International Journal of Database Theory and Application
Vol.9, No.1 (2016), pp.67-76
http://dx.doi.org/10.14257/ijdta.2016.9.1.06

# The Combination Forecasting Model of Auto Sales Based on Seasonal Index and RBF Neural Network

Lihua Yang and Baolin Li*

*School of Economics and Management, Hubei University of Automotive Technology, Shiyan, 442002,China*
*yanglh168@126.com;libaolin108@163.com*

***Abstract***

*To effectively predict auto sales and improve the competitiveness of automotive enterprise, the characteristics of actual auto sales were analyzed, owing to the seasonal fluctuations and the nonlinearity of monthly sales, the combination forecasting model based on seasonal Index and RBF neural network was proposed. The weights of the two single models were computed using mean absolute percentage error and the sum of square error respectively, the result shows that mean absolute percentage error is more effective. Finally, the prediction accuracy of different models was compared based on the criteria of MAPE and RMSE, and the effectiveness of the method was proved, the proposed model can take advantage of the strengths of the two single models, the results indicate that the combination forecasting model suitable for auto sales has high prediction accuracy, which can provide a certain reference to auto sales forecasting.*

***Keywords****: Combination Forecasting, auto Sales, seasonal Index, RBF*

## 1. Introduction

With the rapid development of automotive industry, the automobile market is playing more and more important role on the national economic development. The accuracy of market information is the core competitiveness of auto enterprises, the auto sales forecasting can affect the scientific of decision-making of managers, so it is very important to enterprise decision. Therefore, how to improve the sales forecasting accuracy has become a hot spot for automotive industry, which can help company to improve the competitiveness of marketing.

In recent years, many scholars have researched the sales forecasting methods. In generally, the forecasting methods can be generally divided into two categories, namely, the qualitative forecasting methods and quantitative forecasting methods. The latter methods include some traditional statistics methods, such as moving average, exponential smoothing and multiple regression analysis. In fact, auto sales are affected by many factors, such as economic situation, state policy, the income of the family, and so on. These complicated factors cause the remarkable fluctuation and non-linear characteristics of the historic sales data, so some data don’t have trends and display high fluctuation, to solve the problem, some data mining algorithms are applied to sales forecasting due to the complexity of sales data.

ANN models have been revealed to be more efficient and effective than many traditional statistical forecasting models, Despite the reported advantages, it is relatively more time-consuming for ANN to perform forecasting, a new model which employs both the extreme learning machine and the traditional statistical methods is proposed, a comparison with other traditional methods has shown that this ELM fast forecasting model is quick and effective [1]. To improve the accuracy of fashion retail forecasting, a two-stage dynamic forecasting model is proposed, which is combined with both long-term and short-term predictions [2]. Time series forecasting, as an important tool in many decision support systems, has been extensively studied and applied for sales forecasting over the past few decades, a hybrid forecasting scheme which combines the classic SARIMA method and wavelet transform is proposed [3]. A new market forecasting system for tobacco wholesalers in China based on a new developed market demand forecasting model is presented, this model can help sales company to improve the forecasting accuracy for annual and monthly market demands significantly [4]. Nonlinear and non-stationary evolution and dependencies with diverse macroeconomic variables hinder accurate long-term prediction of the future of automotive sales, a structural relationship identification methodology is presented to identify the dynamic couplings among automobile sales and economic indicators, the empirical results suggest that VECM can significantly improve prediction accuracy of automotive sales for 12-month ahead prediction compared to the classical time series techniques [5].Using the combination forecasting method which based on PLS to predict the province's cigarette sales of the next year, the results show that the prediction accuracy is good, which could provide a certain reference to cigarette sales forecasting in a province[6]. Two strategies for forecasting – using the set of good forecasters in the testing phase versus filtering out the bad forecasters and using the resulting set of forecasters were compared, experimental results with over 30 sales series indicate that our heuristics greatly out-perform those such as the simple mean [7]. An effective multivariate intelligent decision-making model is developed to provide effective forecasts by integrating a data preparation and preprocessing module [8]. Support vector regression was applied to the newspaper or magazines sales forecasting is a superior method [9].

Most researches usually adopt a single model, but the single model has its own limitation, which can only use effective information of a certain. Recently, combination forecasting model has gained a lot of attention. The concept of combination model obtained from different models has been discussed and used in different fields [10]. However, a few studies have been proposed to examine the accuracy of combination model for auto sales. In order to overcome the inaccuracy of the single forecasting model, a new combination model is established to increase accuracy of auto sales forecasting, in which two forecasting models based on seasonal index and radical basis function neural network are introduced. Actual monthly sales data from January 2010 to December 2013 are used to establish the model, actual monthly sales data from January 2014 to April 2014 are used to test the model. These models are compared according to the MAPE and RMSE of forecasting results. This new forecasting model is an extension to the combination prediction method.

The organization of the paper is shown in Figure 1. Firstly, we obtain the related auto sales data, analyzing the data characteristics; secondly, two single models were established based on the selection data; thirdly, the weights of combination model were calculated based on relative error and absolute error respectively, by comparing the predictive accuracy of the different combination models, the relative error is more effective in computing the weights; finally, the forecasting results of three models were compared based on the criteria of MAPE and RMSE, then the optimization model was proposed.

 

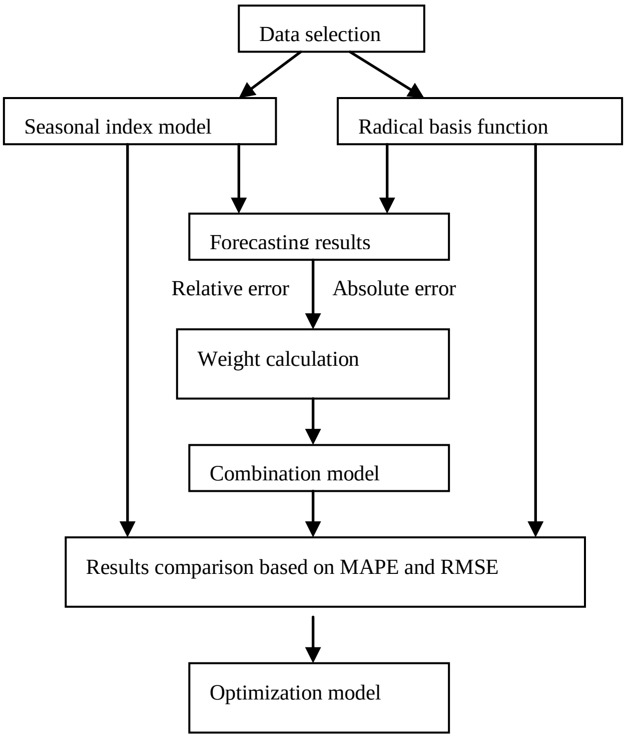


**Figure 1.The Organization of the Paper**

## 2. Establishment of the Two Single Models

### 2.1. Data Selection

In this paper, we obtained monthly auto sales data from China association of automobile manufacturers. To do monthly auto sales forecasting, the models are constructed based on monthly sales data from January 2010 to December 2013, namely 4 years, a total of 48 samples. The forecasting models are tested using the monthly sales data, from January 2014 to April 2014, a total of 4 samples. The monthly sales data are shown in Figure 2.

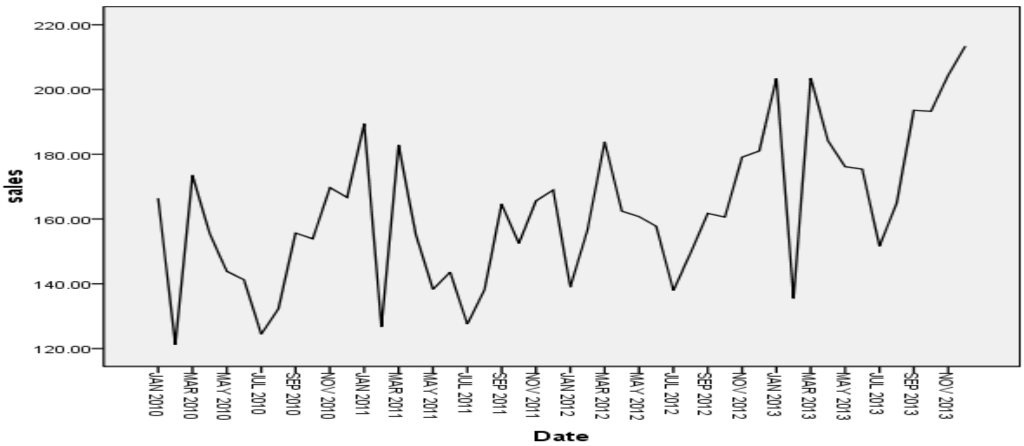


**Figure 2. Monthly Sales Data**

From the above figure, we can see that the series exhibit a remarkable periodic component. One hand, There is an obvious seasonal increase in January and March each year, on the other hand, there is a strongly seasonal decrease in February and July. The time series indicate that seasonal underlying periodicity probably is 12 months. We can also notice that the seasonal variations appear to grow with the upward series trend, which implies a multiplicative model is better than an additive model.

We can gains a more quantitative conclusion about the underlying periodicity from the autocorrelations and partial autocorrelations of the time series, the autocorrelation function in Figure 3. shows a significant peak at a lag of 12, which suggests the presence of an annual seasonal component in the data.

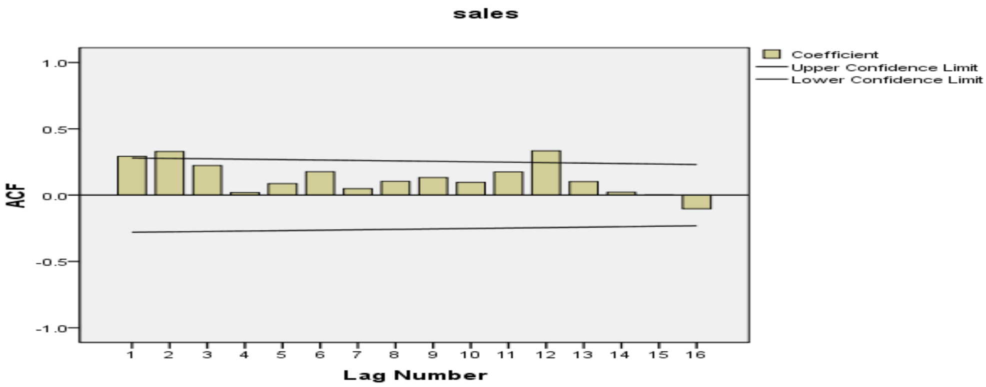


**Figure 3. Autocorrelation Function**

## 2.2. Seasonal Index Model

Time series can be decomposed into four components: the long term trend, the seasonal component, the cyclical component and the irregular component, the relations between the four components are summarized as two models, multiplicative model and additive model. From the above analysis, we can draw the conclusion that multiplicative seasonal adjustment model can be applied to auto sales forecasting. The seasonal decomposition procedure can creates four new variables by analyzing the auto sales original variables, where SAF is seasonal adjustment factors, representing seasonal variation; SAS is seasonally adjusted series, representing the original series removing seasonal variations; STC is smoothed trend-cycle component, which is a smoothed series that shows both trend and cyclic components; ERR is the residual component of the series.

Using the seasonal decomposition procedure, we have removed the seasonal component of a periodic time series to produce a series that is more suitable for trend analysis. Here, the moving average weight is all points equal, the smoothed trend-cycle series are shown in Figure 4, this series show a clear upward trend in the following figure. The seasonal factors series are shown in Figure 5, we can see that there are a number of peaks and troughs.

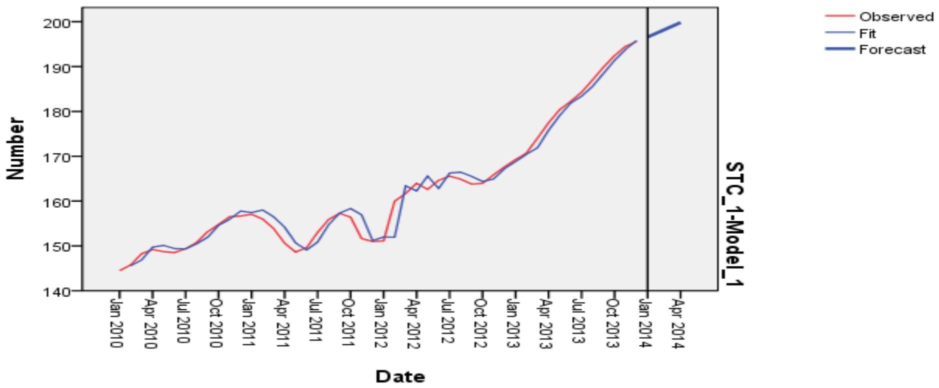


**Figure 4. The Smoothed Trend-Cycle Series**

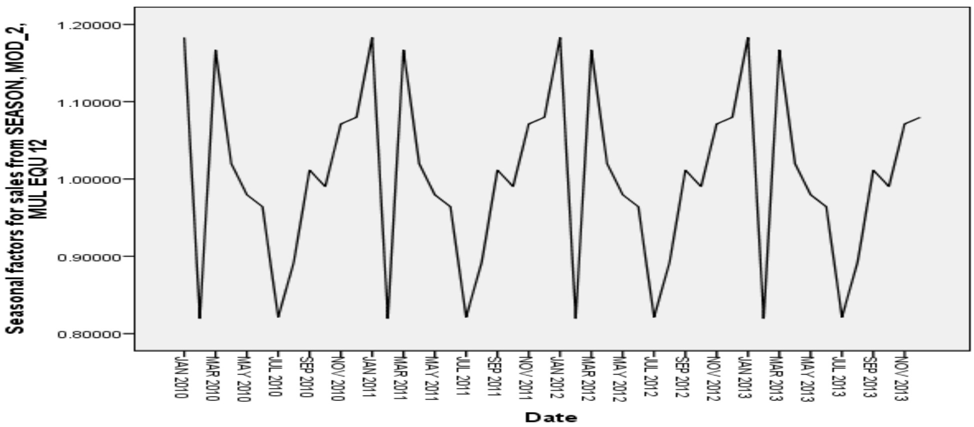


**Figure 5. Seasonal Factors for Sales**

### 2.3. RBF Model

A radial basis function network is an artificial neural network that uses radial basis functions as activation functions,which has many advantages, such as excellent fault tolerance, highly parallel mechanism, and adaptation.

The RBF network is composed of an input layer, hidden layer and an output layer, each layer contains a number of neurons, the input nodes of RBF neural network send signals to the hidden layer, and the output node is usually a simple linear function. As input to output is nonlinear, and the network output is linear to adjustable parameter mapping, the weight value of the network can be recursively calculated, this structure is known as a feedforward architecture because the connections in the network flow forward from the input layer to the output layer without any feedback loops. RBF network is supervised in the sense that the predicted results can be compared with actual values of the target variables.

The input layer of the network contains the predictors, and the output layer contains the responses. From the above analysis, the time series show that seasonal periodicity is 12 months, the variable of nextmonth is selected as the dependent variable, the auto sales of 12 months before nextmonth are selected as independent variables, a numeric variable of traintest is specified as partitioning variable, monthly sales data from January 2010 to December 2013 with a positive value on this variable are assigned to the training sample, monthly sales data from January 2014 to April 2014 with a value of 0 on this variable are assigned to the testing sample, the best number of hidden units is automatically computed according to the smallest error in the testing data, the activation function for the hidden layer is the normalized radial basis function, activation function for the output layer is identity function.

The processing results show that 36 cases were assigned to the training data set, 4 cases to the testing data set. The number of units in the input layer is the number of independent variables, there are 12 neurons in the input layer, likewise, there is 1 unit in the output layer, and there are 3 units in the hidden layer.

## 3. Establishment of combination Forecasting Model

### 3.1. Calculation of the Weight

The study of combination forecasting has become the hot topic in the forecasting field, it was first proposed by Bates and Granger in 1969, they advocated the use of weighted average in the combination forecasting in order to make use of information as much as possible. Information is different to be used in each single forecasting model, and its accuracy is also different, the combination model can take advantage of useful information of different single model, improving forecasting accuracy. How to allocate appropriate weight to single model is the core step of constructing combination model, now the combination forecasting model is generally established based on the criteria of the sum of absolute error, but some researchers point out that the relative error may be more accurate than the absolute error [9]. Considering the forecasting accuracy of relative error, here, the MAPE is proposed to compute the weight of the single model, MAPE represents mean absolute percentage error.

In the combination model, the sum of the weights is constrained to be equal to one, and the value of the weight cannot be less than zero. The principle of calculating weight is that the model of the smaller mean absolute percentage error should have a higher weight, and the model of the higher error of mean absolute percentage should have a smaller weight. Expressed in a formula, it can be written as follows.

$$W_i = M_i^{-1} / \sum_{i-1}^{m} M_i^{-1}, i = 1,2,\cdots,m$$

Where m is the number of model, $W_i$ is the weight of the ith model, $M_i$ is the mean absolute percentage error of the ith model. In the traditional inverse variance method, $M_i$ represents the sum of square error of the ith model, the formula is as follows.

$$M_i = \sum_{t=1}^{n} (y_t - \overline{y_t})^2$$

The combination forecasting model formula is expressed as followings:

$$Y = W_1 Y_1 + W_2 Y_2$$

Where $Y_1$ is the predictive value of seasonal index model, $Y_2$ is the predictive value of RBF neural network. $Y$ is the predictive value of combination forecasting model.

### 3.2. Combination Forecasting

The procedure of establishing the combination model is shown as follows.

1) Firstly, the seasonal index model and RBF neural network are established respectively to predict the monthly auto sales data. The predictive results and relative errors of the two models are shown in Table 1.

 

### Table 1. The Forecasting Results of Two Single Models

| Date_ | Actual sales | Seasonal index | relative error | RBF network | relative error |
|---|---|---|---|---|---|
| JAN 2014 | 215.64 | 232.58 | 0.079 | 200.19 | 0.072 |
| FEB 2014 | 159.64 | 161.96 | 0.015 | 200.11 | 0.253 |
| MAR 2014 | 216.91 | 231.93 | 0.069 | 200.26 | 0.077 |
| APR 2014 | 200.42 | 203.84 | 0.017 | 200.19 | 0.001 |

Here, the measuring unit of auto sales is ten thousand, the formula of relative error is as follows.

$$RE = \left| \frac{\overline{y_t} - y_t}{y_t} \right|$$

Where $\overline{y_t}$ denotes the predicted value, $y_t$ denotes the actual value.

From the Table 1, we can see that the two kinds of model forecasting results were different, because the single model can take advantage of different information of the data.

2) Secondly, after computing the forecasting results of two single models, the weights of two single models in the combination model can be calculated according to the testing data set. On the one hand, the mean absolute percentage error is proposed to calculate the weight of the single model, the weight of the seasonal index model is 0.692, and the weight of the RBF neural network is 0.308, on the other hand, the sum of square error is used to compute the weight of the single model, the weight of the seasonal index model is 0.803, and the weight of the RBF neural network is 0.179. The different weights are shown in Table 2.

### Table 2. The Comparison of the Weights

| Model | The mean absolute percentage error weight | The sum of square error weight |
|---|---|---|
| Seasonal index | 0.692 | 0.803 |
| RBF network | 0.308 | 0.197 |

3) Thirdly, based on the above weights, we can respectively calculate the forecasting result of combination model, the first combination model use MAPE to calculate the weights, the formula is as follows.

$$Y = 0.692Y_1 + 0.308Y_2 \quad (5)$$

The second combination model use the sum of square error to calculate the weights, the corresponding formula is as follows,

$$Y = 0.803Y_1 + 0.197Y_2 \quad (6)$$

The results of different weights are shown in Table 3.

### Table 3. The Forecasting Results of Different Combination Model

| Date_ | Actual sales | Combination Model 1 | relative error | Combination Model 2 | relative error |
|---|---|---|---|---|---|
| JAN 2014 | 215.64 | 222.61 | 0.032 | 226.19 | 0.049 |
| FEB 2014 | 159.64 | 173.71 | 0.088 | 169.49 | 0.062 |
| MAR 2014 | 216.91 | 222.17 | 0.024 | 225.68 | 0.040 |

| | | | | | |
|---|---|---|---|---|---|
| APR 2014 | 200.42 | 202.71 | 0.011 | 203.12 | 0.013 |

By calculation, the MAPE of the combination model 1 is 3.90, and the MAPE of the combination model 2 is 4.11, the former is lower than the latter, the result indicates that relative error is prior to absolute error in the criteria of calculating weight, the MAPE is more effective.

4) Finally, by comparing the forecasting results of the combination model with the two single models, we can draw the conclusion that the mean absolute percentage error and the root mean square error of the combination model are lower than the two single models. In this paper, the MAPE and RMSE are regarded as the criteria for comparing the forecasting precision of different models,

The related formulae are expressed as followings:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{\overline{y_t} - y_t}{y_t}\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\overline{y_t} - y_t)^2}$$

Where $\overline{y_t}$ represents the predicted value of the t time period, $y_t$ represents the actual value of the t time period, n is the total number of observations.

The results are shown in Table 4.

**Table 4. The Comparison of Different Models**

| Criteria | Seasonal index | RBF network | Combination model |
|---|---|---|---|
| MAPE | 4.48 | 10.08 | 3.90 |
| RMSE | 11.51 | 23.20 | 8.36 |

From the Table 4, we can conclude that the precision of the combination model is higher than that of two models in terms of the MAPE and RMSE. Because the combination forecasting model can take into account the more factors and make use of useful information of every single model, so we obtain a more accuracy forecasting result which provide more scientific decision-making. The value of MAPE and RMSE for seasonal index is between RBF network model and combination model, so the seasonal index model has the better accuracy than RBF neural network.

## 4. Conclusion

In this paper, we summarized the achievements of previous studies on forecasting, a new combination forecasting model based on seasonal index model and RBF neural network is proposed, and a new weight calculation method based on mean absolute percentage error is also introduced. The conclusions are described as follows.

1) Two models of seasonal index model and RBF neural network model are used to predict the monthly auto sales, and the results indicate that seasonal index model is better than RBF model in terms of the mean absolute percentage error, different models can achieve different accuracy for forecasting monthly sales due to using the different information from the data.

2) The weights of two models are calculated respectively according to the mean absolute percentage error and the sum of square error. Forecasting results of two different

 

weights combination models are compared, results indicate that the weights of two single models can be appropriately assigned based on MAPE, the proposed relative error method is better than absolute error method on the calculation of weights.

3) The results of three models are compared in terms of mean absolute percentage error and root mean square error, the MAPE and RMSE values of combination model are smallest, the corresponding values of RBF neural network are largest, so the new combination model based on seasonal index and RBF neural network can improve the forecasting accuracy, We believe that the proposed combination model is highly applicable for forecasting sales in the auto industry.

## Acknowledgements

This paper is granted by Special Research Foundation of Young Teachers of Hubei University of Automotive Technology (No. 2012XQ02).The authors would like to thank the editors and the anonymous reviewers for their helpful and constructive comments.

## References

[1] Y. Yu, T. M. Choi and C. L. Hu, "An intelligent fast sales forecasting model for fashion products", Expert Systems With Applications, vol. 38, no. 6, **(2010)**, pp. 376-385.
[2] Y. R. Ni and F. Y. Fan, "A two-stage dynamic sales forecasting model for the fashion retail", Expert Systems With Applications, vol. 38, no. 3, **(2010)**, pp. 1529-1536.
[3] T. M. Choi, Y. Yu and K. F. Au, "A hybrid SARIMA wavelet transform method for sales forecasting", Decision Support Systems, vol. 51, no. 1, **(2011)**, pp. 130-140.
[4] D. J. Fang and W. B. Weng, "Sales Forecasting System for Chinese Tobacco Wholesalers", in Proc. International Conference on Environmental Sciences, China, **(2011)**, pp. 380-386.
[5] A. Sa-ngasoongsong, S. T. S. Bukkapatnam and J. Kim, "Multi-step sales forecasting in automotive industry based on structural relationship identification", International Journal of Production Economics vol. 140, no. 2, **(2012)**, pp. 875-887.
[6] B. Luo, L. Wang and W. W. Yan, "A New Type of Combination Forecasting Method Based on PLS—The Application of It in Cigarette Sales Forecasting", American Journal of Operations Research, vol. 2, no. 3, **(2012)**, pp. 408-416.
[7] V. Murlidharan and B. Menezes, "Frequent pattern mining-based sales forecasting", OPSEARCH, vol. 50, no. 4, **(2013)**, pp. 455-474.
[8] Z. X. Guo, W. K. Wong and M. Li, "A multivariate intelligent decision-making model for retail sales forecasting", Decision Support Systems, vol. 55, no. 1, **(2013)**, pp. 247-255.
[9] X. D. Yu, Z. Q. Qi and Y. M. Zhao, "Support Vector Regression for Newspaper/Magazine Sales Forecasting", Proceeding Computer Science, vol. 17, **(2013)**, pp. 1055-1062.
[10] L. Wan, B. Luo, H.M. Ji and W.W. Yan, "A type of combination forecasting method based on time series method and PLS", American Journal of Operations Research, vol. 2, no.4, **(2012)**, pp.467-472.

## Authors

**Lihua Yang**, She is a Lecturer, School of Economics and Management, Hubei University of Automotive Technology. She received her B.S degree from Yangtze University in 2003, and her M.S. degree from Southwest Jiaotong University in 2006.Her research interests include data mining and decision support.

**Baolin Li**, He is a Lecturer, School of Economics and Management, Hubei University of Automotive Technology. He received his B.S degree from Lanzhou University in 2005, and his M.S. degree from Northwestern Polytechnical University in 2012. His research interests include data analysis and data management.