AI-based College Course Selection Recommendation System: Performance Prediction and Curriculum Suggestion

Yu Hsuan Wu
Department of Computer Science and Information Engineering
National Central University
Taoyuan City, Taiwan
boriscat999@gmail.com

Eric Hsiaokuang Wu
Department of Computer Science and Information Engineering
National Central University
Taoyuan City, Taiwan
hsiao@csie.ncu.edu.tw

***Abstract*—Recent advances of AI applications in various of industries have led to remarkable performance and efficiency. Driven by the great success of datasets and experience sharing, people are exploring more precious datasets with diverse features and longer time range. The promising reasoning information of well-curated student grade datasets is expected to assist young students to find the best of themselves and then improve their learning outcome and study experience. Through data and experience sharing, young students can have a better understanding of their learning condition and possible learning outcomes. Existing course selection systems in Taiwan which offer limited basic enrolling functions fail to provide performance prediction and course arrangement guidance based on their own learning condition. Students now selecting courses with unawareness of their expecting performance. A personalized guide for students on course selection is crucial for how they structure professional knowledge and arrange study schedule. In this paper, we first analyzed what factors can be used on defining learning curve, and discovered the difference between students with different properties and background. Second, we developed a recommendation system based on great amount of grade datasets of past students, and the system can give students suggestions on how to assign their credits based on their own learning curve and students that had similar learning curve. The result of our research demonstrates the feasibility of a new approach on applying big data and AI technology on learning analysis and course selection.**

Keywords—Course Selection, EDM, Score Prediction, Curriculum Recommendation.

I. INTRODUCTION

Artificial intelligence (AI), has been applied in numerous fields and industries in today's society, including marketing, education, security, healthcare, and more. Some AI applications was designed to assist making decisions. Since accumulating experience and extracting formation were very time consuming, we rely on AI to organize huge amount of data and produce a concentrated conclusion [11]. Also, the society has built many platforms for data sharing and information transmission, which allow the accessible data to have more diversity and time range. With the development of AI technology and the accumulation of data, the actual experience sharing is imaginable in the near future.

One way to improve the quality in higher education scheme is by predicting student's academic performance or observing their learning condition and thereby taking early precautions or adjustment to improve student's performance and teaching method. The relevant information or data is hidden with the educational dataset and it is extractable during data mining techniques [5]. In [13], an ITS student model stores information that is specific to each individual learner, which provides information like student's learning outcome and obstacle, and the student model plays a crucial role in planning the training path or adjusting learning method.

One valuable attribute of modern higher education system was discussed in [13], which is the flexibility of learning method and schedule. Each individual should adapt different training path according to what is suitable for them, in other words, a dynamic guidance should be applied to help learners adjusting their training or learning path when facing different learning outcome or expectation. In a university, a student's course selection usually represents his or her training path, which will determine what he or she will learn, and sometimes the difficulty of training path. So, besides the learning method or learning style which are difficult to quantify, a student's course selection is easier to observe, predict and adjust through data mining techniques.

We believe the above two aspects were both crucial in the modern education environment. Academic performance prediction can provide students a better understanding of their learning condition, and student model could help them to adjust their study habit or curriculum. However, limited resources were provided in Taiwan's educational system. When students finished a semester, they cannot review advanced information other than merely numbers of final score. Without any extra information or guidance, an adjustment is hard to make even if the student realizes he/her is facing an obstacle.

Currently, in the research field, a realization of modern student model is actually one step far. A student model based on course selection should contain two components: Student performance prediction and adjustment suggestion. Previous studies provided us a well understanding of student's performance prediction [1, 2, 4, 5, 6]. However, previous studies failed to provide an overview with all students in a college, most of the studies of score prediction only focused on students with same properties like: department, enrollment year, class etc. So, in this paper, we studied the data of all undergraduate students in National Central University, which gave us a holistic perspective on educational data mining. For adjustment suggestion, [8] recently proposed a new approach on optional course selection recommendation. Which filled up one vacancy of modern student model. However, based on

978-1-7281-9362-5/20/$31.00 ©2020 IEEE
DOI 10.1109/IS3C50286.2020.00028

[15], a student's credit amount and other course selection attributes will also affect a student's performance, so the guidance for students on how much credits should they select and how to assign their credits to different courses is also crucial. In this paper, we implemented a course selection suggestion system with performance prediction feature, which can provide a guidance of how to assign credits based on different learning condition and expectations

II. Related Work

In the early 90's, how AI application can be used in education has gained a lot attention due to the progress of data mining and machine learning. An ideal AI mentor was first mentioned in [16], which can provide a personalized learning environment for each learner based on the reaction or response from the learner. Also, a new idea called student model was put forward. A student model could adjust the teaching style, training plan or guidance based on different learner.

In recent years, education data mining has gained a lot of attention. Several studies have accelerated the development of modern learning environment, and provided students and educational institutions a better understanding on student's learning condition. This approach has also been expanded to multiple aspects of education like performance prediction, study habit etc. Performance prediction was divided into many branches based on different predicting target. For example, predicting. the result of some important exam [2], predicting the risk of whether a student will fail on a course [5], predicting overall GPA [4, 6], and other tasks [1]. Many studies were collected and organized in [3], they have provided people a deeper understanding on score prediction. Which can tell that Decision Tree or NN-based algorithm usually has better performance. We could also tell that when using a student's past academic performance or his/her behavior in class as the input feature [3, 4, 5], the accuracy of prediction is satisfying.

However, in [4, 15], we can see a new approach on score analysis or prediction has been proposed. How [4] including the specific course that a student took in the input of their algorithm, which allows people start to view academic performance prediction in the aspect of courses. The analysis of [15] also enhance the understanding of the correlation between course selection and student's score. Later in 2019, a course selection recommender based on score prediction has been proposed [8]. Also, the idea of optimizing the course selection strategy to maximize the academic performance has been implemented, which pushed the realization of student model into a new level.

Combining the above studies, a modern course selection student model was close at hand. By implementing the student model based on course selection, the modern education environment could assist students to adjust their curriculum, and help them to obtain a better learning outcome.

III. Feature Selection

Recent investigators [15] have examined the correlation between several factors and a student's score. Fig 1. Shows that the amount of course has linear relationship with a student's score. Therefore, to find the balance between the course amount and academical performance could be the key of obtaining maximize result.

This research investigated 2322 undergraduate students enrolled between 2012 to 2016 in National Central University.

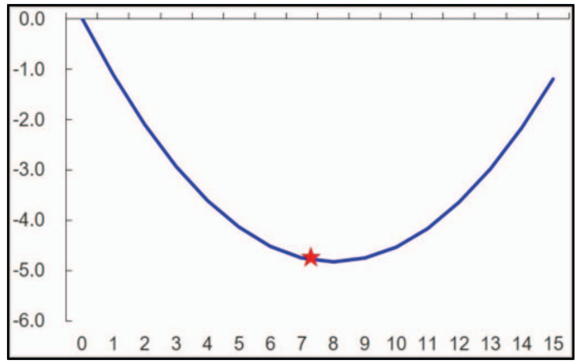

Fig 1. Correlation between course amount and score variation - Analysis of correlation between student's score and factors by Zheng, Bao-Zhi.[15]

Each row of the data represents one student's data of a semester, including these following features, department, current grade, semester, average score of a semester, total credits of this semester, major course credits, selective credits, general credits, whether a student changed his major, current class of a student, class rank of their last semester, the percentage of easy, medium and hard class credits, gender, living area(city), student's nationality, PE class amount, and morning class and afternoon class amount.

A. Student Gender

In [15], academical performance gap between male and female students has been discussed. Female students have higher average score than male students, female students averaging 80.93 out of 100 while male students averaging 76.41 out of 100. Also, from the perspective of percentile-rank (PR), female students also have big advantage to obtain a better PR, Table 1 shows the comparison between two genders.

B. Department

From the statistical aspect which Fig. 2 shows, students in the department of literal arts, management, hakka tends to have higher average score, while departments of science, engineering remaining lower average score. What is worth noted is that among these departments with higher average score, departments of literal arts do not reflect an obvious gender gap of average score, but department of management, hakka still shows a gender gap in average score.

C. Student's Academic Record

From the references [1, 3], we observed that it is a very common method to include a student's performance record as input feature. So, we did a linear regression with student's average score of the previous semester and their current average score as Table 2. Student's past average score does have moderate to strong correlation with their score. By including this feature, we believe it could improve our model.

TABLE I. Student gender versus chances of getting each PR

Gender	PR > 50	PR >75	PR < 25
Male	0.4	0.17	0.31
Female	0.598	0.3	0.15

TABLE II. Linear regression of past score and current score

Prediction Target	R^2
All students	0.44
Without senior students	0.54

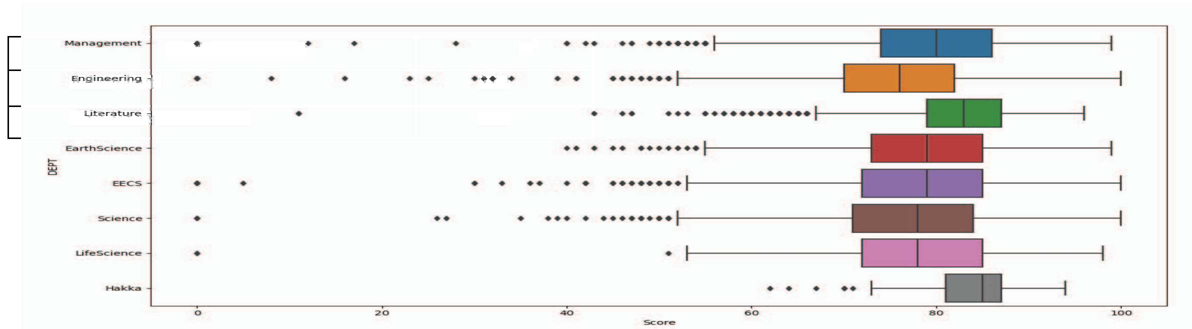

Fig 2. Student average score and their department (*x axis: 0, 20, 40, 60, 80, 100; y axis: from top to bottom: Management, Engineering, Literature, Earth-Science, Electrical Engineering and Computer Science, Science, Life Science, Hakka*)

Worth noted is that, a student's academic ability may not always reflect to their past score, award record, competition record could also show a student's learning ability, so how to quantify such abstract data is also an important topic of education data mining.

IV. Score Prediction and Curriculum Suggestion

A. Score Prediction

According to the organization of [3], we adapted DT as the prediction algorithm, since it has a more stable performance or prediction, we designed a progressive method which an gradually extract a student's learning condition update through semester. We totally trained eight models for score prediction, which maps to eight semesters in four years.

The difference between these models is the amount of feature, while the first model for predicting the first semester only gets 20 features, and the last model maps to the last semester gets 125 features as input, which includes a student's academic record.

For further discussion, we also adopted a cluster method(K-Means) before the prediction, which can divide the data into three parts, and then train and test the three models with divided data. However, this approach did not improve the performance, for this result, we have two assumptions. One is due to the clustering result, perhaps the current features did not perform well in K-Means, or the limited data after division actually dropped many information that model could learn, this assumption is possible to be verified after more data was collected.

B. Curriculum Suggestion

The whole system was designed to be implemented in the real course selection scenario, and the process though each the process was actually similar with a tree structure, so we used the same DT algorithm as score prediction section.

V. Experiment

The result of score prediction was showed in Table 4. The MAE of each semester presents a decline through the first to the sixth semester, which could inference adding the features we selected does increase the performance of model. However, the rise of MAE in last semester could be related to the result of Table 2 that the score of last two semesters could be impacted by other factors. For the course selection, the result was showed in following rows in Table 4. However, a difficult problem in this approach is to obtain the actual expecting score. For the alternative, we used the actual score as their expecting score, which can also extract the correlation between score and course selection information. The result showed that predicting a student's course selection through their record, information and expecting score is highly viable.

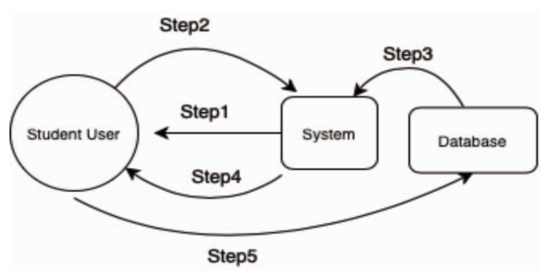

Fig 3. Workflow of how student users employ the system

TABLE III. Instructions of Fig. 3.

Step	Description
1	Output prediction score of the next semester.
2	If unsatisfied, input the expecting score.
3	Retrieve student's information and record from database.
4	Output course selection suggestion.
5	Store the record after the semester ends.

TABLE IV. Prediction MAE of course selection suggestion

Semester	1	2	3	4	5	6	7	8
Score Prediction	5.36	4.19	5.32	4.73	4.67	4.23	5.51	5.95
Total Credits	1.39	1.52	1.78	1.86	2.48	2.38	2.98	3.23
Lecture Hour	0.73	1.08	0.85	1.05	1.26	1.11	0.94	0.62
Selective Credits	0.65	0.76	0.91	0.78	1.33	1.04	0.84	0.67
General Credits	0.53	0.44	1.01	0.91	1.18	1.23	1.13	1.47
Morning Class amount	0.64	0.64	1.27	1.49	1.68	1.8	1.68	1.48
Afternoon Class amount	0.67	0.71	0.71	0.96	1.02	1.14	1.14	1.06

VI. Conclusion

This study completed a system which can be implemented in the course selection phase in NCU. Through this study, we not only attempted a new approach to assist students through course selection, but also showed that it is highly possible to predict a student's performance and course selection. For the future work, we expected to collect more diverse of feature like a student's club activity, attendance etc., and experiment with different time series algorithms, and to implement the system as a feature on the actual NCU course selection system.

Acknowledgment

We would like to thank all colleagues and student team members who contributed to this study. We are grateful to Dr. Chen for his assistance on data analysis, and to Miss Chung, for her advice on the work, and to my classmates Jerry Lee and James Lee for their assistance on data analysis. We thank the editor and series editor for constructive suggestions of an earlier version of this paper.

References

[1] Al-Sudani, Sahar & Palaniappan, Ramaswamy, "Predicting students' final degree classification using an extended profile," *Education and Information Technologies*, Feb., 2019, pp. 1-13.

[2] Ahamed, A T M Shakil & Mahmood, Navid & Rahman, Mohammad, "An intelligent system to predict academic performance based on different factors during adolescence," *Journal of Information and Telecommunication*, vol. 1, Apr., 2017, pp. 155-175.

[3] Kumar, Mukesh & Singh, "A. Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques," *International Journal of Education and Management Engineering*, vol, 6, Nov., 2017, pp. 40-49.

[4] J. Xu, K. H. Moon and M. van der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, Aug., 2017, pp. 742-753.

[5] Hamsa, Hashmia & Indiradevi, Simi & Kizhakkethottam, Jubilant, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm," *Procedia Technology*. vol, 25, Dec., 2016, pp. 326-332.

[6] Y. Meier, J. Xu, O. Atan and M. van der Schaar, "Predicting Grades," in *IEEE Transactions on Signal Processing,* vol. 64, no. 4, Feb., 2016, pp. 959-972.

[7] Rabia, Mahwish & Mubarak, Naima & Tallat, Hira & Nasir, Wajiha, "A Study on Study Habits and Academic Performance of Students," *International Journal of Asian Social Science*, vol, 7, Jan., 2017, pp. 891-897.

[8] L. Huang, C. Wang, H. Chao, J. Lai and P. S. Yu, "A Score Prediction Approach for Optional Course Recommendation via Cross-User-Domain Collaborative Filtering," in *IEEE Access,* vol. 7, 2019, pp. 19550-19563.

[9] Al-Badarneh, Amer & Alsakran, Jamal, "An Automated Recommender System for Course Selection," *International Journal of Advanced Computer Science and Applications,* vol, 7, no. 3, Mar., 2016.

[10] A. Slim, G. L. Heileman, W. Al-Doroubi and C. T. Abdallah, "The Impact of Course Enrollment Sequences on Student Success," *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*, Crans-Montana, 2016, pp. 59-65, doi: 10.1109/AINA.2016.140

[11] Rubens, Neil and Kaplan, Dain and Okamoto, Toshio, "E-Learning 3.0: Anyone, Anywhere, Anytime, and AI," *International Conference on Web-Based Learning*, Jan., 2014, pp. 171-180.

[12] [Carmona, Cristina and Castillo, Gladys and Millán, Eva, "Discovering Student Preferences in E-Learning,", *CEUR Workshop Proceedings*, vol, 305, Jan., 2007.

[13] Floriana Esposito, Oriana Licchelli, Giovanni Semeraro, "Discovering Student Models in e-learning Systems," *Journal of Universal Computer Science*, vol, 10, Jan., 2004, pp. 47-57.

[14] Pistilli, Matthew and Arnold, Kimberly, "In practice: Prudue Signals: Mining real-time academic data to enhance student success," *About Campus*, vol, 15, Jul., 2010, pp. 22-24.

[15] Zheng, Bao-Zhi, "Analysis of correlation between student's score and factors," *Newsletter of Office of Institutional Research of National Central University*, Jun., 2018. Retrieved from: http://ir.ncu.edu.tw/index.php/zh/2017-08-09-14-24-51/2018-04-12-06-11-58/173-1-201804.

[16] Woolf, B.: "AI in Education"; *Encyclopedia of Artificial Intelligence.* Shapiro, S. ed., John Wiley & Sons, Inc., New York, 1992, pp. 434-444.