Thematic Evolution in Sudden Public Events: A Conversation Analysis Perspective on the COVID-19 Pandemic (Postprint)

Authors: Zhai Shanshan, Wang Zuorong, Chen Huan, Pan Ganghui

Date: 2023-04-01T15:51:27+00:00

Abstract

[Purpose/Significance] The integration of conversation analysis theory offers a novel research perspective for topic evolution studies, refining the analytical granularity of topic evolution. Moreover, applying a more sophisticated topic evolution analysis framework to public health emergencies can enhance the efficiency of public opinion guidance and management by regulatory authorities.

[Method/Process] To address limitations in existing research regarding topic identification methods and criteria for topic evolution judgment, this study combines conversation analysis with topic analysis by incorporating both conversation content and conversational organizational structure into the topic evolution analysis process. Using user-generated content (UGC) related to the "COVID-19 pandemic" as the data source for empirical analysis, we conduct topic evolution analysis based on temporal characteristics and discussion intensity to identify evolution patterns of content at different hierarchical levels from the perspective of topic strength. At the level of topic content analysis, we introduce the association rule mining concept from knowledge discovery to uncover reference relationships among corpus contents, and integrate social network analysis methods to determine critical evolution pathways.

[Results/Conclusion] The findings demonstrate that topic content at different hierarchical levels within the network structure exhibits distinct characteristics and exerts significant influence on topic evolution trends. Effective regulation of content at hierarchically important levels can positively contribute to guiding the direction of public opinion.

Full Text

Preamble

Research on Topic Evolution of Public Emergencies from the Perspective of Conversation Analysis: A Case Study of the "COVID-19 Pandemic"

Zhai Shanshan, Wang Zuorong, Chen Huan, Pan Ganghui

School of Information Management, Central China Normal University, Wuhan 430079

Abstract: [Purpose/Significance] The introduction of conversation analysis theory provides a novel research perspective for topic evolution studies, refining the granularity of topic evolution analysis. Simultaneously, a more comprehensive analytical framework for topic evolution is applied to public emergencies, which enhances the efficiency of public opinion guidance for regulatory authorities. [Method/Process] Addressing existing topic identification methods and evolution judgment criteria, this study integrates conversation analysis with topic analysis by incorporating conversation content and organizational structure into the topic evolution analysis process, using user-generated content (UGC) from the COVID-19 pandemic as the data source for empirical analysis. Through temporality and discussion heat-based topic evolution analysis, evolution patterns at different content levels are identified from the perspective of topic intensity. At the content analysis level, the association rule calculation concept from knowledge discovery is introduced to mine reference relationships between corpus contents, with social network analysis methods employed to determine key evolution paths. [Result/Conclusion] The results demonstrate that topic content at different levels in network structures exhibits significant differences and exerts important influence on topic evolution trends. Effective supervision of content at crucial levels plays a positive role in guiding public opinion.

Keywords: Conversation Analysis; Public Emergency; Topic Identification; Topic Evolution; Association Rules

Classification Number: G206
DOI: 10.13266/j.issn.0252-3116.2022.11.010

1. Introduction

In recent years, public emergencies such as the "COVID-19 pandemic" and the "July 20 Zhengzhou extreme rainstorm" have profoundly impacted social stability and economic development. With the advancement of network technology and the proliferation of mobile smart devices, informal information exchange platforms including Weibo, WeChat, short videos, and online communities have gained widespread popularity, with the public increasingly willing to participate in online discussions about event-related public opinion. Compared with traditional media, online public opinion in the big data era is characterized by relative

openness, rapid dissemination, rich diversity, and inherent tendencies. Unlike standardized news corpora or policy texts, massive volumes of user-generated content (UGC)—produced and updated by users themselves and disseminated through network media—not only comprehensively captures users' deeper content preferences but also reveals themes and evolutionary trends that help accurately understand the current status, development patterns, and dynamic trajectories of public emergencies, providing service references for intelligent monitoring, decision support, public opinion guidance, and personalized recommendations by relevant departments.

Current research on topic evolution has been deepened and expanded across multiple dimensions, yet remains limited by macro-level analysis and single-dimensional measurement indicators. Conversation analysis, as an effective means to reveal patterns in informal information exchange, provides theoretical foundations for exploring sociological patterns of human verbal interaction from information exchange data, and offers specific scenarios for continuous evolution research that uses UGC as the source, themes as representation, and identification of latent associations among topic contents as the objective. Building upon this, this study examines public emergencies by combining conversation analysis with topic analysis, introducing both conversation content and organizational structure into the topic evolution analysis process, and treating COVID-19-related UGC data as an asynchronous conversation process on social media for empirical analysis. On one hand, through temporality and discussion heat-based analysis, the hierarchical structure presented in conversation content is utilized to explore evolution patterns at different topic levels. On the other hand, support and confidence from association rules are employed to determine semantic associations and evolution trends between topic contents, thereby identifying key evolution paths and providing reference suggestions for subsequent public opinion guidance, monitoring, management, and prediction.

2. Related Research

2.1 Research on Topic Identification and Evolution in Public Emergencies

Unlike general public events, public emergencies are characterized by suddenness and high destructiveness. Research on their topic identification and evolution analysis constitutes not only an important component of public opinion studies but also provides valuable insights for future monitoring efforts. Current topic identification methods primarily involve co-word analysis and probabilistic models. The first category, co-word analysis-based methods, utilizes co-occurrence relationships between words in text collections to reflect relationship strength, enabling topic clustering and identification, with applications including potential topic mining and addressing missing self-indexed keywords. The second category, probabilistic model-based methods, centers on machine learning algorithms such as the early LDA (Latent Dirichlet Allocation) model. Subsequent research has expanded along two directions: one focusing on adapting the ba-

sic LDA model to different data sources including news texts and UGC data, and the other proposing LDA optimization models like the SECNN model by refining various stages of topic identification.

Research on topic evolution in public emergencies primarily emphasizes semantic similarity calculations to advance evolution analysis. Regarding topic content, scholars have constructed event semantic graphs using social network analysis tools to propose semantic-based frameworks for discovering public health emergency online opinion topics. Others have categorized and analyzed stakeholder concerns in public emergencies by stage, revealing similarities and differences in topic evolution patterns across different stakeholders. Concerning topic intensity, researchers have collected public policy data, combining LDA with discrete-time methods and multiple indicators like discussion heat to compare evolution across different policy topic types. Some scholars have divided public opinion dissemination into four stages, conducting topic extraction and evolution analysis for each stage and proposing microblog text-based management strategies. Beyond single-dimension approaches, researchers have extracted topics from Zhihu platform data, summarizing user focus areas across time periods and analyzing intensity variation trends. Others have analyzed rumor distribution and quantitative characteristics during COVID-19, combining Maslow's hierarchy of needs to explore underlying formation causes. With social media development, public opinion content has become direct expression of people's genuine intentions, making sentiment identification a focus for government departments and researchers, with most studies combining topic content or intensity analysis with sentiment dictionaries like VADER or emotional lexicons.

Comprehensive review of existing literature reveals three limitations: First, analysis levels remain macroscopic, treating UGC as a monolithic dataset while ignoring network organizational structures inherent in social media or online communities, such as the "main post-reply post-threaded reply" hierarchy and its content differences. Second, topic content association methods are simplistic, predominantly using semantic similarity or distance metrics that lack semantic understanding and directional consideration of associations. Third, measurement indicators lack theoretical grounding, relying on temporality, heat, or text similarity without multidimensional standards for continuous evolution. Conversation analysis addresses these gaps by presenting hierarchical structures in UGC data, facilitating exploration of different-level topic evolution patterns, enabling fine-grained content analysis that extends from content to semantic levels, and enriching criteria for continuous evolution determination.

2.2 Conversation Analysis and Its Applications

Conversation analysis examines naturally occurring conversations in daily life, positing that everyday conversations follow certain orders and patterns. Research from this perspective focuses on two aspects: First, analyzing linguistic expressions during conversations, summarizing discourse roles' conversational styles and strategies through pragmatic and semantic feature analysis, with

applications in classroom interaction, doctor-patient communication, psychological counseling, variety show appreciation, and market transactions. Second, examining how different sources of conversational corpora create variations in organizational structure and interaction patterns among discourse roles. With network technology and new media development, more conversational data has shifted online, with online communities becoming primary sources for informal communication. For instance, Li et al. combined social network and content analysis to analyze WeChat group conversations and construct communication networks among participants. Other scholars have selected online academic communities, integrating LDA models to analyze information interaction types and content topology structures, proposing strategies to promote user interaction. Li Yuelin et al. shifted focus to healthcare websites, analyzing factors affecting interaction efficiency through conversation turn data between doctors and patients to provide theoretical guidance for platform development.

In summary, existing topic evolution research measures correlations based on temporal relationships, topic heat, or similarity, suffering from: (1) macro-level analysis ignoring network structures; (2) simplistic content association methods lacking semantic understanding and directionality; and (3) insufficient theoretical grounding for measurement indicators. Conversation analysis introduces hierarchical structures, enables fine-grained semantic-level processing, and enriches evolution determination standards. Therefore, this study incorporates corpus organizational structure from a conversation analysis perspective, mining evolution patterns across different topic levels through topic intensity analysis, while introducing knowledge discovery association calculations at the content level to determine relationship directionality and identify key evolution paths using social network analysis.

3. Research Design

3.1 Overall Research Framework

The emergence of community-based communication applications has transformed data flow patterns and enriched generated content. This study selected the "Novel Coronavirus" board on Baidu Tieba as the data source for three reasons: (1) From the information publisher perspective, Baidu Tieba's over 1 billion registered users ensures UGC fully reflects the genuine needs of the masses; (2) From the information channel perspective, UGC data sources are more extensive than authoritative data from news or official websites; and (3) From the information presentation perspective, the "main post-reply post-threaded reply" hierarchical structure inherent in UGC data constitutes a necessary factor for topic evolution and represents the primary mode of user interaction in current online communities.

The research first establishes a crawler framework based on data source characteristics, performing basic preprocessing including null value handling, stop word removal, and segmentation. Second, the topic model extracts topics from

each post and clusters them into topic clusters. Posts are then assigned to corresponding topics based on feature word-topic correspondence. During topic evolution analysis, on one hand, temporality and discussion heat-based analysis examines topic cluster heat changes and inter-cluster interaction relationships through co-occurrence patterns. On the other hand, feature word pair associations are analyzed and mapped to the topic cluster level to mine semantic-level key evolution paths. Both aspects reflect topic evolution: the former focuses on trend changes in topic attention, while the latter examines content deepening or extension during event evolution. The specific research framework is shown in Figure 1 [Figure 1: see original paper].

3.2 Research Methods

3.2.1 Topic Model Topic models map high-dimensional word collections to low-dimensional topic spaces for dimensionality reduction and concise representation. Existing models fall into two categories: long-text models like LDA, Dynamic Topic Models, and TOT, and short-text models like BTM and WNTM for texts under 10 words. Considering the post volume and length, this study selected the LDA topic model for extraction. LDA's advantages include: (1) extracting effective information from massive texts and assigning topics to each document; (2) using prior probability distributions to avoid overfitting; and (3) providing feature word probabilities within topics for subsequent analysis. Although optimized LDA models offer better accuracy and performance, they remain limited in application scenarios and cannot fully adapt to this research context.

3.2.2 Topic Clustering Method Topic clustering merges topics with minor feature differences to form coherent topic clusters. To improve extraction precision, this study performs clustering analysis on LDA results to create clusters with small internal and large external differences. Current clustering methods include partitioning-based, hierarchical, grid-based, density-based, and model-based approaches. Given LDA's high-dimensional, large-volume output, this study selected partitioning-based methods, specifically using cosine similarity to measure distances between topic feature vectors. Cosine similarity focuses on directional rather than distance or length differences between vectors, achieving precise clustering. The specific formula is shown in equation (3):

$$cos(\theta) = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2} \times \sqrt{\sum_{i=1}^{n}(y_i)^2}}$$

where α and β are n-dimensional vectors.

3.2.3 Topic Evolution Continuity and Its Determination Previous studies typically used the presence of topic-related discussions in time segments as

the evolution continuity criterion—if discussion count is non-zero, evolution continues. However, content association calculations predominantly use semantic similarity, ignoring temporal features and evolution directionality. Moreover, studies often divide corpora into predetermined time segments while overlooking resource structure characteristics within segments. This study argues that evolution continuity determination requires three elements: topic intensity, content association, and network structure. Topic intensity captures temporal change trends, content association focuses on semantic deepening during evolution, and network structure reflects resource structure throughout the process.

(1) **Topic Intensity**: Evolution refers to the changing trend of topic attention heat over time, depicting the lifecycle of public emergencies. Integrating intensity and content analysis improves continuity judgment accuracy and identifies user focus during events. This study follows existing methods by counting corpus numbers for different topics as intensity measures.

(2) **Topic Content Association**: Calculation accuracy critically affects content classification. Ignoring potential content correlations may exclude relevant user comments or incorrectly include unrelated ones. Association directionality further enriches evolution analysis. This study uses support and confidence from association rules as directionality criteria, enabling multidimensional evolution analysis.

(3) **Network Structure**: As a reflection of corpus resource structure, network structure is essential in conversation analysis. Most existing research treats all UGC data as a single entity while ignoring internal resource structure features. In reality, internal structure enriches evolution patterns and provides new analytical dimensions. The "main post-reply post-threaded reply" structure prevalent in most online communities represents the primary interaction mode.

4. Continuous Evolution Process Analysis of Public Emergencies

4.1 LDA-Based Topic Extraction and Topic Cluster Generation

The LDA model completes word clustering through co-occurrence probabilities and 刻画 s document generation using Dirichlet distribution. This study assumes Baidu Tieba post topics follow a hyperparameter Dirichlet prior distribution as shown in equation (1):

$$Dir(\theta_c|\alpha) = \frac{\Gamma(\sum_{t=1}^{T}\alpha_t)}{\prod_{t=1}^{T}\Gamma(\alpha_t)}\prod_{t=1}^{T}\theta_{ct}^{\alpha_t-1}$$

where _{ct} represents the distribution of post c in topic t. Each generated post topic t and topic terms follow distribution *$\phi t \sim Dir(\beta)$; each post c and topic terms follow distribution $\_c \sim Dir(\alpha)$; for each word item n in post c,*

topic item z{cn} ~ Multinomial(_c) and w_{cn} ~ Multinomial(ϕ{z{cn}}). The LDA likelihood model is shown in equation (2):

$$p(W|\alpha,\beta)=\prod_{c=1}^{C}\int p(\theta_c|\alpha)\prod_{n=1}^{N_c}\sum_{z_c}p(z_{cn}|\theta_c)p(w_{cn}|\phi_{z_{cn}})d\theta_c$$

The accuracy of latent topic number setting critically affects model extraction, but LDA cannot automatically generate optimal numbers. Recent studies propose various methods like perplexity, non-parametric automatic training, and Perplexity-var methods, but these suffer from low efficiency and overfitting. This study uses coherence curves to determine optimal topic numbers through data segmentation, probability calculation, measurement confirmation, and averaging.

To reduce LDA result sparsity, this study employs a cosine similarity-based clustering algorithm. A bag-of-words model is constructed from extracted topic feature words, with each topic represented as multiple vectors. Pairwise cosine similarity calculations between topics determine the most appropriate cluster numbers. For main-level posts, the clustering results are shown in Figure 2 [Figure 2: see original paper]. LDA extraction sometimes yields low differentiation—for instance, "infection symptoms" and "physical symptoms" appear as separate topics but share content similarity compared to other topics like news reports, causing dispersion that affects subsequent intensity calculation accuracy.

4.2 Consistency Score-Based Topic Assignment

After obtaining topic clusters, post-topic assignment requires determining which topic(s) each post belongs to. This study uses consistency scores measuring how well each topic matches post content. When two topics show equal probability, manual judgment assigns the topic. After assignment, topic-cluster correspondence is established, and post counts per cluster are tallied for subsequent intensity analysis.

Considering corpus resource structure, different hierarchical levels correspond to their respective topics. When assigning topics to main posts, candidates are extracted only from main-level topics, independent of other levels. This hierarchical assignment ensures precise topic allocation.

4.3 Temporality and Topic Intensity-Based Evolution Analysis

Public emergency topic evolution analysis examines both topic cluster heat and inter-cluster interaction along temporal dimensions. For heat analysis, this study sums frequencies of different topics within each cluster across time slices to reflect discussion heat, summarizing evolution patterns through trend analysis.

The assignment strategy assumes each post belongs to one topic, but reality often involves multiple topics. To study inter-cluster interactions, each post is

assigned up to three topics, using co-occurrence frequency to reflect interaction strength—higher co-occurrence indicates tighter interaction.

4.4 Association Rule-Based Public Emergency Topic Evolution

Association analysis, a common knowledge discovery method, quantifies how item A's appearance depends on item B. Applied to topic evolution, it reveals dependency relationships between feature words. Treating each post as transaction T composed of multiple feature words, a co-occurrence matrix enables association analysis to mine frequently co-occurring word pairs under support and confidence thresholds.

For example, if a drug treatment text contains both "Drug B" and "Disease A" feature words, it suggests Drug B may treat Disease A. With T posts containing N independent feature words from assigned topics, support S measures the probability of words A and B co-occurring in all posts (equation 4), while confidence Co measures the probability of B appearing given A (equation 5):

$$S(A \to B) = \frac{R}{T}$$

$$Co(A \to B) = P(B|A) = \frac{R}{C_A}$$

where C_A is the number of topics containing feature word A, and R is the co-occurring post count.

Support indicates co-occurrence probability but not association strength, so support serves as a filtering condition to identify strong association word pairs, with confidence then measuring topic association relationships. Feature word relationships are categorized into three types: precedence, succession, and parallel (Table 1).

Since feature words represent topic vectors, inter-topic association strength is measured by averaging support sums across feature words. For Topic 1 containing words A, B, C and Topic 2 containing D, E, F, the association strengths Co(tp1→tp2) and Co(tp2→tp1) are calculated as shown in equations (6) and (7):

$$Co(tp1 \to tp2) = \frac{Co(A \to D) + Co(A \to E) + \cdots + Co(C \to F)}{9}$$

$$Co(tp2 \to tp1) = \frac{Co(D \to A) + Co(D \to B) + \cdots + Co(F \to C)}{9}$$

The association strength matrix facilitates further evolution analysis. After calculating inter-topic confidence, a one-mode matrix is constructed: confidence

between identical topics is set to 0, and for topic pairs, the smaller confidence value is replaced with 0. When Co(tp1→tp2) equals Co(tp2→tp1), indicating equivalence, both values are retained.

5. Empirical Analysis

5.1 Data Acquisition and Preprocessing

This study selected the "Novel Coronavirus" board on Baidu Tieba, collecting data from its establishment date (January 21, 2020) to December 21, 2020, totaling 52,025 posts including 7,298 main posts, 20,049 reply posts, and 24,678 threaded replies. To ensure dataset integrity and cleanliness, the study performed simplified/traditional Chinese conversion, null value removal, and pure string data cleaning, then used Python's jieba package with Harbin Institute of Technology's stopword list for segmentation, storing processed content in separate fields for subsequent candidate topic extraction.

After collection and cleaning, 39,563 valid posts remained: 7,280 main posts, 17,950 reply posts, and 14,333 threaded replies. During LDA training, a coherence curve determined 50 as the optimal topic number, outputting topics with corresponding feature words and their contribution probabilities. Table 2 shows the "epidemic spread" topic extraction results for main posts.

5.2 Topic and Topic Cluster Generation

Using extracted topic feature words, a bag-of-words model was constructed to represent topics as vectors. Pairwise cosine similarity calculations determined optimal cluster numbers. Topics were manually summarized and named based on feature word probability distributions, generating main topic clusters across three levels as shown in Table 3 .

For instance, the "emotional expression" cluster appears across all three levels but with distinct characteristics: main posts emphasize personal needs expression and mourning heroes; reply posts focus on hopeful outlooks for improvement; threaded replies tend to praise national or regional policies.

5.3 Temporality and Topic Intensity-Based Evolution Analysis

Data were divided into monthly time slices, with discussion heat as the vertical axis and time as the horizontal axis to plot evolution trends and interaction patterns across three levels (Figures 3 [Figure 3: see original paper] and 4 [Figure 4: see original paper]).

River graphs visualize evolution and interaction, where river width represents topic cluster heat at specific time points, and river intersections indicate co-occurrence relationships (interaction). For main posts, all ten topic clusters peaked twice around April and June 2020. The "epidemic situation" and "control measures" clusters dominated discussion heat during these periods—April

8 marked Wuhan's lockdown lifting, while June saw China's vaccine Phase III trials launch. Interaction analysis shows an "increase then decrease" pattern: low initial interaction, rising co-occurrence frequency with discussion heat, then weakening.

Reply posts show different patterns: heat evolution displays "increase then decrease" without secondary peaks, while interaction evolution remains relatively weak, diminishing as topics die out. Threaded replies show overall declining heat, with high initial interaction that weakens due to decreasing discussion volume.

Three factors influence evolution: (1) **Landmark events** like Wuhan's reopening break existing trends and re-engage users; (2) **Resource structure** causes lag effects across levels—Baidu Tieba's three-level structure means deeper levels accumulate more participants, causing threaded replies to peak early; (3) **Event characteristics** of public emergencies—suddenness causes abrupt heat fluctuations.

5.4 Association Rule-Based Topic and Topic Cluster Evolution Analysis

5.4.1 Evolution Patterns Across Main-Reply-Threaded Levels Based on the association rule calculation method, 1,448 feature word supports were obtained. Support measures word importance in the corpus, while confidence defines relationship types. Although some parallel relationships were extracted, topic-level analysis primarily shows precedence/succession relationships. NetDraw visualizes inter-level evolution relationships (Figure 5 [Figure 5: see original paper]), where T, P, C denote main, reply, and threaded levels respectively. Arrow direction indicates evolution—for example, "T5 => C30" means threaded topic 30 references main topic 5. Circle size represents degree centrality, indicating importance in evolution.

Reply posts generally show higher degree centrality, with P46, P20, and P25 being representative. For main posts, T38 has relatively high centrality pointing to other levels. For threaded replies, C3 shows high centrality as the endpoint of most connections, reflecting resource structure characteristics: main posts generate new topics, reply posts transition and diverge, while threaded replies summarize and deepen previous content.

5.4.2 Topic Cluster Evolution Patterns Following the feature word association strength calculation method, topic associations are mapped to the cluster level using Neo4j (Figure 6 [Figure 6: see original paper]). Core clusters include: main-level "relevant personnel," "epidemic situation," "case data"; reply-level "symptoms," "epidemic form," "case data"; and threaded-level "control measures," "infection and symptoms."

Three key evolution paths are identified: (1) Main "relevant personnel" → Reply "epidemic situation" → Threaded "control measures"; (2) Main "epidemic

situation" → Reply "symptoms" → Reply "treatment"; (3) Main "case data" → Reply "case data" → Threaded "infection and symptoms."

Six evolution patterns are summarized (Table 4), revealing that resource structure significantly influences cluster evolution. Reply-level posts, with generally high degree centrality, play crucial roles in determining evolution direction, suggesting that regulatory departments should focus on comment content rather than just high-influence user posts.

Conclusion

This study integrates conversation analysis into topic evolution research, proposing that evolution continuity should be determined through topic intensity, network structure, and content association. The analysis combines temporality and intensity to identify three influencing factors: landmark events, network structure, and event characteristics. Using association rules with support and confidence, the study determines directional relationships between topic clusters and employs social network analysis to identify core clusters and key evolution paths, providing references for public opinion guidance.

The findings demonstrate that content at different network levels exhibits significant differences that importantly influence evolution trends. Effective supervision of crucial-level content positively guides public opinion. Future research should further explore the dynamic mechanisms of topic evolution and expand to other public emergency types to validate the framework's generalizability.

References

[1] Zhang Ningxi. Application of big data in network public opinion information work for public emergencies[J]. Modern Intelligence, 2015, 35(6): 38-42.

[2] Ma Chao, Zhai Shanshan, Wang Xiao. Topic and topic cluster evolution analysis of informal information exchange from the perspective of conversation analysis[J]. Library and Information Service, 2021, 65(17): 91-100.

[3] Ritzhaup AD, Stewart M, Smith P, et al. An investigation of distance education in North American research literature using co-word analysis[J]. International review of research in open and distance learning, 2010, 11(1): 37-60.

[4] Ba Zhichao, Li Gang, Zhu Shiwei. Research on keyword selection and semantic measurement methods in co-occurrence analysis[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(2): 197-207.

[5] Wang Hongbin, Wang Jianxiong, Zhang Yafei, et al. Research on topic identification methods for topic-imbalanced news text datasets[J]. Data Analysis and Knowledge Discovery, 2021, 5(3): 109-120.

[6] Li Zhen, Ding Shengchun, Wang Nan. Research on network public opinion viewpoint topic identification[J]. Data Analysis and Knowledge Discovery, 2017,

1(8): 18-30.

[7] Qiu Ningjia, Yang Changgeng, Wang Peng, Ren Tao. Research on text topic identification algorithm based on improved convolutional neural network[J]. Computer Engineering and Applications, 2022, 58(2): 161-168.

[8] Shao Qi, Mou Dongmei, Wang Ping, et al. Semantic-based framework for discovering public health emergency online opinion topics[J]. Data Analysis and Knowledge Discovery, 2020, 4(9): 68-78.

[9] An Lu, Du Tingyao, Li Gang, et al. Topic evolution patterns of stakeholders in public health emergencies on social media[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(4): 394-405.

[10] Li Yue. Research on public policy topic evolution in public health emergencies—taking official WeChat of central cities as an example[J]. Intelligence Magazine, 2020, 39(9): 143-149.

[11] Cao Shujin, Yue Wenyu. Topic mining and evolution of public health emergency microblog public opinion[J]. Journal of Information Resources Management, 2020, 10(6): 28-37.

[12] Zhao Rongying, Chang Ruru, Chen Zhan, et al. Topic evolution research of public health emergencies based on Zhihu platform[J]. Journal of Information Resources Management, 2021, 11(2): 52-59.

[13] Yao Aixin, Ma Jie, Lin Ying, et al. Research on rumor evolution and governance strategies in major public health emergencies[J]. Information Science, 2020, 38(7): 22-29.

[14] Liu Yashu, Zhang Haitao, Xu Hailing, et al. Research on multi-dimensional feature fusion network public opinion event evolution topic graph[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(8): 798-806.

[15] Xu Yuemei, Lu Sining, Cai Lianqiao, et al. News topic evolution analysis combining convolutional neural network and Topic2Vec[J]. Data Analysis and Knowledge Discovery, 2018, 2(9): 31-41.

[16] Zhu Xiaoxia, Song Jiaxin, Meng Jianfang. Network public opinion information analysis based on dynamic topic-sentiment evolution model[J]. Information Science, 2019, 37(7): 72-78.

[17] Zhang Guanglu. Characteristics of classroom discourse interaction from the perspective of deep learning: Based on conversation analysis[J]. Journal of Chinese Educational Society, 2021(1): 79-84.

[18] Wang Yafeng, Yu Guodong. Conversation analysis of patients' expanded answers in doctor-patient communication[J]. Foreign Language Teaching Theory and Practice, 2021, 175(3): 108-118.

[19] Hu Wenzhi, Liao Meizhen. Conversation analysis of "reformulation" in Chinese psychotherapy discourse[J]. Journal of Chongqing University (Social

Science Edition), 2013, 19(4): 92-100.

[20] Shen Ruifei. Conversation structure of talk show "Round Table"[J]. Youth Journalist, 2017, 566(18): 72-73.

[21] Zhang Li. Analysis of sales promoters' conversational strategies[J]. Applied Linguistics, 2007, 63(3): 87-93.

[22] Ba Zhichao, Li Gang, Mao Jin, et al. Analysis of network structure, behavior and evolution of internal information exchange in WeChat groups—based on conversation analysis perspective[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(10): 1009-1021.

[23] Li Gang, Li Xianxin, Ba Zhichao, et al. Research on conversation network structure and member role division in WeChat groups[J]. Modern Intelligence, 2018, 38(7): 3-11.

[24] Lu Heng, Zhang Xiangxian, Zhang Liman, et al. Research on user interaction behavior characteristics in virtual academic communities from conversation analysis perspective[J]. Library and Information Service, 2020, 64(13): 80-90.

[25] Li Yuelin, Zhang Jianwei, Zhang Hua. Spiral vs. linear: Research on user-doctor interaction patterns in online health platforms[J]. Journal of the China Society for Scientific and Technical Information, 2021, 40(1): 88-100.

[26] Gui Xiaoqing, Zhang Jun, Zhang Xiaomin, et al. Review of temporal topic model methods and applications[J]. Data Analysis and Knowledge Discovery, 2019, 3(7): 61-72.

[27] Blei DM, Lafferty JD. Dynamic topic models[C]//Proceedings of the 23rd international conference on machine learning. New York: ACM, 2006: 113-120.

[28] Wang XR, McCallum A. Topics over time: A non-Markov continuous-time model of topical trends[C]//Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2006: 424-433.

[29] Cheng X, Yan X, Lan Y, et al. BTM: Topic modeling over short texts[J]. IEEE transactions on knowledge and data engineering, 2014, 26(12): 2928-2941.

[30] Zuo Y, Zhao J, Xu K. Word network topic model: A simple but general solution for short and imbalanced texts[J]. Knowledge and information systems, 2016, 48(2): 379-398.

[31] Guan Peng, Wang Yuefen, Fu Zhu. Research on topic semantic evolution analysis method based on LDA—taking lithium battery field as an example[J]. Data Analysis and Knowledge Discovery, 2019, 3(7): 61-72.

[32] Syed S, Spruit M. Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation[C]//2017 IEEE international conference on data science and advanced analytics (DSAA). Tokyo: IEEE, 2017: 165-174.

[33] Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures[C]//Proceedings of the eighth ACM international conference on Web search and data mining. New York: ACM, 2015: 399-408.

[34] Cheng Yi, Zhu Weikang, Xu Guowei. Improved ORB matching algorithm based on cosine similarity[J]. Journal of Tianjin Polytechnic University, 2021, 40(1): 60-66.

[35] Shi Zhongzhi. Knowledge Discovery[M]. Beijing: Tsinghua University Press, 2002.

Author Contributions:
Zhai Shanshan: Conceptualization, framework design, revision;
Wang Zuorong: Writing;
Chen Huan: Data processing;
Pan Ganghui: Data acquisition and preprocessing.

Note: Figure translations are in progress. See original paper for figures.