An Analysis of Learners' Behavior in MOOC Discussion Area

Jia-Mei YU, Jiong GAO and Lu WANG

School of Mathematics, Southwest Jiaotong University, Chengdu, Sichuan, China

642627461@qq.com, 770379028@qq.com, wanglu@home.swjtu.edu.cn

*Corresponding author

Keywords: Discussion area of MOOC, Learner behavior, Number of speaking, Quality of speaking, Technology of the reptile, Technology of word segmentation.

Abstract. In recent years, the thriving of MOOC has provided a good network learning platform for learners. As an online learning platform for communication, the MOOC discussion area also plays an important role. This paper explores the number of learners' speaking in the discussion area by using the web crawler technique, and studies the speech emotion and quality of the speech in the discussion area by using the text data of the learners' speaking of the 39 courses. The empirical results show that the online communication of the MOOC forum is convenient. The numbers of some individual members' speaking made the average level improve. In addition, we also found that the proportion of posts with lower quality of speech was large and there was a significant gap between the average numbers of speaking of different courses.

Introduction

With the rapid development of Internet’s technology, the use of information technology in the field of education has become increasingly widespread. In particular, the emergence of large-scale online courses (Massive Open Online Courses, referred to as MOOC) have impacted the traditional education model. Online education website not only takes full advantage of the large data technology to provide learners with more rich learning resources, but also provide curriculum-related data support for the instructor in the same time, which laid a solid foundation for the sustainable development of online education.

Compared with traditional education, MOOC focuses on learner autonomy and takes "learning to teach" as its teaching philosophy [1]. Among them, as an important part of online teaching activities, the discussion area provides an interactive learning platform for learners and teachers. The degree of activity is the interaction of teaching activities and results. At present, quantitative researches about MOOC mainly focus on analyses of learning behavior, evaluations of teaching model, personalized learning and so on [2], but the research on the discussion area is less. However, some scholars have suggested that there are some shortcomings in the MOOC discussion area, such as inadequate construction, poor degree of interaction, low participation [4], high repetition rate of the theme [3] and other problems. These shortcomings led to some problems, such as the monotony and limitation of the communication in the discussion area and the interactive role of the discussion area was not fully exploited. Therefore, MOOC platform needs to fully stimulate the leading role of teachers and teaching assistants, enhancing the proportion of core members and influence, establish and improve the evaluation mechanism of the forum assessment [4]. The discussion area is based on textual information. The above conclusions are based on statistical data but lack of research on textual data. But only when we make the qualitative change in the discussion area on the basis of information, can really enhance the role of the discussion area.

In view of this, the article makes statistical analysis through the excavation of the discussion area text data. Due to the significant differences in the knowledge reserves and subjective consciousness of the MOOC learners, the article classifies the learners. Make analysis on different categories of
learners, thus obtaining the factors which restrict the efficient of participation in the discussion. Lastly, we make recommendations to maximize the role of the discussion area on this basis.

Compared with previous studies, the article has the following highlights. In the data type, this paper explores the text data of the discussion area of 39 courses and obtains the numerical data in the text information; In addition, with the qualitative change of the situation in the discussion area, the article also concerned about the number and quality of speaking in the discussion area.

Theoretical Review

In 1959, the famous sociologist Katz put forward the "use and satisfaction theory". Unlike the previous view that "the audience can only passively accept media information and be submerged in information", he argues that different people will choose different media to meet their own differences due to individual characteristics and social impact demand. Once the demand is met, people will be more willing to use the media, or people will give up the media. The theory is more focused on the needs of the audience, emphasizing the initiative of the audience, so the theory is used to study the Internet and other emerging media frequently.

The Number of Speaking

The discussion area is the place which helped learners interact and discuss with other learners and teachers. In the MOOC’s model, the forum provides a platform for learners and the curriculum team (teachers, assistants, technical advisors) to communicate with each other. In the discussion area, the different participants play different roles, such as topic sponsors (initiating a topic), question authors (questions raised) and respondents (reply to questions or other comments). Different roles interact and discuss issues in the discussion area, thus increasing the activity of the discussion area.

The topic of MOOC discussion area mainly involves learning, education management, MOOC platform use, daily communication, chatting and other content [5]. Therefore, from the analysis of the topic, we can find the discussion is often based on learning, education management, the use of the MOOC platform and other aspect to publish questions for help. Through the analysis of the operating mechanism of the MOOC forum, it can be found that the course team and the learners can participate in answering questions. For the content of the question, the teacher can judge its quality and choose whether to adopt the reply eventually.

The originators of the conversation are divided into two types: course team and learner. The purpose of a teacher or an assistant initiates a topic is to set up a curriculum-related question based on a teaching plan for discussion by the learner. During the discussion, teachers or assistants will also pay attention to the performance of students and guide them, thereby enhancing the learning effect. After a period of study, learners tend to publish the topic to express the confusion or experience of the knowledge and solve the relevant issues in the discussion. In this process, teachers or assistants will also post and reply to the contents of the evaluation.

In the above different posts, the participants interact with each other, crashing out of intense sparks to attract more people to join.

![Figure 1. The operation mode of the discussion area.](Figure_1.png)
The Quality of Speaking

In the discussion area, the topic posts are often large and the content of which is not the same. Therefore, based on the content of the topic of discussion, we post the post is divided into three categories, the contents are: learning content, curriculum and other content that has nothing to do with learning. For different speaking, the difference in its content will affect its quality and thus the effects of learners' learning outcomes.

In the partition of topic quality, the higher quality mainly includes the discussion of learning doubts, the extension of curriculum content extension and the discussion of course related content. Among them, the topic of the course content of the expansion helped learners a lot. In this type of speech, learners can deepen their understanding of knowledge, develop their ability to learn independently, stimulate their interest in learning and even gradually form a learning group to play a mutual supervisory role. In contrast, the value of the discussion of the curriculum is small. At the same time this statement is prone to repeat the same type of speech, which has a certain negative impact on the participation of the curriculum team.

![Diagram of the Quality of Speaking]

**Figure 2.** The content of the statement and its corresponding quality.

Research Object

This passage selects 39 courses and 1910429 discussion area data of 42521 learners in discussion forums from MOOCs in China. In order to avoid the speech data differences results from the differences of MOOC platform module construction, the passage randomly selects 39 popular courses which have been accomplished and home page recommended to study. Therefore, the research objects have a strong representation.

Research Method

This passage uses web crawler technology. By the use of Depth-First-Search this passage obtains discussion forum pages of 39 courses and analyses the users’ speech data, access to the numbers of Question posts, discussion posts and reply posts of each student in discussion forum of each course.

Because the core of data mining is the word segmentation, the core of the word segmentation is to establish the parametric model and use the model to fit the data. Therefore, this passage establishes the word segmentation based on HMM. For a given number ‘n’ of states, observing not less than ‘n’ representations for the states. For a given representation sequence, when the state transition probability is known, since the number of state transitions is the product of the total number of
migratory points and the probability of migration, we can achieve the result of the state transition site sequence and calculate the frequency of the words after division. And then use the branch and bound to optimize the discrete points of the points to be split and expand the sum of the frequency. The algorithm is used to cut the original text with the optimal cut point to get the word segmentation result. According to the statement made by the user and the result of the word formation, we establish adjacency matrix and use the PCA method to extract keywords.

According to the pertinence of the content of the discussion, will be related to the course content, and curriculum settings related to the curriculum has nothing to do with the keywords marked positive, neutral and negative feelings of the label, constructed 42,521 linear unrelated 42-dimensional vector $P_1$

We divide the keywords into three kinds: keywords related to curriculum content, related to curriculum set and unrelated to the curriculum, and respectively marked as positive, neutral and negative emotional label, then construct 42521 linearly independent 42-dimensional vectors. Among them, $P_i = [P_{1i}, P_{2i}]$ and $P_{1i} = [x_{1i}, x_{2i}, \ldots, x_{39i}]$ said the number of speaking for all 39 courses of each student. $P_{2i} = [x_{40i}, x_{41i}, x_{42i}]$, said each student speaks positive, neutral and negative statements in the total number of speaking, therefore, $x_{40i} + x_{41i} + x_{42i} = 100\%$.

**Result and Analysis**

**Discussion Area Quantity**

According to the results, draw the average number of students in all 39 courses. The histogram is as follows. It can be seen from the figure that the course number of the course number 18 is the highest number of courses, the representative of the course: higher mathematics I, followed by the course number 36 courses, the representative course: College English (spoken). In contrast, the courses with the lowest number of courses are 20, and the courses are: sports dance and culture. There are obvious differences between the average numbers of speaking in each course, which may be influenced by the attributes of the course, but the role played by the instructor cannot be ignored.

![Figure 3. The average number of discussion (as for each course).](image)

**Learner Discussion Times**

The passage statistics the times of the users’ discussion, classifies the number of discussions, and calculates the proportion of each segment. Obviously, the number of users to speak in the 0-10 this stage accounted for the vast majority. Second, the number of statements in the number of 10-20 is relatively more. In contrast, the number of speakers in the 90-100 number of users less.
Figure 4. Percentage of consumers’ discussion frequency.

Emotion of Consumers’ Words

Using the word segmentation technique to analyze the emotion of the consumers, the results showed that 42521 learners had a large proportion of neutral emotion, which was 0.5563, positive emotion was 0.2223, and negative emotion was 0.2214. Specifically, the number of learners with positive emotions was 20764, and the number of negative learners was 20710. Overall, the user speaks emotionally positively and negatively.

Conclusion and Suggestion

As the main exchanging and interacting platform, MOOCs discussion forum has a strong positive role in promoting learners learning process, however, the specific implementation effect of MOOC discussion forum is unsatisfactory. The passage puts forward some concrete and reasonable suggestions through the study of the number of speakers and the quality of the words in the discussion area.

First of all, enhance the function of the forum platform and system construction. To alleviate the low number of speakers in the majority of the status quo, MOOC learning platform can establish reward system and punishment system to encourage users to speak actively, and can be included as standard into the assessment of performance. MOOC courses are diverse, including design engineering, science, art design, life sciences and other aspects, some categories of courses are professional and needed to use professional software to support the discussion of the normal conduct of the discussion. Now, MOOC discussion forum is less functional and it has obstacles for the learner's speech behavior, thus, to a certain extent, it causes a low number of learners take part in the discussion.

Second, instructors should play an active role in the discussion area. The gap among the average number of discussion in each course is large, besides the attributes of the course itself, the discussion area quantity is related to the teaching behavior and active situation of teachers. In the discussion forum communication, the course team should play its leading role, timely answer the questions of learners, and after school to expand the knowledge to expand and improve the learning effect of learners. At the same time, teachers should select high quality reply posts from the discussion posts, judging the performance of students and setting up appropriate incentive systems will encourage learners to continue to participate in the discussion and enhance the enthusiasm of learners in the discussion area.
Finally, learners should change their learning concepts, learning to share their learning experience. Under the atmosphere of rapid growing internet age, rather than only to receive the course information provided by the teacher team, learners should be good at using a variety of learning tools. In the discussion forum, learners can not only learn to solve their own confusion, but also learn from others excellent way of thinking and learning methods and exchange to share their learning experience to seek the common progress.

Acknowledgement
This research was financially supported by the National Science Foundation, Projects in Sichuan Education Development Research Center of Key Research Base for Humanities and Social Sciences of Sichuan Provincial Education Department (Grant NO. CJF14014), Student Research Training Program in Southwest Jiaotong University (Grant NO. 201704008).

References