Predicting student grades from online, collaborative social learning metrics using K-NN

Matthew Yee-King
Computing, Goldsmiths
University of London
m.yee-king@gold.ac.uk

Andreu Grimalt-Reynes
Computing, Goldsmiths
University of London

Mark d'Inverno
Computing, Goldsmiths
University of London
dinverno@gold.ac.uk

ABSTRACT

We describe a collaborative video annotation system that aims to engage learners in a focused, collaborative process of content sharing and discussion, and explain how it was used in an online creative programming MOOC on Coursera. We explore the use of K-NN (K nearest neighbour) to predict which of a variable number of evenly spaced, final grade bands students will fall into based solely on a feature vector consisting of the total number of UI click and mouseover events they generated during the course. We were able to classify students into pass/fail bands with 88% precision; with 3 grade bands, precision was 77%, going down to 31% with 10 grade bands. Typically, a feature subset containing less than half of the available features provided the best performance.

Categories and Subject Descriptors

K.3.1 [**Collaborative learning**]: K.3.2 Computer science education

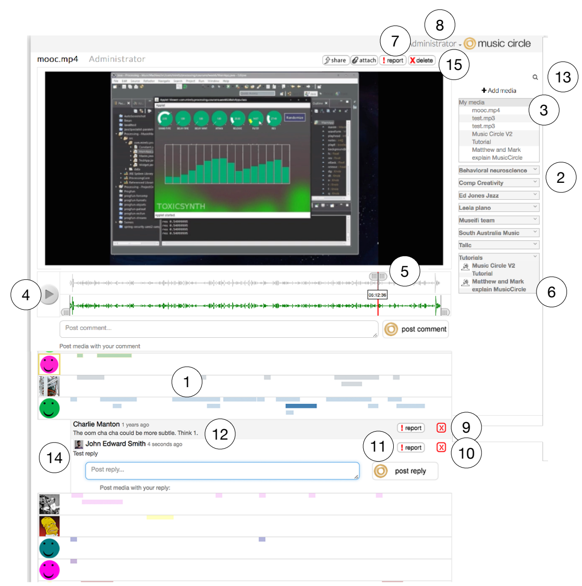
1. INTRODUCTION

Our work is concerned with the development and analysis of systems that enable online, collaborative learning driven by different types of feedback. In this paper, we show how it is possible to predict student grades using user interface telemetry data gathered from a case study involving 993 students who completed all assessments for a creative programming course on MOOC platform Coursera. The students used a collaborative video annotation tool as part of their peer assessment, which we developed as part of an EU funded research project [5]. Previous work with collaborative media annotation systems and grade prediction includes [1, 3] and [2, 4] respectively[1].

2. THE CASE STUDY AND DATA SET

Three times during the course, the students were set a graded peer assessment wherein they had to extend our example programs and create a 5 minute video of themselves explaining their code and running their program. The videos were uploaded to our collaborative video annotation system wherein they could look at each others' videos and create annotations along a 'social timeline'. The system logged click and mouseover events on the UI elements shown in Figure 1, 3,716 unique users logged into our system. Of these, 3558 viewed one or more videos, 827 made one or more comments, and 258 made one or more replies to comments. 2,898 videos were submitted for three separate assessments, and were viewed a total of 112,189 times. 7,370 comments were made, and 978 replies. For this paper, we filtered the data down to all logged click and mouseover events for students who gained a final grade on the course, a total of 993 students.

Figure 1: **A screen shot of the video annotation and discussion system. The numbered labels show all of the elements of the UI for which events are logged automatically.**



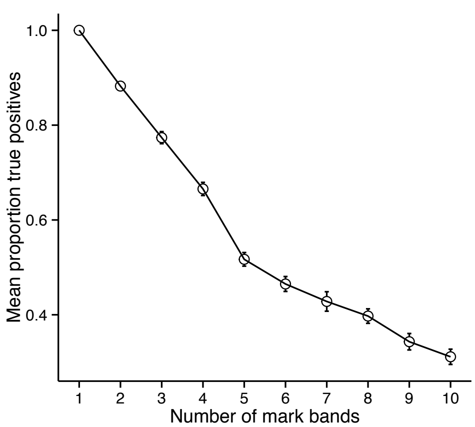
[1]A full version of the paper can be found at http://dx.doi.org/10.13140/RG.2.1.4525.9129

3. ANALYSIS

To predict student grades, we created a 16 dimension feature vector consisting of total numbers of clicks and mouseover events on each of the GUI elements shown in Figure 1 plus the final grade achieved by the student. We began by attempting to correlate individual elements of the feature vector to the final grade but individual correlations were too weak to predict grades, ranging between 0.53 and 0.18. This motivated us to try a multivariate classification approach. For our first analysis, we assigned labels to the students based on which of N evenly spaced grade boundaries they fell into. For example, if $N = 2$, then students were labelled **1** if $final\_grade < 50$ and **2** if $final\_grade \geq 50$. We split the dataset into equally sized training and test sets and attempted to train a K-NN classifier to assign labels to the test set, with varying numbers of mark bands and multiple run cross validation.

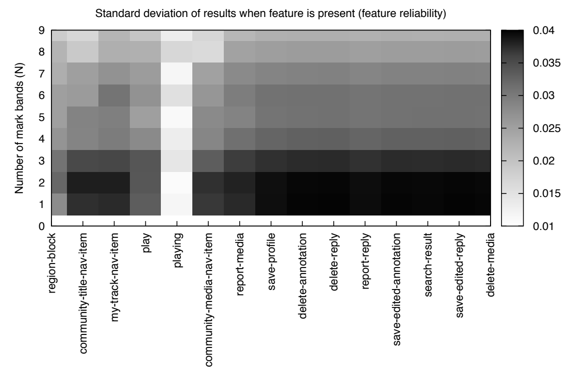
Figure 2 shows the proportion of correctly assigned labels in the test set as number of mark bands N varies from 1 to 10. For example, the pass/fail classification where $N = 2$ achieved 88% true positives. We note that the distribution of marks across the bands has a significant impact on the meaning of accuracy, and that for $N = 2$, for example, there are a large number of examples in each class which are being correctly classified.

Figure 2: **Performance of the classifier with $k = 6$ and number of mark bands $N = 1...10$.**



For our second analysis, we tried out all possible combinations of feature elements to see which combination achieved the highest classification accuracy. Since the number of features was 15, the number of permutations was 32768 (2^15). K was set to 6 and number of mark bands N varied from 1-10. Figure 3 highlights the most reliable features in the feature set by showing how much the prediction score varied (the standard deviation) across the set of all permutations per N which involved that specific feature. To be clear, it does not differentiate between features that reliably provide good or bad results. The most reliable feature was 'playing', which is triggered automatically while a video is playing. The second most reliable feature was 'region block', which is logged when a user clicks on a comment on the timeline to open the discussion thread. More work is needed to understand this result more deeply.

Figure 3: **Heat plot showing the standard deviation in the prediction results when different features are present. Low variation (lighter) is desirable, meaning a feature makes a reliable contribution to the results.**



4. CONCLUSION

We have briefly described a collaborative video annotation tool we have developed. Using interface telemetry data gathered describing click and mouseover events generated by the user interface of the system, we were able use a K-NN classifier to classify students into pass/fail bands with 88% precision; with 3 grade bands, precision was 77%, and with 10 bands it was 31%. We measured the prediction power of different combinations of the features and were able to identify the most reliable features, which relate to playing back videos, exploring content menus and reading comments.

5. REFERENCES

[1] D. Bargeron and J. Grudin. Asynchronous Collaboration Around Multimedia Applied to On-Demand Education. *Journal of Management Information Systems*, 18(4):117—-145, 2002.

[2] C. a. Coleman, D. T. Seaton, and I. Chuang. Probabilistic Use Cases: Discovering Behavioral Patterns for Predicting Certification. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*, pages 141–148, 2015.

[3] E. F. Risko, T. Foulsham, S. Dawson, and A. Kingstone. (CLAS): A New TOOL for Distributed Learning. 6(1):4–13, 2013.

[4] T. Sinha, P. Jermann, N. Li, and P. Dillenbourg. Your click decides your fate: Leveraging clickstream patterns in MOOC videos to infer students' information processing and attrition behavior. 2014.

[5] M. Yee-King, M. Krivenski, H. Brenton, and M. D'Inverno. Designing educational social machines for effective feedback. In *8th International Conference on e-learning*, Lisbon, 2014. IADIS.