Project: Data Mining Piazza

Gary Bean *
gbean@indiana.edu
HID: F16-IU-2000
Gitlab: gbean

Tim Whitson †
tdwhitso@indiana.edu
HID: F16-IU-2007
Gitlab: whitstd

ABSTRACT

Electronic learning tools have become ubiquitous in modern classrooms. One such tool is Piazza, a discussion board where students can interact with each other and ask questions of their instructors. While student/instructor discussion is the main goal of Piazza, the data produced from the discussions can be just as valuable. For this project, we will be data mining Piazza discussions, using Python, and analyzing that data to produce relevant visuals and statistics, using d3js. We wrote three different Python tools for the purposes of:

1. Extracting data from Piazza
2. Scrubbing the data to get important and readable information
3. Analyzing the data

To begin, we wrote a Python program to mine data from Piazza. Piazza has an internal API that it communicates with via XMLHttpRequest, or AJAX, requests to retrieve and display information to the user. We were able to track these requests by monitoring the network requests in Google Chrome Developer Tools. The network requests also display the POST information that is sent to the server. Using the URLs and POST request data, we were able to recreate Piazza's API using the third-party Python "requests" module.

Since the data that is returned by the Piazza API is in JSON format, we wanted to pull out only the pertinent information. JSON data is slightly more complex than data such as CSV, so we wrote Python tools to parse the JSON. The idea behind these tools were for them to be as flexible as possible, so we could feed the data into various types of visualizations.

1. INTRODUCTION

*undergraduate
†undergraduate

Educational Data Mining (EDM) is emerging field of study that is ultimately using large amounts of data to gain insightful knowledge on how we can improve the education system. According to Education Data Mining: A Review of the State-of-the-Art, "EDM is concerned with developing methods to explore the unique types of data in educational settings and, using these methods, to better understand students and the settings in which they learn," [1]. There are a variety of data that this field of study will collect such as: "interactions of individual students with an educational system (e.g., navigation behavior, input to quizzes and interactive exercises) ... data from collaborating students (e.g., text chat), administrative data (e.g., school, school district, teacher), and demographic data (e.g., gender, age, school grades)," [2]. The data from students collaborating with each other and asking their professor questions on a online discussion board will give us understanding and insights, and is one example of how EDM can be utilized in the online classroom.

By data mining the student/teacher discussion on Piazza, we will gain valuable information for every stakeholder. Data that will come out of this include: individual and class statistics (e.g., completion percentages, times), student problem areas, external link usage, classifications, word cloud, and user interaction. From all of this data and analytics, we hope to improve the learning process for the student and to improve the instructing process for the professor. Analyzing this data will allow teachers to see how much a student participates over the duration of a course. Mining this data can also help educational institutions see how quickly and often instructors are communicating with their students.

2. ARTIFACTS

- Gitlab code
- Project paper

3. CONCLUSIONS

4. REFERENCES

[1] C. Romero and S. Ventura. Educational data mining: A review of the state-of-the-art. http://s3.amazonaws.com/academia.edu.documents/30661302/Ve
[2] O. Scheuer and B. M. McLaren. Educational data mining.

http://www.cs.cmu.edu/ bmclaren/pubs/ScheuerMcLaren-EducationalDataMining-EncyOfLearningScience2011.pdf, 2011.