Predicting student performance from multiple data sources

Irena Koprinska, Joshua Stretton and Kalina Yacef

School of Information Technologies, University of Sydney, Sydney, Australia

{irena.koprinska,joshua.stretton,kalina.yacef}@sydney.edu.au

Abstract. The goal of this study is to (i) understand the characteristics of high-, average- and low-level performing students in a first year computer programming course, and (ii) investigate whether their performance can be predicted accurately and early enough in the semester for timely intervention. We triangulate data from three sources: submission steps and outcomes in an automatic marking system that provides instant feedback, assessment marks during the semester and student engagement with the discussion forum Piazza. We define and extract attributes characterizing student activity and performance, and discuss the distinct characteristics of the three groups. Using these attributes we built a compact decision tree classifier that is able to predict the exam mark with an accuracy of 72.69% at the end of the semester and 66.52% in the middle of the semester. We discuss the most important predictors and how such analysis can be used to improve teaching and learning.

Keywords: Computer Science education, student performance prediction

1 Introduction

Novel technology-enhanced teaching tools provide effective solutions to support computer programming courses, and also collect a lot of data about the students. In this paper we describe how the analysis of multiple data sources, combined with student assessment marks, can give insights into student learning.

The context of our study is a first year programming course with 224 students, where different tools were used to support learning: (i) PASTA: an automatic marking and feedback system that allows students to submit their code online, checks this solution against a set of pre-specified tests set by the teacher and provides immediate formative feedback to the student about the passed and failed tests; students can then correct their mistakes and resubmit the solution until the code is correct, (ii) Piazza (www.piazza.com), a mix of discussion board and wiki allowing students to ask and answer questions, and post notes, all under teachers' guidance.

The data captured from these two systems, along with the assessment marks, provides different useful perspectives on student learning: progression in code writing and diagnostics (PASTA), interaction and engagement (Piazza) and student performance (assessment results).

AIED 2015

We extend previous work on predicting student performance, e.g. [1][1], by triangulating data from the three sources above to answer the following questions: 1) What are the characteristics of the high-, average- and low-level performing students?, 2) How accurately can we predict the exam grade? and 3) Is it possible to predict accurately the exam grade before the end of the semester for early intervention?

2 Data

We define three groups of students based on their exam performance: High-level - HDD (High Distinction and Distinction) – exam mark of [75, 100], average-level - CRP (Credit and Pass) – exam mark of [50, 74] and low-level - F (Failing) – exam mark below 50. The number of students was: 64 in HDD, 83 in CRP and 77 in F.

The assessment during the semester consisted of five components: weekly homeworks (10%), two programming tasks and an assignment submitted via PASTA (2% in week 3, 6% in week 6 and 16% in, week 12) and a practical test (16% in week 7) involving writing computer programs in front of the computer. The exam (50%) was conducted at the end of the semester and required writing code for solving problems.

We chose the exam mark as a performance index as the exam is the major and most comprehensive assessment component, conducted under strict conditions which minimizes cheating. The exam mark is highly correlated with the final course mark (r=0.937). Table 1 shows the 23 attributes we extracted from the three data sources.

1. Assessment marks
homework_mark, task1_mark, task2_mark, prac_quiz_mark, assignment_mark – mark (%) awarded for each assessment component.
2. PASTA activity – submission history
Starting and finishing times for assessments
tasks_start, tasks_finish – the average number of days before the due date the student will start and finish the two tasks; *assignment_start, assignment_finish* – the average number of days before the due date the student will start and finish the assignment*; early_task, early_assignment* - 1 if the student starts the tasks/ assignment earlier than the average student; 0 otherwise.
Multiple assignment submissions – improvement and consistency
assignment_only_improvement - 1 if the student's marks for compiling assignment submissions never decrease; 0 otherwise; *assignment_consistency* - goodness of fit (R^2) over each of the student's compiling submissions; range: [1,1], close to 1/-1 - linear increase/decrease in marks over time, close to 0 - random distribution of marks over submissions; *assignment_improvement* - slope of the trendline of the student's assignment marks over each compiling submission, a larger number indicates rapid improvement; *assignment_ first_mark* - mark awarded for the student's first submission for the assignment.
3. Piazza activity – views, questions and answers
piazza_views, piazza_questions, piazza_answers - number of posts viewed, questions asked and questions answered by the student on Piazza; *piazza_activity* – calculated as: (*piazza_views* + 10*(*piazza_questions* + *piazza_answers)* + 5*(*piazza_posts* - *piazza_answers*)) / *total_posts,* where *piazza_posts* is the total number of student contributions (asking or answering a question, or posting a comment), and *total_posts* is the total number of question threads on Piazza; *piazza_active_viewer, piazza_active_questioner, piazza_active_answerer* - 1 if the student has an average or higher number of Piazza posts viewed, questions asked and questions answered; 0 otherwise.

Table 1. Defined attributes to characterise student performance and activity

[1] [1] Romero, C., Ventura, S. Espejo, P. G. and Hervas, C.: Data mining algorithms to classify students, In: International Conference on Educational Data Mining (EDM), pp.8-17 (2008)

3 Results

3.1 What are the characteristics of high-, average- and low-level students?

We computed the mean values for each attribute and group, and also conducted tests for statistical significance of the differences between these mean values. The results (not shown due to space limitation) lead to the following student characteristics:

- High-level students start and finish their assignments early. They tend to start an assignment strongly, and make consistent and significant improvement to their marks with successive attempts. They engage in peer to peer discussions by asking questions and reading the information provided to other students.
- Average-level students start an assessment early if the assessment is not worth a large amount of marks. They start an assignment at a fairly low level, and then make small improvements over time, occasionally decreasing their marks until they are content that they have done enough to stay at the average level. They engage in peer to peer discussions in the same way as the high level students.
- Low-level students start an assessment task with very little time before the due date (on average only 2 days). They start an assignment at a low level; however they tend to make large improvements to their marks as they progress, even if not consistently with each attempt. They do not engage in discussion board activity.

3.2 How accurately can we predict the exam grade?

To investigate how accurately we can predict the exam grade using the extracted attributes, and which of these attributes are the best predictors, we employed a Decision Tree (DT) classifier. DTs are popular for educational data mining as the generated rules provide an explanation about the decision, can be easily understood and applied by teachers and students. Although DTs have an inbuilt mechanism for attribute selection, their performance can benefit from prior attribute subset selection. Starting with the full set of 23 attributes from Table 1, we used several methods for attribute subset selection (automatic such as wrapper and also manual), before generating the DT.

The best DT achieved 72.69% accuracy and is shown in Table 2 (left). It consists of 13 rules using only 7 attributes from all data sources: from the assessment marks: *prac_quiz_mark*, *assignment_mark* and *homework_mark*, from Piazza: *piazza_active_viewer* and *piazza_views*, and from PASTA: *early_assignment* and *assignment_consistency*. Most of these attributes were identified as important for discriminating the three groups of students in Sec. 3. An accuracy of 72.69% is considerably higher than the baseline accuracy of 37.05% (always predicting the majority class CRP, 83/224) and also reasonably high to be used in practical applications.

3.3 Can we predict accurately the exam grade before the end of the semester?

We also investigated how accurately we can predict the exam grade in the middle of the semester, for the purpose of early intervention. We chose the end of week 7, which is approximately the middle of the 13-week semester, and by which time the students have completed the two tasks and the practical test but not the assignment. We did not have the data from Piazza for this point of time. The best DT achieved 66.52% accuracy, which is very promising, and is shown in Table 2 (right). It consists of 10 rules using only 4 of the attributes – 2 assessment marks (*prac_quiz_mark* and *task1_mark*) and 2 activity attributes from PASTA (*early_task* and *tasks_finish*).

<table>
<tr><th colspan="2">Predicting exam grade (HDD, CRP or F)</th></tr>
<tr><th>at the end of the semester</th><th>in the middle of the semester (week 7)</th></tr>
<tr><td>prac_quiz_mark <= 81.88
| prac_quiz_mark <= 54.38
| | assignment_mark <= 97.5: F
| | assignment_mark > 97.5
| | | assignment_mark <= 98.75: CRP
| | | assignment_mark > 98.75: F
| prac_quiz_mark > 54.38
| | piazza_active_viewer <= 0
| | | homework_mark <= 31: CRP
| | | homework_mark > 31: F
| | piazza_active_viewer > 0
| | | assignment_mark <= 35.38: F
| | | assignment_mark > 35.38: CRP
prac_quiz_mark > 81.88
| prac_quiz_mark <= 94.06
| | early_assignment <= 0: CRP
| | early_assignment > 0
| | | assignment_consistency <= 0.79: CRP
| | | assignment_consistency > 0.79: HDD
| prac_quiz_mark > 94.06
| | assignment_mark <= 86.56
| | | piazza_views <= 48: CRP
| | | piazza_views > 48: HDD
| | assignment_mark > 86.56: HDD

Accuracy: 72.69%</td><td>prac_quiz_mark <= 81.88
| prac_quiz_mark <= 54.38: F
| prac_quiz_mark > 54.38
| | task1_mark <= 80.31: F
| | task1_mark > 80.31
| | | early_task <= 0
| | | | prac_quiz_mark <= 79.69
| | | | | tasks_finish <= 0: F
| | | | | tasks_finish > 0: CRP
| | | | prac_quiz_mark > 79.69: F
| | | early_task > 0
| | | | prac_quiz_mark <= 65.31
| | | | | prac_quiz_mark <= 60.31: CRP
| | | | | prac_quiz_mark > 60.31: F
| | | | prac_quiz_mark > 65.31: CRP
prac_quiz_mark > 81.88
| prac_quiz_mark <= 94.06: CRP
| prac_quiz_mark > 94.06: HDD

Accuracy: 66.52%</td></tr>
</table>

Table 2. Predicting exam grade – DT and accuracy using 10-fold cross validation

4 Conclusions

We studied the characteristics of high-, average- and low-level performing students in a first year computer programming course. We triangulated data from three sources, offering different perspectives on student learning: automatic marking system, discussion forum and assessment data. We defined useful attributes that capture the distinctive characteristics of each group and used them to build a compact DT classifier that is able to predict the final exam grade with promising accuracy. Our results can be used to provide regular formative feedback to students during the semester if their current results and behavior are likely to be associated with high-, average- or low-level performance, where the problems are, and what remedial actions should be taken. This includes identifying both students who are at risk of failing and students who are not achieving their goals. Students and teachers can also be made aware of the characteristics of the three groups, which will encourage better learning and teaching.

Acknowledgement. This work was supported by the Human Centered Technology Cluster.