A Systematic Review of Studies on Predicting Student Learning Outcomes Using Learning Analytics

Xiao Hu
Faculty of Education
University of Hong Kong
Pokfulam, Hong Kong
xiaoxhu@hku.hk

Christy W.L. Cheong
Macao Polytechnic Institute
Rua de Luis Gonzaga
Gomes, Macao
wlcheong@ipm.edu.mo

Wenwen Ding
Faculty of Education
University of Hong Kong
Pokfulam, Hong Kong
wynn@connect.hku.hk

Michelle Woo
Faculty of Education
University of Hong Kong
Pokfulam, Hong Kong
mihlwoo@gmail.com

ABSTRACT

Predicting student learning outcomes is one of the prominent themes in Learning Analytics research. These studies varied to a significant extent in terms of the techniques being used, the contexts in which they were situated, and the consequent effectiveness of the prediction. This paper presented the preliminary results of a systematic review of studies in predictive learning analytics. With the goal to find out what methodologies work for what circumstances, this study will be able to facilitate future research in this area, contributing to relevant system developments that are of pedagogic values.

CCS Concepts

• Information systems → Information systems applications
• Applied computing → Education.

Keywords

Systematic review; prediction; methods; performances; learning outcomes; learning context.

1. INTRODUCTION

Existing studies in predictive learning analytics (LA) significantly varied in the techniques being used, the learning environments in which they were situated, and the resultant effectiveness of the prediction models developed. In this study, we collected 39 empirical studies in this field for a systematic review to find out what methodologies work for what circumstances. This study is similar to [1], to some extent, which aimed to find out the types of features important for predicting student learning outcomes and the prediction algorithms used for the purpose. In addition to these methodological considerations, we also examine the contexts, the targets and the performances of the prediction so that effective matching of prediction methodologies and circumstances can be enabled. This study, therefore, aims to facilitate researchers in identifying methodologies to develop prediction models for their purposes. Findings will be useful for future research on enhancing prediction model performance. This study will also contribute to the development of LA systems that use prediction to improve teaching and learning. This paper presented preliminary findings from this study for discussion and planning for subsequent further enhancement.

LAK '17, March 13-17, 2017, Vancouver, BC, Canada
ACM 978-1-4503-4870-6/17/03.
http://dx.doi.org/10.1145/3027385.3029438

2. METHODS

As LA is an interdisciplinary field of study, we conducted an extensive literature search in related disciplines including Computer Science, Electrical Engineering, Education, Learning Information System, and Management. We looked for peer-reviewed journal or conference papers published between 2002 and 2016. These papers should present studies that developed prediction models for student learning outcomes and tested their predictability on empirical data. To select papers for this study, we skimmed through the papers found in the search. Papers of the following categories were excluded: (1) review articles and position papers, (2) studies on intelligent tutoring systems, (3) studies focusing on factors influencing student performance, and (4) papers published in languages other than English.

We developed a coding framework with regard to (1) the teaching and learning environment, (2) the type of learning outcome(s) being predicted, (3) the features extracted for use in the prediction, (4) the predictive algorithms being used, and (5) the performances of the prediction on holdout, unseen data. This framework was developed iteratively using the constant comparative method [2, 3]. We looked at each individual paper retained after the filtering process and then identified/modified the categories in each aspect stated above. The papers were coded by two coders. The results were aggregated and analyzed. Thirty-nine papers[1] were coded so far and the preliminary results are presented below.

3. OVERVIEW OF PAPERS ANALYZED

We identified 45 teaching and learning environments and 14 prediction targets. Over half of the environments identified were online/blended learning and the datasets presented were mostly from electronic sources. Course performance was most commonly noted as prediction targets, followed by student retention/dropout. Course performance was predicted in binary terms (i.e. successful / unsuccessful) or in grades (i.e. A/B/C/D/E) for most cases.

329 types of features in total were used for the prediction in the papers, with 8.44 types being used in average across papers. Note that one feature type can consist of multiple features (each of which is equivalent to a column in a data table). Among the feature types used, 151 unique feature types were identified and they were of the following categories: (1) demographic features, (2) student history record and performance, (3) student record and performance in current course, (4) activity and course features, (5) learning behavior features, (6) self-reported features, and (7) others / unclear features. The top three feature categories having the most feature types included learning behavior features, student record and performance in current course, and demography features. The number of unique feature types in each of these

[1] Papers analyzed is listed here: https://goo.gl/u6HHJL

feature categories tended to be half the size of the feature types used in the papers or even less, except for activity and course features, where the reduction was relatively less significant. This observation indicates that there were significant overlaps on features used among these studies.

To maintain comparability, we selected accuracy, the measure adopted in most of the papers, as the medium of prediction performance comparison (8 papers were therefore excluded). 14 different algorithms were accounted. The top five algorithms used in most papers were Decision Tree (DT), Neural Network (NN), Clustering-based classification, Rule-based algorithms, and Naive Bayes (NB). The papers tended to use more than one algorithm for the prediction. The total number of algorithms accounted was 115, resulting in an average of 3.71 algorithms per paper. The reported accuracies were within the range of 60-98% for most cases, indicating the prediction performance was overall better than that of a random guess approach in binary predictions.

4. PREDICTION PERFORMANCE COMPARISON

We adopted the following performance benchmarks to categorize the papers into two tiers. As each paper may have used multiple algorithms, only the best performance in each paper was considered for this comparison. Papers satisfying one of the following criteria are regarded as Tier-1 for their superior prediction performances (23 papers), others as Tier-2 (17 papers). One paper was categorized into both tiers because it had two different prediction targets.

(1) For Accuracy, Detection Sensitivity, Mean Success Rate and R^2, the best performing algorithm used reaches 90% or above.
(2) For Root-Mean-Square Error (RMSE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Deviation (MAD), the best performing algorithm used reaches 10% or below.

Tier-1 papers involved 26 environments and 27 datasets. Online/ blended learning environments accounted for 77% while all the datasets were drawn from electronic sources. Tier-2 papers involved 19 environments and 20 datasets. Online or blended learning environments were less (58%). While electronic sources were still predominated in Tier 2, these papers also drew on data from non-system-based sources including questionnaire data (10%) and data from external source (5%). This seems to suggest that online environments and electronic data sources may be preferable for better prediction performance.

35% of the Tier-1 papers predicted course performance in binary terms, 22% course final grades, and none course final scores. The respective percentages in Tier 2 were 25%, 15% and 15%. This outcome was reasonable as the prediction tasks are easier when there are fewer available options. Over 60% of the papers predicting course performance in binary terms and in grades were in Tier 1 while 57% of the papers predicting student retention/dropout were in Tier 2. This implied that course performance in binary terms and in grades seemed more predictable as compared to student retention/dropout.

The sample size was generally larger in Tier 1 than Tier 2. 50% of the samples used in Tier 2 were less than 200 while the samples of such size in Tier 1 only accounted for 30%. In other words, Tier-1 papers generally had more examples to train the prediction models. This could be one of the possible reasons for the consequent better performance. The prominent feature types and their ranking in Tier 1 and Tier 2 papers remained the same as the overall picture across all papers. The only exception was self-reported features, which were not used in Tier 1 where learning behavior features were used more often (Tier 1: 3.39 types/paper; Tier 2: 2.25 types/paper). As learning behavior features are automatically tracked by systems in online environments, these findings were in line with the observed prominence of online/blended environments and electronic data sources among Tier-1 papers.

As each paper may have used more than one algorithm to build prediction models, we compared the accuracies attained by all models used in the papers. With 90% accuracy as the benchmark, 30 prediction models were classified into Tier 1 and 85 into Tier 2. In line with their high frequencies of usage, DT, NN and Clustering-based classification had been used most often in both tiers. While the frequencies of NN and Clustering-based classification were similar in Tier-1 models, NN were used more often than Clustering-based classification in Tier 2. Rule-based algorithm, despite its frequent usage, was not used in any of the Tier-1 models while NB had just two such cases. In follow-up studies, the models will be further analyzed in conjunction with the contexts to find out possible interactions of these factors.

5. CONCLUDING REMARKS

This paper presented the preliminary findings of a systematic review of studies in predictive learning analytics, with an objective to find out what prediction methodologies work for what circumstances. At this stage, we found that existing studies tended to predict course performance (successful/unsuccessful), course grades and student retention/dropout in online/blended learning contexts using data drawn from electronic sources. This is understandable as the data are objective and are usually available in a large amount under these circumstances. This explanation is also valid for the prominent feature types accounted in papers with superior prediction performance. Interestingly, self-reported features were not in use in any of the papers with superior prediction performance. Possible reasons might include: 1) these data are hardly scalable as they need extra efforts to collect; and 2) they are subjective in nature which may affect their reliability. As for prediction algorithms and prediction performance, we found that no experiments with superior prediction performance used rule-based algorithms, even though they were used in 29% of the reviewed papers. Decision Tree, Neural Networks and Clustering-based classification were the most frequently used prediction techniques. In the next stage, we will delve further into the prediction experiments to find out what leads to (un-)promising results and examine if there are any correlations between the methodologies and the circumstances in which the studies were situated. We will also continue to expand the pool of papers being analyzed in order to enhance the generalizability of the findings.

6. REFERENCES

[1] Shahiria, A. M., Husaina, W., & Rashida, N. A. 2015.A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422. doi:10.1016/j.procs.2015.12.157

[2] Strauss, A. L. 1987. *Qualitative Analysis for Social Scientists*. Cambridge University Press, London.

[3] Strauss, A. L., & Corbin, J. 1998. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory,* 2nd ed. Sage Publications, Thousand Oaks, CA.