Quality Evaluation Method of Agricultural Talents Distance Education Based on Improved Decision Tree

Qi Wang[1], Guanghai Li[2](✉), and Yang Liu[1]

[1] School of Management, Shenyang Urban Construction College, Shenyang 110167, China
[2] Ministry of Education, Guangxi Normal University, Guilin 541004, China
1850508590@qq.com

Abstract. With the rapid development of information technology and the emphasis on education at all levels of the country and society, wisdom education, as a new application of information technology in the field of education, has a lot of research space. Educational data mining is an interdisciplinary field arising from its application in the field of education. Compared with the traditional educational environment, the current research based on the field of education is no longer lack of student behavior data. As a result, data-rich educational environments have become the norm. Abundant data provides a data base for EDM. To some extent, the sampling of education indicators can improve the problem of unbalanced data, but they also have the problems of low accuracy and insufficient sampling. This paper firstly constructs the l evaluation system of learning quality index quality based on the distance education of agricultural talents, clarifies the changes in learning quality of various groups, and adopts the difference analysis of the random forest algorithm based on an improved decision tree. By comparing with the existing evaluation model, the experimental results indicate that the network model optimized by this algorithm has a better effect on the evaluation of education quality. And that detection accuracy and precision are further improved. It is helpful for educational indicators to develop personalized evaluation and intervention programs. Finally, some suggestions for learners to improve their learning are put forward, and the research results can provide practical guidance for teaching stakeholders.

Keywords: Agricultural education · Online classroom · Education quality · Random forest · Decision tree

1 Introduction

In the case of the combination of the Internet and the field of education, online learning has become the preferred choice of learners, but also provides convenience for learners. Online learning can be watched anytime, anywhere, be recorded and broadcast, and is not limited by location and time. Because the whole online learning process is carried out online, all the learning behavior data of students can be completely retained in the

X. Yuan et al. (Eds.): ICEKIM 2023, AHCS 13, pp. 1168–1176, 2023.
https://doi.org/10.2991/978-94-6463-172-2_123

learning management system, which provides a reliable data source for the study of academic monitoring and evaluation of students' online learning behavior and teaching intervention [1]. The most commonly used method to establish a quality evaluation mechanism is to establish a student score classifier. The common classifier models are decision trees, Naive Bayes, SVM, etc. [2].

The rapid development of computer data processing and analysis technology has greatly improved the efficiency and effectiveness of the use of data in education and teaching and provided an effective way to improve the professional ability of agricultural talents. In the information age, the traditional education evaluation mechanism is moving towards precise evaluation, which can not only expand the sample in an all-round way. It can also acquire a more efficient evaluation system through some available objective information, which is a more accurate process evaluation [3]. Therefore, the evaluation of the application level of agricultural digital education resources is a highly practical activity, and its evaluation and analysis are helpful for decision-makers to predict from the perspective of intelligent science. It can be said that the evaluation through machine learning can support school decision-making, improve the quality of education and optimize education and training services [4].

Many countries and international organizations have investigated and evaluated the professional skills of education quality, including the application ability of digital education resources. The Organization for Economic Cooperation and Development (OECD) launched a large-scale international survey on the quality of education and teaching (TALIS) in 2008. It includes the use of educational curriculum resources [5]. Through interviews, KHanson and others discussed the comfort of searching and using digital resources in education and what kind of training support they want to get [6]. Chen Weiling and others distributed questionnaires through the network to understand the habits and strategies of using resources in the process of application practice [7]. Some of them are based on the platform to collect and evaluate the educational information technology behavior, but there is no evaluation of the application ability of digital educational resources, and more of them are based on the network technology of the educational platform.

This paper starts with the poor application effect of distance education for agricultural talents and the small coverage and weak pertinence of traditional evaluation methods. The paper mines the specific factors that will affect the application ability of distance education resources based on the collected education data, and uses these objective and significant characteristics to build an improved decision tree model to evaluate the application ability of agricultural education resources. Through the comparison and optimization of the model performance, the evaluation method more suitable for the research group was found, and the application feasibility of the evaluation results was demonstrated later, hoping to provide decision support for the education management department to realize the precise monitoring of education quality and training intervention.

2 Establish an Educational Quality Evaluation System

In this paper, the information of distance education courses for agricultural talents is collected for four groups of people, including students, teachers, experts and teaching supervision groups. Through the course evaluation questionnaire system designed by the course evaluation index, some of the indicators of course evaluation are established as displayed in Table 1.

Table 1. Course index evaluation system

First indicator	Second indicator	Contents
Teaching staff	Course leaders and teachers	Teacher's style, academic level and teaching level
	Structure and overall quality of teaching staff	Knowledge Structure, Age Structure, Personnel Allocation and Training of Young Teachers
	Teaching results	Project and scientific research literature
Teaching contents	Course contents	The design is in line with the goal of personnel training, and knowledge, ability and quality are organically integrated.
	Teaching design	Subject design, course design, project design
	Course practice	The content of experimental course can effectively cultivate students' practical ability and innovative ability.
Teaching source	Teaching materials and relevant materials	Construction of teaching materials and related materials
	Practice teaching conditions	Advancement and openness of practical teaching environment
	Network teaching resources	Network teaching resources are rich, auxiliary teaching and learning functions are complete, and can be effectively shared.
Teaching methods	Teaching use effect	Individualized teaching, inquiry teaching, blended and flipped teaching, etc.
	Application of information technology	Using modern information technology to carry out teaching activities

3 Quality Evaluation Model of Distance Education

SPSS Modeler is applied to construct the model and obtain the evaluation results of the model, and then the classification algorithms are screened. According to the results, the evaluation performance of random forest is better than other classification models. Therefore, this paper selects to use the random forest algorithm to improve.

Random forest is an ensemble learning method based on the random subspace technique and a Bagging algorithm. At the same time, the weak classifiers of random forests are CART decision trees, and each tree model will evaluate the input data. The final evaluation result of random forest is to combine the evaluation results of all weak classifiers, and finally select the majority vote of all results as the final evaluation result [8]. The weak classifier of random forest adopts the decision tree as the index to divide the features.

The decision tree is calculated as shown in Eq. 1.

$$\theta_{ini}(p) = \sum_{i=1}^{k} p_k (1 - p_k) \tag{1}$$

where θ_{ini} denotes the decision tree, k denotes the number of classes in the data set, and p_k denotes the probability that the sample belongs to class k. When $k = 2$, Eq. 1 is converted into the decision tree calculation mode of the binary classification problem, as shown in Formula 2.

$$\theta_{ini}(p) = 2p(1 - p) \tag{2}$$

where (p) denotes the probability that the sample belongs to the positive sample.

The decision tree calculation formula for the data set ϕ is displayed in Fig. 3.

$$\theta_{ini}(\phi) = \sum_{i=1}^{k} \frac{|\eta_k|}{|\phi|}^2 \tag{3}$$

where $|\eta_k|$ represents the number of data sets, and the data subsets η_i and η_j are generated according to a certain value of the feature (ϕ), as shown in Formula 4.

$$\phi_1 = \{(x, y)\phi | A(x) = a\} \tag{4}$$

Therefore, the decision tree of a certain feature $A(x)$ under a certain value a can be calculated according to Formula 5. By comparing the decision trees of all values under feature a_i, the value of feature a_i that minimizes the decision tree is selected as the best segmentation point to partition the feature [9].

$$\theta_{ini}(\phi, A) = \frac{|\phi_1|}{|\phi|}(ini(\phi_1)) \tag{5}$$

The algorithm flow of the random forest weak classifier is summarized in Fig. 1.

Input: Training dataset ϕ.

Output: Weak classifier. Starting from the root node, each node of the weak classifier needs to perform the following operations:

(1) Randomly select n features to form a feature set α;

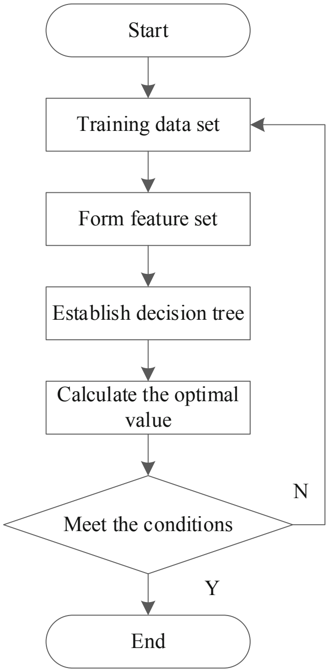

Fig. 1. Random Forest Weak Classifier Algorithm Flow

(2) Calculate a decision tree of all values of the data set under the feature set α' according to Formula 5;
(3) Select the minimum characteristic and the corresponding value of the decision tree as the optimal characteristic and the optimal segmentation point, divide left and right sub-trees according to the selection result, divide the data set ϕ into two subsets according to the optimal segmentation point, and take the two subsets as the input of the left and right sub-trees;
(4) Repeat that steps 1, 2 and 3 for each child node until the end condition of the algorithm is met.

Repeated random under-sampling is applied to the training data of each weak classifier in the ensemble classifier, and random forest, as a representative ensemble learning algorithm, also combines multiple weak classifiers.

Therefore, this paper combines the idea of repeated random under-sampling with the random forest algorithm, and applies repeated random under-sampling to the training data subset of each weak classifier of the random forest.

4 Simulation Experiment

4.1 Experimental Comparison Scheme

Naive Bayes, the Support Vector Machine and the improved evaluation algorithm of decision tree in this paper are selected for preliminary modeling analysis to find out the evaluation algorithm suitable for the learning data. The main reason is that these models have the following significant advantages: (1) they are supervised models, which are easy to control parameters; (2) the model construction is simple and efficient, which is very suitable for the situation of medium sample size (and feature number); (3) the model has high stability and can achieve good evaluation on the set learning features [10].

4.2 Method of Experiment

Select "evaluation variable = grade" and "evaluation attribute = interaction attribute, interest attribute, ability attribute, and knowledge attribute" through the command of "loading data | dividing data set | preparing target input and evaluation variable | data discretization | model building | model application | result display |" in the way of automatic modeling in Rapid Miner. The parameters in the "Split Data" operator were set to 0.6 and 0.4, indicating that 60% of the data was chosen to train the model and 40% of the data was adopted to test the model, respectively, with the default model building automatic parameters [11]. From the two indicators of evaluation accuracy and classification error rate, the suitability of each evaluation model to the case data is judged, and finally, the influence weight value of each attribute on the evaluation result is calculated, which is displayed in Fig. 2.

The online operators complete the construction process of the evaluation model together, replace the "Naive Bayes (Kernel)" operator with the evaluation model in turn,

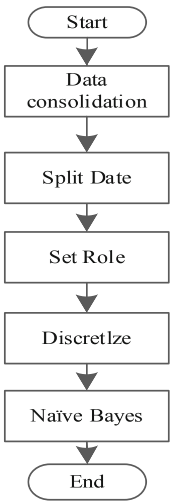

Fig. 2. Evaluation model

and adjust the online and automatic parameter settings of the operators to complete the automatic construction of each evaluation model.

4.3 Comparison of Experimental Results

The comparison results of experimental schemes are displayed in Figs. 3 and 4.

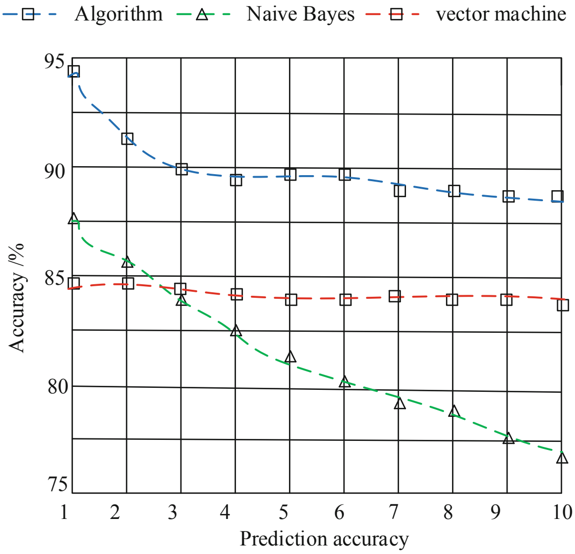

Fig. 3. Comparative analysis of evaluation results of each model

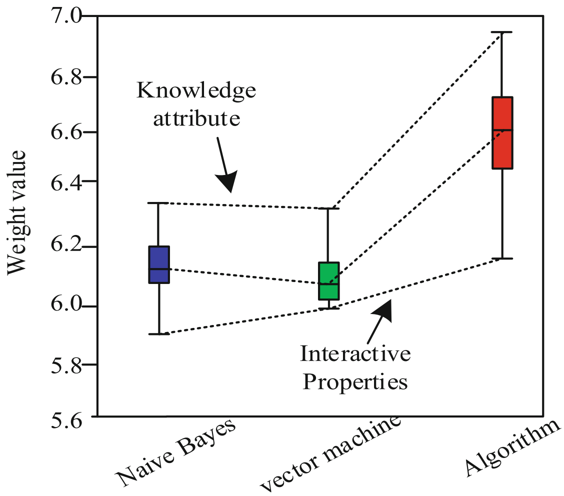

Fig. 4. Comparison of weight ratios

As indicated in Figs. 3 and 4, knowledge attributes and interaction attributes account for a higher weight in each model, which indicates that these two attributes are factors that have a greater impact on academic performance. The evaluation accuracy of several models has reached 0.59 ~ 0.69, of which the accuracy of the algorithm in this paper is the highest: 0.66. The model also has great advantages in training and recognition time. It verifies that the algorithm model in this paper can better explain the learning data.

5 Conclusion

In this paper, the quality evaluation model of agricultural talent education resources based on a random forest algorithm is constructed. The parameters of the model are optimized by using grid search and cross-validation methods to output the optimal evaluation results of each model. The evaluation of education quality can not only promote efficient learning, but also play a very considerable role in improving the quality of courses. It is of great significance to the development of education.

Due to the limitation of current experimental conditions and learning time, the source of all index data is still questionnaire data, which mainly focuses on the analysis of influencing factors, index selection, algorithm implementation and application feasibility of evaluation. If a system can be developed for practical application in education background file management and training records, combining the evaluation results of digital educational resources with the decision-making system effectively can better reflect the application value of this study.

References

1. Tao Y, Zhang Y, Jiang B. DBCSMOTE: a clustering-based oversampling technique for data-imbalanced warfarin dose prediction[J]. BMC medical genomics, 2020, 13(10): 1-13.
2. Prasanthi L S, Kumar R K, Srinivas K. A novel random forest approach using specific under sampling strategy[M]//Data Engineering and Intelligent Computing. Springer, Singapore, 2018: 259-270.
3. Saa A A, Al-Emran M, Shaalan K. Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques [J]. Technology, Knowledge and Learning, 2019, 24(4): 567-598.
4. Wang R, Gan B. Teaching Activity Model of "PHP Website Development Technology" Course Reform Under the Background of Wisdom Education[C]//International Conference on Application of Intelligent Systems in Multi-modal Information Analytics. Springer, Cham, 2020: 252–257.
5. Abu Saa A, Al-Emran M, Shaalan K. Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques[J].Tech Know Learn,2019,24:567–598.
6. Ayse Saliha S, Rabeeh Ayaz A, Hugh C, et al. Modelling MOOC learners' social behaviours [J]. Computers in Human Behavior,2020,107.
7. Riestra-Gonz M, Paule-Ruiz M, Ortin F. Massive LMS log data analysis for the early prediction of course-agnostic student performance[J/OL].Computers & Education,2021,163.
8. Nejatian S, Parvin H, Faraji E. Using Sub-sampling and Ensemble Clustering Techniques to Improve Performance of Imbalanced Classification[J]. Neurocomputing, 2018, 276: 55-66.

9. Wilson C, Rebeca C, Cristóbal R. Multi-source and multi-modal data fusion for predicting academic performance in blended learning university courses- Science Direct[J].Computers & Electrical Engineering, 2021,89.
10. Ali Derakhshan. Tasks, Pragmatics and Multilingualism in the Classroom: A Portrait of Adolescent Writing in Multiple Languages[J].Journal of Pragmatics,2020,168,53–55.
11. Aikens Nikki, Atkins-Burnett Sally, Tarullo Louisa, Malone Lizabeth, Kelly Annalee, Cannon Judith. Preschool boys of color: Portraits of the population served by Head Start [J]. Journal of Applied Developmental Psychology. 70. 101167. https://doi.org/10.1016/j.appdev.2020.101167,2020.