International Journal of Innovative Computing, Information and Control
Volume **13**, Number **2**, April **2017**
 ISSN 1349-4198
pp. 671–677

MODELING AND FORECASTING OF INTERVAL-VALUED TIME SERIES USING FUZZY MODELING AND INTERVAL INFORMATION GRANULARITY

Wenyan Song[1,2], Degang Wang[3], Hongxing Li[3] and Lili Cai[3]

[1]School of Economics
Dongbei University of Finance and Economics
[2]Center for Econometric Analysis and Forecasting, Northeastern Branch
No. 217, Jianshan Street, Shahekou District, Dalian 116025, P. R. China
agathasong@163.com

[3]School of Control Science and Engineering
Dalian University of Technology
No. 2, Linggong Road, Ganjingzi District, Dalian 116024, P. R. China
wangdg@dlut.edu.cn

Received August 2016; revised December 2016

Abstract. *In this paper, modeling and forecasting of interval-valued time series (ITS) are investigated. The principle of justifiable information granularity is utilized to transfer the numeric time series into ITS. And, an interval-valued error-compensated marginal linearization method is proposed to predict ITS. Further, in order to measure the performance of ITS, a novel criterion, which takes forecasting accuracy and interval coverage into account together, is obtained. Numerical experiments illustrate the validity of the proposed method.*
Keywords: Interval-valued time series, Interval information granule, Fuzzy modeling

1. **Introduction.** In researches of financial data analysis, time series can utilize available information effectively to predict variation tendency of some financial indices in the future. For example, in stock market there is a close relationship between the stock price index at a certain time with the historical stock prices. Hence, various time series models are established to forecast stock prices. Since the stock prices are rather changeable, in order to capture the characteristics of the volatility, people usually hope to forecast the variation range of the stock prices over a certain period of time. To solve this problem, granular time series is established. Different from point to point forecasting model, granular time series can deal with linguistic information and interval-valued data.

During the past few decades, prediction of granular time series has become a hot topic. In [1], a granular time series based on fuzzy clustering is proposed for long-term forecasting and trend forecasting of time series. In [2], fuzzy relation is applied to describing and forming granular time series model. In [3], prediction of nonlinear time series with noise data is investigated by fuzzy granular support vector machines. An optimal allocation scheme of information granular is designed so that outputs of the model can provide high level of specificity in [4]. When the state variables are taken as linguistic values, some scholars investigate the forecasting of fuzzy time series. In [5], a hybrid fuzzy granular model is proposed to forecast financial data. In [6], multivariate fuzzy time series is applied to forecasting the stock index information. To improve the forecasting accuracy of fuzzy time series, interval information granular is applied to obtaining the optimal partition of the universe of the state variable in [7]. On the other hand, to forecast the changing ranges of the state variable, some techniques are proposed to handle the forecasting problem of interval-valued time series. In [8], a hybrid scheme based on autoregressive model and neural network is obtained. The former part can describe the linear feature of data, and the latter part can forecast the residual error series and can reflect the nonlinear feature of data. Besides, in [9-11], some fuzzy reasoning methods are proposed to process the interval-valued data.

From existing results, we find out that most methods model and forecast ITS in an invariable time scale. However, forecasting problem of ITS with different time scales still needs to be resolved. For example, when the daily stock price is known, how to forecast the price range over a period is an interesting question which can reflect the developing trend of the stock market. Moreover, how to establish an appropriate criterion to measure the effectiveness of ITS is also an important task. Motivated by above facts, in this paper a novel forecasting method which combines information granular technology and interval-valued error-compensated marginal linearization (I-EMALINE) method is designed for interval-valued financial time series. Compared to some existing methods, the proposed method can handle the forecasting problem of ITS with different time scales. And, a rather comprehensive criterion which synthetically considers the forecasting accuracy and coverage areas, is obtained to evaluate the forecasting capability of the model.

This paper is organized as follows. In Section 2, some preliminaries of fuzzy modeling method and ITS are introduced. In Section 3, modeling and forecasting of ITS are investigated. Then two numerical experiments are reported in Section 4. In Section 5, some conclusions are provided.

2. **Preliminaries.** In this section, we will introduce some preliminary knowledge and basic notations which are used in the paper.

Firstly, we will introduce the mathematical representation of error-compensated marginal linearization method. Consider a group of input-output samples $(x_{1j_1}, \dots, x_{nj_n}, y_{j_1 \dots j_n})$, where $j_i = 1, \dots, p_i$, $i = 1, \dots, n$. By fuzzy C-means (FCM) clustering method, we can divide these data into several sub-regions, denoted by $(j'_1, \dots, j'_n) = \left[x_{1j'_1}, x_{1(j'_1+1)}\right) \times \cdots \times \left[x_{nj'_n}, x_{n(j'_n+1)}\right)$, where $j'_i = 1, \dots, p'_i$, $i = 1, \dots, n$, and $p'_i \le p_i$. Accordingly, we can obtain a group of fuzzy rules as follows:

$$\text{If } x_1 \text{ is } A_{1j'_1} \text{ and } x_2 \text{ is } A_{2j'_2} \text{ and } \cdots \text{ and } x_n \text{ is } A_{nj'_n} \text{ then } y \text{ is } B_{j'_1 \cdots j'_n} \tag{1}$$

$(x_1, \dots, x_n)$ is the input variable and $y \in Y$ is the output variable. $x_{ij'_i}$ is the peak point of triangular fuzzy set $A_{ij'_i}$, and $y_{j'_1 \cdots j'_n}$ is the peak point of fuzzy set $B_{j'_1 \cdots j'_n}$. From [11], a fuzzy model determined by fuzzy rules (1) and error-compensated marginal linearization method can be represented by

$$\begin{aligned} f(x_1, \dots, x_n) = \sum_{j'_1=1}^{p'_1} \cdots \sum_{j'_n=1}^{p'_n} \Bigg(\sum_{i=1}^{n} \Big(& A_{ij'_i}(x_i) \cdot y_{j'_1 \cdots j'_n} \\ & + A_{i\left(j'_i+1\right)}(x_i) \cdot y_{\left(j'_1+1\right) \cdots \left(j'_n+1\right)} - \beta_{j'_1 \cdots j'_n} \Big) \Bigg) \end{aligned} \tag{2}$$

where $\beta_{j'_1 \cdots j'_n}$ is the compensation parameters in local region $(j'_1, \dots, j'_n)$, which can be obtained by solving the following optimization problem:

$$\min_{\beta_{j'_1 \cdots j'_n}} E = \sum_{l=1}^{N} \|f(o_l) - t_l\|^2 \tag{3}$$

where (o_l, t_l) $(l = 1, \dots, N)$ are some known samples which locate in region $(j'_1, \dots, j'_n)$.

Then, we will review some basic concepts of ITS. $\{[x]_t, t = 1, 2, \ldots\}$ is called an interval-valued time series, if the state variable at each time t is taken as an interval number $[x_t^L, x_t^R]$, where $[x]_t$ is the state variable of ITS, and x_t^L and x_t^R are the lower bound and upper bound of the interval respectively. In the following, in order to differentiate ITS from traditional time series, ITS is denoted by $\{[x]_t, t = 1, 2, \ldots\}$ and the latter is denoted by $\{x_t, t = 1, 2, \ldots\}$, where the state variable x_t is taken as a numerical value instead of interval number. To evaluate the forecasting accuracy of ITS, interval mean square error ($\mathrm{MSE^I}$) is defined as

$$\mathrm{MSE^I} = \sum_{t=1}^{K} \left(\left|x_t^L - \hat{x}_t^L\right|^2 + \left|x_t^R - \hat{x}_t^R\right|^2 \right) \Big/ K \tag{4}$$

where K stands for the number of intervals, $[x_t^L, x_t^R]$ is the actual data, and $[\hat{x}_t^L, \hat{x}_t^R]$ is the forecasted value. Notice that an interval number $[x_t^L, x_t^R]$ can be determined by the midpoint m_t and radius ρ_t, where $m_t = \frac{X_t^L + X_t^R}{2}$ and $\rho_t = \frac{X_t^R - X_t^L}{2}$. Hence, ITS can also be represented by the vector time series $\left\{(m_t, \rho_t)^{\mathrm{T}}, t = 1, 2, \ldots\right\}$.

3. **Modeling and Forecasting of ITS.** In this section, we will utilize the principle of justifiable information granularity to transfer data into ITS. After that, we will design the corresponding forecasting method and evaluation criterion for ITS.

Let us consider a numeric type time series $\{x_t, t = 1, 2, \ldots\}$. Firstly, it can be divided into several sub-series by a given time dimension. For example, let x_t denote the daily stock price. If the time dimension is changed into week, then time series $\{x_t, t = 1, 2, \ldots\}$ can be divided into several subsets which include five or less trading day stock prices in each week. Then, for each subset, we can utilize the principle of justifiable information granularity to transfer these five-day stock prices into an interval which represents the variation range of stock prices in one week. From [12], the lower and upper bound of granular interval can be obtained by solving the following optimization problems:

$$\begin{aligned} x_t^L = \arg\max_a \, & (card\,\{x_k \in D \,|\, a \le x_k < med(D)\}) \\ & * (1 - |med(D) - a| / |med(D) - \min\{x_k\}|) \end{aligned} \tag{5}$$

$$\begin{aligned} x_t^R = \arg\max_b \, & (card\,\{x_k \in D \,|\, med(D) < x_k \le b\}) \\ & * (1 - |b - med(D)| / |med(D) - \max\{x_k\}|) \end{aligned} \tag{6}$$

where $D = \{x_1, \ldots, x_n\}$ is the numerical data subset and $med(D)$ is the median of D. In this paper, particle swarm optimization (PSO) method [13] is used to solve optimization problems (5) and (6). Accordingly, ITS $\{[x]_t, t = 1, 2, \ldots\}$ can be obtained.

Further, we will establish I-EMALINE method to forecast ITS. The detailed processes are shown as follows.

At first, ITS is divided into the midpoint series $\{m_t, t = 1, 2, \ldots\}$ and the radius series $\{\rho_t, t = 1, 2, \ldots\}$.

Then, we can obtain a group of fuzzy rules for them:

$$\text{If } m_{t-1} \text{ is } A_{1j_1}^c \text{ and } m_{t-2} \text{ is } A_{2j_2}^c \text{ and } \cdots \text{ and } m_{t-n} \text{ is } A_{nj_n}^c \text{ then } m_t \text{ is } A_{j_1 \cdots j_n}^c \tag{7}$$

$$\text{If } \rho_{t-1} \text{ is } A_{1j_1}^r \text{ and } \rho_{t-2} \text{ is } A_{2j_2}^r \text{ and } \ldots \text{ and } \rho_{t-n} \text{ is } A_{nj_n}^r \text{ then } \rho_t \text{ is } A_{j_1 \cdots j_n}^r \tag{8}$$

where $j_i = 1, \ldots, p_i$ and $i = 1, \ldots, n$. Based on Equation (2), fuzzy model determined by fuzzy rules (7) and (8) is represented as

$$m_t = \sum_{j_i=1}^{p_1} \cdots \sum_{j_n=1}^{p_n} \left(\sum_{i=1}^{n} \left(A_{ij_i}^c(m_{t-i}) \cdot y_{j_1 \cdots j_n}^c + A_{i(j_i+1)}^c(m_{t-i}) \cdot y_{(j_1+1)\ldots(j_n+1)}^c - \beta_{j_1 \cdots j_n}^c \right) \right) \tag{9}$$

$$\rho_t=\sum_{j_i=1}^{p_1}\cdots\sum_{j_n=1}^{p_n}\left(\sum_{i=1}^{n}\left(A^r_{ij_i}(\rho_{t-i})\cdot y^r_{j_1\ldots j_n}+A^r_{i(j_i+1)}(\rho_{t-i})\cdot y^r_{(j_1+1)\ldots(j_n+1)}-\beta^r_{j_1\ldots j_n}\right)\right) \tag{10}$$

where fuzzy sets $A^c_{ij_i}$ and $A^r_{ij_i}$ are chosen as triangular membership functions, $y^c_{j_1\ldots j_n}$ and $y^r_{j_1\ldots j_n}$ are the peak points of fuzzy sets $A^c_{j_1\ldots j_n}$ and $A^r_{j_1\ldots j_n}$. $\beta^c_{j_1\ldots j_n}$ and $\beta^r_{j_1\ldots j_n}$ are the compensation parameters in local regions $(j_1,\ldots,j_n)^c=\left[x^c_{1j_1},x^c_{1(j_1+1)}\right)\times\cdots\times\left[x^c_{nj_n},x^c_{n(j_n+1)}\right)$ and $(j_1,\ldots,j_n)^r=\left[x^r_{1j_1},x^r_{1(j_1+1)}\right)\times\cdots\times\left[x^r_{nj_n},x^r_{n(j_n+1)}\right)$, where $x^c_{ij_i}$ and $x^r_{ij_i}$ are modeling data of midpoint series and radius series respectively, which can determine the local region. Obviously, the scope of each local region and the compensation parameters $\beta^c_{j_1\ldots j_n}$ and $\beta^r_{j_1\ldots j_n}$ will influence the forecasting accuracy of models (9) and (10).

Hence, PSO method is used to optimize the parameters $x^c_{ij_i}$, $x^r_{ij_i}$, $\beta^c_{j_1\ldots j_n}$ and $\beta^r_{j_1\ldots j_n}$ by minimizing the following function:

$$\min E=\left(\sum_{t=1}^{K}\left(\left|x^L_t-\hat{x}^L_t\right|^2+\left|x^R_t-\hat{x}^R_t\right|^2\right)\Big/K\right)\cdot 1\Big/\left(\sum_{t=1}^{K}\mathrm{cov}_t\Big/K\right) \tag{11}$$

where $\left[x^L_t,x^R_t\right]$ is the actual data, $\left[\hat{x}^L_t,\hat{x}^R_t\right]$ is the forecasted value, and K is the number of ITS. In (11), the first term is used to evaluate the forecasting error between the actual data and forecasted data. Within the second term cov_t denotes the number which interval $\left[\hat{x}^L_t,\hat{x}^R_t\right]$ contains the actual numeric data. Different from (4), energy function (11) can not only evaluate the forecasting accuracy but also measure the coverage areas for the output of ITS. In this way, by information granular technology and I-EMALINE method, we can generate ITS using numerical data and establish the corresponding forecasting model.

4. **Numerical Examples.** In this section, two financial datasets[1] are provided to measure the forecasting performance of the proposed method for ITS. The dataset of each stock index price is covered from the period 5/14/2012-5/16/2015, which contains 756 trading days, i.e., 157 weeks. For each stock, the first two thirds of above samples are chosen as the training samples and the remaining samples are used for testing.

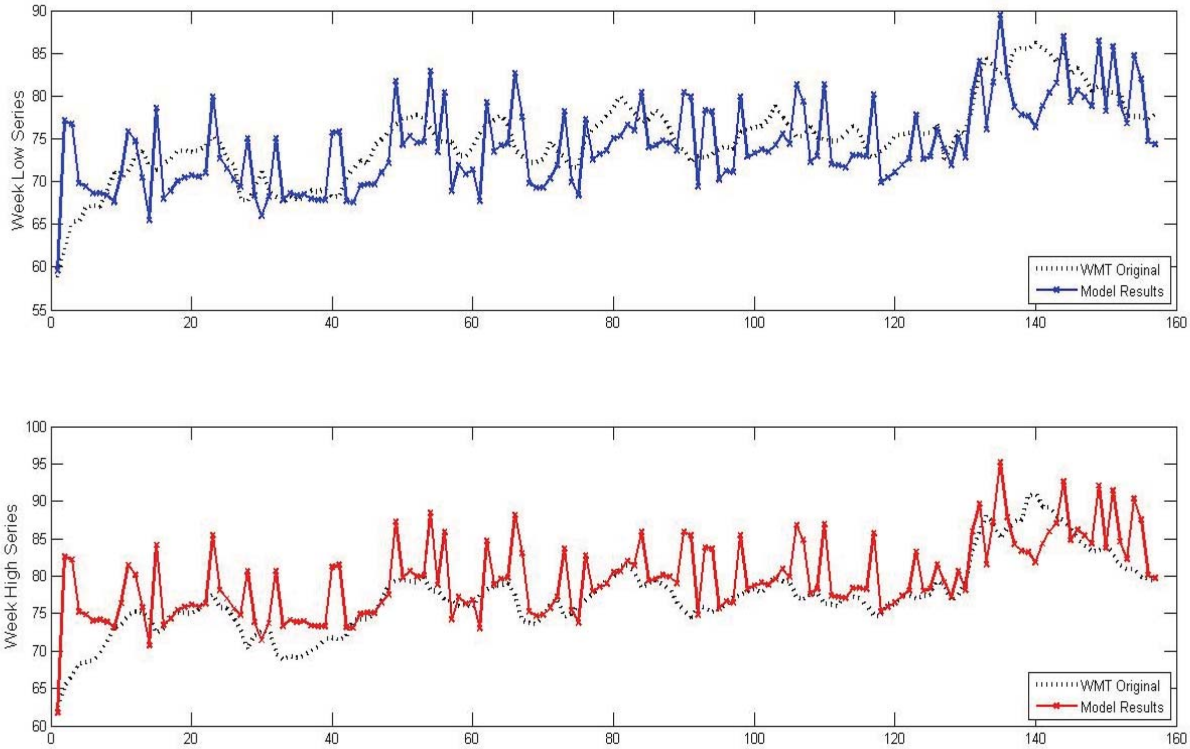
Example 4.1. *Stock price prediction of Wal-Mart Stores Inc. (WMT).*

Based on the principle of justifiable information granularity, we transfer training daily data of WMT into weekly interval series at first. Then by the proposed I-EMALINE method, predictions of the ITS are obtained. Further, we use the model to forecast ITS of the testing phase series. The comparison between modeling results with historical WMT weekly low and high stock index is considered. When the number of rules in I-EMALINE method is 3 or 5, corresponding output series are shown in Figure 1, where the black dotted lines represent the historical weekly trading series, and the thick dash lines are prediction series. The parameters of the particle swarm optimization algorithm used in the experiments are set as follows: maximum number of iterations – 50, number of particles – 10, self recognition coefficient – 1.49, and social recognition coefficient – 1.49.

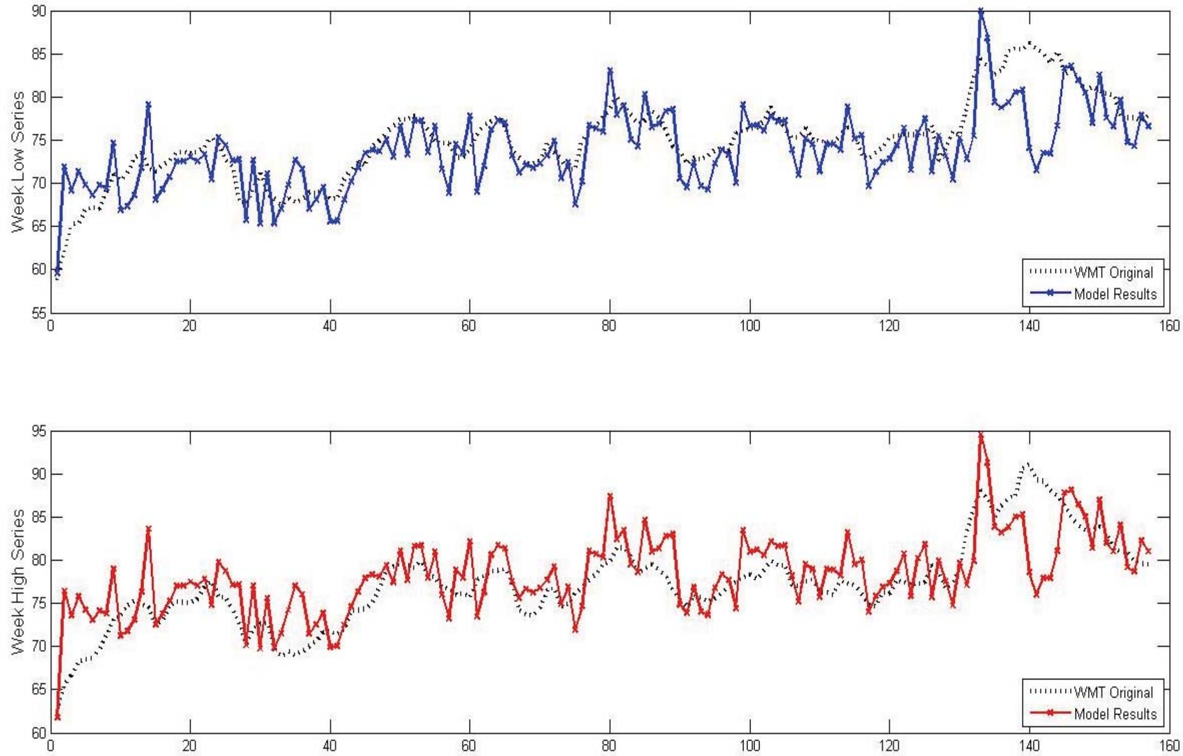
With 3 fuzzy inference rules, the $\mathrm{MSE}^{\mathrm{I}}$ of low and high trading prices between modeling results and real weekly data is 42.54. With 5 rules, the $\mathrm{MSE}^{\mathrm{I}}$ of model outputs is 35.95. It can be seen that by small numbers of rules, I-EMALINE models could have satisfactory results. Moreover, depending on granularity interval information, the tendency of weekly stock price can be effectively captured.

Example 4.2. *Stock price prediction of McDonald's Corp. (MCD).*

[1] New York Stock Exchange is available at http://finance.yahoo.com/.



(a) With 3 rules



(b) With 5 rules

FIGURE 1. Weekly stock price prediction for WMT

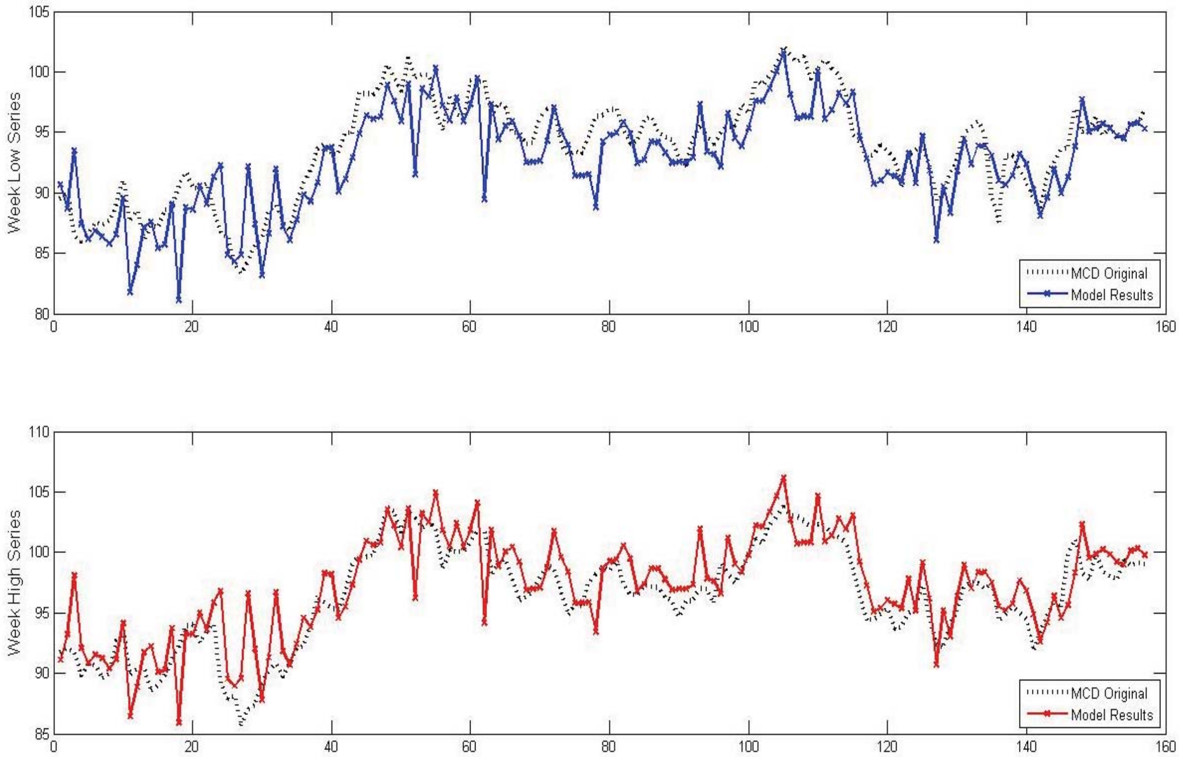


FIGURE 2. Weekly stock price prediction for MCD

When the number of fuzzy rules is supposed to be 7, with the energy function of Equation (11), we can obtain I-EMALINE model and the output interval series is shown in Figure 2. The parameters of the particle swarm optimization algorithm used in this experiment are set as follows: maximum number of iterations – 100, number of particles – 30, self recognition coefficient – 1.49, and social recognition coefficient – 1.49.

The $\mathrm{MSE}^{\mathrm{I}}$ of low and high trading prices between modeling results and real weekly data is 12.79. 79.51% daily trading data of granular information intervals can be included in the modeling prediction intervals. On the other hand, if we use the energy function as Equation (4), then the $\mathrm{MSE}^{\mathrm{I}}$ of corresponding model is 30.23; meanwhile the coverage rate of daily indexes decreases to 36.04%. So it is necessary to integrate the coverage indicator into criterion.

5. **Conclusions.** In this paper, a forecasting method which combines information granular technology and I-EMALINE method is proposed for ITS. Further, a novel evaluation criterion is obtained to measure the forecasting performance of ITS. In the future research, it could be beneficial to compare the forecasting performances of ITS under different energy functions and to improve the forecasting accuracy of ITS in different time scales.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (71571035, 61374118) and the Program for Liaoning Excellent Talents in University (WJQ2014036).

REFERENCES

[1] R. J. Dong and W. Pedrycz, A granular time series approach to long-term forecasting and trend forecasting, *Physica A*, vol.387, pp.3253-3270, 2008.

[2] R. A. AL-Hmouz, W. Pedrycz and A. Balamash, Description and prediction of time series: A general framework of Granular Computing, *Expert Systems with Applications*, vol.42, pp.4830-4839, 2015.
[3] J. Ruan, X. Wang and Y. Shi, Developing fast predictors for large-scale time series using fuzzy granular support vector machines, *Applied Soft Computing*, vol.13, pp.3981-4000, 2013.
[4] X. C. Hu, W. Pedrycz and X. M. Wang, Optimal allocation of information granularity in system modeling through the maximization of information specificity: A development of granular input space, *Applied Soft Computing*, vol.42, pp.410-422, 2016.
[5] M. Y. Chen and B. T. Chen, A hybrid fuzzy time series model based on granular computing for stock price forecasting, *Information Sciences*, vol.294, pp.227-241, 2015.
[6] B. Q. Sun, H. F. Guo, H. R. Karimi, Y. J. Ge and S. Xiong, Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series, *Neurocomputing*, vol.151, pp.1528-1536, 2015.
[7] W. Lu, X. Y. Chen, W. Pedrycz, X. D. Liu and J. H. Yang, Using interval information granules to improve forecasting in fuzzy time series, *International Journal of Approximate Reasoning*, vol.57, pp.1-18, 2015.
[8] A. L. S. Maia, F. A. T. Carvalho and T. B. Ludermir, Forecasting models for interval-valued time series, *Neurocomputing*, vol.71, pp.3344-3352, 2008.
[9] Y. Zhang, Interval-valued variable weighted synthesis inference method for fuzzy reasoning and fuzzy systems, *ICIC Express Letters, Part B: Applications*, vol.4, no.3, pp.511-518, 2013.
[10] W. Zeng and Y. Zhao, Multiple rules interval-valued approximate reasoning based on interval-valued similarity measure set, *ICIC Express Letters*, vol.7, no.4, pp.1265-1271, 2013.
[11] D. G. Wang, C. L. P. Chen, W. Y. Song and H. X. Li, Error-compensated marginal linearization method for modeling a fuzzy system, *IEEE Trans. Fuzzy Systems*, vol.23, no.1, pp.215-222, 2015.
[12] W. Pedrycz and X. M. Wang, Designing fuzzy sets with the use of the parametric principle of justifiable granularity, *IEEE Trans. Fuzzy Systems*, vol.24, no.2, pp.489-496, 2016.
[13] J. Kennedy and R. C. Eberhart, Particle swarm optimization, *Proc. of IEEE International Conference on Neural Networks*, Piscataway, NJ, pp.1942-1948, 1995.