Research and Implement of Course Early Warning System Based on Teaching Behaviour Data

To cite this article: Huansong Yang et al 2019 J. Phys.: Conf. Ser. 1176 042052
Research and Implement of Course Early Warning System Based on Teaching Behaviour Data

Huansong Yang¹, Jingwei Diao², Tao Zhou, Zhengwei Yao, Xingmin Shi and Zhuping Wang

The Institute for Education Data Hangzhou Norma University, Hangzhou, China

*Corresponding author e-mail: diaojingwei393@163.com

Abstract. Digital early warning research based on the course performance of students, is constantly advancing. However, there are a few scholars combing one of specific course, taking the early predication by digging and analysing the data of teaching behaviour and provide real-time intervention. This research, making Java program design as entry point design an overall framework for early warning system based on the data of teaching behaviour, and combines the certain algorithm to implementing the early course predication model with relatively higher accuracy by gathering and analysing students’ past learning data as well as real-time teaching behaviour’s data. Practice indicates that, according to the grade and GPA (grade point average) of pre-course as well as study behaviour during week 4 and week 6, there can be predicted course pass rate of students accurately, especially from week 5, which has high accuracy of predicting. Moreover, the model is able to provide early intervention and automatically sending learning materials for some students who get unsatisfied performance basing on the predicating data, to improve the pass rate for students.

1. Introduction

At present, pre-warn research and practice based on digital learning is in its infancy abroad and domestically, such as course signal system developed by Purdue university, the learning success system of Desire2Learn in the United States, starfish early warning system developed by American starfish enterprises, forecasting and early warning application system developed by China university of electronic science and technology and so on. Course Signals use a new algorithm called SSA (Student Success Algorithm) to monitor students’ learning status, estimating the students who are in learning risk, and pre-warn by guard signal. According to guard signal, teachers intervene to students by email, message, and online news and so on, to promote the healthy growth of students [8]. This research, making Java program design as entry point design an overall framework for early warning system based on the data of teaching behavior by gathering and analyzing students’ past learning data as well as real-time teaching behavior’s data, study for the course prediction model and it is able to predict students’ course performance accurately and provide early intervention for students by early experimental data, leading course results and GPA, which solves some key technology of performance prediction model in course early warning system.
2. Another section of your paper Framework design for course early warning system

![Diagram of course early warning system]

Figure 1. The framework of course early warning system
The framework of course early warning system is shown in Figure 1. It includes four parts: data gathering, predication modeling, early warning present and intervention applying.

3. Course early warning model
Predictive modeling is a critical link in the overall framework of the course early warning system. The following takes the course teaching of Java program design as an example to discuss the process of establishing the course grade prediction model.

3.1. Data source and pre-processing
The school’s course teaching intelligence learning platform provides a virtual simulation teaching environment based on private cloud.

| Name | Student number | Experimental duties | Experiment scores | First time experimental course | Score | Class begins | Submission time |
|------|----------------|---------------------|-------------------|-------------------------------|-------|--------------|-----------------|
| ..... | .....          | 28/28               | 99                | 100                           | 100   | 2017-03-02 12:29 | 2017-03-02 13:43 |
| ..... | .....          | 26/28               | 95                | 100                           | 100   | 2017-03-02 12:29 | 2017-03-02 14:01 |
| ..... | .....          | ......               | ......             | ......                        | ......|              |                 |
The course of Java Programming Design adopts the mixed teaching mode of classroom teaching and online teaching based on digital, in addition to regular classroom teaching, teachers can use the intelligent learning platform of course teaching to carry out practical teaching and students' extracurricular online learning. Teachers can arrange experimental work through the platform, and students can submit experiments on the platform, at the end of the semester, the results of each experiment and the starting time and submission time of the experiment can be derived. In this study, a total of 203 students from five classes were tested, the data exported after the end of the semester is shown in Table 1 (A total of 14 experiments). The completion time of the experiment can be obtained by calculating the starting and ending time of the experiment (In minutes), the initial data source is obtained by removing the useless columns such as the student name and adding the student final total score as shown in Table 2. The overall evaluation score in the initial data source table is discretized by rank (Over 90 is converted to A, 80-90 is B, 70-80 is C, 60-70 is D, and below 60 is E), finally, the preprocessed data table is obtained, as shown in Table 3. The score of each experiment and the completion time of the experiment are the main parts of the final total score, at the same time. Teachers agree that students' scores in each experiment and the completion time of the experiment are the important basis to measure their Java programming ability, it is reasonable to predict the final score of the student from the experimental score and the completion time.

### Table 2. Initial data source.

| First time experimental course | Second time experimental course | Third time experimental course | ...... | Fourteen time experimental course | Final total score |
|-------------------------------|--------------------------------|--------------------------------|-------|---------------------------------|------------------|
| Score                        | Completion time                | Score                          | Completion time | ..... | Score                          | Completion time |
| 100                          | 74                              | 100                            | 45                | 100  | 18                             | 100  | 38                             | 94.8            |
| 100                          | 92                              | 100                            | 104               | 100  | 72                             | 50   | 79                             | 79.4            |
| .....                        | .....                            | .....                          | .....             | ..... | .....                          | ..... | .....                          |                 |

### Table 3. Data table after preprocessing.

| First time experimental course | Second time experimental course | Third time experimental course | ...... | Fourteen time experimental course | Final evaluation grade |
|-------------------------------|--------------------------------|--------------------------------|-------|---------------------------------|-----------------------|
| Score                        | Completion time                | Score                          | Completion time | ..... | Score                          | Completion time |
| 100                          | 74                              | 100                            | 45                | 100  | 18                             | 100  | 38                             | A                |
| 100                          | 92                              | 100                            | 104               | 100  | 72                             | 50   | 79                             | C                |
| .....                        | .....                            | .....                          | .....             | ..... | .....                          | ..... | .....                          |                 |

3.2. *Determine the optimal classification algorithm*

It is a typical classification problem in machine learning to predict the final grade of students through the experiment results and the completion time. The results of Java Programming Design course are predicted by using K-Nearest Neighbor, Decision Tree and SVM, it is found that the SVM algorithm has the lowest error rate.
3.2.1. KNN algorithm.
K is usually takes an integer not greater than 20 in KNN algorithm, different values of K will lead to different error rates. In order to find the K value that is most suitable for the data set, the search range is controlled from 1 to 20 by adjusting the value of K. The test results are shown in Figure 2, when K is taken as 5, 6, 8, 9, 10, the error rate is at the lowest of 0.19, and the prediction result is the best.

![Error rate using KNN algorithm.](image)

**Figure 2.** Error rate using KNN algorithm.

3.2.2. Decision tree algorithm.
As shown in Figure 3, the decision tree is generated in the R language using the rpart package. X04Grade represents the fourth experiment result, X11Time represents the completion time of the 11th experiment, and the remaining analogies, the predicted error rate is 0.26.

![Grade prediction decision tree.](image)

**Figure 3.** Grade prediction decision tree.
3.2.3. **SVM algorithm.**

We use c-classification for the type and combination of four kernel functions to predict in order to obtain the SVM model with the lowest error rate. The error rate is shown in Table 4, it can be seen that the lowest prediction error rate is 0.044 using the classification model and using the radial kernel function.

|          | linear  | polynomial | radial  | sigmoid |
|----------|---------|------------|---------|---------|
| C-classification | 0.161   | 0.191      | 0.044   | 0.206   |

From the above test results, it can be seen that the error rate of SVM algorithm (The kernel function uses radial) is the lowest in performance classification prediction. This indicates that the SVM algorithm has high prediction accuracy in the prediction of small sample data; therefore, in the score prediction of Java Programming Design, SVM algorithm is proposed to be used as the modeling algorithm.

3.3. **Determine a reasonable prediction time**

The above content is to use the data of 14 experiments as the input attribute to predict the final grade, a large number of researchers also make predictions about score before an exam. The purpose of the grade warning is to identify and intervene as early as possible students whose final grade is likely to fail, so make sure that after a few experiments you can predict your final grade more accurately. The SVM algorithm (The kernel function uses radial) was used to predict the final performance of the data of 1 experiment, 2 experiments...... 14 experiments as input attributes, the error rate is shown in Figure 4, it can be seen from the figure that from the 8th experiment, the error rate is stable below 0.059, with better prediction accuracy.

![Figure 4. Error rate predicted by each experiment](image)

3.4. **Optimization of classification prediction model.**

As shown in Figure 4, since the 8th experiment began, the prediction has a higher correct rate, but at this time the course study has passed two months or more, even if the intervention is too late, the intervention effect is relatively poor. The ability to predict more accurately at an earlier time is a key issue to be solved in predictive modelling. The leading course of Java Programming Design is the Foundation of Programming Design. Students’ mastery of programming ability in the course of the Foundation of Programming Design has a great influence on the learning of Java Programming Design course. As shown in Figure 4, the prediction error rate did not drop significantly until the 8th
experiment; the effect of the intervention would be directly affected if student scores could be predicted earlier. It is attempted to test the error rate by adding the GPA and scores of the leading course Foundation of Programming Design on the basis of each experiment data, the results are shown in Figure 5. Table 5 is the data sheet after adding the leading course and GPA on the basis of Table 5.

**Table 5.** At the same time to add the leading course and GPA.

| First time experimental course | Second time experimental course | Third time experimental course | ...... | Fourteen time experimental course | Leading course results | GPA | Final evaluation grade |
|-------------------------------|-------------------------------|-------------------------------|-------|-------------------------------|-----------------------|-----|-----------------------|
| Score                        | Completion time               | Score                         | ......| Score                         | Completion time       |     |                       |
| 100                          | 74                            | 100                           | 18   | ......                         | 100                   | 38  | 2.98                  | 93   | A                     |
| 100                          | 92                            | 100                           | 104  | 100                           | 72                    | 50  | 2.71                  | 61   | C                     |
| ......                        | ......                         | ......                         | ......| ......                         | ......                | ......|                       | ......|                       |

**Comparison of prediction error rates for different experiment times**

![Comparison of prediction error rates for different experiment times](image-url)

**Figure 5.** Comparison of prediction error rate before and after optimization

As shown in Figure 5, compared with the prediction of experimental results and completion time, the predicted error rate decreased significantly from the first experiment to the fourth experiment after adding the leading course or GPA attribute, at around 0.15, but there was no significant change from
the fifth experiment. However, the predicted error rate dropped from the second experiment to less than 0.1 after adding both leading courses and GPA in the input attribute, from the 5th experiment, the predicted error rate dropped to 0, giving well prediction accuracy. This conclusion means that the final grade of the student can be predicted after one month of class, if we intervene in time, it will improve the passing rate of the course.

4. Conclusion

The results have the following implications: (1) In the experimental course teaching, the experimental link is closely related to the passing rate of the course, and the final passing rate of students can be predicted through the early experimental data, so we should pay enough attention to the students' performance in the experiment. The low completion rate of early experiments may lead students to lose confidence in learning and fail the exam. (2) Leading courses and GPA are closely related to the passing rate of courses. Special attention should be given to students with lower grades of leading course and GPA at the beginning of the course, early guidance and intervention can be twice as effective with half the effort. (3) Although this study only takes the Java programming course as an example, the model and algorithm also have a good reference for the prediction and early warning of other courses.

Acknowledgments

This research was supported by Hangzhou Normal University Thinker program and Educational Reform program, this research was also supported by Open Experimental Teaching Environment is Integrated with The Virtual Reality of Multimodal Natural Interaction program which is a national key research and development program, program NO. 2018YFB1004900.

References

[1] Chaokai He, Meng Wu. Analysis and prediction of learning behavior of education Big Data[J]. China Remote Education, 2016, (6):54-59.
[2] Yang, D., Wen, M., and Rose, C. Peer influence on attention in massive open online courses[C]. Proceedings of Educational Data Mining. 2014.
[3] Wen, M., Yang, D., and Rose, C. P. Sentiment analysis in MOOC discussion forums: What does it tell us? [C]. Proceedings of Educational Data Mining. 2014.
[4] Jiaxin You, Zhong Sun. Study on academic performance prediction and intervention of cloud learning platform college students [J]. China Remote Education, 2016, (9):14-20.
[5] Lilin Wang, Yang Ye, Xianmin Yang. Design of online learning early warning model based on big data [J]. Modern educational technology, 2016, 26(7):5-11.
[6] Plunkett J. Starfish supports University of Nebraska-Lincoln’s efforts to improve campus coordination, early identification of at-risk students, and enhanced student planning [OL]. http://www.prweb.com/releases/2014/05/prweb11873587.htm.
[7] Tongtong Li, Luoying Huang, Rui Zou, Fati Wu. Construction of learning intervention model based on education big data [J]. China Electrochemical Education, 2016, (6):16-20.
[8] Huansong Yang, Jiaping Wu. Exploration on Cultivation of Application-Oriented and Innovative Talents in the Information Field under "Internet +" Environment [J]. EURASIA Journal of Mathematics Science and Technology Education, 2017 13(8):5607-5614.
[9] Huansong Yang, Mengyuan Wang, Jiaping Wu. Research on Multiple Complex Data Processing Methods Based on OpenStack Cloud Platform[J]. International Journal of Advanced Engineering Research and Science, 2017, (04).
[10] Jiaping Wu, Zhuping Wang, Huansong Yang. Education data collection, data quality analysis and data preprocessing[J]. Journal of Ningbo Education College, 2016(06).