Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 8 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i08.48025

PAPER

Danmaku-Based Automatic Analysis of Real-Time Online Learning Engagement

Linzhou Zeng[1], Zhibang Tan[1], Yougang Ke[1], Lingling Xia[2,3](✉)

[1]School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China

[2]Academic Affairs Office, Hunan Institute of Science and Technology, Yueyang, China

[3]Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia

linglingxia@hnu.edu.cn

ABSTRACT

In recent years, there has been a rapid growth of online learning in higher education. Apart from professional online course platforms, many online video sharing websites have also provided online learning opportunities for college students. One of the most popular websites among college students in China is Bilibili, a Shanghai-based Chinese video sharing website known for its danmaku commenting system. This system enables users to post scrolling comments synchronized with the video timeline while the video is playing. Which attracts young students due to the lively user interaction. As a result, an increasing number of Chinese students are utilizing online courses on Bilibili as a supplementary learning resource alongside traditional classroom learning. Despite its popularity, online learning faces the challenge of students' lack of participation more than traditional face-to-face learning does. To understand their learning involvement, we propose a novel danmaku-based automatic analysis model that extracts three dimensions of online learning engagement using the Text Mind software. This model enables us to understand the students' learning patterns both as clusters and as individuals. Based on the model results, we present corresponding intervention strategies for different types of students based on their individual engagement characteristics.

KEYWORDS

online learning, student engagement, danmaku, automatic analysis, k-means++

1 INTRODUCTION

Online education has been promoted worldwide in recent years [1]. Many higher-education institutions began offering online courses to meet students' demands for easy access and high flexibility [2]. Despite its growing popularity, online course learning often suffers from low completion rates [3] and low participation levels [4] because it demands more motivation and self-discipline from students compared to traditional face-to-face education [5]. To ensure learning outcomes in online education, an effective way is to monitor the students' engagement with the learning process during online courses [6].

Linzhou Zeng and Zhibang Tan: These authors contributed equally to this work.

Zeng, L., Tan, Z., Ke, Y., Xia, L. (2024). Danmaku-Based Automatic Analysis of Real-Time Online Learning Engagement. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(8), pp. 127–139. https://doi.org/10.3991/ijim.v18i08.48025

Article submitted 2024-01-07. Revision uploaded 2024-02-24. Final acceptance 2024-02-27.

Student engagement typically refers to the interaction between the time, effort, and other relevant resources invested by both students and their institutions intended to enhance learning outcomes [7]. Some researchers attempt to define student engagement in a more specific way. For example, Bond, Buntins, Bedenlier, Zawacki-Richter, and Kerres [8] define engagement as "the energy and effort that students employ within their learning community, observable via any number of behavioral, cognitive, or affective indicators across a continuum." The term "student engagement," which originated from Astin's concept of "student involvement" [9], has become well-established through annual nationwide surveys in North America [10] and Australasia [11]. It has also garnered significant attention in China [12] [13] [14] [15].

With the emergence of various online learning platforms in China, student engagement in online learning has been of interest to Chinese researchers [16] [17] [18]. Although online learning differs from traditional classroom learning in terms of environments (virtual versus physical), the expectation for students' participation and dedication remains constant. Therefore, some researchers directly adopted the three dimensions of engagement (behavioral, emotional, and cognitive) proposed by Fredricks, Blumenfeld, and Paris [19] in the study of online learning engagement [20]. Others, however, added more dimensions, such as social engagement [21].

Information and communication technology (ICT) has equipped online learning environments with special interactive features. The danmaku commenting system, which originated in Japan, is now one of the favorite interaction tools among college students in China when watching online videos [22] [23]. Unlike traditional scroll-down commenting systems, the danmaku commenting system demonstrates significant real-time characteristics [24] [25]. If a piece of danmaku is posted at a certain point in the video, it will only appear when the video is played to that point again. In this way, every piece of danmaku records the real-time response of the sender to that specific video clip. Utilizing such a system, numerous students in online courses can express their opinions and feelings through a series of time-synced comments (TSCs), creating an environment of extensive interaction [26] [27] [28] [29]. In other words, the danmaku by nature contains valuable information on the students' cognitive levels, emotional experiences, and social tendencies [30]. Therefore, it has great potential for evaluating the students' engagement in those courses based on danmaku-supported videos [31].

In this paper, we propose a novel danmaku-based automatic analysis model of real-time online learning engagement. Compared with traditional questionnaire-based analysis of engagement, our danmaku-based analysis is more automatic as it avoids exhausting manual reviews and is more objective by utilizing unbiased computer programs. More importantly, this model enables us to understand the students' learning patterns both as clusters and as individuals. This understanding then empowers us to tailor appropriate intervention strategies for various situations.

2 METHODS

2.1 Data source

We collect the danmaku data from a widely-used danmaku video platform in China called Bilibili. It had over 300 million monthly active users in the second quarter of 2022 and up to 10 billion pieces of danmaku per year as of 2021. On that platform, the online course we mainly focus on is "*Fundamentals of Digital Electronics*" by Prof. Hong Wang. This course consists of 50 episodes of lectures and 1 supplement episode, for a total of 58143 usable TSCs (as of May 3rd, 2022). We also collect and use the danmaku from other courses as a supplement to the danmaku from *Fundamentals of Digital Electronics*. A web crawler was built to scrape this danmaku data. Each entry of the data contains the content text, the sending timestamp, and the user identification number.

We present an example of video frames with danmaku in Figure 1 to illustrate the *Fundamentals of Digital Electronics*. Four typical pieces of danmaku from different users are selected as examples for social, cognitive, positively emotional, and negatively emotional cases, respectively.

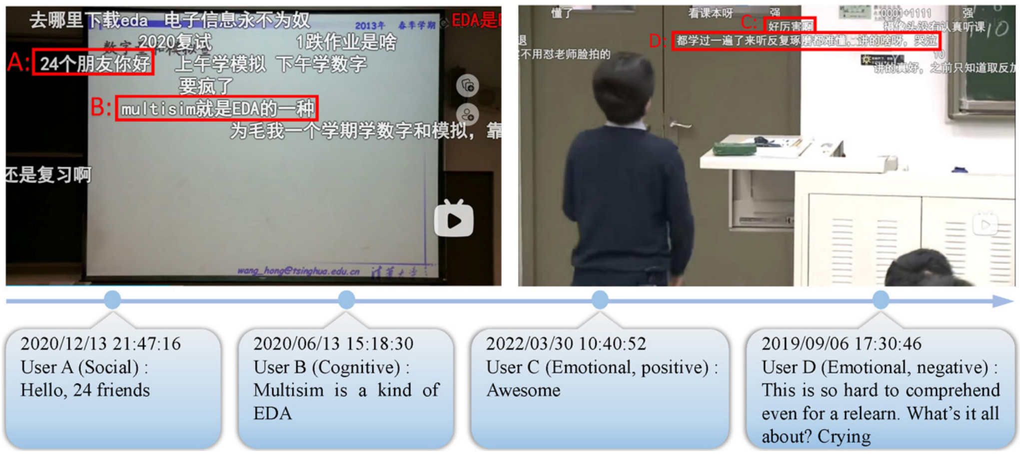

Fig. 1. An example of video frames with danmaku

2.2 Model construction

In this study, we focus on the following dimensions of engagement: cognitive, social, and emotional. Cognitive engagement refers to students' cognitive investment in understanding domain-specific knowledge and focusing on related tasks; social engagement is the degree of students' social interaction with others during the learning process; and emotional engagement reflects students' emotional reactions to their learning experience [32]. We list the details of these dimensions of engagement in Table 1.

Table 1. Details of the dimensions of engagement used in this work

Dimension	Indicator	Meaning	Quantization
Cognitive	CogMech	Cognitive mechanism	The number of cognition-related words
Social	Social	Social interaction	The number of social-interaction-related words
Emotional	PosEmo	Positive emotion	The number of positive-emotion-related words
	NegEmo	Negative emotion	The number of negative-emotion-related words

As mentioned earlier, the danmaku data is real-time and large-scale. In light of these two important characteristics, we propose a danmaku-based automatic analysis model for real-time online learning engagement. The complete diagram of the proposed model is shown in Figure 2. Below, we outline the main steps.

1. Collect and clean the data.
2. Quantify the social, cognitive, and emotional engagement in the danmaku text using the TextMind software [33].

3. Analyze the evolution patterns of the average of each dimension of engagement with episodes and calculate the correlation between any two dimensions.
4. Identify the users who have posted at least 50 pieces of danmaku, and apply the K-means++ clustering algorithm to these users.
5. Project every user into a three-dimensional social-cognitive-emotional space and discover the learning patterns of the outliers.

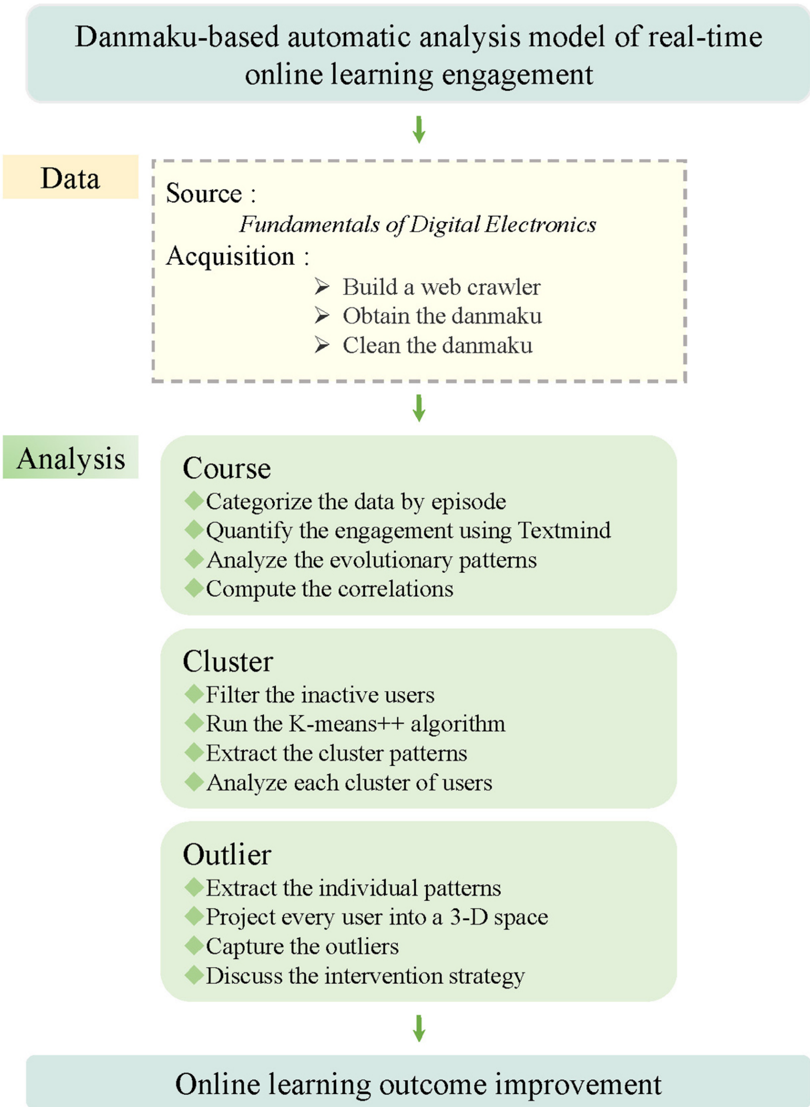

Fig. 2. The model of our danmaku-based analysis

3 RESULTS

3.1 Episode evolution

We calculate the average value of each dimension of engagement per episode and plot these averages against the episode to observe how each dimension of engagement evolves as the course progresses. The results of *Fundamentals of Digital Electronics* are shown in Figure 3, where, for generality, we supplement the results of three more courses: *Principles of Electric Circuits*, *Calculus*, and *Linear Algebra*.

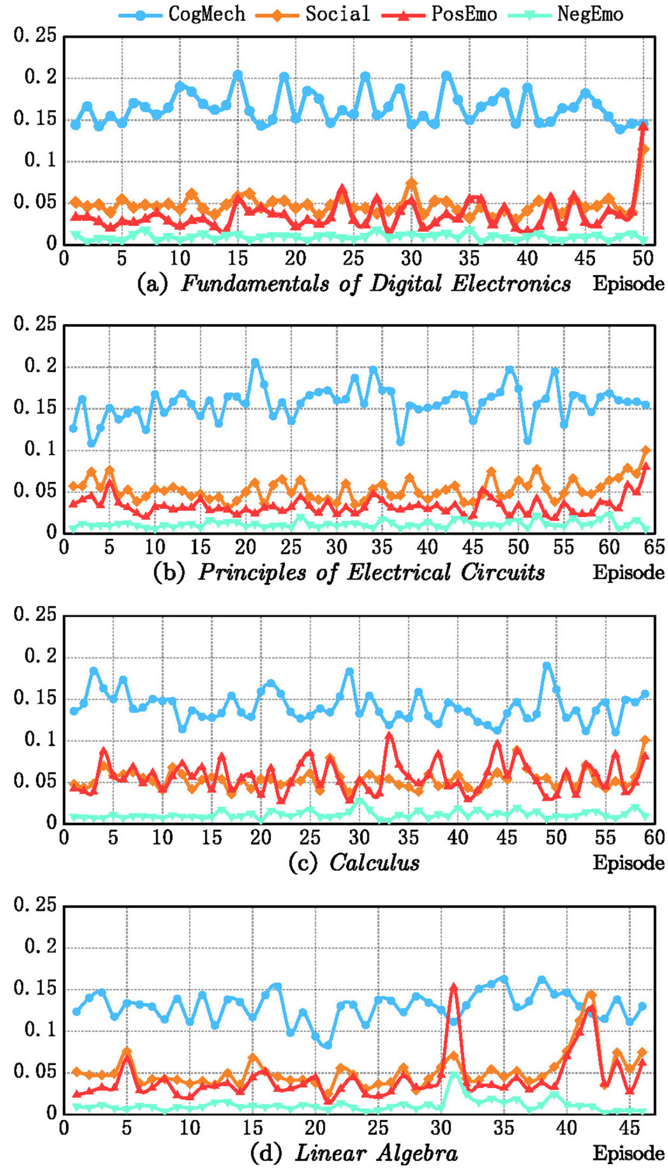

Fig. 3. The evolution of engagements over episodes in four different courses

Taking *Fundamentals of Digital Electronics* as an example, we can see from Figure 3 that the CogMech curve is typically much higher than the others, and the NegEmo curve remains consistently low. This observation indicates that the overall difficulty of this course is moderate. At episodes 7, 27, and 55, the NegEmo curve shows noticeable peaks, while the CogMech curve exhibits severe drops. This suggests that these sections of the course may be challenging for students to comprehend and therefore require careful attention.

Another interesting observation is that there is a remarkable resemblance between the positive motion curve (PosEmo) and the social curve in each course. To analytically demonstrate this similarity, we calculate the Pearson correlation coefficients (PCCs) between dimensions of the engagement. The PCC is a measure of linear correlation between two sets of data and is sometimes referred to as the sample PCC r_{xy} when applied to a sample. For n pairs of paired data, correlation $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, r_{xy} is defined as

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the sample means. This definition of r_{xy} can be rearranged as

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \quad (2)$$

The results of the PCCs are listed in Table 2. It is clear to us that all the correlations between social and PosEmo are over 0.5, some of which are even higher than 0.7 (in *Fundamentals of Digital Electronics* and *Linear Algebra*). These numbers confirm the strong connection between active interaction and positive emotions in online learning.

Table 2. Correlations between engagement level in four different courses

(a) Fundamentals of Digital Electronics				
	CogMech	Social	PosEmo	NegEmo
CogMech	1	–0.16591	–0.23771	0.01143
Social		1	0.753852	–0.1166
PosEmo			1	–0.007842
NegEmo				1
(b) Principles of Electric Circuits				
	CogMech	Social	PosEmo	NegEmo
CogMech	1	–0.088578	–0.070202	0.282772
Social		1	0.68555	0.07913
PosEmo			1	–0.02972
NegEmo				1
(c) Calculus				
	CogMech	Social	PosEmo	NegEmo
CogMech	1	–0.032995	–0.484629	–0.065313
Social		1	0.535068	0.072649
PosEmo			1	0.073273
NegEmo				1
(d) Linear Algebra				
	CogMech	Social	PosEmo	NegEmo
CogMech	1	0.068765	–0.067231	0.117303
Social		1	0.793561	0.141811
PosEmo			1	0.492231
NegEmo				1

3.2 Cluster analysis

We conducted cluster analysis on the danmaku users to divide them into several groups. This allows us to provide tailored suggestions based on each group's learning characteristics. To avoid sparse data, we only consider users who have posted at least 50 pieces of danmaku in *Fundamentals of Digital Electronics*. In total, there are 83 of these users. Every user is considered a set of his or her posted danmaku.

We use the k-means++ algorithm to conduct cluster analysis. It was proposed in [34] as an enhanced version of the widely-used k-means algorithm. With a simple, randomized seeding technique augmentation, the k-means++ algorithm has better accuracy and speed than the conventional k-means algorithm.

One of the major challenges in the practice of k-means++ is determining the optimal number of clusters k. Here, we utilize the well-known elbow method to address this issue. In the elbow method, we calculate the within-cluster sum of squared error (WCSSE) for different values of k, and choose the k for which the value decreases of WCSSE starts to stabilize. If we plot the WCSSE versus k curve, it usually looks like an elbow. By implementing this method on our data, we found the optimal value k to be 4. Subsequentky, we run the k-means++ algorithm to applied the four clusters of danmaku users. The main steps of the k-means++ algorithm is as follows:

1. Choose one center randomly from the data points within the set χ of 83 users.
2. For each data point x not chosen yet, compute $D(x)$, the distance between the point x and the nearest center that has already been chosen, as well as $P(x) = \frac{D^2(x)}{\sum_{x \in \chi} D^2(x)}$ the probability that it x will be chosen as a new center.
3. Choose one new data point at random as the new center, using the probability distribution. $P(x)$.
4. Repeat Steps 2 and 3 until k centers have been chosen.
5. Now that all the k initial centers have been chosen, proceed with the standard k-means clustering algorithm.

In order to understand the characteristics of the obtained clusters, we calculate the average value of every dimension of engagement within each cluster. We present the normalized results in Figure 4, which provides an overall description of each cluster. What's more, we display the most common words in the keyword co-occurrence network in Figure 5. The size of a word in the keyword co-occurrence networks is related to the frequency of that word; the higher the frequency, the larger the size.

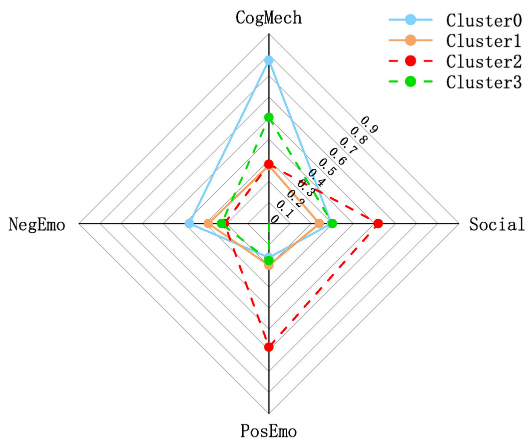

Fig. 4. The average values for engagements of the four clusters

Cluster 0: "hard-workers." This cluster has a significantly higher CogMech value than the other three clusters, as well as a higher NegEmo value. These students feel quite frustrated when they are trying to deive deep into the course materials, as can be seen from Figure 5(a) by their frequently posted words "How," "Why," and "What." On the other hand, they also post a lot of course-related words such as "input," "output," and "trigger" in their danmaku, which shows their efforts in the course study. A representative term in this cluster is "NPEE" (National Postgraduate Entrance Examination in China). It explains their high motivation for learning. We suggest that these students should seek ways to alleviate the academic pressure they are facing.

Cluster 1: "slackers." All the values in this cluster are at low levels. This implies a lack of devotion. Furthermore, we can observe that they seldom pose questions based on the common words in their danmaku, as illustrated in Figure 5(b). Therefore, it is suggested that these students devote a substantial amount of their mind and energy to the course rather than slack off in their studies.

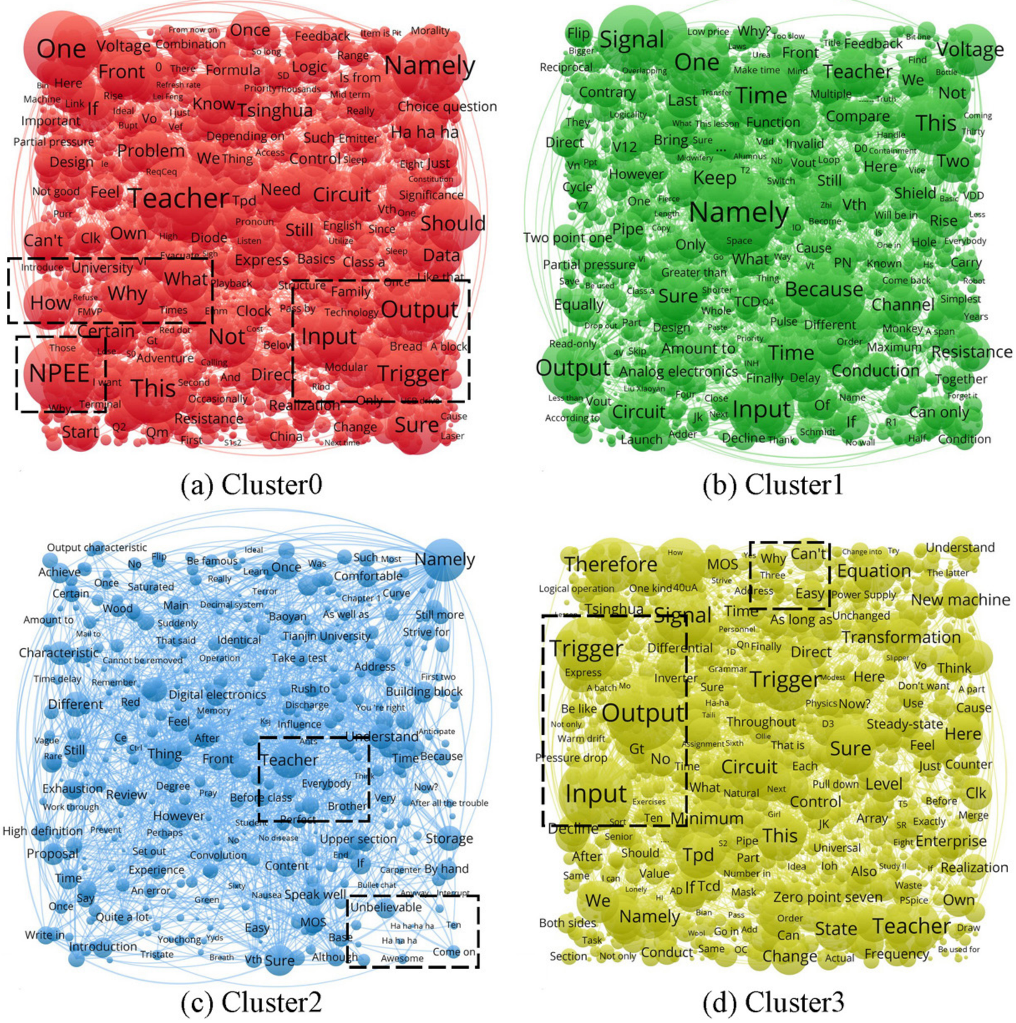

Fig. 5. The co-occurrence networks of the most frequent words posted by the four clusters

Cluster 2: "extroverts." This cluster has significantly higher social and PosEmo values than the others, along with relatively lower cognitive mechanics (CogMech) and negative emotion (NegEmo) values. Based on the most frequent words in these students' danmaku, we notice that they use many vocatives such as "teacher," "everyone," and "bro," and frequently comment on the lectures with words such as "awesome," "unbelievable," and "ha ha ha." The students in this group also enjoy cheering others up by saying, "come on." It is a unique phrase within this cluster, as evident from Figure 5(c). However, few of their frequently posted words are about the course content, indicating that these students prioritize engaging with others over engaging with the course material. Hence, they may need to divert some of their attention from social interaction to the learning process while watching online course videos.

Cluster 3: "talents." This cluster has balanced values across all four aspects. It is quite similar to Cluster 0, according to Figure 5(d), but with a much lower NegEmo value. The most frequent words in the danmaku posted by students within Cluster 3 include course-related terms (e.g., "output," "input," and "trigger") and cognition-related terms (e.g., "Why"). A special term that demonstrates their proficiency in the course of study is "Easy." Some advanced knowledge could be added to in the course enhance their skills.

3.3 Outlier warning

We characterize each user by their engagement values, and they can be represented by a point in a three-dimensional cognitive-social-emotional space. In this area, any outlier user can be easily identified. Note that here we merge the PosEmo value and the NegEmo value into a single value using the formula below.

$$\text{Emotional} = 0.5 + \frac{\text{PosEmo}}{2} - \frac{\text{NegEmo}}{2} \tag{3}$$

Since the PosEmo and both NegEmo range between 0 and 1, the resulting value above also falls between 0 and 1, which we utilize for the social dimension. The closer the result is, the more positive the emotion is; the closer it is to 0, the more negative the emotion. Besides, we use the CogMech value for the cognitive dimension and the social value for the social dimension. The visualization in Figure 6. shows the 83 users who have posted at least 50 pieces of danmaku in the cognitive-social-emotional space, with an example of an outlier highlighted.

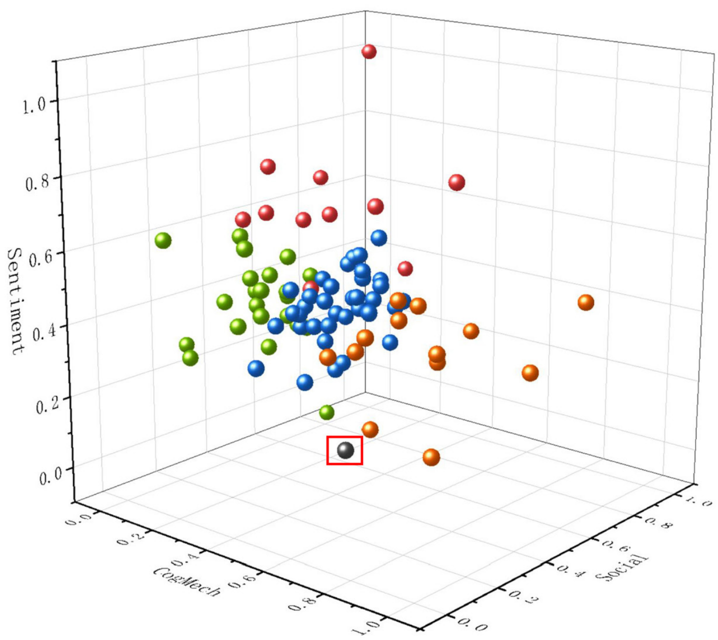

Fig. 6. The positioning of users in a three-dimensional space reveals a distinct outlier user

The outlier points with a frame in Figure 6 correspond to the user with ID number "d182****". For this user, the cognitive-social-emotional coordinate is (0.671748558, 0.128088769, 0.17221438). Specifically, the original PosEmo and NegEmo values are 0.09366473 and 0.74923597, respectively. From these values, we can infer that this user studies diligently (high CogMech) but experiences feelings of exhaustion (high NegEmo). To validate this, we further examine the contents of the danmaku by the user. We observe many texts that express puzzlement and struggle. Below are some examples.

> "Tomorrow I will forget everything. I learned just now, let alone solving the exercises." So disappointed.
>
> "Trying to understand, but the waveform in my mind corrupts every time I revisit this part."
>
> "Not difficult?" Serious"
>
> "Can't do it at all."

All the observations indicate that the outlier user is struggling with the course-work, despite putting in a significant amount of effort. In response to that, both instructional and emotional support should be offered to help this student meet the course requirements.

4 CONCLUSIONS

In this paper, we have introduced a novel danmaku-based automated analysis model for real-time online learning engagement and applied this model to the danmaku from the "*Fundamentals of Digital Electronics" course* and several others. The results provide valuable insights into the learning engagement of students in online courses from three different perspectives: episode evolution, cluster analysis, and outlier warning. We have discovered (1) a high correlation between social engagement and positive emotions, (2) distinct clusters among the course learners, and (3) the negative mood of an outlier user. Some suggestions are provided to learners based on their individual characteristics for online learning.

5 ACKNOWLEDGEMENTS

The authors thank the TextMind team for developing such an easy-to-use Chinese language psychological analysis system.

6 FUNDING

This work was supported by the University-Industry Collaborative Education Program of the Ministry of Education of China (No. 202101138005); the Teaching Reform Research Program of the Hunan Institute of Science and Technology (Nos. 2023B023 and 2021B28); and the Undergraduate Innovation and Entrepreneurship Training Program of Hunan Province (S202310543042).

7 REFERENCES

[1] S. Palvia, P. Aeron, P. Gupta, D. Mahapatra, R. Parida, R. Rosner, and S. Sindhi, "Online education: Worldwide status, challenges, trends, and implications," *Journal of Global Information Technology Management*, vol. 21, no. 4, pp. 233–241, 2018. https://doi.org/10.1080/1097198X.2018.1542262

[2] A. Sun and X. Chen, "Online education and its effective practice: A research review," *Journal of Information Technology Education*, vol. 15, pp. 157–190, 2016. https://doi.org/10.28945/3502

[3] Y. Liu, M. Zhang, D. Qi, and Y. Zhang, "Understanding the role of learner engagement in determining moocs satisfaction: A self-determination theory perspective," *Interactive Learning Environments*, vol. 31, no. 9, pp. 6084–6098, 2023. https://doi.org/10.1080/10494820.2022.2028853

[4] W. You, "Research on the relationship between learning engagement and learning completion of online learning students," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 1, pp. 102–117, 2022. https://doi.org/10.3991/ijet.v17i01.28545

[5] V. Arkorful and N. Abaidoo, "The role of e-learning, advantages and disadvantages of its adoption in higher education," *International Journal of Instructional Technology and Distance Learning*, vol. 12, no. 1, pp. 29–4, 2015.

[6] J. W. Ko, S. Park, H. S. Yu, S.-J. Kim, and D. M. Kim, "The structural relationship between student engagement and learning outcomes in Korea," *The Asia-Pacific Education Researcher*, vol. 25, no. 1, pp. 147–157, 2016. https://doi.org/10.1007/s40299-015-0245-2

[7] V. Trowler, "Student engagement literature review," *The Higher Education Academy*, vol. 11, no. 1, pp. 1–15, 2010.

[8] M. Bond, K. Buntins, S. Bedenlier, O. Zawacki-Richter, and M. Kerres, "Mapping research in student engagement and educational technology in higher education: A systematic evidence map," *International Journal of Educational Technology in Higher Education*, vol. 17, no. 1, pp. 1–30, 2020. https://doi.org/10.1186/s41239-019-0176-8

[9] A. W. Astin, "Student involvement: A developmental theory for higher education," *Journal of College Student Personnel*, vol. 25, no. 4, pp. 297–308, 1984.

[10] G. D. Kuh, "Assessing what really matters to student learning inside the national survey of student engagement," *Change: The Magazine of Higher Learning*, vol. 33, no. 3, pp. 10–17, 2001. https://doi.org/10.1080/00091380109601795

[11] H. Coates, "Development of the Australasian survey of student engagement (AUSSE)," *Higher Education*, vol. 60, no. 1, pp. 1–17, 2010. https://doi.org/10.1007/s10734-009-9281-2

[12] H. Ross, Y. Cen, and Z. Zhou, "Assessing student engagement in china: Responding to local and global discourse on raising educational quality," *Current Issues in Comparative Education*, vol. 14, no. 1, pp. 24–37, 2011. https://doi.org/10.52214/cice.v14i1.11458

[13] H. Yin, "What motivates Chinese undergraduates to engage in learning? Insights from a psychological approach to student engagement research," *Higher Education*, vol. 76, no. 5, pp. 827–847, 2018. https://doi.org/10.1007/s10734-018-0239-0

[14] H. Yin and W. Wang, "Undergraduate students' motivation and engagement in china: An exploratory study," *Assessment & Evaluation in Higher Education*, vol. 41, no. 4, pp. 601–621, 2016. https://doi.org/10.1080/02602938.2015.1037240

[15] Z. Zhang, W. Hu, and O. McNamara, "Undergraduate student engagement at a Chinese university: A case study," *Educational Assessment, Evaluation and Accountability*, vol. 27, no. 2, pp. 105–127, 2015. https://doi.org/10.1007/s11092-015-9213-x

[16] M. Hu and H. Li, "Student engagement in online learning: A review," in *International Symposium on Educational Technology (ISET)*, 2017, pp. 39–43. https://doi.org/10.1109/ISET.2017.17

[17] Y. Wang, Y. Cao, S. Gong, Z. Wang, N. Li, and L. Ai, "Interaction and learning engagement in online learning: The mediating roles of online learning self-efficacy and academic emotions," *Learning and Individual Differences*, vol. 94, p. 102128, 2022. https://doi.org/10.1016/j.lindif.2022.102128

[18] X. Yi, Z. Zhan, and S. Ma, "An evaluation model of mooc learning engagement based on learning behaviors and social network analysis," in *IEEE 2nd International Conference on Educational Technology (ICET)*, 2022, pp. 165–169. https://doi.org/10.1109/ICET55642.2022.9944529

[19] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School engagement: Potential of the concept, state of the evidence," *Review of Educational Research*, vol. 74, no. 1, pp. 59–109, 2004. https://doi.org/10.3102/00346543074001059

[20] M. Hu, H. Li, W. Deng, and H. Guan, "Student engagement: One of the necessary conditions for online learning," in *International Conference on Educational Innovation Through Technology (EITT)*, 2016, pp. 122–126. https://doi.org/10.1109/EITT.2016.31

[21] L. Luan, J.-C. Hong, M. Cao, Y. Dong, and X. Hou, "Exploring the role of online efl learners' perceived social support in their learning engagement: A structural equation model," *Interactive Learning Environments*, vol. 1–12, 2020. https://doi.org/10.1080/10494820.2020.1855211

[22] Y. Bo, "Online learning interaction in danmaku videos: A pilot study," in *Edmedia+Innovate Learning*, 2020, pp. 1262–1270.

[23] X. Chen and S. Feng, "Exploring the relationships between social and teaching presence in video-based informal learning using network analysis," in *Proceedings of the Ninth Acm Conference on Learning@ Scale*, 2022, pp. 340–344. https://doi.org/10.1145/3491140.3528316

[24] T. Peng and T. Wang, "The effect of using the danmaku mechanism in video learning: A pilot study," in *Edulearn 21 Proceedings*, 2021, pp. 2654–2654. https://doi.org/10.21125/edulearn.2021.0579

[25] D. Xi, W. Xu, R. Chen, Y. Zhou, and Z. Yang, "Sending or not? A multimodal framework for danmaku comment prediction," *Information Processing & Management*, vol. 58, no. 6, p. 102687, 2021. https://doi.org/10.1016/j.ipm.2021.102687

[26] Y. Bo, "Learning alone yet together: An exploratory study of the interaction through online danmaku videos," in *IEEE Learning with Moocs (lWMOOCS)*, 2020, pp. 100–104.

[27] C. He, L. He, T. Lu, and B. Li, "Beyond entertainment: Unpacking danmaku and comments' role of information sharing and sentiment expression in online crisis videos," in *Proceedings of the ACM on Human-Computer Interaction (CSCW2)*, 2021, pp. 1–27. https://doi.org/10.1145/3479555

[28] S. Li, H. Zhu, Y. Qian, S. Ren, and B. Fang, "Classification and quantification of danmaku interactions in online video lectures: An exploratory study," *Wireless Communications and Mobile Computing*, vol. 2022, 2022. https://doi.org/10.1155/2022/5656669

[29] B. Yang and H. He, "Learning alone yet together: Enhancing between-learner social connectivity at scale," in *Proceedings of the Ninth Acm Conference on Learning@ Scale*, 2022, pp. 374–378. https://doi.org/10.1145/3491140.3528294

[30] Y. Zhang, A. Qian, Z. Pi, and J. Yang, "Danmaku related to video content facilitates learning," *Journal of Educational Technology Systems*, vol. 47, no. 3, pp. 359–372, 2019. https://doi.org/10.1177/0047239518811933

[31] X. Lin, M. Huang, and L. Cordie, "An exploratory study: Using danmaku in online video-based lectures," *Educational Media International*, vol. 55, no. 3, pp. 273–286, 2018. https://doi.org/10.1080/09523987.2018.1512447

[32] G. Zhu, P. Raman, W. Xing, and J. Slotta, "Curriculum design for social, cognitive and emotional engagement in knowledge building," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, pp. 1–19, 2021. https://doi.org/10.1186/s41239-021-00276-9

[33] R. Gao, B. Hao, H. Li, Y. Gao, and T. Zhu, "Developing simplified chinese psychological linguistic analysis dictionary for microblog," in *International Conference on Brain and Health Informatics*, 2013, pp. 359–368. https://doi.org/10.1007/978-3-319-02753-1_36

[34] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual Acm-Siam Symposium on Discrete Algorithms*, 2006, pp. 1027–1035.

8 AUTHORS

Linzhou Zeng received the B.S. degree from Peking University in 2017, and the M.S. degree from the University of California, Los Angeles in 2018. He has been with the Hunan Institute of Science and Technology since 2019. His research interests lies in data mining applications (ORCID: 0000-0003-2344-8927).

Zhibang Tan is an undergraduate student majoring in Electronic and Information Engineering at the Hunan Institute of Science and Technology. His research interest is educational data mining (ORCID: 0009-0001-7864-3845).

Yougang Ke received the Ph.D degree from Hunan University in 2017 where he then worked as a postdoctoral fellow from 2018 to 2020. He is currently an Associate Professor with the Hunan Institute of Science and Technology. His research interests include text mining and learning analytics (ORCID: 0000-0002-9903-7078).

Lingling Xia received the Master's degree in Educational Technology from Hunan University in 2020. She is now with the Academic Affairs Office of the Hunan Institute of Science and Technology, and also a Ph.D candidate in Educational Technology at the University of Malaya. Her research interest lies in education informatization (E-mail: linglingxia@hnu.edu.cn; ORCID: 0009-0004-4862-6915).