Tseng, S.-F. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Learning Analytics: An Enabler for Dropout Prediction

Shu-Fen TSENG[ad]*, Chih-Yueh CHOU[bd], Zhi-Hong CHEN[cd], Po-Yao CHAO[cd]
[a] *Department of Information Management, Yuan Ze University, Taiwan*
[b]*Department of Information Science & Engineering, Yuan Ze University, Taiwan*
[c]*Department of Information Communication, Yuan Ze University, Taiwan*
[d]*Innovation Center for Big Data and Digital Convergence, Yuan Ze University, Taiwan*
*gssftseng@saturn.yzu.edu.tw

Abstract: A key application of learning analytics is predicting students' learning performances and risks of dropping out. Heterogeneous data were collected from selected school to yield a model for predicting student's dropout. Results from this exploratory study conclude dropout prediction by learning analytics may provide more precise information on identifying at-risk students and factors causing them to be at risk.

Keywords: Learning analytics, dropout, predictive model

1. Introduction

It is gradually recognized the effective use of data and learning analytics are both critical components of digital learning strategies to personalize learning for needed students to increase student retention and achievement in schools (Bienkowski, et al., 2012). Analytics applied to education data can help schools and school systems to better understand how students learn and succeed. Significant improvements in technology tools and resources, and the focus on meeting the needs of individual students through personalized and digital learning have together provided an emerging context in which education systems have the opportunity to advance the values of learning analytics to truly inform teaching and learning.

A key application of learning analytics is monitoring and predicting students' learning performance and addressing potential issues early so that interventions can be provided to identify students who are at risk of failing a course of program of study (Bienkowski, et al., 2012). In particular, one of the key successes of learning analytics is predicting which students are at risk of dropping out. Tobin and Sugai (1999) demonstrated that using data on disciplinary referrals during middle school can predict those who are likely to drop out of high school. Their research supports the use of school records of discipline referrals as a screening device. The work by Bowers (2010) shows that by analyzing student grade history from k-8 or k-12, it can correctly identify over 80% of dropout students.

One of noteworthy cases is that IBM worked with Mobile County Public Schools in Alabama to apply analytics to education data to help school system identify which students were at risk of dropping out. Rapid access to information and analytics means school administrators and principals can make more informed decisions and take appropriate and timely action to develop an individualized response to each student's problems and monitor their progress. Timely intervention based on real-time information is helping Mobile County to keep students on the right path and to lower dropout rates (Centre for Information Policy Leadership, 2013; IBM, 2009).

It has long been recognized by school authorities in Taiwan that dropout students mostly come from disadvantage family (e.g. single parent, minority), who lack of self-discipline, are not interested in school work, and easily befriend juvenile delinquents. There were a total of 5,379 dropouts among k1-k12 students in the year of 2012 in Taiwan. Personal profiling outlined mostly of them dropped out during their middle school periods (89%). Among these dropouts, 58% of them came from single parent household. Major causes of their dropout were categorized into individual reasons (47%), followed by family (24%) and social (17%) problems (MOE, 2012).

School administrators and educators can only analyze these students when they had already dropped out of school. In contrast of taking remedial actions, the newly development of learning analytics could help predict in advance and provide the tools to calculate a student's risk of dropping out of school. In this exploratory study, we collect students' data from heterogeneous sources (ie. grade history, disciplinary referral, class attendance) and employ logistic regression model to identify risk factors of dropping out in one selected school. Then, we compare the findings to profiles of dropout that outlined by conventional method. We further discuss the potential benefits of learning analytics on dropout prediction.

2. Method

Taoyuan county ranks the second highest dropout rate in Taiwan. In this study, we collected data from one middle school in Taoyuan county. The school selected was among one of the highest dropout rate schools. Heterogeneous sources of administrative, academic, and disciplinary data of every student in this school were collected and organized into a database. Descriptive and explanatory statistics were used to determine the relationship between a dependent variable (whether a student was dropped out in current term) and independent variables (e.g. students' family background, academic grade, class absence, and disciplinary referral records in previous term). In particular, logistic regression analysis was employed to create a model for predicting whether a student was at risk of leaving school.

3. Findings

In the year of 2013, there were a total of 712 seventh and eighth grade students in this middle school. Among them, 11 students dropped out of school. Five of them were males and four out of 11 students were minorities. Correlation coefficients showed that dropout was positively related to number of class absence ($r=.363$), number of disciplinary referral ($r=.253$) in previous term, and it was negatively related to academic grade ($r=-.06$) of previous term. Despite the limited number of dropout cases, logistic regression model (table 1) still showed prediction power of disciplinary referral (odd ratio=1.136) and academic grade (odd ratio=.932) of previous term on a students' risk of dropping out of school. Number of class absence in previous term and gender showed insignificant impacts on predicting dropout.

Table 1: Logistic regression of dropping out of school by risk factors.

	Dropout		
	B	Exp(B)	Effect on dropout risk
Constant	.332	1.394	
Male	-.409	.664	
Class Absence (previous term)	.013	1.013	
Disciplinary referral (previous term)	.127	1.136*	Increase risk by 13.6%
Academic grade (previous term)	-.07	.932**	Decrease risk by 6.8%
Cox & Snell R2	.045		
N	712		

4. Discussion

The conventional way of outlining high risk student of dropouts is to analyze demographic and family characteristics of those who had already dropped out. In the year of 2013, among those dropout students in Taoyuan county, 47% of dropouts were from single parent household, 24% of them were minorities, and 6% of them were raised by grandparents. Based on this conventional method, individual and family reasons were ranked the highest causes of student to drop out of school.

By employing method of learning analytic, we collected and organized data from personal and family composition, academic performance and disciplinary records; our finding suggests that we can predict a student's risk of dropping out of school mainly based on his/her school related activity records of previous term. In particular, high number of disciplinary referral and poor academic grade were major factors that help to predict student's risk of dropping out of school. The results suggested instead of targeting individual reasons and family background, student's activities in school can be effectively used to predict a student's risk of dropping out of school. With more longitudinal data of students are collected and more middle schools join the project, we could build a better fit model for dropout prediction and calculate risk of dropout for students. This exploratory study concludes that dropout prediction from learning analytics may provide more precise information on identifying at-risk students and factors causing them to be at risk.

References

Alliance for Excellent Education (2014). Capacity enablers and barriers for learning analytics: Implications for policy and practice. Retrieved from http://all4ed.org/reports-factsheets/capacity-enablers-and-barriers-for-learning-analytics-implications-for-policy-and-practice/

Baker, R.J.d. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (eds.), Handbook on Innovations in Learning (pp.179-109). PA: Temple University.

Bienkowski, M., Feng, M., and Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. Retrieved from http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

Bowers, A.J. (2010). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. Practical Assessment, Research & Evaluation, 15(7), 1-18.

Center for Information Policy Leadership (2013). Big data and analytics: Seeking foundations for effective privacy guidance. Retrieved from http://www.hunton.com/files/Uploads/Documents/News_files/Big_Data_and_Analytics_February_2013.pdf

IBM (2009). Mobile county schools applied analytics to transform student performance. Retrieved from http://www.ibm.com/smarterplanet/us/en/leadership/mobilecounty/

Ministry of Education (2012). Dropout statistic in Taiwan. Executive Yuan, Taiwan.

Romero, C. and Ventura, S. (2010).Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, 40(6),601-618.

Siemens, G. and Baker, R.J.d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. Retrieved from http://users.wpi.edu/~rsbaker/LAKs%20reformatting%20v2.pdf

Tobin, T.J. and Sugai, G. M. (1999). Using sixth-grade school records to predict violence, chronic discipline problems, and high school outcomes. Journal of Emotional and Behavioral Disorders, 7(1), 40-53.