We-Media Facilitating Cultural Transmission: LDA-based Topic Modelling and Sentiment Analysis

Zilong Zhong

Beijing Foreign Studies University

ABSTRACT

In the current study, comments on the YouTube video series “Chinese festival food” were collected using an open-source web crawler. Sentiment analysis was conducted using the VADER sentiment analysis tool, and the topic distribution of sentiment tendencies was explored based on the LDA model. The study found that: (1) high frequency words and their collocations in the comments of Li Ziqi’s “Chinese festival food” video series indicated that viewers greatly appreciated the content. (2) The comments of Li Ziqi’s “Chinese festival food” video series conveyed a positive sentiment. (3) The positive sentiment of Li Ziqi’s “Chinese festival food” video series mainly focused on topics such as food, video, characters, life, and culture. These findings demonstrate that social media has become an emerging and powerful channel for cultural transmission and transcultural communication. This study aims to provide new research methods and ideas for the study of cultural transmission, and shed light on the transmission of culture across the world.

Keywords: we-media, cultural transmission, LDA-based topic modeling, VADER, sentiment analysis.

Classification: FoR Code: 2001

Language: English

Great Britain Journals Press

LJP Copyright ID: 573351
Print ISSN: 2515-5784
Online ISSN: 2515-5792

London Journal of Research in Humanities and Social Sciences

Volume 23 | Issue 18 | Compilation 1.0

We-Media Facilitating Cultural Transmission: LDA-based Topic Modelling and Sentiment Analysis

Zilong Zhong

ABSTRACT

In the current study, comments on the You Tube video series "Chinese festival food" were collected using an open-source web crawler. Sentiment analysis was conducted using the VADER sentiment analysis tool, and the topic distribution of sentiment tendencies was explored based on the LDA model. The study found that: (1) high frequency words and their collocations in the comments of Li Ziqi's "Chinese festival food" video series indicated that viewers greatly appreciated the content. (2) The comments of Li Ziqi's "Chinese festival food" video series conveyed a positive sentiment. (3) The positive sentiment of Li Ziqi's "Chinese festival food" video series mainly focused on topics such as food, video, characters, life, and culture. These findings demonstrate that social media has become an emerging and powerful channel for cultural transmission and transcultural communication. This study aims to provide new research methods and ideas for the study of cultural transmission, and shed light on the transmission of culture across the world.

Keywords: we-media, cultural transmission, LDA-based topic modelling, VADE, sentiment analysis.

Author: Research Institute of Foreign Languages, Beijing Foreign Studies University, Beijing, 100089, China.

I. INTRODUCTION

In today's world, the rapid growth of the economy and the transmission of cultures across borders have led people to share a diverse range of foods. The emergence of the we-media era has further diversified the pathways of cultural communication, making it possible for food cultures from different countries to be exposed to a global audience. Li Ziqi, a Chinese food short video creator, has leveraged we-media platforms to share traditional Chinese food culture with her fans worldwide. To gain insights into how viewers respond to cultural transmission via we-media and their willingness to engage in transcultural communication on social media, we collected and analyzed hot comments from Li Ziqi's "Chinese festival food" video series. Through this analysis, we aim to provide useful insights into the foreign communication of Chinese culture.

As a new technology in the field of text mining, topic mining can automatically annotate and extract representative words, phrases or sentences. This technology has unique advantages in processing large-scale text, quickly and effectively identifying core focuses and valuable information from big data (Kherwa& Bansal, 2019). The Latent Dirichlet Allocation (LDA) model method, based on the Definneti theorem, assumes that each document is composed of several topic, and each topic is defined as the distribution of a vocabulary. LDA model can capture the Dirichlet probability distribution within documents, and estimate the final number of topics of the given document via the Gibbs sampling algorithm (Chen, 2017; De Finetti, 2017). Social networks, as a rich source for analyzing perspectives and behaviors of users, is a popular subject of the LDA model. Guimarães et al. built datasets consisting of tweets of politicians and the replies they received. Taking the 2016 US Election and the UK Brexit as examples, they identified keywords in the relative tweets, and analyzed how the political orientation of the public were influenced by power users (Guimarães et al., 2017). Gao et al. (2022) collected posts on Weibo and Twitter related to public opinions on working from home during the COVID-19 pandemic, andanalyzed the major challenges of working from home by classifying topic via LDA topic modelling. The study conducted by Ginossar et al. (2022) identified YouTube videos sharing tweets about COVID-19 vaccine posted before the vaccine rollout. Through the LDA topic modelling analysis, conspiracy theories dominated forming anti-vaccination frames, seriously impeding the uptake of vaccination against infectious diseases. According to aforementioned applications, LDA topic modelling can be a powerful tool for comprehending the responses and dialogues among users in online communities.

Sentiment analysis or opinion mining refers to the study of subjective texts with emotions, mining the sentimental tendencies and differentiating emotional attitudes. This concept was firstly proposed by Nasukawa and Yi in 2003 (Nasukawa& Yi, 2003). The approaches of sentiment analysis can be broadly divided into three categories, machine learning-based methods (e.g., Duan et al., 2020; Li & Liu, 2014), deep learning-based methods (e.g., Ma et al., 2018; Majumder et al., 2019; Sadr et al., 2020; Wang et al., 2016) and lexicon-based methods (e.g., Dang et al., 2010; Dolores Molina-González et al., 2015; Muhammad et al., 2016; Saif et al., 2016). One significant benefit of lexicon-based approaches is that it does not require training data, while the lexicon employed for sentiment analysis must be matched to the topic of interest (Ligthart et al., 2021). Valence Aware Dictionary and sEntiment Reasoner (VADER) algorithm, as one of the lexicon-based approaches, is a rule-based sentiment analysis tool with an open-source lexicon. VADER algorithm traverses a string of text to determine whether any of the words are included in the VADER lexicon (Bonta et al., 2019; Britzolakis et al., 2020). It is built from a valence-based, generalizable, human-curated gold standard lexicon and does not require any training data. Besides, VADER excels in mining the sentiment from social media, movie or book comments, and commodity evaluations. Park and Seo, for example, collected comments about three artificial intelligence assistants, Siri, Google Assistant, and Cortana, from Twitter, and converted this opinion into sentiment scores via VADER to rank them by statistical analysis (Park & Seo, 2018). Jelodar et al. utilized VADER to analyze user comments of the Oscar-nominated movie trailers on YouTube, and further mine the topics of these comments, assess users' focuses on these movies (Jelodar et al., 2021).

Through an interdisciplinary approach, this study examines the comments left on Li Ziqi's"Chinese festival food" video series using VADER sentiment analysis and LDA topic modelling to investigate the distribution of sentiment and topics among viewers. The aim is to gain insights into the transmission of Chinese culture through foreign communication. This study offers innovative research methods and ideas for exploring cultural transmission and contributes to our understanding of how cultures are shared and transmitted across the globe.

II. MATERIALS AND METHODS

The present study aims to answer the following questions:

- What are the high-frequency words in the comments on Li Ziqi's "Chinese festival food" video series, and what are their collocation characteristics?
- What is the sentiment tendency of the comments on Li Ziqi's "Chinese festival food" video series?
- What are the characteristics of the topic distribution within the comments on Li Ziqi's "Chinese festival food" video series that have a positive sentiment?

The technical route of this study can be generally divided into three steps: data crawling, data processing, and data analysis, as shown in Figure 1. Data processing includes text cleansing and text pre-processing, that is, tokenization and text denoising. Three steps included in data analysis are word frequency statistics, sentiment analysis, and topic mining.

2.1 Data Crawling

Web scraper, YouTube API version 3, in Python scripts was applied to extract hot comments of Li Ziqi's "Chinese festival food" video series (yasmina et al., 2016). The titles of each video were listed in Table 1, and views and the number of hot comments were shown in Figure 2. The collected data was outputted as csv files for further processing. All data was collected at 10:12 on August 12, 2022.

2.2 Data Processing

Data processing was divided into two parts: text cleansing and text pre-processing. Comments without text, comments that are not in English, comments that contain emoticons only, comments with obvious typos, and reduplicative comments were removed in text cleansing. The numbers of comments after the data cleansing of each video were shown in Figure 2.

After data cleansing, the data was integrated into one csv file containing 5510 comments in total, which was imported into Text Analytics Toolbox embedded in MATLAB R2021b for tokenization and denoising. Comments was tokenized into arrays consisting of several tokens. Denoising process included the following steps:

- Removing stop words, including articles (e.g., "a", "an" and "the") and prepositions (e.g., "to", "for", and "of").
- Lemmatization, for example, "watching", "watched" or "watches" into "watch".

2.3 Data Analysis

2.3.1 Word Frequency Statistics

Bag of words (BoW) is one of the frequently used models for natural language processing and text analysis. Different from word embedding model, eigenvectors are transformed from documents rather than single words. The BoW model assumes that a document is merely considered to be a collection of words, with word order, grammar, syntax, and other factors being disregarded. The appearance of each word in the document is independent, not influenced by the appearance of other words (Qader et al., 2019). The co-occurrence matrix is calculated by multiplying the word-count matrix by its transpose, then the co-occurrence network of a specific word can be segmented from the whole co-occurrence map of the document.

2.3.2 Sentiment Analysis

VADER is a method for text sentiment recognition based on lexicon and grammar rules, firstly published in 2014 (Hutto & Gilbert, 2014). The VADER lexicon consists of more than 7,500 lexical features (including adjectives, nouns, adverbs, etc.) showing the polarity and frequency of sentiment, scored from -4 to +4. Different from other proposed sentiment lexicon, the VADER lexicon considers common abbreviations (e.g., "LOL" and "BTW") and slang (e.g., "bruh" and "giggly") to deal with sentiment discrimination of non-standard sentences in the social network environment such as Twitter. The score judged by the VADER algorithm is influenced by punctuations, capitalization, degree modifiers, conjunctions, and negation. The compound score given by VADER algorithm ranges from -1 to +1, where -1 indicates the most extreme negative, and +1 indicates the most extreme positive. The typical thresholds for classifying comments as positive, neutral, or negative are shown in Figure 3.

2.3.3 Topic mining

The LDA model was then applied to mine the underlying topics of the positive comments for determining the effect of LiZiqi's "Chinese festival food" video series on Chinese food cultural transmission. Based on bag of words of documents, LDA assumes that D documents consist of K topics, and the k-th topic are identified by V words. The plate notation of the LDA model is shown in Figure 4. In LDA, the topic distributions in all documents share the common Dirichlet prior α, a K-dimensional vector, and topics over the d-th document follow a Dirichlet distributionθ_d (Liu et al., 2016),

$$\theta_d = Dirichlet(\alpha) \tag{1}$$

Similarly, the word distributions of topics share the common Dirichlet prior η, a V-dimensional vector, and words over the k-th topic follow another Dirichlet distribution β_k,

$$\beta_k = Dirichlet(\eta) \quad (2)$$

For the n-th word in the d-th document, its topic number $z_{d,n}$ can be found in the Dirichlet distribution θ_d,

$$z_{d,n} = multi\left(\theta_d\right) \quad (3)$$

The probability distribution of the $z_{d,n}$-th word, $w_{d,n}$, can be determined from the Dirichlet distribution $\beta_{z_{d,n}}$ of the $z_{d,n}$-th topic,

$$w_{d,n} = multi\left(\beta_{z_{d,n}}\right) \quad (4)$$

The number of topics was determined by the goodness-of-fit of LDA models, evaluated by the perplexity, and the time elapsed for converging. Besides, the t-distributed stochastic neighbour embedding (t-SNE) algorithm was also applied to plot the document topic mixtures to visualize the clustering of similar documents.

III. RESULTS

3.1 High-frequency Word Analysis

MATLAB R2021b software was adopted to plot a histogram of the number of video views (Figure 1), the word cloud of the top 100 words with high frequency in video comments and the co-occurrence network of the high-frequency word "video" (Figure 5).

In terms of the number of views on the YouTube platform for the "Chinese festival food" video series, the video "Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival" had the highest number of views as at 10:12 on 12 August 2022, with 115,382,645 views, followed by "Happiness-filled family dinner, fortune-and-wellness-filled year ahead - New Year Dinner", "Dragon boat zongzi~" and "Chinese New Year's decorations, goods and snacks!", with 18,999,542, 18,212,865 and 16,563,637 views respectively. The total number of views for the series is nearly 200 million (193,663,125), far exceeding the number of views for other videos in the same category on the platform, which reveals that the video series attracts a consistently much attention among similar videos and shows a good dissemination effect.

The size of the word cloud indicates the frequency of the word in the video comments. The words "video", "love", "like", "make" and "beautiful" are the top five most frequent words, with frequencies of 1,503, 1,400, 724, 653, and 646, respectively, among which the words "love", "like", and "beautiful" (examples 1-3) obviously show positive attitudes towards the video series, expressing viewers' much affection and appreciation for the video series.

(1) Woow it was so beautiful!!! I found an amazing person to follow today. Love it so much. Chinese culture is so aesthetic.
(2) I'm not Chinese but I really like the way how they beautifully celebrate their new year. I really love the sweet food, decoration, and her garden. I'm speechless. she is very good artist. I love her art so much.
(3) This is really inspiring, she did everything by her own. Landscape around her is fresh and beautiful. Thank you so much for such a great video.

The lines between high-frequency words and collocations in the collocation network diagram indicate the association between words, and the length of the lines between high-frequency words and collocations demonstrates the strength of the collocation. The high-frequency word "video" includes collocations such as "new", "watch", "like", "love", "beautiful" and "Chinese" (examples 4-7), which are mainly words expressing positive sentiment, indicating that viewers hold a more positive attitude towards videos.

(4) I love watching your videos the place is beautiful, God bless you with a talent it's very

sweet to have your grandma in your videos. I am from Belize.

(5) This is another incredibly good, brilliant, beautiful, and INSPIRING video from Ms. Liziqi (who is also beautiful, hardworking, and brilliant.). We over here in Los Angeles, love your videos so much. The videos are also very therapeutic, soothing, and relaxing. Thank you so much for sharing all your videos, and for all your hard work.

(6) Ziqi is very busy with all kinds of planting work after the Chinese New Year. Everything is great for her, just she doesn't have time to cut a video yet. Be patient, a new video is coming soon.

(7) I've learned more about Chinese culture watchingZiqi videos than I ever have in western schools. The more I watch these videos, the more I become inundated with reasons why capitalist culture is death.

Giventogether, along number of views of Li Ziqi's "Chinese festival food" video series reveals the constant interest of viewers, and the high-frequency words and their collocations in the video comments reflect the much appreciation of viewers for video series.

3.2 Sentiment Analysis

VADER was conducted to measure the sentiment scores of video comments in the current study. Since it is a lexicon and algorithm-based sentiment analysis tool designed specifically for social media content, it could ensure reliability and validity when analyzing the video comment data. The tool adopts the vaderSentimentLexicon function to calculate sentiment scores. The closer the score is to one, the more positive the sentiment is. On the contrary, the closer the score is to minus one, the more negative the sentiment is. A score of zero means neutral sentiment. The number of comments with positive, neutral and negative sentiment was calculated to be 4,655, 640, and 219, respectively. The average scores of positive and negative sentiments were 0.6805, and -0.3093 respectively (Figure 6(a)). The results indicate the high acceptance of Li Ziqi's "Chinese festival food" video series among viewers, with positive, neutral, and negative comments accounting for 84.4%, 11.6%, and 4% respectively (Figure 6(b)).

To further examine the lexical usage characteristics in the positive comments, the word cloud of positive comment was plotted (Figure 7). Among the positive video comments, "love" was the most frequent word, with a frequency of 1,391, uncovering the popularity of the video series among viewers. In addition, words such as "beautiful", "good", "happy", and "amazing" (examples 8-11) stand out of the word cloud, showing a clearly positive sentiment tendency. Hence, it is reasonable to conclude that the video series is highly appreciated by viewers.

(8) Such a beautiful woman so skilled in traditional ways that people don't know anymore. Beautifully filmed.

(9) This is another incredibly good, brilliant, beautiful, and INSPIRING video from Ms. Liziqi (who is also beautiful, hardworking, and brilliant.). We over here in Los Angeles, love your videos so much. The videos are also very therapeutic, soothing, and relaxing. Thank you so much for sharing all your videos, and for all your hard work.

(10) Something about this makes me really happy, but also sad. I live in urban new york and will never be able to experience something like this. I would rather live and work on a farm and do that for my lifetime instead of getting a degree and work my life away in an office.

(11) I couldn't click fast enough when I received the notification. It's always a joy to watch Ziqi's video, wondering what she's going to cook or make. This lady is beautiful, amazing, yet humble and sincere. An amazing human being.

3.3 Topic Mining

To obtain the characteristics of the topic distribution of positive sentiment, the LDA model was built for topic mining. First of all, the optimal number of topics for positive sentiment in video comments was obtained by confusion analysis. The line graph of validation perplexity and time elapsed with the number of topics are shown in Figure 8(a). The number of topics for positive comments was set to 9 since the validation perplexity was lower while the time elapsed was shorter at this point. Then the fitlda function was adopted to match the model solver with the best fit (Figure 8(b)). The solver "cgs" have the lowest perplexity and shorter time elapsed, so "cgs" was selected as the solver to build the LDA model.

The result of topic mining of positive sentiment is shown in Figure 9. Figure 9(a) shows word clouds of each topic, and Figure 9(b) shows the topic mixtures generated by the t-SNE algorithm.

It is vividly depicted that topic 1 focuses on food, with words such as "good", "amazing", and "delicious" expressing the viewers' praise for the food made in the video. Topic 2 is mainly about video, with words such as "beautiful", "amazing", "nice", and "talented" indicating the viewers' approval of the videos' creativity. Topic 3, 4, 7, and 9 concentrate on the main characters, Li Ziqi, grandma, and their puppy, with words "liziqi", "grandma", "ziqi" and "puppy", together with collocations such as "love", "miss", and "cute" showing how much viewers love and are impressed by the three main characters in the video series. Topic 5 is about life, with words such as "great", "beautiful" and "peaceful" reflecting the viewers' desire for a rural and idyllic life. The video is a microcosm of the life of Li Ziqi and his grandmother in the countryside, where the characters live in nature, away from the hustle and bustle of the city. While documenting the process of making food, the video also presents an amazing picture of man and nature living in harmony to viewers in a bustling city. Topic 6 and 8 focus on culture and festivals. The videos share Chinese culture with viewers around the world in a unique way, from the perspective of traditional Chinese festival food. While enjoying the videos, viewers learn about traditional Chinese festivals and experience Chinese culture at the same time.

Among 4655 positive comments, 633 comments (13.6%) were classified into topic 6 and 8 (examples 12-14), indicating that communication and culture are inseparable. Li Ziqi's "Chinese festival food" video series take advantage of we-media to promote the cross-cultural communication effectively and efficiently. The videos originate from her real life, conveying a cultural message that is conducive to viewers' perception on different cultures and more likely to resonate with viewers in a genuine way. The interactive transmission of information through viewers' watching, commenting and sharing reduces the sense of oppression generated by cultural conflict. We-media constructs an innovative platform for viewers to interact with heterogeneous cultures.

(12) It is good if you use nature for such a respect for you and I will visit China at some point to see nature and get to know your culture. Greetings from Germany.

(13) I know we Indian and Chinese are going through war, but I love that there are still people in only these two countries which has culture that can't be compared to any one... our both cultures are purest and graceful love from India.

(14) Watching those videos, I understand how beautiful, ancient and interesting Chinese culture is. Also, as Italian, I confess that Chinese, and eastern chicken in general, is as various as our food! I would like to learn to cook like you!

IV. DISCUSSION

In this interdisciplinary study, comments on the YouTube video series "Chinese festival food" were collected using an open-source web crawler.Sentiment analysis was conducted using the VADER sentiment analysis tool, and the topic distribution of sentiment tendencies was explored based on the LDA model. The Analysis of high-frequency words revealed that the most commonly used words in the comments on video series "Chinese festival food" were "video", "love", "like", etc. The collocations of the high-frequency word "video" included "gift", "wonderful", "talented", indicating a high lof appreciation among viewers for the video series. TheSentiment analysis found that there were 4,655 positive sentiment comments on Li Ziqi's "Chinese festival food" video series, accounting for 84.4% of the comments, with an average score of 0.6805, indicating an overwhelmingly positive sentiment among viewers. Topic mining revealed that the positive sentiment of Li Ziqi's "Chinese festival food" video series mainly focused on food, video, character, life, and culture.

The topics discussed above reveal that viewers ofLi Ziqi's "Chinese festival food" video series are interested not only in the food itself, but also in the characters featured thevideos, their lifestyle, thevideos' creativity and talent, and the culture they convey. Several factors contribute to the immense popularity of Li Ziqi's videos worldwide. First, the videos showcase innovative approaches to making Chinese food by using traditional cooking techniques, especially in the context of traditional Chinese festival food. As food is a universal language and an essential part of people's daily lives, it becomes an effective entry point to attract viewers from different countries.

Second, the videos feature an underlying narrative that satisfies the emotional needs of viewers. The slow-paced and narrative style of the videos, which contrasts the quiet countryside with the bustling city, evoke a wide range of emotional responses from viewers. Third, the videos effectively communicate cultural information, making them a vivid and concrete carrier for sharing traditional Chinese culture with overseas viewers, leaving a deep impression in their minds. Finally, the videos portray a new character, Li Ziqi, who is independent, self-reliant, hard-working, and talented, breaking stereotypes in the minds of viewers and triggering contemplation.In conclusion, Li Ziqi's video series have been well-received worldwide because of the high-quality and talented videos she has created. They provide viewers with a visual feast and an aural treat, reflecting the value of harmonious coexistence between human beings and nature, and enabling viewers to appreciate the spiritual value of shared culture at different stages of life. Therefore, cultural transmission needs to find the right perspective, focus on the narrative, and express in a language that is understandable across the world.

The results of the present study underscore the importance of finding the right perspective, focusing on storytelling, and using a universal language for effective cultural transmission. In addition, the study highlights the crucial role of we-media in facilitating cross-cultural communication and understanding. In the digital age, we-media has significantly reduced the barriers of time and space, bringing people from diverse backgrounds together on a common platform. Platforms such as YouTube have emerged as powerful vehicles for cultural transmission, enabling creators and viewers to connect and share their experiences. Overall, this study presents novel research methods and ideas that can pave the way for a deeper understanding of cultural transmission in the global context.

REFERENCE

1. Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology, 8* (S2), 1-6. https://doi.org/10.51983/ajcst- 2019.8. S2.2037
2. Britzolakis, A., Kondylakis, H., & Papadakis, N. (2020). A Review on Lexicon-Based and Machine Learning Political Sentiment Analysis Using Tweets. International *Journal of Semantic Computing, 14*(04), 517-563. https://doi.org/10.1142/s1793351x2030 0010
3. Chen, L.-C. (2017). An effective LDA-based time topic model to improve blog search performance. *Information Processing & Management, 53*(6), 1299-1319. https://doi.org/https://doi.org/10.1016/j.ipm.2017.08.001
4. Dang, Y., Zhang, Y., & Chen, H. (2010). A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews. *IEEE Intelligent Systems, 25* (4), 46-53. https://doi.org/10.1109/MIS.2009.105
5. De Finetti, B. (2017). *Theory of probability: A critical introductory treatment* (Vol. 6). John Wiley & Sons.
6. Dolores Molina-González, M., Martínez-Cámara, E., Teresa Martín-Valdivia, M., & Alfonso Ureña-López, L. (2015). A Spanish semantic orientation approach to domain adaptation for polarity classification.

Information *Processing & Management, 51* (4), 520-531. https://doi.org/ https:// doi.org/10.1016/ j.ipm.2014.10.002
7. Duan, J., Luo, B., & Zeng, J. (2020). Semi-supervised learning with generative model for sentiment classification of stock messages. *Expert Systems with Applications, 158*, 113540. https://doi.org/https:// doi.org/10.1016/j.eswa.2020.113540
8. Gao, J., Ying, X., Cao, J., Yang, Y., Foong, P. S., & Perrault, S. (2022). *Differences of Challenges of Working from Home (WFH) between Weibo and Twitter Users during COVID-19*. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA. https://doi.org/10.1145/3491101.3519790
9. Ginossar, T., Cruickshank, I. J., Zheleva, E., Sulskis, J., & Berger-Wolf, T. (2022). Cross-platform spread: vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations. *Human Vaccines & Immunotherapeutics, 18*(1), 1-13. https:// doi.org/ 10.1080/21645515. 2021.2003647
10. Guimarães, A., Wang, L., &Weikum, G. (2017, 18-21 Nov. 2017). Us and Them: Adversarial Politics on Twitter. (Ed.), ^(Eds.). 2017 IEEE International Conference on Data Mining Workshops (ICDMW).
11. Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 216-225. https://ojs.aaai. org/index.php /ICWSM/article/view/14550
12. Jelodar, H., Wang, Y., Rabbani, M., Ahmadi, S. B. B., Boukela, L., Zhao, R., &Larik, R. S. A. (2021). A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments. *Multimedia Tools and Applications, 80*(3), 4155-4181. https:// doi.org/10.1007/s11042-020-09755-z
13. Kherwa, P., & Bansal, P. (2019). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems, 7* (24), e2. https://doi.org/10.4108/eai.13-7 2018. 159 623.
14. Li, G., & Liu, F. (2014). Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. *Applied Intelligence, 40* (3), 441-452. https://doi.org/ 10.1007/ s10489-013-0463-3
15. Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review, 54* (7), 4997-5053. https://doi.org/10.1007 /s10462 -021 -09973-3
16. Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus, 5*(1), 1608. https://doi.org/10.1186 s40064-016-3252-8
17. Ma, Y., Peng, H., & Cambria, E. (2018). Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. *Proceedings of the AAAI Conference on Artificial Intelligence, 32*(1). https://doi.org/ 10.1609/aaai.v32i1.1 2048
18. Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and Sarcasm Classification With Multitask Learning. *IEEE Intelligent Systems, 34* (3), 38-43. https://doi.org/ 10.1109/MIS.2019.2904691
19. Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-Based Systems, 108*, 92-101. https://doi.org/ 10.1016/j.knosys.2016.05.032
20. Nasukawa, T., & Yi, J. (2003). *Sentiment Analysis: Capturing Favorability Using Natural Language Processing*. The 2nd International Conference on Knowledge Capture, Sanibel Island, FL, USA. https://doi.org/10.1145/94564 5.945658
21. Park, C. W., &Seo, D. R. (2018, 26-28 April 2018). Sentiment analysis of Twitter corpus related to artificial intelligence assistants. (Ed.),^(Eds.). 2018 5th International Conference on Industrial Engineering and Applications (ICIEA).

22. Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019, 23-25 June 2019). An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges. (Ed.),^(Eds.). 2019 International Engineering Conference (IEC).
23. Sadr, H., Pedram, M. M., & Teshnehlab, M. (2020). Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis. *IEEE Access, 8*, 86984-86997. https://doi.org/10.1109/ ACCESS.2020.2992063
24. Saif, H., He, Y., Fernandez, M., & Alani, H. (2016). Contextual semantics for sentiment analysis of Twitter. *Information Processing & Management, 52* (1), 5-19. https://doi.org/https://doi.org /10.1016/j.ipm.2015.01.005
25. Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016). Attention-based LSTM for aspect-level sentiment classification. (Ed.),^(Eds.). Proceedings of the 2016 conference on empirical methods in natural language processing.
26. Yasmina, D., Hajar, M., & Hassan, A. M. (2016). Using YouTube Comments for Text-based Emotion Recognition. *Procedia Computer Science, 83*, 292-299. https://doi.org/https://doi.org/10.1016/j.pro cs.2016.04.128

Table 1: Titles of videos in Li Ziqi's "Chinese festival food" video series

Video Number	Titles
Video 1	Chinese New Year's decorations, goods and snacks!
Video 2	Dragon boat zongzi~
Video 3	As a kid, I used to eat zongzi wrapped in shells of bamboo shoots
Video 4	Happiness-filled family dinner, fortune-and-wellness-filled year ahead - New Year Dinner
Video 5	Peanut and melon seeds, dried meat, dried fruit, snowflake cake - snacks for Spring Festival
Video 6	Winter cuisine with great meaning—flower shaped shrimp
Video 7	Crust crisp, gravy rich flavor, touched the soul of the Su-style meat moon cake

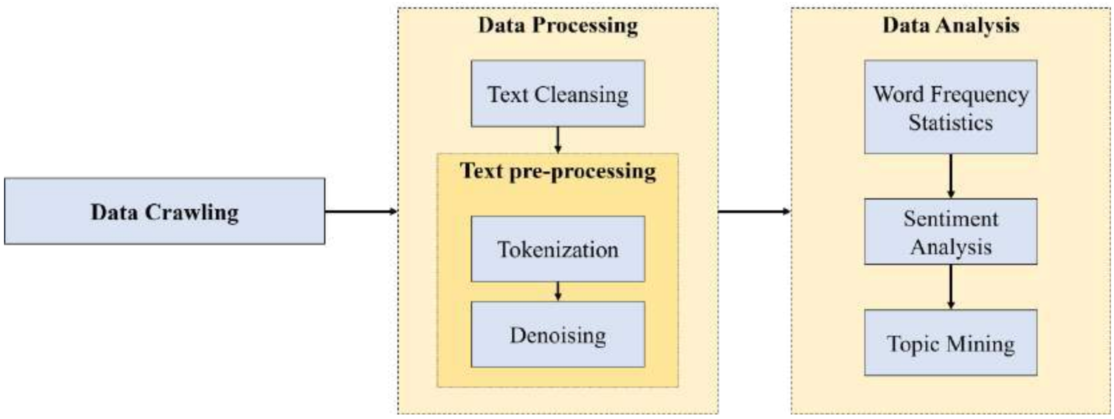


Figure 1: Flow chart of study processes.

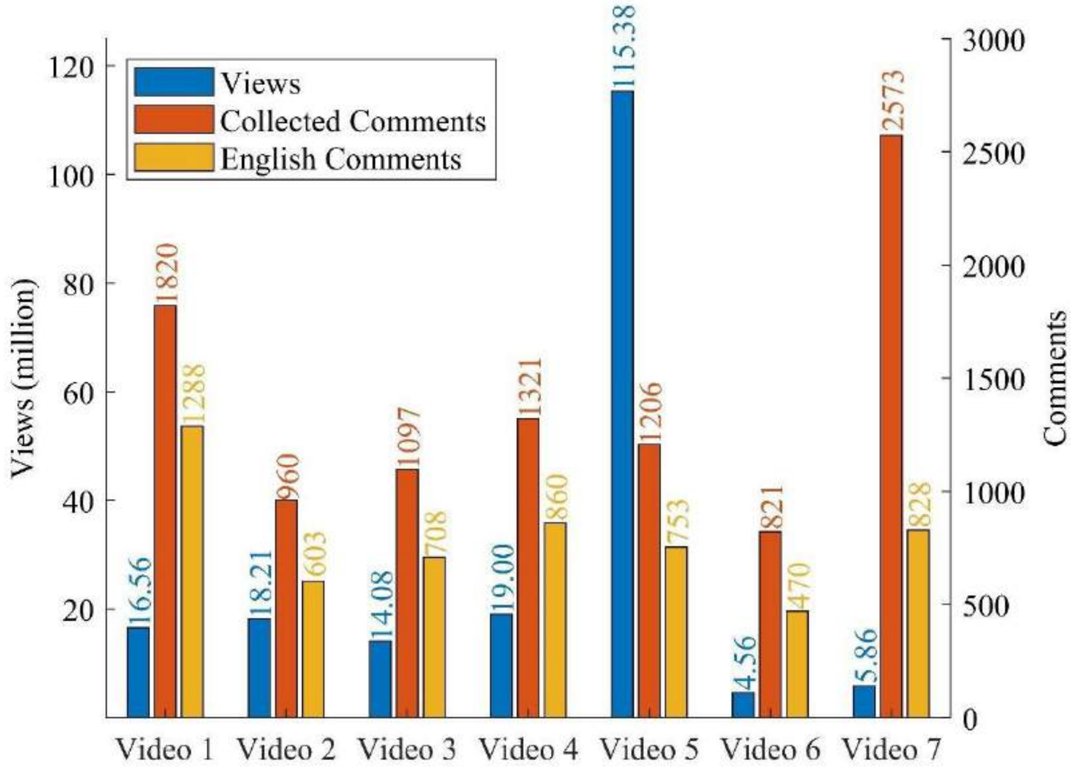


Figure 2: Views, number of collected comments and number of comments after data cleansing.

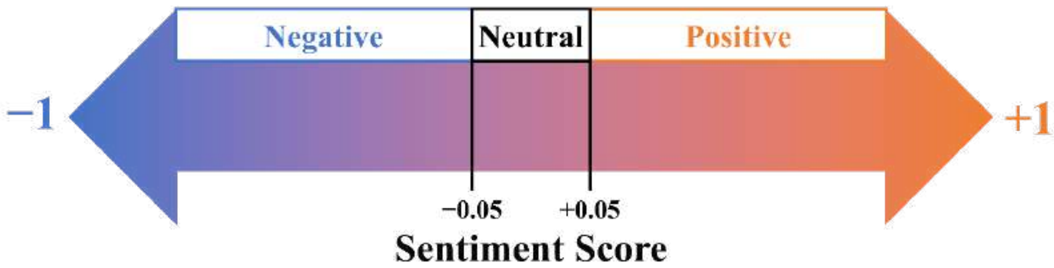


Figure 3: Typical thresholds for classifying comments.

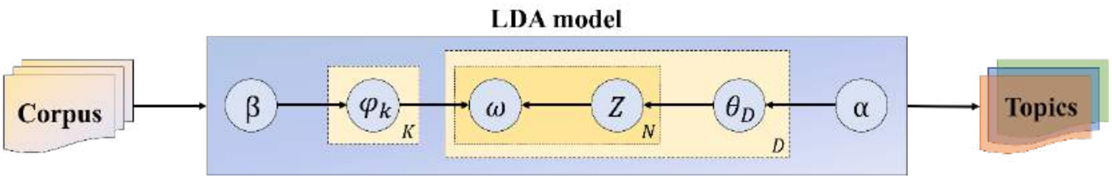


Figure 4: Plate notation of the LDA model.

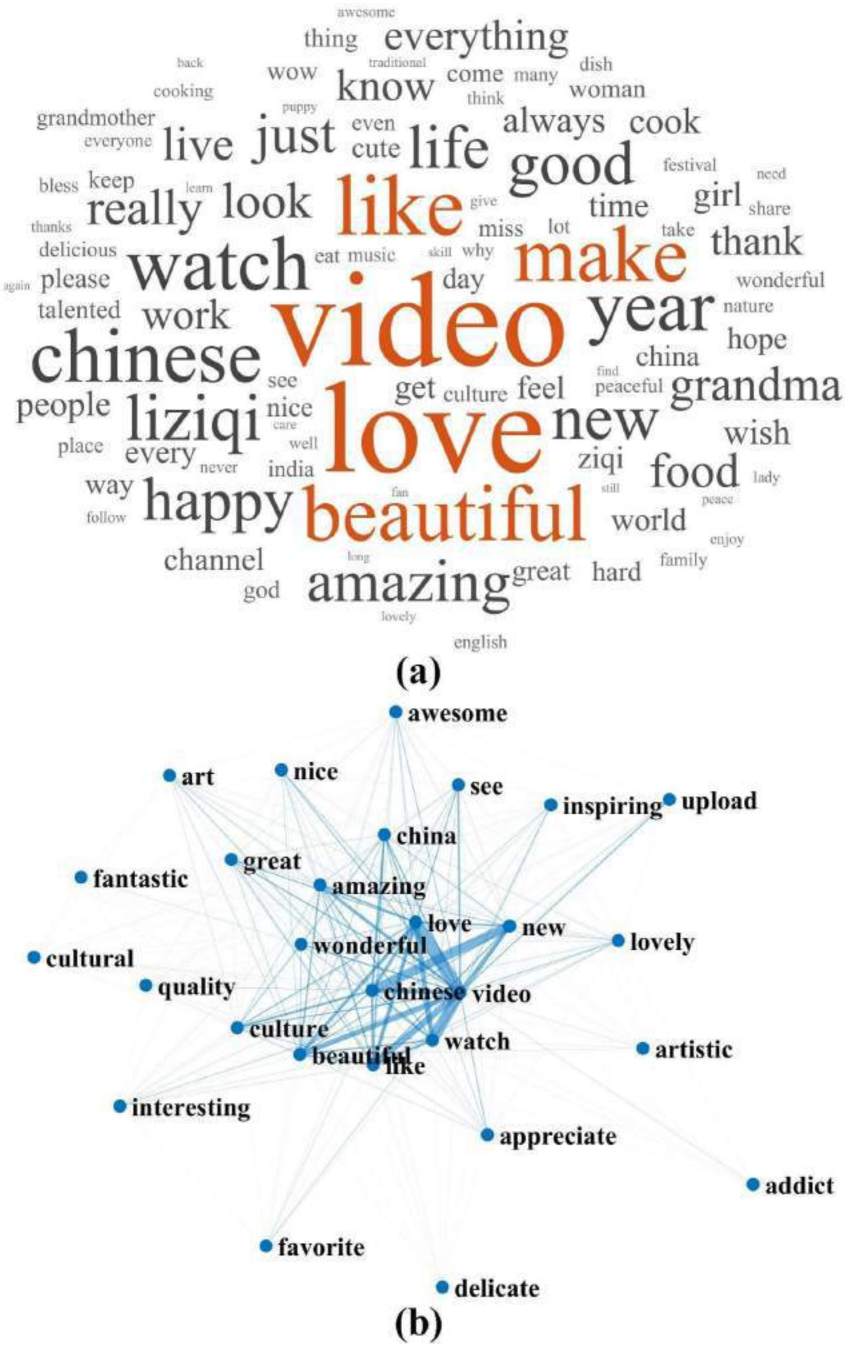


Figure 5: (a) The word cloud of comments and (b) the co-occurrence network of word "video".

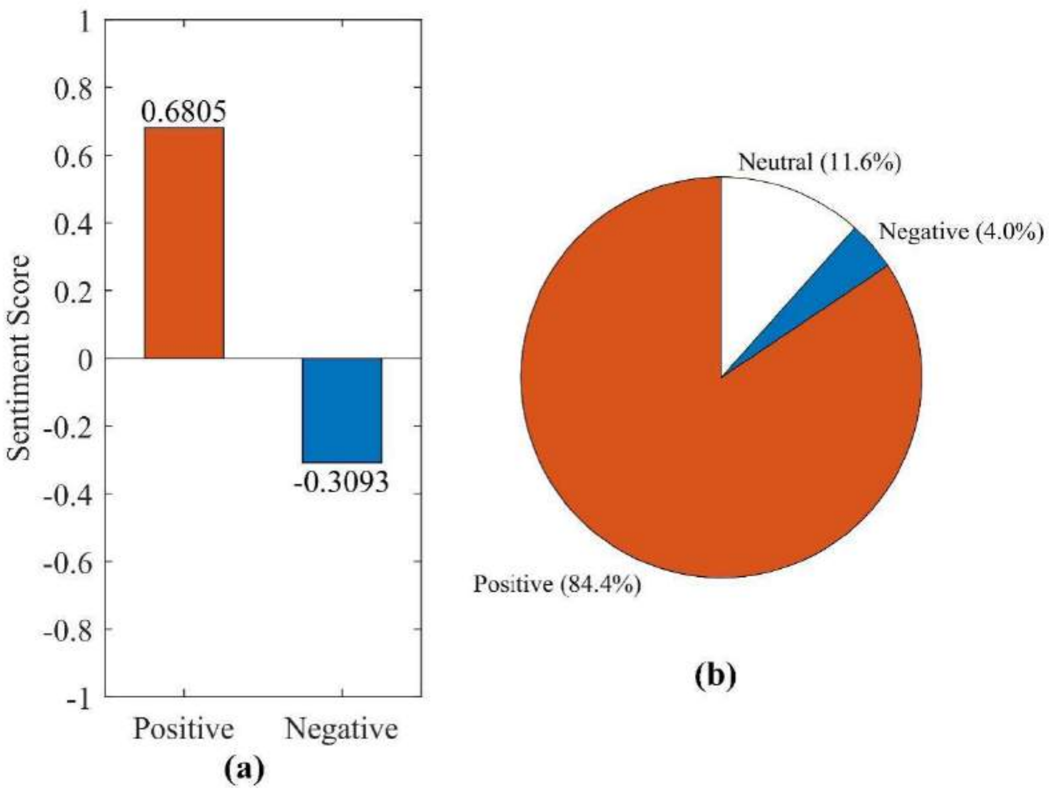
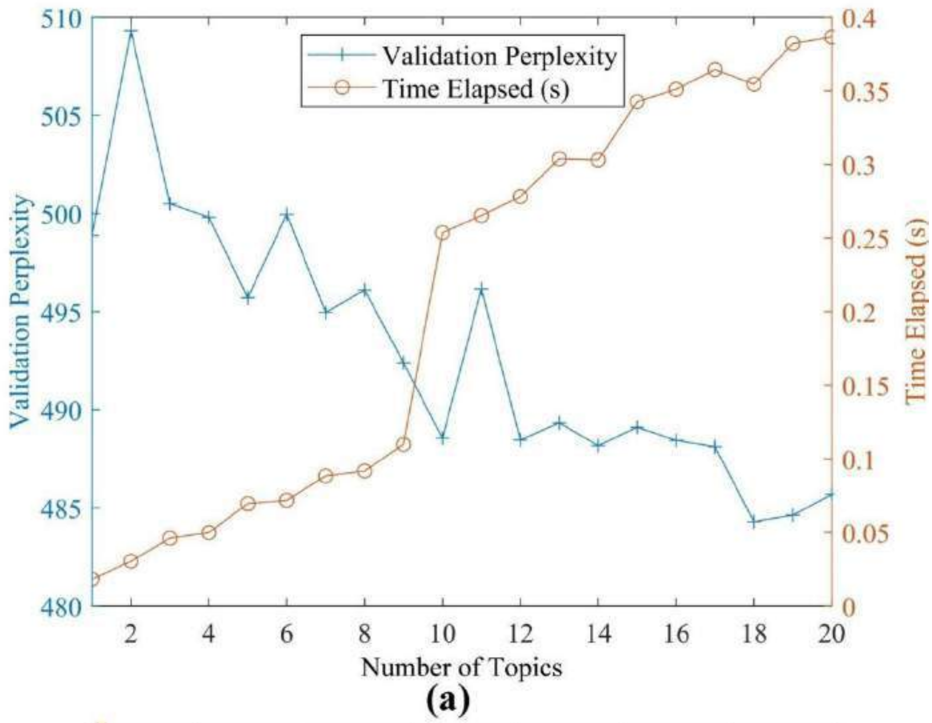


Figure 6: The average scores (a) and the proportion (b) of the positive and negative comments.

Figure 7: The word cloud of positive comments.



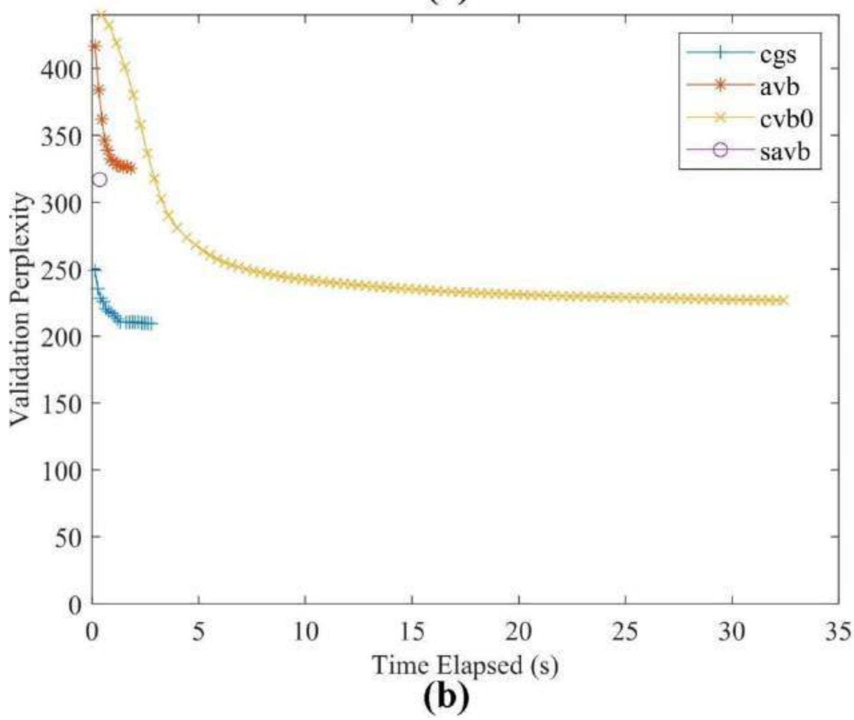


Figure 8: (a) The line graph of validation perplexity and time elapsed with the number of topics. (b) Comparison of various solver of LDA.

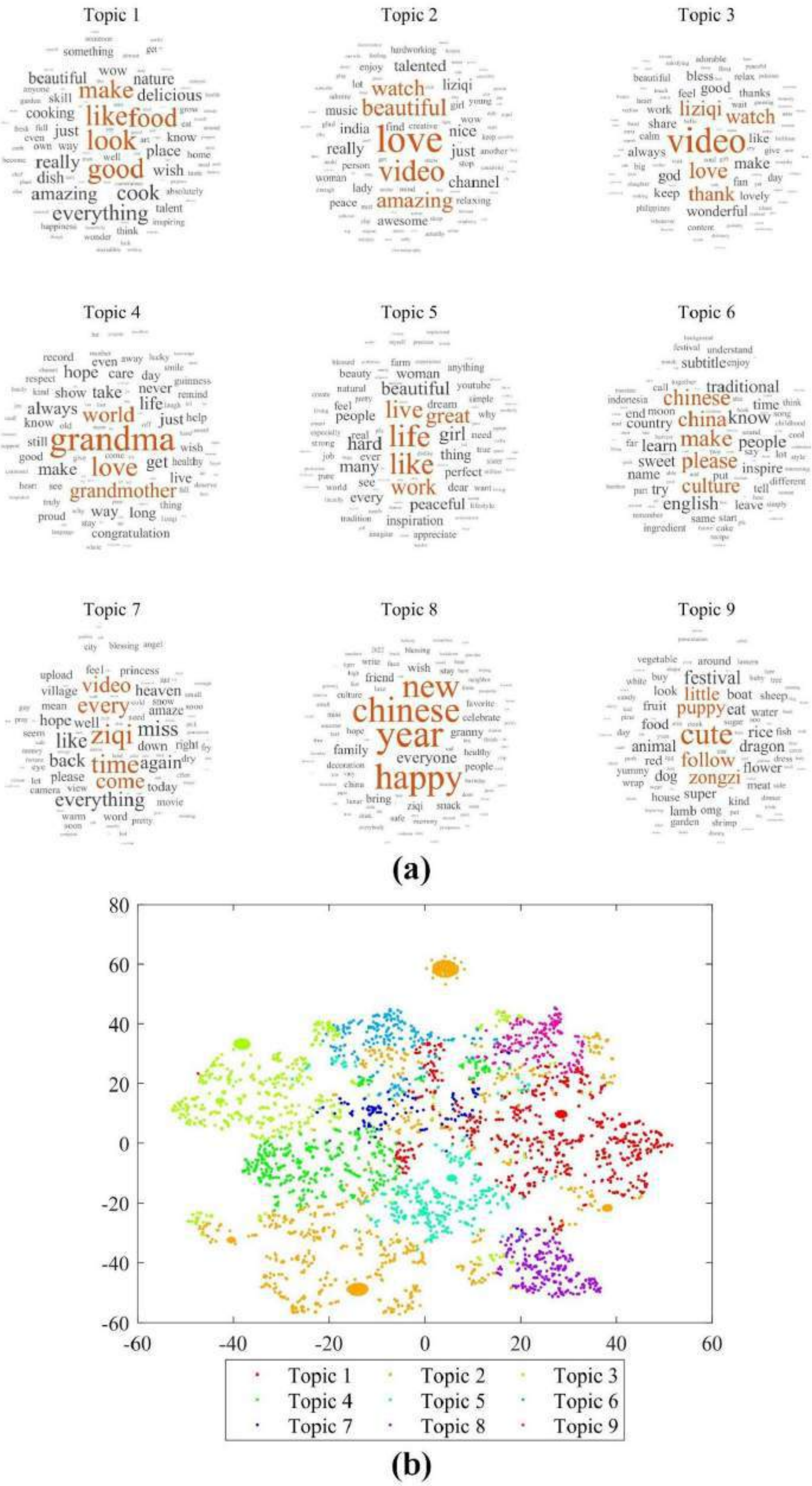


Figure 9: (a) The topic mixtures generated by the t-SNE algorithm. (b) Word clouds of each topic.