Hypernetwork-Based Weibo Similarity and Its Application in Weibo Public Opinion Topic Discovery: Postprint

Authors: Liang Xiaohe, Tian Ruya, Wu Lei, Zhang Xuefu

Date: 2023-04-01T16:15:55+00:00

Abstract

[Purpose/Significance] Accurate calculation of Weibo similarity can improve the efficiency of Weibo topic mining and holds practical significance for public opinion governance and information security. To address the issues of semantic sparsity and high dimensionality in Weibo text, this paper proposes a hyperedge similarity algorithm that incorporates non-textual features of Weibo.

[Method/Process] This study analyzes the mechanism of Weibo public opinion formation, utilizes a hypernetwork model to represent the process of Weibo public opinion topic formation, and constructs a hyperedge similarity algorithm by calculating the similarity of sub-networks at each layer and their contribution to topic formation.

[Results/Conclusion] Research findings indicate that the proposed similarity method helps improve the topic clustering effectiveness of Weibo public opinion information, particularly for Weibo posts with high similarity in textual expression, where it demonstrates clear topic discriminability.

Full Text

Microblog Similarity Based on Supernetwork and Its Application in Microblog Public Opinion Topic Detection

Liang Xiaohe, Tian Ruya, Wu Lei, Zhang Xuefu

Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081

Abstract: [Purpose/Significance] Accurate calculation of microblog similarity can improve the efficiency of microblog topic mining and has practical significance for public opinion governance and information security. Aiming at the

problems of sparse semantics and high dimensionality in microblog text, this paper proposes a super-edge similarity algorithm that incorporates non-textual features of microblogs. [Method/Process] The mechanism of microblog public opinion formation was analyzed, and the process of microblog public opinion topic formation was represented using a supernetwork model. The super-edge similarity algorithm was constructed by calculating the similarity of each subnetwork layer and the contribution of each subnetwork layer to topic formation. [Result/Conclusion] The study found that the proposed similarity method helps improve the topic clustering effect of microblog public opinion information, especially for microblog information with high similarity in textual expression, demonstrating obvious topic differentiation.

Keywords: super-edge similarity; topic detection; supernetwork; microblog

Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2020.11.009

With the advent of the Web 2.0 era, microblogs have flourished, with ordinary netizens, internet celebrities, news media, and government agencies all using microblogs as primary channels for obtaining information and posting comments. Microblogs represent a User-Generated Content (UGC) model where users freely express their views and opinions on events through various media forms including text, emoticons, images, videos, and live streams. This information achieves rapid point-to-surface dissemination through users' following, forwarding, and commenting relationships, easily forming public opinion events. Mining microblog public opinion information is of great significance for predicting future events, ensuring information security, and monitoring public opinion dynamics.

Microblog text is characterized by short content, weak descriptive capacity, and dispersed topics, which poses challenges for microblog similarity research. The main difficulty in current similarity research for short texts like microblogs is data sparsity. This study proposes a super-edge similarity algorithm that incorporates non-textual features of microblogs based on text analysis, expanding the analytical objects of similarity algorithms, achieving deep-level identification of microblog relationships, and improving the accuracy of microblog public opinion topic detection.

2 Related Research

2.1 Microblog Similarity Analysis

Researchers have proposed various similarity calculation methods for short microblog texts, which can be broadly divided into two categories. The first category improves similarity algorithms based on microblog short-text content features, including methods that augment external corpora. For example, A. Islam et al. constructed the longest common subsequence and calculated text semantic similarity using semantic relationships from external corpora. H. Ma et al. mined frequent item sets with co-occurrence and category co-directional re-

lationships from corpora to construct a feature word similarity matrix for short text feature expansion. Other methods mine short text content features, such as H. Wen et al., who performed part-of-speech tagging and concept annotation on feature words, focusing on calculating text semantic similarity. Huang Xianying et al. constructed text common blocks based on word form and meaning, measuring text similarity according to the number and order of word combinations in common blocks.

The second category introduces microblog non-textual features for similarity calculation using social network analysis methods. Among these, user feature introduction is most extensively studied, focusing mainly on building similarity formulas using users' background information and user relationships such as following/followed relationships and the number of common neighbor friends. As research deepens, some scholars have also begun to incorporate temporal features and sentiment features into similarity calculation formulas.

Although these algorithms have improved microblog short-text similarity calculation efficiency to some extent, they still have limitations. Similarity algorithms improved for text content features, while considering semantic information, generally suffer from low accuracy and high time/space consumption when processing microblog texts with limited information and sparse content. Algorithms introducing non-textual features expand microblog information content to some degree, but existing algorithms remain at the simple network level, mostly introducing only single-layer social networks and lacking organic integration of multiple relationship data in the microblog public opinion formation process. Microblog public opinion formation is a complex process; to more accurately mine microblog public opinion similarity, a more comprehensive and effective method is needed to reveal the formation process. Addressing these issues, this study adopts supernetwork concepts and methods for deeper research on microblog public opinion topic similarity, exploring the internal relationships between multiple public opinion elements and topic formation, and proposes a super-edge similarity algorithm for public opinion topic mining.

2.2 Supernetwork Analysis

The term "supernetwork" was first proposed by Y. Sheffi and P. Denning, with A. Nagurney providing a clear definition: networks that are above and beyond existing networks. Supernetworks demonstrate superiority in nesting, multi-layer, multi-level, and multi-attribute aspects, and have been widely applied in supply chains, transportation, finance, and knowledge management. Current supernetwork research mainly focuses on variational inequalities, hypergraphs, and systems science. Internet text research usually belongs to the latter two categories, with this study falling under systems science.

The multi-layer attribute of supernetworks can well describe interactions between networks. Some scholars have attempted to apply supernetwork methods to microblog public opinion research. For example, Shang Yanchao et al. con-

structed a supernetwork model with topic and user dimensions, while Pan Fang et al. further considered the relationship between network community public opinion dissemination networks and social networks, building a microblog anti-corruption supernetwork model. However, these existing supernetwork models contain too little feature information to reveal the microblog public opinion topic formation process, and lack depth in revealing each subnetwork layer. In previous research, we have constructed a microblog public opinion topic detection supernetwork model containing four subnetwork layers based on public opinion dissemination elements. This study further analyzes the relationships between homogeneous and heterogeneous nodes in the supernetwork model and designs a set of super-edge similarity algorithms (SuperEdgeSimilarity), providing a beneficial supplement to existing supernetwork methods for super-edge analysis.

3 Super-Edge Similarity Algorithm

3.1 Microblog Public Opinion Topic Discovery Supernetwork Model

Real-world social events establish user relationships through microblog platforms' following and forwarding mechanisms, achieving information (keywords, sentiment) sharing, dissemination, and exchange, forming microblog public opinion events. The occurrence process of microblog public opinion is similar to real-world emergencies, requiring clarification of the 5W1H (When, Where, Who, Why, How) six elements. A microblog post is a set of keywords published by a user driven by sentiment and external environmental information, while a public opinion event consists of multiple microblog posts. Thus, entities associated with microblog public opinion formation include microblog users (Who), temporal environment (When) as external driving force, sentiment (How) as internal driving force, and keywords (What). Based on this, we construct a microblog public opinion topic discovery supernetwork model containing four subnetwork layers: "Social Subnetwork," "Temporal Subnetwork," "Sentiment Subnetwork," and "Keyword Subnetwork."

(1) Social Subnetwork A (SocialNetwork) represents forwarding relationships among microblog users participating in topic discussions. Nodes are microblog users, and undirected edges are constructed based on forwarding relationships.

(2) Temporal Subnetwork T (TimingNetwork) represents the temporal stages of microblog public opinion evolution. Following lifecycle theory, we divide microblog public opinion evolution into four stages: "incubation period → outbreak period → persistence period → recovery period." Nodes represent evolution stages of public opinion information, with transformation relationships between adjacent stages.

(3) Sentiment Subnetwork S (SentimentNetwork) represents sentiment information contained during public opinion outbreaks. Different sentiments have transformation relationships. This study's sentiment subnetwork con-

tains three nodes: positive sentiment node, negative sentiment node, and neutral sentiment node.

(4) Keyword Subnetwork K (KeywordNetwork) consists of keywords from microblog texts. Connections between keyword nodes indicate that these keywords appear in the same microblog.

In the microblog public opinion supernetwork model, the four subnetwork layers are connected through super-edges (SuperEdge, SE), where SE = {a_i, t_m, s_n, k_j}, representing that user a_i published keyword k_j under external force at time t_m and internal sentiment drive s_n. A super-edge represents a microblog post, defined here as containing one user information, one sentiment information, one temporal information, and multiple keywords.

3.2 Super-Edge Similarity Algorithm

Microblog texts are characterized by short content, casual expression, and non-standardization, resulting in high-dimensional and sparse text vectors. Traditional similarity algorithms cannot accurately measure similarity between short microblog texts. Based on this, this paper proposes a super-edge similarity algorithm that considers keyword similarity, forwarding behavior relationships, sentiment transformation relationships, and temporal stage transformation relationships contained in different super-edges. In the keyword subnetwork, the more similar the keywords contained in two super-edges, the more likely these super-edges are similar. In the social subnetwork, if two super-edges have forwarding relationships or similar forwarding behaviors, their contained keywords are more likely to be similar, making the super-edges more likely to be similar. In the temporal subnetwork, if two super-edges belong to the same temporal stage or have closer temporal stages, they are more likely to be similar. In the sentiment subnetwork, if two super-edges contain the same and similar sentiment tendencies, they are more likely to be similar.

Assuming the microblog public opinion topic discovery supernetwork model has N super-edges, denoted as SE_i ($1 \le i \le N$), and SE_i and SE_j are two super-edges whose similarity is to be calculated, the super-edge similarity algorithm is as follows:

$$\text{SuperEdge}(SE\_i, SE\_j) = \alpha \times \text{sim}\_\alpha(SE\_i, SE\_j) + \beta \times \text{sim}\_t(SE\_i, SE\_j) + \quad \times \text{sim}\_s(SE\_i, SE\_j) + \gamma \times \text{sim}\_k(SE\_i, SE\_j) \quad \text{(Formula 1)}$$

Where $\text{sim}\_\alpha(SE\_i, SE\_j)$ is the social similarity between super-edges SE_i and SE_j, $\text{sim}\_t(SE\_i, SE\_j)$ is the temporal similarity, $\text{sim}\_s(SE\_i, SE\_j)$ is the sentiment similarity, and $\text{sim}\_k(SE\_i, SE\_j)$ is the keyword similarity. α, β, , and γ are weights for social, temporal, sentiment, and keyword similarities respectively, satisfying $\alpha + \beta + \ + \gamma = 1$, with specific values determined using the analytic hierarchy process.

3.3 Calculation of Super-Edge Attributes in Supernetwork Model

(1) Social Similarity sim_α(SE_i, SE_j). Social similarity is calculated using users' forwarding relationships in the social subnetwork. Assuming the social subnetwork of the microblog public opinion topic discovery supernetwork model contains m nodes, p_i P ($1 \leq i \leq m$) is the set of nodes (users) in the social subnetwork. Similarity between any two nodes in P is calculated based on forwarding relationships. Following the Boolean model concept, forwarding relationships in the social subnetwork can be represented by a matrix C = C_i,j, where C_i,j = 1 if nodes i and j have a forwarding relationship, and 0 otherwise (Formula 2).

Using row_i = (C_i,1, C_i,2, ..., C_i,m) (i = 1, 2, ..., m) to represent the forwarding relationship of super-edge SE_i, the social similarity between super-edges SE_i and SE_j is:

sim_α(SE_i, SE_j) = sim_{ij} = (row_i, row_j) / (||row_i|| ||row_j||) (Formula 3)

Where (row_i, row_j) = Σ C_i,j C_i,j and ||row_i|| = (Σ_i)^(1/2).

(2) Temporal Similarity sim_t(SE_i, SE_j). Due to microblogs' rapid forwarding mechanism, public opinion events attract massive forwarding and discussion in short time periods. Similar microblog content is often concentrated in the same time period. This means that during a topic's outbreak period, people frequently use similar keywords for discussion. As discussions deepen, topics evolve and keywords update, but these updated keywords are closely related to the evolved topic, so updated keywords are also similar. Therefore, keywords generated in the same time period are most likely similar, and the closer the temporal stages, the more likely the keywords are similar.

This study divides microblog public opinion evolution stages (t_i) into four stages: incubation period (t_1), outbreak period (t_2), persistence period (t_3), and recovery period (t_4). After determining the temporal stage type, we can measure the temporal evolution relationships contained in different super-edges, i.e., calculate temporal similarity. If SE_i and SE_j are in the same temporal stage (t_i - t_j = 0), their temporal similarity is 1 (completely similar). If they are in different stages, the closer the stages, the greater the temporal similarity. Following the probability model concept, the calculation formula is:

sim_t(SE_i, SE_j) = sim_{ij} = { 1, if t_i = t_j; e^(-|t_i - t_j|), if t_i $\neq$ t_j } (Formula 4)

Where t_i and t_j are different public opinion evolution stages with values in (1, 2, 3, 4). To differentiate similarity between stages, arithmetic sequences are used to assign values to t_i, considering balance among four similarity values: t_1 = 1, t_2 = 3, t_3 = 5, t_4 = 7.

(3) Sentiment Similarity sim_s(SE_i, SE_j). Since microblog public opinion

contains “social pulse” and “public sentiment,” sentiment information reflects microblog self-media characteristics. Microblogs expressing similar views tend to have consistent sentiment tendencies, and microblogs with consistent sentiment tendencies are more likely to be similar.

Sentiment similarity calculation includes three steps:

First, construct sentiment dictionaries and identify sentiment words in super-edges. By analyzing microblog text sentiment characteristics and expression habits, we summarize key features for microblog sentiment polarity judgment, including sentiment words, emoticons, negation words, and degree words. Extracting these sentiment elements helps accurately calculate super-edge sentiment intensity. This study builds a sentiment analysis method including basic sentiment dictionary, negation dictionary, degree adverb dictionary, and emoticon dictionary, drawing on An Lu et al.'s research. The basic sentiment dictionary uses the Chinese Sentiment Vocabulary Ontology provided by Dalian University of Technology. Users often attach emoticons to express sentiment, and analyzing emoticon sentiment polarity can assist sentiment analysis. This study judged and scored 84 custom emoticons from the microblog platform (see Table 1).

Second, calculate super-edge sentiment intensity. Based on the constructed sentiment dictionaries, identify sentiment feature word polarity and intensity, emoticon polarity and intensity, negation word count, and degree adverb adjustment intensity. Drawing on Tang Xiaobo et al.'s sentiment tuple concept, we use sentiment feature tuples to represent each super-edge's sentiment features: S = {sentiment polarity, intensity; emoticon polarity, intensity; negation word count; degree adverb adjustment intensity}. All tuple elements are optional, meaning some super-edges may have empty sentiment tuples. The sentiment intensity formula is:

sent(i) = (-1)^k × Π wei_p(adv) × Σ s(w_j) (Formula 5)

Where sent(i) is super-edge i's sentiment intensity (0 if the tuple is empty). s(w_i) is sentiment intensity calculated from basic sentiment and emoticon dictionaries, considering three polarities: derogatory words with intensities -1, -3, -5, -7, -9; commendatory words with intensities 1, 3, 5, 7, 9; and neutral words with intensity 0. Σ s(w_j) is the sum of all sentiment words and emoticons in super-edge i, with n being the total count. wei(adv) represents degree adverbs within 3 words before/after sentiment words, with wei_p(adv) being the adjustment intensity of degree adverb p. Π wei_p(ddv) represents the product of all m degree adverb adjustment intensities. k is the count of negation words in super-edge i.

Third, calculate sentiment similarity. From step two, we obtain sentiment intensity. The sign (positive, negative, or zero) indicates three possible sentiment polarities: sent(i) $>$ 0 indicates positive sentiment, sent(i) $<$ 0 indicates negative sentiment, and sent(i) = 0 indicates neutral sentiment. After determining polarity and intensity, sentiment similarity is calculated. Let sent(i) and sent(j)

represent sentiment intensities of two super-edges. Smaller differences indicate greater similarity. The sentiment similarity between super-edges SE_i and SE_j is:

sim_s(i, j) = { e^(-|sent(i) - sent(j)|), if sent(i) ≠ sent(j); I, if sent(i) = sent(j) } (Formula 6)

(4) Keyword Similarity sim_k(SE_i, SE_j). Keyword similarity is the measurement object of traditional similarity algorithms. This study selects the classic Vector Space Model for keyword subnetwork representation, TF-IDF for keyword weight calculation, and cosine similarity for measurement. Mapping SE_1 and SE_2 to n-dimensional vector space as SE_1 = (w_1, w_2, …, w_n) and SE_2 = (w'_1, w'_2, …, w'_n), the keyword similarity is:

sim_k(SE_1, SE_2) = Σ w_i × w'_i / (||w_i|| × ||w'_i||) (Formula 7)

Where w_i = tf_{Ti} × idf_{Ti}, tf_{Ti} is the frequency of keyword Ti in SE_1 (TF value), and idf_{Ti} = log(N/n), with N being the total number of super-edges and n being the total occurrences of keyword Ti across all super-edges.

3.4 Feature Weight Calculation Based on Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) can effectively decompose target problems and analyze them from different levels. This study uses AHP to calculate feature weights for super-edge similarity elements in four steps:

(1) Construct hierarchical structure model. By deeply analyzing microblog public opinion topic formation mechanisms, we decompose it into a two-level hierarchical system (see Figure 2 [Figure 2: see original paper]).

(2) Construct comparison matrix. Analyzing multi-feature elements of microblog public opinion topics, keyword features reveal microblog text content and are the main analysis object for topic discovery, thus receiving higher weight. Sentiment features, as part of text content revelation, are secondary important. Social and temporal features influence topic formation from the side and are weaker than the former two, ranking third equally. The comparison matrix is shown in Table 4 .

(3) Calculate relative weights. Using the eigenvalue and eigenvector formula AW = *λ{max} W, we obtain eigenvector W = [0.083, 0.083, 0.333, 0.500] and maximum eigenvalue* λ{max} = 4.0.

(4) Consistency test. Considering both Consistency Index (CI) and Random Consistency Index (RI), the comparison matrix passes the consistency test. Thus, the eigenvector can serve as weights: W = [0.083, 0.500, 0.333, 0.083], meaning social feature weight $\alpha = 0.083$, temporal feature weight $\beta = 0.083$, sentiment feature weight $\gamma = 0.333$, and keyword feature weight = 0.500.

4 Experiments and Analysis

Since similarity algorithm values are subjective, to demonstrate specific efficiency, we apply the super-edge similarity method to clustering problems, measuring similarity calculation effectiveness through clustering results. The process is shown in Figure 3 [Figure 3: see original paper].

4.1 Experimental Data Description

The dataset was collected from Sina Weibo using the keywords "seedless grapes and contraceptives" from August 27 to September 15, 2016, capturing public opinion information about the rumor "seedless grapes are sprayed with contraceptives," including ID, text content, publication time, publishing user, forwarding account, and forwarded content. After removing obviously irrelevant and duplicate microblogs (same username, ID, and time), removing stop words, URLs, and irrelevant symbols ('#', ' [] '), and preprocessing forwarded content by removing @usernames and bringing forwarded content forward, we obtained 3,889 data entries from 3,600 participants.

Information management researchers manually summarized the main themes of the "seedless grapes are sprayed with contraceptives" event and annotated the dataset accordingly. The event contained eight sub-topics: #1 rumor initiation, #2 rumor deepening, #3 and #4 government and research department refutation, #5 rumor destruction, #6 rumor consequences, #7 falsehood analysis, and #8 calls for accountability. The number of microblogs in each sub-topic is shown in Table 5 .

4.2 Experimental Process

We used Python to call the NLPIR segmentation tool from the Chinese Academy of Sciences. To improve segmentation effectiveness, we added specialized terms from Sogou's agricultural lexicon (e.g., gibberellin, plant hormone, animal technology, haploid breeding) to the user dictionary.

Before temporal feature extraction, we determined specific time nodes for public opinion evolution stages. By analyzing publication quantity changes (see Figure 4 [Figure 4: see original paper]), we divided the case into four stages: incubation period (t_1, August 27 to September 3), outbreak period (t_2, September 4-5), persistence period (t_3, September 6-8), and recovery period (t_4, September 9-14). The incubation period showed slight fluctuations but stable month-on-month growth rate; the outbreak period showed a surge on September 4, peaking on September 5; the persistence period showed declining volume with stable growth rate; the recovery period showed fluctuating volume and growth rate but generally low volume.

We selected the k-means algorithm for clustering analysis. While simple and efficient, k-means requires pre-selecting k value (cluster count), which greatly affects results. To eliminate this impact, we combined the elbow method and

silhouette coefficient, using the elbow method to identify the "elbow point" range and selecting the value with maximum silhouette coefficient within this range as k.

4.3 Experimental Evaluation Metrics Design

We used precision (P), recall (R), and their combined F-measure to evaluate results, where precision tests accuracy, recall tests completeness, and F-value comprehensively evaluates both. Formulas are:

P = a / b (Formula 8)
R = a / c (Formula 9)
F = 2 × P × R / (P + R) (Formula 10)

Where a is correctly clustered microblogs, b is total microblogs identified in the category, and c is actual microblogs of that category in the dataset.

We also introduced Effect Improvement (EI) to evaluate new algorithm improvement:

EI = (F_{new} - F_{old}) / F_{old} × 100% (Formula 11)

Where F_{new} is the improved algorithm's F-value and F_{old} is the comparison algorithm's F-value.

4.4 Experimental Results and Analysis

Using the method in 4.2, we determined the optimal k value for both keyword similarity and super-edge similarity algorithms as 8, matching manual cluster count.

To verify our super-edge similarity algorithm's effectiveness for microblog text topic clustering, we compared it against cosine similarity clustering and the common short-text clustering method FIHC. Clustering results for the eight sub-topics are shown in Figure 5 [Figure 5: see original paper]. Our super-edge similarity method achieved higher F-values for all eight sub-topics than the other two methods, demonstrating superior topic identification. Cosine similarity performed particularly poorly, with F-values below 0.5 for all sub-topics except #4.

Examining sub-topic #5 in detail (see Figure 6 [Figure 6: see original paper]), our algorithm (F = 0.80) significantly outperformed cosine similarity (F = 0.20, 300% improvement) and FIHC (F = 0.56, 42.86% improvement). The poor performance of comparison methods on #5 was due to mixing content from sub-topics #1, #2, #4, and #7. Topics #1 and #2 occurred during the incubation period when people began spreading the rumor, while #4, #5, and #7 occurred during outbreak and persistence periods, all discussing the rumor's falsehood from different angles. These sub-topics had extremely similar textual expressions, making them difficult to distinguish using text-only methods. Our super-edge similarity algorithm correctly identified most #5 microblogs

but confused some #4 and #7 content. Further analysis revealed #4 and #5 involved official professional refutation while #7 involved public common-sense refutation, making their keyword, temporal, sentiment, and social features more consistent, causing identification bias.

Overall, cosine similarity achieved F = 0.44, FIHC achieved F = 0.50, while our super-edge similarity achieved F = 0.74, showing significant improvement rates of 68.18% and 48% respectively. This validates that microblog similarity calculation considering only text is insufficient; social forwarding, temporal stage, and sentiment information are all closely related to public opinion topic formation, and effective mining of these features improves topic detection accuracy.

4.5 Conclusion and Discussion

This study uses supernetwork methods to simulate microblog public opinion topic formation mechanisms, extracting four key features closely related to topic formation: microblog users (Who), temporal stages (When), sentiment features (How), and microblog content (What). We constructed a super-edge similarity algorithm incorporating social, temporal, sentiment, and keyword similarities, applied it to microblog clustering, and evaluated it using F-value and improvement rate on Sina Weibo data about the "seedless grapes sprayed with contraceptives" rumor. Results verify the algorithm's effectiveness, providing stakeholders with accurate topic information for risk control.

Future work will focus on four aspects: (1) Deeply analyzing public opinion topic dissemination characteristics to refine nodes and relationships in the supernetwork model, improving applicability, completeness, and effectiveness. Current sentiment and temporal subnetworks are coarse-grained with node counts far different from social and keyword subnetworks, affecting analysis results. (2) Further optimizing subnetwork similarity calculation methods, especially keyword similarity. Our TF-IDF and cosine similarity methods lack semantic association, reducing discriminative power for microblogs with similar social, temporal, and sentiment features. Future work will incorporate semantic similarity algorithms. (3) Expanding experiments to larger, multi-event datasets to verify generalizability. (4) Using supernetwork models to analyze public opinion topic formation and evolution drivers, revealing dissemination patterns to support governance.

References

[1] Li Gang, Xu Wei, Wang Xinping. Combined model for microblog hot event summary extraction based on event elements[J]. Library and Information Service, 2018, 62(1): 96-105.

[2] Liang Xiaohe, Tian Ruya, Wu Lei, et al. Review of microblog topic detection research methods[J]. Library and Information Service, 2017, 61(17): 41-48.

[3] Liao Haihan, Wang Yuefen, Guan Peng. Topic mining and viewpoint iden-

tification of different disseminators in microblog public opinion dissemination cycle[J]. Library and Information Service, 2018, 62(19): 77-85.

[4] Liu Xiaomin, Wang Hao, Li Xinlei, et al. Comparative study of different feature granularities in microblog short text classification[J]. Information Science, 2018, 36(12): 126-133.

[5] Peng Min, Huang Jiajia, Zhu Jiahui. Massive short text clustering and topic extraction based on frequent itemsets[J]. Journal of Computer Research and Development, 2015, 52(9): 1941-1953.

[6] Cui Jindong, Sun Yaoyao, Wang Xin, et al. Research on microblog recommendation method based on Folksonomy and ontology fusion[J]. Information Science, 2015, 33(10): 27-31.

[7] ISLAM A, INKPEN D. Semantic text similarity using corpus-based word similarity and string similarity[J]. ACM transactions on knowledge discovery from data, 2008, 2(2): 1-235.

[8] MA H, DI L, ZENG X, et al. Short text feature extension based on improved frequent termsets[M]. New York: Springer International Publishing, 2016: 169-178.

[9] WEN H, WANG Z, WANG H, et al. Short text understanding through lexical-semantic analysis[C]//Proceedings of the 31st IEEE international conference on data engineering. Seoul: IEEE Computer Society, 2015: 495-506.

[10] Huang Xianying, Chen Hongyang, Liu Yingtao. Short text similarity research and its application in microblog topic detection[J]. Computer Engineering and Design, 2015, 36(11): 3128-3133.

[11] Li Ji, Huang Wei, Guo Sulin. A microblog content recommendation method based on similarity and trust fusion[J]. Library and Information Service, 2018, 62(11): 112-119.

[12] KRISHNAMURTHY B, GILL P, ARLITT M. A few chirps about twitter[C]//WOSP'08 Proceedings of the first workshop on online social networks. Seattle: Association for Computing Machinery, 2008: 19-24.

[13] Lu Peng, Zhang Shanshan, Gao Qingyi. Research on point-weight limited BBV model based on common neighbors[J]. Computer Science, 2014, 41(4): 49-52.

[14] Yan Guanghui, Zhao Hongyun, Ren Yajin, et al. Research on microblog hot topic detection algorithm based on temporal characteristics[J]. Application Research of Computers, 2014, 31(1): 43-46.

[15] Wu Fangzhao, Wang Bingkun, Huang Yongfeng. Microblog data sentiment classification based on text and social context[J]. Journal of Tsinghua University (Science and Technology), 2014, 54(10): 1373-1376, 1383.

[16] SHEFFI Y. Urban transportation networks: equilibrium analysis with mathematical programming methods[M]. Englewood Cliffs: Prentice-Hall, 1985.

[17] DENNING PJ. The science of computing: supernetworks[J]. American scientist, 1985, 73(3): 127-1269.

[18] NAGURNEY A, DONG J. Supernetworks: decision-making for the information age[M]. Cheltenham: Edward Elgar Publishing, 2002.

[19] Ma Jun, Dong Qiong, Yang Deli. Supply chain supernetwork equilibrium model for time-sensitive products[J]. Journal of System & Management, 2015, 24(4): 610-616.

[20] BRICE OL, COMINETTI R, CORTES CE, et al. An integrated behavioral model of land use and transport system: a hyper-network equilibrium approach[J]. Networks and spatial economics, 2008, 8(2/3): 201-224.

[21] Zhu Li, Du Yaqing. Supernetwork model for emergency resource coordination in urban agglomerations[J]. Mathematics in Practice and Theory, 2015, 45(16): 27-37.

[22] Cao Xia, Liu Guowei. Supernetwork analysis of industry-university-research cooperation innovation based on social capital[J]. Journal of Intelligence, 2014, 33(2): 120-127.

[23] Tian Ruya, Sun Wei, Wu Lei, et al. Hypergraph-based analysis of knowledge collaboration characteristics in library and information science[J]. Information Studies: Theory & Application, 2016, 39(10): 25-30.

[24] Shang Yanchao, Wang Hengshan, Wang Yanling. Supernetwork model for information dissemination on microblogs[J]. Technology and Innovation Management, 2012, 33(2): 175-179.

[25] Pan Fang, Bao Yuting. Research on microblog anti-corruption public opinion based on supernetwork[J]. Journal of Intelligence, 2014, 33(8): 173-177.

[26] Liang Xiaohe, Tian Ruya, Wu Lei. Microblog public opinion topic mining method based on supernetwork[J]. Information Studies: Theory & Application, 2017, 40(10): 100-105.

[27] Ma Ning, Liu Yijun. Multi-agent modeling of public opinion evolution based on supernetwork[J]. Journal of System & Management, 2015, 24(6): 785-794, 805.

[28] Zhang Li. Research on a Chinese text clustering method[D]. Harbin: Harbin Engineering University, 2009.

[29] DAKKA W, GRAVANO L, IPIROTIS P. Answering general time-sensitive queries[J]. IEEE transactions on knowledge and data engineering, 2012, 24(2): 220-350.

[30] EFRON M, LIN J, HE J, et al. Temporal feedback for tweet search with non-parametric density estimation[C]//SIGIR'14: Proceedings of the 37th inter-

national ACM SIGIR conference on research and development in information retrieval. New York: ACM Press, 2014: 33-42.

[31] LIN J, EFRON M. Temporal relevance profiles for tweet search[C]//SIGIR'13: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. Dublin: ACM Press, 2013. doi:10.1.1.420.611.

[32] SALTON G, WONG A, YANG CS. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.

[33] Li Mingde, Meng Shengjun, Zhang Hongbang. Research on microblog public opinion dissemination pattern—Based on process analysis[J]. Management Review, 2013, 25(4): 115-124, 157.

[34] An Lu, Wu Lin. Analysis of emergency event microblog public opinion evolution integrating topic and sentiment features[J]. Library and Information Service, 2017, 61(15): 120-129.

[35] Xu Linhong, Lin Hongfei, Pan Yu, et al. Construction of sentiment vocabulary ontology[J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(2): 180-185.

[36] Tang Xiaobo, Lan Yuting. Sentiment analysis of microblog product reviews based on feature ontology[J]. Library and Information Service, 2016, 60(16): 121-127, 136.

[37] Tang Xiaobo, Fang Xiaoke. Research on microblog topic retrieval model based on text clustering and LDA fusion[J]. Information Studies: Theory & Application, 2013, 36(8): 85-90.

[38] Sun Changnian. Research and implementation of text similarity calculation based on topic model[D]. Hefei: Anhui University, 2012.

[39] FAN FJ, GOODMAN ED, LIU ZJ. AHP (analytic hierarchy process) and computer analysis software used in tourism safety[J]. Journal of software, 2013, 8(12): 3114.

[40] MARQUES JP, WU YF, et al. Pattern recognition: concepts, methods and applications[M]. Beijing: Tsinghua University Press, 2002: 67-72.

[41] Wang Jianren, Ma Xin, Duan Ganglong. Improved K-means clustering k-value selection method[J]. Computer Engineering and Applications, 2019, 55(8): 1-8.

[42] CHEN CL, TSENG FSC, LIANG T. Mining fuzzy frequent itemsets for hierarchical document clustering[J]. Information processing & management, 2010, 46(2): 193-211.

Author Contributions

Liang Xiaohe: Proposed the super-edge similarity calculation method, designed the research framework, and drafted the manuscript.
Tian Ruya: Designed subnetwork similarity calculation methods and revised the research.
Wu Lei: Responsible for data processing and cleaning.
Zhang Xuefu: Provided research ideas, designed the overall plan, and revised the final manuscript.

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