

Prediction of Educational Investment in China Based on Artificial Neural Networks and Grey Model GM (1,1)

Xiao Qianyin¹ and Liu Bo²

¹Foreign Language School, Southwest Petroleum University, Chengdu Sichuan, 610500, China

²Department of Planning and Evaluation (Teacher Education and Development Center), Southwest Petroleum University, Chengdu Sichuan, 610500, China

¹conniexqy@qq.com, ¹th100th@swpu.edu.cn

Abstract

Recent years, as China is experiencing a rapid development, the educational investments are undergoing a sharp increase compared to past decades. Educational investment has become a crucial aspect that deeply influencing the future development of a country. What's more importantly, knowing the future investments of education is a crucial affair that decides the future tendency of related projects such as government management, cooperation investment and personnel training. Decision makers cannot make any effective decisions without knowing the future investments of education under all these circumstances. However, it is fairly difficult for people to obtain the predicted data of educational investments without the aid of mathematical and computational modeling. Therefore, in this article, we aim at presenting a series of solutions for the prediction of educational investment for China based on multiple linear regression (MLR), artificial neural networks (ANNs) and grey model GM (1,1). Multiple comparisons are made for deciding whether model should be used under different external conditions. Our research successfully shows that all these models are available for practical applications and scientists and other related people can choose their suitable models alternatively for the sake of making a better prediction under different circumstances.

Keywords: Educational investment; Chinese education; multiple linear regression; grey model GM (1,1); artificial neural networks

1. Introduction

China is undergoing a quick development in recent years [1-3]. To ensure a sustainable development of a country, decision makers should not neglect the development of education [4-5]. Only by ensuring a decent investment of education, can a country have a more vivid development and have more and more qualified persons. Therefore, it is fairly necessary that we should take the change regulation of educational investment into consideration. However, during the past decades, the change regulation of educational investments is not clear, although a large number of related research have been reported. And the impact factors of the educational investment are uncertain because of the complexity of the educational management. So far, there is still no relevant reports that could successfully predict the educational investment for China in a precise method.

In this paper, we aim at presenting a series novel methods for the prediction of educational investment for China. According to the complex linear or non-linear relationship of social indicators, there are some indicators that are commonly considered as the impact factors of educational investments in China. These indicators consist of a series of economical conditions and development of China, including gross domestic product, the number of students in ordinary colleges and Universities, social organizations and citizens to education, as well as social donations and fund-raising education. All these

four respects are considered as crucial indicators according to previous related research. Here, we use these four indicators as the independent variables for the development of relevant linear and non-linear models. And the educational investment is set as the dependent variable. One can only input the data of independent variables to obtain the dependent variable in the computer, using the linear and non-linear models presented by this article.

First, a multiple linear regression (MLR) model was developed in order to analyze the weights of different components. For developing a series of non-linear computational models, artificial neural network (ANNs) as a powerful non-linear machine learning techniques were established. In order to be independent to the independent variables, we also developed a grey model GM (1,1) using the number sequence of the educational investments, in the order of the time.

2. Artificial Neural Network

Artificial neural network (ANN) is an information processing system with interconnected components analogous to neurons [5-10]. The system has splendid nonlinear mapping and a good degree of parallel processing of information. On the basis of mathematical models (also computing models), it can simulate the same features and functions of the biological neural networks. That is to say, an ANN model has the ability to adapt the research objectives continuously to new data and learn from the collected experiences and noise data. Here we present the main structure of the specific ANN model we use as a conclusion in this article.

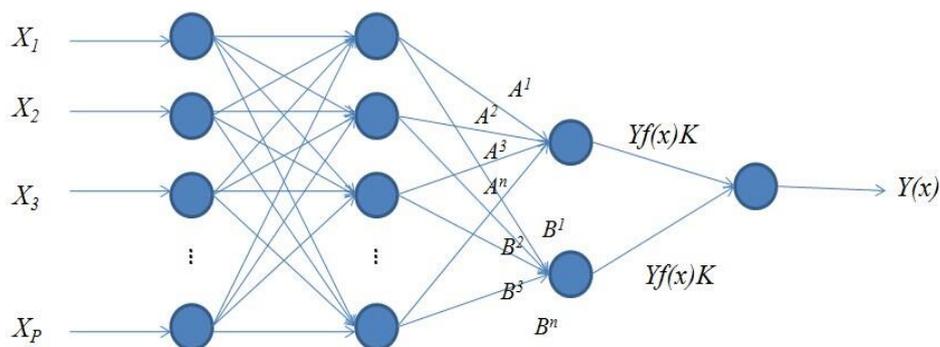


Figure 1. Structure of General Regression Neural Network

Figure 1 shows the main structure of the general regression neural network (GRNN) [11-13]. Our research shows that the GRNN models can robustly generate the precise results for the evaluation of education. An ANN model consists of three parts: the input layer, the output layer and the hidden layer. The input layer is made up of thousand neurons to accept numerous nonlinear input variables [10]. After a series of computing processes, the output variables can be exported from the output layer. Additionally, the hidden layer is included. It is a layer existing between the input and output layer. The more the number of the neurons in the hidden layer has, the greater robustness of the artificial neural network will be. We can input the independent variables and dependent variables for the evaluation, if the data scale is enough, ANN models will generate correct responses and start the process of machine learning, which is what we call "model development". After the model is established, we can input new independent variables to the model and it soon will calculate the dependent variables and present to users. In our research, the dependent variables are the levels of the teachers and students. To ensure the robustness and correctness of the ANN models we established, we develop the models for teachers and students respectively. The models have been trained over 100 times in order to confirm their robustness.

3. Grey Model GM (1,1)

Grey Model GM (1,1) [14-17] is an excellent mathematical approach that is mainly used for predicting in uncertain systems. Original data can be changed into new forms for the sake of finding out regulations of the change of the system, which is mainly determined by a series of iteration. Prediction based on GM (1,1) is thoroughly independent to the independent variables. GM (1,1) is widely used for the prediction of the values which change with a certain gap of time.

In our research, because the educational investments of China is planned by each year, and the data is enough for the modeling of GM (1,1) model. Therefore, GM (1,1) model is what we consider as a suitable and time-saving model for the prediction.

4. Results and Discussion

In this section, the MLR, ANN models the GM (1,1) model are presented for the prediction of educational investments respectively, based on the data provided by reference [18]. Results and establishing processes of all the three kinds of models are presented in the separated sections.

4.1. Development of the MLR Model

Here, we use SPSS to develop an equation for the change regulation of educational investments in China. The 4 indicators were presented by x_1 to x_4 , whereas y is the educational investments. The multiple linear regression equation (Equation (1)) is shown as follows

$$y = 0.047x_1 + 3.781x_2 - 7.257x_3 - 2.914x_4 - 1325.869 \quad (1)$$

Equation (1) is the result of MLR modeling, based on the existing data of educational investments as well as the four indicators, which shows that x_2 , x_3 , x_4 , have significant impacts to the educational investment of China (which presented as y in the equation). Figure 2 illustrates the comparison of actual values versus the predicted values using the multiple linear regression method:

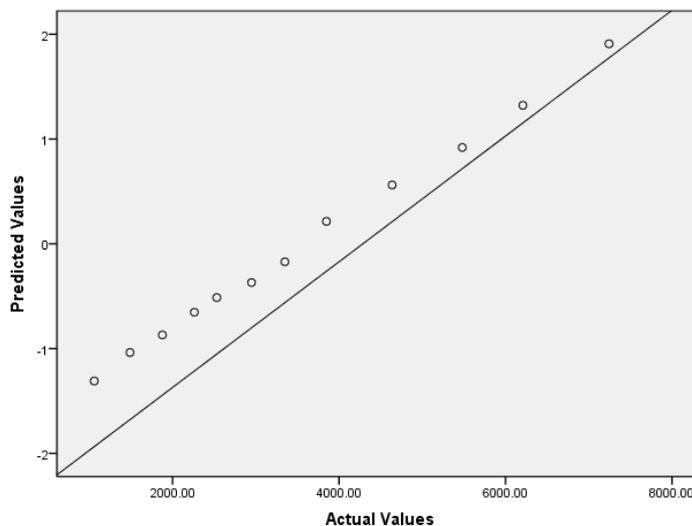


Figure 2. Actual Vales Versus Predicted Values via the MLR Model

Figure 2 illustrates that the MLR model has a decent result during the prediction. However, we still consider that the results are not precise enough because the linear

function cannot directly fit the relationship between the educational investments and the four indicators. Therefore, we decided to use non-linear functions to fit the data. Artificial neural networks is a powerful machine learning techniques that can fit the data into a non-linear form. In Section 4.2, we present the detailed results of the training and testing processes of relevant ANN models.

4.2. Development of ANN Models

For the sake of developing a series of ANN models, we used the same indicators and dependent variable. Best net search method was used for finding a suitable ANN model. We used the general regression neural network (GRNN) [19-23] and multi-layer feed-forward neural network (MLFN) [24-27] to develop the models and undertake the best net search comparison (see Table 1). The number of nodes of MLFN models were set from 2 to 25 in order to find out the most suitable nodes of MLFN model for the prediction.

Table 1 shows the best net search results of ANN models for the prediction of the emission of HC in the first system:

Table 1. Best Net Search Results of Models for the Prediction of Private Vehicle Ownership in Chinese Area

ANN Model	Trained Samples	Tested Samples	RMS Error	Training Time
GRNN	8	4	254.56	0:00:00
MLFN with 2 Nodes	8	4	462.55	0:00:34
MLFN with 3 Nodes	8	4	1121.79	0:01:05
MLFN with 4 Nodes	8	4	1242.45	0:01:42
MLFN with 5 Nodes	8	4	2887.77	0:04:54
MLFN with 6 Nodes	8	5	48437.35	0:26:16
MLFN with 7 Nodes	8	4	438.72	0:34:28
MLFN with 8 Nodes	8	4	7116.37	0:00:31
MLFN with 9 Nodes	8	4	9775.28	0:00:33
MLFN with 10 Nodes	8	4	1390.46	0:00:38
MLFN with 11 Nodes	8	4	4370.56	0:00:39
MLFN with 12 Nodes	8	4	3259.50	0:00:47
MLFN with 13 Nodes	8	4	6869.84	0:00:46
MLFN with 14 Nodes	8	4	4927.99	0:00:48
MLFN with 15 Nodes	8	4	5121.20	0:00:53
MLFN with 16 Nodes	8	4	7709.36	0:00:58
MLFN with 17 Nodes	8	4	17625.61	0:01:03
MLFN with 18 Nodes	8	4	6990.54	0:01:06
MLFN with 19 Nodes	8	4	3216.10	0:01:04
MLFN with 20 Nodes	8	4	11387.71	0:01:05
MLFN with 21 Nodes	8	4	12657.03	0:01:12
MLFN with 22 Nodes	8	4	2720.15	0:01:18
MLFN with 23 Nodes	8	4	10426.98	0:01:23
MLFN with 24 Nodes	8	4	7360.01	0:01:29
MLFN with 25 Nodes	8	4	12855.64	0:01:38

Table 1 indicates that the GRNN is the best model for prediction, with an lowest RMS error (254.56) in the result list. Figure 3 is illustrating the change regulation of the RMS errors of MLFN models in accordance with the change of nodes:

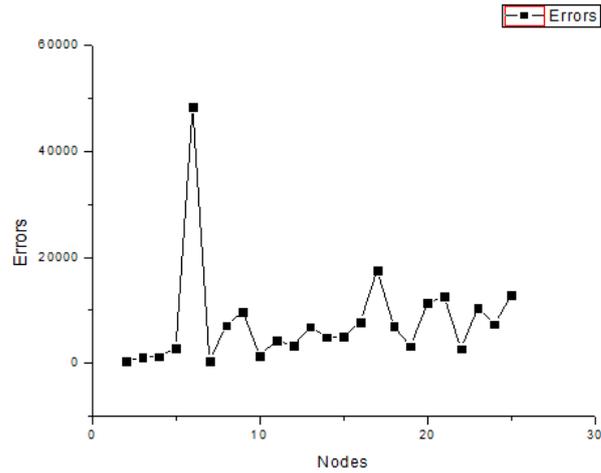


Figure 3. RMS Errors versus Nodes of MLFN Models

Figure 3 shows that there is a significant fluctuation of RMS errors when changing the nodes of MLFN model. The fluctuation presented by Figure 3 indicates that the MLFN model is not as suitable as the GRNN model. And the overall RMS errors of MLFN models during our experiments are not as low as that of GRNN model. Therefore, we strongly believe that the GRNN model is the best model during the process of best net search. In addition, the results below presents the training and testing process of the GRNN model only because the GRNN model is believed to be the only suitable ANN model during our detailed repeated experiments.

Here, we present the results of the GRNN model in Figure 4 to Figure 9:

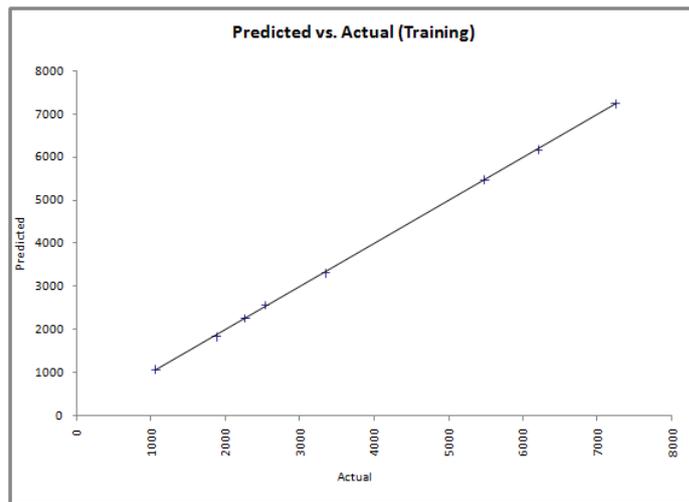


Figure 4. Predicted Values versus Actual Values after Training in GRNN

Figure 4 shows that the training process is extremely robust. The regression line presents a regulation of proportional function. Hence, this result show that the training process is correct and precise. It also confirms that Figures 5 and 6 is reliable.

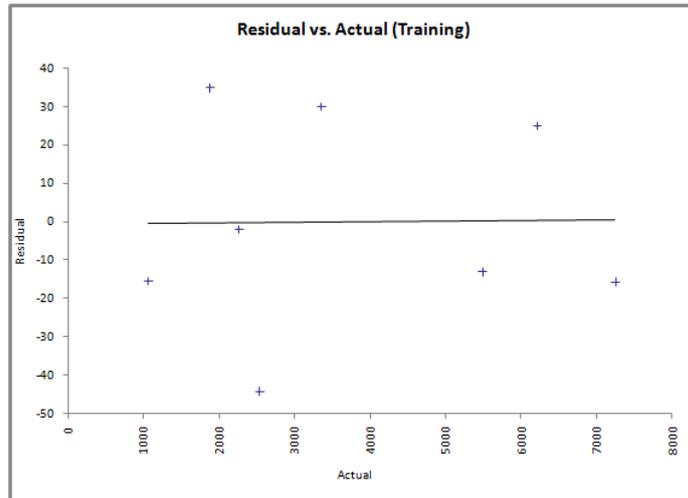


Figure 5. Residual Values versus Actual Values after Training in GRNN

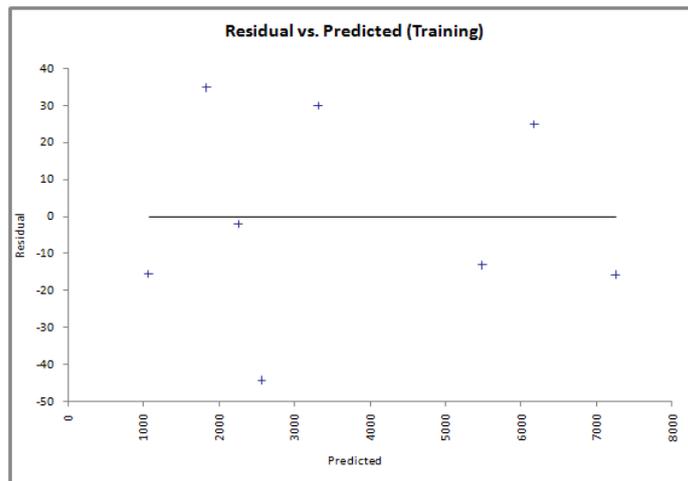


Figure 6. Residual Values versus Predicted Values after Training in GRNN

Figures 5 and 6 show that the residual values are comparatively low. The training process is fairly robust, which shows that the testing results presented by the following three figures are reliable.

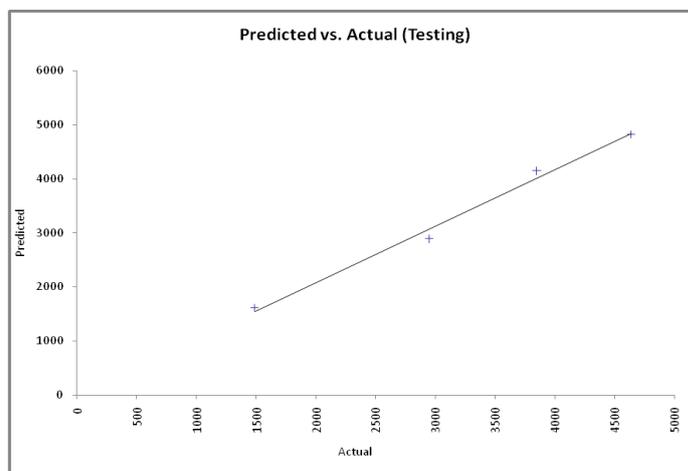


Figure 7. Predicted Values versus Actual Values after Testing in GRNN

Figure 7 shows that the training process is ideal. The regression line perfectly presents a regulation of proportional function. This result indicates that the testing process is correct and precise. It also suggests that following results presented by Figures 8 and 9 are reliable.

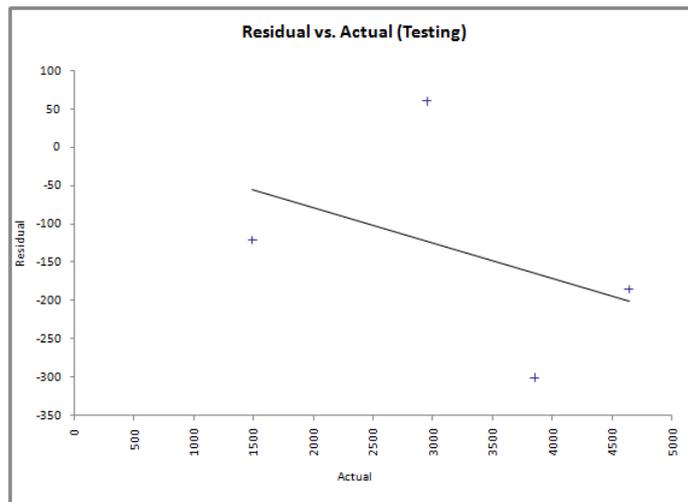


Figure 8. Residual Values Versus Actual Values after Testing in GRNN

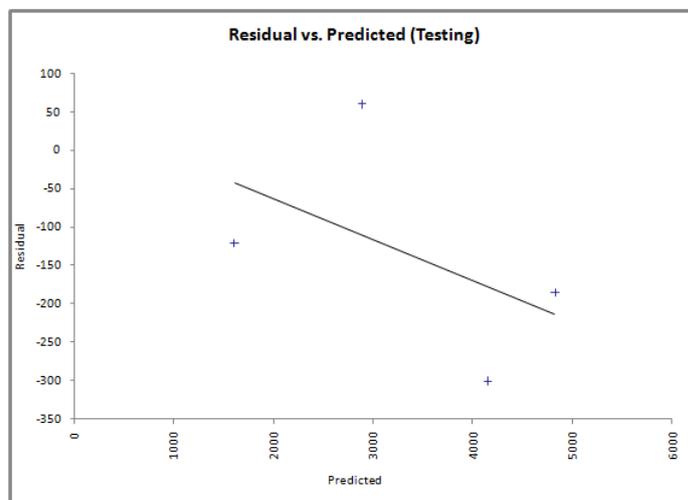


Figure 9. Predicted Values versus Actual Values after Testing in GRNN

Figures 8 and 9 illustrate that the residual values are comparatively low. The testing process is fairly robust, which suggests that the GRNN model is robust during the prediction of educational investments of China, under the circumstance of the data set we used for the training and testing.

4.3. Development of Grey Model GM (1,1)

A GM (1,1) model based on the sequence of educational investments was established using relevant software. 8 data was used for the model development, whereas 4 data was used for testing. Testing results are presented by Figure 10.

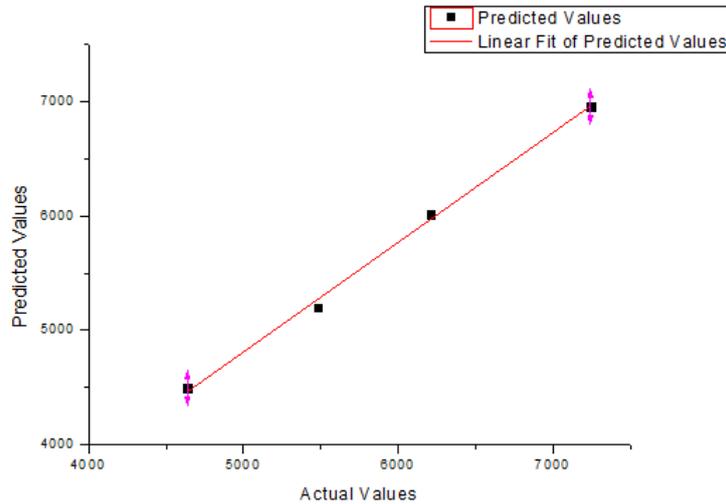


Figure 10. Testing Result of Grey Model GM (1,1) (Actual Values versus Predicted Values)

Figure 10 presents the testing results of GM (1,1) model for the prediction of educational investment in China. Results show that the testing situation of GM (1,1) model is nearly as good as that of GRNN model, with a very suitable result. One of the advantages of GM (1,1) model is time-saving, and it can be used for predicting more data in the future years. However it can only be used for the prediction to the following future years because GM (1,1) model is only perfectly suits for a exponential prediction. If the educational investment is not corresponding to the exponential changes, in the long run, the prediction result may fail to correspond to actual results.

To sum up, it can be clearly seen that all the three models (MLR, ANN and GM (1,1)) can be used for the prediction of the educational investments in China. However, the choice of the certain model in practical applications should be up to a series of constraints. MLR model is not as precise as GRNN and GM (1,1) models. However, using Equation (1) is very convenient for users, which is one of the advantages of the MLR model. ANN model has the advantage of robustness and correctness to the prediction because of the powerful and reasonable non-linear fitting. However, it has a risk of over-fitting. What's more importantly, the best net search requires time, which is the obvious advantages. Grey model GM(1,1) is quick and precise during our experiments. However, it can only be used for a "short" prediction. If users want to predict a long list of data, the GM (1,1) model may not be quite suitable.

5. Conclusion

As China is undergoing a quick development, educational investments are experiencing a rapid increase. Educational investment has become an important respect that thoroughly impacting the sustainable development of a country. Knowing the further investments of education is a crucial thing that determines the future trend of relevant projects like government management, cooperation investment and personnel training. Decision makers cannot make any effective decisions without knowing the future investments of education under all these circumstances. Hence, in this paper, we aim at reporting a series of solutions for the prediction of educational investment in China based on multiple linear regression (MLR), artificial neural networks (ANNs) and grey model GM (1,1). Multiple comparisons are made for deciding whether model should be used under different external conditions. Our research successfully shows that all these models are available for practical applications and scientists and other related people can choose their suitable

models alternatively for the sake of making a better prediction under different circumstances.

Acknowledgments

We appreciate Foreign Language School and Teacher Education & Development Center of Southwest Petroleum University for the great supporting.

This paper is supported a project approved by Provincial Key Research Base of Philosophy & Social Science and the Research Base of Humanities & Social Sciences of Sichuan Province, affiliate to Educational and Developmental Center of Sichuan Province in China West Normal University. The project is approved in 2014, titled as “Research on the Formulation and Evaluation Method of the Teacher-Quality Assessment System in Technological Universities” (Principle Investigator: Liu Bo, Grant No. CJF14039).

This paper is also supported by Higher Education Quality and Teaching Reform Project of Sichuan Province, 2013-2016. (No.:197) “Study on Young Teachers’ Training and Development in Provincial Governed Universities”.(SWPU Project No.:X15021301003). Project No.: Letter Issued by Sichuan Provincial Education Department (2013)781.

References

- [1] Y. L. and P. L., Economic Research Journal, vol. 3, (2003).
- [2] J. Monson, The China Quarterly, vol. 200, (2009).
- [3] G. Montinola, Y. Qian and B. R. Weingast, “Federalism”, World Politics, vol. 48, no. 1, (1995).
- [4] B. Jickling, The Journal of Environmental Education, vol. 23, no. 4, (1992).
- [5] T. Gylfason, European Economic Review, vol. 45, no. 4, (2001).
- [6] B. Samanta, K. R. A. Balushi, Mech. Syst. Signal Pr., vol. 17, no. 2, (2003).
- [7] W. G. Baxt, Ann. Int. Med., vol. 115, no. 11, (1991).
- [8] M. Shao, X. J. Zhu and H. F. Cao, Energy, vol. 67, (2014).
- [9] C. H. Aladag, A. Kayabasi and C. Gokceoglu, Neural Comput. App., vol. 23, no. 2, (2013).
- [10] J. Van Schependom, G. Nagels and W. Yu, Schizophrenia Bulletin, (2013).
- [11] E. C. Santos, E. D. Armas and D. Crowley, Soil Bio. Bioche., vol. 69, (2014).
- [12] H. Yip, H. Fan and Y. Chiang, Auto. Const., vol. 38, (2014).
- [13] C. M. Hong, F. S. Cheng and C. H. Chen, Int. J. Elect. Power & Energy Syst., vol. 60, (2014).
- [14] M. Mao and E. C. Chirwa, “Technological Forecasting and Social Change”, vol. 73, no. 5, (2006).
- [15] L. Dang, L. Sifeng and D. Yaoguo, Eng. Sci., vol. 8, (2003).
- [16] G. J. Tan, Syst. Engineering-Theory & Practice, vol. 4, no. 18, (2000).
- [17] N. Xie and S. Liu, Syst. Eng.-Theory & Practice, vol. 1, no. 14, (2005).
- [18] Y. Wang and M. Wu, Times Finance, vol. 372, no. 3, (2008).
- [19] D. F. Specht, Neural Networks, IEEE Transactions, vol. 2, no. 6, (1991).
- [20] D. Tomandl and A. Schober, Neural Networks, vol. 14, no. 8, (2001).
- [21] H. M. Kandirmaz, K. Kaba and M. Avci, International Journal of Photoenergy, (2014).
- [22] H. Li, W. Leng, Y. Zhou, F. Chen, Z. Xiu and D. Yang, The Scientific World Journal, (2014), pp. 478-569.
- [23] T. Masters and W. Land, Computational Cybernetics and Simulation, IEEE, vol. 3, (1997).
- [24] D. Svozil, V. Kvasnicka and J. Pospichal, Chemometr. Intell. Lab., vol. 39, no. 1, (1997).
- [25] A. Y. Shamseldin, A. E. Nasr and K. M. O'Connor, “Hydrology and Earth System Sciences Discussions”, vol. 6, no. 4, (2002).
- [26] D. Yang, H. Li, C. Cao, F. Chen, Y. Zhou and Z. Xiu, Journal of Spectroscopy, (2014).
- [27] J. R. M. Smits, W. J. Melssen and L. M. C. Buydens, Chemometr. Intell. Lab., vol. 22, no. 2, (1994).

Authors



Xiao Qianyin, female, 1981-11, Pengxi County, Sichuan Province, Lecturer of Southwest Petroleum University, Master’s degree in Foreign Linguistics and Applied Linguistics. Southwest Petroleum University of Foreign Language School, Chengdu Sichuan, 610500, China. Research Direction: Teacher Education and Development. Email: conniexqy@qq.com



Liu Bo, male, 1981-09, Huili County, Sichuan Province, Lecturer of Southwest Petroleum University, Master's degree in engineering and laws. Office Director of the Planning and Evaluation Department. Southwest Petroleum University Department of Planning and Evaluation (Teacher Education and Development Center), Chengdu Sichuan, 610500, China .Research Direction: Education of Ideology and Politics for College Students and Higher Education Management. Email: th100th@swpu.edu.cn