Text analysis with LIWC and Coh-Metrix: Portraying MOOCs Instructors

Junyi Li
Central China Normal University
University of Pennsylvania
junyili@mails.ccnu.edu.cn

Lijun Sun
Central China Normal University
lijunsunccnu@gmail.com

Yun Tang
Central China Normal University
tangyun@mail.ccnu.edu.cn

Xiangen Hu
Central China Normal University
University of Memphis
xiangenhu@gmail.com

ABSTRACT

To date, most MOOCs in major platforms (e.g. Coursera and edX) are xMOOCs, which means teacher speech is still the major part of these MOOCs. Therefore, it is necessary to evaluate the quality of lecture and to explore the relationships between lecture quality of MOOCs and learning outcomes. The present study attempted to explore the lecture styles of instructors in MOOCs by using text analysis. One hundred and twenty-nine course transcripts were collected from Coursera and edX. We also collected public data of course evaluation from the largest MOOC community in China (mooc.guokr.com) Linguistic inquiry and word count (LIWC) and Coh-Metrix were used to extract text features including self-reference, tone, affect, cognitive words, and cohesion. After combined students' comments with clustering analysis, results indicated that four different lecture styles emerged from 129 courses: "mediocre", "boring", "perfect" and "enthusiastic". Significant difference was found between four lecture styles for the notes taken, but significant differences were not found for the course satisfaction and discussion posts. Future studies should exam whether different lecture styles have impacts on students' engagement and learning outcomes in MOOCs.

Keywords

MOOCs; Lecture styles; Instructors; Text analysis

1. INTRODUCTION

Massive open online courses (MOOCs) have attracted much attention in the recent years. They provide not only free courses from high prestige universities, but also the freedom of learning for learners all over the world. Major MOOC platforms, such as Coursera, FutureLearn, edX, and Open2Study, are well received by most learners. The reason why MOOCs become a popular way to learn is that it provides each individual learner with opportunities to engage with the materials via formative assessments and the ability to personalize her learning environment (Evans, Baker & Dee, 2016).

Researchers from different discipline have conducted many studies focused on MOOCs learners, including course completion, quality of interaction, student engagement, and collaborative learning in MOOCs (Andres et al., in press; Wang & Baker, 2015). However, the complexities of teaching have been largely absent from emerging MOOC debates (Ross et al., 2014). After all,

MOOC is quite different from traditional class in many aspects. For example, MOOC instructors were motivated by a sense of intrigue, the desire to gain some personal rewards, or a sense of altruism; they were challenged by difficulty in evaluating students' work, encountering a lack of student participation in online forums, being burdened by the heavy demands of time and money, and having a sense of speaking into a "vacuum" due to the absence of student immediate feedback (Hew & Cheung, 2014). Some instructors found it difficult to teach when not facing a real audience of students (Allon, 2012). To date, most MOOCs in major platforms (e.g. Coursera and edX) are xMOOCs, which is a highly structured, content-driven course and designed for large numbers of individuals working mostly alone, teacher speech is still the major part of these MOOCs. Therefore, it is necessary to evaluate the quality of lecture and to explore the relationships between lecture quality of MOOCs and learning outcomes. Some researchers have tried to build models to automatically predict if certain course content would show up by using natural language processing (Araya et al., 2012). Based on the mentioned above, the present study attempted to explore the lecture styles of instructors in MOOCs by using text analysis.

2. METHOD

2.1 Data Collection

Transcripts from 129 courses (humanities: 24.8%, social science: 38%, science: 37.2%) were collected from Coursera and edX. We also collected public data of course evaluation from the largest MOOC community in Mainland China (mooc.guokr.com). This community offered online learners a platform on which they could voluntarily evaluate MOOCs and share their opinions with fellow online learners. The data set we used included course satisfaction, the number of asynchronous discussion posts per course, notes taken per course, the number of followers per course, to name a few.

2.2 Extracting Text Features

Two text analysis tools (i.e. LIWC and Coh-Metrix) were used to extract text features from 129 course transcripts. According to previous studies, self-reference (I, me, my), affect (positive emotion and negative emotion), tone, cognitive words, and cohesion were extracted. Other features like words per sentence and big-words (words are longer than 6 letters) were also viewed as complexity measure of teacher speech.

1

2.3   Data Analysis

Clustering analysis and ANOVA were conducted by using RapidMiner and SPSS. We first transformed all the text features into Z score, then performed k-means algorithm with euclidean distance in RapidMiner. The k value was assigned with a value from 2 to 6, because of comprehensibility.
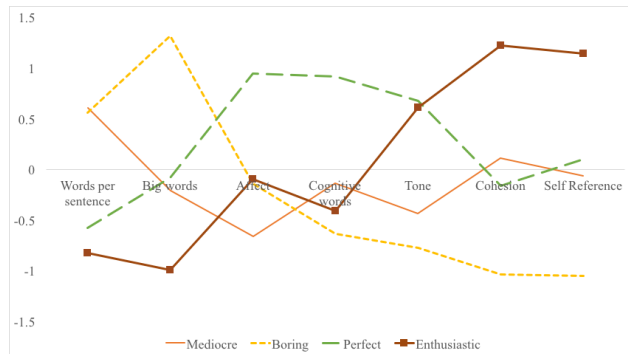

Figure 1 The four lecture styles in MOOCs

3.   RESULTS AND DISCUSSION

Four clusters were found, and there were 42, 27, 36, and 24 courses in each cluster respectively. We then checked the students' comments of these courses in Guoke MOOC community, and assigned label to these clusters (Figure 1).

Concretely, instructors who used the most self-reference words (I, me, my), short sentences, and the least big-words were perceived as agreeable and enthusiastic by students (Cluster 4: Enthusiastic). Instructors who used the least self-reference words, long sentences, the most big-words, and showed a low cohesion were perceived as boring by students (Cluster 2: Boring). Instructors who used the most cognitive words to help students to understand and used medium level of self-reference words, big-words and showed medium cohesion were labeled as "perfect" (Cluster 3). Courses used the most of long sentences and showed average level in other dimensions were labeled as "mediocre" (Cluster 1). No significant differences were found between four lecture styles for the course satisfaction ($F = .76$, $p = .52$, $\eta2 = .02$) and discussion posts ($F = 1.39$, $p = .25$, $\eta2 = .03$). However, significant difference was found for notes taken ($F = 2.80$, $p = .4$, $\eta2 = .06$). Concretely, the number of notes taken in "perfect" style was much more than "mediocre". Notes taken can stand for the cognitive processing of learners to some extents. These results suggested that the "perfect" lecture style may be more likely to encourage students' engagement. Since the discussion posts, notes taken and course satisfaction data in the present study were acquired from a third-party platform, further evidence are needed to verify these results. Future studies should examine whether the four lecture styles have different impacts on students' engagement and learning outcomes (e.g. academic performance and course completion) in MOOCs.

4.   ACKNOWLEDGEMENTS

5.   REFERENCES

[1] Evans, B. J., Baker, R., & Dee, T. S. 2016. Persistence Patterns in Massive Open Online Courses (MOOCs). *Journal of Higher Education,* 87, 206-242.

[2] Andres, M., Baker, R., Gasevic, D., & Spann, G. in press. Replicating 21 Findings on Student Success in Online Learning. *Technology, Instruction, Cognition, and Learning.*

[3] Wang, Y., & Baker, R. 2015. Content or Platform: Why do students complete MOOCs? *Journal of Online Learning and Teaching,* 11, 17-30.

[4] Ross, J., Sinclair, C., Knox, J., Bayne, S., & Macleod, H. (2014). Teacher experiences and academic Identity: The missing components of MOOC pedagogy. *Jounral of Online Learning and Teaching*, 10, 57-69.

[5] 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.

[6] Allon, G. 2012. Operations Management, Udemy. *Chronicle of Higher Education,* 59, B10-11.

[7] Araya, R., Plana, F., Dartnell, P., Soto-Andrade, J., Luci, G., Salinas, E. and Araya, M. 2012. Estimation of teacher practices based on text transcripts of teacher speech using a support vector machine algorithm. *British Journal of Educational Technology*, 43, 837-846.