Behavior Analytics, Sentiment Analysis, and Topic Detection of Danmaku from Online Electronics Courses on Bilibili

Linzhou Zeng, Zhibang Tan, Lingling Xia, Yu'an Xiang, and Yougang Ke*

***Abstract*—Danmaku data from an online course contains implicit information about the students, the teacher, and the course itself. To discover the information, we design a behavior-sentiment-topic mining procedure, and apply it on the danmaku from two electronics courses on Bilibili, a popular video sharing platform in China. The procedure enables us to obtain behavior patterns, text sentiments, and hidden topics, of those danmaku comments effectively. Results show similarities and differences between the danmaku from *Fundamentals of Analog Electronics* and that from *Fundamentals of Digital Electronics*. Some interesting observations are given according to the results. For example, students tend to experience an emotional upsurge right before the end of a course, which is due to their fulfilment for completing the course. Based on the observations, we make some suggestions for students, teachers, and platforms on how to improve the learning outcomes using the results of danmaku analysis.**

***Index Terms*—E-learning, danmaku, time-sync comments, educational data mining, learning analytics.**

I. Introduction

E-learning, a.k.a online learning, has become popular among college students in recent years. Apart from massive open online course (MOOC) providers, video sharing platforms such as YouTube, have launched online video courses uploaded by their users. Free of charge and easy to playback, these online courses are favored by some learners. College students in China often use a video sharing platform named Bilibili for entertainment and study, which attracts young people with its danmaku system [1]. Danmaku, originating in Japan, is a series of time-sync comments (TSCs) that scroll on the top of a video stream. It serves as an interaction between the users through time and space, and also a channel for the users to express their opinions and attitudes [2]. Compared with traditional forum-type comments, danmaku is integrated into the videos and highlighted in various colors, which attracts much more of the viewers' attention and participation [3]. With the huge amounts, danmaku is a mine of information about both the videos and the viewers. Recently, researchers have contributed to the field of TSC analysis, such as studying user behaviors [4] and exploring social interactions [5]. Some researchers have seen the potential of danmaku in online learning. Yao *et al.* utilized the information about the use of danmaku to improve the users' participation and interaction in online video learning [6]. Leng *et al.* investigated the impact of danmaku on learning outcomes based on the eye gaze data [7]. However, to the best of our knowledge, the TSC analysis on specific courses in electrical and electronics engineering (EEE) is still seldom seen in literature. Mining the TSCs from online courses benefits all parties involved in online learning who share a common goal of improving the quality of teaching and learning. For students, the behavior patterns hidden in the data let us know how to help them allocate their time more properly. For teachers, the sentiments conveyed by the TSCs, whether positive, neutral, or negative, are an important feedback on the lectures. And for platforms, the topics of the TSCs show what the viewers mainly care about with respect to online course videos.

In this paper, we develop a behavior-sentiment-topic mining procedure to analyze the TSCs from two EEE core courses, analog electronics and digital electronics, on Bilibili. Results reveal the students' learning regularities, sentimental variations, and topical interests. We summarize and investigate some interesting observations to draw useful conclusions. This paper fills the gap between the TSC analysis and the EEE course videos, giving some guidance on the development of online learning in EEE education. The remainder of the paper is organized as follows. Section II specifies the mining procedure shown in Fig. 1. Section III presents the acquisition and prepossessing of TSC data, and Section IV discusses the analysis and mining of the data. Conclusions are drawn in Section V.

II. Mining Procedure

A. Behavior Analytics

Behavior analytics is a statistical method to analyze how people behave or act based on user data. It is a simple but useful way to extract some behavior patterns. As for danmaku video viewers, sending a TSC is an important action of them while watching a video. Temporal patterns of their watching behaviors can be inferred from the relation between the number of TSCs and time, i.e., how many TSCs there are in different time intervals. Depending on the definition of the time axis, the meaning of that relation is different. For example, in [2], the time axis was defined pertaining to the running time of a video, and the relation was about how the number of comments changes as the video proceeds. In the behavior part of our mining procedure, we focus on the hourly variation of the number of TSCs in a day, so the time axis we use is the 24 hours. That variation is closely associated with the daily patterns of students learning behaviors. This is because a TSC is typically sent by a user while he/she is watching the video. Hence, we assume the

Manuscript received August 8, 2022; revised September 4, 2022; accepted September 21, 2022.

Linzhou Zeng, Zhibang Tan, Yu'an Xiang, and Yougang Ke are with the School of Information of Science and Engineering, Hunan Institute of Science and Technology, Yueyang, 414006, China.

Lingling Xia is with the Faculty of Education, University of Malaya, 50603, Kuala Lumpur, Malaysia.

*Correspondence: yougangke@hnu.edu.cn

doi: 10.18178/ijiet.2023.13.2.1800

distribution of the TSCs' sending time is the same as that of the users' watching time (namely the students' learning time). To obtain the distribution, we format the timestamps and then count the number of TSCs per hour, regardless of the date.

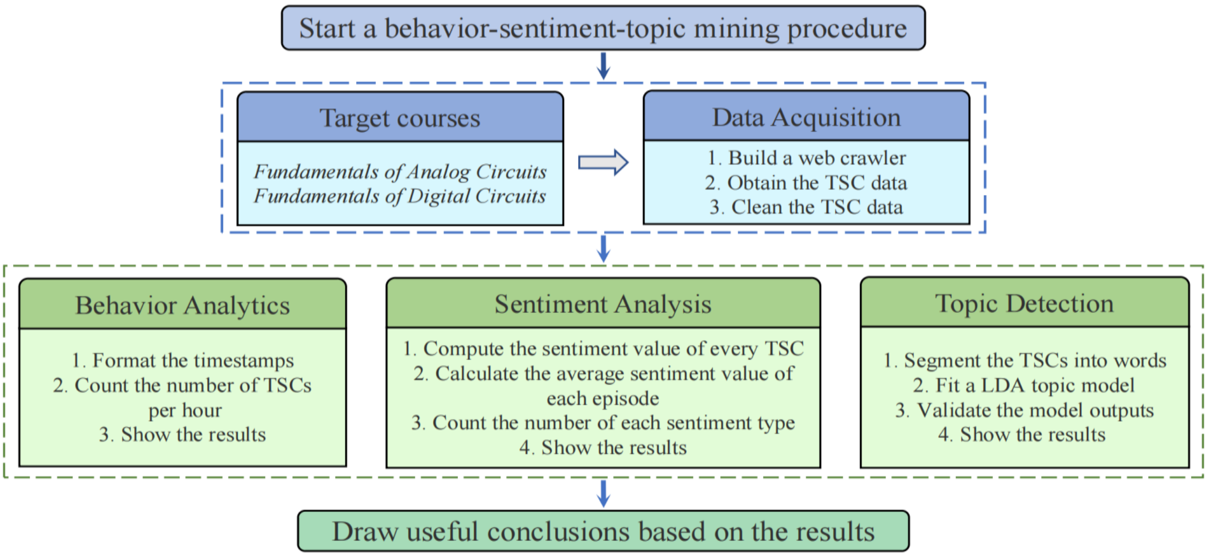


Fig. 1. A behavior-sentiment-topic mining procedure.

B. *Sentiment Analysis*

Sentiment analysis is a computational process of texts to categorize the perceived writer's attitude as positive, negative, or neutral [8]. Current methods for sentiment analysis can be broadly divided into two categories, sentiment lexicon (or dictionary) based methods and machine learning based methods [9]. The design of the sentiment lexicon is critical to the accuracy of this kind of methods. In contrast, the machine learning based methods regard sentiment analysis as a supervised learning problem. The texts are manually labeled and then used to train a sentiment classifier. Some researchers have conducted sentiment analysis on danmaku. Li *et al.* designed a sentiment dictionary and used it with Naive Bayes (NB) for sentiment analysis of danmaku [10]. Wang *et al.* proposed an improved Bi-LSTM based model to identify the emotions of danmaku messages [11]. Besides, a framework of entity-level sentiment analysis on danmaku video comments was proposed in [12].

In the sentiment part of our mining procedure, we use a Python library called SnowNLP to process the TSC texts in Chinese for sentiment analysis [13]. It provides a Naive Bayes algorithm that computes the sentiment value of each TSC text, ranging from 0 to 1. The closer to 1 the value is, the more positive the sentiment is; the closer to 0 the value is, the more negative the sentiment is. The mapping from the sentiment value to the sentiment type is defined as follows. If 0<Value<=0.33, then Type is negative; if 0.33<Value<=0.66, then Type is neutral; if 0.66<Value<1, then Type is positive. After mapping the sentiment value of every TSC text to one of the three types, we count the number of each type and calculate its proportion to the total number of TSCs. This gives us a summary of the sentiments. Then, we calculate the average sentiment value of each episode in the two courses. The averages are plotted in a chart, which shows what kind of impressions the course gives to students as it goes on.

C. *Topic Detection*

Topic detection is an automatic technique to discover the topics in a large collection of text data [14]. A topic contains a cluster of words that frequently occur together [15]. Topic modeling aims at connecting words with similar meanings and distinguishing between uses of a word that has multiple meanings [15]. Many probabilistic topic models have been proposed, such as vector space model (VSM), latent semantic indexing (LSI), probabilistic latent semantic analysis (PLSA), and latent Dirichlet allocation (LDA) [16]. Among them, LDA is one of the most popular models [17]. It is a generative probabilistic model for collections of discrete data such as text corpora [18], and can be used to discover a user-specified number of topics shared by documents within a text corpus. A few studies have explored the topic modeling on danmaku data. Wu *et al.* proposed a temporal and personalized topic model to extract the video tags from the danmaku comments [19]. Bai *et al.* introduced a joint online non-negative matrix factorization model to detect latent topics with automatically exploiting danmaku comments [20]. A question that remains unanswered is whether the LDA performs well on danmaku. Previous studies have shown its limitations on short texts like microblogs [21]. Considering this question, we conduct some other analysis on danmaku, including the term frequency analysis and the keywords co-occurrence analysis, to see if the results agree well.

In the topic part of our mining procedure, we use a Python library named pyLDAvis to detect the topics and visualize the results. As the name suggests, it is a Python implementation of LDAvis, an interactive LDA topic model visualization tool by Sievert and Shirley [22]. This tool allows users to manually adjust relevance metric by changing the value of a parameter λ to find the optimum set of terms under a particular topic. When λ is set close to 0, the most exclusive terms under a topic will be considered more relevant to that topic; when λ is set close to 1, the most frequent terms under a topic will be considered more relevant to that topic. The most relevant terms for a topic will be displayed by LDAvis. These outputs, for validation, are compared with other results, that is, the term frequency and the keywords co-occurrence.

III. Data Acquisition and Preprocessing

We collect the TSCs on Bilibili as raw data. The reasons why we choose Bilibili as data source are as follows. For one thing, it is a leading danmaku video website in China with over 200 million daily active users. For another, it is a rich supply of TSCs, the total number of which has been 10 billion up to Nov. 29th, 2021.

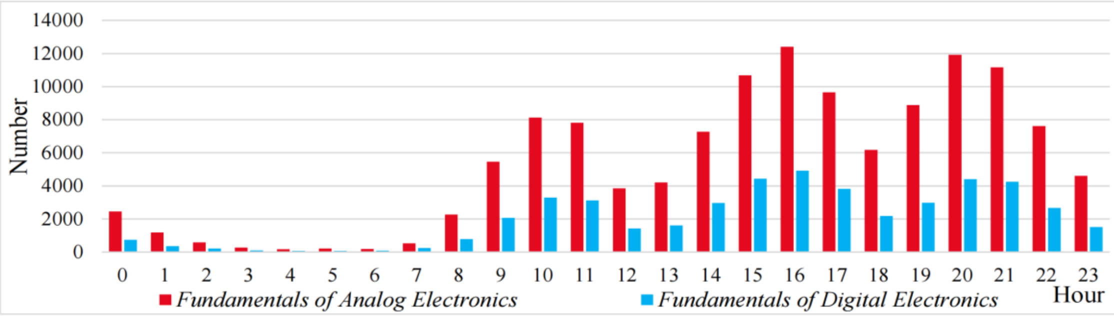


Fig. 2. Distribution of the number of TSCs by hour in a day.

The online courses we look into are *Fundamentals of Analog Electronics* [1] and *Fundamentals of Digital Electronics*[2]. These two courses on Bilibili have about 4 million and 2 million views in total, providing nearly 130 thousand and 50 thousand TSCs, respectively (ending Feb. 23rd, 2022). We build a web crawler in Python to scrape the TSC data of those courses on Bilibili. The data are saved as a comma-separated values (CSV) file. Missing data, irrelevant data, and abnormal data are then removed manually. The numbers of usable TSCs are 127785 and 46256, respectively.

The numbers of Analog and Digital Electronics viewers (that have sent at least one TSC) are 76850 and 27587, respectively. The number of viewers that have watched both is 1509, which is 2% of the Analog Electronics viewers and 5.5% of the Digital Electronics viewers.

IV. Data Analysis and Mining

A. Behavior Analytics

We transform the timestamps in the TSC data into standard format "yyyy-MM-dd'T'HH:mm:ss". Then, we count the number of TSCs per hour on a 24-hour basis regardless of the date. The results are presented by a bar chart in Fig. 2.

According to Fig. 2, we know that the TSCs from the two courses have resemble distributions. The distributions are both trimodal (there are three humps, one small, two large) and left skewed (there appears to be a longer tail on the left side). Furthermore, the time axis (24 hours of a day) can be divided into the following stages. The initial nine hours (0 to 8) is a trough; the next five hours (9 to 13) is a growth; the following five hours (14 to 18) is the first peak; the final five hours (19 to 23) is the second peak. Based on this division of stage, we sum the numbers of TSCs in each stage and list the results in Table I.

As shown in Table I, each type of stages has its own characteristics. In the trough stage, only about 6% of TSCs are sent by users throughout the 9 hours, which means that students hardly stay up late watching course videos. As for the growth stage, its percentage of TSCs increases to over 23%. After that, two peak stages have approximately 60% of TSCs in total. Peak I is from 14 to 18, and peak II is from 19 to 23. This tells us that students tend to do online learning in the afternoons and in the evenings before midnight.

TABLE I: Stages of the Distribution of the Number of TSCs

Stage	Hour	*Fundamentals of Analog Electronics*		*Fundamentals of Digital Electronics*	
		Number	Percentage	Number	Percentage
Trough	0-8	7,901	6.2%	2,605	5.6%
Growth	9-13	29,468	23.0%	11,044	23.9%
Peak I	14-18	46,200	36.2%	17,469	37.8%
Peak II	19-23	44,216	34.6%	15,138	32.7%
Total		127,785	100%	46,256	100%

B. Sentiment Analysis

We compute the sentiment value of every TSC text and obtain the corresponding sentiment type as defined in Section 2.2. To show the overall proportion, we calculate the percentage of each sentiment type, as depicted in Fig. 3. From the chart, we notice that the sentiments are mostly positive or at least neutral. Negative sentiments are only about 20% for both the courses. This indicates that most of the students feel good about these two courses, but still some of them have trouble dealing with the course contents.

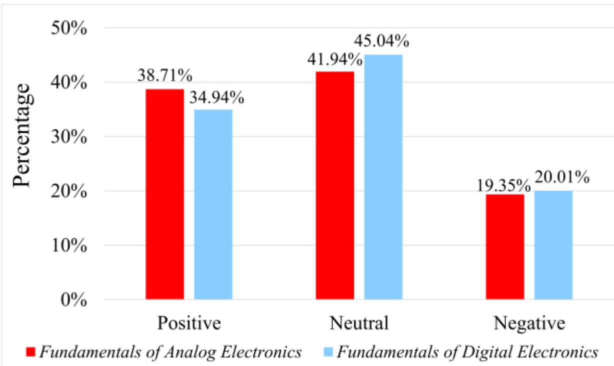


Fig. 3. Overall proportion of the sentiments of TSCs.

There are 46 episodes in *Fundamentals of Analog Electronics* and 51 episodes in *Fundamentals of Digital Electronics*. In order to see how the sentiments vary as the courses proceed, we calculate the average of sentiment values of the TSCs in each episode, and plot the averages in Fig. 4.

[1] https://www.bilibili.com/video/BV1Gt411b7Zq
[2] https://www.bilibili.com/video/BV18p411Z7ce

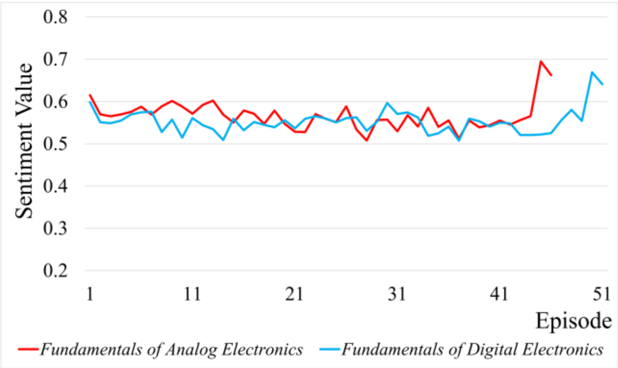


Fig. 4. Average sentiments of the TSCs by episode.

As can be seen in Fig. 4, for the two courses, the sentiments vary similarly by episode. At the very beginning, the emotions are running relatively high, which is due to the students' excitement about starting a new course. However, shortly they feel frustrated at the difficulty of lectures, as can be seen from the decrease of the sentiment values. In the middle of the courses, their sentiments fluctuate yet mostly neutral. The fluctuations mainly come from the difficulty differences between episodes. It is interesting that their emotions reach a highest point when the courses are about to end. This is because the students feel satisfied with their upcoming accomplishment of an online course.

C. Topic Detection

We select the LDA method for topic detection. A tricky problem with this method is to determine the number of topics, which is important to the quality of text mining. To solve the problem, the perplexity has been widely used to find the optimal number of topics [23], which represents the uncertainty of a text document pertaining to a particular topic in a fitted topic model. It has been proved in theory that the smaller the perplexity is, the better the model is. Hence, we may try a range of the number of topics, compare the corresponding perplexities, and then choose a number with the minimal perplexity as the optimal. In practice, however, some other factors need to be taken into consideration. For instance, the number of topics should not be too large for the ease of interpretation. Based on the ideas above, by multiple attempts, we select the number of topics to be 3 for the data from *Fundamentals of Analog Electronics*, and 4 for that from *Fundamentals of Digital Electronics*. The LDA models are then fitted after 100 iterations. After several tests, we decide to set the parameter λ as 0.5. An example of the LDA outputs are visualized in Fig. 5, where original Chinese texts have been replaced by English translations.

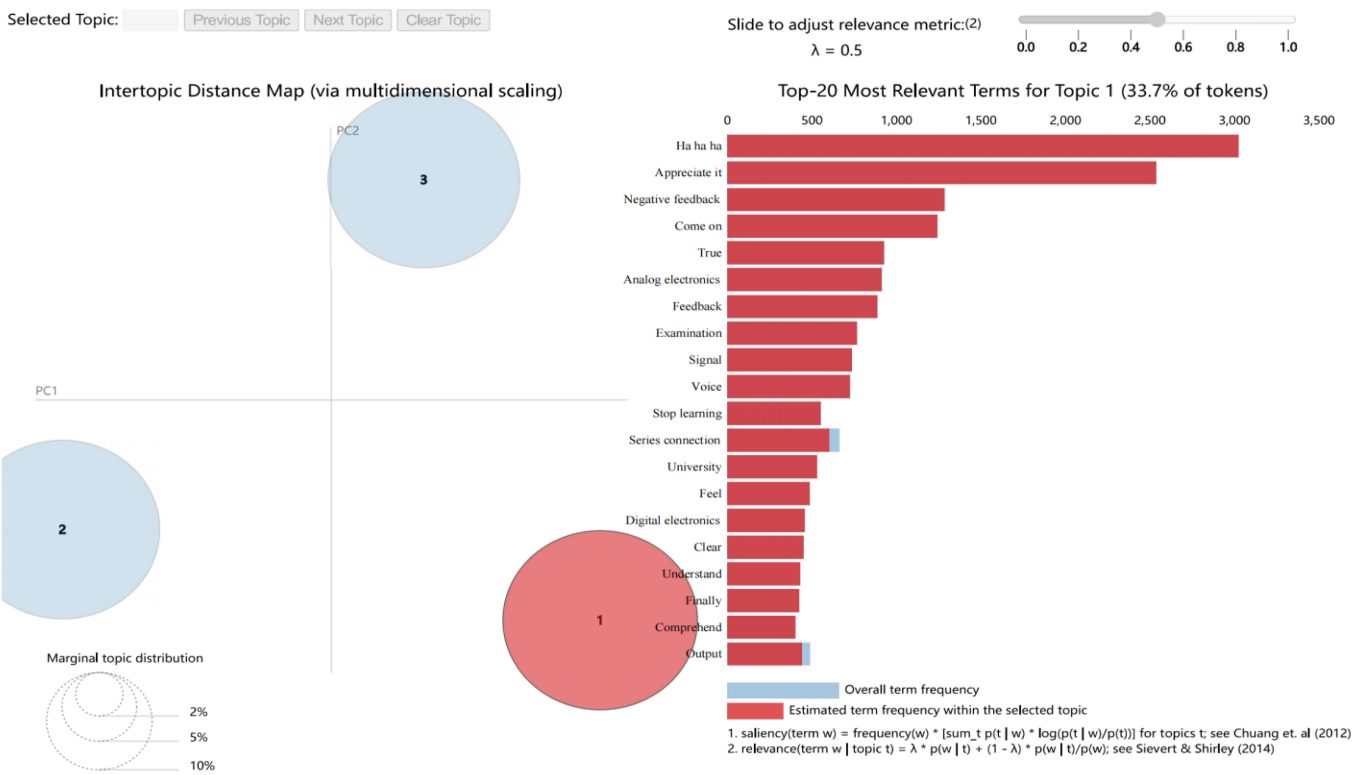


Fig. 5. Visualization of Topic 1 for the TSCs from *Fundamentals of Analog Electronics*.

It can be seen from Fig. 5 that there are two panels in the visualization by LDAvis. The left one gives a global view of the fitted LDA topic model, where the topics are presented as circles in a two-dimensional space. The distance between circles is a two-dimensional projection of the distance between topics, and the area of a circle represents the prevalence of that topic. The three circles shown in Fig. 5 are almost equidistant and similar in size. Besides, no overlapping exists between any two of the circles, which justify our selection of the number of topics. The right panel of the visualization lists the top-20 most relevant terms for Topic 1. For each term, the corpus-wide frequency and the topic-specific frequency are given by a pair of overlaid bars in different colors. One can choose from the circles on the left panel to see the most relevant terms for the corresponding topic on the right. To avoid redundancy, in the paper, we only include the visualization of Topic 1 for the TSCs from *Fundamentals of Analog Electronics* as an example. Contents of the other two topics, as well as those of the four topics from *Fundamentals of Digital Electronics*, are gathered in Table II. Note that the LDAvis always displays the terms in the text language, so we have translated all the topic contents into English in Table II.

TABLE II: TOP-20 MOST RELEVANT TERMS FOR ALL THE LDA TOPICS WITH Λ = 0.5

Fundamentals of Analog Electronics			*Fundamentals of Digital Electronics*			
Topic 1	Topic 2	Topic 3	Topic 1	Topic 2	Topic 3	Topic 4
Ha ha ha	Teacher	Thanks	Thanks	Appreciate it	Ha ha ha	Teacher
Appreciate it	Circuit	Current	Come on	Tsinghua	Output	Trigger
Negative feedback	Du du	Really	Gain admission	Signal	Brothers	Feel
Come on	Voltage	Can't understand	Can't understand	Circuit	Voltage	Ha ha ha ha
True	Ha ha ha ha	Cute	Kaoyan	Time	Real	Guru
Analog electronics	Resistance	Cheers	Input	Data	Analog electronics	Two
Feedback	Easy	Goodbye	Cheers	Clothes	Prof. Wang	Finally
Examination	Brothers	Awesome	Complete	Success	2021	End
Signal	Amplify	This is	Capacitance	Ah ah ah	This is	Easy
Voice	Study	Anybody	Understand	Sixth edition	Digital electronics	Comprehend
Stop learning	Input	Complete	Anybody	Not in the exam	A lesson	Design
Series connection	Parallel connection	Classmate	Hong Wang	University	2022	Exactly
University	Exactly	Gain admission	Awesome	School	System	Classmate
Feel	Capacitance	2020	State	Smoothly	Diode	Clear
Digital electronics	Kaoyan	Ah ah ah	Secondary examination	Go go go	High level	Analysis
Clear	Miss	Why	Recharge	Hope	Read a book	Goodbye
Understand	Two	Diode	Examination	Change	Reverse	A lecture
Finally	2021	Guru	In the book	Period	Seem	Pulse
Comprehend	Thing	In class	Principle	Counter	Functioning	Study
Output	Grounding	SJTU	Cool	Cute	Count	Video

In order to ascertain that the LDA outputs are reliable, we count the term frequency from the raw data to find out the most frequent terms directly. Table III lists the top-10 most frequent terms among the TSCs. Compare Table II and Table III, and we notice repetitions of terms like “Teacher”, “Current”, and “Voltage”. Such repetitions confirm the validity of the LDA results. Comparing the left half and the right half of Table III, we observe that the most frequent words from the two courses have a lot in common. For example, the top 1 word is the same “Teacher”, and the words “This”, “Exactly”, “Not”, etc. are on both the lists. The main difference lies in the rest of the words. Specialized terms, such as “Current” and “Voltage”, appear more frequently in *Fundamentals of Analog Electronics* than in *Fundamentals of Digital Electronics*.

TABLE III: TOP-10 MOST FREQUENT TERMS OF THE TSCS

Fundamentals of Analog Electronics			*Fundamentals of Digital Electronics*		
Word	Times	Percentage	Word	Times	Percentage
Teacher	7787	2.4032%	Teacher	2330	1.9057%
This	2624	0.8098%	Precisely	1028	0.8408%
Ha ha ha	2433	0.7509%	This	977	0.7991%
Precisely	2298	0.7092%	Tsinghua	940	0.7688%
Current	2076	0.6407%	One	814	0.6658%
Circuit	2071	0.6392%	Nice	778	0.6363%
Voltage	2055	0.6342%	Ha ha ha	768	0.6281%
Not	2021	0.6237%	Not	736	0.6020%
No	1662	0.5129%	No	647	0.5292%
Nice	1565	0.4830%	We	598	0.4891%

To better understand the connections between words, we conduct the so-called keywords co-occurrence analysis, where a keyword is defined as a word that frequently appear together (co-occur) with other words. The co-occurrence analysis shows how the keywords are related to one another. The analysis generates a co-occurrence matrix and renders a co-occurrence network accordingly. To carry out the analysis on the TSC data we have collected, we run Python programs to compute the co-occurrence matrices, and then use software tools (including VOSviewer, Pajek, and Gephi) to obtain the co-occurrence networks presented in Fig. 6.

As illustrated in Fig. 6, a keywords co-occurrence network consists of nodes and edges, where a node represents a keyword and an edge connecting two nodes indicates that those two keywords have appeared together at least once. If a keyword co-occurred with another one in a text, then we say this keyword has one connection. The size of a node is proportional to the number of connections that the keyword has to others. The more connections a keyword has, the larger the corresponding node is. In Fig. 6a, the largest nodes are “Voltage” and “Current”. These two words not only appear frequently as discussed before, but also have the most connections to other words. Hence, we may say that “Voltage” and “Current” are important keywords in *Fundamentals of Analog Electronics*. In Fig. 6b, the largest nodes are “Output” and “Input”, and they are followed by “Teacher”, “Voltage”, etc. This shows the difference between the most frequent words and the most connected words in *Fundamentals of Digital Electronics*. It is worth noting that the two networks share many identical keywords, e.g., “Voltage”, “Teacher”, and “Signal”. The words “Voltage” and “Signal” reflects the interrelation in topics between the courses. In contrast, the word “Teacher” shows that students often express their greetings or comments to the lecturer. In each of the courses, the students’ interactions via danmaku focus on the course content itself along with the instructor, which demonstrates their attention on the course study.

Different colors in a network represent different clusters of words. The more often two words co-occur in TSC texts, the more likely those two nodes are to be in a same cluster. There are 5 clusters in each network presented by Fig. 6, and all the clusters have been reshaped into aligned columns for clarity. The last column to the right in both Fig. 6a and Fig. 6b reflects the students' feelings about the teacher's lecture, such as “Great”, “Very well”, and “Can't understand”. The other columns are mostly about specialized terms in electronics, each of which has its own focus. As an example, in Fig. 6a, the first column on the left is built around the semiconductor physics. Compared with the LDA algorithm, the keywords co-occurrence analysis clusters the words in a more explainable way. Yet it is worth mentioning that these two methods output plenty of words that are exactly the same without any prior co-adjustment. The keywords co-occurrence analysis serves as not only a demonstration but also a supplement to the LDA topic model. For example, it can be easily noticed from Fig. 6 that the edges are denser among the nodes on the left and sparser to the right, which indicates that specialized terms are more likely to appear together than are other kinds of words. Such connections are not that obvious if we look at the LDA visualization.

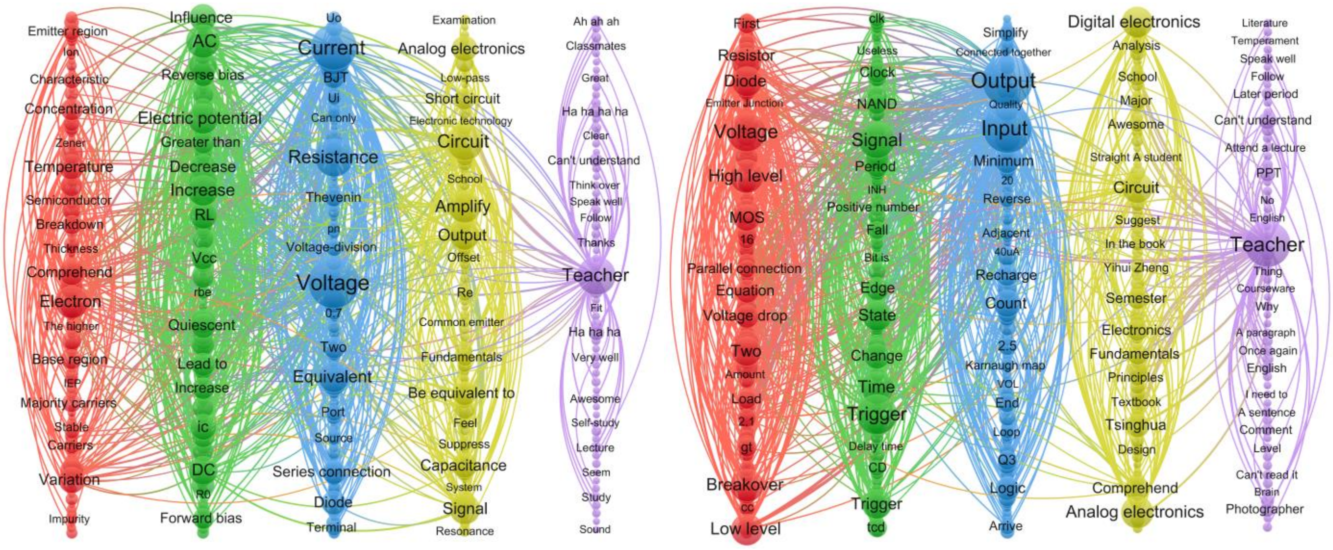


Fig. 6. Keywords co-occurrence networks based on the TSCs from (a) *Fundamentals of Analog Electronics* and (b) *Fundamentals of Digital Electronics*.

V. Conclusion

In the paper, we have built a behavior-sentiment-topic mining procedure and applied it on the TSC data from two online electronics courses. Some useful conclusions are as follows.

Firstly, the TSCs' distributions reflect the students' learning regularities in a day. A significant number of TSCs are posted at 4 p.m. and 8 p.m., which reflects the students' daily routine for studying. However, research has shown that students learn more in classes that are scheduled in the morning than in the afternoon [24]. Therefore, we suggest that students adjust their schedule to make the most of the morning studying hours of a day.

Secondly, the TSCs' sentiments embody the students' emotional experiences of the lectures. Most of the students feel good, or at least not bad, about the lectures. However, at certain points of the lectures, they post a great deal of danmaku that expresses their puzzlement over the teaching materials. Such points should be paid attention by teachers.

Thirdly, the TSCs' topics exhibit the students' major concern for online courses. It may seem weird in the first place that the students turn to danmaku video platforms to learn a course online that they have probably learned at school. As a matter of fact, they are seeking a plain explanation for the difficult parts of the course, as well as an interactive learning environment. This is an inspiration to the platforms that wish to expand their online learning sections.

To prove the wide applicability of the conclusions above, our future work is to collect more danmaku from other online courses, in different disciplines, and on various platforms. Then we will dive into the study of student engagement (e.g., cognitive engagement, emotional engagement, behavioral engagement, and social engagement) based on danmaku, and consider different teacher intervention strategies for different student engagement patterns.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

YK supervised the research; LZ, ZT, LX, and YX conducted the research; ZT and YX collected the data; LZ and ZT analyzed the data; LZ wrote the paper; LX checked and commented on the paper; all authors had approved the final version.

Funding

This work was supported by the University-Industry Collaborative Education Program of the Ministry of Education of China (No. 202101138005), the Undergraduate Innovation and Entrepreneurship Training Program of the Hunan Province (No. S202110543049), and the Teaching Reform Research Program of the Hunan Institute of Science and Technology (No. 2021B28).

References

[1] B. Yang, “Learning alone yet together: an exploratory study of the interaction through online danmaku videos,” in *Proc. 2020 IEEE Learning with MOOCS (LWMOOCS)*, pp. 100-104, 2020.

[2] Q. Bai, Q. V. Hu, L. Ge, and L. He, “Stories that big danmaku data can tell as a new media,” *IEEE Access*, vol. 7, pp. 53509-53519, 2019.

[3] Q. Wu, Y. Sang, S. Zhang, and Y. Huang, "Danmaku vs. forum comments: understanding user participation and knowledge sharing in online videos," in *Proc. 2018 ACM Conference on Supporting Groupwork*, pp. 209-218, 2018.

[4] J. Zhang, Y. Yang, and X. An, "Enabling learning interaction through 'bullet screen' videos," *Distance Education in China*, no. 11, pp. 22-30+79-80, 2017.

[5] Q. Wu, Y. Sang, and Y. Huang. "Danmaku: a new paradigm of social interaction via online videos," *ACM Transactions on Social Computing*, vol. 2, no. 2, pp. 1-24, 2019.

[6] Y. Yao, J. Bort, and Y. Huang, "Understanding danmaku's potential in online video learning," in *Proc. 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 3034-3040, 2017.

[7] J. Leng, J. Zhu, X. Wang, and X. Gu, "Identifying the potential of danmaku video from eye gaze data," in *Proc. 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, pp. 288-292, 2016.

[8] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: a survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014.

[9] Z. Nasim, Q. Rajput, and S. Haider, "Sentiment analysis of student feedback using machine learning and lexicon based approaches," in *Proc. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS)*, pp. 1-6, IEEE, 2017.

[10] Z. Li, R. Li, and G. Jin, "Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary," *IEEE Access*, vol. 8, pp. 75073-75084, 2020.

[11] S. Wang, Y. Chen, H. Ming, H. Huang, L. Mi, and Z. Shi, "Improved danmaku emotion analysis and its application based on bi-lstm model," *IEEE Access*, vol. 8, pp. 114123-114134, 2020.

[12] Q. Bai, K. Wei, J. Zhou, C. Xiong, Y. Wu, X. Lin and L. He, "Entity-level sentiment prediction in danmaku video interaction," The Journal of Supercomputing, vol. 77, no. 9, pp. 9474-9493, 2021.

[13] L. Jia and Z. Li, "Emotional analysis on the public sentiment of students returning to university under covid-19," in *Proc. 2020 16th International Conference on Computational Intelligence and Security (CIS)*, pp. 238-242, IEEE, 2020.

[14] Z. Mottaghinia, M.-R. Feizi-Derakhshi, L. Farzinvash, and P. Salehpour, "A review of approaches for topic detection in twitter," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, no. 5, pp. 747-773, 2021.

[15] R. Alghamdi and K. Alfalqi, "A survey of topic modeling in text mining," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 1, 2015.

[16] B. V. Barde and A. M. Bainwad, "An overview of topic modeling methods and tools," in *Proc. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 745-750, IEEE, 2017.

[17] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169-15211, 2019.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. 1, pp. 993-1022, 2003.

[19] B. Wu, E. Zhong, B. Tan, A. Horner, and Q. Yang, "Crowdsourced time-sync video tagging using temporal and personalized topic modeling," in *Proc. 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 721-730, 2014.

[20] Q. Bai, Q. Hu, F. Fang, and L. He, "Topic detection with danmaku: a time-sync joint nmf approach," *International Conference on Database and Expert Systems Applications*, pp. 428-435, Springer, 2018.

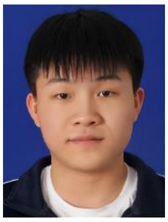
[21] L. Huang, J. Ma, and C. Chen, "Topic detection from microblogs using t-lda and perplexity," in *Proc. 2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*, pp. 71-77, IEEE, 2017.

[22] C. Sievert and K. Shirley, "Ldavis: A method for visualizing and interpreting topics," in *Proc. the Workshop on Interactive Language learning, Visualization, and Interfaces*, pp. 63-70, 2014.

[23] M. Hasan, A. Rahman, M. Karim, M. Khan, S. Islam, and M. Islam, "Normalized approach to find optimal number of topics in latent dirichlet allocation (lda)," in *Proc. International Conference on Trends in Computational and Cognitive Engineering*, pp. 341-354, Springer, 2021.

[24] N. G. Pope, "How the time of day affects productivity: Evidence from school schedules," *Review of Economics and Statistics*, vol. 98, no. 1, pp. 1-11, 2016.

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 include data mining applications.



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.

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 is education informatization.

Yu'an Xiang is currently pursuing his Master's degree at the Hunan Institute of Science and Technology. His research interests include text mining, sentiment analysis, and deep learning.

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.