Measuring the Evolutionary Hierarchy of Online Public Opinion Viewpoint Clusters: An Empirical Study Postprint

Authors: Yan Lu, Yang Gang, Zhao Jiangyuan

Date: 2023-04-01T16:02:59+00:00

Abstract

[目的/意义] This study proposes and constructs an evolution level for viewpoint clusters in online public opinion to characterize the evolutionary degree of collective opinion states among online public opinion audiences as time and situations evolve, which holds significant theoretical and practical implications for the guidance, control, and precise steering of online public opinion. [方法/过程] A measurement model for the evolution level of viewpoint clusters in online public opinion is constructed based on LDA and CNN neural networks, with the "Zhai Tianlin CNKI Incident" serving as an experimental case to validate the effectiveness of this indicator. [结果/结论] The evolution level of viewpoint clusters in online public opinion can effectively reflect the evolution of collective opinion states in online hotspot events; while displaying attribute values across three dimensions, it also captures the evolutionary degree of viewpoint clusters relative to their previous time-node state. The measurement results accurately demonstrate the various evolution peaks of event viewpoints, providing a novel guiding direction for relevant departments to conduct targeted guidance of collective opinions in online public opinion.

Full Text

Measurement and Empirical Study on the Evolution Level of Internet Public Opinion Clusters

Yan Lu, Yang Gang, Zhao Jiangyuan

School of Management, Jilin University, Changchun 130022

Abstract: [Purpose/Significance] This study proposes and constructs an evolution level framework for Internet public opinion clusters to describe the evolutionary degree of group opinion states among Internet public opinion audiences over time and with changing events, which holds important theoretical and practical significance for Internet public opinion guidance and precise steering. [Method/Process] Based on LDA and CNN neural networks, we construct a measurement model for the evolution level of Internet public opinion clusters, and use the "Zhai Tianlin CNKI Event" as a case study to verify the effectiveness of this evolution level indicator. [Result/Conclusion] The evolution level of Internet public opinion clusters effectively reflects the evolution of group opinion states in Internet hotspot events. While displaying attribute values across three dimensions, it also captures the evolutionary degree of opinion clusters compared to their previous state. The proposed evolution level measurement accurately identifies various evolution peaks in event opinions, providing a new guiding direction for relevant departments to conduct targeted guidance of group opinions in Internet public opinion.

Keywords: Internet public opinion; opinion cluster; opinion evolution; evolution level; measurement model

Classification Number: G203
DOI: 10.13266/j.issn.0252-3116.2021.23.012

With today's rapid innovation in Internet technology, the vigorous development of new media has driven the high-speed growth of Internet users, forming a complex and ever-changing ecological environment for Internet public opinion. Individual opinions in Internet public opinion are formed based on individual standpoints, concepts, and cognitive levels, exhibiting strong subjectivity. During the formation of individual opinions, influenced by other viewpoints, interactions and mutual influences among individuals facilitate the emergence of several mainstream opinions regarding certain Internet hotspot events. This abstract concept of mainstream group opinions can be considered as a cluster structure formed by multiple individual opinions with identical or similar positions—namely, opinion clusters. The evolution of opinion clusters serves as an indicator of their emotional tendencies and group cohesion, where the level of evolution reflects the degree of change in opinion clusters. This indicator effectively captures the evolutionary state of Internet public opinion across temporal and event development dimensions, including changes in emotional sentiment, influence capacity, and opinion volume of opinion clusters, providing reference basis for targeted and precise guidance of public opinion trends and control of public opinion development by Internet public opinion management entities. This paper clarifies the concept of opinion clusters and their evolutionary attributes at the theoretical level, constructs an evolution level indicator and measurement model for Internet public opinion clusters at the practical level, and uses the "Zhai Tianlin CNKI Event" as a research case to verify the universality and accuracy of the opinion cluster evolution level and measurement model.

Related Research

2.1 Internet Public Opinion Mining

Opinion mining primarily extracts the emotional tendency and semantic information of viewpoints. Some researchers have used K-nearest neighbor classifiers and naive Bayes classifiers combined with biosignal detection methods to obtain user opinions. Other scholars have utilized word2vec to vectorize opinion texts before inputting them into Support Vector Machines (SVM) and Long Short-Term Memory (LSTM) models to train opinion mining models capable of recognizing new viewpoints. Deep learning-based opinion mining methods currently represent the mainstream approach in research. One study constructed a multi-layer embedded CNN model with a Dropout mechanism to enhance local semantic feature recognition capability, effectively mining texts containing certain viewpoints. Another researcher combined word vectors with multi-scale convolutional neural networks to classify the emotional tendencies of Internet public opinion viewpoints, fusing three-scale convolution units into one-dimensional vectors, demonstrating good performance in practical opinion mining tasks.

In summary, research on Internet public opinion mining primarily utilizes various classifiers to mine group opinions in Internet public opinion. However, relatively few studies analyze how group opinion attributes change over time and with event development. Combining semantic perspectives, emotional types, and dissemination capacity to comprehensively mine Internet public opinion viewpoints facilitates multi-angle analysis of current Internet public opinion evolution states. Research in this field can provide practical guidance for the guidance and control of Internet public opinion evolution.

2.2 Internet Public Opinion Evolution

Current research on Internet public opinion evolution mainly demonstrates opinion evolution from a propagation perspective. Some researchers have constructed multi-dimensional Internet public opinion topic graphs from the perspectives of user information and text content, combining topic graphs for feature evolution and visual analysis of Internet public opinion to represent opinion evolution characteristics. Scholars have proposed and constructed an Internet public opinion evolution index, using text clustering results and clustering validity as bases to propose discrimination standards for Internet public opinion evolution, describing the process of mainstream opinion evolution and new opinion generation during Internet public opinion development. One study constructed a continuous opinion interaction model between adjacent nodes based on social network models to represent opinion evolution states during propagation. Current research also analyzes the temporal evolution characteristics of Internet public opinion viewpoint propagation based on infectious disease models to identify outbreak nodes and achieve unsupervised early warning.

In summary, current domestic and international scholars' research on Internet public opinion evolution primarily focuses on using social network models or propagation models to represent the evolutionary state of opinion clusters, without quantitatively representing the numerical characteristics of mainstream opinions among broad Internet public opinion audiences in hotspot events through data. The quantitative direction of Internet public opinion cluster evolution remains a research gap.

Measurement Model for Internet Public Opinion Cluster Evolution Level

3.1 Measurement Process for Opinion Cluster Evolution Level

The implementation 思路 for measuring opinion cluster evolution level is as follows: First, divide Internet public opinion data into opinion clusters and quantify the emotional intensity, cluster scale, and propagation range of each opinion cluster as evolutionary attributes. Then, construct opinion cluster evolution chains based on semantic similarity of keywords in opinion clusters to represent the continuous state of opinion clusters on a certain topic over time. Finally, calculate the evolution level of opinion clusters on the evolution chain by comparing changes in evolutionary attributes at various time nodes.

The key to measuring opinion cluster evolution level lies in opinion cluster division. First, it is necessary to determine the emotional classification of Internet public opinion viewpoints, then divide opinion clusters for positive and negative emotional viewpoints. This approach solves the problem that after LDA divides opinion clusters, they contain nearly equal numbers of positive and negative emotional viewpoints and cannot well reflect the emotional tendencies of opinion clusters. Therefore, the opinion cluster division process consists of three steps: First, use a method combining convolutional neural networks and sentiment dictionaries to classify the emotional sentiment of Internet public opinion data and calculate emotional intensity. Then, slice positive and negative emotional corpora using fixed time windows (daily slices in this study). Finally, use the perplexity index to determine the optimal division number K for positive and negative opinion clusters in each time slice, and use the LDA model to divide opinion clusters for positive and negative emotional corpora separately.

The construction process of the opinion cluster evolution level measurement model is: Determine opinion polarity: divide positive and negative emotional viewpoints and calculate emotional intensity. Divide opinion clusters: use the LDA model to determine the optimal number of opinion cluster categories based on perplexity indicators, dividing positive and negative opinion clusters. Quantify opinion cluster evolution attributes: quantify three evolution state attributes—propagation range, emotional intensity, and cluster scale. Construct opinion cluster evolution chains: determine evolution relationship chains between opinion clusters at previous and current time nodes based on semantic similarity of keywords in opinion clusters. Calculate opinion cluster evolution level: compare changes in evolution attributes of opinion clusters on the same evolution chain across time nodes to obtain the evolution level.

3.2 Emotional Intensity Measurement for Opinion Clusters

This study uses Convolutional Neural Networks (CNN) to classify the emotional sentiment of Internet public opinion data, combined with sentiment dictionaries to calculate the emotional intensity of each viewpoint. CNN can accurately classify the emotional sentiment of Internet public opinion viewpoints, while sentiment dictionary methods can represent emotional intensity numerically. Combining both yields viewpoint emotional intensity. The CNN emotional classification method is similar to CNN image processing: convolutional layers extract features, pooling layers reduce neuron quantity, and fully connected layers serve as classifiers outputting probabilities. First, a dictionary of words and frequencies is constructed based on word frequency, with higher-frequency words ranked higher. The top 10,000 words are retained to accelerate training speed. At this point, Chinese vocabulary has been converted into data types readable by the model. Then, convolutional layers read sentences using moving steps of three, four, and five words as convolution kernels, which can perfectly present sentence semantic connotations. Cross-entropy function serves as the loss function to calculate training losses. Finally, feature vectors extracted by multiple convolution kernels are expanded and connected together, with fully connected layers added to output categories. CNN model parameters are adjusted based on loss function and accuracy changes. The CNN model parameters are shown in Table 1 .

From Figure 1 [Figure 1: see original paper], it can be seen that after 30,000 iterations, the loss function of the CNN model using Table 1 parameters has converged, and classification accuracy reaches a high level. After confirming the emotional intensity of Weibo public opinion viewpoints, the BosonNLP sentiment dictionary is used, combined with stop words, negation words, and degree adverb dictionaries to calculate emotional intensity of Weibo public opinion viewpoint text content. Python's jieba segmentation package segments the test set corpus, removes stop words, and matches segmented words with the sentiment dictionary to finally obtain sentiment values containing scores.

3.3 Opinion Cluster Scale Measurement

For opinion cluster scale measurement, this study selects the LDA model as the opinion cluster division model. As an unsupervised model, although its short text classification effect is inferior to long texts, it can better handle the problem that large-scale Internet public opinion information cannot be manually annotated. It provides classification results for each document in the document set in the form of probability distributions. The theme of Internet public opinion opinion clusters can be considered as the common opinion reaction within the cluster. The idea of dividing Internet public opinion opinion clusters based on LDA assumes that each word in the Internet public opinion ontology describes a certain opinion cluster theme with a certain probability, while opinion cluster themes are described by probability distributions of a set of feature words. Different themes may contain the same vocabulary but with different cluster probabilities. Therefore, the high-frequency word set of each opinion cluster theme can be regarded as the embodiment of the topic's potential semantics. After obtaining the final opinion cluster theme probability distribution, the viewpoint ontology-cluster theme matrix and cluster theme-word probability matrix are output after iteration completion. Words are inserted into the cluster theme list according to their probability of belonging to the theme to constitute the feature word set representing the theme.

In the opinion cluster division process, the optimal number of opinion clusters in each time unit (K value) affects the accuracy of opinion cluster division. One problem with using the LDA model for opinion cluster division is determining the optimal classification number for the corpus. Therefore, using the LDA topic model for classification requires determining the number of categories. This study adopts the commonly used perplexity index to determine the optimal number of opinion clusters in each time period to reflect classification credibility. Perplexity measures how well a probability distribution or probability model predicts samples. By comparing the 优劣 of two probability distributions or models in predicting samples, the optimal model is selected. Perplexity has excellent effectiveness in evaluating clustering algorithm performance, so the optimal number of LDA opinion clusters can be selected by comparing perplexity. Perplexity calculates from three methods: probability distribution perplexity, probability model perplexity, and word segmentation perplexity. For natural language processing models, Perplexity per word (word segmentation perplexity) is typically selected. On test set Dt, the perplexity expression is shown in Formula (1):

$$Perplexity(D_t) = exp\left(-\frac{\sum_{d=1}^{D} \log p(w_d|M)}{\sum_{d=1}^{D} N_d}\right) \tag{1}$$

Where M refers to trained model parameters (theta and phi in the LDA model, i.e., viewpoint-opinion cluster matrix and viewpoint-feature word matrix), Nd is the number of words in viewpoint d, and wd is the word vector form of viewpoint d in test set Dt. From Formula (1), the numerator term in brackets represents the probability that viewpoint d in the test set belongs to the model. Larger values indicate smaller perplexity and higher model performance.

When solving for the optimal number of opinion clusters, automatic identification of the optimal number is required. However, curves composed of discrete points cannot directly extract the optimal cluster number. Therefore, this study uses curve fitting to transform the discrete point function into a continuous function and calculates its second derivative. Using the property that a zero second derivative indicates an inflection point, the first lowest point of the perplexity fitting curve is taken as the optimal number for Internet public opinion opinion cluster scale division. The perplexity curve and fitting curve for Internet public opinion opinion clusters are shown in Figure 2 [Figure 2: see original paper].

After determining the optimal number of Internet public opinion opinion clusters in each time slice, LDA is used to divide opinion clusters for positive and negative emotional viewpoints in each time period, simultaneously obtaining opinion cluster scale attribute data.

3.4 Opinion Cluster Propagation Range Measurement

The calculation 思路 for propagation range is: separately calculate the user propagation capacity and viewpoint propagation capacity of opinion clusters, then combine them to obtain the opinion cluster propagation range. User propagation capacity is represented by the total account interaction attributes of Internet public opinion users—i.e., the sum of their following count, follower count, and Weibo post count. Each indicator value combined with its weight and the user's verification coefficient yields user propagation capacity. After calculating individual user propagation capacities within an opinion cluster, the opinion cluster's user propagation capacity is obtained. Therefore, the calculation of opinion cluster user propagation capacity Cuser is shown in Formula (2):

$$C_{user} = index_{rz} \sum_{i=1}^{N} (n_{gz} w_{gz} + n_{fs} w_{fs} + n_{wb} w_{wb}) \tag{2}$$

Where indexrz is the verification coefficient of users within the opinion cluster, N is the number of users in the opinion cluster, ngz, nfs, nbw represent the following count, follower count, and blog post count of users in the opinion cluster respectively, and wgz, wfs, wbw are the weights for each indicator.

Viewpoint propagation capacity is calculated similarly to user propagation capacity, represented by the total interaction attributes of Internet public opinion viewpoints—i.e., the sum of likes, reposts, and comments received. Combined with indicator weights and viewpoint data type coefficients, the viewpoint propagation capacity is obtained. The opinion cluster's viewpoint propagation capacity is the accumulation of individual viewpoints. Therefore, the calculation of Internet public opinion opinion cluster's viewpoint propagation capacity Copinion is shown in Formula (3):

$$C_{opinion} = index_{type} \sum_{j=1}^{M} (n_{dz} w_{dz} + n_{zf} w_{zf} + n_{pl} w_{pl}) \tag{3}$$

Where indextype is the data type bonus for viewpoints within the opinion cluster, M is the number of viewpoints in the opinion cluster, npl, ndz, nzf represent the comment count, like count, and repost count of viewpoints in the opinion cluster respectively, and wpl, wdz, wzf are the weights for each indicator.

The propagation capacity of Internet public opinion opinion clusters is obtained by adding user information volume and viewpoint information volume. The calculation of Internet public opinion opinion cluster information volume Ccluster is shown in Formula (4):

$$C_{cluster} = C_{user} + C_{opinion} \tag{4}$$

3.5 Construction of Opinion Cluster Evolution Chains

According to opinion dynamics theory, individual opinion formation is influenced by other individual opinions, and emotional tendencies lean toward viewpoints that individuals identify with. Opinion cluster evolution chains can be considered as: on the temporal dimension, opinion clusters at subsequent time nodes are influenced by previous node opinion clusters, exhibiting high semantic consistency. Therefore, when two opinion clusters at different time points have the highest semantic similarity, they can be considered to have an evolutionary relationship across the temporal dimension, with the later time node's opinion cluster being the continuation of the earlier time node's opinion cluster. The construction process for opinion cluster evolution chains is shown in Figure 3 [Figure 3: see original paper].

This study uses the TextRank algorithm to rank segmented words in opinion cluster text content, selecting the top 500 words as the basis for similarity comparison. Then, the TF-IDF method calculates the TF-IDF values of words in the corpora of previous and current time nodes. Cosine similarity compares opinion clusters between two time nodes pairwise. After similarity calculation, an n×m similarity matrix is obtained, as shown in Formula (6). The matrix values represent the semantic similarity between each opinion cluster in the later time node and each opinion cluster in the previous time node. By extracting the two opinion clusters with the highest similarity, the opinion cluster evolution chain is constructed. This method eliminates interference from numerous useless words in semantic similarity comparison while comparing top-ranked words better reflects the correlation between opinion clusters.

First, each opinion cluster's Internet public opinion viewpoints are merged, with the integrated corpus serving as the opinion cluster text content. Each opinion cluster is represented by a long text summarizing all viewpoints within that cluster. If an event's lifecycle is T days, then T = {1, 2, 3, …, t}. The nth opinion cluster at time node t contains a keyword corpus set Wn = {w1, w2, w3, …, wi}, while the mth opinion cluster at previous time node t-1 contains a keyword corpus set Wm = {w1, w2, w3, …, wj}. The TextRank algorithm reranks all opinion cluster keyword corpora, selecting the top 500 words (padding with spaces if the corpus contains fewer than 500 words). TF-IDF values are calculated for each word in each opinion cluster corpus at both time nodes.

TF-IDF values characterize a word's importance to a document in the corpus, increasing with frequency in the document but decreasing with frequency across the corpus. In this study, the TF-IDF values of sorted opinion cluster keywords serve as the vector representation of opinion clusters. The TF-IDF value is the product of TF and IDF values. Term frequency (TF) refers to how many times a word appears in a document, typically normalized to prevent long text interference. After calculating TF-IDF values for the top 500 keywords representing each opinion cluster, the opinion clusters at time node t and their keyword TF-IDF values form the time node's opinion cluster vector. Each value in the vector represents a keyword's TF-IDF value, with 500 words forming a 500-dimensional vector. After vectorizing opinion clusters, cosine similarity algorithm can calculate similarity between two opinion clusters, as shown in Formula (5):

$$cos\_similarity = \frac{\sum_{i=1}^{k}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{k} x_i^2} \times \sqrt{\sum_{i=1}^{k} y_i^2}} \tag{5}$$

Where k is the vector dimension, xi and yi are the ith dimension vector values of the previous and current time nodes respectively. Cosine values closer to 1 indicate smaller vector angles and higher similarity. After calculating similarity between two time nodes' vectors, an n×m similarity matrix is obtained, as shown in Formula (6):

$$Matrix_{similarity} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{m1} & s_{m2} & \cdots & s_{mn} \end{bmatrix} \tag{6}$$

Where n is the number of opinion clusters in the later time node and m is the number in the previous time node. Columns in the matrix represent similarity values between the nth opinion cluster of the later day and all opinion clusters of the previous day. The row with the maximum value in each column indicates which opinion cluster in the previous time node the later time node's opinion cluster evolved from. After comparing opinion cluster similarity across all time nodes in an event, the evolution chain of opinion clusters is constructed, and evolution state attributes of opinion clusters on the chain are extracted. In reality, an opinion cluster from the previous day may have multiple evolution chains—i.e., multiple opinion clusters in subsequent time nodes are all influenced by the previous cluster, representing a splitting phenomenon in evolution.

3.6 Construction of Opinion Cluster Evolution Level

After constructing opinion cluster evolution chains, the evolution attribute scores of opinion clusters on the chain can be calculated. Analysis of existing events' opinion cluster evolution reveals that when the growth of propagation range, emotional intensity, and cluster scale is less than 100%, the connotation and state attributes of opinion clusters basically remain unchanged, indicating low evolutionarity. When growth is between 100% and 200%, the evolution is more obvious with prominent evolutionary characteristics. When growth exceeds 200%, opinion clusters exhibit extremely high evolutionarity, with most opinion connotations having changed. This study uses piecewise functions to represent opinion cluster evolution attribute scores.

The propagation range evolution attribute score Lct is shown in Formula (7):

$$L_{ct} = \begin{cases} 1 & \text{if } 0 \leq \frac{C_t - C_{t-1}}{C_{t-1}} < 1 \\ 2 & \text{if } 1 \leq \frac{C_t - C_{t-1}}{C_{t-1}} < 2 \\ 3 & \text{if } \frac{C_t - C_{t-1}}{C_{t-1}} \geq 2 \end{cases} \tag{7}$$

Where Ct is the propagation range value of the opinion cluster at time node t on the evolution chain, and Ct-1 is the propagation range value at the previous time node t-1.

The emotional intensity evolution attribute score Lst is calculated as shown in Formula (8):

$$L_{st} = \begin{cases} 1 & \text{if } 0 \leq \frac{S_t - S_{t-1}}{S_{t-1}} < 1 \\ 2 & \text{if } 1 \leq \frac{S_t - S_{t-1}}{S_{t-1}} < 2 \\ 3 & \text{if } \frac{S_t - S_{t-1}}{S_{t-1}} \geq 2 \end{cases} \tag{8}$$

Where St is the emotional intensity value of the opinion cluster at time node t on the evolution chain, and St-1 is the emotional intensity value at the previous time node t-1.

The cluster scale evolution attribute score Lot is shown in Formula (9):

$$L_{ot} = \begin{cases} 1 & \text{if } 0 \leq \frac{O_t - O_{t-1}}{O_{t-1}} < 1 \\ 2 & \text{if } 1 \leq \frac{O_t - O_{t-1}}{O_{t-1}} < 2 \\ 3 & \text{if } \frac{O_t - O_{t-1}}{O_{t-1}} \geq 2 \end{cases} \tag{9}$$

Where Ot is the cluster scale value of the opinion cluster at time node t on the evolution chain, and Ot-1 is the cluster scale value at the previous time node t-1.

This study divides the evolution level of Internet public opinion opinion clusters into 14 levels, with each level further divided based on the three evolution attribute scores of propagation range, emotional intensity, and cluster scale. The evolution attribute scores Lct, Lst, and Lot are calculated through Formulas (7), (8), and (9) respectively. By calculating the evolution attribute scores for each time node and subsequent node on the opinion cluster evolution chain and matching them with the opinion cluster evolution level classification table, the evolution level of opinion clusters is obtained. The classification of Internet public opinion opinion cluster evolution levels is shown in Table 2 .

Empirical Study

4.1 Data Source Selection and Collection

To verify the applicability of the Internet public opinion opinion cluster evolution level, this study uses the "Zhai Tianlin CNKI Event" as the empirical case. In this event, the first peak was reached when the Ministry of Education responded to the event and launched an investigation, while the second surge occurred when the Beijing Film Academy released investigation results and revoked Zhai Tianlin's doctoral degree, after which 热度持续走低进入蔓延期. The two 热度高潮 in this event were both due to turning points in the situation, with official responses pushing the event to its peak. Netizens' discussions focused on Zhai Tianlin's alleged doctoral thesis plagiarism, the Beijing Film Academy's investigation results, and the Ministry of Education's response. Key time nodes from February 9 to February 15 in the "Zhai Tianlin CNKI Event" are shown in Figure 4 [Figure 4: see original paper].

Using a web crawler tool with "Zhai Tianlin" as the keyword, relevant Internet public opinion data was collected from the Weibo platform spanning February 9, 2019 to April 9, 2019. After data cleaning—including removing pure emoji or symbol posts, eliminating identical prefix/suffix patterns using regular expressions, and removing duplicate posts from the same blogger—27,433 valid data points were obtained. The relationship between viewpoint quantity and time in the "Zhai Tianlin CNKI Event" is shown in Figure 5 [Figure 5: see original paper].

As shown in Figure 5, the number of Weibo public opinion viewpoints in the "Zhai Tianlin CNKI Event" peaked on February 12 and February 15, then rapidly declined. The trend of viewpoint quantity change basically matches the event's 热度发展趋势, indicating that the collected data effectively describes the event's development.

4.2 Data Processing and Analysis

(1) Opinion Cluster Evolution Attribute Measurement. First, event Weibo data must be divided into positive and negative emotional viewpoints based on emotional intensity. After emotional classification, 5,153 positive emotional viewpoints and 22,280 negative emotional viewpoints were obtained, showing that the majority of Internet public opinion users held negative sentiments toward actor Zhai Tianlin's doctoral thesis plagiarism. Evolution attribute values of opinion clusters were then calculated. Due to the long time span of the event, this study selected two key time nodes—the event start date of February 9 and the last 热度高峰 on February 15—to demonstrate the evolution attribute measurement data. The positive and negative Internet public opinion opinion cluster attribute information for February 9 is shown in Table 3 and Table 4 .

As shown in Tables 2 and 3, both positive and negative Internet public opinion opinion cluster quantities were 7 on February 9. Since the event was in its early fermentation stage, some netizens and Zhai Tianlin's fans were unaware of the incident, resulting in several opinion clusters expressing expectations for Zhai Tianlin's works, recognition of his acting skills, or content about his activities, along with some advertising content posted in Zhai Tianlin's super-topic sections. Considering that this "noise" also has propagation influence, this study retained it. Positive opinion clusters had relatively low attribute values, with generally small cluster scales. In contrast, negative opinion clusters had significantly larger attribute values, most notably the fifth negative opinion cluster, which had the largest propagation range but small cluster scale, indicating that many opinion leaders had already participated in the discussion and expressed negative viewpoints. Semantic connotations revealed that opinion clusters at this time already pointed to Zhai Tianlin's academic paper plagiarism, his acceptable acting skills but lack of morality, and demands for explanations.

On February 15, when the Zhai Tianlin CNKI Event reached its peak, the positive and negative opinion cluster attribute information is shown in Table 5 and Table 6 .

As shown in Tables 4 and 5, when public opinion 热度 reached its peak, the evolution attribute values of various opinion clusters were extremely high, with cluster scale and information volume increasing dozens of times compared to February 9, indicating widespread attention to the event. Semantic connotations summarized from opinion cluster keywords show that positive opinion clusters had almost no Weibo public opinion users defending Zhai Tianlin; instead, positive viewpoints mostly consisted of advertising posts in various Zhai Tianlin super-topics or topic sections. Meanwhile, negative opinion clusters' semantic connotations evolved from discussions about the event itself to the event's impact on social order.

(2) Construction of Opinion Cluster Evolution Chains. After measuring the evolution attributes of opinion clusters in this event, evolution chains can be constructed by cross-comparing opinion clusters from two consecutive time nodes and associating those with the highest keyword similarity. Due to the long time span of the collected event data, it is difficult to display all evolution chains. Therefore, this study uses opinion cluster data from February 9 to February 15 to demonstrate evolution states. After constructing the evolution chains, 23 positive opinion cluster evolution chains and 22 negative opinion cluster evolution chains were obtained. Due to the large number of evolution chains, this experiment selected one positive and one negative evolution chain to demonstrate empirical results. The top 10 keyword changes in the positive and negative evolution chains between February 9 and February 15 are compared in Table 7 .

As shown in Table 6, the semantic connotations of keywords in both positive and negative evolution chains remained basically consistent during this period, indicating that both chains represent temporal continuations of their respective viewpoints. By comparing evolution attribute values of opinion clusters on the evolution chain, the evolution level can be obtained. Larger numerical increments indicate obvious evolution phenomena in the viewpoint represented by the chain. The evolution attributes and evolution levels of the measured positive and negative opinion cluster evolution chains are shown in Table 8 and Table 9 .

The cluster categories in Tables 6 and 7 represent which category of opinion cluster constitutes the evolution chain at each time period. Opinion clusters in the evolution chain all represent the same viewpoint theme. Opinion clusters after February 10 are continuations of the starting opinion cluster from February 9, while 0 indicates no subsequent evolution. The evolution attributes of the chain are fixed at certain values, with the three evolution attribute values being accumulations of opinion clusters on the chain. As shown in Tables 6 and 7, evolution levels were high on both February 11 and 12, with significant growth in emotional intensity, cluster scale, and propagation range, indicating obvious evolution phenomena on these two days. This occurred because on February 11, the Beijing Film Academy announced the establishment of an investigation team for Zhai Tianlin's paper plagiarism, triggering large-scale discussions among netizens on February 11 and 12 and forming a peak in Internet public opinion 热度. The opinion cluster evolution level demonstrates good effectiveness in identifying peaks in Internet public opinion evolution.

4.3 Discussion of Empirical Results

The measurement results of Internet public opinion opinion cluster evolution level using the "Zhai Tianlin CNKI Event" as the empirical case show that the evolution level measurement model can provide accurate data on various aspects of opinion clusters, including propagation range, emotional intensity, group scale, and semantic connotations. By monitoring the evolution attribute values of opinion clusters, the most influential opinion clusters in current events can be quickly identified. Deconstructing these clusters can extract key user information for targeted viewpoint guidance strategies. Furthermore, calculating the opinion cluster evolution level can clearly identify the evolution state of opinion clusters, providing a basis for rapid response and targeted guidance in Internet public opinion management.

Analysis of empirical results reveals: The Internet public opinion opinion cluster evolution level can be applied to identify the evolution degree of group viewpoints in Internet public opinion across the temporal dimension. For example, in the "Zhai Tianlin CNKI Event," new developments on February 12 generated a public opinion hotspot. Measurement of the event's opinion cluster evolution level identified February 12 as having the highest evolution level, demonstrating that this method effectively identifies evolution peaks in opinion clusters. The evolution attribute values of Internet public opinion opinion clusters can also serve as reference basis for public opinion guidance. Empirical data shows that negative emotional intensity opinion clusters had attribute values far exceeding positive ones, with clusters addressing netizens' primary concerns having much larger attribute values than others. Therefore, in actual public opinion management work, targeted control can be implemented based on opinion cluster attribute data. Additionally, the number of evolution chains at a given time point indicates the quantity of viewpoint groups in the event—more evolution chains mean more complex group viewpoints. Internet public opinion opinion cluster evolution level and attribute data can clearly represent current Internet public opinion evolution states. By comparing data from previous time nodes, changes in emotion, scale, semantic connotation, and propagation range of opinion clusters on the same evolution chain can be understood, providing diversified data support.

This paper theoretically clarifies the concept of Internet public opinion opinion clusters, providing new 思路 for measuring Internet public opinion evolution. At the practical level, it constructs a measurement model for Internet public opinion opinion cluster evolution level, using the "Zhai Tianlin CNKI Event" as a case study to verify the accuracy and reliability of the evolution level indicator. Results show that the Internet public opinion opinion cluster evolution level can accurately reflect the change level and evolution degree of group thinking in Internet public opinion, demonstrating good applicability in Internet public opinion supervision and early warning.

This study also has limitations: Opinion cluster division is based on the LDA model, and division accuracy needs improvement; This study mainly identifies evolution of text-based Internet public opinion opinion clusters, lacking recognition of evolution states in multimedia content such as images and videos; The proposed method cannot predict future evolution trends of opinion clusters, only measuring evolution levels at current time nodes. Therefore, in future research, the authors will optimize these limitations by improving opinion cluster division accuracy, adapting to multimedia Internet public opinion opinion cluster evolution perception, and using neural networks to predict future evolution trends of opinion clusters, enabling better management of Internet public opinion by relevant departments.

References

[1] Han Zhongming, Li Mengqi, Liu Wen, et al. Research on aspect-level opinion mining methods for online reviews[J]. Journal of Software, 2018, 29(2): 417-441.

[2] Sun Hong, Li Quanqi, Zhao Na. Opinion mining and tendency analysis based on double-layer tree support vector machine[J]. Intelligent Computer and Applications, 2021, 11(3): 44-47.

[3] Tang Hongting, Cai Xiuding, Zhang Yanlin, et al. Research on user idea mining method in enterprise open communities based on deep learning[J/OL]. Systems Engineering—Theory & Practice. [2021-09-21]. http://kns.cnki.net/kcms/detail/11.2267.N.20210527.1546.009.html.

[4] Zhang Liu, Wang Xiwei, Huang Bo, et al. Research and experimental study on sentiment classification model of Weibo comments based on multi-scale con-

volutional neural network of character and word vectors[J]. Library and Information Service, 2019, 63(18): 99-108.

[5] Chen Jianyao, Xia Lixin, Liu Xingyue. Evolution and visual analysis of network public opinion features based on topic graph[J]. Information Science, 2021, 39(5): 75-84.

[6] Huang Wei, Zhu Zhenyuan, Xu Yejing, et al. Construction and empirical analysis of Internet public opinion evolution index[J]. Library and Information Service, 2019, 63(20): 26-33.

[7] Saito K, Ohara K, Yamagishi Y, et al. Learning diffusion probability based on node attributes in social networks[C]//Foundations of intelligent systems-international symposium. Berlin: Springer berlin heidelberg, 2011.

[8] Zhou Qiping, Yang Fang. Research on unsupervised early warning mechanism of Internet public opinion based on SIS model[J]. Information Science, 2019, 37(8): 51-55.

[9] Yang Ding, Yang Aimin. A Chinese text sentiment classification method based on sentiment dictionary and naive Bayes[J]. Computer Application Research, 2010, 27(10): 3737-3739.

[10] Blei D M, Ng A Y, Jordan M L, et al. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 11(3): 993-1022.

[11] Freeman R M. Observing language changes in aging and Alzheimer's speech using information theory techniques[J]. Dissertations & theses-gradworks, 2015, 11(1): 21-26.

[12] Su Jiongming, Liu Baohong, Li Qi, et al. Online rating population prediction based on opinion dynamics[J]. Computer Engineering, 2014, 40(10): 155-160.

[13] Wang Meifang, Liu Peiyu, Zhu Zhenfang. Feature selection method based on TFIDF[J]. Computer Engineering and Design, 2007, 28(23): 5795-5796, 5799.

[14] Gao Junfeng, Huang Wei. Calculation of emotional polarization degree of opinion clusters in Internet public opinion field[J]. Library and Information Service, 2019, 63(10): 106-114.

Author Contributions:
Yan Lu: Designed and modified research methods, conducted experiments, analyzed data, wrote the paper.
Yang Gang: Proposed research direction and 思路.
Zhao Jiangyuan: Collected and organized data, analyzed data, revised article structure.

Abstract: [Purpose/significance] Proposing and constructing the evolution level of Internet public opinion opinion clusters to describe the evolution degree of group opinion states of Internet public opinion audiences over time and events is of great theoretical and practical significance for Internet public opinion guidance and precise guidance. [Method/process] Based on LDA and CNN neural Internet, the paper constructed a level measurement model of Internet public opinion cluster evolution, and took "Zhai Tianlin CNKI event" as the experimental object to verify the effectiveness of the evolution level index. [Result/conclusion] The evolution level of Internet public opinion clusters can well reflect the evolution of group opinion states of Internet hotspot events. It can show the attribute values of three dimensions and also reflect the evolution degree of opinion clusters compared with the previous time node state. The measurement results of opinion cluster evolution level proposed in this paper accurately reflect each evolution peak of event opinions, providing a new guiding direction for relevant departments to target and guide the opinions of Internet public opinion groups.

Keywords: Internet public opinion; opinion cluster; opinion evolution; evolution level; measurement model

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