Investigating the Relationship between Learners' Cognitive Participation and Learning Outcome in Asynchronous Online Discussion Forums

Zhi Liu[1], Shiqi Liu[2], Cuishuang Zhang[2], Zhu Su[1], Tianhui Hu[1] and Sannyuya Liu[1,2]
[1]National Engineering Laboratory for Educational Big Data, Central China Normal University, Luoyu Road 152, 430079 Wuhan, China
[2]National Engineering Research Center for E-Learning, Central China Normal University, Luoyu Road 152, 430079 Wuhan, China

Keywords: Online Discussion, Learning Analytics, LDA Topic Modeling, Regression Analysis.

Abstract: In asynchronous forums of Blended Learning and E-learning, learners' cognitive participation, such as knowledge construction and critical-thinking dialogues, is a crucial factor for their learning outcome, which has not yet been further exploited. This study investigated learners' cognitive behaviors and implicit content derived from posts by using a mixed-method of text mining and statistical analysis. We adopted a content analysis approach to manual coding learners' cognitive behaviors in a Blended Learning discussion forum. Then we proposed an improved topic model called Cognitive Behavior Topic Model (CBTM) to detect learner's semantic content between three achievement groups (High/Medium/Low). Moreover, we performed a statistical analysis to investigate the relationship among cognitive behaviors, cognitive content, and learning outcome. The results showed that the high achievement group's cognitive behavior frequency in all categories is higher than the other, and effective order of behaviors with the learning outcome is "constructive > active > interactive". The "application practice" related topic is more effective for learning outcome than "theoretical discussions". Specifically, when the cognitive content changes from "theoretical discussion" to "application practice", or the number of posts on the same cognitive content-related topic is large, the high-level cognitive behaviors bound to the topic content will increase significantly. Therefore, this study could provide new insights into theoretical and practical implications.

1 INTRODUCTION

Online Asynchronous forums in Blended Learning and E-learning provide learners and educators with an interactive learning environment for cognitive participation such as Q&A and argumentation (Almatrafi & Johri, 2019; Ezen-Can, Boyer, Kellogg, & Booth, 2015). In those discussions, cognitive participation refers to the information processing of cognitive content by cognitive behaviors (Gerrig, 2012). Cognitive high-level participation is a crucial factor that contributes to learning outcome (Chi & Wylie, 2014). In recent years, researchers have studied the effects of conversational cognitive behaviors (Galikyan & Admiraal, 2019; Wang, Wen, & Rosé, 2016; Wang, Yang, Wen, Koedinger, & Rosé, 2015) on E-learning asynchronous forums. However, cognitive behaviors and cognitive content, as two closely related components, have not been used by association modeling to analyze the impact of cognitive participation on learning outcome, especially in blended learning environments where learners are more connected. Moreover, the present cognitive content analysis requires a large amount of manual manipulation (Atapattu, Thilakaratne, Vivian, & Falkner, 2019), lacking automated algorithms to analyze the cognitive participation process, which has a negative impact on large-scale, real-time, automatic cognitive detection and intervention. the large amount of discussion data generated by learners in blended learning asynchronous forums provides the basis for research.

The purpose of this research is to design a Cognitive Behavior Topic Model(CBTM) based text mining algorithm to automatically analyze cognitive behaviors and cognitive content in the asynchronous forum, and further study the relationship between the

Liu, Z., Liu, S., Zhang, C., Su, Z., Hu, T. and Liu, S.
Investigating the Relationship between Learners' Cognitive Participation and Learning Outcome in Asynchronous Online Discussion Forums.
DOI: 10.5220/0009338900260033
In *Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020) - Volume 2*, pages 26-33
ISBN: 978-989-758-417-6

two factors and the learning outcome, to capture higher-order cognition participation in providing insights. It will provide effective algorithms and suggestions for online learning discussions. The organizational structure of this article is as follows: we review related works on discourse cognitive behaviors, content, and behavior-topic models in Section 2. The study design and methodology are described in Section 3. The analysis results and findings of the study are presented in Section 4. Section 5 summarizes and discusses the findings of this study.

2 RELATIONS WORKS

Interactive-Constructive-Active-Passive (ICAP) (Chi & Wylie, 2014) framework presents a guide for evaluating the cognitive engagement of contributions within online communities. Although initially developed to understand classroom conversational data, this framework has also been proved to be effective in online environments to understand learners' cognitive participation, which has been utilized within MOOC studies to measure the association between course materials and discussion contributions (Wang et al., 2015).

As an explicit form of learners' thinking and knowledge processing, interactive discourse is an essential basis for discriminating learners' cognitive patterns, knowledge construction levels, and independent inquiry ability (Wang et al., 2015). SPOC, as a kind of restricted learning community, has produced a large number of cognitive discourse samples related to learning behaviors in its forum. It is a crucial carrier reflecting the knowledge construction, cognitive strategies, and interactive quality of learners. A large number of studies have laid the foundation for the analysis of cognitive discourses in MOOC. For example, Wang et al. (2015) adopted a content analysis approach to analyze learners' cognitively appropriate behaviors in a MOOC discussion forum and further explored the relationship between the quantity and quality of that participation with their learning gains. To interpret what kind of discussion behaviors could help learners from the semantic level. Wang et al. (2016) proposed to trigger the cognitive behaviors of higher-order thinking through ICAP cognitive coding scheme, and found that learners who displayed more higher-order thinking behaviors learned more through more in-depth engagement with course materials posted by their discussion behaviors. Moreover, many researchers investigated the relationship between cognitive behaviors and learning outcomes in terms of quantitative and probabilistic models. Galikyan & Admiraal. (2019) verified the effectiveness of individual learner's cognitive behaviors in predicting learner's learning effect through multiple regression analysis in an asynchronous discussion community and discussed the complex dynamics of knowledge construction in pre-service teacher education. Atapattu et al. (2019) proposed a fusion neural word embedding (Doc2Vec) model to automatically identify teachers' cognitive participation in MOOC communities, such as active participation, constructive participation, etc. They explored the content of constructive cognitive involvement in 67 cases. Recently, some studies did excellent works in modeling cognitive behavior from the perspective of the topic model. For example, Qiu et al. proposed an LDA-based (Blei, Ng, & Jordan, 2003) behavior-topic model, which combined the users' subject interest and behavior patterns. They proved that the model could obtain more dominant behavioral indicators (Qiu, Zhu, & Jiang, 2013). Peng et al. proposed Behavior Emotion Topic Model (BETM) to detect reviews' semantic content between two achievement groups (completers and non-completers), the results showed that posting comments is an essential behavior of the completers, while replying and giving peer-review"praise" arenecessary behaviors of the non-completers (Peng & Xu, 2020). The above researches on behaviors and discussion topics only focus on the semantic modeling of shallow behaviors (such as approval, replying, posting, and other discourse operations) and seldom on the semantic modeling of the content of discourse behaviors from the cognitive level. In addition, the existing research has not explored the relationship between cognitive behavior and learning outcomes in the forum of hybrid courses, nor has it used automated methods to analyze the impact of discourse content themes and cognitive behavior on learning outcomes. Thus, in order to fix the shortage of automatic semantic analysis of discourse cognitive behaviors in the field of learning analytics, we propose the CBTM-based text mining method, which not only models the semantic content with the topic factors from the cognitive level of discourse (compared with the shallow behavior of BETM, CBTM is a discourse behaviors encoded by ICAP framework), but also conducts a specific group dialogue-oriented cognitive behavior analysis method through topic probability modeling.

3 RESEARCH METHODS

3.1 Research Questions

Asynchronous forum discussions can support learners' cognitive participation activities such as knowledge construction and critical-thinking dialogues. To promote learners' high-order cognitive participation in the forum and improve learning effect, it is necessary to understand the topic content corresponding to cognitive behaviors and their relations to learning achievement. This study will focus on the following issues:

(1) What is the relationship between different categories of cognitive behaviors and the learning outcome?

(2) What is the relationship between different types of the cognitive content-related topic and the learning outcome?

(3) What is the relationship between the cognitive content-related topic and cognitive behaviors?

3.2 Dataset

The forum data set is retrieved from the second-year undergraduate course "Introduction to Psychology" in the fall of 2014 offered by a university from China. This course serves as an introductory course in psychology. The first goal is to enable college students to understand the knowledge system of psychology, to master the basic concepts and principles of psychology, and to use the scientific knowledge of psychology to understand and analyze people's psychological phenomena. The second goal is to enrich the knowledge structure of college students, help learners to better understand themselves and self-improvement, and enhance the psychological quality of college students. The third goal is to improve students' teaching ability and quality of daily life. Learners can participate in discussions initiated by instructors in online forums after class. They can also initiate and participate in discussions freely in Chinese. The issues discussed include highly specialized "theoretical discussions" discussions and "application practice" discussions. This course offers ten teaching classes, with a total of 490 learners participating.

Learners in this course are in non-psychological majors (e.g., literature, philosophy, and sports). We assume that their initial level of psychological knowledge is the same, so they are not pre-tested. The total score of this course is 100 points. First, online learning accounts for 15% of the time. Second, the participation in asynchronous forum discussions and electronic assignments accounted for 30%. Third, the counts and quality of face-to-face discussions in offline classrooms accounted for 25%. Fourth, the final exam account for 30%, and the content of the exams is about the understanding and application of basic psychological knowledge. The performance evaluation aims to assess learner participation and knowledge from online and offline, respectively, and can effectively assess learners' learning outcomes.

The average learner score is 84.4 points, the highest score is 97 points, the lowest score is 32 points, and the standard deviation is 6.9. For the difference analysis (Kelley, 1939), we marked the first 29% (n = 144) of the score as the high achievement group, the medium 54% (n = 214) as the medium achievement group, and the last 27% (n = 132) as the low achievement group.

Table 1: Coding rule of Cognitive Behaviors.

Cognitive Behavior	Coding rule	Example
Active	Learner repeats or explain or cite information that is existed in the textbook.	心理学是研究心理和行为的科学。/Psychology is the science of psychology and behavior.
Constructive	Learners ask new questions or express new ideas or compare examples.	我觉得创造力是可以被培养的，但可能也有一定程度是遗传的。/I think creativity can be cultivated, but it may also be inherited to a certain degree.
Interactive	Learner acknowledgment or debate peer' contribution, or build a new idea from it.	同意你的看法，我觉得一个人的性格应该是先天和后天两者结合影响的。/I agree with you, I think a person's personality should be influenced by both innate and acquired.

Learners' postings generated a total of 7,574 discussions about learning content. Their maximum number of posts is 158, the minimum number of posts is 1, the average number of posts per person is 15.0, the standard deviation is 22.3, and the average length of each post is 141.5 words.

3.3 Research Data Analysis

3.3.1 Cognitive Behaviors Coding

The forum discussion reflects the cognitive behaviors and content of the learners. As for the first research question, we studied the cognitive behaviors in the forum and analyzed their frequencies and relationship with learning achievement. we used coding rules that are adjusted from the IACP framework by Wang et al. (2015), which were verified useful for research about the MOOC forum. The simplification coding rules are shown in Table 1. Then, we invited three graduate learners majoring in education technology to learn the coding rules and randomly select 50 posts to conduct coding of cognitive behaviors. The consistency Kappa coefficient reached 0.69. Next, through discussion and clarification on some ambiguous samples, their differences were resolved through the discussion. After the three coders reached a consensus, they simultaneously coded 7574 discussion texts related to the learning content.

3.3.2 Cognitive Behavior Topic Model

Before text modeling, we used HanNLP (hankcs, 2019) for Chinese word segmentation. At this time, named entity words in the field of psychology were added to the user dictionary. Finally, removing stop words, the text is used as the input data for the model.

In the Cognitive Behavior Topic Model (CBTM), the cognitive behavior is considered as a factor that drives a topic. For a large-scale data set, researchers or instructors tend to understand the topics discussed by learners and corresponding cognitive behavior patterns.

As is shown in Figure 1, the modeling process of CBTM is as follows: we assume that when writing a discussion post, each learner will choose a topic of cognitive content (such as human motivation), then choose a cognitive behavior processing pattern (such as construction) for the content. This post is composed of a series of words and semantic content that fits the topic and behavior in a post (e.g., how to enhance children's intrinsic motivation in teaching?). Among them, α, β, γ represents the Dirichlet prior parameters of the author-topic distribution θ , topic-words distribution ϕ , and topic-cognitive behaviors distribution π . According to the conditional dependencies of Bayesian networks, the joint probabilities of topics, words, and cognitive behaviors can be described formally in equation (1).

1. For each topic $t = 1,...,T$
   - Draw $\pi_z \sim Dirichlet(\gamma), \phi_z \sim Dirichlet(\beta)$
2. For each learner $l = 1,...,L$
   - Draw topic distribution $\theta_l \sim Dirchlet(\alpha)$
   - For $l's$ m-th post , $m = 1,...,M_l$
     - Draw a topic from $z_{l,m}$ from θ_l
     - For each word $n = 1,..., L_{l,m}$
       - Draw $w_{l,m,n} \sim \phi_{z_{l,m}}$
     - Draw a posting behaviour $c_{l,m} \sim \pi_{z_l}$

Figure 1: Generative process of CBTM.

$$p(w,z,c \mid \alpha,\beta,\gamma) = p(w \mid z,c,\alpha,\beta,\gamma)\, p(c \mid z,\gamma) p(z \mid \alpha) \quad (1)$$

We used the Gibbs sampling method to estimate the hidden variables in the CBTM. The parameters are estimated as follows in equation (2).

$$\theta_{l,t} = \frac{n_l^t + \alpha}{\sum_{t=1}^{T} n_l^t + T\alpha}, \phi_{t,w} = \frac{n_t^w + \beta}{\sum_{w=1}^{V} n_t^w + V\beta}$$
$$\pi_{t,c} = \frac{n_t^c + \eta}{\sum_{c=1}^{C} n_t^b + C\eta} \quad (2)$$

In order to solve the research question 1, we conducted a multiple regression analysis of cognitive behaviors and the learning outcome (The final score), analyzed the cognitive behaviors differences between different achievement groups, and obtained the influence of each of three cognitive behaviors on the learning outcome. To solve the research questions 2 and 3, we developed a cognitive behavior text mining model to calculate the cognitive content-related topic probability of each learner and then performed the multiple regression analysis and difference analysis to uncover the relationship between these topic probabilities and learning outcome. Futhremore, the cognitive content-related topic differences were investigated in terms of different achievement groups. Finally, we examined the differences between these cognitive-behavior patterns corresponding to content-related topics.

4 RESEARCH RESULTS

4.1 The Relationship of Cognitive Behaviors And Learning Outcome

As shown in Table 2, by the comparison of the mean values and the post-hoc test results between different achievement groups, we can find that the high achievement group (HAG) exhibited significantly higher-frequency constructive, positive, and interactive behaviors than the other two achievement groups. The average times that high-achievement learners took constructive, active, and interactive (16.382, 4.542, 2.979) were significantly larger than those of the medium-achievement group (MAG) (9.051, 2.089, 1.481) and the low-achievement group (LAG) (6.462, 0.985, 1.439), respectively. Looking into the cognitive behavior models of the high and medium achievement group, the comparative relationship of “Construct>Active>Interact” can be revealed. For the low achievement group, the internal cognitive pattern tended to be the “Construct > Interact > Active”.

The regression analysis of the cognitive behaviors and the learning outcome shows that constructive behaviors have a significantly higher regression coefficient to learning outcome than the positive and interactive. The constructive and positive regression coefficients are 0.317 and 0.092, respectively. Interactive has shown a non-significant correlation with the learning outcome, but the post-hoc result is significant.

Table 2: The impact of cognitive behavior on learning achievement.

Cognitive behaviors	Number of cognitive behaviors						F	η^2	Post-hoc test	Regression
	HAG		MAG		LAG					
	Mean	SD	Mean	SD	Mean	SD				
Active n=1231	**4.542**	11.127	2.089	4.104	0.985	2.238	10.422^{**}	0.041	HAG > LAG*** HAG> MAG**	0.092^{*}
Constructive n=5357	**16.382**	21.038	9.051	12.128	6.462	11.411	16.534^{**}	0.064	HAG > LAG*** HAG> MAG ***	0.317^{***}
Interactive n=986	**2.979**	5.943	1.481	3.109	1.439	2.761	7.002^{**}	0.028	HAG > LAG* HAG> MAG **	--

($p<0.001^{***}$, $p<0.01^{**}$, $p<0.05^{*}$)

Table 3: Topics and their keywords for regression analysis of learning outcomes.

Topic label	Regression	Top 10 words with the highest probabilities
T13 Study and work with psychology	0.249^{***}	心理学/ Psychology (0.064), 学习/ Learning (0.018), 生活/Life(0.008), 研究/ Research (0.008), 知识/Knowledge(0.008), 理论/Theory(0.008), 了解/Understanding(0.007),老师/Teacher(0.007), 学生/Learner(0.005)
T11 motivation and emotions	0.129^{**}	动机/Motivation(0.018), 影响/Impact(0.010), 情绪/Emotion(0.010), 压力/Stress(0.010), 努力/Effort(0.009), 失败/Failure(0.009), 意识/Consciousness(0.009), 因素/Factor(0.009), 成功/Success(0.007), 成就/Achievement (0.007)
T28 intelligence and creativity	0.103^{*}	智力/Intelligence(0.060), 能力/Ability(0.022), 创造力/Creativity(0.014), 情绪/Mood(0.011), 情商/Emotional intelligence(0.010), 关系/Relationship(0.009), 心理/psychological(0.007), 流体智力/Fluid intelligence(0.007), 记忆/Memory(0.006), 年龄/Age(0.006)
T17 Negative reinforcement of behaviorism	0.094^{*}	强化/Strengthen(0.060), 学习/Learning(0.022), 刺激/Stimulate(0.011), 条件/Condition(0.010), 反应/Reaction(0.009), 负/Negative(0.007), 作用/Effect(0.007), 心理/Psychological(0.006), 潜意识/Subconscious(0.006)

($p<0.001^{***}$, $p<0.01^{**}$, $p<0.05^{*}$)

4.2 The Relationship of Cognitive Content Topics and Learning Outcome

We used the proposed CBTM model to detect 30 topics and calculate the probability of each learner's engaging about each topic. Then, we conducted the regression analysis on the occurrence probabilities of topics and the learning outcome.

The result is shown in Table 3. The three columns represent a topic number and simplified semantics, regression coefficient, and top 10 highest-probability topic words. In this table, the most significant four topics predicting the learning outcome are ranked by regression coefficients. When the content of the topic words changes from application practice-related discussions to theory-related discussions (from T13 to T11, T28, and T17), the regression coefficients of the topics decrease in order. The content of T13 is related to the use of psychological knowledge for improving the quality of study and work, and this topic has the highest regression coefficient to learning outcome (0.249). T11 involves how to increase motivation and regulate emotions in the pursuit of personal success and achievement. T28 is represented by the thinking or mind-related terms such as "intelligence", "emotional intelligence", "memory" , and "fluid intelligence" . T17 involves behaviorism-related words such as "stimulus", "negative reinforcements", and "subconsciousness", it seems that this topic has the lowest regression coefficient to learning outcome (0.094). In general, the more professional and theoretical their semantic content is, the lower the regression coefficient will be.

As shown in Table 4, to further explore the differences in cognition-related topics between different achievement groups, we conducted a difference analysis of them. The topic probability value (T13=0.053, T11=0.034, T28=0.035, and T17=0.035) and post-hoc test results of the high-achievement group were significantly higher compared with the other achievement groups. Specially, the high achievement group were more likely to express the content of "application practice" than "theoretical discussions".

4.3 Relationship between Cognitive Behaviors and Cognition-related Topics

In order to study the relationship between cognitive behaviors and cognition-related topics, we selected six cognitive topics with unambiguous semantics. The proportions of three cognitive behaviors for each topic were also calculated. They were listed in ascending order from right to left according to the number of posts related to the cognitive content-related topic, as shown in Figure 2.

When the number of posts related to the cognitive content topic increases, the proportion of higher-order cognitive behaviors (interactive and constructive) also increases. T11 is related to "motivation and emotion." Its proportion of higher-order cognitive behavior is 74.8%, and its number of posts is 229. T2 is related to "the personality of children." Its proportion of high-order cognitive behaviors is 94.8%, and its number of posts is 859. For other topics, as the number of posts on the same cognitive content-related topic increases, the proportion of higher-order cognitive behaviors also increases.

As the semantic of topic is trandferred from "theoretical discussion" to "application practice," the proportion of higher-order cognitive behaviors would gradually increase. T2, T13, and T28 are about the application of psychology in teaching. The average proportion of higher-order cognitive behaviors is 93.2%. T20, T17, and T11 are about the terms of psychological theory, and the average proportion of

Table 4: Differences of significant topics between different achievement groups.

Topic label	Cognitive content topic probability						Post-hoc test
	HAG		MAG		LAG		
	Mean	SD	Mean	SD	Mean	SD	
T13	**0.053**	0.035	0.047	0.023	0.036	0.014	HAG > LAG*** HAG> MAG***
T11	**0.034**	0.012	0.032	0.010	0.031	0.006	HAG > LAG*** HAG> MAG***
T28	**0.035**	0.012	0.032	0.008	0.032	0.008	HAG > LAG*** HAG> MAG***
T17	**0.035**	0.011	0.033	0.014	0.031	0.008	HAG > LAG*** HAG> MAG***

($p<0.001^{***}$, $p<0.01^{**}$, $p<0.05^{*}$)

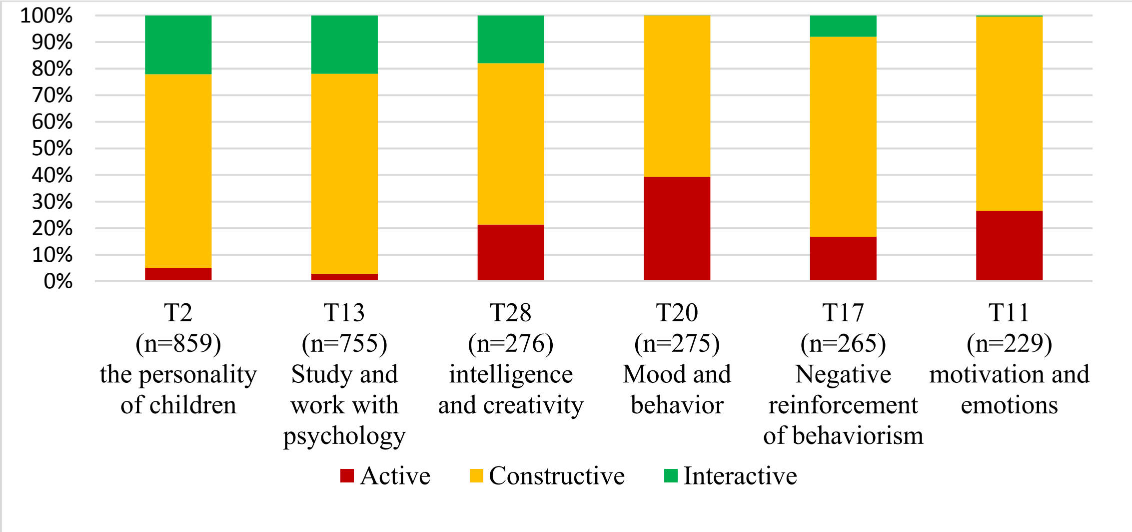

Figure 2: Behavior Model of Cognitive Topic.

higher-order cognitive behaviors reaches 72.1%. The proportion of higher-order cognitive behaviors corresponding to the topic of " application practice " is significantly larger than those corresponding to " theoretical discussion."

5 SUMMARY AND DISCUSSION

In this study, we designed a cognitive behavior topic model to analyze the relationship between cognitive behaviors, cognitive content, and learning outcome in discussion texts of asynchronous learning forums, and reached the following conclusions:

To answer question 1, in a blended learning environment, Constructive behavior have a greater impact on the learning outcome than active and interactive. Learners in the high achievement group took all categories of cognitive behaviors. This conclusion in the Blended Learning environment is consistent with Wang's results in the E-learning environment(Wang et al., 2015) . But it is different from ICAP theory(Chi & Wylie, 2014), i.e., learning outcomes related to interactive behaviors are greater than that related to other cognitive behaviors. Through the results of instructional design and text mining, we can infer that the first reason may be that Wang et al. defined knowledge test scores as learning outcomes, and our learning outcomes were mainly indicated by learning activity participation and knowledge understanding. None of these studies have used critical thinking and interactive skills as teaching goals (they are related to interactive ones). The second reason may be that the cognitive content related to high interactive behaviors involves the basic knowledge points, not the content of the examination. Therefore, educational discourse analysis needs to be combined with educational goals and instructional design, otherwise the causality of the teaching process cannot be efficiently understood. The reason might be that the interactive has fewer times of occurrence, or the related discussed content is only a regular knowledge point.

To answer question 2, The "application practice" topic related has a greater impact on the learning outcome than the "theoretical discussions" topic content. The high achievement group seems to pay more attention to the topic content related to the "application practice" of psychology, and the general topic has the highest regression coefficient to learning outcome, indicating that the appropriate application of knowledge in discussions can effectively promote the learning outcome. One of the goals of this course is to apply psychological knowledge to work and life, and the results of the automatic text mining algorithm are consistent with the teaching goals, so the algorithm can be used as an effective component of a learning management system. On the other hand, by only counting cognitive behavior, it is difficult to reflect the level of a learner's cognitive participation fully.

To answer question 3, When the cognitive content changes from "theoretical discussion" to "application practice," or the number of posts on the same cognitive content-related topic is increased, the high-level cognitive behavior bound to the topic content will increase significantly. Existing studies cannot calculate the cognitive behavior patterns of cognitive content using the original Lda(Ezen-Can et al., 2015; Wang et al., 2016), and the semantic probability of cognitive content cannot be obtained based on the LIWC dictionary method (Moore, Oliver, & Wang, 2019). In this study, cognitive behavior and content are jointly modeled, which can effectively provide teachers with timely and profound dialogue analysis results. Therefore, the content of the instructor-directed discussions needs to be adjusted to adapt to the learners' future work to promote higher-order cognitive participation and learning achievements.

However, there are still some limitations to this study. First, there may be some imbalances between different types of cognitive behaviors within the discussion posts due to specific instructional and interactive design and forum activity settings. Moreover, learning is an information-processing process, and future work also needs to consider the evolution of cognitive behaviors and discussed content in terms of the time dimension.

ACKNOWLEDGMENTS

This work was supported by the Research Funds from National Natural Science Foundation of China (grant number: 61702207, 61937001, 61977030, 31600918), National Key Research & Development Program of China (grant number: 2017YFB1401303),Hubei Provincial Natural Science Foundation of China (grant number: 2018CFB518).

REFERENCES

Almatrafi, O., & Johri, A. (2019). Systematic Review of Discussion Forums in Massive Open Online Courses (MOOCs). *IEEE Transactions on Learning Technologies*, *12*, 413–428.

Atapattu, T., Thilakaratne, M., Vivian, R., & Falkner, K. (2019). Detecting cognitive engagement using word embeddings within an online teacher professional development community. *Computers & Education*, *140*, 103594.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 993–1022.

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*, 219–243.

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 146–150. Poughkeepsie, New York: ACM Press.

Galikyan, I., & Admiraal, W. (2019). Students' engagement in asynchronous online discussion: The relationship between cognitive presence, learner prominence, and academic performance. *The Internet and Higher Education*, *43*, 100692.

Gerrig, R. J. (2012). *Psychology and Life*. Pearson.

hankcs. (2019). *Hankcs/HanLP* [Python]. Retrieved from https://github.com/hankcs/HanLP

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, *30*, 17–24.

Moore, R. L., Oliver, K. M., & Wang, C. (2019). Setting the pace: Examining cognitive processing in MOOC discussion forums with automatic text analysis. *Interactive Learning Environments*, 1–15.

Peng, X., & Xu, Q. (2020). Investigating learners' behaviors and discourse content in MOOC course reviews. *Computers & Education*, *143*, 103673.

Qiu, M., Zhu, F., & Jiang, J. (2013). It Is Not Just What We Say, But How We Say Them: LDA-based Behavior-Topic Model. *Proceedings of the 2013 SIAM International Conference on Data Mining*, 794–802. Society for Industrial and Applied Mathematics.

Wang, X., Wen, M., & Rosé, C. P. (2016). Towards triggering higher-order thinking behaviors in MOOCs. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 398–407. ACM Press.

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating How Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains. *International Educational Data Mining Society*.