A Web based dashboard for MOOC instructors

Submitted by
Zhang Yichi
Department of Electrical & Computer Engineering

In partial fulfillment of the
requirements for the Degree of
Bachelor of Engineering (Computer Engineering)
National University of Singapore

B.Eng. Dissertation

A Web based dashboard for MOOC instructors

By

Zhang Yichi

National University of Singapore

2017/2018

Project ID: H0791490

Project Supervisor: Prof Min-Yen Kan

Deliverables:

Report: 1 Volume

ABSTRACT

Massive Open Online Courses (MOOCs) aim to achieve open education by providing a platform in which unlimited number of participants can access the filmed lectures and teaching materials via Internet. To enable interaction between MOOC instructors and its participants, MOOC platforms normally provide discussion forums for them. With the facts that responding to a participant's post can improve his or her learning outcome and there is a large participant-instructor ratio, our work is to help instructors to select the posts that are worth responding to benefit the most other participants. This is achieved by a supervised classifier trained with a large size of sample data from more than 60 MOOC forums. My work for the first semester is to develop a dashboard to integrate with the machine learning model and develop a web crawler for Integrated Virtual Learning Environment (IVLE) to get more training data from university-level MOOC platform for the classifier.

ACKNOWLEDGMENTS

I would like to express my most sincere appreciation to my project supervisor and project administrator, Prof Min-Yen Kan, and his PhD student, Muthu Kumar Chandrasekaran. Throughout this project, two of them have given me necessary advice and guidance to help me familiarize the project. As a beginner in web application, I cannot possibly reach this step without the guidance from both of them.

CONTENTS

ABSTRACT **3**

ACKNOWLEDGMENTS **4**

1. INTRODUCTION **6**

1.1 IMPORTANCE 7

2. LITERATURE REVIEW **8**

3. IMPLEMENTATION PROGRESS **9**

3.1 OVERVIEW 9

3.2 DASHBOARD 10

3.2.1 Design *10*

3.2.2 Challenges *10*

3.2.3 Results *11*

3.3 CRAWLER 12

3.3.1 Design *12*

3.3.2 Challenges *12*

3.3.3 Results *13*

4. PLAN **13**

5. REFERENCES **15**

1. INTRODUCTION

Massive Open Online Courses (MOOCs) are the platforms in which unlimited number of participants can access the filmed lectures and teaching materials via Internet. It was first developed and introduced in 2006 as a distance education tool and became popular in 2012 (Pappano, 2012). It does not only allow people from every part of the world to have the opportunity to attend the best lectures in the world, but also provide those who cannot afford universities with a chance to make the change in their life. Some examples of the popular MOOC platforms are Coursera, edX and Udacity. Figure 1.1 gives an overview of how MOOC platforms look like. In addition to the traditional form of teaching model, MOOCs usually provide interactive forums for the students and teaching staffs to interact with each other. The teaching staffs in this report refer to MOOC instructors, teaching assistant and mentors. Currently, discussion forums are the only method for them to do so. This is how it works. A participant can access video lectures, hands-on practices and online assessments from MOOC platforms. If any participant encounters problems or has some doubts, he or she can simply go to the course's respective discussion forum and post the question on it. Those discussion forums usually contain a very large volume of threads and posts from thousands of participants. It is fair to say that every participant is trying to compete with each other and get his or her post to be read and answered by the teaching staffs.

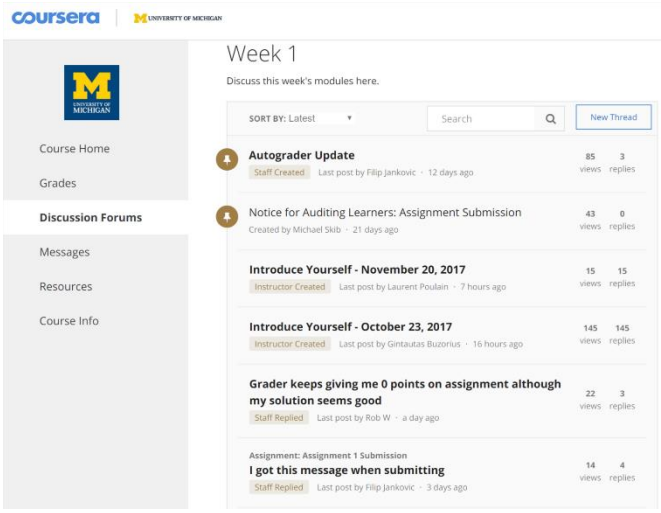

Figure 1.1 Screenshot of Coursera website

Figure 1.2 Screenshot of Xuetangx website

1.1 Importance

It has been argued that teaching staffs' intervention to participants' question and discussion posts in MOOC forums can improve participants' learning outcomes (Chen, Phang, Zhang, & and Cai, 2016). Therefore, each MOOC platform is now dedicated to maximizing its teaching staffs' interaction coverage on its discussion forums. However, with its easy accessibility, MOOC platforms often find it difficult to let teaching staffs have interactions with every individual participant. This is because that for each course or discussion forum, teaching staffs are usually outnumbered by the participants. On the other hand, the teaching staffs on the platforms can only assign limited time on it because they are usually full-time lecturers or professors in universities. The most effective way to minimize the effect of failing to cover all

posts is to select the most representative question or discussion posts to respond. In another word, by responding to those particular posts, most of the other participants can be benefit from it as well (Chandrasekaran, Epp, Kan, & Litman, 2017). However, to find those posts that can maximally benefit other students, the teaching staffs have to read through every post first. That process is very time-consuming. The posts on MOOC platforms are normally displayed in threads and sorted by date. None of the sorting or display methods can significantly simplify the selecting process. Moreover, some teaching staffs might not be experienced enough to identify those worth-answering posts.

2. LITERATURE REVIEW

Earliest in 2013, Catropa has highlighted the low completion rate on MOOCs (Catropa, 2013). This triggered scholars to wonder the factors causing it. Many of them have paid their attention to importance of intervention from MOOC instructors and believe it to be an important issue to tackle (Chandrasekaran, Kan, Tan, & Ragupathi, 2015). Indeed, according to a report, the lack of interaction can make individual participants think that they are "ignored" or they have posted some "stupid" questions (Dolan, 2014). Those negative feelings may cause the participants to drop out from using MOOCs (Zhou, 2017).

To tackle the issue discussed above, Chandrasekaran et al. proposed to use a machine learning model to help interpret if a post is worth responding. He has used a large size of sample data from 61 MOOCs to train a binary classifier which can identify the "high-value" posts from hundreds of threads. In his report, he has discovered that there could be a significant improvement in the prediction if the model knows the forum type (Chandrasekaran, Kan, Tan, & Ragupathi, 2015).

However, the vocabulary used in various courses can be very different. The prediction result would become less accurate if used for a new course from an unseen discipline. Chandrasekaran then proposed to use discourse-based signals to categorize the posts on discussion forums and eventually identify what form of intervention would bring the maximal benefit to most students (Chandrasekaran, Epp, Kan, & Litman, 2017). From Chandrasekaran's research, it is shown that the Penn Discourse Treebank (PDTB) tags generated by an automated discourse parser for post contents have improved the prediction task significantly. This has helped his classifier to be more independent from course-specific terms or domain-specific vocabularies.

3. IMPLEMENTATION PROGRESS

3.1 Overview

There are two implementation tasks for this semester – a dashboard and a web crawler:

To bring the whole research project alive, we need to develop a discussion forum dashboard for instructors to visualize the results from the back-end machine learning model and access posts from MOOC platforms.

The Integrated Virtual Learning Environment (IVLE) is the MOOC platform for the NUS community. It is designed to facilitate and supplement teaching at the NUS. To improve the accuracy of the machine learning model and prepare it for an actual demonstration within the National University of Singapore (NUS), the model must be trained using relevant data from NUS. Therefore, I need to develop a web crawler to extract the posts from IVLE.

3.2 Dashboard

3.2.1 Design

The major components of the dashboard should contain a view of complete discussion forums, a control panel to manage the posts and a login system. Within the discussion forums view, the instructors will be able to see a score attached to each post indicating how much the other students can benefit by responding to the post. The posts can be sorted by their scores so that the instructors do not need to browse through all the posts and can start responding the high priority posts first. The Entity-Relationship Diagram of the dashboard can be referred in figure 3.1 below.

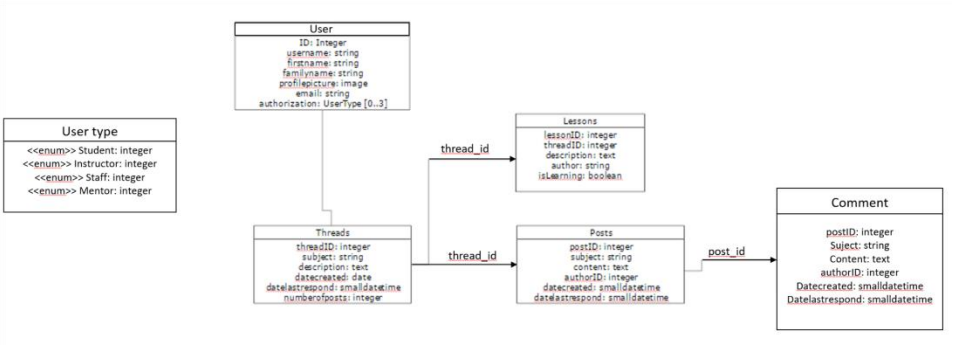

Figure 3.1 ER Diagram for forum dashboard

3.2.2 Challenges

The main challenge of doing this part of implementation is the lack of relevant experience and knowledge. It is my first web application project and I must use React.js and Bootstrap to develop the dashboard as a beginner. My limited experience has made the progress slow in the beginning.

3.2.3 Results

A skeleton dashboard is developed using React.js. It has functions of authentication, view threads, posting new posts and choose forums according to schools and categories. There is another dashboard developed by a research intern. Previously, Chandrasekaran has used that dashboard to illustrate the result from his machine learning model. Therefore, we decided to use his dashboard for the final integration and the progress of my dashboard can be stalled for now to prevent the double work. The work of my dashboard and the dashboard we will integrate with can be seen in the figure 3.2 and figure 3.3 below.

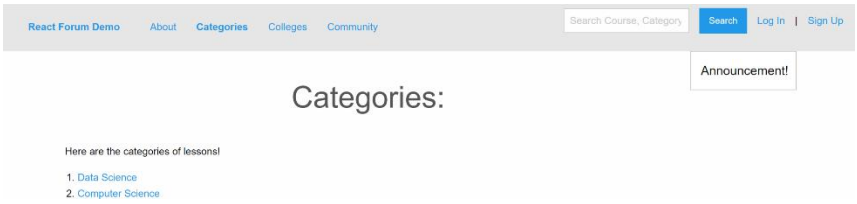

Figure 3.2 Screenshot of my dashboard

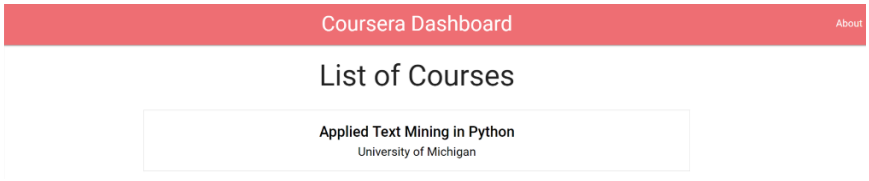

Figure 3.3 Screenshot of the integrated dashboard

3.3 Crawler

3.3.1 Design

The structure of IVLE's entity relationship model is slightly different from the ER diagram from figure 3.1 above. I use a bottom-up approach to find the key information that I need. To find the author, date, title and content of a post, I need the thread_id to which the post belongs. To find the thread_id of a thread, the respective forum_id is required. To obtain the forum_id, I need to find the course_id of the course to which the forum belongs. Therefore, after the crawler manages to log in to my IVLE account, it will crawl the key information from the course list to individual posts. The data crawled for each course will be organized into objects and stored into a json file. In the last step, the json files will be converted into sqlite databases.

3.3.2 Challenges

The training data's database structure for Chandrasekaran's machine learning model is designed for the data crawled from Coursera. The data fields required for training can be referred from the figure 3.4 below. However, there are a lot of differences in the database structure between the data crawled from IVLE and the data from Coursera. For instance, in Coursera, it is clearly indicated whether the post is responded by an instructor, mentor or student in the data field 'answerBadge'. In IVLE, the type of the user who responds to a post is not identified. I have to import a name list of all the teaching staffs and match the responder's username to the name list to identify if the post is responded by instructor or student. In addition, due to the difference in database format between the two platforms, I need to re-organize the data crawled from IVLE to accommodate the Coursera database format.

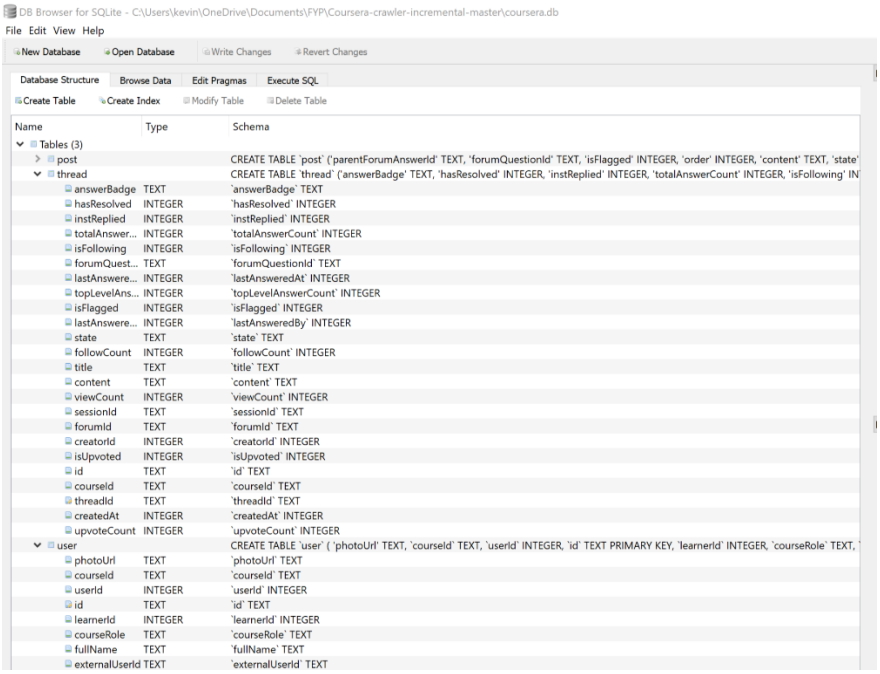

Figure 3.4 Database table for the data crawled from Coursera

Another challenge is that IVLE has 2 types of forum display as shown in the figure 3.5 and figure 3.6. I have to enable the crawler to recognize the forum types and develop two crawling methods to accommodate each type of forums.

3.3.3 Results

In the end, I have managed to overcome the challenges above and delivered the IVLE database to train the machine learning model.

4. PLAN

- Integrate the database crawled from IVLE with the dashboard. I need to make sure that the dashboard can fully display the data in the correct format and structure.

- Beautify the dashboard and make it ready for trials. I need to make sure that the dashboard's interface is user-friendly and decent for teaching staffs to use.
- In my final thesis, I need to include a full evaluation of the integrated application as my research component.
- If the progress goes faster than expected, I would need to study the machine learning model deeper and do some research so that I can contribute my input to the model.

5. REFERENCES

Catropa, D. (2013, Febuary 24). *Big (MOOC) Data*. Retrieved from Inside Higher Ed: https://www.insidehighered.com/blogs/stratedgy/big-mooc-data

Chandrasekaran, M. K., Epp, C. D., Kan, M.-Y., & Litman, D. (2017). Using Discourse Signals for Robust Instructor Intervention Prediction. *the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).* San Francisco.

Chandrasekaran, M. K., Kan, M.-Y., Tan, B., & Ragupathi, K. (2015). Learning Instructor Intervention from MOOC Forums: Early Results and Issues. *8th International Conference on Educational Data Mining*. Madrid.

Chen, H., Phang, C. W., Zhang, C., & and Cai, S. (2016). WHAT KINDS OF FORUM ACTIVITIES ARE IMPORTANT FOR PROMOTING LEARNING CONTINUANCE IN MOOCS? *PACIS 2016 Proceedings*, (p. 51).

Dolan, V. (2014). Massive online obsessive compulsion: What are they saying out there about the latest phenomenon in higher education? *The International Review of Research in Open and Distributed Learning*, 15(2), pp. 268-281.

Pappano, L. (2012, Novamber 2). *The New York Times.* Retrieved from The New York Times Web: http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=0

Zhou, J. (2017). Exploring the factors affecting learners' continuance intention of MOOCs for online collaborative learning: An extended ECM perspective. *Australasian Journal of Educational Technology*, 33(5).