J. Electrical Systems 20-2 (2024): 304-309

[1] Wei Zhang

[2,*] Xiaohong Hu

An Algorithm Course Grade Warning Model Based on Ensemble Learning

Journal of Electrical Systems

Abstract: - This paper analyzes the massive multidimensional teaching process data generated in the blended teaching of algorithm design and analysis courses. In response to the problems of uneven data set categories and distribution differences in curriculum data among students at different levels and majors, the data is processed unbiasedly to construct new curriculum ability features. A model based on ensemble learning is constructed to provide intelligent early warning for students' learning status. Teaching application found that the model can effectively identify students with learning difficulties and provide effective assistance for teachers' differentiated teaching and guiding students' curriculum learning.

Keywords: Algorithm Design and Analysis, Blended Teaching, Performance Warning, Integrated Model

I. INTRODUCTION

Algorithms research is one of the core issues in computer science, and most universities offer courses on algorithm design and analysis for both undergraduate and graduate students. The course content covers a wide range of disciplines, based on mathematics, physics, and computer science, but also extends to applied disciplines such as engineering and control. The difficulty of the course has caused many students to feel intimidated. Blended teaching is a combination of face-to-face teaching and online teaching[1], which has been widely used in curriculum teaching after more than 30 years of development[2,3]. Integrating blended teaching into the algorithm curriculum teaching process can provide learners with personalized learning opportunities and is an effective way to improve educational effectiveness and enhance students' knowledge mastery and practical abilities.

The current blended teaching model has generated a large amount of multi-dimensional teaching process data. How to mine valuable information such as learning habits, learning effects, and learning needs from these data, establish effective curriculum performance warning models, and provide better educational services and guidance for students is the key to improving the effectiveness of curriculum teaching. Many meaningful works have been made using intelligent analysis methods for performance prediction models. In 2010, Pardos[4] et al. proposed a Bayesian network-based knowledge tracking model that uses data from online education platforms to model and predict students' knowledge mastery levels, and provides personalized teaching recommendations based on the prediction results. He[5] et al. studied the prediction method of academic performance in large-scale online courses, and eliminated the distribution differences between different prediction models through transfer learning to improve the prediction accuracy of the model. Godwin[6] et al. observed students' task completion behavior, collected students' behavioral data before, during, and at the end of the semester, studied the relationship between students' attention allocation patterns, static characteristics, teacher teaching design, and school type, and provided suggestions and guidance for improving students' classroom attention based on the experimental results. Krejcar Ondrej[7] proposed a supervised machine learning method to construct a student achievement analysis and prediction model, using students' historical academic performance to predict their final grades. The model uses a variety of experimental tools such as decision trees, random forests, support vector machines, and logistic regression to compare the effectiveness of classification and regression techniques in predicting student achievement. Hussain[8] et al. used regression models and decision tree algorithms to predict academic performance based on online learning data and historical education records from learners in seven regions, providing valuable information for education administrators and teachers based on experimental results. Badgley[9] et al. analyzed the age, family, occupation, and previous academic performance of 99 students using multiple regression methods to explore the correlation coefficients between these factors and academic performance.

However, in practice, performance warning still faces many challenges, which makes it difficult for existing methods to achieve ideal warning effects. Firstly, compared to normal samples (i.e. samples without warning risks), warning samples (i.e. samples with potential warning risks) are often less, and there is an imbalance in the number

[1] Department of Computer Science and Technology, Beihua University, Jilin, Jilin132013, China.
[2] Department of Computer Science and Technology, Beihua University, Jilin, Jilin132013, China.
Corresponding author: Xiaohong Hu

of positive and negative samples in the warning data set, which affects the accuracy of the warning model. In addition, algorithm courses are offered at various levels and professional stages for graduate and undergraduate students, and there are distribution differences in course data among students of different levels and majors, making it difficult for existing models to adapt to complex data from students of different majors. In response to the above practical situation, this paper proposes an academic warning method based on ensemble learning, which constructs new curriculum ability features after unbiased processing of data. The feature represents students' learning abilities in different aspects, which can eliminate the impact of different majors on the prediction model. Finally, Stacking algorithm is used for performance warning prediction.

II. Data Processing

A. Data Collection and Preprocessing

This article selects undergraduate students and graduate students majoring in software engineering from four majors in the School of Computer Science and Computational Engineering from 2019 to 2021 as the research subjects. After excluding student data with makeup, retake, and withdrawal, a total of 1,862 valid data were screened. The online teaching platform for this course mainly includes course lecture videos, experiment links, engineering practice cases, and various question banks, which can support students to engage in personalized learning and online self-testing. Before and after class, students use the platform to preview, answer theoretical questions, and conduct experiments. Data collection is divided into two parts: student achievement and process data. Academic achievement collection is imported through the educational administration system service information platform, while process data is imported from the learning platform to collect students' online learning data throughout the entire process, including students' pre-course preview page counts, preview answer scores, check-in times, classroom quiz scores, homework and project discussion performance, and practice project scores. Due to the uneven distribution of raw data, that is, 17% of students have final exam scores below 60 points and 43% of students have final exam scores above 70 points, random sampling is conducted on students with final exam scores above 70 points to reduce the number of samples, while under-sampling is used on students with final exam scores below 60 points to increase the number of samples. After sampling, the proportion of three types of students in the sample data is 1:1:1 to ensure the unbiased nature of the data. After conducting a correlation analysis on the raw data, five dimensions are selected as the five basic attributes for academic warning: preview page count, check-in time, homework after class, classroom quiz, and project discussion participation.

B. Construction of Performance-Related Features

Different majors have different knowledge backgrounds, and there are also significant differences in learning characteristics between graduate and undergraduate students, which are reflected in the significant distribution differences in course data. In order to characterize the learning abilities of students at different levels and majors in different aspects and eliminate the impact of these factors on the prediction model, a course ability feature is defined, which is the regularization of data after multiplying the relative ranking of students in the same major and the score of practical projects. Through normalization, it not only improves the interpretability and generalization of the prediction model. The final feature correlation analysis is shown in Table 1.

Table 1 Data feature correlation

	Preview page number	Sign-in time	Classroom quiz	homework	Discuss participation	course ability
Paged of preview	1					
Sign-in time	0.08345	1				
Classroom test	0.11486	0.09732	1			
homework after class	0.14382	0.08391	0.42018	1		
Discuss the participation	0.23931	0.23905	0.21091	0.33173	1	
course ability	0.1038	0.2615	0.1345	0.4629	0.3648	1

C. Standardized Data Processing

Different attribute data have different formats, quality, and completeness, which requires uniformity and standardization. Specific standard settings are as follows: Discussion participation standard: Statistics the total number of times students participate in discussions during the semester, then calculate the total number of discussions initiated by the course during the semester, and finally sum up and average. Quantification standard for preview pages: Statistics the number of pages students preview each time during the semester, as well as the total number of pages for this preview courseware. Divide the number of preview pages by the total number of pages to obtain the preview rate for one preview, and finally sum up and average the preview rates for each preview. Quantification standard for check-in time: A check-in time is generated for each class by the system, and the difference between the student's personal check-in time and the check-in deadline is calculated. The difference value greater than the threshold time is set to be negative; the difference value less than the threshold time is set to be positive, thus obtaining a check-in value for one class. Finally, sum up all check-in values and average them. Quantification standard for classroom quiz: Statistics the scores of students completing classroom quizzes each time during the semester, then calculate the total score for each quiz, obtain the scoring rate for this quiz, and finally sum up and average the scoring rates for this semester. Quantification standard for homework: Statistics the scores of students completing homework each time during the semester, with a total score of 100 points for each assignment. Obtain the scoring rate for this quiz, and finally sum up and average the scoring rates for this semester.

D. Risk Indicators

According to the operation of the course, students who score 60 or more on the final exam are considered qualified. Therefore, the warning is divided into three categories

- High-risk students. Students with academic performance below 60 points are defined as high-risk students. This group of students lacks enthusiasm for learning, has a weak grasp of basic knowledge, and is at high risk in terms of academic performance. Early warning should be provided during the course of study to ensure that they improve their learning level during the remaining semester and can avoid academic risks;
- Moderate-risk students. Students with academic performance between [60,70) are defined as moderate-risk students. These students are at risk in terms of course learning and need to be identified and warned as early as possible to help them complete their learning tasks and successfully complete the course.
- Risk-free students. The academic performance of this group of students is above 70 points, and there are no academic problems. They can study seriously without being reminded, and their learning status is stable, without the need for academic warning.

III. MODEL CONSTRUCTION

Due to the difficulty of adapting a single model to complex data, poor resistance to noise in data with high noise levels, and the potential for overfitting. Stacking ensemble learning method has been proven to effectively improve prediction ability[10].This article uses the Stacking algorithm, which can better leverage the differences between models, to construct an ensemble model to improve prediction accuracy. Taking a classification tree with a depth of 3 as the base learner, we construct Random Forest[11,12], GBDT[13], and XGBoost[14,15] classification models.

The prediction steps are as follows, and the specific process is shown in Figure 1

- Randomly select 80% of the data as the training set, 10% of the data as the validation set, and the remaining 10% of the data as the test set. The training set and validation set are mainly used for the establishment and optimization of ensemble learning models, while the test set is used for model performance evaluation.
- Use 5-fold cross-validation to train the random forest classification model, AdaBoost algorithm model, and GBDT algorithm model separately to obtain the prediction results Pi,j (i=1,2,3; j=1,2,…,5) of each classifier on the test set.
- After performing 5-fold cross-validation, use the trained first-layer model to predict the entire test set, and then average the 5 prediction results on the test set to obtain Ti.
- The output results of the three base models in the first layer are spliced to obtain the training set P4 = (P1, P2, P3) and test set T4 = (T1, T2, T3) for the second-layer logistic regression model.
- Train the second-layer learner using the training set P4 to perform logistic regression and predict T4, obtaining the prediction results for the final test set.

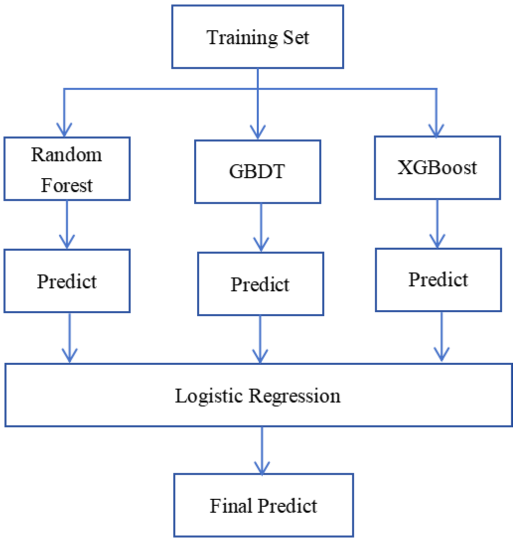

Figure 1: Structure of the Prediction Model

IV. Model Prediction Results and Analysis

This article uses four commonly used performance evaluation metrics for classification models, Accuracy, Precision, Recall, and F1-score, to evaluate the performance of the prediction model. The higher the four classification model performance evaluation metrics, the better the model. First, using Random Forest, LightGBM, and XGBoost algorithms to establish a single model to analyze and predict student data, evaluate the prediction effect of the model on students of various categories, and compare the prediction effect of students across different models. To illustrate the prediction effect of the integrated model, the three single models are fused using the Stacking fusion mechanism to establish a two-layer Stacking fusion structure, and the prediction effect of the fused model is analyzed and compared with that of the three single algorithm models.

A. Prediction Performance of a Single Classifier

Tables 2, 3, and 4 show the prediction performance of each single classifier on the test set. From the results, Random Forest has a relatively accurate prediction for low-risk students, but overall its performance on the test set is not good enough. For early warning of course grades, the most important task is to identify as many high-risk students as possible for early intervention. Therefore, the recall rate is more meaningful than the accuracy rate of model evaluation. However, the recall rate of Random Forest for high-risk students is only 0.6121.

The various indicators of GBDT are more than 6% higher than those of Random Forest, which is mainly due to the flexible processing of various types of data, such as discrete and continuous value data types, by the GBDT algorithm.

XGBoost has the best overall performance, which may be due to the fact that XGBoost adds a regularization term to the cost function to control model complexity. The regularization term reduces the variance of the model, making the learned model simpler and preventing overfitting.

Table 2 Prediction Results of Random Forest

Early warning type	Accuracy	Precision	Recall	F1-score
high risk	0.5921	0.6219	0.6121	0.6169
Low risk	0.6293	0.7117	0.6023	0.6524
No risk	0.6105	0.6832	0.6146	0.6471

Table 3 GBDT prediction results

Alert type	Accuracy	Precision	Recall	F1-score
high risk	0.7133	0.7239	0.7465	0.7351
Low risk	0.6943	0.7043	0.7361	0.7198
No risk	0.7333	0.7239	0.7465	0.7350

Table 4 XGBoost Prediction Results

Early warning type	Accuracy	Precision	Recall	F1-score
high risk	0.8043	0.8245	0.8234	0.8239
Low risk	0.7356	0.8243	0.8573	0.8404
No risk	0.7834	0.8136	0.8246	0.8191

B. Integrated Model Prediction Performance

Based on the model structure and experimental process mentioned in this article, a Stacking based multi model fusion was performed on three single models, Random Forest, GBDT, and XGBoost. The results of student prediction using the constructed fusion model are shown in Table 5. Compared with the single model, the accuracy of the integrated model for high-risk and low-risk early warning prediction has been significantly improved, and the recall rate of both categories is also greater than 80%. In addition, besides the accuracy and F1 score of excellent categories being below 80%, the prediction accuracy, recall, and F1 score of risk-free categories have all reached over 80%, indicating that the integrated model in this article has good discrimination and prediction ability for each category.

Table 5 Prediction Results of Integrated Model

Early warning type	Accuracy	Precision	Recall	F1-score
high risk	0.8646	0.8643	0.8854	0.8747
Low risk	0.8793	0.8618	0.8742	0.8679
No risk	0.8523	0.8823	0.8745	0.8784

V. MODEL APPLICATION

Based on the algorithm course learning warning model constructed above, we process and analyze student data in the course teaching process to generate student warning results, which can be viewed by both students and teachers through the system during the mid-course stage, thus providing early warning for students.

The quantitative evaluation of students' learning outcomes can to some extent illustrate the role of early warning models in improving students' learning outcomes. Guided by the curriculum objectives, we emphasize student process evaluation and establish a diversified evaluation system, including classroom tests, classroom interactions, homework, course-based extended papers, and final exams. The proportion distribution of each evaluation method is shown in Table 6.

Table 6. Proportion distribution of evaluation methods for undergraduate courses

course objectives	Performance in ordinary times	homework	course paper	Course Quiz
Understand the basic algorithm principles	5	5	5	5
Basic algorithm optimization method	5	2	5	8
Solving simple application cases	8		10	10
Conduct efficiency analysis on the solution	16	8		8
Total amount	34	15	20	31

Comparing the final exam scores of the 2020, 2021, and 2022 undergraduate computer science majors at Beihua University, the average scores of students after adopting the early warning model were 80, 85, and 86, respectively. The failure rates of the courses were 10.76%, 8.29%, and 7.35%, respectively, which also illustrates the forward-looking nature of the early warning model for learning expected results.

VI. CONCLUSION

This study aims to analyze students' online learning data for the algorithm course of blended learning, design and construct an early warning model based on ensemble learning, and apply it in the learning system of the course. Practice has shown that through the use of this model, teachers can timely understand students' learning status, implement targeted intervention and guidance for students with learning difficulties, reduce the failure rate of students in the course, and effectively improve teaching efficiency. Students can also use the model to understand their own learning status and problems, and identify the risk of declining performance or failure to meet standards.

ACKNOWLEDGMENT

This work was funded by the following projects:

1. Project of Jilin Provincial Department of Education (Grant No.: JLJY202331291428)
2. Excellent Course for Graduate Students of Beihua University Grant No.: JPK2022007)
3. Graduate Teaching Reform Project of Beihua University (Grant No.: JG2022003)

REFERENCES

[1] D. Randy Garrison, Heather Kanuka, Blended learning: Uncovering its transformative potential in higher education, The Internet and Higher Education, vol. 7, Issue 2, 2004, pp. 95-105, ISSN 1096-7516, https://doi.org/10.1016/j.iheduc. 2004.02.001.

[2] Ngoc Thuy Thi Thai, Bram De Wever, Martin Valcke, The impact of a flipped classroom design on learning performance in higher education: Looking for the best "blend" of lectures and guiding questions with feedback, Computers & Education, Vol. 107, 2017, pp. 113-126, ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2017.01.003.

[3] Wendy W. Porter, Charles R. Graham, Kristian A. Spring, Kyle R. Welch, Blended learning in higher education: Institutional adoption and implementation, Computers & Education, Vol.75, 2014, pp. 185-195, ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2014.02.011.

[4] Pardos, Z.A., Heffernan, N.T. (2010). Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. In: De Bra, P., Kobsa, A., Chin, D. (eds) User Modeling, Adaptation, and Personalization. UMAP 2010. Lecture Notes in Computer Science, vol .6075. pp. 255-266. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13470-8_24

[5] He, J., Bailey, J., Rubinstein, B., & Zhang, R. (2015). Identifying At-Risk Students in Massive Open Online Courses. Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (1). pp. 1749-1755. https://doi.org/10.1609/aaai.v29i1.9471

[6] Karrie E. Godwin, Ma. V. Almeda, Howard Seltman, Shimin Kai, Mandi D. Skerbetz, Ryan S. Baker, Anna V. Fisher, Off-task behavior in elementary school children, Learning and Instruction, Vol. 44, 2016, pp. 128-143, ISSN 0959-4752, https://doi.org/10.1016/j.learninstruc.2016.04.003.

[7] Bujang SDA, Selamat A, Krejcar O, Maresova P, Nguyen NT. Digital Learning Demand for Future Education 4.0—Case Studies at Malaysia Education Institutions. Informatics. 2020, vol. 7 (2): 13. https://doi.org/10.3390/informatics7020013

[8] Hussain, S., Khan, M.Q. Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. Ann. Data. Sci. Vol. 10, pp. 637–655 (2023). https://doi.org/10.1007/s40745-021-00341-0

[9] W. T. Mao, G. S. Wang, L. L. Kou, and X. H. Liang, "Deep domain-adversarial anomaly detection with one-class transfer learning," IEEE/CAA J. Autom. Sinica, vol. 10, no. 2, pp. 524–546, Feb. 2023. doi: 10.1109/JAS.2023.123228

[10] Wanli Xing, Xin Chen, Jared Stein, Michael Marcinkowski, Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization, Computers in Human Behavior, Volume 58, 2016, Pages 119-129, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2015.12.007.

[11] Mohammed S. Ozigis, Jorg D. Kaduk, Claire H. Jarvis, Polyanna da Conceição Bispo, Heiko Balzter, Detection of oil pollution impacts on vegetation using multifrequency SAR, multispectral images with fuzzy forest and random forest methods, Environmental Pollution, Vol. 256, 2020, 113360, ISSN 0269-7491, https://doi.org/10.1016/j.envpol. 2019. 113360.

[12] Breiman, L. Random Forests. Machine Learning. Vol. 45, pp.5–32 (2001). https://doi.org/10.1023/A:1010933404324

[13] Jerome H. Friedman. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 2001, vol. 29 (5). pp. 1189-1232.

[14] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System CA.Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp.785- 794, 2016.

[15] HOWARD E, MEEHAN M, PARNELL A. Contrasting prediction methods for early warning systems at undergraduate level. The Internet and Higher Education, 2018, vol. 37. pp. 66-75.