Using Learning Analytics for Providing Personalized Content and Feedback in Large Classes

Hassan Khosravi hkhosrav@cs.ubc.ca and Kendra Cooper kendra.m.cooper@gmail.com

Introduction

Educational environments continue to evolve (traditional, on-line, blended, flipped classrooms, MOOCs) to better meet the needs of diverse students, often enrolled in large classes. As the class size grows providing timely, personalized feedback to individuals becomes more difficult for the instructor, as does recommending personalized content to help each student effectively master the learning objectives. Currently, little is available to assist instructors on this issue.

A novel framework based on learning analytics is introduced that has the potential to use available data on students to create a more personalized learning experience tailored towards each individual’s needs. Part of the framework is presented in more detail, which generates personalized feedback for students. Historical data from an engineering course at UBC is used in the validation.

Flipped Course Example

All engineering students at UBC must take an introductory course (APSC 160) in C programming, which focuses on program design and problem solving. This course is fully flipped: students are provided with screencasts (voice over PowerPoint) that introduce the material to be covered in the subsequent class.

The data collected spans lectures, labs, and midterm material:

- For each lecture, we have used: (1) the number of times screencasts were watched, (2) the grade received for the in-class clicker quiz, (3) the grade received for the in-class group exercises, and (4) a Boolean indicator of whether or not the provided sample solution was accessed by the student.
- For each lab, we have used: (1) the number of times they viewed the content of the pre-lab, and (2) the grade received for the lab.
- For each student, we have used the two midterm exam grades, and 6 Boolean indicators on whether the files on practice midterms and solutions were accessed.

Machine Learning Engine

Supervised Learning: for the studies on identifying the weaknesses and strengths of students, we used Weka’s implementation of linear regression with 10-fold cross-validation.

Unsupervised Learning: for the studies on uncovering relationships and patterns, we used Weka’s implementation of k-means.

Overview of the Personalized Education Framework

Student Data

Current students
Interaction with the learning environment
In-class activities
Collective and summative assessments
Discussion boards
Pre and post surveys
Interviewing students
Ongoing marking exams
Students’ status (motivation level, emotion)
Current students from other courses
Previous offerings of the course

Intelligent Learning Assistant

Machine learning tool
Supervised learning
Unsupervised learning

Capabilities

Identify weaknesses and strengths of students
Uncover interesting relationships and patterns and outliers
Find correlations among various factors of the course
Identify general misconceptions
Find students with similar behavior

Outcomes

Provide personalized content and feedback
Provide general feedback
Improve/add teaching resources
Recommend study buddies to students

Flipped Course Example

For each student, we have used the two midterm exam grades, and 6 Boolean indicators on whether the files on practice midterms and solutions were accessed.

For each lab, we have used: (1) the number of times they viewed the content of the pre-lab, and (2) the grade received for the lab.

Identifying Weaknesses and Strengths of Students

(data) that they leave behind, which could be in many forms. In this paper we consider data produced thorough their interaction with the learning environment as well as their performance in formative and summative assessments.

How well students learn the content of a block can be predicted by the footprints (data) that they leave behind, which could be in many forms. In this paper we consider data produced thorough their interaction with the learning environment as well as their performance in formative and summative assessments.

Comparing the results of a few different approaches for predicting the grade of the second midterm:

- Total number of screencasts views (1-Sc):
- Screencast views (k-Sc):
- Summative assessments (Sa):
- Summative assessments (k-Sa + Sa):
- Personalized feedback model (Pf):

The following characteristics are a good representative of students in each of the clusters, which are supported by students’ in-class clicker quiz and lab scores.

G1: Consists mostly of students that work hard, do all exercises and attend in-class lectures, and still perform relatively well.

G2: Consists mostly of students that work even harder, watch even more screencasts, but perform poorly (averaging around 34 points out of 60).