PSO-BP Neural Network Grade Prediction Model Based on Bagging Ensemble Learning

To cite this article: Hongyi Li et al 2018 J. Phys.: Conf. Ser. 1069 012103

You may also like

- Effect of Bagging on Fruit Quality of Three Pummelo (Citrus grandis Osbeck) Cultivars
  Yan Wang, Xiaoke Fu, Wen He et al.

- Handling imbalance data in churn prediction using combined SMOTE and RUS with bagging method
  Eka Pura Hartati, Adiwijaya and Moch Arif Bijaksana

- An Ensemble Learning Imbalanced Data Classification Method Based on Sample Combination Optimization
  Yuxin Wang
PSO-BP Neural Network Grade Prediction Model Based on Bagging Ensemble Learning

Hongyi Li¹, Xinhang Li² and Di Zhao¹,*
¹LMIB, School of Mathematics and Systems Science, Beihang University, Beijing 100191, China
²School of Software, Beihang University, Beijing 100191, China
*Corresponding author (Email: zdhyl2010@163.com)

Abstract. Performance prediction is a very basic but important part of the application of educational big data, and it plays an important role in education management. For performance prediction, this paper innovatively integrates Bagging ensemble learning, PSO particle swarm optimization and neural network algorithm, proposes a Bagging-based PSO-BP neural network model, and uses real data to verify the model data. In the low-dimensional performance prediction problem has got very good results.

1. Introduction
It is widely accepted that the development of big data benefits many fields from providing people with more convenient ways to solve problems. There are more and more areas where big data is used for predictive analysis. However, as the cornerstone of social development, the application of big data in education is scarce [1].

In college education, student’s grade is a very important indicator. The average dropout rate for undergraduates in American universities is as high as 40%. Except for a small number of students with excellent grades who actively withdraw from school, most of them are unable to continue their studies due to unsatisfactory grades. Analysis of personal information, learning behaviors, and past performance could help future performance predictions and academic warnings for students who may have a crisis [2]. Traditional academic warning tends to focus only on the total score of a semester, but ignores many obscure phenomena. It has a large lag in the actual situation and does not play a good role in early warning. With big data analysis, students' behaviors and past performance trends can be evaluated in real time [3] to determine whether there will be declines in their performance in the future. Under the conditions where the academic performance will seriously affect the social status of a career, using big data to deliver early warning could be an efficient way of educational management.

At present, the analysis of performance management still remains in the microscopical stage [4]. It is difficult to predict and analyze individual performance. The algorithms used for individual grade prediction mainly include decision tree [5], the Bayesian network, neural network [6], support vector machine, etc. However, the improved algorithms based on these machine learning algorithms failed to perform well in grade prediction. Junyu Long [7] proposed a grade prediction model based on mean clustering and decision tree. Although the C4.5 decision tree algorithm can achieve ideal results in model training speed and classification results, there are too many factors affecting the performance. It is easy to lead to some problems such as too big tree and too many branches. At the same time, the C4.5 algorithm can only deal with discrete problems and cannot convert discrete results into continuous results, which is not intuitive and accurate. Jianming Huang [8] proposed to use Bayesian networks for grade prediction. Bayesian network is one of the most effective theoretical models for
uncertain knowledge and inference in fitting with probability. Bayesian networks can be used for correlation analysis and accurately obtain the implicit connection between results. But the Bayesian network is still an NP-complete problem, and the model training is very complicated. Yi Zhang [9] described the application of neural network in big data. Neural network has higher requirements for data. It requires a large amount of data training to obtain good results and could be easily trapped in local optima. The grade prediction algorithm based on support vector machines proposed by Li Zhang [10] is difficult to put into practical use. Support vector machines perform well when there is a small amount of data, but it is too difficult to construct the kernel function by pure mathematical theory when the amount of data increases. So it has no satisfying effect that can be used in practice.

This paper proposes an optimized neural network algorithm based on BP neural network. It makes a contribution to existing algorithms. The innovations of this paper are mainly in the following aspects.

- Integrate Bagging operation in integration learning, PSO particle swarm optimization algorithm, and BP neural network algorithm to optimize the neural network continuously.
- It avoids the problem that existing grade prediction algorithms can only predict discrete values and cannot predict continuous values for accurate prediction.
- The generalization of the model is enhanced, and the accuracy of data fitting is significantly improved.

2. PSO-BP Neural Network Based on Bagging Ensemble Learning

The main steps of PSO-BP neural network algorithm based on Bagging ensemble learning are as follows: Each time a set of samples are randomly selected from the training sample set to form a new dataset. PSO-BP neural network uses this dataset to train. Weight and bias adjustment in the process of training uses the PSO algorithm to conduct a iterative search. The end of the iteration is to get the approximate optimal solution and then the solution is used by the neural network for secondary training to get the best fitting results. Get several neural network classifiers using the same method and the final classifier is selected by voting.

Using Bagging and PSO to optimize BP neural network can significantly improve the generalization ability and training speed of the model to obtain higher accuracy. The specific process of PSO-BP neural network based on Bagging integrated learning is as the figure shows.
Bagging algorithm creates multiple learners. Create a new dataset as large as the original dataset by randomly selecting elements from the original dataset. Process the data to ensure that only the corresponding performance data is retained. At the same time, the set is divided into two parts, one for training and the other for testing.

Build a neural network. In order to guarantee the training effect and training speed, it is proposed to choose a feed-forward neural network which only includes two hidden layers as the basis. The number of input layer nodes is the dimension of the vector, and the number of output layer nodes is 1, which means it only contains one output.

Particle coding is based on neural network structure. The weight and bias parameters of each layer of the neural network are used as the optimization dimension, and iterative optimization of coding is performed.

PSO parameter initialization. The number of particles in a particle swarm is determined based on the problem requirements, usually 10-1000 particles. Learning factors, particle velocity, and position should be initialized. Since the particles need to optimize the optimization dimension, the location information should be initialized to satisfy the range of values and weights in the neural network. The evaluation function of PSO fitness value should be determined. The evaluation function of neural network basically includes Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Percentage Error (MAPE). In order to meet the accuracy of neural network, the root means square error is selected as an evaluation function in this problem. The root means square error expression is as follows.

$$ fitness = RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} $$  \hfill (1)
• PSO iterate to find the optimal solution. The encoded information in the particle is decoded and then substituted into the neural network. The neural network runs to calculate the fitness value and the particle updates according to the fitness value until the approximate optimal solution is found. The solution is regarded as the approximate training result of the neural network.

• Neural network training. The accuracy of the PSO optimization algorithm is limited. Therefore, it is necessary to continue the neural network training based on the approximate optimal solution obtained by PSO. Since it is already in the vicinity of the optimal solution, the problem of falling into the local optimal solution is almost avoided, and the training time is greatly shortened.

• Bagging integrated learning. For the multiple PSO-BP neural network learners obtained, voting is performed by analyzing prediction errors. The learner with the most votes is the final Bagging-PSO-BP neural network model.

3. Experiments Analysis

3.1. Experiment Data Introduction
The experimental data comes from the results of some subjects of eleven sophomores in the College of Software of the author's school. Considering the importance and relevance of the courses, the compulsory and strong correlation of course results were analyzed to get the following results table.

The algorithm takes the values of the five columns as input when the last one as output. The neural network fits the last column of data according to the input of the first five columns, evaluates and updates the parameters with the MSE as a loss function. Select the first 8 rows of the data as the training set and the last 3 rows as the test set for neural network fitting. In order to ensure the fitting effect, choose a neural network with two hidden layers, where each layer consists of 20 neurons and the activation function is tansig. The final output layer selects purelin as activation function to guarantee the prediction range of the output value.

Table 1. Subjects score of strong correlation of eleven students.

| Math  | Linear algebra | Data structure | Computer composition | Java | Algorithm |
|-------|----------------|----------------|----------------------|------|-----------|
| 98    | 96             | 97             | 97                   | 87   | 100       |
| 100   | 97             | 79             | 96                   | 89   | 87        |
| 66    | 84             | 83             | 83                   | 93   | 65        |
| 83    | 98             | 85             | 93                   | 89   | 74        |
| 91    | 96             | 86             | 87                   | 85   | 90        |
| 92    | 91             | 79             | 88                   | 86   | 74        |
| 90    | 88             | 87             | 87                   | 87   | 76        |
| 87    | 93             | 70             | 92                   | 89   | 99        |
| 98    | 99             | 93             | 98                   | 93   | 98        |
| 94    | 97             | 89             | 96                   | 93   | 95        |
| 97    | 98             | 96             | 92                   | 91   | 96        |

3.2. Experimental Performance Evaluation Index
It is acknowledged that MSE and MAE perform well explaining the effectiveness of neural network algorithm. Therefore, MSE and MAE are also selected as performance evaluation indicators for comparative experiments. By comparing the MSE and MAE indexes of common BP neural network and Bagging-PSO-BP neural network, the optimization performance of Bagging-PSO-BP neural network is evaluated.
3.3. Experiment results

Matlab is selected as the programming language for algorithm testing. The standard BP neural network has a very high degree of fitting to the sample set after training. The specific parameters are as follows.

| Table 2. BP neural network training results. |
|---|---|---|---|
| Epoch | Time | Performance | Gradient |
| 48266 | 0:19:42 | 7.11e-13 | 0.00154 |

| Table 3. BP neural network test results |
|---|---|
| Predict Value | Actual Value |
| 98.4070 | 98 |
| 98.9706 | 95 |
| 91.8320 | 96 |

The test results of BP neural network are analyzed. MSE=11.1012 and MAE=2.8485 are obtained. However, the MSE in the BP neural network fitting to the training set is only 7.11e-13. Its advantage is high-precision while the disadvantage is an unideal result. After many times of model training, it is found that the MSE always stays between 10-30, the MAE is between 2-10 and the prediction difference is about ±6 points.

| Table 4. Bagging-PSO-BP neural network training results. |
|---|---|---|
| Time | Performance | Gradient |
| 0:00:41 | 1.12e-11 | 0.00328 |

Compared with the BP neural network model, it can be seen that although the Bagging-PSO-BP neural network is slightly lower than the BP neural network in the accuracy of the training set, training times and training time are greatly decreased.

| Table 5. Bagging-PSO-BP neural network test results |
|---|---|
| Predict Value | Actual Value |
| 97.4772 | 98 |
| 95.3992 | 95 |
| 97.6201 | 96 |

According to the test results in the table, MSE is 1.0191; MAE is 0.8474, significantly lower than BP neural network.

Compared with the standard BP neural network, Bagging-PSO-BP neural network can significantly reduce the time of model training. At the same time, the training degree and accuracy of the test set are better. It can be roughly estimated that the introduction of PSO and Bagging makes neural network steadier. The degree of over fitting is reduced, the generalization is enhanced, and the processing of special data is more excellent.

However, what cannot be ignored is that only a small-scale dataset was used in the objective problem experiment. Meanwhile, in order to speed up the training speed of the neural network, only the courses with a strong correlation were selected for the experiment. So Bagging-PSO-BP neural network could not be rigorously proved better in efficiency and accuracy in grade prediction. Further optimization of Bagging-PSO-BP neural network model should be achieved so that the model can truly perform large-scale multi-dimensional performance prediction.
4. Conclusion
This paper combines three common machine learning algorithms: Bagging ensemble learning, PSO particle swarm optimization algorithm, and BP neural network algorithm. It proposes a high-precision algorithm for predicting target grade based on previous learning performance, which solves the difficulty of continuous values and generalization issues. Compared with decision tree and Bayesian network algorithm, this algorithm can perform continuous value prediction, that is, predicting percentage scores rather than graded results; compared with the basic neural network algorithm, the training time is significantly reduced and the prediction accuracy is improved.

At present, the application of educational big data is still relatively poor. A large amount of valuable information has not been well utilized. Through the analysis of educational big data, the level of personalized education can be significantly improved, and the emergence of education management can be solved. Nowadays, the development of education meets a bottleneck. Traditional education methods cannot meet the growing demand for education. Combining big data and education is the only way for the new round of reform and development in education.

5. Acknowledgments
The project was supported by the National Natural Science Foundation of China (Grant no.61771001).

6. References
[1] X.T. Cui 2017 Research on Early-warning Algorithm for University Achievements in Big Data Environment Journal of Tonghua Normal University vol. 12 pp. 75–78
[2] Q. Wang 2015 University Education Early Warning Management Based on Data Mining Technology China Management Informationization vol. 18 pp. 242 - 244
[3] B. Wang 2007 Grade Data Mining and Analysis Based on Multi-Strategy Computer Technology and Development vol. 2 pp. 200 - 204
[4] H. Wang & P. Liu 2015 Application of improved association rule algorithm in early warning of student performance Computer Engineering and Design vol. 3 pp. 679 - 682
[5] K. Wu 2017 Modeling and Research of Early Learning System for Student Achievement Based on Machine Learning Journal of Taiyuan Urban Vocational College vol. 12 pp. 178-180
[6] Y. Chen 2016 Research and implementation of result prediction based on genetic neural network Modern Electronics Technique vol. 5 pp. 96 - 100
[7] J.Y. Long 2014 Analysis of Student’s Achievements Based on Mean Cluster and Decision Tree Algorithm Computer and Modernization vol. 6 pp. 79 - 83
[8] J.M. Huang 2013 Application of Bayesian Network to Predicting Students' Achievement Computer Science vol. Z3 pp. 208 - 282
[9] Y. Zhang 2017 Big Data Analysis Using Neural Networks Journal of Sichuan University(Engineering Science Edition) vol. 1 pp. 9 - 18
[10] L. Zhang 2016 National matriculation test prediction based on support vector machines Journal of University of Science and Technology of China vol. 47 pp. 1 - 9