Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

Research on Network Video Hot Event Discovery System

Ying Li*, Jiale Yuan

New Media College, Communication University of Zhejiang, Hangzhou, China
**Corresponding author email:916029788@qq.com*
1970581092@qq.com

ABSTRACT. *In this paper, the network video hot event discovery system integrates domestic mainstream network video website data (e.g. Youku, Tudou, Ku6, etc.) and Internet TV data. This research studies the bottom layer of network video multi-source information extraction, filtering, aggregation, storage, etc. To realize video semantic analysis and processing, the research further studies the multi-layer semantic unified representation model of multi-source heterogeneous video data. On this basis, aiming at providing decision support for government departments to monitor video public opinion and emergency intervention, the research also studies network emergency discovery technology, hot event real-time tracking technology, and hot development trend prediction technology. Finally, relying on the above research results, the research develops a cloud computing-based network video hot event monitoring and analysis system and then demonstrates the application in network media.*

KEYWORDS: *Public Opinion Analysis; Hot Events; Network Video*

1. Introduction

With the rapid development of the domestic network video industry in recent years, the network video has become one of the most popular network services in China. According to the 43rd "China Internet Development Report" issued by the China Internet Network Information Center in February 2019, China's network video users reached 612 million by the end of 2018 and the network audio-visual content has become a spiritual and cultural consumption of China's Internet users. Hot events caused by network video and the video transmission of hot events are increasing year by year. [1] Not only has it become a new trend in network video marketing, but also an urgent issue facing the current research on network public opinion.

The multi-source heterogeneous information contained in the network video and the behavior information of the video users form an important channel for data

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

acquisition in the era of big data. The mining and application of multi-source video big data are getting more and more attention from academia, industry and government departments. [2][3][4][5].For example, Netflix mined video big data to discover user needs, and launched the "House of Cards" episode with great commercial success[6].Hot spot event discovery is the application of topic discovery and tracking (TDT) technology in the actual field [7] [8].

In recent years, the researches on text data mining such as news, forums, and microblogs [9] [10], monitoring hots and sensitive events [11], and analyzing the opinions and sentiments of netizens [12] [13] [14] have emerged in large numbers , trying to meet the urgent needs of government departments for public opinion monitoring, early warning and intervention [15] [16] [17].

Existing research mainly relies on the integration of computer science and journalism. The research on network hot events is basically focused on the grabbing analysis of text information. For example, by studying message propagation models and viral infection models, Wang Wei et al. [18] proposed a topic propagation model based on Weibo fan relationships, user activity, and influence in order to distinguish which users are more likely to be affected and which users are in the "immune" state in the process of topic communication, and then predict the spread of topics or public opinion. Li Biao analyzed the basic structural characteristics of hot event propagation in cyberspace from the aspects of centrality, intermediary, density and subgroups [19] and studied the attention duration and critical threshold of the six stages of hot events in the incubation period, outbreak period, spread period, repeated period, mitigation period, long-tail period [20]. Zhong Ying and Yu Xiucai analyzed the propagation law of the major network public opinion events in time, space, groups involved, content involved, information sources, and dissemination from the perspective of network public opinion guidance. [21].

There are few researches and technical application of current network video hot events. The audio-visual language symbols of video have the characteristics of image, intuitiveness, multi-dimensionality, complexity, and ambiguity that are different from linear text language. Analysis techniques based on text discourse still have many limitations and can’t meet the needs of network video hot event analysis. Cloud computing-based network video hot event monitoring and analysis system can be more effective and more accurate.

2. System architecture and design

2.1 Basic principles

Firstly, cloud computing-based network video hot event monitoring and analysis system collects network video content information and user information from video websites, Internet TV stations, and social media. Secondly, it intelligently analyzes and extracts relevant information, and implements automatic deduplication and denoising. Thirdly, through the cloud storage architecture design to achieve big data management functions. And then through BoF modeling, LDA modeling, semantic

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

feature extraction and other technologies to achieve the unified representation of multi-source heterogeneous data. Finally combined with the network video hot spot aggregation analysis engine and network video hot event monitoring and analysis application platform, to realize the analysis and monitoring of hot events on the Web application site based on B / S architecture and APP applications based on IOS or Android systems.

2.2 System functional architecture design

The functional architecture design of the cloud-based network video hot event monitoring and analysis system is shown in Figure 1.

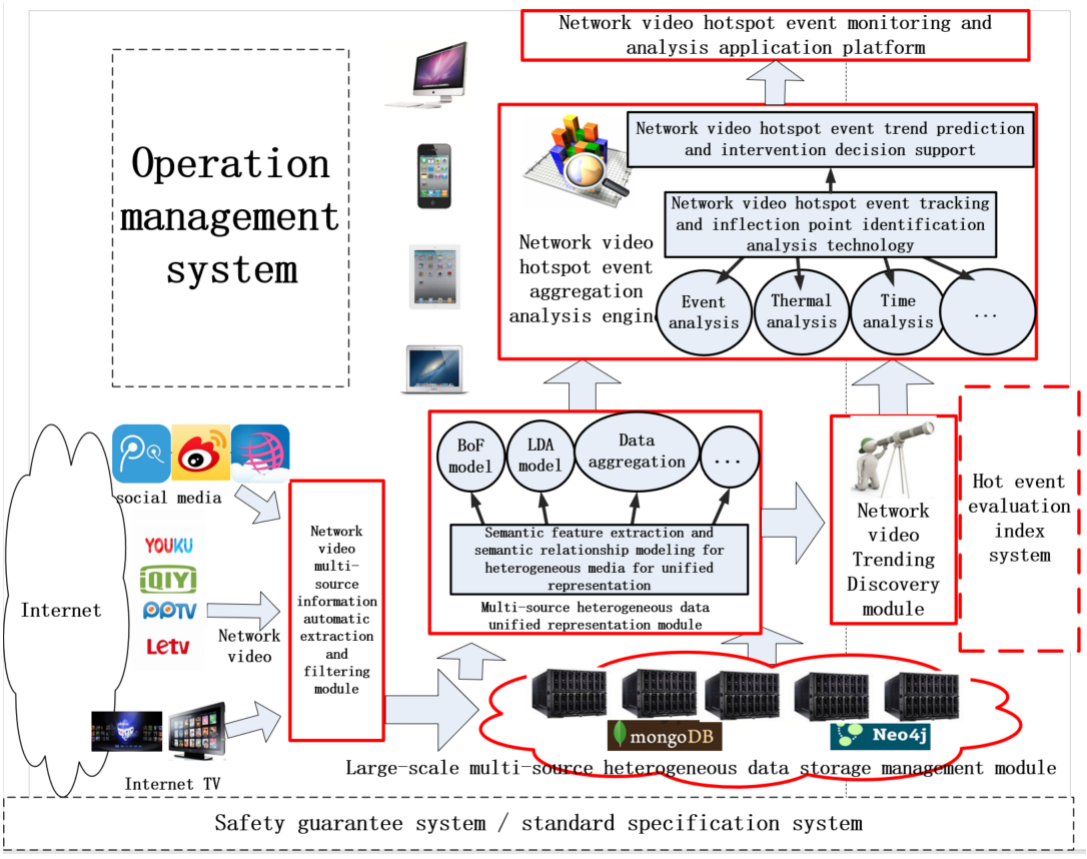

Figure 1 System Total Function Architecture Design

The main functional modules of the system include:

(1) Automatic extraction and filtering module of network video multi-source

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

information

The module is primarily used to capture content information (e.g. text, images, videos, etc.)of web videos from massive, irregular video sites, network television stations, and social media, user behavior information (e.g. clicks, downloads, favorites, reviews, sharing, invitations, labels, uploads, scoring, etc., It even includes page dwell time, scrolling, hovering, fallback, etc.),and intelligently analyzes and extracts relevant information, and automatically de-sicsizes noise, including semantic element extraction, semantic network construction, multi-channel monitoring, spam filtering, consistency analysis, incremental extraction and other functions.

(2) Large-scale multi-source heterogeneous data storage management module

The module is at the data layer, enabling efficient storage management of multi-source heterogeneous data. It includes functions such as Semantic Web distributed storage and indexing, high-speed parallel Semantic Web access, Semantic Web evolution process management, and multiple NoSQL database storage management. The module adopts cloud storage architecture design to realize big data management function.

(3) Unified representation module of multi-source heterogeneous data

This module realizes the unified representation of multi-source heterogeneous data. Including BoF modeling, LDA modeling, topic distribution graph construction, semantic feature extraction, data aggregation and other functions.

(4) Network video hot spot discovery module

This module mainly realizes the construction and management of the evaluation index system of network video hot events, and intelligent discovery of network video hot events.

(5) Aggregation analysis engine for network video hots

This module is at the core layer of the system, and realizes the tracking and prediction of network video hots and sensitive events. Contains functions such as event analysis, heat analysis, time analysis, event tracking and event prediction.

(6) Network video hot monitoring and analysis application platform

This module is a terminal application program for monitoring and analyzing network video hot events, including Web application sites based on B / S architecture and APP applications based on IOS or Android systems.

3. System technical route

3.1 Hot event semantic element extraction and network video multi-source data extraction and filtering

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

This part focuses on the extraction, filtering and storage management technologies of network video multi-source data from the bottom data layer. First, it researches the semantic element extraction technology and semantic web representation model of hot events. Then under the guidance of the semantic web, it extracts and filters the video multi-source data (video content, user behavior and video sources) related to hot events; Finally, it achieves efficient management of multi-source heterogeneous data with the help of cloud database in NoSQL data management and the study of cloud database management key technologies.

(1) Extraction of semantic elements of hot events and construction of semantic web

Multi-source heterogeneous media contains a large amount of high-dimensional, strong correlation and multi-semantic heterogeneous information. The research analyzes the correlation of this information can unearth many potential hot events. The research also studies how to quickly perceive hot events from massive multi-source heterogeneous media information and extract its semantic elements to construct a semantic web. First of all, collecting a large amount of multi-source heterogeneous data, mine the correlation of the multi-source heterogeneous data on the original attributes from the sample data, and eliminate conflicts, remove redundancy, and simplify the expression of data; On this basis, the fusion and knowledge reasoning of multi-source heterogeneous data in the semantic layer form a standardized multi-level data, features, and knowledge description framework that can support deep processing and mining of data, cross-reference semantic analysis and reasoning of multi-source heterogeneous data, semantic annotation and semi-supervised machine learning to Obtain semantic information; Finally, a multi-source heterogeneous data semantic network model based on complex network graphs is established to study the recognition method of complex semantic concepts and the mining technology of media semantic relationship.

(2) Automatic and efficient extraction, filtering, and consistency analysis of multi-source data

Under the guidance of the Semantic Web, research and develop automatic data extraction methods with higher performance, higher scalability, and higher accuracy. Automatically extract and filter video data related to network events (such as text, images, videos, etc.) and user behavior data (such as clicks, downloads, collections, comments, sharing, invitations, annotations, uploads, ratings) , Even including page dwell time, scrolling, hovering, backing, etc.) and video source data (video service operators and video hosts); Aiming at the dynamic nature of big data, study capture methods of changing data, and then propose incremental data extraction method; On this basis, Consistency analysis is performed from the model layer and instance layer, covering aspects such as integrity constraints, naming conflicts, structural conflicts, inconsistent time selection, missing data, and data summary conflicts.

(3) Cloud storage technology for large-scale multi-source heterogeneous data

Cloud database is a database deployed and virtualized in a cloud computing environment providing a powerful means for NoSQL data management. However,

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

cloud database is weak in transaction support, and some analytical applications require a certain degree of transaction semantic guarantees. At the same time, many analytical applications have cross-relational and NoSQL database access requirements, which requires the integration of cloud databases RDBMS and various NoSQL databases. Therefore, it is necessary to study the key technologies of cloud database management for analytical applications to provide the underlying support for further mining and analysis. Specific research contents includes: large-scale semantic web distributed storage and indexing technology, high-speed parallel semantic web access technology, management of the evolution process of the semantic web, cloud database architecture design integrating RDBMS and multiple NoSQL databases, and analysis based on CAP theory NoSQL database execution mechanism, large-scale heterogeneous data preprocessing technology, large-scale media data common processing technology, etc.

3.2 Unified representation of multi-source heterogeneous media data and hot spot event discovery based on multi-layer semantic model

The multi-source information of network video is extremely heterogeneous, and the types include multi-source information such as video, image, and text. The difference in data representation form is a huge difficulty for semantic analysis. How to build a unified representation model is the current research difficulty, and it is also one of the key links for the success of the hot spot discovery task. To study the problem of semantic analysis of multi-source heterogeneous data, the first is introduce a multi-layer semantic model, reduce the semantic gap layer by layer, and realize the unified representation of multi-source heterogeneous data; Secondly, propose a network video hot event evaluation system to study real-time Efficient hot spot event discovery method.

(1) Semantic feature extraction for unified representation and semantic relationship modeling of heterogeneous media

Faced with massive amounts of multi-source heterogeneous media data, a unified data representation model will lead to changes in information representation. So it can bring theoretical breakthroughs and great application value. The traditional method is to calculate the word frequency of the dictionary to extract the keyword description, but the keyword cannot accurately express the semantics, and cannot reflect the emotion and behavior mode of the specific user. In order to solve this problem, first of all, the research studies the feature package model (BoF) representation method of image and video data, and the topic model (LDA) representation method of text data. So as to obtain the topic statistical distribution map of heterogeneous media; Then the research studies semantics based on human cognitive theory and compressed perception method to obtain a sparse representation of multi-source heterogeneous data; Finally, on the basis of semantic categories, it finds the matching pattern between the heterogeneous media perception features. Specific research includes: evaluation method of modal-category connection based on text perception feature and image feature perception; common semantic space mapping method of two modal features to category

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

information representation; construction of semantic association model between heterogeneous media in common semantic space.

(2) Evaluation index system and discovery technology of network video hot events

Network social networks, including network video sites, have attracted large-scale users to share various types of information with their convenient information release and acquisition methods. At the same time, they have become a convenient channel for people to connect and organize. On the basis of the event semantic web model and the unified representation model of video multi-source data, the system builds a hot event evaluation system first. It also explores the evaluation indicators such as attention, explosion, penetration, continuity, importance to establish a quantitative system for hot events; Secondly, it models the event discovery problem as a classification problem and studies the classifier machine learning framework, especially the problems of fusion of multiple classifiers and imbalance of cluster samples.

3.3 Tracking and prediction of network video hots and sensitive events

Based on the event discovery based on video semantic analysis, it can further research on the evolution and prediction of hot spots and sensitive events. First of all, starting from the fragmented information brought by the user community, the fragmented information aggregation technology is used to quickly and accurately track the trend of hot event events, and supplemented by the time series analysis of inflection point recognition technology. Secondly, multi-dimensional analysis of the occurrence and development of events is conducted to construct a prediction model based on matrix decomposition to realize the prediction of sudden behavior.

(1) Tracking and inflection point identification and analysis technology of network video hot events

The infinitely subdivided audiences and the extremely small amount of information carried by a single network video, so it makes hot event information fragmented. Research on fragmented information aggregation technology plays an important role in tracking the evolution of hot events. The specific technologies involved in fragmented information aggregation includes: cross-platform hot event information association analysis, aggregated information quality control, content filtering and incremental update technology, sentiment analysis, etc. On the basis of obtaining the time series data of event evolution, the inflection point indicates that the event has changed greatly, and it is often the appropriate time for manual intervention. Therefore, this section also studies the inflection point recognition technology based on time series analysis.

(2) Network video hot event trend prediction and intervention decision support

In order to realize the trend prediction of hot events, it is necessary to grasp the rules of the generation and development of hot events. So the research conducts statistical analysis through a large number of actual cases to obtain the correlation

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

between the occurrence of sudden behavior and the time of related events, the exposure rate of the event, the degree of emotional fluctuation, the proportion of negative emotions, the number of participating individuals, the duration of hot spots, etc. Then it studies the prediction model based on matrix decomposition to realize the prediction of sudden behavior. Finally, it provides network video hot event early warning, and provide decision support for manual intervention at the sudden change point (inflection point) and the predicted future mutation point.

4. System application and prospect

The network video hot event monitoring and analysis system based on cloud computing can be applied to the evolution, prediction and intervention of hots and sensitive events. First of all, starting from the fragmented information brought by the user community, the fragmented information aggregation technology is used to quickly and accurately track the trend of hot event events, and supplemented by the time series analysis of inflection point recognition technology. Secondly, it conducts multi-dimensional analysis of the occurrence and development of events, constructs a prediction model based on matrix decomposition, realizes the prediction of sudden behavior. It also conducts statistical analysis through a large number of actual cases to obtain the correlation between the occurrence of sudden behavior and the time of related events, the exposure rate of the event, the degree of emotional fluctuation, the proportion of negative emotions, the number of participating individuals, the duration of hot spots, etc. Then it studies the prediction model based on matrix decomposition to realize the prediction of sudden behavior. Finally, it provides network video hot event early warning, and provide decision support for manual intervention at the sudden change point (inflection point) and the predicted future mutation point.

5. Conclusion

The rapid development of the network video industry makes network video play an increasingly important role in the spread of hot events, the discovery of social needs, and the monitoring and control of public opinion. The cloud video-based network video hot event monitoring and analysis system can automatically extract and filter multi-source data information of network video, extract semantic elements of hotspot events, build a unified representation of multi-source heterogeneous media data and hotspot event discovery model based on multi-layer semantic model, realize tracking and prediction of network video hot-spots and sensitive events, and solve the limitations of traditional text-based discourse analysis technology. Finally it achieves the purpose of providing decision support for government departments to conduct video public opinion supervision and emergency intervention.

Acknowledgement

Supported by “College Students Innovation and Entrepreneurship Training

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

Program" (Grant No. 201911647014)

References

[1] Chen Si, Lu Qinghe, Ma Xiaoyi: The breaking point, pain point and inflection point of China's network video development, Contemporary Communication, 2019 (3): 56-58+69

[2] Zaïane O R, Han J, Li Z N, et al. MultiMediaMiner: a system prototype for multimedia data mining[C]//ACM SIGMOD Record. ACM, 1998, 27(2): 581-583.

[3] Yang Fuzheng, Wan Shuai. Research status and development trend of network video quality assessment technology [J]. Journal of Communications, 2012, 33 (4): 107-114.

[4] Matsumoto Y, Uchida H, Hagimoto M, et al. Manycore processor for video mining applications[C]//Design Automation Conference (ASP-DAC), 2013 18th Asia and South Pacific. IEEE, 2013: 574-575.

[5] Liu J, Javed O, Cheng H, et al. Detecting Complex Events from Big Video Data[J]. E-LETTER, 2013, 15(5).

[6] Jenner M. Is this TVIV? On Netflix, TVIII and binge-watching[J]. New Media &Society, 2014: 1461444814541523.

[7] J.Allan. Introduction to Topic Detection and Tracking in Topic Detection and Tracking:Event-based Information Organization[R]. Kluwer Academic Publishers,2002:1-16.

[8] Hong Yu, Zhang Yu, Liu Ting, etc. Evaluation and research review of topic detection and tracking [J] Journal of Chinese Information Processing, 2007,21 (6): 71-87.

[9] Guo Jianyong, Cai Yong, Zhen Yanxia. Topic discovery based on text clustering technology. Computer Engineering and Design [J], 2008 (6): 1426-1432.

[10] Huang Xiaobin, Zhao Chao. Application of Text Mining in the Analysis of Internet Public Opinion Information. Information Science [J], 2009 (1): 94-99.

[11] Zhao Xujian, Yang Chunming, Li Bo, Zhang Hui, Jin Peiquan, Yue Lihua, Dai Wenkai. A news topic evolution mining method based on feature evolution [J]. Chinese Journal of Computers, 2014 (4): 819-832.

[12] Huang He. Sentiment analysis of Sina Weibo based on semantic sentiment space model. International Conference on Management Science and Engineering (ICMSE'13)[C], 2013: 206-211.

[13] J. Bollen, H. Mao and X. Zeng. Twitter mood predicts the stock market. Journal of Computational Science[J], 2011, 2(1): 1-8.

[14] Yang Liang, Lin Yuan, Lin Hongfei, Discovery of Hot Events on Weibo Based on Sentiment Distribution, Journal of Chinese Information, 2012 (1): 84-90

[15] Jiang Shenghong. The formation and development, status quo and guidance of public opinion hots on the Internet [J] .Theory Monthly, 2008 (4): 34-36.

[16] Tao Jianjie. Improving the emergency response mechanism of network public opinion linkage. Party and government forum [J], 2007 (9): 28-30.

[17] Liu Yi. Internet public opinion and the transformation of government governance paradigm. Frontier [J], 2006 (10): 140-143.

ISSN 2616-5775 Vol. 3, Issue 1: 150-159, DOI: 10.25236/AJCIS.2020.030116

[18] Wang Wei, Li Ruiguang, Zhou Yuan, Yang Wu. Microblog burst topic propagation prediction algorithm based on user and node scale [J]. Journal of Communications, 2013 (S1): 84-91.
[19] Li Biao: Research on the Spatial Structure and Characteristics of Internet Event Communication——Taking 40 Internet Hot Events in Recent Years as an Example, Journalism and Communication Research, 2011 (3): 90-99.
[20] Li Biao: Research on the Stages and Thresholds of the Dissemination of Internet Events—Taking 34 hot Internet public opinion events in 2010 as an example, International Press 2011 (10): 22-27
[21] Zhong Ying, Yu Xiucai, Analysis of Major Internet Public Opinion Events and Their Transmission Characteristics from 1998 to 2009, Journalism and Communication Research, 2010 (4): 45-52