

Understanding Student Engagement in Large-Scale Open Online Courses: A Machine Learning Facilitated Analysis of Student's Reflections in 18 Highly Rated MOOCs

Khe Foon Hew, Chen Qiao et Ying Tang

Volume 19, numéro 3, juillet 2018

URI : <https://id.erudit.org/iderudit/1051258ar>
DOI : <https://doi.org/10.19173/irrodl.v19i3.3596>

[Aller au sommaire du numéro](#)

Éditeur(s)

Athabasca University Press (AU Press)

ISSN

1492-3831 (numérique)

[Découvrir la revue](#)

Citer cet article

Hew, K., Qiao, C. & Tang, Y. (2018). Understanding Student Engagement in Large-Scale Open Online Courses: A Machine Learning Facilitated Analysis of Student's Reflections in 18 Highly Rated MOOCs. *International Review of Research in Open and Distributed Learning*, 19(3).
<https://doi.org/10.19173/irrodl.v19i3.3596>

Résumé de l'article

Although massive open online courses (MOOCs) have attracted much worldwide attention, scholars still understand little about the specific elements that students find engaging in these large open courses. This study offers a new original contribution by using a machine learning classifier to analyze 24,612 reflective sentences posted by 5,884 students, who participated in one or more of 18 highly rated MOOCs. Highly rated MOOCs were sampled because they exemplify good practices or teaching strategies. We selected highly rated MOOCs from Coursetalk, an open user-driven aggregator and discovery website that allows students to search and review various MOOCs. We defined a highly rated MOOC as a free online course that received an overall five-star course quality rating, and received at least 50 reviews from different learners within a specific subject area. We described six specific themes found across the entire data corpus: (a) structure and pace, (b) video, (c) instructor, (d) content and resources, (e) interaction and support, and (f) assignment and assessment. The findings of this study provide valuable insight into factors that students find engaging in large-scale open online courses.

Copyright (c) Khe Foon Hew, Chen Qiao, Ying Tang, 2018



Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

<https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

<https://www.erudit.org/fr/>

July – 2018

Understanding Student Engagement in Large-Scale Open Online Courses: A Machine Learning Facilitated Analysis of Student's Reflections in 18 Highly Rated MOOCs



Khe Foon Hew, Chen Qiao, and Ying Tang

The University of Hong Kong

Abstract

Although massive open online courses (MOOCs) have attracted much worldwide attention, scholars still understand little about the specific elements that students find engaging in these large open courses. This study offers a new original contribution by using a machine learning classifier to analyze 24,612 reflective sentences posted by 5,884 students, who participated in one or more of 18 highly rated MOOCs. Highly rated MOOCs were sampled because they exemplify good practices or teaching strategies. We selected highly rated MOOCs from *Coursetalk*, an open user-driven aggregator and discovery website that allows students to search and review various MOOCs. We defined a highly rated MOOC as a free online course that received an overall five-star course quality rating, and received at least 50 reviews from different learners within a specific subject area. We described six specific themes found across the entire data corpus: (a) structure and pace, (b) video, (c) instructor, (d) content and resources, (e) interaction and support, and (f) assignment and assessment. The findings of this study provide valuable insight into factors that students find engaging in large-scale open online courses.

Keyword: MOOCs, massive open online courses, engagement, text mining, machine learning

Introduction

Online learning allows students to gain access to education despite spatial and temporal restraints. One of the most significant developments of online learning is the emergence of massive open online courses, or MOOCs for short. Taking a MOOC is convenient, flexible and economic for any individual with an Internet connection. According to the statistics presented by Class Central, a free directory of online courses that helps users find and track MOOCs, there were more than 6,850 MOOCs being offered by 700 universities as of December 25, 2016 (Shah, 2016). This data is presented in Figure 1.

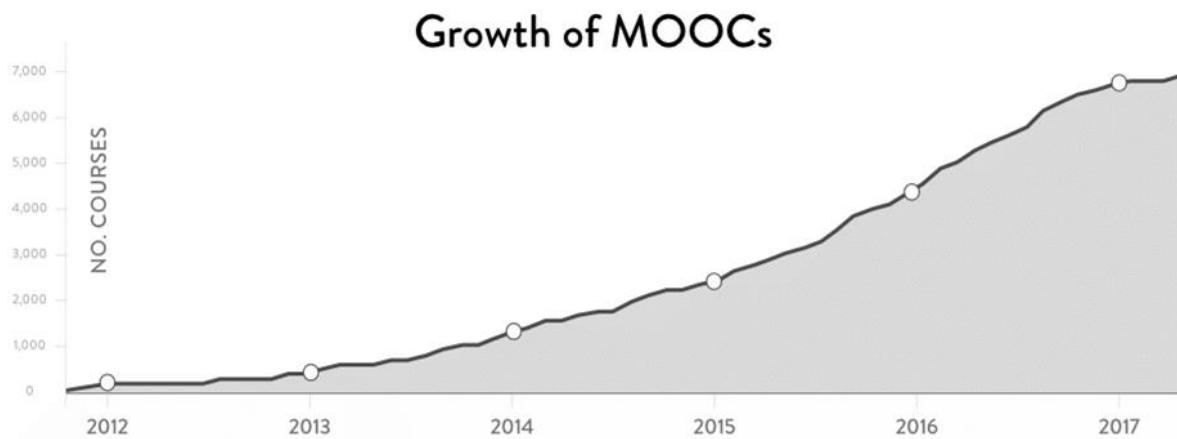


Figure 1. Growth of MOOCs as presented by Shah (2016). From “By the numbers: MOOCs in 2016,” D. Shah, 2016 (<https://www.class-central.com/report/mooc-stats-2016/>). In the public domain.

Although the emergence of MOOCs has fueled much attention among researchers and educators around the world, our understanding of student engagement in these large open online courses is still limited (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Compared to conventional online courses, the task of engaging students in large-scale open online learning environments is often more challenging (Hew, 2016). In conventional online courses, learners usually share the same academic goals, are familiar with one another, and are supervised closely by the teacher (Chiu & Hew, 2018). However in MOOCs, learners do not know most of their peers, are not supervised by the teacher, and are under no expectation to complete the course (Chiu & Hew, 2018).

The main purpose of this study was to identify which aspects of MOOCs participants found engaging, by analyzing a large data set of participant qualitative comments. More specifically, we set out to offer a new contribution by testing a set of five machine learning automatic classification models (k-Nearest Neighbors, Gradient Boosting Trees, Support Vector Machines, Logistic Regression, and Naïve Bayesian). The best performing model was employed to analyze 24,612 reflective sentences posted by 5,884 students, who participated in one or more of 18 highly rated MOOCs. To the best of our knowledge, this is the first work that mobilized a repertoire of analytical and technological resources in the fields of text data mining and machine learning to analyze a large dataset of MOOC students’ reflective comments. The scalable algorithmic approach in machine learning freed up human labor, and enabled us to analyze large data corpus at a scale that would be infeasible by human annotations.

Before explaining our use of the machine learning automatic classifier in detail, we first describe the aspects of a technology-based environment, which may aid in student engagement, by using the framework of engagement theory (Kearsley & Schneiderman, 1998). Second, in the Literature Review section, we provide a brief review of previous MOOC research, followed by a discussion of the current research gaps regarding student engagement in MOOCs. Third, in the Method section, we explain in detail how we selected 18 highly rated MOOCs, collected participant reflective comments of these MOOCs, and analyzed the comments. Finally, we present the results, followed by the discussion and conclusion.

Engagement Theory

Student engagement may take many forms, such as attending classes (behavioral engagement), asking questions (cognitive engagement), and/or expressing enjoyment towards the course activities or instructors (emotional engagement; Fredricks, Blumenfeld, & Paris, 2014).

One frequently cited theory that serves as a useful conceptual framework to understand teaching and learning in a technology-based environment is Engagement Theory (Kearsley & Shneiderman, 1998). Engagement theory posits three primary elements to accomplish student engagement: (a) Relating, (b) Creating, and (c) Donating. The role of technology in this theory is to help facilitate engagement in ways that may be difficult to achieve otherwise (Kearsley & Shneiderman, 1998).

The first element, *Relating*, emphasizes peer interaction whereby students exchange ideas or opinions with other students, enabling learners from different backgrounds to learn from one another (Kearsley & Shneiderman, 1998). The second element, *Creating*, refers to the “application of ideas to a specific context” (Kearsley & Shneiderman, 1998, p. 20), such as students discussing a case study on a wiki (Hazari, North, & Moreland, 2009). The third element, *Donating*, refers to the use of authentic learning environment that has strong connections to the real world (Kearsley & Shneiderman, 1998). This principle is particularly valuable for adult learners, who expect immediate application of knowledge learned in class. By accomplishing authentic tasks, students can transfer in-class content and see the immediate implementation of this knowledge (Kearsley & Shneiderman, 1998). The authenticity of a task can boost students’ satisfaction and motivation (Keller, 1987). Since its inception, engagement theory has been referred to in a variety of conventional online education contexts (Beldarrain, 2006; Bonk & Wisner, 2000; Hazari et al., 2009; Knowlton, 2000; Sims 2003). However, hitherto, engagement theory has not been used to analyze how MOOCs engage participants. As within a conventional e-learning course, learning in MOOCs also happens online. However, as previously explained, MOOCs and conventional e-learning courses are dissimilar in their nature. MOOCs are characterized by free access, and massive open participation. Students can choose to enroll or drop out of MOOCs at any time they wish without incurring any penalty. Do the elements espoused in Engagement theory also apply to MOOC-specific contexts? Which element (i.e., relate, create, donate), if any, is considered most engaging by MOOC students? What additional elements are considered engaging to MOOC students? Answers to these questions can help enrich our understanding of MOOC engagement as well as extend our perspective of Engagement Theory.

Literature Review

In this section, we provide a brief review of previous MOOC studies. This is followed by a discussion of the current knowledge gaps pertaining to student engagement in MOOCs.

Currently, most previous research studies on MOOCs can be parsimoniously grouped into five major categories: (a) impact of MOOCs on institutions, (b) student motives for signing up for MOOCs and reasons for dropping out, (c) instructor motives and challenges of teaching MOOCs, (d) click-stream analysis of log data, and (e) types of MOOCs. Each of these categories will be briefly discussed in the following paragraphs.

The advent of MOOCs has caused concerns to many universities and libraries. Gore (2014), for example, examined the new challenges faced by librarians in light of MOOCs, and discussed a number of challenges that librarians may face as MOOCs become more widespread. These challenges include licensing and copyright, and delivering remote services. Lombardi (2013) examined the types of decisions undertaken by Duke University, an institution which attempted to capitalize on the opportunities and challenges presented by MOOCs. . Decisions revolved around questions such as how well does partnership with Coursera (a MOOC platform provider) align with the University's academic goals, and how might this partnership promote a sustainable model to advance open education as a social good?

Hew and Cheung (2014) reviewed 25 studies to understand the motivation and challenges of instructors' and students' use of MOOCs. Their study revealed four main reasons of student motives for signing up for a MOOC, including the wish to learn something new, to expand existing knowledge reservoir, to challenge themselves, and to get a MOOC completion certificate. Reasons for students dropping out include difficulty in understanding the subject material, insufficient support, and having other priorities over the course.

Instructors' motives for teaching MOOCs include the desire to enhance their professional reputation, to provide opportunity for students around the world to access their courses (Kolowich, 2013). Challenges of teaching MOOCs include lack of student feedback, lack of online forum participation, and heavy burden of time and effort in developing and implementing the MOOCs (Hew & Cheung, 2014). In a study by Baxter and Haycock (2014) regarding MOOC online forum participation, most students only posted intermittently in the online forum. Those who considered themselves frequent contributors only accounted for about 6% of 1,000 randomly selected students (Baxter & Haycock, 2014).

Other studies used click-stream data to investigate student online activities during MOOCs. For example, previous studies found that more students watched videos than worked on course assignments, and that the number of student participation deteriorated steadily as the weeks progressed (Coffrin, de Barba, Corrin, & Kennedy, 2014). Previous studies also found that frequency of forum postings and quiz attempts positively correlated with student MOOC grades (Coetzee, Fox, Hearst, & Hartmann, 2014; de Barba, Kennedy, & Ainley, 2016). Other studies attempted to propose models or methods to predict MOOC dropout (Kloft, Stiehler, Zheng, & Pinkwart, 2014).

Finally, other scholars focused on examining the different types of MOOCs. Essentially, MOOCs can be parsimoniously classified into either xMOOCs or cMOOCs. xMOOCs follow a cognitive-behavioral approach (Conole, 2013), while cMOOCs are modeled after the notion of connectivism (Daniel, 2012; Rodriguez, 2012). xMOOCs typically come with a syllabus, a course content that consists of readings, discussion forums, assignments (e.g., quizzes, projects), and pre-recorded instructor lecture videos (Hew & Cheung, 2014). The syllabus, course content, readings, forums and assignments in xMOOCs are predefined by the instructors before the commencement of the course (Hew & Cheung, 2014). In cMOOCs, however, students define the actual course contents as the course progresses, and there is no fixed syllabus

(Rodriguez, 2012). Participants organize their own learning according to different learning goals, and to interact with others, while an emphasis is placed on personalized learning through a personal learning environment (Conole, 2013; Rodriguez, 2012). This could result in more than one topic being examined concurrently (Hew & Cheung, 2014).

Research Gaps

Although the aforementioned studies have provided us with a useful understanding of MOOCs, they fall short of explaining the reasons why participants find a course or certain parts of a course engaging. Many researchers have begun questioning the validity of using traditional metric such as completion or dropout rate to measure whether a MOOC is engaging or not. As previously described, MOOCs are open courses that are usually offered free of charge, and learners who sign up are under no obligation whatsoever to complete the course. Due to time constraints and work commitment, learners may only complete certain course activities (Kizilcec, Piech, & Schneider, 2013) and still find the activities engaging.

Therefore, in order to understand which aspects of MOOCs students find engaging, we need to analyze students' reflective comments about the MOOCs. So far to the best of our knowledge, only two studies were found that focused specifically on student reflection data of MOOCs. Hew (2015) analyzed guidelines of improving the quality of online teaching and learning from four professional councils, as well as qualitatively analyzed 839 participants' comments of two highly rated MOOCs using the grounded approach. By synthesizing the policy guidelines and actual opinions of online learners, Hew (2015) proposed a rudimentary model of engaging online students, covering six dimensions: course information, course resources, active learning, interaction, monitoring of learning, and making meaningful connections. The findings of this study show us an emerging picture of what is valued by MOOC students.

Another grounded approach study analyzed 965 course participants' reviews on three top-rated MOOCs in the subjects of literature, arts and design, and programming language to find out what factors about the course engaged students, and contributed to their favorable consideration of the online learning experience (Hew, 2016). Five factors were listed: problem-centric learning, instructor accessibility and passion, active learning, peer interaction, and helpful course resources. This study set a foundation of what are worthwhile factors to consider when instructors prepare and deliver a MOOC.

The limitation of these two study lies in the fact that only two or three MOOCs were inspected respectively, which, as the author noted, is not sufficient to warrant strong conclusions (Hew, 2016). Nevertheless, the results of these two studies provide a useful conceptual basis for other researchers to conduct studies to examine students' reflection comments.

Method

This study aims to answer the following question: What elements pertaining to the course design or the instructor did students find enjoyable, helpful in learning the materials, or motivational (motivating them to take part in the activities)? In this section, we explain how we chose the 18 highly rated MOOCs, collected the MOOC participants' reflective comments, and analyzed the comments. An overview of the whole data collection, processing, and analysis procedure is shown in Figure 2.

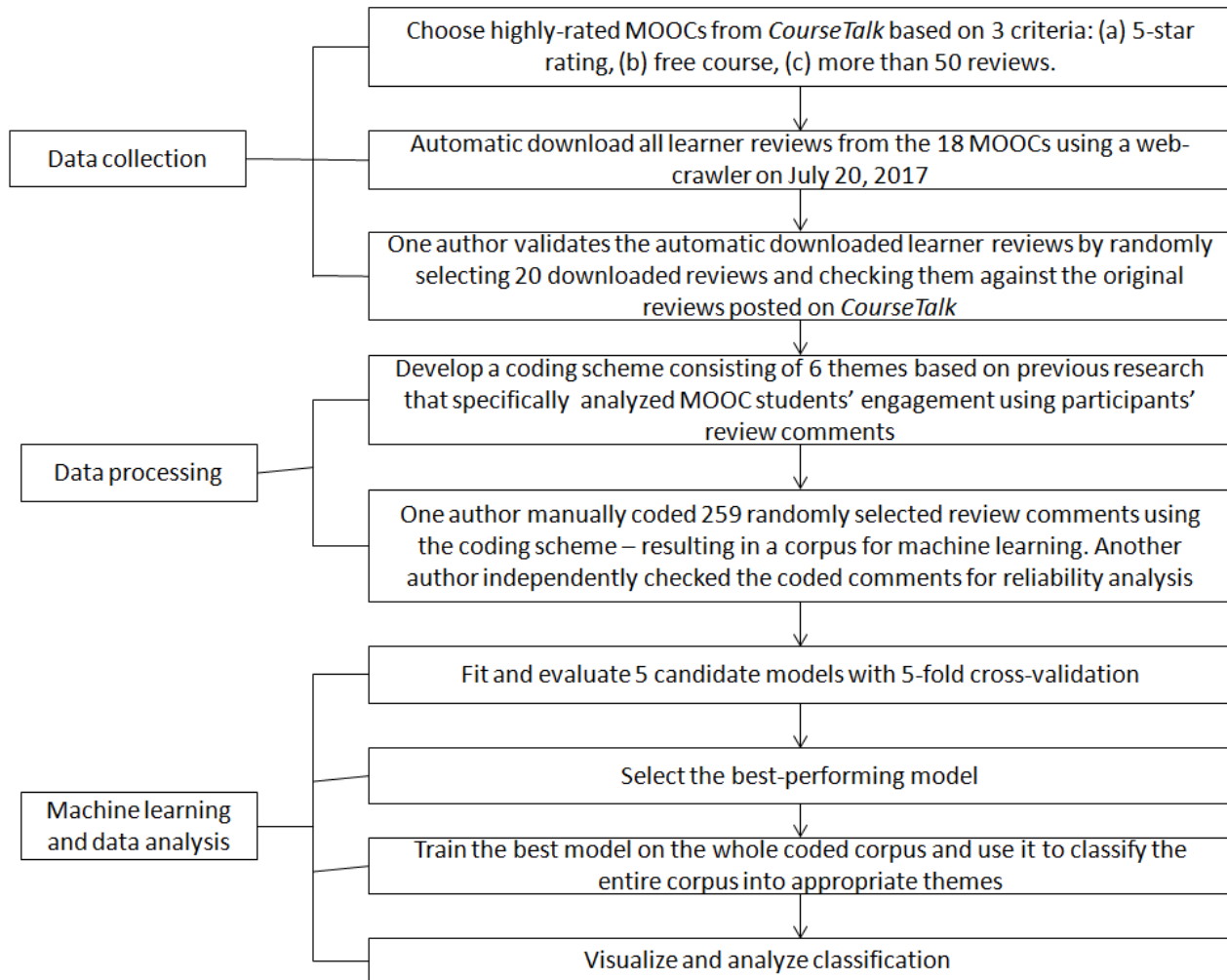


Figure 2. Overview of the method.

Data Collection

Highly rated MOOCs were sampled because they exemplified good practice or teaching strategies. We selected “highly rated” MOOCs from *Coursetalk*, an open user-driven aggregator and discovery website that allows students to search and review various MOOCs. We defined a highly rated MOOC as a course that received an overall five-star course quality rating, and received at least 50 reviews (from different learners) within a particular subject discipline. Using reviews from many different participants provided data triangulation, which promotes trustworthiness of the ratings (Shenton, 2004).

Coursetalk was chosen because it is considered the largest platform connecting learners to courses (Business Wire, 2014). When this study was conducted, it listed more than 50,000 courses. Visitors to the website can rate a course from three dimensions: course content, course instructor, and course provider. One to five stars can be given, with five stars being the highest quality. The website will calculate an individual user’s rating across the three categories, and further aggregate an overall rating among all participants. Therefore, a course will have a general rating score, as well as a detailed list of ratings from specific participants. In addition, participants can write down review comments.

Using data from *Coursetalk*, we searched all 282 subject disciplinary areas, and applied the following selection criteria to identify eligible student reviews: free of charge, rated five-star, and with more than 50 reviews. As of July 20, 2017, 18 highly rated MOOCs were identified. We provide an overview of the 18 MOOCs in Table 1.

One of the authors wrote a web crawler to automatically download all the reviews of the 18 MOOCs. Given the URLs of the 18 course pages, the crawler was launched at 9:46 p.m. on July 20th, 2017 and obtained a total of 5,884 learners' reflective posts, with 24,612 sentences generated on or before that day. Of the 5,884 learners, 90.9% completed one or more MOOCs, 8.3% were currently taking a MOOC, and 0.8% were drop outs. Once the entire data had been downloaded, another researcher randomly selected 20 of the downloaded reviews and checked them against the original reviews posted on *CourseTalk*. This procedure served to establish the reliability of the data collection process. The percent agreement was 100%.

Table 1

MOOCs Reviewed in This Study

| | Subject discipline area | MOOC title | University | Ratings and number of reviews | Purpose of the course |
|----|-------------------------|--|------------------------------|-------------------------------|--|
| 1 | Computer sciences | An Introduction to Interactive Programming in Python | Rice University | 5*, 3077 reviews | Develops simple interactive games (e.g., Pong) using Python. |
| 2. | Social sciences | The science of everyday thinking | The University of Queensland | 5*, 1167 reviews | Explores the psychology of our everyday thinking, such as why people believe weird things, and how we can make better decisions. |
| 3. | Natural sciences | The science of the solar system | Caltech | 5*, 303 reviews | Explores Mars, the outer solar system, planets outside our solar system, and habitability in our neighborhood and beyond. |
| 4. | Environmental sciences | Introduction to environmental science | Dartmouth College | 5*, 233 reviews | Surveys environmental science topics at an introductory level, ultimately considering the sustainability of human activities on the planet. |
| 5. | Design | Design: creation of artifacts in society | University of Pennsylvania | 5*, 217 reviews | Focuses on the basic design process: define, explore, select, and refine. Weekly design challenges test student ability to apply those ideas to solve real problems. |
| 6. | Literary art | Modern and contemporary American poetry | University of Pennsylvania | 5*, 171 reviews | Introduces modern and contemporary U.S. poetry, with an emphasis on experimental verse, from Dickinson and Whitman to the present. |

| | | | | | |
|-----|-------------------|---|--|-----------------|--|
| 7. | Cybersecurity | Cybersecurity fundamentals | Rochester Institute of Technology | 5*, 168 reviews | Introduces essential techniques in protecting systems and network infrastructures, analyzing and monitoring potential threats and attacks, devising and implementing security solutions. |
| 8. | Statistics | The analytics edge | Massachusetts Institute of Technology | 5*, 149 reviews | Explores the use of data and analytics to improve a business or industry. |
| 9. | Management | Finance: time value of money | n.a. | 5*, 90 reviews | Introduces the logic of investment decisions and familiarize with compounding, discounting, net present value, and timeliness. |
| 10. | Computer sciences | HTML5 coding essentials and best practices | The World Wide Web Consortium (W3C) | 5*, 84 reviews | Explains the new HTML5 features to help create great Web sites and applications in a simplified but powerful way. |
| 11 | Management | u.lab: leading from the emerging future | Massachusetts Institute of Technology | 5*, 82 reviews | Introduces a method called Theory U, developed at MIT, for leading changes in business, government, and civil society contexts worldwide. |
| 12. | Art and culture | Drawing nature, science and culture: natural history illustration 101 | The University of Newcastle, Australia | 5*, 81 reviews | Introduces essential skills and techniques that form the base for creating accurate and stunning replications of subjects from the natural world. |
| 13 | Biology | Introduction to biology-the secret of life | Massachusetts Institute of Technology | 5*, 75 reviews | Explores the mysteries of biochemistry, genetics, molecular biology, recombinant DNA technology and genomics, and rational medicine. |
| 14. | Literary art | Comic books and graphic novels | University of Colorado Boulder | 5*, 75 reviews | Presents a survey of the Anglo-American comic book canon and of the major graphic novels in circulation in the United States today. |
| 15. | Social sciences | Justice | Harvard University | 5*, 68 reviews | Explores critical analysis of classical and contemporary theories of justice, including discussion of present-day applications. |
| 16. | Computer sciences | Mobile computing with app inventor—CS principles | Trinity College | 5*, 51 reviews | Explains the open development tool, <i>App Inventor</i> , to program on Android devices, and some of the fundamental principles of computer science. |
| 17 | Social sciences | International human rights law | Université catholique de | 5*, 51 reviews | Examines a wide range of topics including, religious |

| | | | | | |
|----|-------------|-----------------------------|--------------------------------|----------------|---|
| | | | Louvain | | freedom in multicultural societies, human rights in employment relationships, etc. |
| 18 | Environment | Climate change: the science | University of British Columbia | 5*, 50 reviews | Introduces climate science basics such as flows of energy and carbon in Earth's climate system, how climate models work, climate history, and future forecasts. |

Data Processing

To train the machine learning algorithm, we first developed a coding scheme consisting of the following six themes based on previous grounded approach studies that specifically analyzed MOOC students' engagement using participant review comments (Hew, 2015; 2016). These themes, which are not listed in order of importance or priority included:

- Theme 1: Structure and pace,
- Theme 2: video,
- Theme 3: instructor attributes,
- Theme 4: content and resources,
- Theme 5: interaction or community, and
- Theme 6: assignment and assessment.

Table 2

Themes and Descriptors

| Theme | Descriptors |
|---------------------------|--|
| Structure and Pace | Clear objective, duration, structure and syllabus |
| Video | Videos, captions, choice of speed variation, video-integrated quiz |
| Instructor attributes | Instructor knowledge, instructor passion, instructor humor |
| Content and resources | Examples or case studies that relate to the real world, problem-solving centric, relevant and up-to-date course content and resources, Availability of transcript and pdf documentation, slide notes |
| Interaction and support | Student-student interaction, instructor-student interaction, course support |
| Assignment and assessment | Use of active learning strategies such as mini-projects, exercises, quizzes, questions, feedback |

To fulfill the training purpose, an annotated "instructional" dataset is required to form the training materials, consisting of positive and negative cases, on which the automatic machine classifiers "learn" through adapting their model parameters to fit the dataset. A sample of sentences (259) were randomly extracted from the newly collected dataset, and human raters were recruited to manually label these texts using the aforementioned theme labels. One human rater independently labeled the texts using the labels. This was then independently examined by another human rater. Discrepancies among the raters were resolved through discussion.

For example, 25 positive cases of Theme 6 "Assignment and Assessment" were annotated by the human rater, and they were used to train the automatic classifiers. In the training process, the computer might automatically learn that positive cases of this theme may contain certain linguistic cues such as: "(+) quiz" or

“(+) task.” When the machine analyzed the data independently afterwards, if a case contained “(+) quiz” or “(+) task”, most likely the computer would classify it into “Assignment and Assessment”. For example, the comment “the quizzes are detailed and complete and require a bit of extra programming and the mini-project and peer reviews require a few hours of extra effort” was classified as positive within this theme. In the meanwhile, other cases without these cues were possibly considered negative.

To simplify the problem in our scenario, we leveraged the one-versus-the-rest treatment (Bishop, 2006, p. 182) for multi-label classification task, and decomposed the problem into six binary classification tasks, each of which corresponds to a specific theme. To explain, using the one-versus-the-rest treatment, we trained six independent classifiers for the six themes. One classifier could only decide whether a case was positive or negative under the corresponding theme. In other words, it only focused on one theme and did not consider other themes. The six classifiers were adopted independently to process the same data corpus. Note that in our case, the theme labels were not exclusive, hence we were free from the issue of ambiguous classification regions proposed in Bishop (2006, pp. 182-183).

However, regarding a single theme, the fact that all its non-positively-labeled cases were taken as negative cases greatly enlarged the proportion of negative-to-positive data ratios, resulting in data unbalance. To counteract the data unbalance, we conducted data expansion for positive samples in each theme class. Specifically, we extracted all sentences in the rest of the corpus containing the theme-dependent cue terms in a term list (summarized by the annotators). Immediately after that, the annotators checked the new data and removed wrong samples. The sampling and validation procedure repeated until the new cases were all valid. The statistics for the dataset for each theme before and after data expansion are summarized in Table 3. As a result of the data expansion procedure, the sub corpus for each theme was balanced.

Table 3

Annotated Corpus Statistics Before and After Data Expansion

| | Theme 1 | Theme 2 | Theme 3 | Theme 4 | Theme 5 | Theme 6 |
|-----------------------------------|---------|---------|---------|---------|---------|---------|
| Positive samples before expansion | 37 | 28 | 68 | 78 | 36 | 25 |
| Negative samples before expansion | 222 | 231 | 191 | 181 | 223 | 234 |
| Corpus size after expansion | 444 | 462 | 382 | 362 | 446 | 468 |

Data Analysis and Testing of Models

The machine learning goal in our scenario can be formalized as follows. The objective is to fit a series of binary classifiers $\mathbf{f} = [f_1, \dots, f_N]$ (where f_i corresponds to the i th theme, and $N=6$ is the number of themes) on the annotated dataset D , and then apply \mathbf{f} to predict theme labels on the rest texts. Given a text \mathbf{x} and the trained classifiers \mathbf{f} , the themes of \mathbf{x} are predicted as $\mathbf{y} = \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_N(\mathbf{x})]$, where $\mathbf{y} \in \{1,0\}^N$ is a N -dimensional vector of zeros and ones with the i th dimension indicating whether the i th theme is assigned to the text. Values 1 and 0 in \mathbf{y} designate true and false respectively.

As the data representation building block, the design matrix (Murphy, 2012, p. 2) in our context is implemented with TF-IDF features, i.e., a matrix of which columns and rows represent terms and documents respectively, and each cell in the matrix records a value computed on the frequency of the term in the document (measuring its degree of popularity within the document), weighted by the term's inversed document frequency (measuring its degree of rarity in the whole corpus) (Wu, Luk, Wong, & Kwok, 2008). TF-IDF is a common technique of text representation and can filter out stop words and keep the most discriminant terms (Robertson, 2004).

On the other hand, due to what has been claimed in the *no free lunch theorem* (Box & Draper, 1987, p. 424), we could not guarantee a best performed classification model type beforehand, since there is not a best model which can outperform all the other models on all problems. As a result, we prepared a set of candidate model types and expected to find the best performing one. The candidate classifiers included:

1. k-Nearest Neighbors (KNN) (Altman, 1992),
2. Gradient Boosting Trees (GBT) (Friedman, 2001),
3. Support Vector Machines (SVM) (Cortes & Vapnik, 1995),
4. Logistic Regression (LR) (Cox, 1958), and
5. Naïve Bayesian (NB) (Murphy, 2012).

To evaluate the classification performance in a comprehensive stance, we adopted five metrics. In these metrics, the first four measurements are calculated on the confusion matrix (Stehman, 1997), which records predicted labels and ground truths and is primarily for measuring correctness from different angles. The kappa value (Cohen, 1960) measures the classification consistency between the learning classifiers and the human annotators:

1. Accuracy: the proportion of correct predictions (both positive and negative) in all cases;
2. Precision: the proportion of correct predictions in all cases predicted as positive.
3. Recall: the proportion of true positive cases predicted as positive in all true positive cases;
4. F1: the harmonic mean of precision and recall, which balances the measurements of 2) and 3)
5. Cohen's kappa: the agreement between two raters (the machine trained classifier and human annotators in our scenario (Cohen, 1960)).

We implemented the machine learning and evaluation experiment using Python programming language. The classifiers were implemented with the Scikit-learn packageⁱ, and the texts were segmented, tokenized and cleaned with the spaCyⁱⁱ natural language processing toolkit, before the design matrix was constructed via the Scikit-learn TF-IDF feature extraction tool.

The experiment was conducted on an Ubuntu16.04 system equipped with Intel Core i5-4460 3.20GHz CPU and 16GB memory. We arranged five-fold cross-validation (Bishop, 2006, p. 33) for each classifier on each theme, so that each classifier on each theme obtained five test scores on each metric. Finally, the metric scores were averaged, and the best performed classifiers on each metric and each theme identified (Table 4).

Table 4

Best Performing Automatic Machine Learning Classifiers on Each Theme

| | Theme 1 | Theme 2 | Theme 3 | Theme 4 | Theme 5 | Theme 6 |
|-----------|---------|---------|---------|---------|---------|---------|
| accuracy | GBT | GBT | GBT | GBT | GBT | SVM |
| precision | NB | SVM | LR | LR | LR | NB |
| recall | KNN | KNN | KNN | KNN | KNN | KNN |
| f1 | GBT | GBT | GBT | GBT | GBT | SVM |
| kappa | GBT | GBT | GBT | GBT | GBT | GBT |

In general, the candidate classifiers performed differently in different metrics and for different themes. Whereas, the GBT classifier performed better than all the other classifiers on kappa metrics, demonstrating its superb consistency with the human annotators. The same classifier was also good in f1 scores, meaning that it was relatively balanced in precision and recall. Finally, GBT was very accurate in making positive and negative predictions, with the accuracy score outperforming most of the other classifiers except for Theme 6. Due to its stable and superb performance, we selected GBT as the prediction model for later use. Table 5 presents the performances of GBT.

The GBT achieved sound values in accuracy, f1, precision and recall. In addition, its kappa values were all above 0.61, indicating its substantial agreement with human annotators (McHugh, 2012). All the metrics demonstrate the promising quality of the classifier.

Table 5

Performance Metrics for GBT

| | Theme1 | Theme2 | Theme3 | Theme4 | Theme5 | Theme6 |
|-----------|--------|--------|--------|--------|--------|--------|
| accuracy | 0.8079 | 0.9630 | 0.8289 | 0.8056 | 0.8467 | 0.8991 |
| precision | 0.7868 | 0.9820 | 0.8773 | 0.8625 | 0.8967 | 0.9148 |
| recall | 0.8464 | 0.9435 | 0.7684 | 0.7333 | 0.7836 | 0.8845 |
| f1 | 0.8152 | 0.9623 | 0.8171 | 0.7894 | 0.8361 | 0.8980 |
| kappa | 0.6158 | 0.9261 | 0.6579 | 0.6111 | 0.6934 | 0.7982 |

Results

Full Dataset Prediction and Analysis

Adopting the GBT classification model, we trained six GBT classifiers for each of the six themes, and used them to classify the remaining texts according to the appropriate theme categories. Figure 3 shows how frequently each theme was found in the participants' reflective sentences; the more frequent a certain theme was mentioned in the participants' reflective comments, the bigger the circle size of the theme. Figure 3 also indicates how likely each theme co-occurred with other theme (mentioned together by the participants); the closer the distance between the themes are shown in Figure 3, the more likely they are mentioned together by

the participants. The distances between the themes shown in Figure 3 were not designed a priori to the analysis. To explain how the distance between themes was computed, we provide the following illustration.

Suppose we have the following text: "The exercises are related to the videos and allow the student to progress week after week". This text was inferred correctly by the automatic classifiers to contain Theme 2 (Video) and Theme 6 (Assignment and assessment). The results were then represented with a vector with six dimensions (6-D) indicating the existence or nonexistence of six themes. In this case, the example text would be marked with a 6-D vector (0, 1, 0, 0, 0, 1), representing Theme 1 (leftmost) to Theme 6 (rightmost), where each position represents whether the theme is on (using 1) or off (using 0). In our example, the 2nd and 6th dimensions of the vector were turned on (using 1), indicating the text contained Theme 2 and Theme 6. The other themes (i.e., Themes 1, 3, 4, 5) were turned off (using 0). Now assuming we have the following results from Texts 1 to 4 (Table 6):

Table 6

Examples of Texts and Themes

| Text | Theme1 | Theme2 | Theme3 | Theme4 | Theme5 | Theme6 |
|--------|--------|--------|--------|--------|--------|--------|
| Text 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Text 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| Text 3 | 0 | 1 | 1 | 1 | 0 | 0 |
| Text 4 | 1 | 1 | 1 | 0 | 0 | 0 |

We find that Themes 2 and 3 are more likely to co-occur (they co-occurred in Texts 1, 3, 4), so we could consider Themes 2 and 3 as more inclined to be semantically closer than the other themes. Moreover, we can find that the closeness could be measured by the similarity of the columns of the themes in the above matrix. The Theme 2 column, which is (1,1,1,1) is only one bit different from Theme 3 column, i.e., (1,0,1,1).

Below we give an example showing how we can calculate theme distances via their column vectors. We denote \mathbf{a} , \mathbf{b} to be any column vectors of two themes in Table 6, then using the formula of Euclidean distance metric (which can be used to compute the distance of two vectors), the distance of \mathbf{a} , \mathbf{b} can be computed as: $\text{dist}(\mathbf{a}, \mathbf{b}) = \sqrt{((\mathbf{a} - \mathbf{b})^T (\mathbf{a} - \mathbf{b}))} = \sqrt{((a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_N - b_N)^2)}$, where a_i, b_i ($i = 1, 2, \dots, N$) are the i th elements of the column vectors \mathbf{a} and \mathbf{b} respectively, $\sqrt{(\cdot)}$ is the square root function, and N is the dimension of the column vectors ($N = 4$ in our example, since we have only 4 texts). We can compute the distance between Theme 2 and Theme 3, and other themes, using the above formula, and obtain for example:

$$\begin{aligned} \text{dist}(\text{Theme2}, \text{Theme3}) &= \sqrt{((1-1)^2 + (1-0)^2 + (1-1)^2 + (1-1)^2)} = \sqrt{(0+1+0+0)} = 1, \\ \text{dist}(\text{Theme2}, \text{Theme5}) &= \sqrt{((1-0)^2 + (1-0)^2 + (1-0)^2 + (1-0)^2)} = \sqrt{(1+1+1+1)} = 2, \\ \text{dist}(\text{Theme2}, \text{Theme1}) &= \sqrt{((1-1)^2 + (1-0)^2 + (1-0)^2 + (1-1)^2)} = \sqrt{(0+1+1+0)} \approx 1.414, \end{aligned}$$

Therefore, we can see that the vector of Theme 2 is closer to Theme 3 than to Themes 1 and 5. So when two themes are more likely to co-occur (mentioned together in the participants' comments), their column vectors are closer.

To make it possible to visualize the high-dimensional theme vectors, we conducted Principle Component Analysis (PCA) (Wold, Esbensen & Geladi, 1987) for the matrix and managed to reduce and project the column vectors into two-dimensional space, as presented in Figure 3. Note that PCA not only projects the vectors into the 2-D space, but also maximally keeps the vector closeness information in the original high dimensional space. With the 2-D coordinates yielded by PCA, we are able to draw the themes on x-y axis and generate Figure 3.

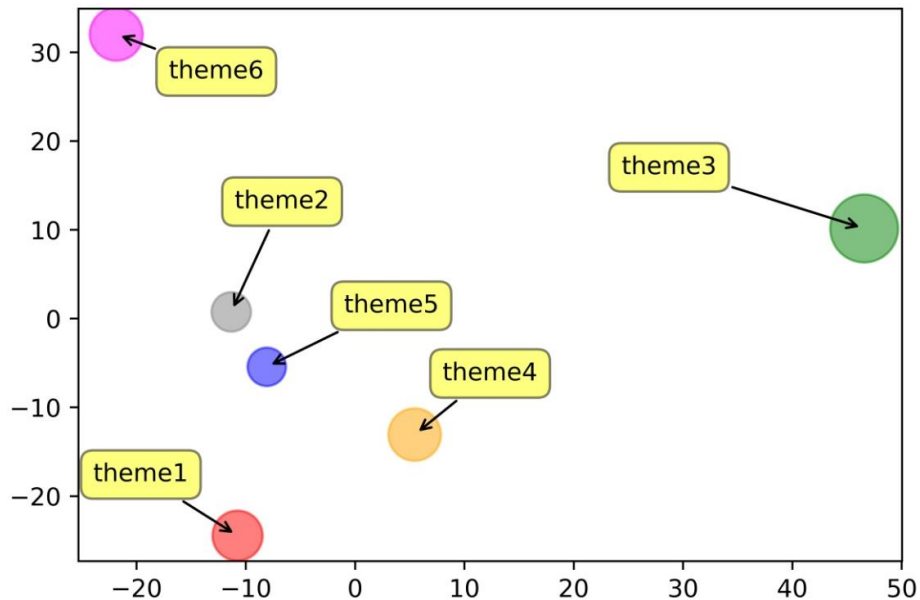


Figure 3. Visualization of the relatedness of the themes. Circle size is proportional to the percentage of the theme in the corpus, and the distances between circles indicate their relatedness in terms of co-occurrence.

It is beyond the length of this paper to list out every single student comment related to each of the seven themes. We therefore provide some representative examples to provide the reader with a closer look at each of the six themes (see Table 7). The main findings can be summarized as follows:

- a) The most frequently mentioned theme was instructor attributes. The other commonly mentioned themes were course content and resources, assignment and assessment, and structure and pace.
- b) Two particular instructor attributes stood out among the many student comments: instructor's passion about the subject as well as teaching it, and instructor's sense of humor.
- c) Students enjoy course content and resources that emphasized real-world application or problem-solving.
- d) Students desire moderately challenging courses assignments that require them to apply the contents learned. Easy assignments or questions that merely test factual recall are disliked. Assignments that are fun and enjoyable (e.g., building simple games) make the tasks more engaging to students.
- e) Student prefer short lecture videos ranging from about five to ten minutes long. Videos consisting of different instructors were perceived to be more engaging. Guest speakers' appearances in videos are generally welcomed and appreciated by students.
- f) Use of in-video quizzes helps sustain students' attention to lecture content.
- g) Students prefer a course structure that builds on each lesson progressively from simple to difficult, and provides options for students to do either the minimum or go deeper into the subject.

Table 7

Themes and Examples of Students' Comments

| Theme | Examples of students' comments |
|--------------------------|---|
| 1. Structure and pace | <ul style="list-style-type: none"> • A very well paced course, with a well thought progression from easy to more difficult. • They provide you with a very structured sequence of steps that make it possible to do what seems like impossible at first. • There is room to do just the minimum, which is good for many, or go a bit deeper. • One has the option of doing the homework that the CalTech students are doing. • The course started slowly and each week builds on the last; very good pace to learn. |
| 2. Video | <ul style="list-style-type: none"> • The video lectures were nice bite sized pieces that I could do on my commute or lunch hour, and re-watch if needed. • The video lectures are very interesting, usually in very short chunks with plenty of examples, demonstrations etc, so it's easy to understand the principle being taught. • I like the short video segments - usually 1-2 main concepts taught. • The video segments were engaging and kept to around 5 minutes each, which allow you to come and go to the class easily. • Having 3 different lecturers every week is quite useful as they do have different teaching styles. • I especially like to in-video quizzes that are auto-graded. • These in-video quizzes help minimize mind-wandering; they also test the concepts or ideas taught to help me remember and understand them better. |
| 3. Instructor | <ul style="list-style-type: none"> • Instructors are obviously very passionate about the topic and have spent a lot of time preparing information for this course! • I also liked the added humor in the lectures to avoid the material from becoming dull. • The instructors enjoy teaching the class so that makes the class easy to immerse yourself into. • The teacher teaches with passion and encourages his students. • All professors have some good sense of humor that makes the learning fun! • His [the instructor] enthusiasm is contagious and it helped to motivate me to complete the quizzes |
| 4. Content and resources | <ul style="list-style-type: none"> • Lander [the instructor] provided multiple real world examples (diseases, court cases, discoveries made in the last decade) that made the material very relevant and modern. • Every discussed concepts of machine learning field was accompanied with real world problems and solutions. • We learned how to apply a design process to solve problems we saw in the world around us. • It is the only course of which knowledge can easily be applied to the real world. • There are actual field trips to power plants, water plants, and farms that help make the material real. • I like the combination of video, transcript and pdf documentation to allow students to review. • The way this class is taught helps all types of learners by having transcripts, captions and videos. |

| Theme | Examples of students' comments |
|------------------------------|--|
| 5. Interaction and support | <ul style="list-style-type: none"> • The support provided for this class is also very sufficient, with a forum, code clinic, support email and the TA who work very hard to try and help out in the forum. • The special activities like live webcasts and facebook interactions from the instructor and TAs provide an extra feeling of connectedness and commitment. • I also really appreciated his [instructor] and his team's direct interaction with students. • The course has a great support group with message boards, TA's, and instructors who are fully invested in helping you learn. • There is also a "code clinic" where the staff provides more direct help to the students. |
| 6. Assignment and assessment | <ul style="list-style-type: none"> • The problem sets were very helpful and challenged me to think logically and apply the material in real world kind of ways. • The exercises made you think and were challenging enough to keep ones interest up. • Progressive weekly assignment that are specifically designed to help students understand concepts taught in lecture. • Small questionnaires that arise at the end of each topic or end of each workweek are very good because they allow us to see which subject we lacked a better understanding and thus focus on it. • Marking scheme for each assignment is clear and easy to understand most of the time. • It was very helpful to have a short "check your understanding" multiple choice questionnaire after each section. |

Discussion

It is interesting to note that all 18 MOOCs can be classified as xMOOCs. To briefly recall, xMOOCs typically come with a syllabus, a course content that consists of readings, discussion forums, assignments (e.g., quizzes, projects), and pre-recorded instructor lecture videos (Hew & Cheung, 2014). The syllabus, course content, readings, forums and assignments in xMOOCs are predefined by the instructors before the commencement of the course (Hew & Cheung, 2014). The results of the present study demonstrate that xMOOCs, which come with a clear instructor-defined course structure, are perceived more positively than cMOOCs by students. Students appreciate a course structure with a well thought progression of pace from easy to more difficult as the weeks advance.

Overall, the most frequently mentioned themes that participants perceived as engaging were instructor attributes. The most common means for an instructor to present course materials in a MOOC is through lecture videos (Young, 2013). Even though instructors may feel awkward being on videos, they can still engage students if they are excited or enthusiastic about the subject matter (Young, 2013). An instructor's enthusiasm for teaching the subject can help break the boredom of watching videos, and even motivate students to complete the activities, as indicated in student feedback : "His [the instructor] enthusiasm is contagious, and it helped to motivate me to complete the quizzes." (Student)

As indicated in student feedback, students also value an instructor's humor because it helps break the boredom and make the lesson more enjoyable, as shown in the following student comments. Use of humor can help arouse students' attention, increase student liking of the professor, establish a positive rapport with

students, and motivate students to participate in the course (Wanzer, 2002): “Their geek sense of humor helped the classes be very interesting and enjoyable” (Student) and “I also liked the added humor in the lectures to avoid the material from becoming dull” (Student).

Course content and resources that emphasized real-world application or problem-solving were other themes that participants perceived as engaging. As outlined by Dillahunt, Wang, and Teasley (2014), a majority of MOOC learners are adult learners who have at least a bachelor’s degree and are employed. Adult learners will be more engaged in learning when new content that is presented is applicable to real-life situations (Knowles, Holton, & Swanson, 2011). This finding implies that instructors should emphasize application of content to real-world practices over mere transmission of information. Instructional strategies such as showing real-life problems to which the principles or solutions taught in the course can be applied, and presenting practical tips are particularly useful because these elements provide valuable add-on insights to students’ learning. In addition, instructors should provide text-based resources such as video transcripts, video captions, and pdf documentations to help students review the course content.

Interestingly, despite the commitment-free nature of MOOCs (e.g., no actual course credit or course fees), students still desire moderately challenging courses assignments that require them to think or apply the concepts or principles learned. One possible explanation for this may be offered by the achievement goal theory. According to achievement goal theorists, there are two main types of goals: (1) the performance goal, which focuses on exhibiting ability in comparison with other people; and (2) the mastery goal, which focuses on developing competence in a particular topic or area (Ames, 1992; Meece, Blumenfeld, & Hoyle, 1988). Since learners in a MOOC do not know most of their peers, it is unlikely that they are motivated by performance goals. We posit therefore that learners in MOOCs are more likely to be motivated by mastery goals. Adopting a mastery goal is believed to produce a desire for moderately challenging tasks, a positive stance toward learning, and enhanced task enjoyment (Elliot & Church, 1997). The implication here is that instructors should avoid simple assignments that merely test factual recall. Instead, instructors should employ strategies such as asking students to apply the concepts learned to solve some real-world problems. The activities should also have varying levels of difficulty so that students can choose an activity that matches their personal ability, while simultaneously providing them with an opportunity to accomplish more difficult tasks in order to master a particular topic or skill.

Other themes that participants perceived as engaging included interaction and support, and video lectures. As reported by participants, interaction and support from the tutors (instructors and/or teaching assistants) helped foster cognitive engagement, which can assist student learning of the topic. Not all the 18 MOOCs have the same degree of interaction and support. MOOCs that had relatively more participant comments about interaction and support such as the Interactive Python Programming and American Poetry courses used one or more of the following strategies:

- a) Providing an opportunity for students to interact with peer raters regarding their submitted assignments. For example, the instructor of the American Poetry MOOC provided an opportunity for students to interact with peer reviewers (e.g., seek clarification) regarding their submitted assignments. Students in the Interactive Python Programming MOOC could discuss the quiz problems with other students with the only rule being that they could not post explicit answers in the forum (Warren, Rixner, Greiner, & Wong, 2014).

- b) Organizing live online sessions for instructor and fellow students to exchange answers and ideas. For example, observation of the American Poetry MOOC revealed that the teaching staff of the said MOOC offered live webcasts every week for students around the world to join the live discussions through the telephone, Facebook, and course forums.
- c) Hold weekly one-hour virtual office hours. For example, the teaching staff of the American Poetry MOOC organized one-hour virtual office hours every week in the discussion forum to answer student questions. The Python MOOC used a professional help desk service (<http://helpscout.net>) to manage student email inquiries. Specifically, this desk service routed student help requests to a course website called “Code Clinic” so that the teaching staff could respond to them (Warren et al., 2014). The professional help desk service provides several useful features such as collision detection to prevent duplicate replies, an auto reply to let students know their email request has been received, auto tracking of student requests, and customizable stock responses to common questions (Warren et al., 2014; <http://helpscout.net>). These features helped the instructors managed the students’ email inquiries (Warren et al., 2014).

A variety of video production styles were used in the 18 MOOCs. These video production styles may be categorized under one or a combination of the following labels (Guo, Kim, & Rubin, 2014, p. 44):

- a) Slides: PowerPoint slide presentation with voice-over.
- b) Code: video screencast of the instructor writing code in a text-editor.
- c) Khan-style: video of instructor drawing freehand on a digital tablet.
- d) Classroom: video captured from a live classroom lecture.
- e) Studio: instructor recorded in a studio with no audience.
- f) Office desk: close-up shots of instructor’s head filmed at an office desk.

Despite the different video production styles, the following three findings should be noted:

- a) First, students prefer short videos. Findings from a study that analyzed 6.9 million video watching sessions across four MOOCs revealed that the median video watching time was six minutes (Guo et al., 2014). This therefore suggests that instructors should segment videos into short chunks, shorter than six minutes (Guo et al., 2014).
- b) Second, to help minimize student mind wandering during video watching instructors should embed short-recall quiz into the video. For example, the instructors of the Interactive Python Programming and Science of the Solar System MOOCs interrupted the videos with in-video quizzes. Several psychological research studies (Szpunar, Khan, & Schacter, 2013; Szpunar, Jing, & Schacter, 2014) have found that participants who answered short-recall questions after each short video segment (interpolated tests), retained more information at the end of the lecture as compared to the participants who did not receive any recall questions, or the those who merely watched the questions-with-given-answers. Compared to participants in the control group who were not provided interpolated tests, mind wandering occurred less (half as much) for participants who answered the interpolated tests (Szpunar et al., 2013).

- c) Third, lecture videos that featured different instructors' faces (e.g., appearances of guest speakers) were perceived to be more engaging because it helped break the boredom of watching the same instructor throughout all the sessions.

Conclusion

Despite the worldwide attention attributed to MOOCs, scholars still understand little about student engagement in these large open online courses. This study offers a new original contribution by analyzing the reflective comments posted by 5,884 students who participated in one or more of 18 highly rated MOOCs in order to identify the reasons why participants find a MOOC or certain parts of a MOOC engaging. These 18 highly rated MOOCs were chosen from a pool of 282 subject disciplinary areas, having successfully fulfilled the following selection criteria: free-of-charge, rated five-star, and with more than 50 reviews. In this section, we discuss several implications for distance education theory, research, and practice. We conclude by describing the limitations of the present study.

First, the theoretical contribution of this paper lies in its examination of the elements espoused by Engagement Theory, as well as extending our current perspective of Engagement Theory in the context of large-scale fully online courses such as MOOCs that have no requirement for face-to-face attendance. We found that the elements of Creating, and Donating in Engagement Theory (which refer to the application of ideas, and to the use of real-world contexts respectively) (Kearsley & Shneiderman, 1998) were two of the themes found in the MOOCs participants' reflective comments. Many participants of the 18 highly rated MOOCs reported that the use of moderately challenging assignments that require them to apply the concepts or principles learned, instead of merely asking them to recall factual information, helped students learn the subject material better. Participants also reported that the use of content and resources focusing on real-world examples or problems made the course material very relevant. This helped bring tangible meaning to the concepts or principles taught, which sustained students' interest, and enabled them to learn the material more easily because they could see how the principles or theories learned might be applied in real-life.

Contrary to expectation, the Engagement Theory element of Relating, which emphasizes peer interaction (Kearsley & Shneiderman, 1998), was one of the least mentioned themes found in the MOOCs participants' reflective comments. This implies that MOOC students do not seem to attach much importance with respect to the need for peer interaction in large-scale open online courses when compared to traditional online or face-to-face classes. It is likely that the anonymous nature of MOOCs, along with job or family responsibilities diminishes student expectations of course interaction with their peers.

The present findings suggested that MOOC student engagement is promoted when certain instructor attributes are present, namely the instructor's ability to show enthusiasm when talking about the subject material, and the instructor's ability to use humor. These instructor attributes, which formed the most frequently mentioned theme perceived as engaging by MOOC participants, extend our current perspective of Engagement Theory in the context of large-scale fully online courses, which rely primarily on an instructor presenting the subject materials through videos. Although the inclusion of an instructor's face can give a more personal and intimate feel to the video lecture (Kizilcec et al., 2014), how an instructor projects himself or herself (e.g., by showing interest in teaching the material) seems to play a more important role than merely putting a face in the video.

Second, this study contributes to distance education research by proposing and testing five scalable algorithmic models. This is the first work, to our knowledge, that mobilized a repertoire of analytical and technological resources in the fields of machine learning and text data mining to analyze a large dataset of MOOC students' reflective comments. Specifically, we found the Gradient Boosting Tree algorithm (Friedman, 2001), to be the best performing model. The detail technical procedure provided in this study will be of great interest to other researchers who are similarly keen in this type of research methodology.

Third, this study contributes to distance education practice by highlighting several practical solutions to other instructors of large open online courses, as well as those teaching traditional e-learning classes. For example, the various strategies to support instructor-student interactions, and practical tips of using video lectures (as described in the Discussion section) can offer possible solutions for traditional e-learning courses that might otherwise be overlooked.

We conclude the present article by highlighting three limitations. First, it should be noted that highly rated courses may not necessarily be the most effective ones. It is beyond the scope of this study to examine causal effect between course effectiveness (e.g., learning performance) and user ratings. Second, this study did not examine the participants' disaffection of using MOOCs. Exploring student disaffection may offer information that can complement our overall understanding of student engagement. We therefore invite other researchers to conduct this investigation. Third, *CourseTalk* did not provide any indication on which student comments came from students taking the MOOC for course credit or from students who were taking it for other reasons. This precludes an investigation of how the comments from these students may differ. Despite the aforementioned limitations, we believe that the findings of this study provide valuable insight on the specific elements that students find engaging in large-scale open online courses.

Acknowledgment

This research was supported by a grant from the Research Grants Council of Hong Kong (Project reference no: 17651516).

References

- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185. <http://dx.doi.org/10.2307/2685209>
- Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *Proceedings of the 23rd international conference on World Wide Web*, 687-698. <http://dx.doi.org/10.1145/2566486.2568042>
- Ames, C. (1992). Achievement goals, motivational climate, and motivational processes. In G. Roberts (Ed.), *Motivation in Sports and Exercise* (pp. 161-176). Champaign, IL: Human Kinetics Books.
- Baxter, J. A., & Haycock, J. (2014). Roles and student identities in online large course forums: Implications for practice. *The International Review of Research in Open and Distributed Learning*, 15(1), 20-40. <http://dx.doi.org/10.19173/irrodl.v15i1.1593>
- Beldarrain, Y. (2006). Distance education trends: Integrating new technologies to foster student interaction and collaboration. *Distance Education*, 27(2), 139-153. <http://dx.doi.org/10.1080/01587910600789498>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Singapore: Springer.
- Bonk, C. J., & Wisher, R. A. (2000). *Applying collaborative and e-learning tools to military distance learning: A research framework*. (Technical Report #1107), US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical model building and response surfaces*. New York, NY: John Wiley & Sons.
- Business Wire. (2014). Coursetalk and edX collaborate to integrate online course reviews platform. Retrieved from <http://www.businesswire.com/news/home/20140417005360/en#.VCohj2eSx8F>
- Chiu, K. F., & Hew, K. F. (2018). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australasian Journal of Educational Technology*, 34(4), 16-28. <http://dx.doi.org/10.14742/ajet.3240>
- Coetzee D., Fox, A., Hearst, M.A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? In *Proceedings of CSCW 2014*, ACM Press, 1176-1187. <http://dx.doi.org/10.1145/2531602.2531657>
- Coffrin, C., de Barba, P., Corrin, L., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of LAK 2014*, ACM Press (2014), 83-92. <http://dx.doi.org/10.1145/2567574.2567586>

- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. <http://dx.doi.org/10.1177/001316446002000104>
- Conole, G. (2013). MOOCs as disruptive technologies: Strategies for enhancing the learner experience and quality of MOOCs. *RED, Revista de Educación a Distancia*. Número 39. <http://dx.doi.org/10.6018/red/50/2>
- Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine Learning*, 20(3), 273-297. Retrieved from http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf
- Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 215-242. Retrieved from https://www.jstor.org/stable/2983890?seq=1#page_scan_tab_contents
- Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox, and possibility. *Journal of Interactive Media in Education*, 3. <http://dx.doi.org/10.5334/2012-18>
- de Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. *Journal of Computer Assisted Learning*, 32, 218-231. <http://dx.doi.org/10.1111/jcal.12130>
- Dillahunt, T. R., Wang, B. Z., & Teasley, S. (2014). Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education. *International Review of Research in Open and Distance Learning*, 15(5), 177-196. <http://dx.doi.org/10.19173/irrodl.v15i5.1841>
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology*, 72, 218-232. <http://dx.doi.org/10.1037/0022-3514.72.1.218>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence. *Review of Educational Research*, 74, 59-119. <http://dx.doi.org/10.3102/00346543074001059>
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of statistics*, 1189-1232. Retrieved from <https://statweb.stanford.edu/~jhf/ftp/trebst.pdf>
- Gore, H. (2014). Massive Open Online Courses (MOOCs) and their impact on academic library services: Exploring the issues and challenges. *New Review of Academic Librarianship*, 20(1), 4-28. <http://dx.doi.org/10.1080/13614533.2013.851609>
- Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41-50). ACM. <http://dx.doi.org/10.1145/2556325.2566239>

- Hazari, S., North, A., & Moreland, D. (2009). Investigating pedagogical value of wiki technology. *Journal of Information Systems Education*, 20(2), 187-198. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.714&rep=rep1&type=pdf>
- Hew, K. F. (2015). Towards a model of engaging online students: Lessons from MOOCs and four policy documents. *International Journal of Information and Education Technology*, 5(6), 425-431. <http://dx.doi.org/10.7763/IJiet.2015.V5.543>
- Hew, K. F. (2016). Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCs. *British Journal of Educational Technology*, 47(2), 320-341. <http://dx.doi.org/10.1111/bjet.12235>
- Hew, K. F., & Cheung, W. S. (2014). Students' and Instructors' Use of Massive Open Online Courses (MOOCs): Motivations and Challenges. *Educational Research Review*, 12, 45-58. <https://doi.org/10.1016/j.edurev.2014.05.001>
- Kearsley, G., & Shneiderman, B. (1998). Engagement theory: A framework for technology-based teaching and learning. *Educational technology*, 38(5), 20-23. Retrieved from https://www.jstor.org/stable/44428478?seq=1#page_scan_tab_contents
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2-10. <http://dx.doi.org/10.1007/BF02905780>
- Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 170-179). New York, NY: ACM. <http://dx.doi.org/10.1145/2460296.2460330>
- Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning models. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 60-65. Association for Computational Linguistics. Retrieved from <http://www.aclweb.org/anthology/W/W14/W14-41.pdf#page=67>
- Knowles, M. S., Holton, E. F., & Swanson, R. A. (2011). *The adult learner: The definitive classic in adult education and human resource development*. (7th ed.). New York: Elsevier Inc.
- Knowlton, D. S. (2000). A theoretical framework for the online classroom: A defense and delineation of a student-centered pedagogy. *New Directions for Teaching and Learning*, 2000(84), 5-14. <http://dx.doi.org/10.1002/tl.841>
- Kolowich, S. (2013). The professors who make the MOOCs. *Chronicle of Higher Education*, 59(28), A20-A23. Retrieved from <http://publicservicesalliance.org/wp-content/uploads/2013/03/The-Professors-Behind-the-MOOC-Hype-Technology-The-Chronicle-of-Higher-Education.pdf>

- Lombardi, M. M. (2013). The inside story: Campus decision making in the wake of the latest MOOC Tsunami. *MERLOT Journal of Online Learning and Teaching*, 9(2), 239-248. Retrieved from <https://search.proquest.com/docview/1500421747?pq-origsite=gscholar>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276-282. <http://dx.doi.org/10.11613/BM.2012.031>
- Meece, J. L., Blemenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514-523. <http://dx.doi.org/10.1037/0022-0663.80.4.514>
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. Cambridge, MA: MIT press.
- Robertson, S. (2004). Understanding inverse document frequency: On theoretical arguments for IDF. *Journal of Documentation*, 60(5), 503-520. <http://dx.doi.org/10.1108/00220410410560582>
- Rodriguez, C. O. (2012). MOOCs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses. *European Journal of Open, Distance and E-Learning*. Retrieved from <http://www.eurodl.org/index.php?p=archives&year=2013&halfyear=2&article=516>
- Shah, D. (2016, December 25). *By the numbers: MOOCs in 2016*. Retrieved from <https://www.class-central.com/report/mooc-stats-2016/>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22, 63-75. <http://dx.doi.org/10.3233/EFI-2004-22201>
- Sims, R. (2003). Promises of interactivity: Aligning learner perceptions and expectations with strategies for flexible and online learning. *Distance Education*, 24(1), 87-103. <http://dx.doi.org/10.1080/01587910303050>
- Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1), 77-89. [http://dx.doi.org/10.1016/S0034-4257\(97\)00083-7](http://dx.doi.org/10.1016/S0034-4257(97)00083-7)
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16), 6313-6317. <http://dx.doi.org/10.1073/pnas.1221764110>
- Szpunar, K. K., Jing, G. H., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3(3), 161-164. <http://dx.doi.org/10.1016/j.jarmac.2014.02.001>
- Wanzer, M. B. (2002). Use of humor in the classroom: The good, the bad, and the not-so-funny things that teachers say and do. In J. L. Chesebro & J. C. McCroskey (Eds.), *Communication for Teachers* (pp. 116-125). Boston: Allyn & Bacon.

- Warren, J., Rixner, S., Greiner, J. & Wong, S. (2014). Facilitating human interaction in an online programming course, in *Proceedings of SIGCSE 2014*, ACM Press, 665-670.
<http://dx.doi.org/10.1145/2538862.2538893>
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37-52. [http://dx.doi.org/10.1016/0169-7439\(87\)80084-9](http://dx.doi.org/10.1016/0169-7439(87)80084-9)
- Wu, H. C., Luk, R. W. P., Wong, K. F., & Kwok, K. L. (2008). Interpreting tf-idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3), 13.
<http://dx.doi.org/10.1145/1361684.1361686>
- Young, J. R. (2013). What professors can learn from “hard core” MOOC students. *Chronicle of Higher Education*, 59(37), A4. Retrieved from <https://www.chronicle.com/article/What-Professors-Can-Learn-From/139367>

i <http://scikit-learn.org/stable/>

ii <https://spacy.io/>

