http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (15): 3046-3057, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3046.3057

Influence on Normalization and Magnitude Normalization for Price Forecasting of Agricultural Products

Zhu Quanyin, Pan Lu, Yin Yonghua and Li Xiang
Faculty of Computer Engineering, Huaiyin Institute of Technology, Huaian 223005, China

Abstract: In order to obtain suitable model and increase accuracy in price forecasting, data pretreatment approach via data normalization and the order of magnitude normalization which has the effect on price forecasting is proposed in this paper. The new data preprocessing method that normalized or normalizing the order of magnitude of the original data based on the proposed approaches are described in detail. The data of agricultural products extracted from Website based on improved Back Propagation (BP) neural network and Support Vector Machine (SVM) are utilized the proposed normalization methods and obtain the different results. Experiments demonstrate that the best forecasting average accuracy based on SVM is improved 0.33 percent by normalization and 0.35 percent by normalization of magnitude 10 and the best forecasting average accuracy based on improved BP neural network with no normalization is the best one, but the best of normalization of magnitude 100 cab be lifted 0.66 percent compare with the average accuracy of no normalization. Experiments demonstrate that the proposed approach performance and proves a new data pretreatment method via normalization is meaningful and useful for the model research of price forecasting.

Key words: Normalization, price forecasting, support vector machine, BP neural network, agricultural products

INTRODUCTION

Price forecasting has been a hot research and topic. Nowadays, price changes have great effect on both businesses and consumers. Price forecasting is meaningful and useful not only for businesses but also for consumers. It will bring an active market which brings the benefits to the businesses even the government officials. In particular, it has a significant impact on the electronic market. On the one hand, replacement of electronic products is growing fast, especially the products of computers and mobile phones. Quick update brings a severe unstable of price. Faced with the unstable prices, years of experience of the businesses are also becoming useless. Under the rapid development of the economic environment at present based on the Websites, judgment ahead of the market will make businesses' works fruitful or in vain.

But then, the psychology of consumers also changes with the economic development. The unstable prices make some consumers become rational. It shows that the demand for consumers is unstable when the reaction of consumers is projected onto the market.

However, big enterprises continue to quickly update their products, the businesses can only cut prices constantly and passively and the consumers can only wait to see.

Without a doubt, price forecasting is worth to study. At home and abroad, great deals of research are done in different fields of price forecasting, such as electricity market (Pindoriya *et al*., 2008; Lira *et al*., 2009; Lei and Shahidehpour, 2010; Catalao *et al*., 2011), oil futures (Rotering and Ilic, 2011) and gold futures (Xu *et al*., 2009) etc. A variety of methods have been proposed for different fields such as forecasting power prices using a hybrid econometric model (Gonzalez *et al*., 2012; Gradojevic and Gencay, 2011); forecasting stock price based on wavelet and TSK Fuzzy rules (Chang and Fan, 2008); a dynamic demand forecasting in smart grids (Motamedi *et al*., 2012); theoretical foundations for efficient coordination of wind power and price-responsive demand (Ilic *et al*., 2011); short-term forecasting of stock price based on support vector regression (Chen *et al*., 2012), etc. And all of these methods have gotten the ideal results in the papers proposed. In all of methods that have been proposed, most of them are concerned about the modifications on the existing prediction algorithm. They change the various parameters of the algorithm, or modify existing algorithms directly from the internal. Of course, part of them is also concerned on the data pretreatment, but the magnitude of the data they used is the same, such as gold, electricity price and futures etc.

In this study, the data of eight types of agricultural products are used which are beef, soybean oil, mutton,

Corresponding Author: Zhu Quanyin, Faculty of Computer Engineering, Huaiyin Institute of Technology, Huaian 223005, China

blend oil, white sugar, rice, pork and flour. The data are all extracted from the Website (http://www.abchina.com/cn/RuralSvc/Information/RealtimePrice/AgriculturalMarkets_Information/) from January 2012 to December 2012, a total of one year of data. The orders of the original data are 1, 10, 100, 1000 and 10000. The magnitude of the original data is normalized to different orders. Experiments with these pre-processed data based on SVM and improved BP neural network obtain different results. Experiments with the pretreatment data the order of which is one obtain the best results. The results prove that the novel approach is a good idea and further researches on it are meaningful.

NOTATIONS AND IMPROVEMENT MODELS

Notations: Some definitions used in this paper are given as follows:

Single errors of predicted value:

$$e_t = Y_t - \hat{Y}_t, \quad t = 1, 2, \cdots, n \tag{1}$$

Relative errors of single predicted value:

$$\tilde{e}_t = \frac{e_t}{Y_t} = \frac{Y_t - \hat{Y}_t}{Y_t}, \quad t = 1, 2, \cdots, n \tag{2}$$

Mean Absolute Errors (MAE):

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| = \frac{1}{n}\sum_{t=1}^{n}\left|Y_t - \hat{Y}_t\right| \tag{3}$$

Improved SVM algorithm: SVM is a supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis.

Assume time sequence data normalized is $\hat{A}_{order} = \{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n\}$, any element is $\forall \hat{x}_i \in \hat{A}_{order}, i \in [1, n]$ and normalized data is $\hat{A}_{order} = \{\hat{A}_0, \hat{A}_1, \hat{A}_{10}, \hat{A}_{100}, \hat{A}_{1000}, \hat{A}_{10000}\}$. For the orders of the original data of agricultural products have five orders of magnitude, so the normalization defines as $\hat{A}_0$, one order of magnitude normalization defines as $\hat{A}_1$ and so on. So the equations can be denoted as follows:

Polynomial (homogeneous):

$$k_{order}(\hat{x}_i, \hat{x}_j) = (\hat{x}_i \bullet \hat{x}_j)^d \tag{4}$$

Polynomial (inhomogeneous):

$$k_{order}(\hat{x}_i, \hat{x}_j) = (\hat{x}_i \bullet \hat{x}_j + 1)^d \tag{5}$$

Gaussian radial basis function:

$$k_{order}(\hat{x}_i, \hat{x}_j) = \exp(-\gamma\|\hat{x}_i - \hat{x}_j\|^2) \tag{6}$$

$\gamma>0$, sometimes parameterized using $\gamma = 1/2\sigma^2$.

Hyperbolic tangent:

$$k_{order}(\hat{x}_i, \hat{x}_j) = \tanh(k_{order}\hat{x}_i \bullet \hat{x}_j + c) \tag{7}$$

For some (not every) is $k > 0$ and $c < 0$.

Improved BP neural network approaches: BP NN always uses activation functions. First of all, pre-processing the original data must be done to get the experimental data. The pretreatment data of A_1 is given by Eq. 8.

$$\hat{A}_1 = \{\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n\} \tag{8}$$

Then, z is defined as the number of a forecast sample and D is defined as the number of the forecast value. The training function of the model is given by equation (9) and the forecasting function of the model is given by equation (10). Both of them are built-in functions of the Matlab.

$$NET = newff(P,T,NEURON),\ NET' = train(NET,P,T) \tag{9}$$

$$NET\ (Test) \tag{10}$$

Then, j is defined as the number of the different Neuron of BP neural network. When forecasting the first value, many different Neuron are given. And the different Neurons are given by the Eq. 11.

$$Neurons = \{neuron_1, neuron_2, \ldots, neuron_j\} \tag{11}$$

Then, different net will be trained to forecast the first value with the Neurons which is given by equation through the Eq. 12.

$$net_{lj} = newff(P,T,neuron_j),\ net_{lj} = train(net_{lj},P,T) \tag{12}$$

In the Eq. 12, P is the training sample set and T is the test forecast value set of training. A forecasting sample Test is necessary to be built. And Test is given by Eq. 13.

$$Test = [t_1, t_2, \cdots, t_z], t_z \in \hat{A}_1 \tag{13}$$

Then, the forecast values of the (n+1) day of A_l can be gotten through the Eq. 13. Y_{lj} is defined as the forecast values of the (n +1) day of A_l and it is given by Eq. 14.

$$Y_{lj} = net_{lj}(Test) \quad (14)$$

In Y_{lj}, the best value is selected to be the forecast value of the (n +1) day of Y_{lj} and it is also selected to be the first forecast value y_l. Three best Neuron in the Neurons are selected according to the three best forecast values in Y_{lj}. And three best Neuron is given by Eq. 15.

$$Bneuron_{l1} \in Neurons, Bneuron_{l2} \in Neurons, Bneuron_{l3} \in Neurons \quad (15)$$

A custom weight W is given by equation (16) and in this paper W= [2, 4, 2].

$$W = (w_1, w_2, w_3) \quad (16)$$

Bneuron is defined as the best Neuron and it is given by Eq. 17.

$$Bneuron = \frac{Bneuron_{11} * w_1 + Bneuron_{12} * w_2 + Bneuron_{13} * w_3}{w_1 + w_2 + w_3} \quad (17)$$

Then, when forecasting the rest values, a constant network is being used. The net is given by Eq. 18.

$$net = newff(P,T,Bneuron),\ net = train(net,P,T) \quad (18)$$

When forecasting the rest values, the previous forecast value will be seen as the one of the values of the forecast samples for the next prediction. And it is given by Eq. 19-21.

$$(t_1, t_2, \ldots, t_{z-1}) = (t_2, t_3, \ldots, t_z) \quad (19)$$

$$t_z = y_{lm} \quad (20)$$

$$Test = [t_1, t_2, \ldots, t_z] \quad (21)$$

Finally, Y_{lv} is defined as the all forecast values and it is given by Eq. 22.

$$Y_{lv} = net(Test) \quad (22)$$

EXPERIMENTS PREPARATIONS

The experimental original data are all extracted from the Website (http://www.abchina.com/cn/RuralSvc/Information/RealtimePrice/AgriculturalMarkets_Information/) from January 2012 to December 2012, a total of one year of data. The orders of the original data are 1, 10, 100, 1000 and 10000. One week has an original data which express average piece weekly.

Table 1: The defining of agricultural products

PD numbers	Product name	PD numbers	Product name
PD1	Beef	PD5	White Sugar
PD2	Soybean Oil	PD6	Rice
PD3	Mutton	PD7	Pork
PD4	Blend Oil	PD8	Flour

Table 2: The defining of price forecasting model

MD	Price Forecasting Model
MD1	Improved BP neural network with no normalization
MD2	Improved BP neural network with normalization
MD3	Improved BP neural network with normalization of magnitude 1
MD4	Improved BP neural network with normalization of magnitude 10
MD5	Improved BP neural network with normalization of magnitude 100
MD6	Improved BP neural network with normalization of magnitude 1000
MD7	Improved BP neural network with normalization of magnitude 10000
MD8	SVM model with no normalization
MD9	SVM model with normalization
MD10	SVM model with normalization of magnitude 1
MD11	SVM model with normalization of magnitude 10
MD12	SVM model with normalization of magnitude 100

Some improvements are done based on the original model on the basis of our previous research work (Zhu *et al*., 2010; Zhu *et al*., 2011a; Zhu *et al*., 2011b; Zhu *et al*., 2011c; Zhu *et al*., 2012a; Zhu *et al*., 2012b; Wu and Zhu, 2012; Zhu *et al*., 2012c). In past research, many improved methods have been proposed based on original models which are applied to some different fields, such as frequency algorithm and its application in stock index movement prediction (Zhang *et al*., 2012), optimization of neural networks using variable structure systems (Mohseni and Tan, 2012) and divide and concur and difference-map BP decoders for LDPC codes (Yedidia *et al*., 2011) etc.

Different types of agricultural products are defined as PD1 to PD8 and different types of magnitude normalization using SVM or improved BP neural network model are defined as MD1 to MD12 which are shown in Table 1 and 2, respectively.

The simulation software is the MATLAB of MathWorks. The software version is R2011b (7.13.0.564).

EXPERIMENTAL RESULTS

The experiment results which are gotten with the preprocessed data based on the improved algorithm are shown above which are SVM model and improved BP Neural Network. In order to get the influence of our proposed data pretreatment approach, we select different normalization methods for SVM model and improved BP Neural Network respectively. The actual price define be AP and the forecast price define be FP respectively.

Experimental Results for Each Agricultural Product: Eight different types of agricultural products are opted to experiment on the SVM model and improved BP Neural Network respectively. Figure 1 to 8 shows the experimental results respectively.

Experimental MAE for Each MD: The MAE of ten MD is calculated based on different type of agricultural products which SVM model and improved BP Neural Network are used and the experimental results shown as Fig. 9-20, respectively.

Experiments analysis: Analysis results of twelve methods can be drawn that all of them are experimental results. The effect of magnitude differences in the original data of agricultural products based on improved BP neural network is not obvious. Certainly, the result of experiments based on improved BP neural network is still very meaningful. Among the better results, the best result is still the result of the experiment which is done with the data the magnitude of which is normalized to the different order the same as the experiment based on SVM model. Relative to the great changes based on the SVM model, the order of magnitude of the data has a bit effect on the BP neural network. So a little improve proves the new method is general to SVM model. One step further, the new idea and the new method is general to other algorithms.

Table 3 shows the average MAE of eight PD via twelve MD. Table 4 shows the each week average MAE of price forecasted via twelve MD.

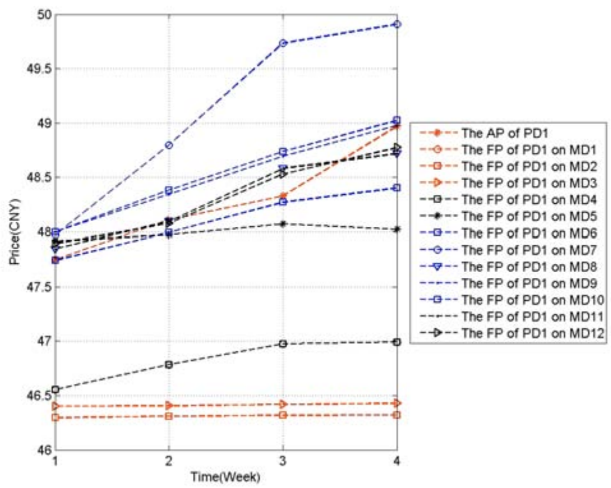

Fig. 1: The forecast price of PD1 on ten MD

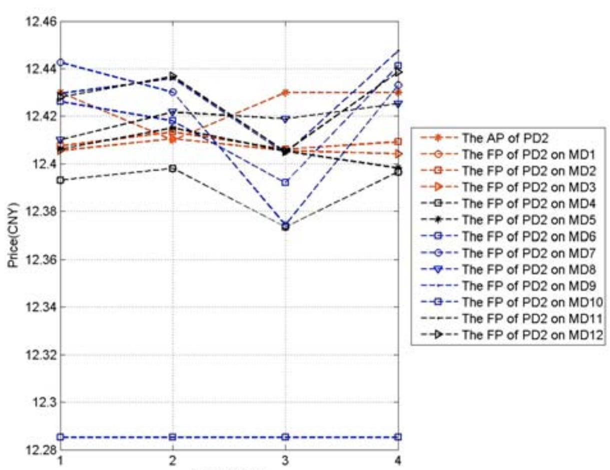

Fig. 2: The forecast price of PD2 on ten MD

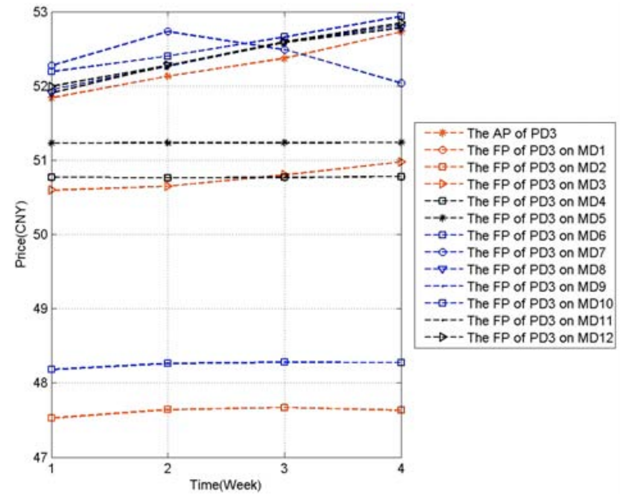

Fig. 3: The forecast price of PD3 on ten MD

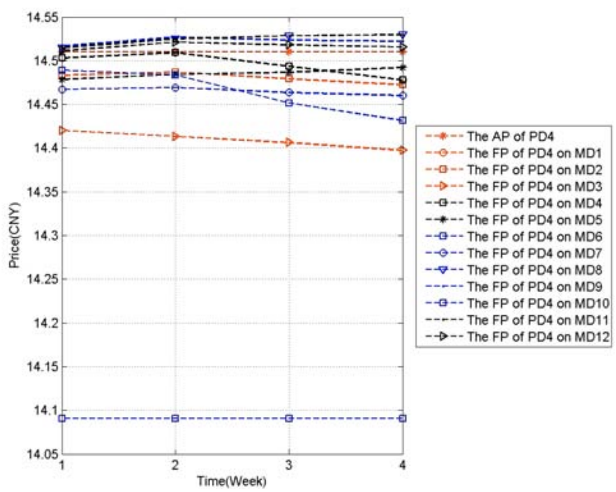

Fig. 4: The forecast price of PD4 on ten MD

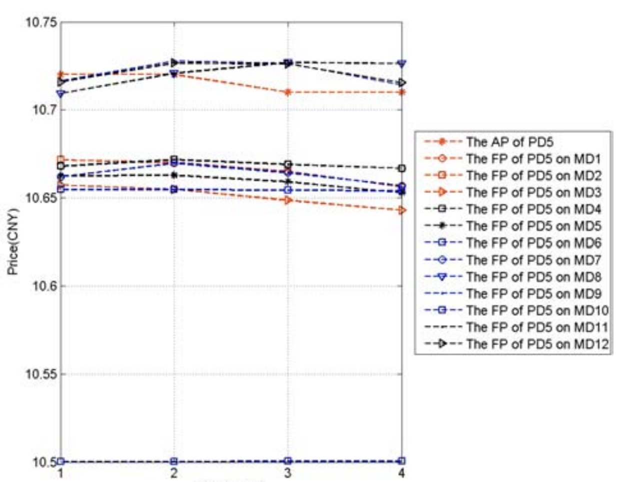

Fig. 5: The forecast price of PD5 on ten MD

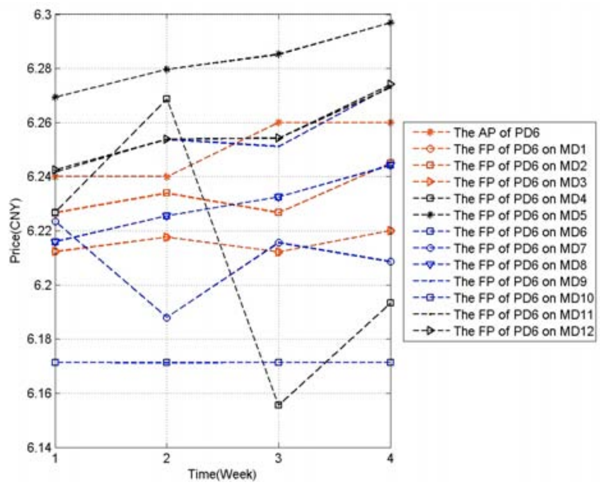

Fig. 6. The forecast price of PD6 on ten MD

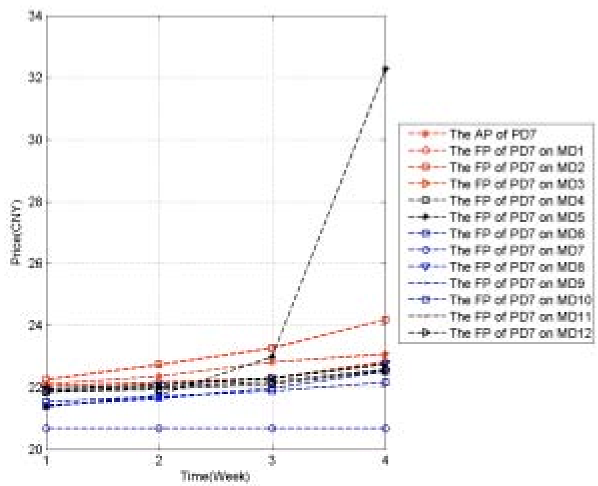

Fig. 7. The forecast price of PD7 on ten MD

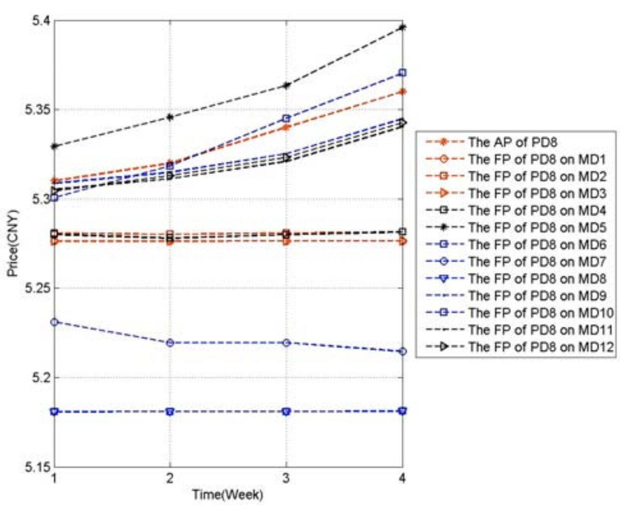

Fig. 8. The forecast price of PD8 on ten MD

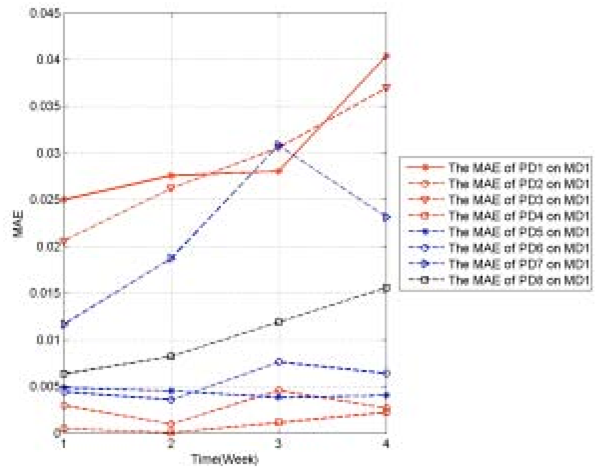

Fig. 9. The MAE of different PD on MD1

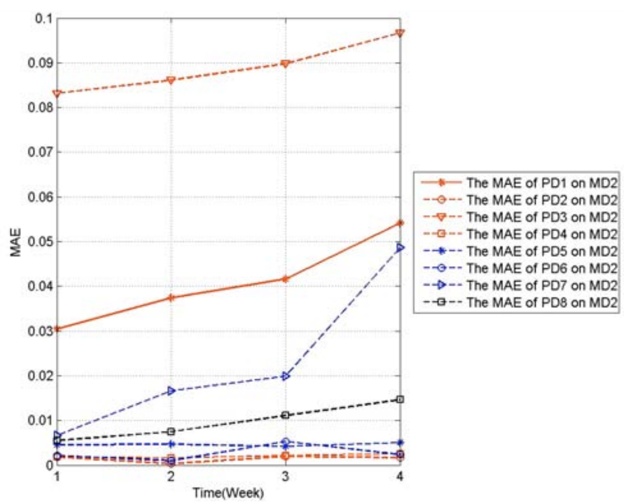

Fig. 10. The MAE of different PD on MD2

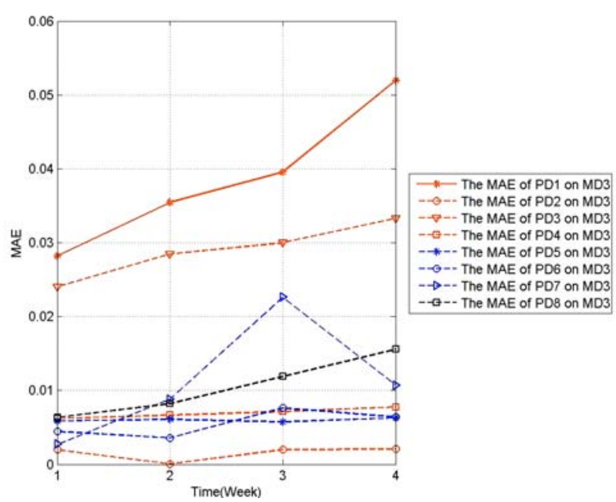

Fig. 11. The MAE of different PD on MD3

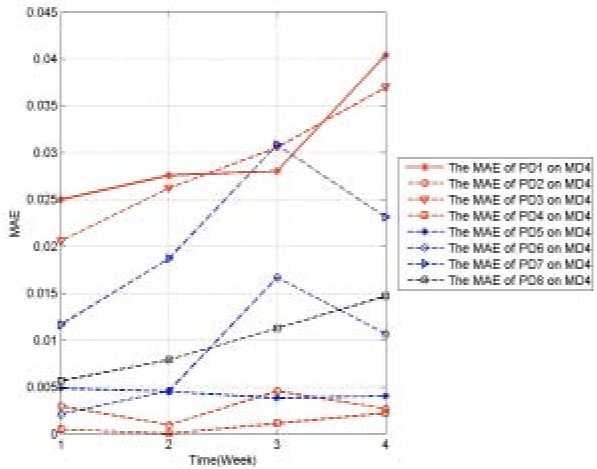

Fig. 12. The MAE of different PD on MD4

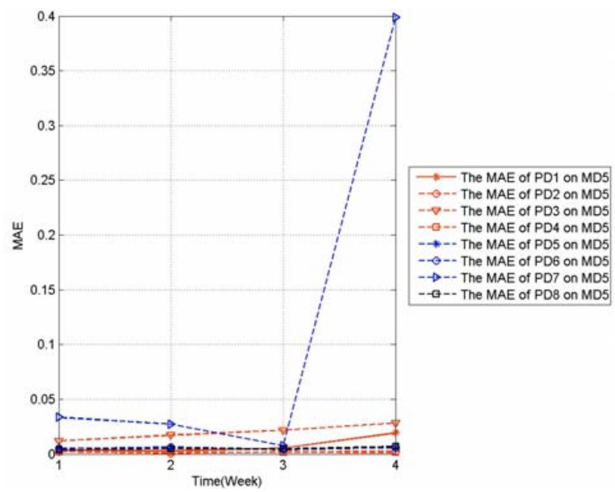

Fig. 13. The MAE of different PD on MD5

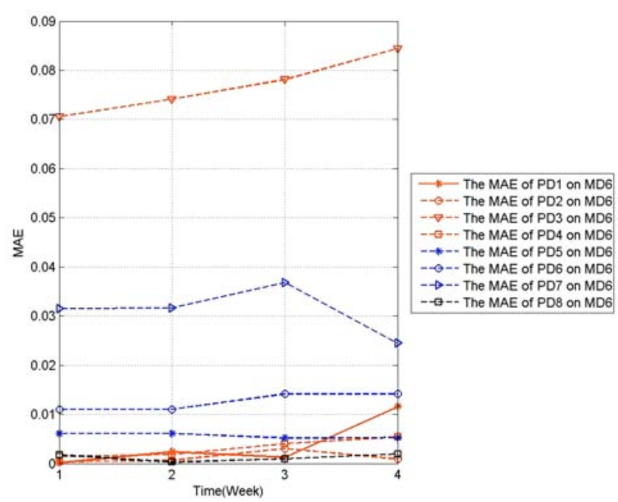

Fig. 14. The MAE of different PD on MD6

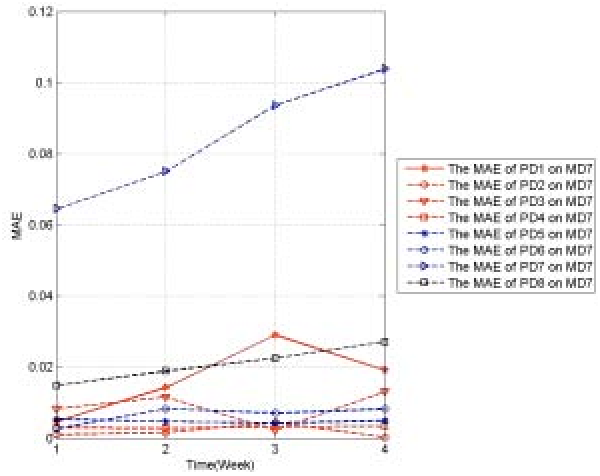

Fig. 15. The MAE of different PD on MD7

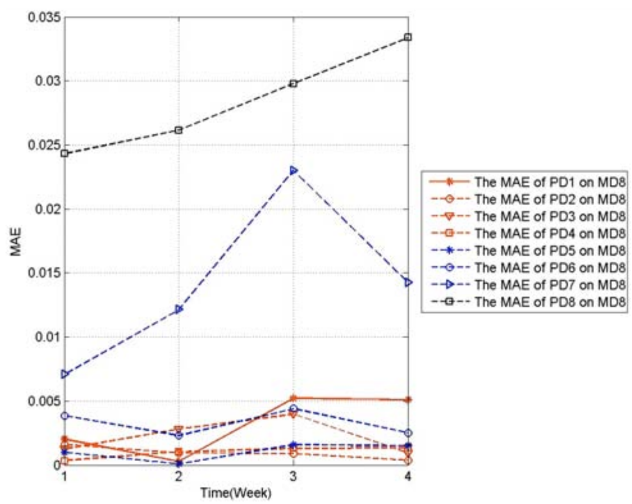

Fig. 16. The MAE of different PD on MD8

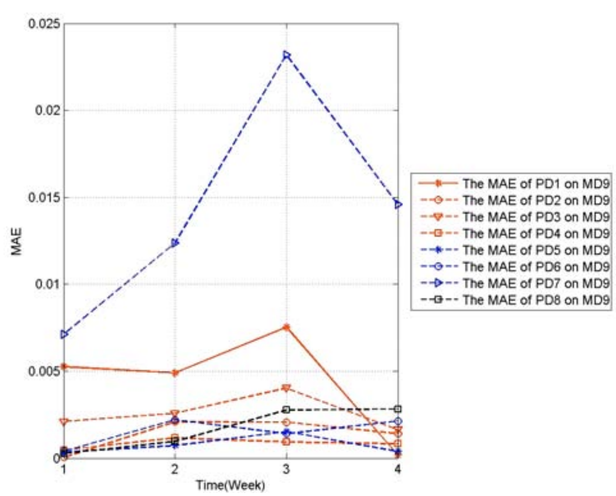

Fig. 17. The MAE of different PD on MD9

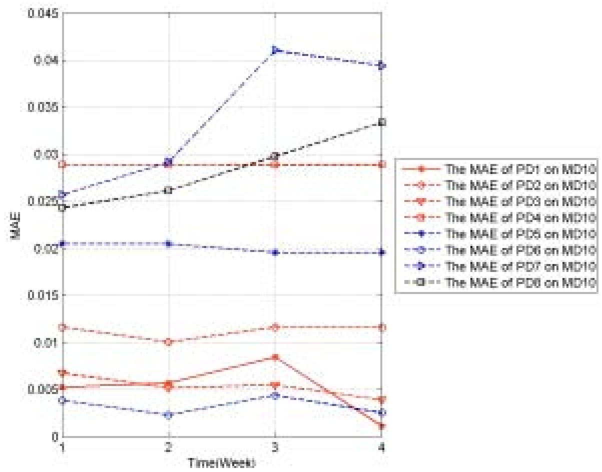

Fig. 18. The MAE of different PD on MD10

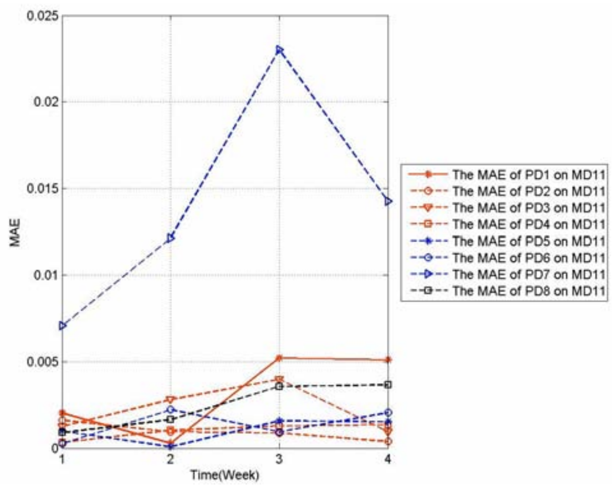

Fig. 19. The MAE of different PD on MD11

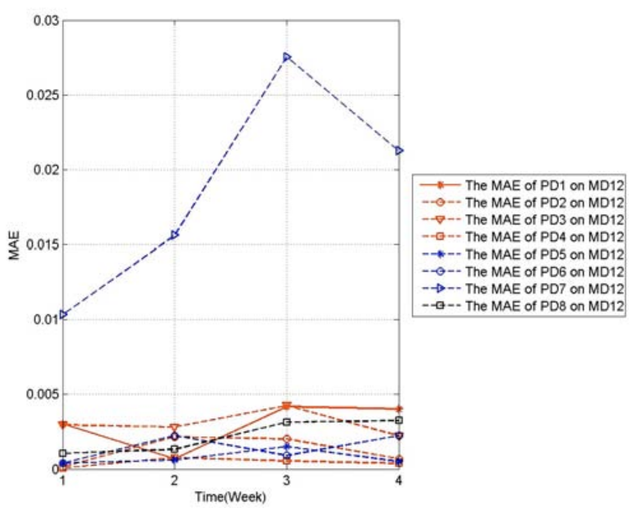

Fig. 20. The MAE of different PD on MD12

Table 3: The defining of agricultural products

PD	MAE of FP	PD	MAE of FP
PD1	Beef	PD5	White Sugar
PD2	Soybean Oil	PD6	Rice
PD3	Mutton	PD7	Pork
PD4	Blend Oil	PD8	Flour

Table 4: The defining of agricultural products

MD	MAE		MD	MAE	
MD1	FiW	0.009545	MD7	FiW	0.013049
	SW	0.011225		SW	0.017148
	TW	0.014815		TW	0.020821
	FoW	0.016425		FoW	0.022515
	Ave	0.013003		Ave	0.018383
MD2	FiW	0.017016	MD8	FiW	0.005183
	SW	0.019395		SW	0.005716
	TW	0.021998		TW	0.008768
	FoW	0.028212		FoW	0.007430
	Ave	0.021655		Ave	0.006774
MD3	FiW	0.009966	MD9	FiW	0.001997
	SW	0.012160		SW	0.003377
	TW	0.015819		TW	0.005431
	FoW	0.016750		FoW	0.002999
	Ave	0.013674		Ave	0.003451
MD4	FiW	0.009168	MD10	FiW	0.015868
	SW	0.011313		SW	0.015991
	TW	0.015867		TW	0.018664
	FoW	0.016845		FoW	0.017554
	Ave	0.013298		Ave	0.017019
MD5	FiW	0.008314	MD11	FiW	0.001806
	SW	0.008237		SW	0.002642
	TW	0.006399		TW	0.005060
	FoW	0.058489		FoW	0.003656
	Ave	0.020360		Ave	0.003291
MD6	FiW	0.015356	MD12	FiW	0.002297
	SW	0.016010		SW	0.003275
	TW	0.017944		TW	0.005510
	FoW	0.018523		FoW	0.004320
	Ave	0.016958		Ave	0.003850

From Fig. 1-8 and Fig. 9-20, we know the best FP is not the same for the different MD, it told us that if we want to get the best FP, we should to do more experiments. But from Table 3 and 4, we know that if we not always want to get the best, we can use the one model to get the FP, but we need to normalize the original data. However, from Table 3 and 4 we can also get that the different data normalization method may suitable different product. That may used by our experience, but from tables we can know no normalization or normalization of magnitude 100 for improved BP neural network is the better and the no normalization or normalization of magnitude 10 for SVM method may the better way to save time and may suitable our requirements.

CONCLUSIONS

The market of agricultural products changes very fast. The price forecasting models is focused on many researchers. The data mining and pretreatment is an important way to improve the forecast models. We do lots of experiments in order to reveal the effect on price forecast model. So we process the original data depend on normalization and magnitude normalization. After experiments, we find that the normalization can affect the MAE and speed on the different model. Experiments demonstrate that the best forecasting average accuracy which is gotten based on SVM obtains 99.67% which is normalized of magnitude 10 and the best forecasting average accuracy which is gotten based on improved BP neural network obtains 98.70% which is no normalized, but the highest accuracy only one week can get 99.36 which is normalized of magnitude 100. Experiments demonstrate that the proposed approach performance and proves a new data pre-treatment method is meaningful and useful for the model research of price forecasting.

ACKNOWLEDGMENTS

This study is supported by the National Sparking Plan Project of China (2011GA690190), the Major Program of the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (11KJA460001); the fund of Huaian Industry Science and Technology of China (HAG2011052, HAG2011045, HAG2012086, HASZ2012046).

REFERENCES

Catalao, J.P.S., H.M.I. Pousinho and V.M.F. Mendes, 2011. Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting. IEEE Trans. Power Syst., 26: 137-144.

Chang, P.C. and C.Y. Fan, 2008. A hybrid system integrating a wavelet and TSK fuzzy rules for stock price forecasting. IEEE Trans. Syst. Man Cybernetics Part C: Appl. Rev., 38: 802-815.

Chen, X., Z.Y. Dong, K. Meng, Y. Xu, K.P. Wong and Ngan, 2012. Electricity price forecasting with extreme learning machine and bootstrapping. IEEE Trans. Power Syst., 27: 2055-2062.

Gonzalez, V., J. Contreras and D.W. Bunn, 2012. Forecasting power prices using a hybrid fundamental-econometric Model. IEEE Trans. Power Syst., 27: 363-372.

Gradojevic, N. and R. Gencay, 2011. Financial applications of nonextensive entropy. IEEE Signal Process. Mag., 28: 116-141.

Ilic, M.D., X. Le and J.Y. Joo, 2011. Efficient coordination of wind power and price-responsive demand. IEEE Trans. Power Syst., 26: 1875-1884.

Lei, W. and M. Shahidehpour, 2010. A hybrid model for day-ahead price forecasting. IEEE Trans. Power Syst., 25: 1519-1530.

Lira, F., C. Munoz, F. Nunez and A. Cipriano, 2009. Short-term forecasting of electricity prices in the colombian electricity market. IET Generation, Transmission Distribution, 3: 980-986.

Mohseni, S.A. and A.H. Tan, 2012. Optimization of neural networks using variable structure systems. IEEE Trans. Syst. Man Cybernetics Part B: Cybernetics, 42: 1645-1653.

Motamedi, A., H. Zareipour and W.D. Rosehart, 2012. Electricity price and demand forecasting in smart grids. IEEE Trans. Smart Grid, 3: 664-674.

Pindoriya, N.M., S.N. Singh and S.K. Singh, 2008. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans. Power Syst., 23: 1423-1432.

Rotering, N. and M. Ilic, 2011. Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets. IEEE Trans. Power Syst., 26: 1021-1029.

Wu, B. and Q.Y. Zhu, 2012. Week-ahead price forecasting for steel market based on RBF NN and ASW. Proceedings of the 3rd International Conference on Software Engineering and Service Science, June 22-24, 2012, Beijing, pp: 729-732.

Xu, G.X., B.C. Shia, Y.B. Hsu, P.C. Shen and K.H. Chu, 2009. To integrate text mining and artificial neural network to forecast gold futures price. Proceedings of the International Conference on New Trends in Information and Service Science, June 30-July 2, 2009, Beijing, China.

Yedidia, J.S., Y. Wang and S.C. Draper, 2011. Divide and concur and difference-map BP decoders for LDPC codes. IEEE Trans. Inform. Theory, 57: 786-802.

Zhang, L.M., N. Liu and P.Y. Yu, 2012. A novel instantaneous frequency algorithm and its application in stock index movement prediction. IEEE J. Selected Topics Signal Process., 6: 311-318.

Zhu, Q.Y., Y.Y. Yan, J. Ding and Y. Zhang, 2010. The commodities price extracting for shop online. Proceedings of the International Conference on Future Information Technology and Management Engineering (FITME), Volume: 2, October 9-10, 2010, Changzhou, China, pp: 317-320.

Zhu, Q.Y., Y. Y. Yan, J. Ding and J. Qian, 2011a. The case study for price extracting of mobile phone sell online. Proceedings of the IEEE 2nd International Conference on Software Engineering and Service Science, July 15-17, 2011, Beijing, China, pp: 282-285-10.1109/ICSESS.2011.5982309.

Zhu, Q.Y., S.Q. Cao, J. Ding and Z.Y. Han, 2011b. Research on the price forecast without complete data based on web mining. Proceedings of the 10th International Symposium on Distributed Computing and Applications to Business, Engineering and Science, October 14-17, 2011, Karlsruhe, Germany, pp: 120-123.

Zhu, Q.Y., S.Q. Cao, P. Zhou, Y.Y. Yan, H. Zhou, 2011c. Integrated price forecast based on dichotomy backfilling and disturbance factor algorithm. Int. Rev. on Comput. and Software, 6: 1089-1093.

Zhu, Q.Y., P. Zhou, S.Q. Cao, Y.Y. Yan, J. Qian, 2012a. A novel RDB-SW approach for commodities price dynamic trend analysis based on Web Mining. J. of Digital Inform. Manage., 10: 230-235.

Zhu, Q.Y., J.J. Zong and Y.H. Yin, 2012b. Agriculture products price forecasting with grey method and neural network. Int. Rev. on Comput. Software, 7: 1363-1369.

Zhu, Q.Y., J. Ding, Y.H. Yin and P. Zhou, 2012c. A hybrid approach for new products discovery of cell phone based on web mining. J. Inform. Comput. Sci., 9: 5039-5046.