Topical Classification of Comments in MOOC Discussion Forums

Bita Akram, Devarshi Pratap Singh, Pranjal Deka, Simerdeep Singh Jolly
Department of Computer Science
North Carolina State University
Raleigh, NC 27606
{bakram, dsingh4, pdeka, sjolly}@ncsu.edu

Project Idea

By the emergenc of information era, massive open online courses (MOOCs) are becoming a prevalent source of education, specially higher education [1]. Many studies have attempted to underestand students' interactions with MOOCs through pattern recognition and data-mining. For example, a study demonstrated in [2] used trace-data to build behavior patterns of low and high achieving students. In another study [3], the relevance of the course threades are ranked using linear regression and generative models. In this project, we aim to build a classfier to conduct topical clustring on students' discussion forums. This will help instructors and also higher education researchers to better analyze learner's learning progress and to predict grades and completion of the courses. Analyzing trace data obtained from MOOCs also inform the desing of next generation of MOOCs.

Dataset

Our dataset consists of 487 samples of conversation data in a MOOC discussion forum. In particular, it consists of comments shared between learners in the forum. It also contains several other attributes like the timestamp of the comment, the title of the discussion, the category of the discussion, the type of comment (eg. question, statement, reflection etc) and the level of critical thinking the comment represents.

Software

We use Lightsise along with python to apply feature extraction.We will also write code in Python to pre-process and parse forum discussions and finally to implement our learning algorithm. We will make use of R, MATLAB and Weka. We will also implement a user-facing API to demo our results.

Work Division between Group Members

Devarshi and Bita - Pre-processing, feature selection and data transformation (R, MATLAB, Weka)
Devarshi and Simerdeep - Implementation of text extraction and final algorithm (Python)
Pranjal and Bita - Run, test and evaluate the model

Midterm Milestone

We plan to complete pre-processing, transformation and division of dataset into training and testing dataset by midterm. We also plan to prepare initial evaluation results for different classification models that we will implement.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

References

[1] Ezen-Can, Aysu, et al. "Unsupervised modeling for understanding MOOC discussion forums: a learning analytics approach." Proceedings of the fifth international conference on learning analytics and knowledge. ACM, (2015).

[2] Brinton, Christopher G., et al. "Learning about social learning in MOOCs: From statistical analysis to generative model." IEEE transactions on Learning Technologies 7.4 (2014): 346-359.

[3] Anderson, Ashton, et al. "Engaging with massive online courses." Proceedings of the 23rd international conference on World wide web. ACM, (2014).