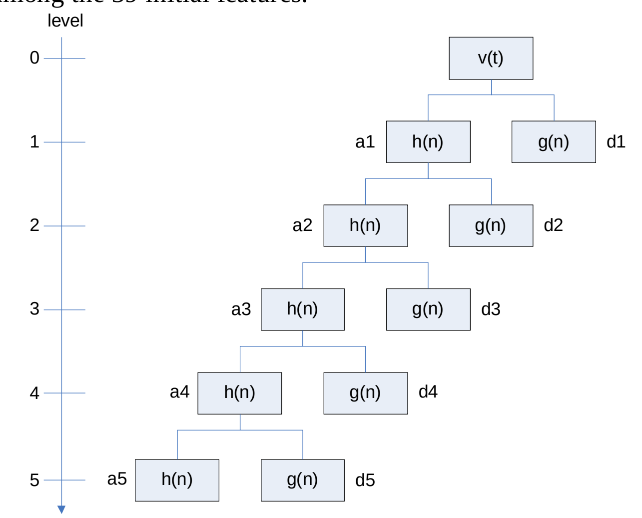

Figure 1: Decomposition of Wavelet Transforms with Scale Level 5

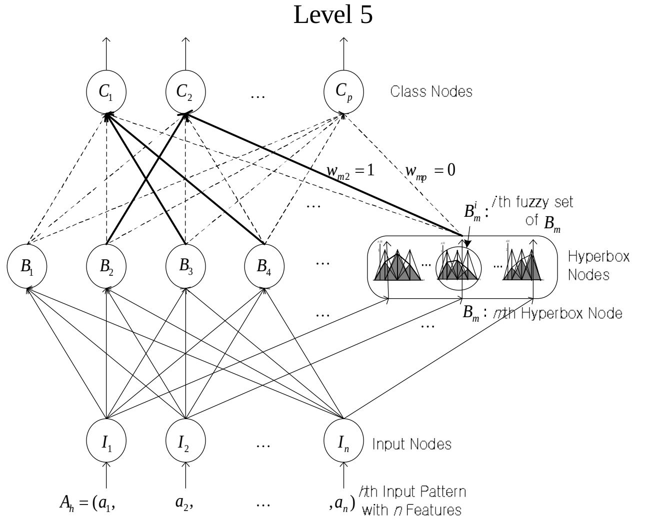

Figure 2: Structure of NEWFM

3. NEURAL NETWORK WITH WEIGHTED FUZZY MEMBERSHIP FUNCTION (NEWFM)

NEWFM is a kind of fuzzy neural networks using the bounded sum of weighted fuzzy membership functions (BSWFMs). Figure 2 explains the structure of the NEWFM that is composed of input, hyperbox, and the class layer. An hth input can be used as $I_h=\{A_h=(a_1, a_2, a_3, a_4, \ldots, a_n), class\}$, where *class* is classification node and A_h is n features of an input. The 5 coefficients extracted by the decomposition of wavelet transforms were used as inputs in Figure 2.

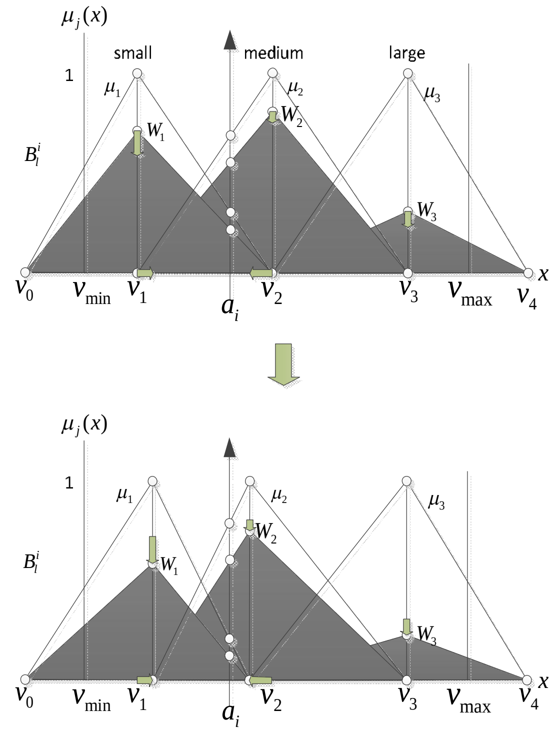

Figure 3: Example of before and after *Adjust*(B_l) operation

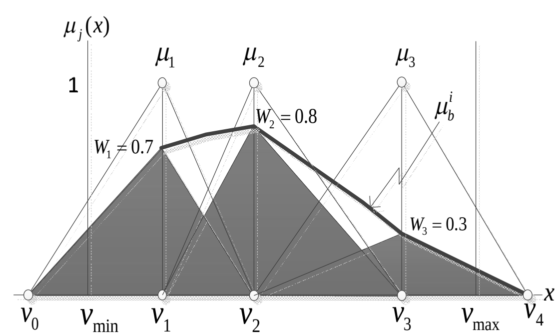

Figure 4: Example of the 3 BSWFMs

The Adjust(B_l) operation adjusted the weights and the center of membership functions in Figure 3. W_1, W_2, and W_3 are moved up or down, v_1 and v_2 are moved up to a_i, and v_3 stays in the same position. After finishing Adjust(B_l), each of all fuzzy sets in hyperbox node B_l in Figure 2 contains three *weighted fuzzy membership functions* (*WFMs*). The WFM means grey membership functions in Figure 4. The *bounded sum* of WFMs (BSWFM) in the ith fuzzy set of $B_l^i(x)$ denoted as $\mu_b^i(x)$ defined by:

$$\mu_b^i(x) = \sum_{j=1}^{3} B_l^i(\mu_j(x)). \tag{2}$$

The BSWFM means bold line in Figure 4. The two BSWFMs graphically show the difference between the rise and fall of KOSPI for each feature.

Table 1: KOSPI experimental data

Set	Year								Total
	1991	1992	1993	1994	1995	1996	1997	1998	
Training instances for GANN	234	236	237	237	235	235	234	234	1882
Selected instances for GAIS	74	71	87	66	93	86	93	85	655
Holdout instances for GANN & GAIS	58	58	59	59	58	58	58	58	477

Table 2: Used features

Used features [13]	Used features in NEWFM
CCI, RSI, Stochastic, etc (11 features)	5 wavelet coefficients

Table 3: Comparison of performance results of [13] with NEWFM (Accuracy)

Year	GANN(%)[13]	GAIS(%)[13]	NEWFM(%)
1991	53.45	72.41	72.41
1992	56.90	58.62	62.07
1993	59.32	59.32	62.72
1994	57.63	61.02	61.02
1995	65.52	67.24	65.52
1996	65.52	77.59	79.31
1997	58.62	58.62	67.24
1998	56.90	68.97	70.69
Average	59.23	65.47	67.62

4. EXPERIMENTAL RESULTS

This paper used the KOSPI experimental group from 1991 to 1998 of [13] for the evaluation of the classification performance of the KOSPI rise and fall after one day, as shown in Table 1. In Table 1, the general artistic natural network (GANN) with the general algorithm (GANN) was tested using about 80 percent of the total data from 1991 to 1998, and the general algorithm approach to instance selection for artificial neural network (GAIS) selected a gene algorithm example (GANN).

Table 2 compares the 12 features used in the experiment in [13] with those used in NEWFM. In this paper, the classification performance is obtained by using the BSWFM produced by NEWFM, and by this classification performance, the low-importance features were eliminated and minimized with five features. These five feature were used to forecast the rise and fall of the KOSPI in one day.

Table 3 shows the NEWFM classification performance comparison for GANN and GAIS. In Table 1, when tested with data for each year used in Training instance for GANN, NEWFM shows better performance than GANN for all years and means. GAIS averaged 65.47% of the total data, using only the selected data, while NEWFM showed 67.62% improvement.

5. CONCLUSION

Using NEWFM and NAMM, this paper proposes a BSWFM for forecasting KOSPI rise and fall after one day. In addition, $CPC_{n,m}$ (Current Price Change of day n) was proposed as a stock technical index for forecasting. In addition, by using the NAMM, 38 coefficients are extracted from the latest 32 days $CPC_{n,m}$ by wavelet transforms, and 5 coefficients are selected from 38 coefficients. The NAMM minimizes the number of features by removing features that are unnecessary or adversely affect the classification results. Tests showed a 67.62 percent forecast for the KOSPI's rise and fall one day later. Using these results in the future, research will be conducted on forecasting the timing of the purchase and automated trading systems.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1F1A1055423).

REFERENCES

1. Tao Zhao, Jiahao Liu, Songyi Dian, Rui Guo, Shengchuan Li, "**Sliding-Mode-Control-Theory-Based Adaptive General Type-2 Fuzzy Neural Network Control for**

Power-line Inspection Robots," Neurocomputing 401, pp. 281-294, 2020. https://doi.org/10.1016/j.neucom.2020.03.050

2. Haiyue Wang, Chao Luo, Xingyuan Wang, "**Synchronization and identification of nonlinear systems by using a novel self-evolving interval type-2 fuzzy LSTM-neural network**," Engineering Applications of Artificial Intelligence 81, pp. 79-93, 2020. https://doi.org/10.1016/j.engappai.2019.02.002
3. Chao Luo, Chenhao Tan, Xingyuan Wang, Yuanjie Zheng, "**An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction**," Applied Soft Computing 78, pp. 150-163, 2019. https://doi.org/10.1016/j.asoc.2019.02.032
4. Aaron Don M. Africa, Francis Xavier Asuncion, Janos Lance Tiberio and Raymund Miguel Francisco A. Munchua, "**Sensor-based Traffic Control Network with Neural Network Based Control System**," International Journal of Advanced Trends in Computer Science and Engineering 8, pp. 983-989, 2019. https://doi.org/10.30534/ijatcse/2019/01842019
5. Gennady G. Kalach and Gennady P. Kalach, "**Navigation System Based on the Fuzzy Logic Expert System**," International Journal of Advanced Trends in Computer Science and Engineering 8, pp. 2693-2698, 2019. https://doi.org/10.30534/ijatcse/2019/02862019
6. Jer-Guang Hsieh, Jyh-Horng Jeng, Yih-Lon Lin, Ying-Sheng Kuo, "**Single index fuzzy neural networks using locally weighted polynomial regression**," Fuzzy Sets and Systems 368, pp.82-100, 2019.
7. Yue Hou, Long Zhao, Huaiwei Lu, "**Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution**," Future Generation Computer Systems 81, pp. 425-432, 2018. https://doi.org/10.1016/j.future.2017.08.041
8. Ahmad Bagheri, Hamed Mohammadi Peyhani, Mohsen Akbari, "**Financial forecasting using ANFIS networks with Quantum-behaved Particle Swarm Optimization**," Expert Systems with Applications 41, pp. 6235-6250, 2014.
9. Shian-Chang Huang, "**Integrating spectral clustering with wavelet based kernel partial least square regressions for financial modeling and forecasting**," Applied Mathematics and Computation 217, pp. 6755-6764, 2011. https://doi.org/10.1016/j.amc.2011.01.096
10. Shanoli Samui Pal, Samarjit Kar, "**Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory**," Mathematics and Computers in Simulation 162, pp. 18-30, 2019.
11. Sang-Hong Lee, "**Feature selection based on the center of gravity of BSWFMs using NEWFM**," Engineering Applications of Artificial Intelligence 45, pp. 482-487, 2015. https://doi.org/10.1016/j.engappai.2015.08.003
12. Sang-Hong Lee, Joon S. Lim, "**Parkinson's disease classification using gait characteristics and wavelet-based feature extraction**," Expert Systems with Applications 39, pp. 7338-7344, 2012. https://doi.org/10.1016/j.eswa.2012.01.084
13. Kyoung-jae Kim, "**Artificial neural networks with evolutionary instance selection for financial forecasting**," Expert System with Applications 30, pp.519-526, 2006. https://doi.org/10.1016/j.eswa.2005.10.007
14. Anand Parey, Amandeep Singh, "**Gearbox fault diagnosis using acoustic signals, continuous wavelet transform and adaptive neuro-fuzzy inference system**," Applied Acoustics 147, pp. 133-140, 2019. https://doi.org/10.1016/j.apacoust.2018.10.013