MOOClens: A Peek into MOOCs for Picking MOOCs

Gaurav Kumar Singh, Abhay Doke, Varun Kumar, Savita Bhat, Niranjan Pedanekar[1]
Tata Research Development and Design Centre, Tata Consultancy Services Limited, Pune, India 411013

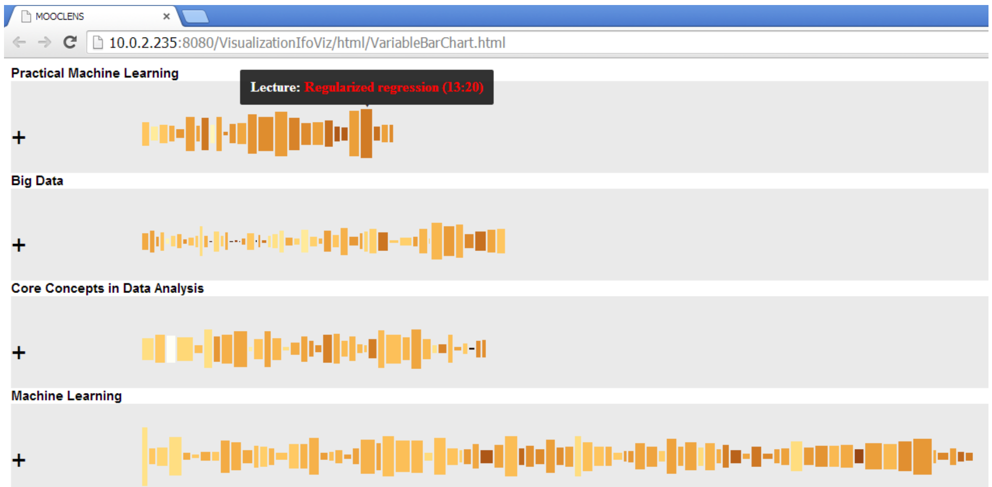

Figure 1: Lecture stacks of 4 online courses using MOOClens, *viz.* A) Practical Machine Learning (uppermost), B) Big Data, C) Core Concepts in Data Analysis and D) Machine Learning

Abstract

A Massive Open Online Course (MOOC) is a series of video lectures delivered by experts and hosted on the internet for viewing. Given the variety of MOOCs available for a variety of subjects, choosing a MOOC can be a challenge for a learner. In this paper, we propose MOOClens, a visualization tool for comparing and exploring courses using parameters such as difficulty, length of lectures and conceptual content. It uses a horizontal stack visualization for course and lecture overview, and a heat map for understanding the important concepts in a course.

Keywords: MOOCs. Stack visualization. Heat map. Text mining.

Index Terms: Human-centered computing; Visualization; Visual application domains; Information visualization.

1 Introduction

A Massive Open Online Course (MOOC) typically consists of a series of online video lectures delivered by experts. A large number of MOOCs are available and can reach a massive number of learners at low cost. Coursera, the largest provider of MOOCs, hosts more than 600 courses. Multiple MOOCs by different instructors are often available for the same subject. MOOCs vary in parameters such as duration, number of lectures, number of quizzes, difficulty level, workload and coverage of concepts. Choosing the right MOOC is an important problem for learners as evident by presence of course selection sites such as MOOCAdvisor, Mooctivity and Class Central. These sites provide search on selection parameters to narrow down the learner's choice. Yet, the reduced choice often involves a number of courses to choose from. These sites do not offer help for exploring concepts in the course to provide further assistance in choosing. A tool such as VAST MM [1] allows visualization of concepts across video lectures of a course, but does not provide an overview of courses and a means of comparing one course to another.

In this paper, we propose MOOClens, a tool for learners to visualize and compare MOOCs in terms of overall characteristics as well as content. We attempt to provide answers to the following important questions using MOOClens:

- How difficult is a course / its lectures? How long are they?
- How does the course progress in terms of difficulty?
- What are the most important concepts in a course / lecture?
- Which other courses / lectures teach the same concepts?

2 Data for MOOClens

For demonstrating MOOClens, we used 4 MOOCs from Coursera obtained by querying the term *machine learning*. We obtained length of lectures and lecture titles for each course. We also acquired course transcripts providing subtitles for the lectures with timestamps for identifying concepts taught in the lectures. For each course, we found the concepts in each lecture by locating words or phrases with a valid Wikipedia entry using Wikipedia Miner [2]. We found the importance of each concept in each lecture using term frequency-inverse document frequency (*tf-idf*) metric [3] to assign more importance to concepts occurring frequently in a lecture, but not so frequently in other lectures. We used two metrics for content difficulty, viz., unfamiliarity and concept count. For unfamiliarity, we used the negative of the global familiarity metric [4] based on the number of pages referring to a concept's page in Wikipedia. We

[1] n.pedanekar@tcs.com

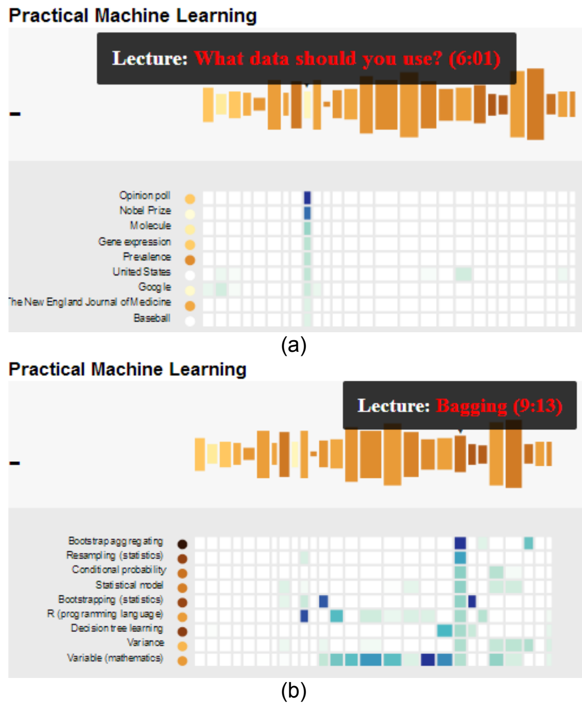

Figure 2: Concept maps for easy (a) and difficult (b) lectures

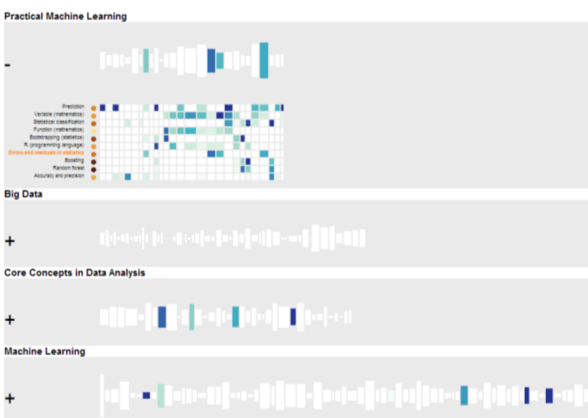

Figure 3: Where does a concept appear in all other courses?

then calculated the average unfamiliarity for each lecture by weighing each concept unfamiliarity with its frequency in that lecture. For concept count, we counted the number of unique concepts in a lecture. A higher concept count in a shorter lecture with higher average unfamiliarity typically indicated higher difficulty for a lecture. We automated the process for data collection and made it scalable.

3 Design of MOOClens

For providing an overview of MOOCs, we used a horizontal stack visualization as seen in Figure 1. Each 'lecture stack' represented a course and consisted of a series of rectangles, each representing a lecture. The width of the rectangle corresponded to the duration of the lecture, while the height to the unique concept count in that lecture. So, the aspect ratio of the rectangle indicated the concept density of the lecture. The taller the rectangle, the more was the concept density, while the flatter the rectangle, the less was the concept density. The colour of the rectangle corresponded to the average unfamiliarity of the lecture, darker hues representing more unfamiliar concepts. The stack was centred along the horizontal centre line of the rectangles. A mouse over each rectangle provided a tooltip with details about the lecture such as title and duration.

Figure 1 shows the lecture stacks for four machine learning MOOCs. Course A had the smallest duration, while course D had the longest. Course A had longer lectures and more concepts in each lecture. The average unfamiliarity for the most of course A was high as indicated by the darker colour. In contrast, course B had smaller lectures, covered less number of concepts per lecture and used more familiar concepts. So, an advanced learner who wanted to learn the subject in shorter time could go for course A, while a beginner could go for the easier course B. A learner wanting a more comprehensive understanding with more time to spare could go for course D. It can be noted that the courses became more difficult as they progressed.

After an overview, the learner would want to explore and compare the courses in terms of content. For looking into individual lectures, we provided a click-based drill-down on lectures as seen in Figure 2. Upon clicking a lecture in the stack, a concept heat map was shown that consisted of rows of rectangles with each rectangle representing a lecture and each row corresponding to a concept. We showed top 10 concepts having the highest average importance in the lecture as 10 rows. The colour of each rectangle showed the importance of a concept in the lecture, white corresponding to low importance and dark blue to high importance. We placed a filled circle next to the concept name indicating the unfamiliarity of the concept, darker circles being more unfamiliar. Figure 2(a) shows top 10 concepts in an easy lecture ('*what data should I use?*') which were mostly examples such as *molecule* and *baseball*. Figure 2(b) shows top 10 concepts in a difficult lecture ('*bagging*') which were advanced concepts in machine learning such as *bootstrapping*. For understanding the most important concepts in the whole course, we provided a similar concept heat map listing the most important topics in the whole course upon clicking an expand button.

A learner would need to compare courses in terms of concept coverage. For this we provided a drill-down, where hovering on a concept name from a course (shown in orange text) highlighted lectures in the lecture stacks for other courses with a colour indicating its importance in the lectures. For example, Figure 3 shows that a top concept from course A appeared in some lectures in courses C and D as an important concept. This feature can be used to find concepts unique to a course or common among courses.

4 Conclusion

In this paper, we presented MOOClens, a tool to compare and explore MOOCs. It allows learners to compare courses in terms of difficulty, duration and conceptual content, and make a choice. In future, we would like to conduct experimental evaluation of MOOClens with a variety of MOOCs and learning material.

References

[1] Alexander Haubold, and John R. Kender, "Vast MM: Multimedia Browser for Presentation Video", Proc. of the 6th ACM Intl. Conf. on Image and Video Retrieval, ACM, pp. 41-48, 2007.

[2] David Milne, and Ian H. Witten, "Learning to Link with Wikipedia", Proc. of the 17th ACM Conf. on Information and Knowledge Management, ACM, pp. 509-518, 2008.

[3] Gerard Salton, and Chung-Shu Yang, "On the Specification of Term Values in Automatic indexing", Journal of Documentation, 29(4), pp. 351-372, 1973.

[4] Gaurav K. Singh, et al., "Assessing the Need of Augmenting Video Lectures with Supporting Information." Learning and Collaboration Technologies. Designing and Developing Novel Learning Experiences. Springer International Publishing, pp. 238-249, 2014.