Research on Academic Prediction and Intervention From the Perspective of Educational Big Data

Xiaoming Du, Jiangsu University of Science and Technology, China*
Shilun Ge, Jiangsu University of Science and Technology, China
Nianxin Wang, Jiangsu University of Science and Technology, China

ABSTRACT

In the context of education big data, it uses data mining and learning analysis technology to accurately predict and effectively intervene in learning. It is helpful to realize individualized teaching and individualized teaching. This research analyzes student life behavior data and learning behavior data. A model of student behavior characteristics is constructed, and a robust multi-task learning method is used to construct an academic prediction model. According to the prediction results, different intervention measures are taken for students with academic excellence and academic difficulties. Finally, it takes the one-semester blended teaching course of a certain university as an example. The research results show that in terms of predictive models, through the analysis of student behavior characteristics data, the model can accurately identify the learning status of students. In terms of intervention, it can play a positive role in promoting students with high learning and can effectively promote students with learning difficulties.

KEYWORDS

Academic Intervention, Academic Prediction, Education Big Data

INTRODUCTION

Academic prediction and intervention are hot research topics in the education field. With the application of cloud computing and big data in the education field, the use of data mining and learning analysis techniques to accurately predict and intervene in school is an effective way to achieve personalized teaching in accordance with their aptitude. The “Education Informatization 2.0 Action Plan” points out that the important feature of education informatization 2.0 is the proposal and application of educational big data (Xiaoqing & Zhanjie, 2018). “China Education Modernization 2035” proposes to use modern technology to accelerate the reform of the talent training model and realize the organic combination of large-scale education and individualized training (Zhengce, 2019). Education big

DOI: 10.4018/IJICTE.315763 *Corresponding Author

data refers to the collection of structured data, semi-structured data and unstructured data generated in the whole education process, including data in the learning process, campus life and education management process (Fang & Yanshen, 2018). Through data mining and learning analysis techniques, valuable information and knowledge can be mined fro these data so as to discover learning rules and predict potential problems (Cuibi et al., 2020).

Certain results have been achieved in the prediction of students' academic performance. Some scholars have used campus behavior to predict student performance (Zhijia et al., 2018), others some scholars have analyzed students' online learning behavior to predict academic performance (Suartama et al., 2019; Yizhou & Qiong, 2018). However few of them have comprehensively studied both campus behavior and learning behavior. Besides, the results of academic predictions are mostly based on students' total scores or a certain course, instead of considering the correlation and difference between different courses. For the application of prediction results, most scholars focus on the intervention of students with learning difficulties, without recommending excellent students with learning ability. Therefore, this paper proposes to describe students' behavior from multiple perspectives, such as online learning, classroom learning, extracurricular learning and campus behavior. The robust multi-task learning method is used to predict the learning situation of multiple courses at the same time, meanwhile the learning intervention is carried out for students with learning difficulties according to the prediction results, and resources are recommended to outstanding students.

LITERATURE REVIEW

Regarding the core question of "How to accurately predict students' academic performance and conduct prescription intervention?", scholars have carried out research on behavior characteristics, model construction and timely intervention, and began to pay attention to the diversity and comprehensiveness of data types, and the real-time nature of academic performance prediction and intervention.

Students' Behavior Characteristics

The emergence of information technology has made the sources of student behavior data wider, the granularity of collection smaller, and the analysis methods diversified. Which have of different sources and low granularity, it is necessary to generalize the behavior characteristics of students through different algorithms, Liu Bopeng used partial mutual information method and correlation rule algorithm to dynamically select students' basic attributes, campus attributes and historical scores attributes, and then used support vector machine to predict students' scores (2019). On the basis of personality psychology and learning analysis theory, Li Youzeng used the data mining algorithm based on frequent pattern tree to construct the student behavior analysis model with five dimensions, including students' basic information, classroom learning, extracurricular learning, campus life and entertainment (2018). Chen Zijian jointly determined the influencing factors of academic performance by calculating the correlation coefficient between all individual data attributes and academic performance categories and calculating the information gain rate of all attributes (2017). Some scholars not only summarized the behavioral characteristics of the students, but also calculated the degree of influence of different characteristics on academic performance and the relationship between the different characteristics, Through machine learning and regression algorithms, Sun Faqin found that learning attitude, learning time level and investment level are the main factors affecting academic performance, while other factors have little or no impact on academic achievement (2019). Wu Qing selected the learning behavior data of distance learning platform and used association rule algorithm to mine the internal law among learning style, learning behavior and learning achievement (2015). Zhao Huiqiong used multiple regression method to determine 6 early warning factors affecting students' learning performance from 21 learning activity variables, and constructed regression equations affecting learning performance (2017). Li Hongyan used the stepwise regression method to extract the influence of knowledge input, learning attitude, and test attention on students' academic performance, and explored the relationship between these indicators (2019).

Academic Predictions

Academic prediction refers to the input of data generated by students in the process of study and life, the calculation of the predictive model constructed by educational data mining technology and learning analysis technology, and the output of possible academic performance at the end of the study. Domestic and foreign scholars have studied academic prediction under different scenarios from the perspectives of prediction techniques, theoretical models and realization of prediction systems. With the expansion of data from structured to unstructured, prediction methods have also changed from educational statistical analysis methods to machine learning methods such as neural network, cluster analysis, regression analysis, decision tree and deep learning, among which regression analysis is the most widely used method. Based on the Big Five personality theory, Zhang Qi used multiple linear regression analysis to analyze the correlation between learning behavior indicators and learning results of groups with different personality traits (2019). Romero used cluster analysis to predict the course passing (2008).Some scholars compared and analyzed the accuracy of the algorithm. Hu Zuhui analyzed the relationship between students' online behavior and learning quality by using decision tree, association rule and logistic regression, and found that the association rule method had the highest accuracy (2017). In the aspect of prediction theory model, Wu Fadi extracted learners' personalized behavior characteristics through content analysis, designed the three-layer learning result classification principle of target, process and result, and designed the learning result prediction framework based on personalized behavior characteristics and result classification principle (2016). Mou Zhijia based on the theories of personalized learning, cognitive psychology and teaching system design. Taking learner data as the center, the theoretical framework of learning outcome prediction is designed from the aspects of learning emotion, learning experience, learning performance, learning preference, personality characteristics and learning attention (2019). Jin Yifu built the academic early warning LAOMA model by collecting data from courses, classes and extracurricular activities and using outlier data mining method (Binyamin et al., 2019). In terms of prediction system application, the online learning management system (LMS) integrates multiple functions such as course design, learning management, online examinations, etc., and evaluates and predicts learning effects through various auxiliary tools (Yifu et al., 2016). Leah P. Macfadyen of Columbia University used social network analysis and multiple regression methods to design an "early warning system" to identify students at risk (Xiaoqing et al., 2012). Purdue University started the course signal system to improve the graduation rate and freshman retention rate. The Student Success Algorithm was used to predict students who might fail in their studies and intervene to help them complete their studies successfully (Macfadyen & Dawson, 2010).

Academic Intervention

Academic intervention is to analyze the students' learning behavior and predict their learning results in the context of big data education, and carry out prescription intervention for students whose learning results are at risk, so that they can successfully complete their studies. Academic interventions can be divided into group interventions and individual interventions according to different intervention targets. According to different learning scenarios, they can be divided into offline interventions (traditional classroom learning scenarios) and online interventions (online learning scenarios). With the development of information technology and the promotion of online open courses, researchers have begun to pay attention to individual online learning and mixed learning interventions based on learning analysis,According to the change of correct frequency and error frequency with time, Zhu Zhiting implemented different degrees of intervention to achieve the effect of accurate teaching (Arnold & Pistilli, 2012). Wang Kaihua analyzed intervention strategies and designed the learning intervention model of adaptive learning system from the perspectives of time dimension, course developer, teacher and learner (Zhiting et al., 2017). Li Tongtong designed intervention strategies, intervention timing and intervention methods from four aspects: learning style type, learning progress level, learning interaction level and academic achievement level (Gaihua & Gangshang, 2019). Ding Mengmei analyzed the influence of deep level and line level teaching intervention on the students. It is found that the teaching intervention has a positive promoting effect on the students whose studies is in the danger zone, besides, the deep teaching intervention is more effective than the shallow teaching intervention (Mengmei et al., 2017; Tongtong et al., 2016).

It can be seen from the above relevant studies that the link between big data in education and modern education is getting closer and closer. Data mining technology and learning analysis technology have promoted the rapid development of educational informatization and made a lot of achievements. However, there are still shortcomings in data sources, analysis methods, and scope of intervention. For example, the data sources are relatively single, limited to learning platform data without considering traditional curriculum data, or limited to learning behavior data without considering campus behavior data. Aiming at the above problems, this study constructed a three-stage cycle structure with smaller data collection granularity, more accurate academic prediction and more comprehensive academic intervention so as to enrich the research of big data in education.

METHODOLOGY

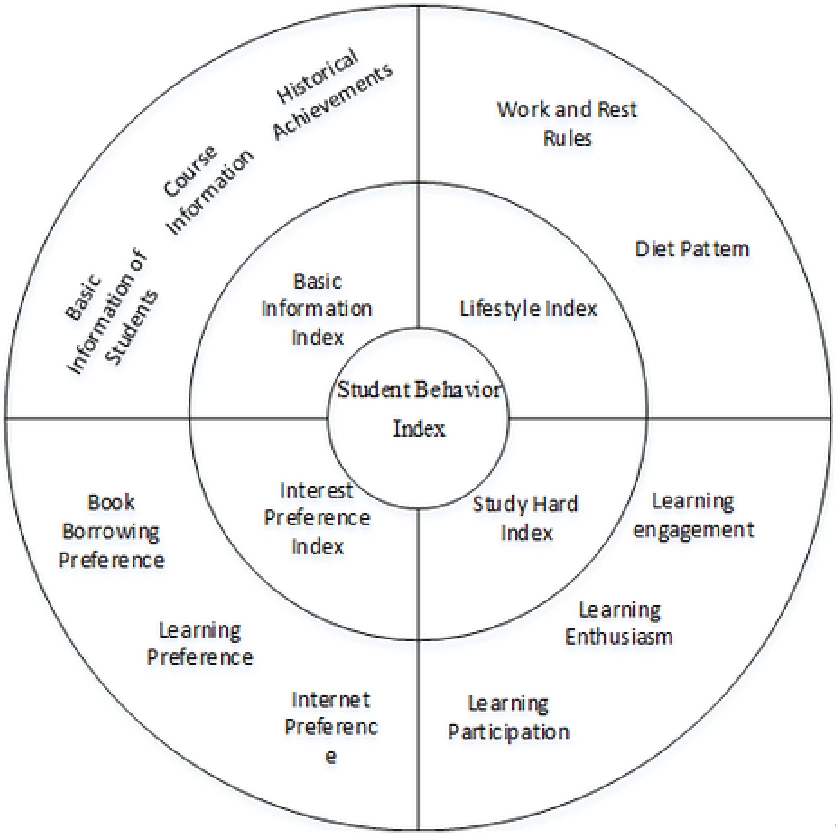
Data Acquisition and Preprocessing

The collection and storage of students' behavior data is the basis of analysis, which requires the aggregation and sharing of scattered and heterogeneous data sources for standardized processing. Based on the concept of xapi activity description, the flow of student behavior activity is constructed based on student behavior events. A behavior record of the students is called an event, which includes subject, object, acquisition equipment, time and result. A group of continuous events of the same type of acquisition equipment within a certain time is called an activity, and the activity includes subject, object, acquisition equipment, start time, end time and result. The flow of student behavior and activities can effectively depict the traces of information flow that the students move through different learning and living spaces. According to this standard, the encapsulation and storage of activity flow of data can comprehensively analyze the problems of students in the whole process of learning and living. The construction of student behavior activity flow requires data cleaning, data integration, data reduction and other processes. In order to improve data quality, problems such as missing values, outliers and data inconsistencies must be dealt with in a timely manner to lay the foundation for further research.

Student Behavior Indicators

The selection of student behavior indicators is a very important link in the prediction of academic results, which have an impact on the accuracy and interpretability of the prediction results. It is necessary to extract the core data of learning and life based on the previous data collection to form variables (attributes), and combine the variables into behavior indicators. The behaviors of students in campus life and learning process are characterized by diversity and complexity. When selecting behavior indicators, they should be both scientific and feasible, and combined with existing research to analyze indicators that have an impact on academic results. This research extracts basic information indicators, living habits indicators, study effort indicators and interest preference indicators from three levels of students' basic information, campus life and learning. These four indicators are relatively independent but interrelated organic whole, and each indicator contains variables of different granularity, so as to comprehensively reflect the individual characteristics of students and constitute the student behavior indicator model which shown in Figure 1.

Figure 1. The Model of Student Behavior Index



Basic information indicators, also known as static indicators, refer to the characteristics of students at the beginning of learning. It includes students' basic information, score information, grade point average, make-up examination and retake course information, etc. It reflects which reflect the changing process of students' historical scores and academic level, and have strong predictive ability at the beginning of learning (Hu et al., 2014).

The indicators of life habits include working, rest habits and eating patterns. Regular life habits are conducive to the control of students' study time and are closely related to their studies. Sleeping and rest habits were calculated by the number of late returns, absenteeism and the number of network traffic in the early morning,the calculation formula is $Z = \frac{w_i + q_i + l_i}{\sum p_i}$, where w_i indicates the number of late returns, q_i indicates the number of absences, l_i indicates the number of times of network traffic in the early morning, $\sum p_i$ means the sum of all times, The smaller the Z value, the better the habit. Eating regularity uses real entropy to count the regularity of breakfast, lunch and dinner, the calculation formula is $E = (\frac{1}{n}\sum_i ¼_i)^{-1} \ln n$, n represents the length of the sequence, $¼_i$ represents the length of the shortest subsequence that appears for the first time from the i-th position,The smaller the E value, the stronger the regularity (Yi & Jian, 2018).

The learning effort index is analyzed through the behavioral data of students' online learning and offline learning, and it is from the three dimensions of learning engagement, learning enthusiasm and learning engagement. The learning engagement includes attendance and online learning completion. The online learning completion is calculated by the following formula: $W = \sum_{i=1}^{m} \frac{t_i * f_i}{z_i}$, where t_i represents the length of the i-th course, f_i represents the number of the i-th course, z_i represents the total length of the i-th course. The study enthusiasm includes the times of preview before class, the times of answering questions and discussing in class, the times of taking study notes and the times of submitting homework after class. Learning participation is mainly calculated from online interaction, and its calculation formula is $C = \frac{p_i + a_i + r_i}{q}$,where p is the number of posts,a is the number of replies,r is the number of followers,q is the total number of posts.

Interest preference indicators are mainly used for the selection of intervention methods and the push of resources. It is described by three dimensions: borrowing preference, learning preference and surfing preference. The preference of borrowing books is calculated according to the categories and authors of students' borrowing books. The main statistical methods of students' online learning include learning time, learning place, resource type and learning strategy. Internet preference is counted from three characteristics: Internet time, Internet traffic and Internet times, and intervention restrictions are carried out for students who exceed the standard threshold (Zuhui & Quan, 2017).

Predictive analysis of academic achievement

Most of the existing predictions of students' academic performance are based on the random forest, neural network and other single-task machine methods. The prediction granularity is relatively large, and it is rarely used to predict each course of students. If modeling is conducted separately for each course, due to the relatively small sample data and the vulnerability of the model to data changes, problems such as redundant repetitive training and model generalization performance degradation will result. Multi-Task Learning (MTL) trains and optimizes multiple tasks in parallel, and shares information through features, examples, or parameters in the learning process to improve the generalization of a single task and the whole.

Suppose there is T course, and the training data set is $\left\{\left(X_1, y_1\right), \left(X_2, y_2\right), \ldots, \left(X_t, y_t\right)\right\}$, $X_i \in R^{d*n_i}$ is the characteristic matrix of a certain course, d denotes the dimension of students' behavioral characteristics, n_i denotes the number of students in the course, $y_i \in R$ denotes the result of the course. The prediction of T courses using the multi-task learning method can be expressed as formula(1):

$$f\left(X\right) = WX \approx Y \tag{1}$$

W represents the model weight matrix, the row represents the feature vector, and the column represents the course vector. Parallel multi−task learning between multiple courses is realized through the parameter W.

The optimal solution of the weight matrix W is the goal of multi-task learning, as shown in formula (2). The first term is the loss function, and the second term is the regularization term. The learning goal is achieved by minimizing the error and regularization of the weight matrix.

$$\min_{w} \sum_{i=1}^{T} \frac{1}{Tn_i} X_i^T w_i - y_i^{\,2} + {}^{3} w_* \tag{2}$$

Formula (2) assumes that the correlation between multiple tasks, but no correlation between some classes, so this article use robust multitasking learning algorithm modeling, robust multitasking learning tasks can be divided into abnormal related tasks and abnormal tasks (not related task), the corresponding parameter W is divided into two parts $W = P + Q$,and optimization solution formula changes into:

$$\min_{W} \sum_{i=1}^{T} \frac{1}{Tn_i} X_i^T w_i - y_i^{\,2} + \lambda_1 P_* + \lambda_2 Q^T{}_{1,2} \tag{3}$$

Low rank matrix P can distinguish related tasks by kernel norm, $P_* = \sum_{i=1}^{n} \sigma_i$, σ_i is a singular value. Sparse matrix Q performs the task of selecting abnormal tasks by the sum of L_2 norms of column vectors, $Q^T{}_{1,2} = \sum_{i=1}^{d} Q^T{}_2$.

The proximal gradient algorithm is used to solve formula (3). Let $f\left(W\right) = \min_{W} \sum_{i=1}^{T} \frac{1}{Tn_i} X_i^T w_i - y_i^{\,2}$, which is convex and differentiable, $g\left(W\right) = \lambda_1 P_* + \lambda_2 Q^T{}_{1,2}$, which is convex and non-differentiable. The minimization of formula (3) is calculated iteratively, which is shown in formula (4):

$$W_{k+1} = \mathrm{Prox}_{h,\gamma_k}\left(W_k - \gamma_k \nabla f\left(W_k\right)\right) \tag{4}$$

Among them, $0 < \gamma < 1$ is the descent step size, $W_k - \gamma_k \nabla f\left(W_k\right)$ represents the gradient descent, and $\mathrm{Prox}_{h,\gamma_k}\left(\cdot\right)$ represents the near-end operator, and the point in the space is projected onto the convex set h.

The robust multi-task learning algorithm is summarized as follows

Academic intervention model

According to the academic prediction results, the academic intervention model provides two types of intervention strategies, one is to provide active intervention strategies for students with academic risks, and the other is to provide adaptive recommendation strategies for students with excellent academic performance (Fati et al., 2019). The academic intervention model is shown in Figure 2.

Active intervention strategies include the identification of intervention target factors, the matching of intervention strategies, the choice of intervention timing and methods, and the feedback of intervention effects. The identification of intervention target factors is determined by the analysis of the above-mentioned student behavior indicators. The matching of intervention strategies was carried out by identifying the characteristics of poor students' behaviors, including the reminder of life irregularities, the comparison of learning progress, the suggestion of learning paths, the recommendation of learning resources, the cooperative learning prompts, and the learning competitive strategies. The choice of timing and method of intervention includes the choice of individual intervention and group intervention, the choice of online intervention and offline intervention, the choice of shallow intervention and deep intervention, the choice of intervention at the beginning of learning and the intervention in the learning process. Intervention effect analysis analyzes the academic situation after the intervention. If there is no improvement, the strategy, timing and method need to be changed.

Algorithm 1. Robust multi-task learning algorithm

Input: The training set $\left\{\left(x_i^{(t)}, y_i^{(t)}\right)\right\} i = 1, \ldots, N, t = 1, \ldots, T$

Initialization: $W^{(t)} = 0, t = 1, \ldots, T \quad {}^{3}{}_{k}, {}^{2} \epsilon \left(0, 1\right)$

For k=1 to T:

While True do:

$\text{Set } Z = \text{Prox}_{h, {}^{3}{}_{k}}\left(W_k - {}^{3}{}_{k} \nabla f\left(W_k\right)\right)$

If $f\left(Z\right) \leq f\left(W_k\right) + \nabla f\left(W_k\right)^T \left(Z - W_k\right) + \frac{{}^{3}{}_{k}}{2} Z - W_{k2}^{2}$:

Exit the loop

else:

$\gamma_k = \beta * \gamma_k$

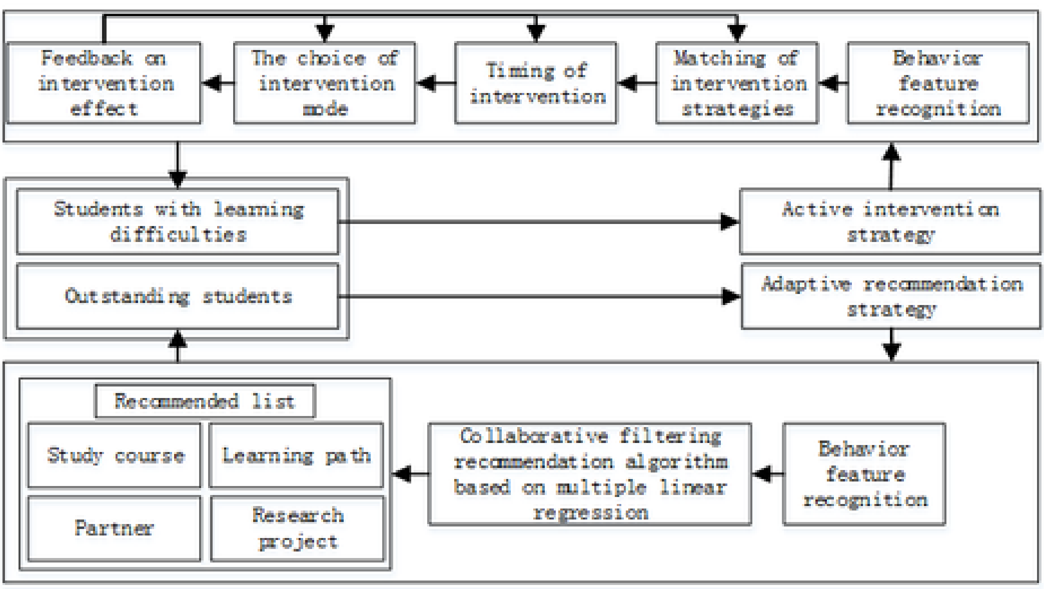
End while

$W_k = Z$

End for

Output: W

Figure 2. The Model of Academic Intervention



By identifying the behavioral characteristics of academically excellent students and adopting collaborative filtering recommendation algorithm, the adaptive recommendation strategy recommends the list of learning resources for them, including learning paths, learning courses, cooperative partners, teachers' scientific research projects, students' innovation and entrepreneurship projects, etc. For the sparse data problem of the collaborative filtering recommendation algorithm, multiple linear regression algorithms are used to predict and fill the vacancy.

RESULT

Study Design

The research object is 90 sophomore university students majoring in information management and information systems in a university. They have taken a compulsory course "Python Programming", which uses a hybrid teaching model combining traditional classrooms and MOOC education platforms. Students complete basic knowledge learning, collaborative discussion and testing in traditional classes, and learn relevant resources, upload task works and interactive evaluation on MOOC platform. This course is divided into four tasks. Task 1: Basic Python grammar, data types and basic operations; Task 2: Program control structure; Task 3: Application of functions; Task 4: Object-oriented programming. Students will conduct a comprehensive assessment after the end of each task, in which the stage test score accounts for 60% of the task score, student mutual evaluation accounts for 20%, and teacher evaluation accounts for 20%. At the beginning of the course, the students' performance on each task was predicted based on their historical scores, life behavior and study effort behavior data. The robust multi-task learning algorithm and neural network algorithm designed in this paper are used to predict and compare the performance of different algorithm. Starting from the third task, students with academic difficulties were randomly divided into an experimental group and a control group based on the prediction of student performance, and the experimental group was intervened while the control group was not intervened.

Academic Prediction Control

Obtain the course information and past grades information from the educational administration system. The original data of students' life and diet were obtained from the one-card management system, and the sleeping, and rest and diet rules of each student were calculated according to the above formula. According to the data of the students in the classroom and the MOOC learning platform, calculate the learning engagement, learning enthusiasm and learning reference of each course. Students' basic information indicators and living habits indicators are shared data indicators for all courses which is study effort indicators are exclusive data indicators for each course. Table 1 shows student behavior indicators.

The academic prediction target is the students' performance, which is divided into three categories, namely excellent, normal and risky. The students' data are randomly divided into training set and testing in the proportion of 2:1. Accuracy rate, recall rate and comprehensive analysis indicators are used to evaluate the prediction effect. In order to detect the algorithm proposed in this article, the single-task learning algorithm and the multi-task learning algorithm are used for comparison during the training process. For the multi-task learning algorithm proposed in this article, the learning effort behavior data of other courses are introduced to predict multiple courses simultaneously analysis. The performance analysis table of learning prediction results is shown in Table 2. From the comprehensive analysis indicators in Table 2, it can be seen that the multi-task learning prediction algorithm is superior to the single-task learning prediction algorithm, and the multi-task learning algorithm can simultaneously predict the scores of multiple courses. As learning data increases, the prediction index of task 2 performs better than that of task 1.

Table 1. Student behavior index data

student ID	Share index data			Python programming			Data structure		
	Historical grade point	Work and rest rules	Diet pattern	Engagement	Positivity	Reference degree	Engagement	Positivity	Reference degree
1	3.62	0.08	1.33	0.92	0.23	0.32	0.94	0.31	0.92
2	3.23	0.12	1.52	0.90	0.18	0.25	0.87	0.23	0.91
3	3.48	0.06	1.89	0.95	0.35	0.31	0.93	0.24	0.95

Table 2. Performance analysis table for learning prediction results

Task	Task 1			Task 2		
Index	Accuracy	Recall rate	Comprehensive analysis index	Accuracy	Recall rate	Comprehensive analysis index
Single task learning prediction	0.72	0.61	0.68	0.76	0.57	0.71
Multi-task learning prediction	0.76	0.62	0.73	0.82	0.60	0.81

Implementation of Academic Intervention

The prediction algorithm proposed in this paper was used to predict task 3. According to the prediction results, the students were divided into three groups: excellent students, good students, and students with learning difficulties, with 30 students in each group. The Poor learning group was divided into experimental group and control group with 15 students in each group, and the learning excellence group was also divided into experimental group and control group with 15 students in each group. The experimental group was intervened, while the control group and the good learning group were not intervened. For the experimental group students with learning difficulties, a combination of online and offline intervention was carried out. The intervention content included early warning, learning resource recommendation and learning content answering, and group resource recommendation was made for the experimental group students with excellent learning skills. The paired T-test was used to analyze the average scores of the second and third tasks, as shown in Table 3, the results of the first round of academic intervention,After the task three is over, conduct the second round of academic intervention according to the prediction. Table 4 shows the results of the second round of academic intervention. From Table 3 and Table 4, it can be seen that the students with excellent grades in the experimental group and the control group have little change. This is mainly because the students with excellent studies are familiar with the content of the assessment and cannot distinguish between high and low academic performances. The performance of the non-intervention students with learning difficulty group declined, while the performance of the students in the intervention group improved significantly, which shows that the teaching intervention has played a positive role in promoting students' performance.

DISCUSSION AND CONLUSION

This study analyzes students' life and learning status through three cycles: data collection, academic prediction and academic intervention. Through case analysis and comparison, we can see that on the one hand, the problem of small amount of data in the prediction process can be overcome by collecting

Table 3. Results of the first round of academic intervention

		Number of students	Average score for task 2	Average score for task 3	Variation	T	Sig.
Excellent learning group	Intervention	15	83.4	85.2	Increase	-0.963	0.231
	No intervention	15	83.5	84.1	Increase	-0.361	0.425
Good learning group	No intervention	30	72.8	73.5	Increase	-0.465	0.563
Poor learning group	Intervention	15	42.3	54.6	Increase	-3.243	0.002
	No intervention	15	41.5	39.1	Decrease	0.364	0.089

Table 4. Results of the second round of academic intervention

		Number of students	Average score for task 2	Average score for task 3	Variation	T	Sig.
Excellent learning group	Intervention	15	85.2	89.2	Increase	-1.327	0.042
	No intervention	15	84.1	85.6	Increase	-0.236	0.364
Good learning group	No intervention	30	73.5	72.6	Decrease	0.185	0.314
Poor learning group	Intervention	15	54.6	65.9	Increase	-2.57	0.004
	No intervention	15	39.1	38.3	Decrease	0.196	0.196

data between different courses, while that there is a correlation between courses offered in the same semester. On the other hand, through Intervention can improve students' academic performance.

1) Multi-angle data collection is helpful to understand students' behavior more comprehensively. This research collected data from multiple perspectives, such as students' life behavior and learning behavior, online learning behavior and offline learning behavior, classroom learning behavior and extracurricular learning behavior, then collected students' behavior events into students' behavior activity flow, and constructed students' behavior characteristics according to students' behavior activity flow. Collecting students' behavior data from multiple angles can achieve students' portraits more accurately, enabling students to understand themselves more comprehensively, and helping students discover existing problems in time.
2) The learning behavior data of different courses can improve each other's predictive indicators. This research uses single-task learning model and multi-task learning model to predict academic performance. The comprehensive analysis index of the multi-task learning model is significantly higher than the single-task learning model, and students have certain similarity in learning efforts of different courses. In order to improve students' academic performance, it is necessary to strengthen the cultivation of students' living habits and learning habits.
3) Academic interventions can effectively improve academic performance. Both the excellent learning group and the learning difficulty group have improved their academic performance after intervention and recommendation. The performance of all students in the excellent learning group has improved, indicating that this group of students has good living and learning habits. Meanwhile the improvement of students after the recommendation of learning resources are better than that of students without intervention, indicating that the recommendation has played a certain role in promoting their learning resources, and gradually guide them to become learning leaders. The performance of the students in the learning difficulty group improved significantly after intervention, and reached the level of passing the course, while the performance of the students without intervention maintained a downward trend. Therefore, for students with learning difficulties at the beginning of the course, they should be intervened in time, as soon as possible, some learning resources with low difficulty learning resources and learning paths can be recommended for them to stimulate their interests in learning and improve the learning effect.

In this study, a robust multi-task learning model was adopted to improve the predictive composite index of each course by integrating structured and unstructured data such as basic information, living habits, learning efforts and interest preferences, and academic intervention was conducted according to the predicted results to improve academic performance. Due to the small number of course samples and limited data collected, in the following work, first of all, the collection of traditional classroom students' learning data will be further strengthened, and then intelligent intervention tools will be adopted to reduce the workload of researchers and improve the timeliness of student intervention. Finally, different intervention strategies are designed according to the characteristics of courses and students.

ACKNOWLEDGMENT

This research was supported by National Educational Science Planning for the 13th Five-Year Plan of China (No. ECA180463) and the National Natural Science Foundation of China (No. 71972090).

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012) Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. doi:10.1145/2330601.2330666

Binyamin, S. S., Rutter, M., & Smith, S. (2019). Extending the Technology Acceptance Model to Understand Students' Use of Learning Management Systems in Saudi Higher Education. *International Journal of Emerging Technologies in Learning*.

Bopeng, L., Tiecheng, P., & Hong, Y. (2019). *Research on application of early warning of students' achievement based on data mining. Journal of Sichuan University*.

Cuibi, Shuliang, & Yong. (2020). Application of big data technology in blended teaching of college students: A case study on rain classroom. *International Journal of Emerging Technologies in Learning*.

Fang, Y., & Yanshen, L. (2018). Application Research of "User-Centered" Educational Data Mining. *e- Education Research*.

Faqin, S., & Rui, F. (2019). *A Research on Online Learning Achievement Factors Based on Learning Analysis*. China Educational Technology.

Fati, Baoyuan, & Shihua. (2019). The Research Framework of Learning Habits Dynamics Based on Big Data in Education. *China Educational Technology.*

Fati, W., & Zhijia, M. (2016). *The Design Research of Learning Outcomes Prediction Based on the Model of Personalized Behavior Analysis for Learners*. China Educational Technology.

Gaihua, W., & Gangshang, F. (2019). *Prediction of online learning behavior and achievement and design of learning intervention model*. Distance Education in China.

Hongyan, L., Shengfan, X., & Xuanshu, L. (2019). *Prediction model of distance education students' academic performance under the background of big data*. China Adult Education.

Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing Early Warning Systems to Predict Students' Online Learning Performance. *Computers in Human Behavior*, *36*, 469–478. doi:10.1016/j.chb.2014.04.002

Huiqiong, Z., Qiang, J., & Wei, Z. (2017). Empirical Research of Predictive Factors and Intervention Countermeasures of Online Learning Performance on Big Data-based Learning Analytics. *e-. Education Research*.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. doi:10.1016/j.compedu.2009.09.008

Mengmei, D., Minhua, W., & Xinjia, Y. (2017). *Research on teaching intervention based on academic achievement prediction*. Distance Education in China.

Qi, Z., Hongmei, W., & Lu, Z. (2019). *Predictions for personalized learning from a learning analytics perspective*. Distance Education in China.

Qing, W., Ruguo, L., & Quanyu, W. (2015). *Empirical Research on E-Learning Behavior based on Association Rules*. Modern Educational Technology.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. doi:10.1016/j.compedu.2007.05.016

Suartama, I. K., Setyosari, P., & Ulfa, S. (2019). Development of an Instructional Design Model for Mobile Blended Learning in Higher Education. *International Journal of Emerging Technologies in Learning*.

Tongtong, L., Luoying, H., & Rui, Z. (2016). *The Construction of a Learning Intervention Model Based on Big Data Analytics*. China Educational Technology.

Xiaoqing, G., Jinliang, Z., & Huiying, C. (2012). Learning analysis: Emerging data technology. *Journal of Distance Education*.

Xiaoqing, G., & Zhanjie, W. (2018). Learning technology system architecture enters 2.0 era. *Computer Education*.

Yi, C., & Jian, G. (2018). Orderliness predicts academic performance: Behavioral analysis on campus lifestyle. *Journal of the Royal Society, Interface*.

Yifu, J., Tao, W., & Zishi, Z. (2016). Design and Analysis of Learning Alert System in Big Data Condition. *China Educational Technology*, *349*(2), 69–73.

Yizhou, F., & Qiong, W. (2018). *Prediction of academic performance and risk: a review of literature on predpredicative indicators in learning analytics*. Distance Education in China.

Youzeng, L., & Hao, Z. (2018). *Smart Campus Big Data Education Application Based on Students' Behavior Analysis Model*. China Educational Technology.

Zhengce. (2019). http://www.gov.cn/zhengce/2019-02/23/content_5367987.htm

Zhijia, M. (2019). Design and Empirical Research of Learning Outcome Prediction Based on CIEO Analysis from the Perspective of Learning Computing. *e-*. *Education Research*.

Zhijia, M., Yuting, L., & Dahu, Y. (2018). The Design and Implementation of Learning Warning System based on Learning Behavior Data in Blended Learning Environment. *Journal of Distance Education*.

Zhiting, Z., Yanyan, S., & Hongchao, P. (2017). Understanding Cultural Implications of Big Data in Education. *e- Education Research*.

Zijian, C., & Xiaoliang, Z. (2017). *Research on Prediction Model of Online Learners' Academic Achievement Based on Educational Data Mining*. China Educational Technology.

Zuhui, H., & Quan, S. (2017). *Research on Internet behavior analysis and data mining of College Students*. Distance Education in China.