Research Article

Broad Learning-Based Optimization and Prediction of Questionnaire Survey: Application to Mind Status of College Students

Lin Yu ¹ and Shejiao Ding²

¹School of Marxism, Northwestern Polytechnical University, China

²School of Humanities, Economics and Law, Northwestern Polytechnical University, China

Correspondence should be addressed to Lin Yu; yulinsodme@163.com

Received 9 July 2018; Accepted 12 September 2018; Published 9 October 2018

Guest Editor: Jing Na

Copyright © 2018 Lin Yu and Shejiao Ding. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mind status of college students is important since it can reflect how the public opinion is going. Only with the accurate prediction, the corresponding actions can be conducted to prevent the situation from going worse. This paper focused on the data analysis using the recent developed broad learning method to obtain the learning model and then the prediction can be done. Firstly, the questionnaire related to the ideological state is designed. Secondly, the data are collected and classified using the typical questions and answers. Thirdly, for each pair of the question and the answer, the score is obtained and considered as data training of the system. Fourthly, the input and the output are selected according to the key questions and conclusions. Finally, the broad learning using flat network is employed for data analysis without deep structure. Tests show that the design using broad learning can efficiently deal with the regression problem and the learning network can be used for prediction.

1. Introduction

Social phenomenon is complex and challenging since the potential dynamics cannot be known exactly or cannot be obtained. For example, it is noted that civic education is a basic education subject that all countries around the world attach high importance to. Civic education in the UK mainly includes moral education, political education, and religious education, while in the USA, it is implied in patriotic education and moral education. In Japan, civic education is called “social science,” while it is called “life education” and “state consciousness education” in Singapore, and called “ideological and political education” in China. Due to the difference of the countries in state nature, political system, cultural tradition, education system, etc., their civic education differs quite a lot in form, content, and method. However, the purpose of civil education in whatever culture or system is the same: the ruling

class tries to have purposeful and planned impact on the citizens by applying a certain political ideas and moral norms, so that the citizens form social moral practice in line with the needs of the ruling class. From the implementation effect of the education of all countries, we can see that the civic education of each country presents specific goals according to the national needs and by integrating the characteristics of the times. Therefore, “political, ideological, and advancing with times” are the characteristics of the civic education of each country.

Due to the significance of civic education, it is important to know how the status is. Furthermore, with analysis, it is key point to find some relation between factors and the results, while in this way, some actions can be taken before the consequence comes. Using questionnaire to obtain the quality of the status is a useful way. The questionnaire should be designed with specific aim so that the results can reflect a certain intention. However,

the data is not easy to analyze. If only the mean value and the ratio are calculated, only the status of interviewers can be reflected while the deep information cannot be digged out for test and the design cannot be used for prediction. For comprehensive analysis, the goal for questionnaire survey is not restricted to the statistics. The possible target is to model the public opinion while the prediction is available and then the possible action can be taken. With the abovementioned consideration, the study using machine learning is of great interest. In literature, many methods have been analyzed using the bioinspired technique such as genetic algorithm (GA) [1], neural networks (NNs) [2, 3], fuzzy logic system (FLS) [4], particle swarm optimization (PSO) [5, 6], extreme learning machine (ELM) [7–9], deep learning [10], and broad learning [11]. The methods are mimicking the behavior of biology. For example, the GA is using initialization, selection, and variants. ELM is using the feedforward NN, and the basic idea is using the way of random assignment for generalization. Deep learning is using the learning data representations. Recently, the broad learning is gaining attention. The method is using random vector functional link NN, and the key point is to expand the network by randomly adding the functions instead of increasing the depth. In [12], the feature nodes are replaced with Takagi-Sugeno fuzzy system and the fuzzy broad learning system is proposed.

For data analysis and system control, NN is widely studied in [13, 14]. For data analysis, in [15], the complex sale forecasting problem is studied due to the requirement of reliable prediction where the fuzzy NN is used while the initial weights are generated by genetic algorithm. In [16], the fuzzy NN learning is proposed while the continuous genetic algorithm is used to improve the performance. One case of global supply chain management is studied with data collection, construction of learning method, and decision integration model. In [17], the study is on importance-performance analysis while the relationship is nonlinear and multicollinearity where the backpropagation NN is used to train the model. In [18], similar work has been done while the case study is with gathering data of customer perception about focal delivered service, fuzzy set design, and attributes' implicity. In [19], the market segmentation is studied using self-organizing feature maps NN and genetic K-means algorithm. In the test, the procedure includes questionnaire design, importance analysis, satisfaction analysis, performance evaluation of the satisfaction, and the discussion. In [20], the restaurant service recovery is studied to see how much the consumer will return with the recovery plan. In the design, support vector machine and multilayer perceptions are used to predict the consumer expectations. In [21], the learner's preference for the visual complexity on small screens of mobile computers is studied by NN. Using RBF-NN model, the accuracy can be guaranteed while the investigation efforts can be avoided. In [22], the fruit fly optimization algorithm is adopted to optimize the NN model while the principal component regression is used for the questionnaire survey.

From the abovementioned discussion, the questionnaire survey is widely used in many applications since it includes the principal component while the data indicates the potential nonlinear model. Using questionnaire survey is an effective way to analyze the status of mind. The challenge is on how to design the questionnaire survey and how to analyze the data. Many methods are adopted for data analysis and optimization where the basic idea is using NN while some other evolutionary algorithms are used to make the NN more efficient. However, the structure and the parameters should be selected step by step. Deep learning is widely studied recently, and the great breakthrough has been achieved. But the design is time-consuming. Broad learning is interesting with randomly adding new functions while updating the results using pseudoinverse method. In this way, though the flat work might grow fast, the computation is not increasing so fast.

In this paper, the work is on questionnaire survey-based data analysis and prediction. The mind status of college students is collected while the inputs and the output for the training network are selected. Furthermore, the BL algorithm is employed for optimization. The obtained model is verified via test data. The organization of the paper is given as follows. In Section 2, the questionnaire survey is discussed. In Section 3, the broad learning is described and the data optimization is given. In Section 4, the conclusions and future works are discussed.

2. Questionnaire Survey and Analysis

This survey takes some college and university students in Shaanxi Province as the objects determine the total sample size required for the survey according to the conservative formula for calculating the sample size and the average expected sample size and samples randomly by the level. A total of 1236 questionnaires are distributed, 1202 effective questionnaires are recovered, and the effective recovery rate is 97%. The questionnaires do a comprehensive survey to the ideological and political quality of the college and university students, including their outlook in life and values, dynamic focus on social events, and degree of ideological and political identity. Statistical analysis is conducted after the questionnaire data is input into the database through the SPSS software.

The three dimensions of the ideological and political quality of the college and university students are interconnected and complementary. Therefore, when evaluating the ideological and political quality level of the college and university students, we need to take the three dimensions as the correlation coefficients with the same importance (the three dimensions, respectively, account for 33.3%), take 2.5 scores (totally 5 scores) as the total score for the evaluation standard through the expert consultation method according to the confidence interval and actual needs, work out the general mean distribution map, and determine poor ideological and political quality index (2.5–3) of the college and university students, quite poor ideological and political quality index (3–3.5), average ideological and political quality index (3.5–4), quite

high ideological and political quality index (4–4.5), and high ideological and political quality index (4.5–5), so as to judge the overall ideological and political quality level of the college and university students, wherein the ideological and political quality level of 64.8% college and university students is [4, 5]. Therefore, we can see that the current ideological and political quality level of the college and university students is quite high and at the middle and upper levels.

In the analysis, it is interesting to see that college students are more interested in career-related topics such as “New joint cooperation with Jingdong Company” (ratio 30.12%) “Accommodation Condition Improvement with Air-conditioner” (ratio 49.33%) while high technology-related news which can exhibit the university’s reputation is gaining attention such as “The Unmanned Aerial Vehicle Show in National Day” (ratio 45.01%). Oppositely, they pay little attention to the speech on politics (ratio 15.47%) or the results of Outstanding Young Investigator Award of National Natural Science Foundation (ratio 21.13%).

3. Data Representation and Broad Learning

In the questions, 34 items are included for the data regression. In the question, different answers are labeled as “1–5.” Furthermore, the gender is considered while the grade is used as input. Male is labeled as “0” while female is labeled as “1.” For different grades, different number “1–6” is used as a label. In this way, the input dimension is 36. The status is considered as regression output, while in the questionnaire survey, the option is 4 levels between “positive” and “negative.” Considering the diversity, 404 samples are used for data analysis and learning.

For the broad learning, it is proposed in [11], the main idea is using pseudoinverse to update the learning weight since the new weights W_{n+1} are obtained from the previous weights W_n using new enhancement nodes with the following equation:

$$W_{n+1} = \begin{bmatrix} W_n - db^T Y_n \\ b^T Y_n \end{bmatrix}. \quad (1)$$

Furthermore, the sparse features can be included to solve the optimization problem.

The learning scheme is given as follows. Consider that the training sample is $X_{304 \times 36}$ while the output is $Y_{304 \times 1}$. For $j=1, j \leq m$, the parameters are randomly generated as W_{h_j} and β_{h_j} . Furthermore, the new function is calculated as follows:

$$H_j = \xi(XW_{h_j} + \beta_{h_j}). \quad (2)$$

The enhancement node group is written as $H^m = [H_1, H_2, H_3, \dots, H_m]$. The critical step is to calculate the pseudoinverse as follows:

$$A^m = [X|H^m]. \quad (3)$$

Furthermore, the weights are calculated as follows:

$$W^m = (A^m)^+ Y. \quad (4)$$

If the training error threshold is not satisfied, the new generation of $W_{h_{j+1}}$, $\beta_{h_{j+1}}$, H_{m+1} , and A^{m+1} will be randomly given. The process will continue until the training error threshold is satisfied.

Remark 1. The idea is interesting since the adding of nodes is easy while the calculation will not take up too much time. In this way, the broad learning can be applied in an efficient way.

In the test, the performance using broad learning is given. Two kinds of learning are shown as training accuracy and testing accuracy. For the all 404 samples, if “ P ” samples are used for test, the rest “ $404 - P$ ” samples will be employed for model training.

In Figures 1–3, the cases using 20 test samples are studied. The results show that with the nodes adding, the training accuracy is increasing while the testing accuracy is changing. But in all, the trend of the accuracy is good and the difference is due to the characteristics of randomness of the algorithm.

In Figures 4–6, the cases using 30 test samples are studied. The training accuracy increases with nodes adding while the testing accuracy is changing. This is dependent on the quality of the test samples. It can be concluded that the quality is important for data analysis while the broad learning can efficiently fulfill the requirement of regression.

In Figures 7–9, the cases using 40 test samples are studied. The training accuracy increases with nodes adding while the testing accuracy is stable.

Through the 3 sets of tests, it can be known that the broad learning-based design can efficiently build the relation between the input signals and the output information, though the model is nonlinear and unknown. During the process, the random vector functional link neural network is generated using the function and the network can grow with more features while the computation efficiency remains very high. For the prediction, due to the randomness of the samples, the accuracy is also highly dependent on the data quality.

4. Conclusions and Discussions

In this paper, the broad learning is given as data analysis and optimization. The questionnaire survey data is firstly analyzed and then the label is given for the inputs and the output. Using broad learning, the design shows the performance of the learning algorithm. It is also shown that data quality is important for accuracy.

For future work, the pruning of the nodes is interesting. Also due to the randomness, the performance is not so stable. Maybe some statistics can apply on the randomness, and the results can be more robust. More complex questionnaire

FIGURE 1: Performance using 20 test samples—case 1.

FIGURE 2: Performance using 20 test samples—case 2.

FIGURE 3: Performance using 20 test samples—case 3.

FIGURE 4: Performance using 30 test samples—case 1.

FIGURE 5: Performance using 30 test samples—case 2.

FIGURE 6: Performance using 30 test samples—case 3.

FIGURE 7: Performance using 40 test samples—case 1.

FIGURE 8: Performance using 40 test samples—case 2.

FIGURE 9: Performance using 40 test samples—case 3.

can be designed to collect more information, and the learning scheme is expected to obtain some critical features for mental test.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognition*, vol. 33, no. 9, pp. 1455–1465, 2004.
- [2] S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1994.
- [3] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568–576, 1991.
- [4] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," in *MVL '76 Proceedings of the sixth international symposium on Multiple-valued logic*, pp. 196–202, Logan, Utah, USA, May 1976.
- [5] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, pp. 320–324, Washington, DC, USA, July 1999.
- [6] R. C. Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, vol. 1, pp. 81–86, Seoul, South Korea, May 2001.
- [7] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [8] H. J. Rong, Y. S. Ong, A. H. Tan, and Z. Zhu, "A fast pruned-extreme learning machine for classification problem," *Neuro-computing*, vol. 72, no. 1–3, pp. 359–366, 2008.
- [9] H. J. Rong, G. B. Huang, N. Sundararajan, and P. Saratchandran, "Online sequential fuzzy extreme learning machine for function approximation and classification problems," *IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics)*, vol. 39, no. 4, pp. 1067–1072, 2009.
- [10] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [11] C. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep architecture," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 1, pp. 10–24, 2018.
- [12] S. Feng and C. P. Chen, "Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification," *IEEE Transactions on Cybernetics*, pp. 1–11, 2018.
- [13] H. Jiang and Y. Dong, "A novel model based on square root elastic net and artificial neural network for forecasting global solar radiation," *Complexity*, vol. 2018, Article ID 8135193, 19 pages, 2018.
- [14] X. Lv, X. Li, J. Cao, and P. Duan, "Exponential synchronization of neural networks via feedback control in complex

- environment,” *Complexity*, vol. 2018, Article ID 4352714, 13 pages, 2018.
- [15] R. Kuo, “A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm,” *European Journal of Operational Research*, vol. 129, no. 3, pp. 496–517, 2001.
 - [16] R. Kuo, S. Hong, Y. Lin, and Y. Huang, “Continuous genetic algorithm-based fuzzy neural network for learning fuzzy if-then rules,” *Neurocomputing*, vol. 71, no. 13–15, pp. 2893–2907, 2008.
 - [17] W.-J. Deng, W.-C. Chen, and W. Pei, “Back-propagation neural network based importance–performance analysis for determining critical service attributes,” *Expert Systems with Applications*, vol. 34, no. 2, pp. 1115–1125, 2008.
 - [18] W.-J. Deng and W. Pei, “Fuzzy neural based importance–performance analysis for determining critical service attributes,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3774–3784, 2009.
 - [19] R. Kuo, Y. An, H. Wang, and W. Chung, “Integration of self-organizing feature maps neural network and genetic k-means algorithm for market segmentation,” *Expert Systems with Applications*, vol. 30, no. 2, pp. 313–324, 2006.
 - [20] H.-C. Huang, “Using artificial neural networks to predict restaurant industry service recovery,” *International Journal of Advancements in Computing Technology*, vol. 4, no. 10, pp. 315–321, 2012.
 - [21] L.-T. Wang and K.-C. Lee, “The study of learners’ preference for visual complexity on small screens of mobile computers using neural networks,” *Turkish Online Journal of Educational Technology*, vol. 13, no. 2, pp. 168–180, 2014.
 - [22] S.-M. Lin, “Analysis of service satisfaction in web auction logistics service using a combination of fruit fly optimization algorithm and general regression neural network,” *Neural Computing and Applications*, vol. 22, no. 3–4, pp. 783–791, 2013.