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.

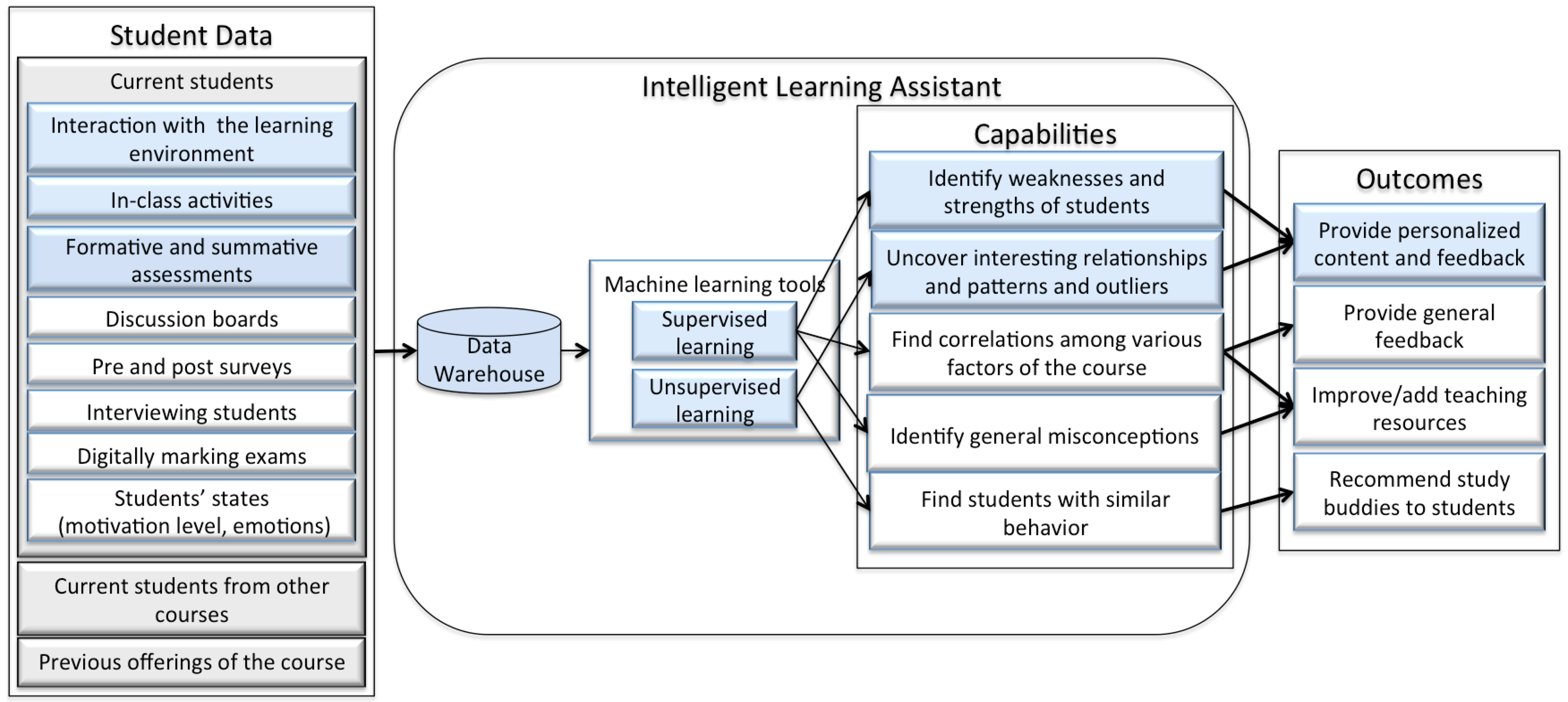
•For each student, we have used the two midterm exam grades, and 6 Boolean indicators on whether the files on practice midterms and solutions were accessed.

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.

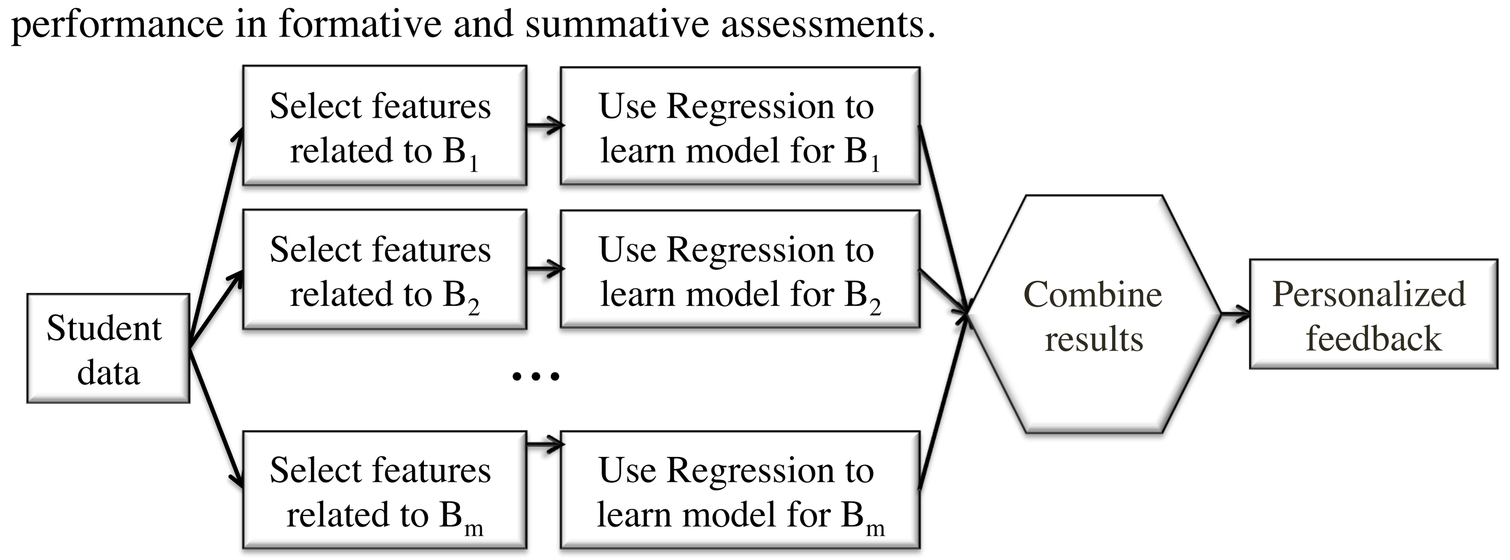
Unsupervised Learning: for the studies on uncovering relationships and patterns, we used Weka's implementation of k-means.

Overview of the Personalized Education Framework



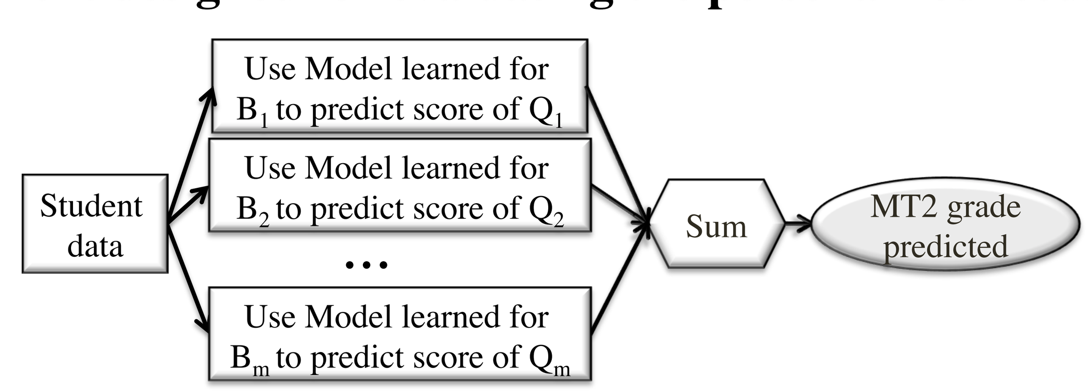
Identifying Weaknesses and Strengths of Students
(using supervised learning)

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.



Personalized feedback
- Covers progression on learning objectives that is going well and needing attention
- Identifies one or more tasks for the student to work on (e.g., material to review, problem exercise)

An experiment designed for evaluating the personalized feedback model



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)
•Screencast views + summative assessments(k-Sc + Sa)
•Personalized feedback model (Pf)

	1-Sc	k-Sc	Sa	k-Sc + Sa	Pf
Correlation coefficient	0.13	0.32	0.66	0.68	0.71
Root relative squared error	98%	94%	75%	73%	68%

Uncovering Relationships and Patterns
(using unsupervised learning)

Students who infrequently participated in lectures.

	C1	C2	C3
# students	11	17	6
Normalized in-class mark	0.37	0.32	0.23
Screencast views	265	299	813
Normalized Worksheet solution	0.64	0.59	0.86
Exam access	1.00	0.24	0.67
Midterm two score	48.09	36.41	35.33

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.

•C1: Consists mostly of students that have some prior programming experience, which is why they are able to watch fewer videos than the rest of the class, participate in fewer lectures, and still perform relatively well.

•C2: Consists of mostly students with little or no programming experience that are trying to put in minimum effort to pass the course. Similar to students from C1, they watch fewer videos than the rest of the class, participate in fewer lectures; however, their grade is significantly lower than those in C1.

•C3: Consists of students that have no prior programming experience, who intend to replace lectures with screencasts. Even though they watch far more screencasts than the rest of the class and consistently check the provided solutions, they are doing very poorly on the exams. These students are not studying the content effectively, and are not grasping the material from the screencasts alone.

Students who watch an overabundance of screencasts.

Data from previous study indicates that some students are substituting lecture time with screencasts, but this does not mean all students who watch an excessive number of videos are skipping out on lectures

	G1	G2
# students	34	30
Normalized in-class mark	0.95	0.78
Screencast views	872	1018
Normalized Worksheet solution	1	0.98
Exam access	1	0.76
Midterm two score	51.18	33.87

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 and participate in almost every lecture, and in turn are rewarded (aver- aging over 51/60). These are the motivated keeners that are willing to work hard to get a good grade.

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