Machine Learning in Educational Technology

Laksh Advani, Kylee Budai, Erik Holbrook, Rashmi Shetty, Ankit Srivastava, Varsha Teratipally

Abstract

Much research has been done to improve educational technology. Generally, though, it focuses on cognitive and behavioral sciences and ignores the potential benefits of integrating machine learning techniques. Being able to classify questions as helpful, unhelpful, and misleading using learning techniques could lead to a significant increase in the overall effectiveness of current-day educational technology. Furthermore, an in-depth conceptualization of students' understanding and cognitive models would allow instructors to better assist in the learning process at an individual level.

In this work, we seek to apply machine learning techniques, namely prediction and clustering, as an avenue to model the abilities of students in their educational environments, and to elucidate commonalities among problems at multiple levels of granularity. We found success in both predicting student success on particular problems across domains, as well as revealed descriptive information about the structure and similarity of schools, students, and problems.

Introduction

The goal of this project is to investigate trends at the level of individual student, classroom, and across problems. We will apply a variety of unsupervised techniques to discover common errors. Woot Math (www.wootmath.com) is an educational website that provides math lessons for students in elementary and middle school. Access to the site is purchased by the school and students participate in a computer-lab setting under the supervision of a teacher. The goal of Woot Math is to provide a fun, educational experience that allows students to gain experience with new concepts in math.

The layout of each problem is presented below. Students work out solutions using the canvas and their responses are judged based on total correctness.

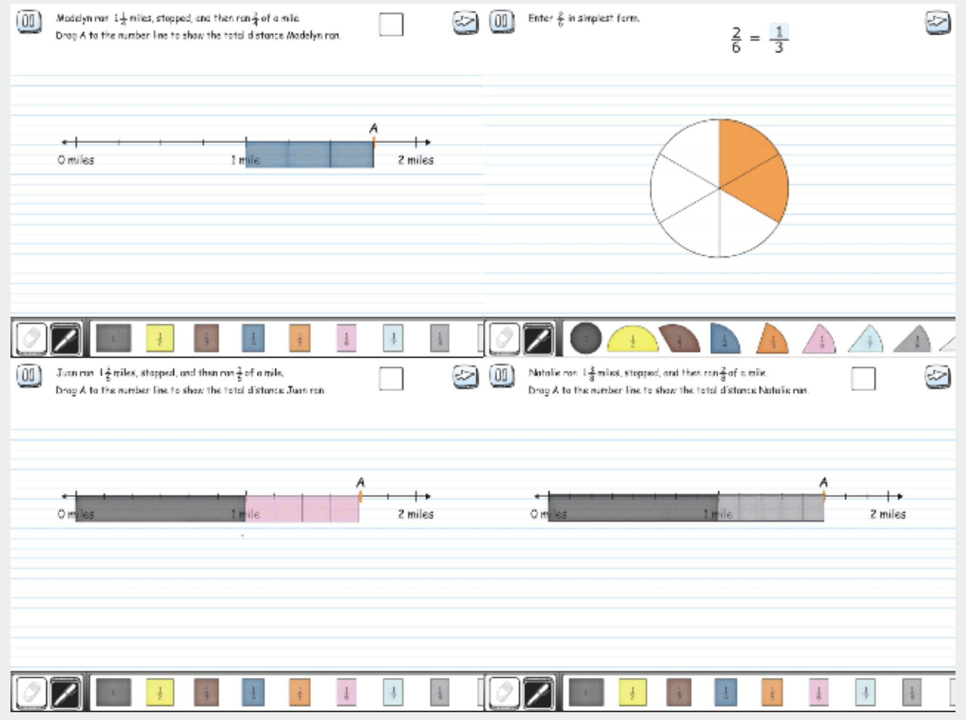


Figure 1. Some examples of problem "canvases" where the students work on problems.

Woot Math provided us with a data base with ~1 million entries. Each entry represents a student answering a single problem. Attributes include but are not limited to subject, time spent, correctness, difficulty, student grade, and a handful of internal measures.

Student ID	Student Grade	School ID	Problem ID	Subject	Time Spent	Time Elapsed	Correct	Difficulty [0,1]
F8W8W9U0R2U1	3	F5asdncA2	t8suuCs7vN.l31_9_parta.b6c1c8dLXx	Fractions	24904	56930	0	.5792
22Q6W2E2A1	6	AFJ3fa022	x9tw0UBBWG.set1.f4F0t0H7hk	Decimals	43222	78220	1	.8346

Figure 2. An example of the important fields from the dataset.

Prediction Methods, Results, and Analysis

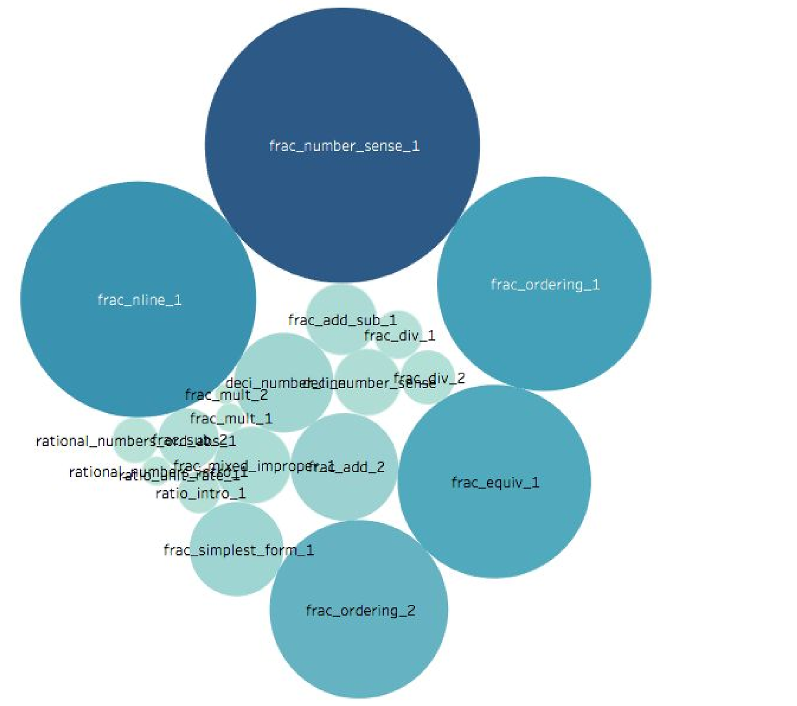


Figure 4. Visualization of the relationship between the internal difficulty measure and number of correctly answered problems for different lessons. The color represents difficulty - darker colors are perceived as being easier problems. The size of the circles represents the number of correct answers.

Feature Selection:

We ran recursive feature elimination with Logistic Regression to rank features based on performance. This helped us to eliminate poorly performing features and focus on more potentially beneficial ones..

Analysis of Internal Measures:

$$P\left(correct \mid diff\right) = P\left(correct \mid \sum_{i}^{L} \frac{nright[i]}{ntotal[L]}\right)$$

One of our research areas was the analysis of the company's own indicators like the 'difficulty' feature, this measure is assessed and assigned by subject matter experts within Woot Math. The 'difficulty' rating gives a rough estimate of the likelihood of success for students, based on the specific concept and how it relates to the overall lesson. Using logistic regression to predict the correctness of each question using the only "difficulty" parameter resulted in a 90% accuracy. This agreed with the accuracy that resulted from a logistic regression using the total number of correctly answered problems over the total number of problems instead of the difficulty measure. This led us to conclude that perceived difficulty from the students agrees with the internal difficulty measure.

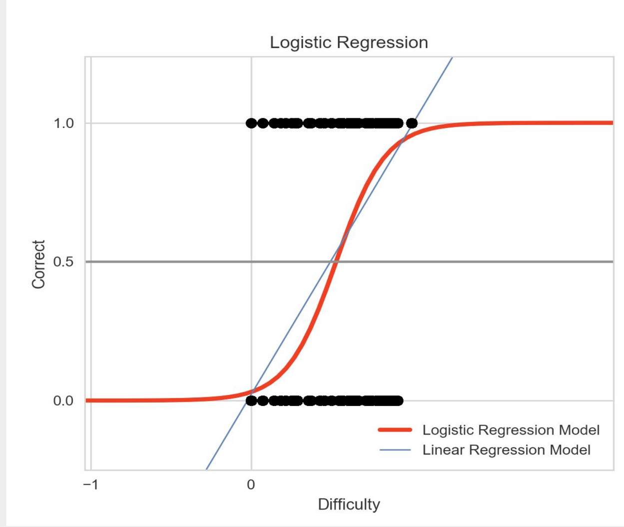


Figure 3. Logistic Regression with Difficulty and Correctness

Time Analysis: Further, in order to get a time at which questions are classified as difficult we used a SVM to split the data into hard and easy questions, from this we could gather that questions that take more that 3 minutes to complete are generally classified as difficult.

Predicting the Occurrence of a Bonus Problem:

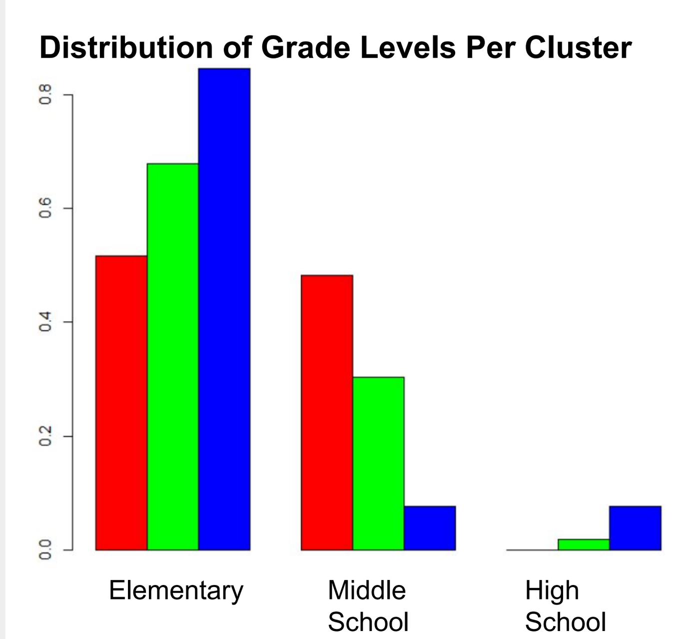
If a student is performing well, he is presented with a bonus problem. Since being given a bonus problem serves as an indicator of well performing students, we decided predict the occurrence of a bonus problem. We used Principal component analysis to extract significant features such as:

1) Time Spent
2) Difficulty Measure
3) Problem ID
4) Number of Correct Answers
5) Number of Unattempted Problems
6) Number of Incomplete Answers

Using these features as covariates, we used Naive Bayes' as a baseline model. To improve accuracy, we used a Support Vector Machine as well as Logistic Regression. These methods yielded the accuracies shown below.

Method	Accuracy
Naïve Bayes	78.26%
SVM	77.25%
Logistic Regression	88.3%

Clustering of Schools



Aggregating the data from each of the individual 98 school allowed us to cluster schools based on performance using k-prototypes, a clustering method that allows for categorical data.

Difficulty, mastery mean, and mastery standard deviation are scores assigned to each student based on performance in a given lesson. A student is given a mastery score based on previous performance as well as current performance. Clustering on these attributes allowed us to determine how well these scores represented student and school performance.

Assuming that schools generally are either elementary schools, middle schools, or high schools, we were able to characterize schools as such and look at the distribution of school levels in each cluster.

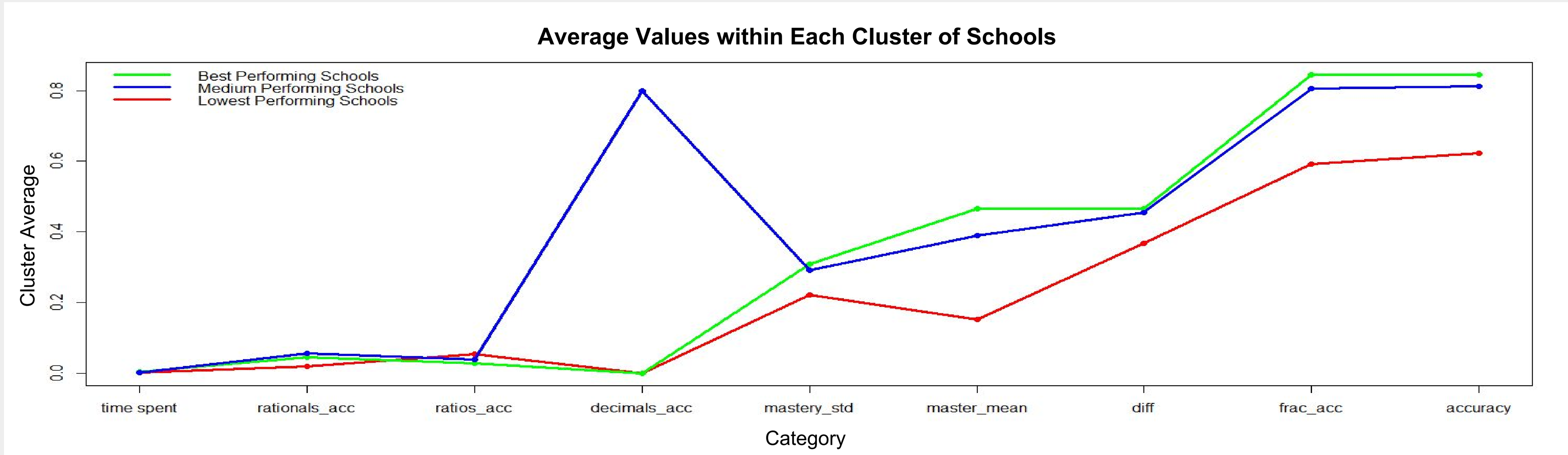


Figure 5. Visualization of cluster centers for each of the three clusters on aggregated school data.

Conclusions

In this work, we were able to determine the benefit of clustering for schools in determining the relative skill levels of students. Through prediction techniques, we verified the effectiveness of the Woot Math internal difficulty score and we were able to discover high-achieving students through the identification of bonus problems.

Future Work

Were we to extend this work, we would like to focus on better developing a cognitive model for the students. As of this writing, we were only able to use high level features and not well develop a deep set of well-engineered features to use as representation of student knowledge.

Each attempt of a problem generates a screenshot of the canvas where the student performed their work. This image set clearly contains a plethora of information; we initially intended on using these in our knowledge set, but this would require at the very least a heavy use of image processing, and probably would further entail using a convolutional machine learning approach to glean any useful information.

Finally, our techniques show useful aspects of the dataset already in existence in Woot Math servers. Our final future task would be to help the team implement our most effective methods into their production environment to effectively utilize our techniques to provide business value.

Acknowledgements

We would like to sincerely thank Woot Math for their generous contributions to this project. They have substantially assisted us in the process of data acquisition, cleaning, and modification, as well as providing us with areas research focus. We would also like to thank Professor Chenhao Tan and the CU Boulder Computer Science Department for providing us the opportunity to pursue and present this work.