Acta Technica 62 (2017), No. 4A, 247–254

Research on multi-input hamacher-anfis ensemble model applied in stock price forecast

Zhang Fengyi[2], Liao Zhigao[3]

Abstract. Stock market is a complex, evolutionary and nonlinear dynamic system. Forecasting stock prices has been regarded as one of the most challenging applications of modern time series forecasting. This paper proposed multi-input Hamacher-ANFIS (Adaptive-Network-Based Fuzzy Inference System based on Hamacher operator) ensemble model to forecast stock prices in China's stock market, and achieved good prediction performance. We selected five stocks with the largest total market capitalization respectively from Shanghai and Shenzhen Stock Exchange respectively, measured their historical volatility in the same time period, and the weights of performance of each stock forecasting model based on above volatility; Then, the experiment was totally repeated for 100 times for each data set, and we calculated the composite value of R^2 of the testing set according to the weight that we got before. The experimental results validate that: (1) In the aspect of prediction performance of the stock price, multi-input Hamacher-ANFIS model is superior to other models; (2) Compared with the non-ensemble forecasting strategy, the ensemble strategy of Hamacher-ANFIS model has significant advantages.

Key words. ANFIS, multi-input hamacher operator, stock price forecasting, ensemble learning.

1. Introduction

Stock market is a complex, developing and non-linear dynamic system [1], and stock price is one of the most important indicators of macro economy as well as a barometer. A large number of studies have shown that the high noise, non-linear, unstructured, complex influential factors and the large number of human factors

[1]Acknowledgement - The authors would like to thank the editors and the anonymous reviewers for their insightful and constructive comments and suggestions that have led to this improved version of the paper. The work was supported in part by Guangxi Graduate Education Innovation Project (YCSZ2014203).

[2]Workshop 1 - Master, Guangxi University of Science and Technology, Liuzhou, Guangxi, China; e-mail: `zhangfengyi90@163.com`

[3]Workshop 2 - Prof., Guangxi University of Science and Technology, Liuzhou, Guangxi, China; e-mail: `liaozhigao@126.com`

involved make it difficult to predict the change of stock price [2]. However, as a dynamic system with randomness and non-linearity, the price change of the stock market presents the characteristics of coexistence of chaos and regularity. Stock price data is time series data, which is a set of dynamic data sequences that are interchangeable over time and have complex changes. A large number of analysis found that Chinese stock market volatility has long-term memory, which means there is price change law [3] [4], so that it is feasible to predict future stock prices using historical price information. The forecast for stock prices has been seen as one of the most challenging applications of time series prediction. Accurate price forecasts can provide investors with a large and effective trading strategy. For a long time, investor have been hoping to find more effective stock price forecasting methods to reduce financial risk. Various prediction methods have been proposed and implemented.

To be a commonly used extension of ANN, Adaptive Neural Network Fuzzy Inference System (ANFIS) was proposed by Takagi and Sugeno (1985) [5], and has been originally used in industrial controlling. It is an excellent fusion of ANN and Fuzzy Inference System in information processing and control. It is a new fuzzy inference system structure which combines neural network and fuzzy reasoning. Zhou and Xing (2002) proposed the ANFIS integration strategy of the prefilter to be used in the forecast of stock price sequence, and achieved good performance [6]. However, such improved model hasn't yet applied in stock price forecast.

This study proposes ensemble Hamacher-ANFIS model, to predict the stock price of ten stocks that has the largest present value in the Chinese stock market, and the historical volatility of each stock is used as the weight of the new model to predict the performance. The actual data validation of Shanghai and Shenzhen based on volatility weighted and ten cross test shows that: (1) multi-input Hamacher-ANFIS model is superior to other conventional models in forecasting performance of stock price. (2) Compared with the non-ensemble forecasting strategy, ensemble forecasting strategy is significantly dominant. Based on our work, further study of stock price forecast can be done on

2. Proposed Model

Hamacher T-norm as a kind of T-norms with parameter, satisfies boundary conditions, commutativity, associativity and monotonicity. The parameter of Hamacher T-norm is also monotonous, and its expression is given below:

$$T_\lambda(x,y) = \frac{xy}{\lambda + (1-\lambda)(x+y-xy)} \tag{1}$$

where $\lambda > 0$. Especially, when $\lambda = 0$, Hamacher T-norm equals to algebraic product T-norm.

It is easy to recognise that algebraic product T-norm is an special Hamacher T-norm which has a constant parameter λ. However, employing a constant parameter λ is not always appropriate. For any rule , there must be a corresponding parameter λ suited for it. It is wise to use back-propagation algorithm to determine the corresponding λ.

The output of layer-2 $O_{2,3(i-1)+j}$ refers to the result of intersection operation between $\mu_{Ai}(x)$ and $\mu_{Bj}(y)$, which means the membership degree that x1 belongs to Ai and x2 belongs to Bj. It is common to use algebraic product T-norm "$*$" to deal with the membership degree in intersection operation, but as is well-known that algebraic product T-norm is not proper in any situation. What (**??**) shows is that, algebraic product T-norm is a special Hamacher T-norm whose parameter is constant to 1. So modifying the parameter to suit to the data pairs is a meaningful way to overcome the dilemma. It is not easy to determine the value of λ that should be served in Hamacher T-norm to handle intersection operation. Iliadis et al. have tried to use other constant λ to obtain better performance but not all always resulted in good situation [1]. It is a good solution to make ANFIS to adaptively select its own λ for each rule. If ANFIS could select λ for each rule respectively ,according to the training data pairs, it is more likely to fit to the performance curve and close to the inherent law.

The proposed model differs from regular ANFIS in two points:

1. It makes Hamacher parameter variable and adaptive by back-propagation algorithm, and needs to calculate the gradient with respect to each parameter and input.

2. It's combined with subtract clustering and employ it to determine the amount and value of each rule.

As is given above, the gradient of Hamacher T-norm's parameter and inputs have been achieved by Proposition 3 and Proposition 4. Different from the regular ANFIS, the output of layer-2 for a proposed model which has 3 inputs is given below:

$$O_{2,3^2(i-1)+3(j-1)+k} = T_\lambda(A_3) \tag{2}$$

where. $A_3 = \{\mu_{A_i}(x), \mu_{B_j}(y), \mu_{C_k}(z)\}$

Each rule gets the same place in regular ANFIS, because their λ have been uniformly set to 1, which hammers the system to find the most significant rule adaptively. However, proposed model's each rule with different λ in the end can lay the foundation for measuring the importance of itself. Both the IF part and the THEN part correlate to the λ and the principle of updating λ is to minimize the error, which guarantees that updated λ is harmonious to the system. w_i is the weight of i^{th} rule and decreasing with respect to λ according to Equation(2). It means that the less λ leads to bigger w_i, then bigger w_i, so the i^{th} rule plays a more important role in proposed model.

In addition, this model involves the field that the others have never touched upon. This field is attached to the improvement in fuzzy reasoning, and it could be combined with the improvement both in fuzzy reasoning and in other process, because it provides a new methodology for handling intersection operation.

With the variable and adaptive parameter, the prediction ability of proposed model may be improved; the parameter is modified according to the gradient and so as to fit to the inherent law. Empirical study will be given in next section, which proves that the proposed one overweighs the regular ANFIS.

3. Experiment Setup

We used historical stock information from the Shanghai Stock Exchange and Shenzhen Stock Exchange, and five stocks with the largest total present value in each exchange are respectively selected, with transaction information from January 15, 2013 to March 7, 2016. The names and the codes of the stocks are as follows: Shanghai: China National Petroleum Corporation (601857), Industrial & Commercial Bank of China (601396), Sinopec (600028), The Agricultural Bank of China (601288) and Bank of China (601988); Shenzhen: Hikvision (002415), GREE electric (000651), Ping An Bank (000001), (002027) and HEDY Holding (002027) and Vanke A (000002). Model inputs include the closing price, opening price, maximum price, minimum price, and transaction value of each share on day T - t. The daily closing price of each stock is used as the output of the model T. Among them, T is the time point at which the stock price is to be predicted, and t is the predicting time in advance.

The mean of the performance indexes of the ten independent, repeated experiments is used to measure the performance of the model. The experiment takes 60% of each stock's sample data as the training set, 20% as the validation set, and 20% as the test set. The training set is a sample set for learning. The model modifies the corresponding parameters by measuring their errors in the training set. The error of the model on the validation set is directly related to when the training process is terminated, while the performance of the test set is the final evaluation of the model performance. The training set, test set and test set are different in each independent repeated experiment. Among them, the R? index of the test set is used to evaluate the performance of the model:

$R_{j,k}^2 = 1 - \frac{\sum_{j=1}^{n}(\hat{y}_{j,k}(i)-y_j(i))^2}{\sum_{j=1}^{n}(y_j(i)-\overline{y}_j)^2}$

where k=1,2,…,10 represents the experimental serial number, $\hat{y}_{j,k}(i)$ indicates the ith predictive value of the kth independent repeated experiment of the jth sample set. $\overline{y}_j$ denotes to the average value of the test set for the jth sample set. For each stock, an average of ten independent repeated experiments is used as the R? of the stock. With the historical volatility of each stock as the weight, the R? is weighted and the final composite R index is obtained: $p_i = ln(y_i + \Delta t/y_i), i = 1, \cdots, n-1;$, Δt is the sampling interval, taking $\Delta t = 1$. The matrix R contains all the performance index of R?. The weights of each stock and the weights of each experiment are recorded by matrix W and matrix T respectively.

$$R = \begin{bmatrix} R_{1,1}^2 & \cdots & R_{1,10}^2 \\ R_{2,1}^2 & \cdots & R_{2,10}^2 \\ \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots \\ R_{10,1}^2 & \cdots & R_{10,10}^2 \end{bmatrix} \tag{3}$$

$$W = \begin{bmatrix} \frac{\sigma_1}{\sum_{j=1}^{10}\sigma_j} & \cdots & \frac{\sigma_{10}}{\sum_{j=1}^{10}\sigma_j} \end{bmatrix} \tag{4}$$

$$T = \begin{bmatrix} 0.1 & 0.1 & \cdots & 0.1 \end{bmatrix} \tag{5}$$

The comprehensive performance index of the model is:

$$CR^2 = T \times R \times W^T \tag{6}$$

Considering that different integration patterns affect the performance of the model, we tested 8 multivariate Hamacher-ANFIS predicting models named M1, M2, M3, ... M8, as shown in Table 1. The parameters are corrected using the LM (Levenberg-Marquardt) algorithm and the LM-LSE (Lease Square Error) algorithm. When integrated, the number of models defaults to 50. In order to compare the predictive performance of the models, SVM (support vector machine, including S1 and S2, respectively corresponding to the non-integrated strategy and integrated strategy) and RBFNN (radial basis function neural network, including B1 and B2, respectively corresponding to the non-integrated strategy and integrated strategy) are also used to predict for the same data.

Table 1.Definitions of each multivariate Hamacher-ANFIS model

Property	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8
Updated λ								
Algorithms	LM-LSE	LM-LSE	LM	LM	LM-LSE	LM-LSE	LM	LM
Ensemble								

4. Results

Table 2 shows 8 mean and standard deviations of the 100 independent R models of 1 day in advance. It shows that: (1) there is no significant difference between the non-integrated models; (2) there is no significant difference between the integrated models; (3) all the integrated models are better than non-integrated models.

Table 2.The compound R? and standard deviations of all the models (considering 1 day in advance)

Sample set	M1	M2	M3	M4	M5	M6
GREE	0.908(0.121)	0.973(0.013)	0.911(0.015)	0.9(0.191)	0.871(0.191)	0.960(0.015)
ICBC	0.920(0.073)	0.970(0.045)	0.912(0.125)	0.959(0.008)	0.957(0.023)	0.980(0.021)
Hikvision	0.825(0.131)	0.981(0.081)	0.843(0.034)	0.95(0.087)	0.923(0.103)	0.927(0.064)
ABC	0.964(0.014)	0.961(0.015)	0.976(0.054)	0.976(0.013)	0.975(0.015)	0.973(0.023)
Ping An Bank	0.975(0.046)	0.978(0.014)	0.946(0.015)	0.956(0.021)	0.978(0.117)	0.946(0.131)
HEDY holding	0.977(0.010)	0.996(0.046)	0.969(0.053)	0.948(0.122)	0.932(0.078)	0.963(0.101)
Vanke A	0.836(0.111)	0.983(0.001)	0.839(0.052)	0.983(0.003)	0.966(0.027)	0.982(0.003)
Sinopec	0.908(0.030)	0.930(0.000)	0.900(0.045)	0.9(0.090)	0.830(0.090)	0.930(0.005)
CNPC	0.930(0.030)	0.9300.045)	0.903(0.035)	0.959(0.008)	0.953(0.030)	0.980(0.030)
Bank of China	0.835(0.000)	0.980(0.080)	0.840(0.004)	0.95(0.083)	0.930(0.000)	0.933(0.034)

Table 3.The composite R indexes and standard deviations of all the models (considering different time in advance)

Time Interval	M1	M2	M3	M4	M5	M6
1	0.942(0.071)	0.985(0.017)	0.942(0.057)	0.986(0.017)	0.952(0.058)	0.988(0.019)
2	0.92(0.115)	0.979(0.019)	0.952(0.048)	0.981(0.016)	0.959(0.042)	0.983(0.02)
3	0.941(0.054)	0.973(0.019)	0.924(0.092)	0.968(0.02)	0.95(0.047)	0.971(0.02)
4	0.944(0.041)	0.962(0.02)	0.915(0.08)	0.964(0.019)	0.939(0.064)	0.963(0.023)
5	0.926(0.048)	0.946(0.025)	0.902(0.078)	0.952(0.023)	0.926(0.049)	0.956(0.02)
6	0.932(0.038)	0.943(0.023)	0.926(0.05)	0.947(0.022)	0.929(0.048)	0.95(0.017)
7	0.942(0.07)	0.985(0.017)	0.942(0.057)	0.986(0.017)	0.952(0.058)	0.988(0.019)
Time Interval	M7	M8	S1	S2	B1	B2
1	0.982(0.018)	0.984(0.016)	0.937(0.069)	0.979(0.118)	0.964(0.079)	0.897(0.065)
2	0.972(0.021)	0.978(0.018)	0.888(0.163)	0.922(0.168)	0.956(0.044)	0.888(0.143)
3	0.946(0.064)	0.965(0.02)	0.864(0.163)	0.876(0.199)	0.817(0.171)	0.877(0.15)
4	0.923(0.052)	0.958(0.023)	0.814(0.088)	0.789(0.038)	0.788(0.074)	0.854(0.162)
5	0.922(0.076)	0.957(0.022)	0.789(0.125)	0.757(0.126)	0.785(0.118)	0.817(0.102)
6	0.935(0.028)	0.951(0.024)	0.779(0.27)	0.754(0.106)	0.771(0.045)	0.787(0.029)
7	0.982(0.018)	0.984(0.016)	0.937(0.69)	0.979(0.118)	0.964(0.079)	0.897(0.065)

Table 3 shows the composite R? and standard deviations of all models under different predicting times in advance. It can be found that with the increase of predicting time in advance, the CR? mean decreases gradually, the standard deviation tends to be stable, and the predicting performance declines. It is difficult to make relative evaluation of these models intuitively. It also shows that: (1) there is no significant difference between the non-integrated multivariate Hamacher-ANFIS models; (2) there is no significant difference between the integrated multivariate Hamacher-ANFIS models; (3) there is no significant difference between the 4 control models; (4) all the integrated multivariate Hamacher-ANFIS models are better than non-integrated multivariate Hamacher-ANFIS models; (5) all the integrated multivariate Hamacher-ANFIS models are better than other control models. In summary, we can conclude from the results that: (1) the predicting performance of stock price of the multivariate Hamacher-ANFIS model is better than SVM and RBFNN; (2) when using multivariate Hamacher-ANFIS, the integrated predicting method is better than non-integrated predicting method.

5. Conclusions

In this study, we verified that ensemble multi-input Hamacher-ANFIS can improves the accuracy of the multi-time-in-advance stock price forecast. Taking volatility as the weight and taking the true historical price of the stock in Chinese stock market as sample, 100 independent repeated experiments are conducted on each stock data. Finally the compound R? is obtained to measure the performance of the model. The statistical test of the experimental results show that: (1) in the performance of stock price prediction, the multivariate Hamacher-ANFIS model is overwhelming other conventional; (2) compared to the single predicting strategy, ensemble predicting strategy is significantly dominant. In addition, there is a huge space for improvement in multivariate Hamacher-ANFIS. How to further improve the model to be closer to the economic forecast and management practice is the future research direction.

References

[1] Y. S. Mostafa, A. F. Atiya: *Introduction to financial forecasting.* Applied Intelligence (1996), No. 6, 205–213.

[2] J. W. Hall: *Trading on the edge: neural, genetic, and fuzzy systems for chaotic financial markets.* New York: Wiley (1994).

[3] T. Takagi, M. Sugeno: *Fuzzy identification of systems and its applications to modeling and control.* IEEE Trans.on Syst.man & Cybern *15* (1985), No. 1, 387–403.

[4] P. Zhou, G. Xing: *Research on stock price trend prediction.* Computer engineering *28* (2002), No. 1, 136–137.

[5] W. Q. Huang, X. T. Zhuang: *A network analysis of the Chinese stock market.* Physica A Statistical Mechanics & Its Applications *388* (2009), No. 14, 2956–1964.

[6] H. H. Chu, T. L. Chen, C. H. Chen, C. C. Huang: *Fuzzy dual-factor time-series for stock index forecasting.* Expert Systems with Application *36* (2009), No. 1, 165–171.

Received November 16, 2016