Educational Data Mining and Analyzing of Student Learning Outcome from the Perspective of Learning Experience

SHU Zhong-mei
School of Education, Sun Yat-sen University
No. 135, Xingang Xi Road
Guangzhou, 510275, P. R. China
issszm@mail.sysu.edu.cn

QU Qiong-fei
School of Education, Sun Yat-sen University
No. 135, Xingang Xi Road
Guangzhou, 510275, P. R. China
quqf@mail.sysu.edu.cn

FENG Lu-qi
School of Education, Sun Yat-sen University
No. 135, Xingang Xi Road
Guangzhou, 510275, P. R. China
fenglq3@mail2.sysu.edu.cn

ABSTRACT

Student Learning Outcome (SLO) has been a hot issue in the research fields of higher education quality assurance and institutional research. Based on the classic college student development theories and students' learning experiences, this paper combines two educational data mining techniques, regression analysis and neural network modeling to explore the influential mechanism of SLO by the means of empirical analysis. Finally, from the perspective of learning analytics and considering students' individual factors and university factors, a predication model of SLO is constructed to provide universities with essential reference to enhance SLO and teaching quality.

Keywords

Student Learning Outcome; Educational Data Mining; Learning Experience; Learning Analytics

1. INTRODUCTION

SLO has been a hot issue in the international research fields of higher education quality assurance, providing important evidence to reflect the educational effectiveness of a university [1]. A major part of the assessment of SLO is to obtain information and evidence of learning outcomes with qualitative or quantitative measurement methods. However, the diverse categories and levels of universities and colleges, as well as the complicated learning objectives and learning process, result in the multi-dimension and complexity of college-SLO.

It is recognized that the academic research on SLO has made a marked progress. In particular, research that explores models of SLO and their influential mechanism regarding the environments and conditions of schools, students' individual characteristics and student engagement in learning, contribute significantly to the decisions on how to promote SLO. Yet, there is little research examining SLO from the perspective of learning experience, for instance, student emotion, behavior and cognition. Empirically, students' participation in school activities and their rich experiences will influence their learning outcomes to a certain extent. And through those activities and experience they will have various cognitive experiences and respond with different emotional reactions. The data involved are huge and extensive that it becomes difficult for traditional statistical methods to discover the hidden laws. As a result, emerging data statistics and analysis methods, such as educational data mining, are urgently needed. In order to better understand the influential factors of SLO, this research uses stepwise regression analysis and neutral network to mine data of SLO, and reconstruct the meaning of mining results from the perspective of learning analytics. The ultimate purpose is to establish the intricate relationship between student learning experience, school characteristics, student characteristics and SLO, and thereby provide crucial reference for improving the training quality of university talents.

2. INSTRUMENT OF THE SURVEY

Research has shown that questionnaire survey precisely reflects the overall level of SLO by indirectly measuring learning outcomes with students' self-reports. This study devises the Sun Yat-sen University Student Learning Status Survey based on Astin's and Pace's student survey assessment models [2, 3]. Students' relevant experiences are emphasized, for instance, how much they involve in learning and work hard on it, and their interactions with teachers. Besides, students' emotional learning outcomes and general-knowledge education outcomes are also included. The survey divides student learning experiences and outcomes into 6 dimensions, illustrated in Figure 1. The coefficient of internal consistency among items in each dimension is above 0.9, demonstrating a high degree of reliability. This research analyzes the university and student factors' impacts on SLO in the logical framework of "Inputs-Learning-Outcomes", and from the standpoint of the broad learning experiences like student emotion, behavior and cognition.

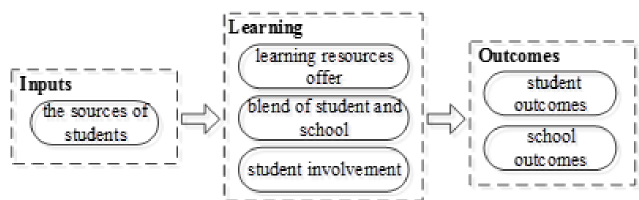

Figure 1. The Instrument of the Survey.

3. ANALYSIS

Educational data mining shows its potential value in determining the influential factors of SLO, such as identifying student characteristics and the dimensions of learning experiences that really affect learning outcomes in datasets. This paper uses educational data mining to deal with SLO data, interprets the mining results from the perspective of learning analytics, and explores the influential mechanism of SLO imposed by student learning experience, school characteristics, and student characteristics.

Data of this study came from the campus-wide online survey officially conducted in 2012, which was part of the Sun Yat-sen University Student Learning Status Survey Project, covering 36 departments and 33,000 undergraduates. These students completed the items in the questionnaire under the circumstance without stress. A total of 7,051 questionnaires were returned with a 21.2%

response rate, representing a considerable satisfaction compared with the international response rate of questionnaires. This research focuses on the samples of undergraduates, and selects 6,673 effective questionnaires out of the total returned with an effective rate of 94.6% based on principles such as response time and questionnaire quality standard.

3.1 Determining the Influential Factors of SLO with Regression

Stepwise regression provides an approach to identify the concrete experiences relevant to SLO. Specifically, it groups together similar items among the total 227-item in the survey, identifies those related to learning outcomes using forward and backward stepwise regression, thoroughly examines the residual plot and the diagnostics, and ultimately determines 17 independent variables in the multivariate regression model. 4 dimensions with 17 variables in student learning experiences are important factors affecting SLO, which are respectively the availability of learning resources, student involvement in learning, campus culture and school outcomes, including university factors like assessment of coursework and major study experience, guidance of academic norms, equal cultures and the atmosphere of cultivating multiple abilities, together with student factors like self-regulated learning, activity engagement, extra-curricular reading, thesis writing, peer communications, discussion contents, student-teacher interactions, academic activities and allocation of personal spare time. While school outcomes combining student and university factors, for instance, satisfactions towards school experiences and capability cultivation, as well as overall satisfaction, also have certain influential power on student learning outcomes.

These findings are consistent with Vincent Tinto's theoretical model on dropout problems of college students [4]. Whether students could obtain better learning outcomes depends on how well they fit their own experiences and objectives into the academic and social systems within the school system.

3.2 Optimizing Predictions of Learning Outcomes by Neutral Network Modeling

On the basis of the above analysis and for the purpose of increasing predictive accuracy, this study optimizes predictions of learning outcomes by neutral network modeling. The authors consider the influential factors of SLO determined by the regression analysis as experts' prior knowledge, and then further promote the intelligent processing through optimization model of neutral network, achieving the optimized prediction of SLO.

The number of input-layer nodes is determined by the number of factors affecting SLO, that is, 17 independent variables identified by regression analysis are considered as the input-layer nodes in the neutral network. Meanwhile, SLO is identified as the output-layer node. Finally, the optimized number of hidden-layer nodes is set to be 7 with the method of cut-and-trial. As such, the optimum topological structure for predicting SLO based on the neutral network model is 17-7-1.

Among the 6,673 effective questionnaires from the student learning condition survey, 8 samples with missing values have been eliminated. The 6,665 samples left are divided into two subsets, with 4,680 training samples (70.2%) for network model training, and 1,985 testing samples (29.8%) for testing whether the model meets the fundamental function required. Given the minimum relative change of training deviation is .0001 and that of training error is .001, and considers 1, 985 samples as testing samples. Through experiments, the output O represents the prediction of SLO, which ranges within [0, 1]. According to the overall evaluation of SLO by the 5-level rankings, i.e., A($0.9<=O$), B($0.8<= O <0.9$), C($0.7<= O <0.8$), D($0.6<= O <0.7$) , E($O <0.6$), the output has a sound consistency and accuracy with the self-assessment results of learning outcomes by the respondents. Table 1 shows a random list of the comparisons between network output and students' self-assessment results. Due to the large number of samples, Table 1 does not exhibit all the results.

Table 1. Sample Comparison

Testing Sample	Output	Expected	Relative deviation	Level
Sample 1	0.53	0.53	0.00	E
Sample 2	0.71	0.72	0.01	C
Sample 3	0.64	0.66	0.02	D
Sample 4	0.63	0.62	0.01	D
Sample 5	0.73	0.72	0.01	C
Sample 6	0.66	0.65	0.01	D
Sample 7	0.82	0.83	0.01	B
Sample 8	0.92	0.92	0.00	A
Sample 9	0.90	0.91	0.01	A

4. CONCLUSIONS

Predictive model of student learning outcomes which is constructed by the combination of neutral network and experts' prior knowledge established in the regression analysis, has a simple structure, better objectivity and accurate predictive effects. The trained neutral network model mentioned above could be used to scientifically and appropriately predict and evaluate student learning outcomes.

5. ACKNOWLEDGMENTS

We would like to acknowledge the funding on Key project of Ministry of Education (China) grant DIA130303 as well as funding on NSFC grant 61202345.

6. REFERENCES

[1] Terenzini, P.T. Assessment with open eyes: pitfalls in studying outcomes [J]. The Journal of Higher Education, 1989, 60(6): 644-64.

[2] Astin A W. Assessment for Excellence: the Philosophy and Practice of Assessment and Evaluation in Higher Education [M]. Phoenix, AZ: Oryx, 1993.

[3] Pace,C.R. Quality, Content, and Context in the Assessment of Student Learning and Development in College[M]. Washington, DC: Office of Educational Research and Improvement. 1986.

[4] Tinto, V. Dropout from Higher Education: A Theoretical Synthesis of Recent Research [J]. Review of Educational Research, 1975 (45):89-125.