



Event Evolution Graph

Based on Social Media Comments

by

Weng Xu

Supervisor: Renuga Jayakumar

September 2023

DECLARATION

I, Weng Xu declare that I am the sole author of this Project; that all references cited have been consulted; that I have conducted all work of which this is a record, and that the finished work lies within the prescribed word limits.

This work has not previously been accepted as part of any other degree submission.

Signed : Weng Xu.....

Date : 15/09/2023.....

FORM OF CONSENT

I Weng Xu hereby consent that my Project, submitted in candidature for the *M.Sc. Data Science and Analytics*, if successful, may be made available for inter-library loan or photocopying (subject to the law of copyright), and that the title and abstract may be made available to outside organisations.

Signed Weng Xu

Date 15/09/2023

Where necessary replace *M.Sc. Data Science and Analytics* with your named award.

Abstract

Netizens have become more comfortable expressing their opinions on an event due to the widespread use of social media and the popularity of the Internet. Online public opinion has also increased significantly. It is a collection of netizens' opinions and views that can reflect users' most direct thoughts. Comment video data is one of the simplest and most direct forms of netizen commentary on an event, which can be a great source of objective information and a visual representation of current social issues.

The Event Evolution Graph is used to realize the evolutionary research of online public opinion public events in this paper. By using the Event Evolution Graph, we can clearly visualize public opinion's evolution path and summarize and summarize the evolution law and evolution characteristics of network public opinion more conveniently and conveniently. It is an Event Evolution Graph with causal relationships that can be used to illustrate the evolution path of public opinion.

This study uses BILIBILI video website comments as the data source, and selects the recent popular online public opinion event "local government debt". Natural language processing and other methods are used to identify the causal relationship in the public opinion text, take the cause event as the source node and the result event as the target node, and construct a theory map based on the transitivity of the causal relationship. And integrate TF-IDF

+ K-means clustering method, based on Word2vec + K-means clustering method, based on Doc2vec + K-means clustering method, based on TF-IDF + K-means++ clustering method, based on Word2vec + K-means++ clustering method, based on Doc2vec + K-means++ clustering. By contrasting the use of the based on Word2vec + K-means++ clustering method, an Abstract Event Evolution Graph of things is generated.

Contents

1. Introduction	1
2. Literature Review	7
2.1. Online Public Opinion in Public Emergencies.....	7
2.2. Characteristics of Online Public Opinion in Public Emergencies ..	9
2.3. Event Evolution Graph.....	12
2.3.1 Rule-based Templates	12
2.3.2 Feature-based Relationship Identification.....	13
2.3.3 Machine Learning-Based Relation Identification	14
2.3.4 Event Generalization	15
2.3.5 Graph.....	16
3. Research and Development Methodology	20
3.1. Research Methodology	20
3.1.1 Knowledge Graph and Event Evolution Graph	20
3.1.2 Data Mining and Collection	21
3.1.3 Natural Language Processing.....	22
3.1.4 Relationship Extraction.....	24
3.1.5 Event Generalization	25
3.1.6 Abstract Event Evolution Graph	38
3.2. Experimental Design.....	39
3.2.1 Experimental Hypothesis	39

3.2.2 Experimental Inputs	40
4. Design and Implementation	41
4.1. Design	41
4.1.1 Experimental Process	41
4.1.2 Data Mining and Collection	43
4.1.3 Natural Language Processing.....	43
4.1.4 Relationship Extraction.....	44
4.1.5 Event Evolution Graph.....	47
4.1.6 Clustering Algorithm Improvement Evaluation.....	48
4.1.7 Abstract Event Evolution Graph	54
4.2. Implementation	55
4.2.1 Data Mining and Collection	55
4.2.2 Natural Language Processing.....	56
4.2.3 Relationship Extraction.....	57
4.2.4 Event Evolution Graph.....	58
4.2.5 Abstract Event Evolution Graph	59
5. Analysis	61
6. Results and Discussion.....	63
7. Conclusions and Recommendations	64
8. Reflection	65
References	66

1. Introduction

With the rapid rise of technology in our lives, platforms such as Twitter and Facebook have become major hubs for social interaction, giving users the opportunity to express their opinions as well as to share information on an event. Users are able to share, access and spread information through these social networks. It is not uncommon for these platforms to become hotbeds of public discourse during times of crisis, triggering lots of interaction between users. The topic-based search tool becomes an effective tool for the retrieval of information, tracking of events, and analyzing topics in this digital ecosystem. It is important to note that user reviews are both positive and negative, and provide a true picture of real-world events and serve as an indicator of pressing societal issues at hand. A natural disaster, an accident, an incident in the public health system, or an incident in the social security system that occurs suddenly, causes or may cause serious social harm, and requires emergency handling measures to be taken. Recent years have seen an increased number of emergencies occur, and due to the severe losses caused by various emergencies, the government has issued a large number of laws, regulations, and normative documents on emergency response.

The aim of this study is to develop a method for visualizing user interactions by analyzing hidden topic data from social media. We analyze the distribution of topics in user reviews and quantify the prevalence of

specific tags. The graphical representations of online discussions will allow us to clearly understand their trajectory, as well as to draw patterns and trends in the evolution of topics from these discussions. Social networks tend to be characterized by the change in topics of discussion, which are often influenced by multiple factors that interact at different stages of the theme's development, which means that the theme's development trajectory is unpredictable to a certain extent. The current academic inquiry mainly focuses on analyzing the internal factors and universal patterns underlying change in public opinion. Public opinion development stages are generally identified by examining the temporal sequence and periodicity of public opinion events and further exploring the changes in public opinion during these stages by examining the stages of public opinion development. In these studies, the general law of public opinion change in each stage is discussed by analyzing the temporal sequence and periodicity of public opinion events, which then allows the studies to categorize the different stages of public opinion development. Despite this, existing research has barely explored the logical relationship between public opinion events and their interrelationship and development trajectory, and in terms of the logical relationship between public opinion events, the research has not yet been exhaustive enough. In order to provide strong support for decision-making, it is crucial to conduct in-depth research on the evolution path of online public opinion so that we can be more aware of major events and control key links in public opinion

management. Social media platforms are expected to play a key role in public opinion management in the future, especially in terms of focusing on major events and controlling key actors and influencers.

In the present day, the research pertaining to online public opinion prediction mainly focuses on the prediction of public opinion development trends, emotional trends and risk levels. The majority of these predictions are made by analyzing the postings and discussion of netizens to predict how public opinion might develop over the coming years, which has an important role for public opinion warnings on social media platforms as well. In order to provide enterprises or organizations with more targeted risk decision-making and emergency management based on a deeper understanding of the evolution of emergency events from a variety of perspectives, such as event development and risk subjects, the analysis perspective can be further enriched and the accuracy of the analysis results can be improved, so as to provide enterprises or organizations with more accurate data for making risk decisions and managing emergency situations. In the current social media environment, evolutionary analysis of public events and emergency management research remain a relatively new field. It is true that a number of studies have been conducted to study the impact of social media public events or risk case studies and lessons learned from specific events, but these studies still face significant challenges in the context of big data. To start with, traditional expert-centered methods of analysis and decision-making are

unable to deal with large amounts of data generated by social media. Therefore, we must use more sophisticated and automated tools and techniques in order to analyze this data effectively. As a result of the current research, neither the evolution mechanism of events has been deeply explored from a data-driven perspective, nor has an elaborate research system been established. Additionally, social media public events are subject to a great deal of subject matter and influence factors, as well as the obvious derivation and secondary nature of these events, making the evolution mechanism of events extremely complicated. It creates significant challenges and pressures for emergency management measures that are routine in nature. As a result, it has become a hot topic in academia to use reasonable and effective methods in order to discover and analyze the evolutionary mechanisms behind social media public events. The evolution of events must be understood and managed in more depth in order for us to develop new tools and techniques so that we can analyze and predict trends in events in a better manner. In order to do that, we need to investigate the evolutionary mechanisms of events in greater depth.

There are a number of algorithms used to study online public opinion, which are usually based on machine learning and statistics. In the field of statistics, we often rely on methods like regression analysis, differential equation analysis, and analytic hierarchy; and in the field of machine learning, neural networks, gray system theory, support vector machines, and Markov

chains have also been widely used. To build these models, these algorithms mostly use numerical data, so that they can accurately predict trends, emotional tendencies, and potential risks of public opinion. However, they do not have the ability to predict specific events, since they are based mainly on numerical data. As a logical directed graph model, the Event Evolution Graph provides a new perspective for solving this problem. As a result of this model, nodes represent different events, while edges show the causal connections between these events, thus providing an excellent opportunity to gain a more detailed understanding of the nature of the problem as well as to determine a more accurate public opinion of the network. There are, however, still some limitations to the traditional theory Event Evolution Graph of causality in showing the essence and universal law of causality in a particular field. In response to these findings, the concept of abstract events evolved into a concept that is not only a general name for a series of specific events, but also ignores secondary factors such as time, place, and concentrations more on the universal laws of cause and effect in the subject area. As a result of this kind of abstract event, the "abstract Event Evolution Graph " provides the theoretical and methodological foundation needed for the prediction of online public opinion events to be more accurate. In this study, crawler technology is used to extract public event descriptions that are unstructured from social media on the web. These descriptions are extracted through crawlers. The next step in this process is to perform preliminary processing by means of

natural language processing techniques, such as word segmentation and syntactic analysis. Using pattern matching technology, it is possible to identify explicit relationships between events as well as the trigger words that are associated with events, as a means of achieving the generalization of events by combining traditional methods. As a final step, we construct a social media public Event Evolution Graph based on the event information in the text, which is then displayed through a visual interface in order to create a structured Event Evolution Graph. In the field of social media emergencies, this provides a valuable reference for the design of the Event Evolution Graph.

2. Literature Review

Online public opinion in public emergencies is a complex and challenging subject. These emergencies have unique characteristics such as their suddenness, potential harm to society and the public, and their public nature. These events can be classified into four importance levels based on their nature, depth of impact, difficulty of control, and scope of impact: particularly important, important, relatively large, and ordinary. A striking feature is their suddenness, which makes it difficult to predict and determine their cause in a short period of time. In addition, these events may lead to losses in various fields, such as politics, economics and other fields, posing a threat to the environment and people's lives. The literature review part provides the corresponding support required for this paper by describing and summarizing the process and related methods of constructing the map of predecessors.

2.1. Online Public Opinion in Public Emergencies

Public emergencies are characterized by their unexpected occurrence, their potential for harm to society and the public, and their public characteristics[1]. There are four levels of significance in these events: particularly significant, significant, relatively large, and ordinary, depending on the nature of the events, the depth of the impact, the difficulty of control, and the scope of the impact.

This type of event has the following salient characteristics: 1) Sudden nature, which makes it hard to predict it in advance and determine its causes in a short period of time. During the event, unforeseen changes occur due to external environment and human intervention in addition to uncontrollable objective factors at the beginning[2][3]; 2) Harmful, these events may result in losses to society in various fields, including politics, economics, etc., and pose a threat to the environment and people's lives. Especially in the network environment, when the event occurs suddenly, public panic may result; if information about the event cannot be conveyed within a reasonable amount of time, false information may spread[4][5][6]; 3) A wide distribution of events in a network increases the scope and complexity of the network. Because society is diverse and the borderless network and a variety of application software are so popular, public emergencies are usually unconventional. These events become more complicated as a result of their structure and development model. There may be multiple departments involved in handling an accident, and one incident may also draw attention to other issues, making it difficult for standard emergency plans to be followed[7].

As a result of public emergencies being unpredictable and unexpected, they often spark extremely complex public opinion on the Internet. In addition to being urgent, these events also carry a great deal of uncertainty, which causes widespread concern and heated discussion among netizens[8].

During such times, the public is desperate for information and to express their opinions. Further, the public will participate actively in the discussion of the incident on the Internet in addition to official news releases, official information releases and comments from the government and major media. In order to maintain public trust, governments need to respond in a timely manner and act quickly and effectively. If not handled appropriately, it can severely damage the credibility of the government. To ensure the health of the online public opinion environment, the government should adopt stable, universally applicable, yet flexible management strategies to deal with these incidents.

2.2. Characteristics of Online Public Opinion in Public Emergencies

There are many characteristics of public opinion that are unique to the Internet, including its universal nature, its unexpected occurrence, its bias in direction, and its polarity in spread. A 2008 study by Tang Xiliang concluded that online public opinion reflects Internet-related characteristics like interaction, real-time, concealment, visibility, autonomy, manageability, and emotion, since it is generated and expanded in the context of the Internet. Logic and illogicality, individualization and collective extremes, etc[9].

It is the speed of information dissemination in time and space that has markedly improved the difference between public opinion and traditional public emergencies on the network. In addition to changing the way we live,

the Internet has also had a two-sided impact. Often, public emergencies happen at unexpected times and at unexpected places, and their cycles from the beginning, outbreak, peak, and gradual subsidence are usually short, and their development pattern is often unpredictable as well[10]. Public opinion has a high degree of freedom in the network due to the suddenness of events and limited information. In the network, both government agencies and independent media can be disrupted or make inaccurate decisions. It is more evident in the network that public opinion is sudden in a network because of the suddenness of emergencies and the openness of online public opinion. Frequently updated websites and headlines with interesting content can often lead to misinterpretations, as website content is updated to attract attention[11].

In the interim, given that online regulation is still being revised, some people spread misinformation, post extreme views on social media sites, or express themselves emotionally. The "Free Opinion Square" of this network is a place where people can share their opinions and thoughts in the virtual world despite hiding their identities and no network surveillance, as the user's identity is kept secret. A different perspective is that the anonymity of the Internet also enhances the realism of speech and encourages netizens to express and vent their emotions. Psychology professor Freud divided the human personality structure into three parts: id, ego, and superego[12].

Many external influences interact with one another, but they are

independent. When public opinion is formed, it becomes very difficult for the government to control, resulting in a spiraling situation. The qualitative analysis method has been adopted by scholars, for example Zhang Fang et al. [13] proposes a corresponding information dissemination model based on the traditional infectious disease model, showing that rumors spread as quickly as the butterfly effect in emergencies, gradually transforming from a chaotic state to one that is consistent. Public opinion behavior is influenced by state policies, public opinion, and the media, according to Kam C. Wong [14]. The purpose of Bart W. Edes (2000) is to examine how government information officers influence public opinion and analyze their abilities to guide public opinion in order to ensure transparency and openness[15].

First, the paper clarifies how public emergencies are defined and classified, discusses the unique characteristics of online public opinion regarding public emergencies, and provides theoretical support for further studies. There is no doubt that public emergencies are often unexpected and affect society and the general public significantly. Thus, we can deeply explore the causes of different events and formulate coping strategies that reduce their negative social consequences according to their characteristics. A majority of Internet users' views, emotions, and behavioral tendencies are reflected in the online public opinion about public emergencies in the Internet age of big data and the Internet. An obvious characteristic of this type of opinion is that it is sudden, secrecy, diverse, and difficult to control.

2.3. Event Evolution Graph

By identifying and extracting connections between events from relevant text, we can build a networked knowledge framework that ties together different event concepts. By extracting event relationships, we can discover their internal laws and understand their development process. An event is the core of this method, and logical relationships between events are extracted, including causal relationships, continuous relationships, temporal relationships, and hierarchical relationships. Specifically, the construction technology of the affair map is closely related to the extraction of event causality and event continuum.

2.3.1 Rule-based Templates

Real information extraction involves relating each event to a different entity, so there is also a complex relationship between them. Events do not occur independently, and they tend to have semantic links. The causality of events is one of the key elements in explaining how events relate to one another, revealing the order of events and the causal connections that affect each other over time.

Since the Chinese text was unique, the method of causality identification was adaptively improved after it was absorbed and borrowed from the international academic community. A comprehensive extraction method combining part-of-speech labeling, syntactic analysis, and predetermined

causal templates has been proposed by researchers like Ittoo[16]. After extracting rules from text with causal relationships, the method then probes potential relationships in other sentences using these templates. In Kozareva's method[17], causal triggers are detected and relationships between nouns are analyzed to determine if a text is logical. The identification of textual relationships has also been done by scholars such as Sorgente[18] and Zhao S[19] using rule-based template techniques.

The use of local syntactic characteristics and rules in the identification of causal components within sentences has been successfully achieved by domestic scholars like Li Yuequn in research on open source Chinese texts for security[20]. The same time, Qiu J further categorizes causal text into two categories, explicit and fuzzy. As well as identifying logic, a Bayesian classifier was used to identify logic in addition to extracting relationship pairs based on the rules of the extracted sentences. Experiments proved the method to be effective[21].

2.3.2 Feature-based Relationship Identification

As part of their research in Chinese text analysis, domestic scholars focus primarily on the relationship between sentences, and use dependent syntactic analysis techniques to do so. The team of Li Lishuang[22], for instance, developed a planar kernel system based on eigenvectors and a structural kernel system based on syntactic trees to extract entity relations. In

experiments on Chinese datasets, this method improves relational extraction performance. Guo Xiyue improved his experiments on news texts by combining syntactic and semantic features to identify relationships between entities[23]. The analysis was further improved by applying causal connective word features constructed from nucleus similarity of sentences containing connective words, as well as conventional features[24].

2.3.3 Machine Learning-Based Relation Identification

A matter map describes relationships between entities as its core, which is why relationship extraction is crucial in the construction process. We discussed earlier that relational extraction consists of two main categories: traditional machine learning models and deep learning-based tools. It is possible to differentiate template-based and statistical-based machine learning methods within traditional machine learning methods. In addition, statistics-based methods include those based on linguistic characteristics, graph technologies, and subspace embedding strategies. In order to recognize relationships, machine learning is often used in conjunction with rule template methods and feature-based approaches. A classifier is used to identify relationships between sentences after forming a rule template or generating features.

Ittoo et al. were able to identify partial-global relationships through a minimal supervision algorithm in a study by Riaz, M., et al[25]. Through

supervised training on large amounts of data, Radinsky et al. developed three classes of classifiers by dividing causal connectives into rule templates[26]. They then built relationship recognition rules and extracted causal relationships from 150 years of news headlines[27]. Using artificially constructed dependency syntax rules, Sorgente et al. extracted possible causal pairs and filtered out noisy word pairs for non-causal pairs by using Bayesian classifiers[28].

2.3.4 Event Generalization

Generally speaking, when events are generalized, they can represent a whole class of events that occur at a higher level of abstraction. The process of event representation involves extracting structured data from unstructured text, which is the basis for generalizing events.

VerbNet was used to classify verbs by Radinsky; nouns were generalized using Wikipedia conceptual words. By calculating a cosine similarity between the body of the news article and the body of the Wikipedia article, ambiguity is further eliminated[27]. In addition, he clustered events that were semantically similar together using hierarchical clustering. Similarly, Zhao et al. propose using VerbNet to classify verbs and WordNet to classify nouns. By converting events to verb classes and upper places, events can be removed that are too specific[19].

2.3.5 Graph

While the affair map has only been around for a few years, its research has been focused mainly on statistical script learning and event relationship recognition for years. Statistical script learning, which is a prediction method based on known events, was first developed by Chambers et al. in 2008[29]. Based on known events, it can determine what is likely to occur next based on known events. A chain structure is formed by placing events in chronological order in the text in this approach. Moreover, event-centric knowledge representation was also proposed by Glavas et al. Glavas et al. defined the concept of an "event graph", a structured representation based on events. An event graph is a representation of events where vertices represent mentions, edges represent time, and relationships between events are different[30]. Additionally, they designed an automated map-building system. Generally speaking, an event map looks very similar to a map of events from a structural standpoint.

Our understanding and analysis of events and the relationships between them is improved by the introduction of the event map. An event graph, in contrast to traditional knowledge graphs, focuses more on the relationships between events, such as succession, cause and effect, conditions, etc., rather than entities and their attributes and relationships. Our understanding of this relationship is enhanced and we can see the logic and patterns behind events when we characterize it.

An explanation of why A causes B could be provided by a knowledge graph, while a matter graph would provide a further explanation of why A causes B, and why A causes B under certain conditions. We can use the Facts Map to better understand the causes and potential effects of an event. Additionally, we can use the Facts Map to predict the future course of events. Inferring subsequent events that may follow an event can be done by analyzing the logical relationships between events, making predictions more accurate. Overall, the Facts Map offers a more comprehensive and in-depth tool for understanding and analyzing events and their relationships, thus providing stronger support for decision-making[31][32].

Several large-scale knowledge graphs, such as Wikipedia and Dbpedia, have emerged since the concept of graph was first proposed. The affair map is constructed by analyzing the structure of triples[33]. In spite of the large amount of unstructured text stored in real-world, these maps are primarily constructed by using structured data. However, manual extraction is time-consuming and costly, so it is often underutilized due to the challenges of extracting information from this unstructured data. In that time, machine extraction technology was not very accurate, and it was difficult to apply to industrial applications[34]. In recent years, the method of automatically extracting knowledge from unstructured text to build a knowledge graph has gradually become mainstream thanks to rapid advancements in deep learning, particularly natural language processing. In order to automatically construct

a knowledge graph, two core elements need to be recognized: named entities and relationships.

Traditional knowledge maps primarily describe entities, but Mai et al. describe the real world as another key element[35]. By using edge-weighted PageRank and sampling strategies, we established an answer system for knowledge question questions using a spatial knowledge embedding learning model based on knowledge maps. By aggregating different data sources, Kim creates a knowledge graph by combining the entertainment ontology with the K-POP knowledge map[36]. According to Chi et al., events and their interrelationships often do not receive systematic explanation[37]. They propose an ensemble management model for scientific metadata based on knowledge graphs for analyzing knowledge graphs in intelligent education.

Using news data, Rospocher et al. created the Event Atlas in 2016, which provides a powerful complementary tool to static knowledge in the TK Atlas that captures long-term development and historical trajectories of a large number of entities[38]. In 2018, Simon Gottschalk developed a cross-lingual knowledge graph for Event Hubs that not only incorporated events at its core, but also modeled their temporal relationships[39]. Harbin Institute of Technology professor Liu Ting and his colleagues also introduced the concept of an affection map, which focuses on events and models causal, conformity, and transition relationships among events[40]. It has been demonstrated that Matter Maps can be used to predict events well.

It is noteworthy that the Affair Map has been constructed using the best possible technology, and its advantages have been proven in analyzing changes in public opinion in networks. In spite of this, it is important to note that public opinion differs from events. Although the structure of the "public opinion" text may be looser compared with the normativity of the "event" text, and its core information may be more scattered, its structure is clearer and easier to comprehend. The connections between "event" texts are more complex because they involve more participants and roles.

3. Research and Development Methodology

3.1. Research Methodology

3.1.1 Knowledge Graph and Event Evolution Graph

In essence, the Event Evolution Graph is a logical knowledge base that describes the laws and patterns of evolution between events in terms of logical knowledge. Graph structure of this program is a directed cyclic graph in which nodes represent events and directed edges indicate succession and causal relationships.

As far as knowledge expression is concerned, the knowledge graph as well as the Event Evolution Graph are the fundamental forms of expression, however there are some significant connections between them as well as differences between them. There is a lot of information about entities, abstract concepts and their relationships in the knowledge graph. However, most of its research focuses on noun entities and their interrelationships, and the depiction of knowledge is not dynamic. On the other hand, the Event Evolution Graph focuses on the temporal sequence and causal logic of predicate events and their interrelationships. Its description emphasizes how these events are connected on a time and spatial scale. I believe these two approaches are functionally different. While the knowledge graph is focused primarily on answering the questions of "who, what, when, and where"; the Event Evolution Graph is focused more on answering the questions of "why,

how". Structurally, both approaches use directed cyclic graphs for expressing knowledge. It is worth noting that the knowledge in the knowledge graph is determined, with fact triples $\langle \text{entities, relationships, entity} \rangle$, $\langle \text{entities, attributes, attribute values} \rangle$ as the main knowledge representation. This knowledge state is relatively static, and has high requirements for knowledge accuracy and real-time. In contrast, the knowledge in the theory Event Evolution Graph is represented in events, metaset, and logical relationships. Its knowledge logic has uncertainty and probabilistic transfer relationships, which is more conducive to mining the development mode of events and conducting decision-making knowledge reasoning.

3.1.2 Data Mining and Collection

It is a resource devoted to annotating breaking news events that appear in Chinese coverage by using the CEC, or Chinese Events Corpus. According to the National Overall Emergency Response Plan for Public Emergencies, the corpus is constructed based on the classification system of five major types of emergencies: earthquakes, fires, traffic accidents, terrorist attacks, and food poisoning, among others. The first step in constructing this corpus was to select news reports from five kinds of breaking news events as raw data. Using XML as the annotation format, all annotation results are saved in a corpus and are saved as XML files. They are then text-preprocessed, analyzed, and event-annotated.

This study required the construction of the required corpus manually as the CEC corpus does not cover the event of interest. As such, the crawler program written by the octopus program crawled the BILIBILI video website in real-time, providing information on epidemics and video commentary on them. After the crawled content had been processed, JieBa and HIT LTP tools were used to perform natural language processing such as word segmentation, clause segmentation, and part-of-speech labeling on the crawled content. When working with crawled data, you need to ensure that labels extend from the word level to the sentence level. In addition to data cleansing, word segmentation, and deduplication, you have to ensure that labels extend from the word level to the sentence level of the data.

3.1.3 Natural Language Processing

It is necessary to segment words in Chinese sentences because unlike words in English sentences, which are usually separated by spaces, Chinese sentences do not have the obvious separation sign, which leads to the necessity of word segmentation. There are two main types of Chinese word segmentation algorithms. The dictionary method identifies words in sentences by comparing them with words in a predefined dictionary. The statistical learning method identifies words in sentences by comparing them to words from a set of dictionary entries. There are several tools that are well known for segmenting words: NLPIR (Chinese Academy of Sciences), LTP

(Harbin Institute of Technology), Tsinghua University's THULAC, and Jieba, for example. A combination of Jieba and LTP was used to perform Chinese word segmentation and part-of-speech labeling in this study. Combining dictionaries with statistical learning methods can help word segmentation be more accurate and efficient. Furthermore, term statistics can be used to create a prefix dictionary, which enables the generation of a directed graph of the possible word compositions in a sentence. The dynamic programming method is used as a final step in determining the best word segmentation combination based on the frequency of each word.

Using crawler technology, data acquisition may contain noise and duplicates. Therefore, the first step in preprocessing is to eliminate duplicates and noise from the data. In addition, we use the Jieba word segmentation method to segment clauses, words, and part-of-speech labels on the collected corpus using the Jieba method. In experiments, verbs play an important role as triggers, and through part-of-speech labeling, verbs can be identified accurately from sentences. Verbs are important for experimental purposes. As a result of the part-of-speech labeling, you will see a number of pattern templates that describe a statement after the part-of-speech labeling is complete. Due to the relatively complicated grammatical structure of Chinese, tagging parts of speech is not a straightforward task. In contrast, an English word can usually be identified directly by its morphological changes as to what part of speech it belongs to. Nevertheless, in Chinese, the same form of

the word may have different parts of speech, which, undoubtedly, makes part-of-speech labeling even more challenging.

3.1.4 Relationship Extraction

A technique to extract logical connections between events is referred to as event relationship extraction. By using this method in depth, we can provide a more detailed description of the trajectory of events. In addition to helping to reveal changing patterns of events, this in-depth understanding also provides early warning of their potential development.

There are nodes in a network public opinion Event Evolution Graph that are used to represent events, which means that in order to construct an abstract Event Evolution Graph, one needs to extract events from that network public opinion Event Evolution Graph. In order to form the correlation logic between events, it is necessary to identify the starting and ending events in order to form event combinations. As part of the construction of a Event Evolution Graph of things and the discovery of the internal connections between events, it is necessary to investigate the logic of events in order to gain a better understanding of them. By combining Chinese causal sentences comprehensively, we are able to obtain a rule template that is constructed based upon the syntactic pattern of these sentences. Using the template, a set of decision rules can be used to determine whether the data constitutes a causal sentence. Patterns are a regular pattern with causal triggers,

Constraints are syntactic limits of a sentence, and Priority is an option that you can choose from when multiple templates match. In this context, Pattern represents a regular pattern with causal triggers, Constraints represent syntactic constraints, and Priority represents a choice for priority when multiple templates match. The use of explicit causal connectives is often necessary when it comes to identifying causal event pairs. Among the existing research, rule templates for causal determination are considered to be the most effective method. As a result of the research carried out in this study, the rule template that is required to be able to determine the causal relationships of events is constructed, and the causal pairs in the research data are extracted in order to reveal the causal relationships between major emergencies in social media public opinion.

This rule template is derived by aggregating relational words from Chinese text to form causal sentences. As a next step, you can apply this template to new sentences in order to determine whether or not they contain a relational word and mark them as causal sentences. You should continue using the rule template for the marked causal sentences, separating the cause and effect parts with keywords, and subsequently extracting relevant events from the sentences, which are then used as nodes in the model.

3.1.5 Event Generalization

While it is true that this diagram can be used to illustrate the logical

relationship between events, it cannot capture the universal causal pattern as a directed diagram describing the logical relationship between events. In particular areas, this model is crucial to understanding the pattern of progress in online public opinion. Thus, a simplified and universal causal model can be formed by reorganizing the events on the Event Evolution Graph of events. The abstract picture of things, which is derived from the universal relational model, is something we call the "abstract Event Evolution Graph." By using the abstract picture of things, we are able to understand more clearly the evolution of network public opinion in a particular area. If we are trying to predict network dynamics, the abstract matter graph can provide us with a lot of useful information when it comes to event relationships and weightings.

Our goal is to introduce the concept of event generalization in order to establish a link between the event graph and the abstract event graph. In event generalization, two semantically close words are generalized to the same event, thereby representing them as one entity. As a result of this strategy, the complexity caused by the diverse event vocabulary is reduced, and similar events in the event graph are represented as one node, thereby improving the generalization capability of the event graph. Our reasoning for using the clustering method was based on the fact that it makes sense to group sample points with similar characteristics into a group, which is in line with the idea that events can be generalized, so we decided to use this method to attain this.

An abstract truth Event Evolution Graph is an analysis technique that

aims to generalize various events in the field of online public opinion so as to indicate the causal connection between these events and other events that are more generalized. In order to classify the sample points into the correct cluster, the key principle of clustering is to match the sample points with their nearest cluster centers, and then to describe a series of specific events using a highly representative event, and to formulate an abstract picture of public opinion in a network. By using this method, it is possible to sample the events and group them into clusters. K-means clustering uses techniques such as TF-IDF vector, Word2vec vector, and Doc2vec vector when converting text into numeric vectors to optimize the algorithm in order to achieve the best possible generalization results.

1) TF-IDF

There are several ways in which words can be ranked according to the significance they hold within a document set or corpus by using the TF-IDF technique. When a word or phrase appears frequently in a particular document but appears less frequently in other documents, it indicates that the word or phrase has significant classifying capacity. This word, or phrase, is therefore of high value when it comes to classifying documents. TF stands for Word Frequency, and is used to determine the number of times a word is found in a document. It is known as the inverse document frequency, or IDF, and it is used to assess the universality of a given word. That is to say, if a word is frequently occurring in most documents, its IDF value is low, whereas if it

occurs in fewer documents, its IDF value is relatively high.

TF-IDF word vectors use TF-IDF values to represent a text or document based on their weighting. It is possible, for instance, to construct a vector from the TF-IDF value of each word within a document to be used to classify or cluster the text. The distinctive feature of the TF-IDF is that it can effectively filter out common but low-information words such as "and", "is". The vector is suitable for subsequent tasks, such as text classification or clustering. In spite of being ubiquitous in a large number of documents, these words are not very informative. When used in combination with TF-IDF, these words are reduced in weight, whereas words that appear frequently in a particular document but are rare in other documents are given a higher weight, resulting in a more accurate description of the document topic. During the study, each event is treated as a document. The words in the document represent the words in the event, and are used to explain the event. There is a length difference between the word bag and the word bag of the event graph. The word bag contains all the words of all the events in the event graph. The TF-IDF vector representation of each event can be found in the word bag, as shown in Figure 1.

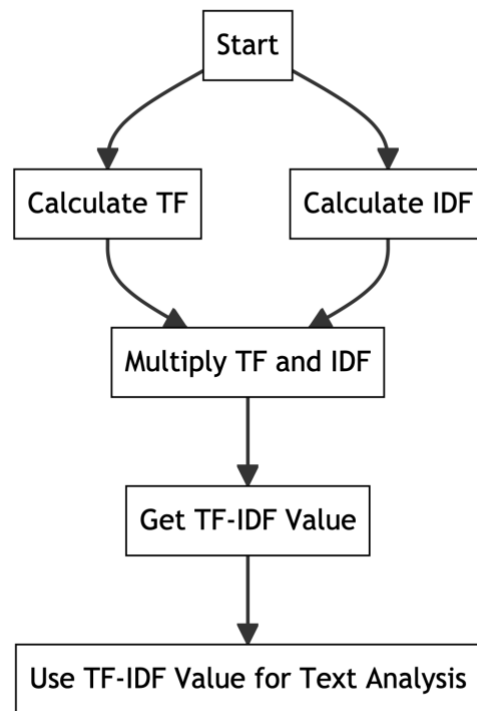


Fig. 1. The Schematic Flow Chart of TF-IDF Method

2) Word2vec

By using the bag-of-words model, the TF-IDF method defines the dimensions of an event vector based on the number of words in the bag. The number of words determines the dimensional length of the event vector. It is possible for the dimensions of an event vector to become quite large when there are a large number of vocabularies in the event. Some even refer to this phenomenon as so-called "dimensional disasters". Using Word2Vec, on the other hand, each word is converted into a fixed-sized vector. The vectors represent the semantic associations between words, such as meaning, antonym, and context, in which the words are connected.

The Word2Vec method generates vectors based on word context

relationships in the text, thus overcoming the limitations of the TF-IDF strategy. As soon as the corpus has been learned, every word is converted into a vector representation and stored in a vector space. The basic principle is that the proximity between words is primarily determined by the distribution of the words adjacent to them. In-depth exploration of the relationships between words is enabled by the effective capture of semantic information in word vectors.

A CBOW model predicts a target word (or central word) based on its surrounding vocabulary using the given context (i.e. surrounding vocabulary). Specifically, the model takes the word vector of the context vocabulary as input and attempts to predict its corresponding target word word vector. For example, in the sentence $\text{sent} = (\text{context}(A), A)$, the model uses the information in $\text{context}(A)$ to predict the vector representation of the word A . The CBOW network consists of three main layers: input, hidden projection, and output. On the other hand, the Skip-Gram model is the exact opposite of CBOW. With CBOW and Skip-Gram, target words and context are predicted using the word vectors of the target words as inputs. These strategies are both different in terms of how they are predicted. In order to choose the right model, you should consider the requirements of your application and the characteristics of your data.

3) Doc2vec

A vectorized representation of a document can be obtained using the

Doc2Vec algorithm. Doc2Vec produces a fixed-length vector for a document or paragraph, unlike Word2Vec, which only generates vectors for words. In order to use this algorithm, word and document vectors must be learned in parallel. A new element called "Document ID" or "Paragraph ID" was introduced during the training process, which is used to train the document with its vocabulary. Thus, the model captures not only word semantics, but also the overall context and semantic structure of the document as a whole.

For many machine learning algorithms, fixed-length vector input is a requirement, and bag-of-words models have become widely used to vectorize short text. There are, however, two obvious shortcomings of the model: first, it doesn't take into account word order, so sentences that contain the same word in different orders are interpreted as identical representations; second, it doesn't capture semantic information about words. As a result of these problems, the Doc2vec model, which is based on word2vec, was developed. Among its many advantages is its ability to process sentences of various lengths without presetting sentence lengths as training samples. In addition to overcoming the structural limitations of the bag-of-word model, Doc2vec generates vector representations for documents as an unsupervised method. When Doc2vec predicts document vectors, it also captures semantic information when compared with word2vec. Consider the existing training samples where each is considered as a sentence.

The Doc2Vec system has two main variants: Distributed Memory (DM)

and Distributed Bag of Words (DBOW). The DM model is an extension of the continuous bag-of-words model used by Word2Vec. As part of the DM strategy, the center word is not only predicted according to the context word vector, but also according to the document vector. A document ID and related context words are inputs into the model, and it tries to predict the center word. By focusing on nearby contexts of words as well as global semantics of the document, this strategy ensures that the learned document vectors reflect the document's overall structure. A sliding window samples fixed-length phrases from a sentence when processing it, and one of those phrases becomes the target word and the others the context. A new vector X is created by averaging or accumulating the context word vectors along with the sentence's paragraph vector. The vector X is then used to predict the target word. Doc2vec differs from word2vec primarily in the way its input layer is built. Doc2vec introduces a new vector called Paragraph vector, a special word vector that serves as a memory mechanism in Doc2vec. It ignores the existence of other words in the sentence when training in the bag-of-words model, which focuses only on a portion of vocabulary in a sentence. Due to this, the generated vector is just a simple accumulation of each word vector, so it is not possible to capture word order in the text. In Doc2vec, the Paragraph vector helps solve this issue. As the Paragraph vector is shared across multiple trainings of the same sentence, even though a part of words are still selected for training through the sliding window, it gradually incorporates the whole

semantic information. Each training session not only vectorizes words, but also continuously optimizes the Paragraph vector so that it reflects the sentence's core meaning more accurately. As with word2vec's CBOW model, Doc2vec's DM model's training mechanism employs stochastic gradient descent methods to adjust model parameters.

With Doc2vec, not only can you get the vector representation of each word, but you can also get the vector representation of each sentence. So, how do you get a vector representation of a new sentence with Doc2vec? As a matter of fact, every new sentence will have its corresponding Paragraph vector randomly initialized and then be added to the pre-trained model. As a result of this process, only the Paragraph vector is iteratively updated, leaving the word vectors in the model and the softmax weight parameters in the projection layer unchanged. Since only the Paragraph vector is updated, the computation time required to predict a new sentence's vector representation is shorter. DBOW is based on the Skip-gram model in Word2Vec. It predicts randomly selected words within a document using document vectors. In other words, the model predicts the vocabulary within the document when given a document ID without relying on specific contextual vocabulary. Rather than drilling down into specific contextual details, this approach is centered around grasping the overall semantic content of the document.

A document can be represented using two different vector representation methods: DM and DBOW. Both of these methods take into account the

semantic content, but have slightly different objectives. This model emphasizes the combination of context and document, as the input is the sentence vector of paragraph ID, which is the same as DM. A document's overall semantics are more important in DBOW. To predict words with multiple classifications, stochastic gradient descent samples a text window every iteration and randomly selects a word from that window to create a Paragraph vector. A similar model to Skip-gram, Doc2vec is built on Word2vec. Doc2vec can predict new sentence vectors in addition to training words and sentence vectors, and is comparable to Word2vec in this respect. The Doc2vec model differs from Word2vec in that an additional Paragraph vector sentence vector is added to the input layer. In each training, a single weight is assigned to this vector. In this way, the trained Paragraph vector will gradually stabilize each sentence after several trainings to form the gist of the sentence. This enables us to train the sentence vectors we require. To predict a new sentence vector, the model must be retrained. In this step, the word vectors and projection layers to output layers of the model are fixed, and a gradient descent method is used to obtain the paragraph vectors only. Consequently, when predicting a new sentence, it must also be incorporated into the model. You can continuously iterate and find it, which will be faster than training. A specific topic is implicitly represented in the obtained text collection. As a result of event generalization technology, the event's core content can be highlighted, allowing for the exploration of potential event

knowledge in a more effective manner. To illustrate the principle, consider a paragraph, a set of sentences, containing the id and event in the event. In order to obtain the vector representation sets X_p and X_w , the Doc2vec model is used to Event Evolution Graph the paragraph vector into a word vector. Although the paragraph vector and word vector have the same dimensions, they come from two separate vector spaces. A sentence's ID and paragraph vector are fixed during model training. In this way, text similarity can be calculated by taking into account the overall context while predicting each word, allowing the semantic information of the entire sentence to be used when calculating text similarity. A starting centroid vector of k sentence samples is then randomly selected from the event, and the silhouette coefficient method is then used to determine the optimal value for k . The cluster division and clustering process is completed when all k centroid vectors are stable and unchanged, and the maximum number of iterations is N . Calculate the semantic similarity between all sentences and all centroids.

4) K-means

To be grouped into the corresponding cluster, sample points must be matched to their nearest cluster centers using K-means clustering. As a result of this method, similar events can be summarized and a series of specific events can be described by a highly representative event, thereby forming an abstract picture of network public opinion. The events are sampled as text samples, or clusters, in this process.

With K-means clustering, n data points are divided into k clusters with a value less than n , with the average of their internal data points representing each class cluster. This algorithm takes a step by step approach. To optimize the cluster center, this algorithm begins by estimating the cluster center, which is then continuously grouped into their nearest cluster center and re-estimated. Until the cluster center has stabilized, this process continues iteratively. While K-means works well in many scenes, it requires a preset number of clusters (k), and the initial cluster center may affect the algorithm's results. To solve these problems, it is often recommended to run the algorithm multiple times, with each time using a different initial cluster center, and then select the optimal solution based on the results of each run. This clustering algorithm divides the dataset by iteratively optimizing the centroid location at the beginning. The centroid is initialized at the beginning of the algorithm. Once the centroid has been identified, each data point has been assigned a centroid based on the nearest neighbor principle. A centroid is updated to reflect its assigned data point's mean value. Until the centroid position is no longer changing significantly or until a predetermined number of iterations have been reached, these two steps (allocation of data points and updating centroids) are repeated. An algorithm's goal is to minimize the distance between all data points and their corresponding centroids, which is commonly measured using Euclidean distances.

5) K-means++

K-means++ is an improved version of K-means designed to reduce the poor clustering effects caused by random selection of starting centroids. Its main objective is to provide a more stable initialization method for centroids. With K-means++, the first centroid is randomly selected in the dataset, and subsequent centroid selection is based on the distance between it and the already selected centroid. Based on the square of its distance to the nearest determined centroid, a data point's probability of being selected as the next centroid is proportional. By using this strategy, the K-means process can be improved in terms of convergence efficiency and clustering effect by maximizing the distance between the initial centroids. With K-means++, events with high similarity can be correctly classified in clustering tasks involving serialized data, such as time series and logs. In large-scale or complex event datasets, it is easier to get high-quality clustering results if you use the K-means++ strategy. As opposed to traditional random selection methods, K-means++'s cluster center selection method is its core advantage. While the initial cluster centers are still randomly selected, the algorithm does not select all K cluster centers at once. First, the algorithm selects one point from the sample set to serve as the first cluster center. After selecting the remaining $n-1$ points, the next cluster center is selected, but the selection process is not entirely random. The selected cluster is based on a mechanism that ensures that points closer to the center are more likely to be selected, while points closer are less likely to be selected. The process continues until

K cluster centers are selected.

Its main advantage lies in its ability to accelerate convergence and enhance clustering. With K-means++, convergence is often achieved in fewer iterations, leading to a more optimized solution because the algorithm selects the initial centroid effectively. Furthermore, K-means++ has a greater possibility of finding a superior local optimal solution due to the rationality of its starting centroid, thereby improving clustering. It is particularly useful when dealing with large data sets or performing K-means multiple times for optimal clustering results.

3.1.6 Abstract Event Evolution Graph

In this study, an Abstract Event Evolution Graph is proposed as a means of showing the correlation between abstract events after generalizing events. However, current clustering methods do not specify representative events for each class cluster, so it proposes an abstract affair Event Evolution Graph to show how abstract events are related. To define labels for each class cluster, this study selected high-frequency words in the class cluster to represent abstract events. In the abstract story Event Evolution Graph, events in the matter Event Evolution Graph are viewed from a high-level perspective, the specific causal pattern is transformed into a more general pattern, and the logical evolution of common events in online public opinion can be seen. Specific events become abstract events after event generalization, and

specific event relationships become universal relationships between abstract events.

An Abstract Event Evolution Graph of public opinion and affairs is constructed based on generalized event causality, with nodes connected by arrows form a chain of causal relationships. A-B and B-C represent causality, where node A causes node B and node C, and they form a causal chain. Nodes A,B,C are precursors to nodes B, and nodes C are successors to nodes B. An arrow pointing in the direction of the development of events indicates that one event results from another.

3.2. Experimental Design

3.2.1 Experimental Hypothesis

Event Evolution Graph provides a clear view of the characteristics and development of network public opinion and offers a new research area to study online public opinion. As a result of the affection Event Evolution Graph, a theoretical framework for managing network public opinion can be developed, allowing appropriate strategies to be adopted for disseminating public opinion, which allows the affair Event Evolution Graph to expand its application scope.

Through the use of text mining, natural language processing, and complex networks, the Event Evolution Graph shows public opinion's development law and path in detail. The Event Evolution Graph compensates

for the limitations of time series and regression analysis research methods by introducing methods such as event reasoning. The construction of this Event Evolution Graph can also help government departments and enterprises understand online public opinion trends. The Event Evolution Graph allows us to predict the direction of online public opinion and take appropriate action. The government will be better able to manage online public opinion, mitigate social risk, and maintain social stability as a result of this.

3.2.2 Experimental Inputs

These long conversations and short sentences include the causal relationship information required for this article. Relevant text comments are gathered from social media based on topic keywords. These texts contain both the text information and the time when the text was posted.

4. Design and Implementation

4.1. Design

4.1.1 Experimental Process

This Event Evolution Graph provides a solid basis for analyzing and predicting the evolution path of public opinion by revealing the origin and development process of Internet public opinion events. Using public comments, views, and opinions on specific online public opinion events, this graph provides the core data. As a result, these comments and opinions are more objective than news texts, reflecting the public's true views. It is, however, necessary to preprocess these unstructured and noisy data. This section introduces data acquisition and preprocessing. Identify causal sentences from preprocessed text and extract events from it. A network public opinion and affairs Event Evolution Graph was constructed using gephi software. Figure 2 illustrates the construction process for the network public opinion and Event Evolution Graph.

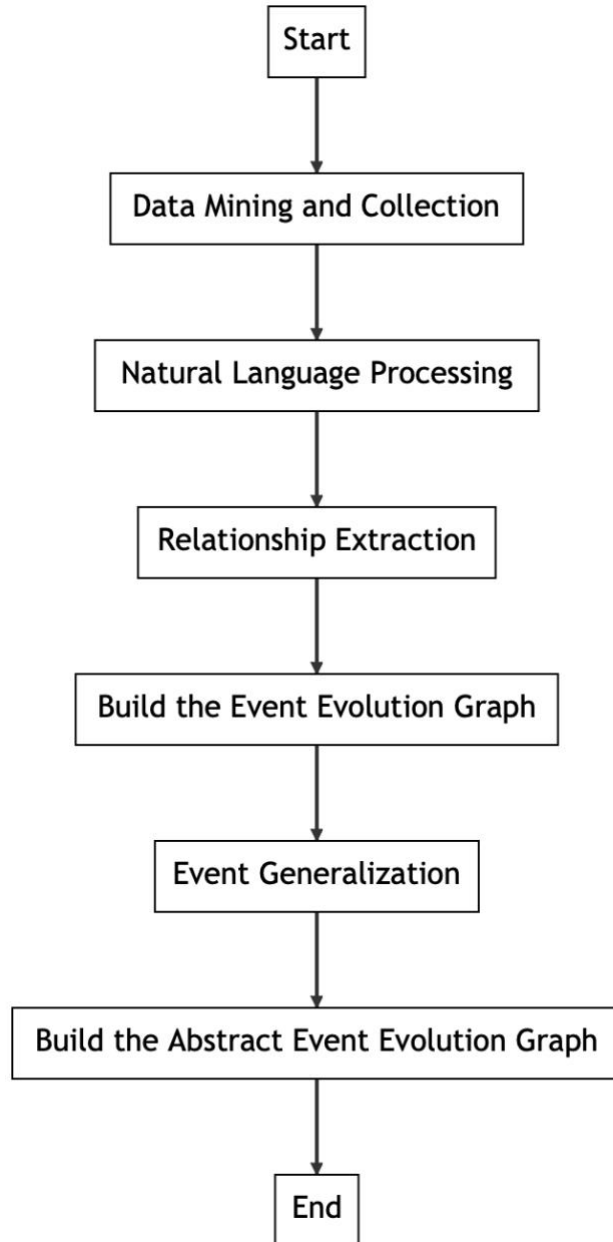


Fig. 2. The Diagram of Experimental Process

Using the data collected from online public opinion events, preprocess data, such as deduplication, denoising, word segmentation, and part-of-speech labeling, then create a network public opinion Event Evolution Graph based on these data. As part of this study, we summarize the existing relationship recognition research and propose a template for identifying

Chinese causal relationships in order to further improve the atlas. Using these rules, causal sentences in text are judged and pairs of events are extracted. This study also proposes a method of predicting online public opinion events, as well as constructing the network public opinion and Event Evolution Graph using Gephi tools. In order to create an abstract Event Evolution Graph of network public opinion, the clustering method is used to generalize specific public opinion events because the Event Evolution Graph based on specific events may not fully reflect the logical evolution relationship between events. Using the weight of future nodes at that node, you can predict the probability of a new event occurring by finding nodes similar to that event in an abstract graph. For empirical research, this study chose specific social media events to verify the efficacy of the above methods.

4.1.2 Data Mining and Collection

Identify the keywords associated with the events that need to be crawled. Each event may have more than one keyword associated with it. Using the keywords, find the BILIBILI URL corresponding to the event, then use that URL as the crawler's initial URL. Stop the crawler until there is no more data on pages containing different result data based on this URL.

4.1.3 Natural Language Processing

Due to the duplication, noise, and other factors involved in crawler data collection, the first preprocessing step is deduplication and denoising of the

data. As part of the preprocessing process, the corpus data is segmented into clauses, words, and parts of speech using the Jieba word segmentation technology.

4.1.4 Relationship Extraction

There are three main methods for recognizing relationships: recognizing rules, recognizing features, and recognizing relationships using machine learning. The prompt word strategy is adopted, the corresponding rule templates are developed, and the relationships in the text are then determined based on these templates. Using machine learning technology, a classifier can be trained to identify the relationship between sentences by using text features and semantic connections; this method identifies the relationships between sentences accurately.

Due to the pluralistic, mostly implicit influence factors in network public opinion systems, the feature-based method of relationship recognition has limitations in this system, and relying only on a single causal relationship for identification can be difficult to achieve the ideal effect. The high efficiency of machine learning methods, on the other hand, is gradually gaining favor among researchers. There have been few studies combining computer graphics and network public opinion recognition, and machine learning methods are often not comprehensive enough to cover specific fields of text. Event Evolution Graph offer us a visual way to illustrate the causal

relationship between public opinion events online, providing a solid foundation for future public opinion predictions. We need to explore the universal causal pattern behind specific events and use artificially constructed rule templates to identify the relationship between them rather than focusing on the specific causal relationships between them. Using rule templates, these events can be identified and generalized into abstract events. With abstract Event Evolution Graph, we are able to discover the universality, frequency, and simplicity behind specific causal relationships, which can help us predict more general outcomes, rather than only concrete results.

When the data is analyzed and compiled according to the syntactic pattern of the Chinese causal sentences, the rule template determines whether it is a causal sentence. Specifically, the rule template follows the format of Pattern, Constraint, Priority >, in which Pattern contains causality trigger words, while Constraint contains syntactic constraints in some sentences. The authors of this paper compiled 9 types of syntactic templates based on syntactic patterns found in existing literature. Some of the templates are listed in Table 1.

Table.1. The Part of Causal Sentence Template

<i>Dependency Syntax</i>	<i>Extraction Rules</i>	<i>Causality Identifiers Word (Example)</i>
<i>By Effect Regression</i>	<i>pattern=re.compile(r's? (%s)/[p c]+\s(.*)(%s)/[p c]+\s(.*)' % (word[0], word[1]))</i>	<i>['The reason is', 'because'], ['The reason is', 'due to'], ['The reason is', 'owing to']</i>
<i>From Cause to Effect</i>	<i>pattern=re.compile(r's? (%s)/[p c]+\s(.*)(%s)/[p c]+\s(.*)' % (word[0], word[1]))</i>	<i>['Because', 'thereby'], ['Because', 'for this reason'], ['Given that', 'so'], ['Due to', 'for this purpose'], ['Unless', 'only then'], ['Only if', 'then'], ['Owing to', 'to the extent that'], ['If', 'then'], ['Due to', 'thus'], ['Because', 'therefore']</i>

It is possible to determine whether a sentence is a causal sentence, using nine rule templates generated from the generated syntactic template, as shown in Table 2. When writing, use the rule template to match the reason and effect parts of the sentence, such as sentence because[sentence1], sentence2], cause event1, and result event2. In part-of-speech labeling, each part of the sentence is divided into two small sentences, from which the cause event and the result event are extracted.

Table.2. The Rule Templates

	Rule Templates
1	$\langle \text{Conj} \rangle \{ \text{Effect} \}, \langle \text{Conj} \rangle \{ \text{Cause} \}$
2	$\langle \text{Conj} \rangle \{ \text{Cause} \}, \langle \text{Conj} \rangle \{ \text{Effect} \}$
3	$\{ \text{Cause} \}, \langle \text{Conj} \dots \rangle \{ \text{Effect} \}$
4	$\{ \text{Cause} \}, \langle \text{Verb} \text{Adverb} \dots \rangle \{ \text{Effect} \}$
5	$\langle \text{Prep} \dots \rangle \{ \text{Cause} \}, \{ \text{Effect} \}$
6	$\{ \text{Cause} \}, \langle \text{Verb} \text{Adverb} \dots \rangle \{ \text{Effect} \}$
7	$\langle \text{Conj} \dots \rangle \{ \text{Cause} \}, \{ \text{Effect} \}$
8	$\{ \text{Effect} \} \langle \text{Prep} \dots \rangle \{ \text{Cause} \}$
9	$\{ \text{Effect} \} \langle \text{Conj} \dots \rangle \{ \text{Cause} \}$

4.1.5 Event Evolution Graph

Graphs representing causal relationships between network public opinion events are known as network public opinion Event Evolution Graph. When a public opinion event is the cause, a target event connected to it is considered the direct consequence of the causal event. In order to create a complete Event Evolution Graph, causality can be connected to form a pair of causal events. Gephi is a popular open source web analysis tool that can be used to create and analyze multiple networks. A spreadsheet with the identified pairs of events was first constructed in this study, and then the data was imported into Gephi for verification to make sure that there was no overlap. The transitivity of causality can be automatically identified with

Gephi so that a graph can be automatically generated based on whether the source event is the target event for other event pairs.

4.1.6 Clustering Algorithm Improvement Evaluation

Additionally to Events and Denoters, CEC uses XML as the annotation format, which has six important data structures (tags): Time, Location, Participant, and Object. An event is described by the Event variable; the Denoter variable indicates the time, place, participants, and objects indicate the indicators and features of that event. Additionally, the CEC corpus, despite its small size, features comprehensive event and element labeling. A class of events, then modeled accordingly as event ontologies, is composed of their relationships and composition, and then they are analyzed as the cognitive unit of event ontology. To study the ontology, it is necessary to build a corpus first, so the emergency corpus is selected on the Internet for the event labeling of the corpus, and the classification system of emergencies includes three levels: 4 categories at the first level (natural disaster N, accident disaster A, public health events P, social security events S), 33 subcategories at the second level, 94 subcategories at the third level, mainly including five categories: earthquakes, fires, traffic accidents, terrorist attacks, and food poisoning. A total of 332 articles.

The original algorithm can pose a number of problems when it comes to event clustering. When generalizing events, the original algorithm is

improved in two ways, by using both word vectors and clusterings, and then using the CEC data set as a benchmark, these algorithms are combined for testing, and then compared to the annotated data based on the cluster classification of each event in CEC, and the indicators are calculated. All events in the CEC data set are used as clustering sample points, respectively using TF-IDF + K-means clustering method, based on Word2vec + K-means clustering method, based on Doc2vec + K-means clustering method, based on TF-IDF + K-means++ clustering method, based on Word2vec + K-means++ clustering method, based on Doc2vec + K-means++ clustering method generalizes events, and compares the generalization results with artificial clustering results to evaluate the accuracy of the improved algorithm.

As part of clustering, the goal is to reduce the similarity between clusters and increase the similarity within clusters. The sample points should be at the closest distance to the cluster center, and the sample points should be at the closest distance to the cluster center. For a sample set S , which contains n sample points, arbitrarily select two different sample points M, N , use C to represent the results given by the clustering algorithm, and P to represent the results of clustering labeled by the CEC dataset, for these sample points, there are four relationships: a: M, N belong to the same cluster in C and P ; b: M, N are in the same cluster in C and in different clusters in P ; c: M, N are divided into different clusters in C and the same clusters in P ; d: M, N are in different clusters in C and P . where a, b, c, d represent the corresponding numbers of

these relationships.

1) In order to determine whether classification models predict positive classes accurately, precision is an important metric. This metric measures the percent of samples that the model predicts as positive when they actually are. A mathematical definition of precision can be expressed as the ratio between a true class (TP) and its sum with a false positive class (FP). The TP represents the size of the actual sample that was correctly marked as positive by the model, while the FP represents the number of samples that were actually negative but were mistakenly marked as positive. A high accuracy rate indicates the model is highly accurate in predicting positive outcomes.

2) A classification model's ability to detect positive samples is evaluated by recall, which is a metric that measures how many of the actual positive class samples are correctly classified as positive. In mathematic terms, recall is the ratio between the true class (TP) and the false negative class (FN). As a result, TP represents the number of samples that are positively marked by the model, while FN represents the number of samples that are positively marked by the model but mistakenly labeled as negative. An accurate labeling of positive samples indicates a high recall value.

3) In addition, F1-scores are a comprehensive evaluation index that reflects the comprehensive effect of Precision and Recall on a classification model. As its name implies, this metric aims to achieve a balance between the two measures in one single metric. The F1-score is especially important in

scenarios with uneven data distribution because it prevents misunderstandings caused by high and low measures. F1-score values above 50 indicate that recall and precision results are satisfactory.

4) FMI calculates the geometric mean between clustering results and actual category labels as an indicator of the effects of the clustering algorithm. When the FMI reaches 1, the output matches the actual class label exactly. FMI is a number that ranges from 0 to 1, with 1 being the ideal clustering effect and 0 being the least ideal case.

5) In situations where the actual category labels are known, Rand metrics can be used to evaluate clustering algorithms. The indicator measures the percentage of all sample pairs correctly clustered. Considering two sets of sample pairs: one derived from the clustering results, the other based on the category labels. This equation calculates the degree of consistency between these two sets of sample pairs, which is whether they are both in the same cluster or scattered in different clusters. Mathematically, Rand has a range of values from 0 to 1. As the Rand value indicates, when there is a perfect match between the cluster output and the actual category labels, while at 0 there is a very poor match.

6) As a measure of similarity between two sets, the Jaccard metric calculates the ratio between the intersections of two sets and the unions of them. A Jaccard Index is often used in clustering and classification tasks to assess a model, particularly if the true class labels are known. Usually, a

Jaccard Index value near 1 indicates that the two collections are very similar; a Jaccard Index value near zero indicates that there is little similarity between them. When the Jaccard Index of two collections is 1, they have only one element in common, while when the Jaccard Index of two collections is 0, they have no elements in common.

The CEC dataset consists of 322 events clustered using the six clustering techniques discussed earlier. The final evaluation indicator score is derived from the average clustering results for each technique as shown in Table 3.

Table.3. Clustering Algorithm Improvement Evaluation

	<i>Precision</i>	<i>Recall</i>	<i>F1 scores</i>	<i>FMI</i>	<i>Rand</i>	<i>Jaccard</i>
<i>TF-IDF + K-means</i>	93.7872	94.2265	94.0607	90.2018	92.7134	90.3083
<i>Word2vec + K-means</i>	97.0118	97.1108	97.2564	94.7439	94.6504	94.8197
<i>Doc2vec + K-means</i>	92.1238	92.5225	92.2522	84.6576	87.2596	85.0126
<i>TF-IDF + K-means++</i>	94.1132	94.7742	94.3050	91.6130	93.1050	91.5791
<i>Word2vec + K-means++</i>	97.8446	97.9375	97.8862	95.8020	94.7234	95.8761
<i>Doc2vec + K-means++</i>	92.7044	93.0524	92.7627	85.2432	87.2430	86.1324

TF-IDF is based on the concept that words appear more often in certain documents and less frequently in others, which is why words that appear more frequently in one document have a higher degree of differentiation for text classification. A TF-IDF-based event representation, however, has obvious limitations: events consist of multiple words containing different numbers, resulting in a sparse vector of events. Furthermore, while the strategy emphasizes the importance of word frequency, it ignores the semantic

meanings of words, which might result in semantically similar events being far away during the calculation despite the emphasis on word frequency. When constructing word vectors, Word2vec considers the context information of words, which is an effective complement to TF-IDF.

Word2vec represents events significantly more efficiently than K-means clustering using TF-IDF. The F value of the Word2vec method is significantly higher than that of the K-means clustering technique. Additionally, the Rand value, the Jaccard coefficient, and the FMI value have increased significantly. These results indicate that Word2vec is more efficient at generalizing events than TF-IDF. Despite its slightly inferior event generalization, the improved algorithm based on Word2vec is not significantly different from the algorithm based on Doc2vec, despite its slightly inferior event generalization. A comprehensive analysis of the data table shows that the K-means++ clustering technique outperforms the K-means algorithm in terms of performance, resulting in the decision to adopt Word2vec + K-means++ for the event generalization task.

4.1.7 Abstract Event Evolution Graph

It is possible to classify similar events into clusters after processing them by an improved clustering algorithm. The abstract affection graph does not specify representative events for each class cluster due to the fact that this method does not specify representative events for each class cluster. Each

class cluster must be assigned a label to indicate the interrelationships of abstract events.

As part of this study, we constructed events based on words that appear frequently in each class cluster to represent abstract events reflected in the class clusters. At a higher level, the abstract atlas describes events' causal relationships, transforming concrete causal patterns into general ones. As a result of this transformation, an event's logical evolution in online public opinion is revealed. We can further transform a Event Evolution Graph into an abstract Event Evolution Graph by generalizing concrete events into abstract events, thereby transforming the relationship between concrete events into a universal relationship between abstract events.

4.2. Implementation

4.2.1 Data Mining and Collection

The Octopus software crawled the video commentary text on bilibili video website for the month of January 2023 through the present, and retrieved 3,926 texts, some of which are displayed in Table 4.

Table.4. The Data Text

	<i>Time</i>	<i>Text</i>
<i>1</i>	2023-07-26 17:03	<i>How to change the local fiscal mode to an operational one? Investment competition among local governments. Local governments collect direct taxes through corporate taxation, and local governments in areas without property collect direct taxes. How to deal with local debt? The central government rescued. The local government bank bears the part of the revaluation loss by revaluation of the acquisition of local assets.</i>
<i>2</i>	2023-07-21 04:37	<i>In fact, debt is not terrible, terrible distribution system, it is not clear that macro calculation money cannot create consumption, because too few flows into the hands of those who create debt, total factor productivity declines, and it is not sustainable at all.</i>
<i>3</i>	2023-07-07 18:13	<i>If money is sent directly without organizing production, no assets are formed to generate inflation; The currency of goods needs to expand at the same time, not the currency; Key distribution, labor-capital distribution ratio; Otherwise, macro liabilities become cost transfers; The state should issue treasury bonds to engage in the collective economy of the state-run economy and rural revitalization.</i>

4.2.2 Natural Language Processing

Similar events are classified into the same class of clusters after processing by an improved clustering algorithm. The abstract affection graph does not specify representative events for each class cluster because we need

to assign a label to each class cluster in order to display the interrelationships among abstract events. This study constructed abstract events by selecting words that appear frequently in each class cluster.

This transformation reveals the common pattern of event logical evolution in online public opinion by transforming concrete causal patterns into more general causal patterns at a higher level. An abstract Event Evolution Graph describes causal relationships at a higher level, transforming concrete causal patterns into more general causal patterns. We can transform the story map into an abstract Event Evolution Graph by generalizing concrete events into abstract events, thereby transforming the relationship between concrete events into a universal relationship between abstract events.

4.2.3 Relationship Extraction

In Table 5, we find 1372 causal event pairs based on the description of the event, and each pair of events can be connected to form a connecting edge. The events with only one word, for example "get" and "think," were excluded because they are difficult to convey the meaning of a complete event clearly. There is a direct causal relationship between each pair of events. In this study, we call the network structure the theory map, since causality is transitive and causes and effects are connected when two pairs of events share the same cause and effect.

Table.5. The Part of Examples of Causal Sentences

	<i>Cause</i>	<i>Tag</i>	<i>Effect</i>
<i>1</i>	<i>Too few inflows into the hands of creating indebtedness, and total factor productivity declines.</i>	<i>Because</i>	<i>Not sustainable at all.</i>
<i>2</i>	<i>A certain link in the economic circulation chain is blocked.</i>	<i>If</i>	<i>Debt is increasing, and the problem of procrastination is widening.</i>
<i>3</i>	<i>Central banks should expand their balance sheets, issue bonds and give money to citizens.</i>	<i>Promote</i>	<i>Consumption of excess products.</i>

4.2.4 Event Evolution Graph

The above causal events are imported into the software Gephi, and the software automatically draws a map according to the node file and edge file, and sets the gravity value according to the concentration, and gathers important events and relationships to the center, that is, the general event center topic and secondary topic can be seen, and the general evolution law of network public opinion can be obtained, as shown in Figure 3.

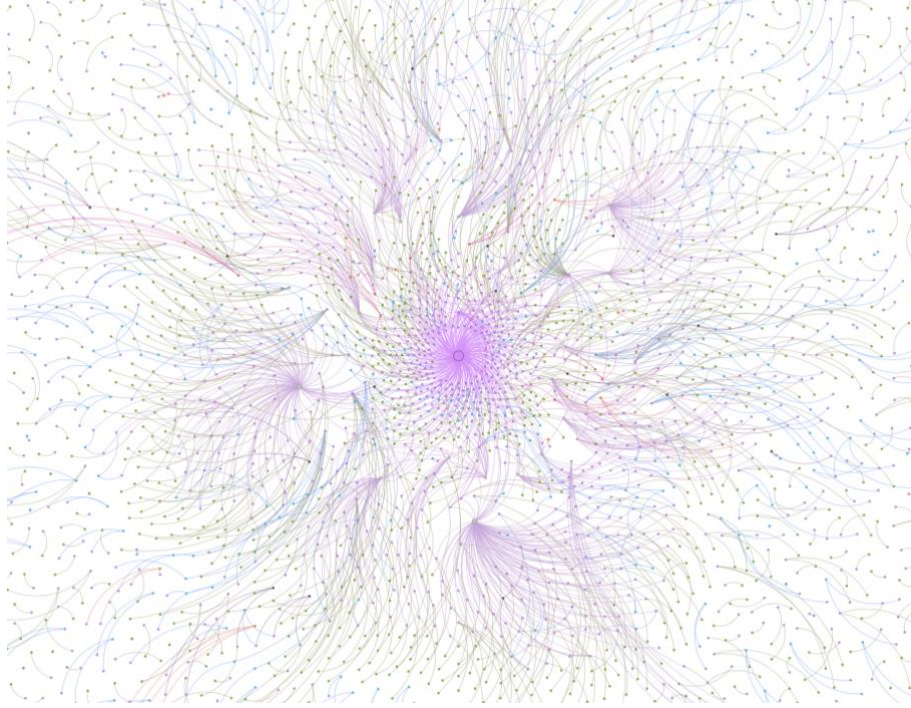


Fig. 3. The Event Evolution Graph

4.2.5 Abstract Event Evolution Graph

In order to generalize event data, we chose a K-means++ clustering algorithm that was optimized using word2vec. As shown in Figure 4, the elbow method was used to experiment with K values ranging from 2 to 15, resulting in six clustered clusters. The K value was determined to be 6, which indicates that there are six clustered clusters in our dataset. By segmenting words and counting word frequencies, we are able to create labels for six events within each cluster. As a result of the generalization of the incident, six main nodes were formed on the medical network's public opinion and affairs map. Figure 5 provides an abstract picture of network public opinion based on causal relationships between specific events in the graph.

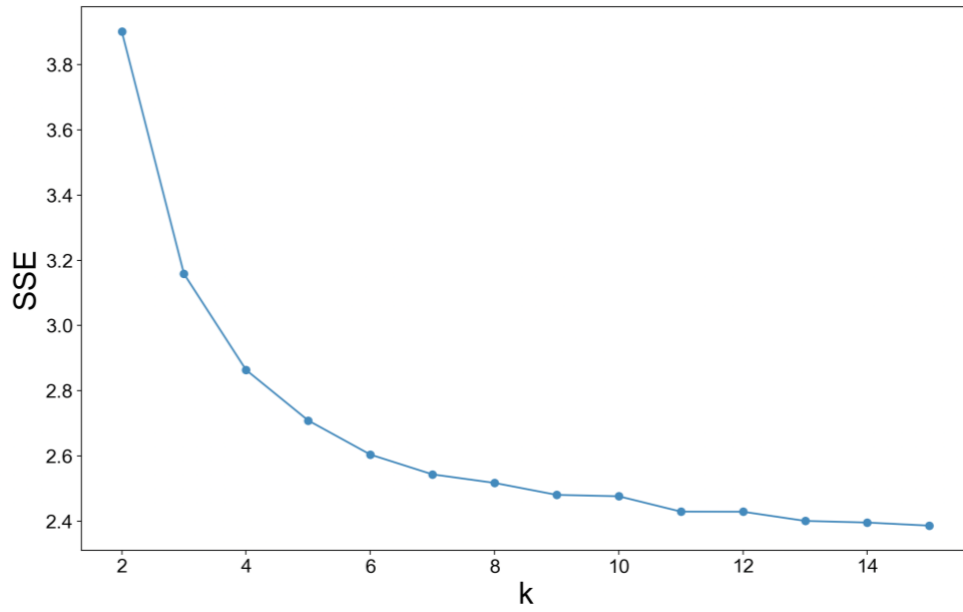


Fig. 4. The K Value

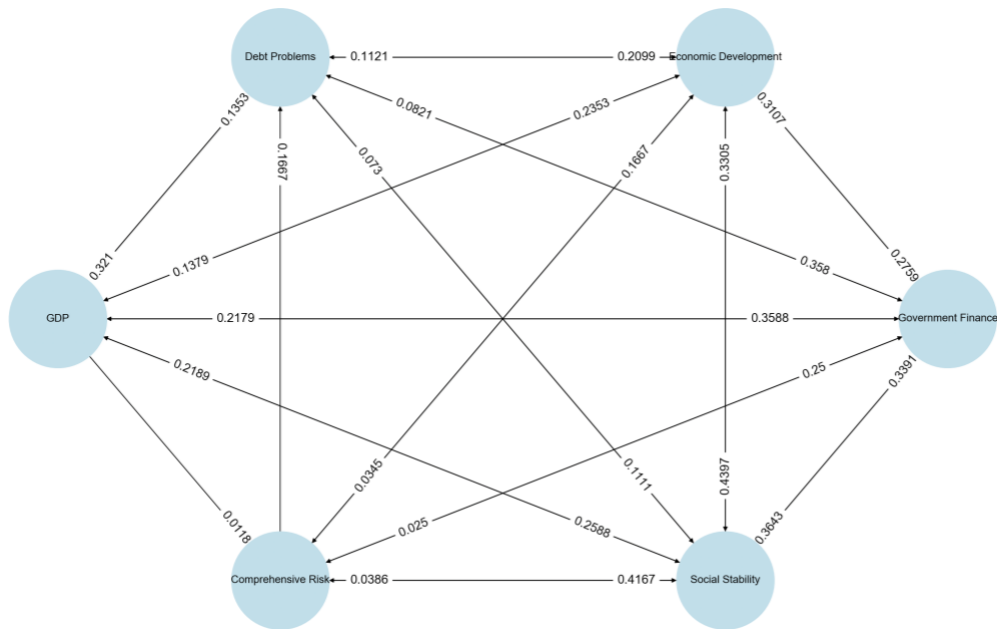


Fig. 5. The Abstract Event Evolution Graph

5. Analysis

This paper proposes an online public opinion research method based on events graph to help public opinion workers better control the development of online public opinion. A visual map depicting the logical evolution relationship between network public opinion events is constructed by analyzing the causal relationship between public opinion events, extracting causal event pairs from it, and summarizes the evolution characteristics and evolution path of network public opinion as a result. As a result of the improved clustering algorithm, events are generalized and abstract Event Evolution Graph are constructed, which establishes a solid foundation for prediction of public opinion in networks. The proposed network public opinion prediction method is then empirically tested for its accuracy and effectiveness.

It describes methods for creating network public opinion graphs. Researchers primarily conduct research using foreign text data, but Chinese texts differ considerably from foreign texts in terms of causal relationships. This article elaborates on the construction process of the network public opinion map. Chinese texts express causal relationships in a richer and more diverse way than foreign texts.

We summarize and analyze causal sentences' characteristics, provide a syntactic template for identifying causal sentences, and describe a method for

extracting text events based on previous research. To achieve better event generalization, an improved clustering algorithm is proposed to enhance the application scope of the affairs map and to serve as a reference for scholars conducting research on it. Because of the methods of event representation, distance calculation, and centroid update, the original clustering method has deficiencies such as missing semantic factors, uneven clustering results, and poor accuracy. By using word2vec as an event representation, using event similarity as distance between events, and updating the centroid method, this paper improves the K-means++ clustering algorithm to achieve better event generalization. Experiments have shown that the clustering method based on word2vec is better than the clustering method based on TF-IDF + K-means, the clustering method based on Word2vec + K-means, the clustering method based on Doc2vec + K-means, and the clustering method based on TF-IDF + K-means++. The accuracy of clustering methods and clustering methods based on Doc2vec + K-means++ is improved.

Through the results, as shown in Figure 5, we can see the causal logic between "Debt Problems", "Economic Development", "Local Government Finance", "Comprehensive Risk", "Social Stability" and "GDP", and the probability of occurrence, and we can get the context in the evolution of network public opinion and know the relationship between the main topics.

6. Results and Discussion

Using the characteristics of the evolution of network public opinion, this paper examines how events are constructed, how relationships are extracted, how events are generalized, and how events are visualized in a graph. Getting causally logical information from online public opinion is successful, as it is consistent with previous research and the expectations of experimental hypotheses.

In order to discover how events evolve, causality can be identified dynamically. It is possible to identify the variable relationship between events when public opinion occurs through the evolution path of causality, and then analyze the whole process of public opinion generation, development, rise, and retreat based on that relationship. It is possible to identify the root cause of an outbreak accurately by tracking the development path of causal events, and by tracking the development trend, relevant departments can take action and warn. It is especially important to control the development of an incident when emotions of netizens fluctuate greatly to prevent a crisis from occurring.

7. Conclusions and Recommendations

There are multiple levels of dissemination and complexity in the evolution of online public opinion. The public opinion incident itself plays a significant role in the multi-level nature of dissemination. As a result of the evolution of various causes and events over time, there is a relationship between the inheritance of the network public opinion and affairs map. Among those events can easily result in secondary crises arising. At the same time, different time nodes and the evolution of follow-up events can easily lead to different directions in public opinion evolution. In this way, the relevant departments can analyze the public opinion development trend at the first point of study, monitor the development of public opinion, and eliminate the negative effect of negative public opinion events through the exploration of key nodes in the network public opinion map.

In spite of this, this article has some shortcomings, and future studies are needed. The clustering methods in this paper cannot be compared because there are fewer corpora for network public opinion emergencies, but you can crawl more sentences in the future, expand the corpus, and use the marked corpus in the appropriate field. In addition, a new rule template method based on machine learning must be integrated with the existing relationship extraction method and a new method proposed for comparison.

8. Reflection

In this study, the method in the process of generating the Event Evolution Graph needs to be further optimized, so as to reduce the generation of free nodes in the Event Evolution Graph, find a more linked causal chain, and find other logical relationship extraction methods to generate Graph of different relationships and enrich the semantic relationships of the Event Evolution Graph.

It is currently common for online public opinion prediction methods to focus on analyzing public opinion evolution trends and scales, but they cannot accurately predict subsequent events of online public opinion, leading enterprises and governments to be passive when responding. This paper proposes an abstract Event Evolution Graph based method for predicting network public opinion evolution in order to solve this problem. This method allows relevant departments to take measures in advance based on the prediction results, provide more active guidance in the direction of online public opinions, and achieve real prevention of government corruption before it takes place.

References

- [1] X. Ma, P. Xue, Y. Chen, et al., "Construction and trend analysis of social media crisis theme evolution model," *Journal of Library and Information Work*, vol. 65, no. 13, pp. 77-86, 2021.
- [2] J. Kong, G. Teng, S. Wang, et al., "Research on the impact of public opinion respondents' responses on netizens' emotions," *Journal of Library and Information Work*, vol. 64, no. 18, pp. 89-96, 2020.
- [3] X. F. Dong, Y. Lian, X. Y. Tang, et al., "The Damped Oscillator Model (DOM) and Its Application in the Prediction of Emotion Development of Online Public Opinions," *Expert Systems with Applications*, vol. 148, pp. 113268, 2020.
- [4] S. Stieglitz, M. Mirbabaie, B. Ross, et al., "Social Media Analytics – Challenges in Topic Discovery, Data Collection, and Data Preparation," *International Journal of Information Management*, vol. 39, pp. 156-168, 2018.
- [5] S. Chen, J. Mao, G. Li, et al., "Uncovering Sentiment and Retweet Patterns of Disaster-related Tweets from a Spatiotemporal Perspective – A Case Study of Hurricane Harvey," *Telematics and Informatics*, vol. 47, pp. 101326, 2020.
- [6] Z. Sun, H. Zhao, Z. Wang, "Research progress on online public opinion and its theme association relationship path analysis," *Journal of Library and Information Work*, vol. 65, no. 7, pp. 143-154, 2021.
- [7] Z. Y. Xing, X. H. Su, J. M. Liu, et al., "Spatiotemporal Change Analysis of Earthquake Emergency Information Based on Microblog Data: A Case Study of the

- “8.8” Jiuzhaigou Earthquake," *International Journal of Geo-information*, vol. 8, no. 8, pp. 359, 2019.
- [8] G. Atluri, A. Karpatne, V. Kumar, "Spatio-temporal Data Mining: A Survey of Problems and Methods," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1-41, 2018.
- [9] X. Tang, "Research on online public opinion of sudden public events in China," Ph.D. dissertation, University of Electronic Science and Technology of China, Chengdu, 2008.
- [10] X. Su, B. Spierings, M. Dijst, et al., "Analysing Trends in the Spatiotemporal Behaviour Patterns of Mainland Chinese Tourists and Residents in Hong Kong Based on Weibo Data," *Current Issues in Tourism*, vol. 23, no. 12, pp. 1542-1558, 2020.
- [11] K. Pichotta and R. J. Mooney, "Using Sentence-Level LSTM Language Models for Script Inference," 2016.
- [12] X. Chen, M. Xia, J. Cheng, et al., "Trend prediction of internet public opinion based on collaborative filtering," in *International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, IEEE, 2016, pp. 583-588.
- [13] F. Zhang, G. Si, and P. Luo, "Research on psychological models in information dissemination modeling and simulation," *Computer Simulation*, vol. 30, no. 2, pp. 165-169, 2013.
- [14] K. C. Wong, "The making of the USA PATRIOT Act II: Public sentiments, political gamesmanship, media patriotism," *International Journal of the Sociology of Law*, vol. 34, pp. 105-140, 2006.

- [15] B. W. Edes, "The role of government information officers," *Journal of Government Information*, vol. 23, pp. 455-469, 2000.
- [16] A. Ittoo and G. Bouma, "Minimally-supervised extraction of domain-specific part-whole relations using Wikipedia as knowledge-base," *Data & Knowledge Engineering*, vol. 85, no. 3, pp. 57-79, 2013.
- [17] Z. Kozareva, "Cause-effect relation learning," in *Workshop Proceedings of TextGraphs-7 on Graph-based Methods for Natural Language Processing*, Association for Computational Linguistics, 2012, pp. 39-43.
- [18] A. Sorgente, G. Vettigli, and F. Mele, "Automatic extraction of cause-effect relations in natural language text," 2013, pp. 37-48.
- [19] S. Zhao, Q. Wang, S. Massung, et al., "Constructing and Embedding Abstract Event Causality Networks from Text Snippets," 2017, pp. 335-344.
- [20] Y. Li, W. Mao, and F. Wang, "Causal knowledge extraction for domain open source text," *Computer Engineering and Science*, vol. 32, no. 5, pp. 100-104, 2010.
- [21] J. Qiu, Y. Du, and Y. Wang, "Extraction and Representation of Feature Events Based on a Knowledge Model," in *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, IEEE Computer Society, 2008, pp. 219-222.
- [22] L. Li, Y. Dang, J. Zhang, et al., "Research on Chinese entity relation extraction based on combined kernel," *Journal of Information Science*, vol. 31, no. 7, pp. 702-708, 2012.
- [23] X. Guo, T. He, X. Hu, et al., "Chinese entity relation extraction based on syntactic

- semantic features," *Journal of Chinese Information Processing*, vol. 28, no. 6, pp. 183-189, 2014.
- [24] S. Zhao, T. Liu, S. Zhao, et al., "Event causality extraction based on connectives analysis," *Neurocomputing*, vol. 173, no. P3, pp. 1943-1950, 2016.
- [25] M. Riaz and R. Girju, "Another look at causality: discovering scenario-specific contingency relationships with no supervision," in *IEEE Fourth International Conference on Semantic Computing*, IEEE, 2010, pp. 361-368.
- [26] A. Ittoo and G. Bouma, "Extracting explicit and implicit causal relations from sparse, domain-specific texts," in *International Conference on Natural Language Processing and Information Systems*, Springer-Verlag, 2011, pp. 52-63.
- [27] K. Radinsky, S. Davidovich, and S. Markovitch, "Learning causality for news events prediction," in *International Conference on World Wide Web*, ACM, 2012, pp. 909-918.
- [28] A. Sorgente, G. Vettigli, and F. Mele, "Automatic extraction of cause-effect relations in natural language text," vol. 1109, pp. 37-48, 2013.
- [29] N. Chambers and D. Jurafsky, "Unsupervised learning of narrative event chains," in *Proceedings of ACL-08: HLT*, 2008, pp. 789-797.
- [30] G. Glavas and J. Snajder, "Construction and evaluation of event graphs," *Natural Language Engineering*, vol. 21, no. 4, pp. 607-652, 2015.
- [31] X. Shan, S. Pang, X. Liu, and J. Yang, "Analysis of the evolution path of online public opinion based on the event-reason graph: Taking medical public opinion as an example," *Information Theory and Practice*, vol. 42, no. 09, pp. 99-103+85,

2019.

- [32] R. Girju, "Toward social causality: An analysis of interpersonal relationships in online blogs and forums," in Proceedings of the ICWSM2010, Menlo Park: AAAI Press, 2010, pp. 66-73.
- [33] D. Kinzler and L. Pintscher, "Wikidata," in The International Symposium, 2014.
- [34] A. E. Auer, "DBpedia: A Nucleus for a Web of Open Data," in Semantic Web, International Semantic Web Conference, Asian Semantic Web Conference, Iswe + Aswe, Busan, Korea, November, 2007.
- [35] G. C. Mai, B. Yan, K. Janowicz, et al., "Relaxing unanswerable geographic questions using a spatially explicit knowledge graph embedding model," in Proceedings of the 22nd AGILE Conference on Geographic Information Science, Cham: Springer, 2020, pp. 21-39.
- [36] H. Kim, "Building a K-Pop knowledge graph using an entertainment ontology," Knowledge Management Research & Practice, vol. 15, no. 2, pp. 305-315, 2017.
- [37] Y. Chi, Y. Qin, R. Song, et al., "Knowledge graph in smart education: A case study of entrepreneurship scientific publication management," Sustainability, vol. 10, no. 4, pp. 995, 2018.
- [38] M. Rospocher, M. V. Erp, P. Vossen, et al., "Building event-centric knowledge graphs from news," 2016.
- [39] S. Gottschalk and E. Demidova, "EventKG: A Multilingual Event-Centric Temporal Knowledge Graph," in European Semantic Web Conference, 2018.
- [40] C. S. M. Granroth-Wilding, "What Happens Next? Event Prediction Using a

Compositional Neural Network Model," 2016.