Comprehensiveness, Preciseness and Interconnectedness: How to Evaluate International Public Opinion Based on Cross-media Data Mining on the Internet

Deng Tianqi\textsuperscript{1,}\textsuperscript{*}, Hou Xiaotian\textsuperscript{2,} and Liu Yumo\textsuperscript{3,}c

\textsuperscript{1}Institute of Communication Studies, Communication University of China, No. 1 Dingfuzhuang East Street, Chaoyang District, Beijing, China;
\textsuperscript{2}School of Journalism and Communication, Shanghai International Studies University, 1550 Wenxiang Road, Shanghai, China;
\textsuperscript{3}School of International Studies, Communication University of China, No. 1 Dingfuzhuang East Street, Chaoyang District, Beijing, China

\textsuperscript{*},\textsuperscript{a}dengtianqi@cuc.edu.cn;\textsuperscript{b}houxiaott1995@163.com;\textsuperscript{c}liuyumo2009@163.com

\textbf{Abstract.} With emerging new media technologies, the analytic focus of research on international public opinions has switched from single content form, such as text, audio, image or video into multiple forms of media integrated together in cyberspace or in physical space. The depth and frequency of monitoring international public opinion on topics related to China are improving rapidly. This paper pinpoints the inadequacies of current monitoring activities and puts forward ways to analyze international public opinions on topics related to China based on technologies, such as web indexing, deep learning and data mining. Thematic discovery based on big data analytics, real-time surveillance and monitoring of trending topics, intelligent prediction of popularity, intelligent analysis of emotional valence, etc. can effectively support the establishment of a system using multimedia data to monitor and analyze international public opinion.

1. Introduction

Thanks to the rapid development of artificial intelligence and data technology, all kinds of information can be stored fundamentally as data, available for analysis with intelligent tools. For example, indexes of social friendliness, happiness and pressure are calculable after collection and surveillance of public opinion and data in related areas, such as the relationships people have on social networks. Inspired by extant paradigms of evaluating online public opinion, some scholars shared their insights on how to apply artificial intelligence into practice and its value.

Evaluating international public opinion refers to the process of judging and analyzing the value of international public opinion. The implementation and enforcement of national policies depends on the preciseness of this process which can help improve the government’s capacity for international crisis management. Against the background of artificial intelligence, the volume of information has increased exponentially, experience-based methods are inadequate for the analysis of a large amount of cross-media data to evaluate public opinion. Therefore, this paper will focus on theoretically and practically updating the evaluation of cross-media public opinion on the Internet in the age of big data.
2. The inadequacies of evaluations of international public opinion on China-related topics based on cross-media data

For a long time, contemporary western media have been in center stage in studies of international public opinion. The development of the Internet put social media platforms, such as Facebook, Twitter, Weibo, Wechat, to the forefront of voicing international public opinion. The emergence of new types of media platforms reshaped the international media landscape. A good understanding of the public opinion revealed on Internet media or social media is conducive to the judgment of the effectiveness of China’s international communication and the international public opinion landscape China is facing.

China started to evaluate international public opinion in the 1990s in the form of human readers or listeners collecting information from newspapers, magazines, radio, etc. and then writing reports. With the development of Internet technologies since the 21st century, the emergence of artificial intelligence during recent years has posed great challenge for China’s evaluation of international public opinion. Traditional methods are outdated in the current environment. Online communities are developing rapidly and the Internet is playing a key role in people’s life as an important information source[1]. China has established an evaluation system of public opinion which can collect information about, do qualitative analysis of and monitor public opinion, users’ behavior, and public image.

However, the system is still lacking in the following aspects. First, it lacks timeliness. During monitoring, information delay is very common. Second, the system lacks cognitive abilities. In other words, it’s not intelligent enough. Third, the system is not fully capable of dealing with international emergencies without instant reaction to the latest high-publicity international issues. Besides, there’s still room for improvement, such as developing more functions and modes. Considering the increasing complexity of the international public opinion landscape and the high regularity of black swan events, it’s very important to establish a dynamic evaluation system which can collect information from, intelligently analyze social media, news websites, do text analysis and emotion detection and automatically issue alerts on trending topics for 24 hours.

3. The necessity and technologically determined characteristics of the evaluation of cross-media international public opinion

The development of Internet technologies has blurred the boundaries between different media types. In a post-truth age, a country shapes the images of other countries based on its national interest and influences public opinion which goes through different phases. Moreover, traditional thinking and methodology of influencing public opinion and social regulation is more and more outdated due to a large amount of non-structural data in a new media environment. Only with the help of interconnected machine learning can attitudinal textual analysis of public opinion and analysis of the influences of public opinion based on big data come into being.

3.1. A sharp increase in non-structural data

The diversity of communicators on social media and the availability of communication resources deconstruct the monopoly of media establishment on communication channels and guarantee the public’s right to express. The Social Media Age is characterized by diverse forms of dialogues. People of different social roles can participate in the production, transmission and re-transmission of diverse contents. Public opinion on social media is also diversified after discussion and communication among a large number of users which may lead to an avalanche of opinions, amplifying the non-structural characteristic of data[2]. In other words, the amount of data increases exponentially. Therefore, the traditional data analysis method with a focus on structural text cannot meet the demands of analyzing sharply increased non-structural data nor the goals of in-depth analysis, value unearthing and application. The accessibility of social media warrants the irreversible increase of non-structural data which requires more intelligent and effective technologies to unearth, process and analyze data so that the problem of managing public opinion can be effectively solved.
3.2 Potential interconnected machine learning
The inherent high accessibility of modern media increases the possibility of cross-media and cross-platform communication. The interconnected information centers around one event that engenders semantically bonded public opinion on global media platforms. The worlds of public opinion on global media platforms have connections with each other via the Internet, symbolized by netizens’ social networks, the network of trending issues, and the interaction network between the trending issues and netizens. Closely connected networks results in the gradual disappearance of information islands and the appearance of cross-media data which are fundamentally characterized by interconnectedness.

Though media platforms all over the world are developing their technologies to improve the relevancy between contents and queries and increase the semantic relatedness of different factors in a trending issue, diversified media platforms allowing no fully open access to their information bases and immaturity of intelligent data technologies make an effective establishment and application of the interconnected relationship between diverse data impossible. Therefore, it’s very important and necessary to focus on the interconnectedness of different forms of data when evaluating cross-media public opinion. Thanks to the development of artificial intelligence, machine technology can copy and strengthen the deep learning framework of a brain. Artificial intelligence can study the different characteristics of and unearth the hidden connections and the interconnected relationships among heterogeneous data, such as text, image, audio, and video by means of cluster computing, classification computing, association rule mining, data fusion through statistical matching so that learning the interconnectedness of cross-media data could be possible.

3.3 Predictability of the development of international public opinion
International public opinion on a trending topic changes as time goes. The emotive reaction of the audience and the way media communicate go through different phases which can be roughly divided into dormancy, incubation, explosion, transmission, re-transmission, remission and dissipation, etc. Development phases of international public opinion are shown in Figure 1. Public opinion on certain issues may cause distress, anxiety, betrayal, social upheaval and division among international communities, which requires the evaluation of public opinion to predict its future development and seize the window of opportunity to lead and manage it[3]. When public opinion goes astray or out of control, at a right time, emergency methods should be taken to respond to the public demand and guide the public opinion. Lessening the negative effect of out-of-control public opinion has become the new goal and focus of predicting and responding to public opinion. Artificial intelligence is more and more applied into monitoring and prediction to avoid out-of-control public opinion with an artificial neural network as the basic framework and summarizing the development pattern and model of public opinion where structurally reconstructed big data are input. Natural language processing, pattern recognition and machine learning can analyze the past and current characteristics of public opinion and reasonably predict its development to provide reliable and valuable basis for government and industry stakeholders to make policies.
Methods of evaluating cross-media international public opinion

Evaluation of cross-media international public opinion is a complicated process. A new landscape of public opinion requires accordingly upgrading strategies and mechanisms to process public opinion. Measuring the influences of key opinion leaders, analyzing receivers’ personal preferences and comprehensively evaluating the effectiveness of cross-media communication are conducive to understanding how international public opinion evolves on the Internet and predicting its trajectory.

4.1 Intelligent detection of public opinion based on big data

The key to evaluating international public opinion is to check relevancy between keywords and search results. Therefore, improving the intelligent search engines of international public opinion should focus on relevancy rankings and sorting the search results into order from higher to lower according to their correspondence with the keywords. Due to the development of computer and communication technologies, the models of searches can be classified into information search and information browsing.

Against the background of big data, intelligent search engines should expand and improve their ability to search for keywords in news based on their meanings, for example, synonyms of the keywords and other words close to the core concepts. Besides, it’s also necessary to analyze the characteristics of journalistic texts collected from searching for the basic ideas of the keywords[4]. In conclusion, intelligent searches for cross-media international public opinion include five steps. First, web indexing will collect needed information in the form of web pages and do real-time monitoring. HTML parser can help clean up the messy HTML found on Web and bring order to links, time and headlines. Second, using split() method can sort out keywords. Third, extract the common keywords from the sample web pages and allocate them a numerical value or vector. Fourth, Term Frequency–Inverse Document Frequency (TFIDF) will help effectively express the thematic topics of the texts. Vector Space Model will visualize the keywords as vector spaces. At last, Trending Topic Discovery in search engines will produce a matrix related to the sample texts to get the latest cross-media international public opinion. Models of intelligent searches based on big data are shown in Figure 2.
4.2 Monitoring and tracking trending topics that engender international public opinion

Technologies of monitoring and tracking trending topics deal with information flow on media, such as news outlets, radio or television and track the reports on new topics to help people deal with information overload. In the Internet Age, public opinion explodes in a short time. Among infinite international public opinion, spotting trending topics accurately and tracking its transmission pathway and trajectory can help evaluate the development of public opinion[5]. Discovery and detection of international emergencies relies on a comprehensive study of the source of trending topics, users’ participation and the regularity of their activities, etc. Tracking relevant public opinion needs to compare the similarity between new topics and original ones.

The foundation for tracking international opinion lies in the discovery of thematic topics which requires differentiating new topics from the old ones and monitor newly added topics. The goal is to actively track certain keywords of news in infinite public opinion and unearth texts related to the news keywords. Tracking trending topics relies on two ways: first, on the basis of Vector Space Model, applying search technology to construct corresponding query expressions which can be used to see if unprocessed international opinions matches processed ones; second, establishing improved algorithms with the technology of text classification, such as K-Nearest Neighbors (KNN) Algorithm and Decision Tree Algorithm.

4.3 Intelligent prediction of the popularity of a trending topic

The popularity of an international trending topic determines how to react to public opinion on the Internet and reflects how much attention an international trending topic is gaining. How large the information flow of the public opinion on a topic on Internet platforms, such as news websites, news apps, Wechat and Weibo determines the popularity of a trending topic. Sources of international public opinion in the age of the Internet are shown in Figure 3. In the age of the Internet, international public opinion is presented in a non-structural mixture of text, audio, video and image which contribute to the explosion of international public opinion.

The key to intelligent prediction of the popularity of a trending topic is to understand the contexts and build a computable model and, in other words, to associate similar contents based on big data analytics and build correlation models for data clustering with the purposes of comprehending real-life situations and contributing to understanding the meaning of language constructs. Besides, it’s necessary to recognize images and videos, analyze their popularity with the help of ImageNet, an image database, and predict the popularity of potential trending topics based on the characteristics of cross-media contents and related analysis in order to automatically unearth and intelligently evaluate the potential
high-publicity international topics. Process of intelligent prediction of the popularity of an international event is shown in Figure 4.

Intelligent prediction of the popularity of a trending topic targets non-structural visual and aural contents and automatically predict the popularity of trending information in the forms of audio, video and image so that it’s possible to infer the possibility and intensity of an explosion of international public opinion.

4.4 Intelligent analysis of emotive reaction based on big data
Intelligent analysis of emotive reaction based on big data refers to the judgment of the public’s attitudes towards a trending topic, normally categorized into positive, negative and neutral, etc. The extant analysis of international public opinion on the Internet mostly uses semantic analysis to categorize words that convey emotions, which means focusing on emotive words, negative words, and adverbs, etc, in the texts, to calculate the emotional valence of an event felt by the participants. However, such a method is inadequate for analyzing public opinion on cross-media online contents or judging the emotional valence of non-structural data. Moreover, it’s also relatively lacking in accuracy.

Deep learning, a new way of machine learning, solves the problem of unsupervised emotional clustering of large-scale texts. Typical deep learning includes Feedforward Neural Networks, Convolutional Neural Networks, and Recurrent Neural Networks, etc. Intelligent analysis based on big data of cross-media international public opinion necessitates the use of deep learning to make a comprehensive study of keywords, images and videos and judge the emotional valence with the help of non-linear deep learning. Moreover, intelligent calculation is important as well. Calculating the emotional valence of public opinion is a complicated process with many variables, such as time, nationality, context and addressee, etc. The calculation includes three steps: first, use tools of web indexing to obtain texts of emotional valence, and then transform the natural language into a form recognizable and readable to computers. At last, use Emotional Content Classifier to calculate the emotional valence of the texts.

5. Conclusion
Intelligent analysis of cross-media contents is an important part of the new generation of artificial intelligence. How to react to the high-publicity international emergencies on cross-media platforms has become an key issue in state governance and international relations. Nowadays, information on the Internet is adulterated with rumors or unsubstantiated claims which catch a lot of attention from the international communities and negatively affect China’s national image. Therefore, it’s urgent to effectively predict trending topics on cross-media platforms and minimize the risks of negative public opinion from the source. The development of digital technologies, such as big data, artificial intelligence and cloud computing can effectively contribute to the prediction, discovery and evaluation of trending topics that engender international public opinion. This paper thinks the sharp increase of non-structural data, the development of interconnected machine learning, and the predictability of the trajectory of
international public opinion pose huge challenge to traditional evaluation of public opinion methodologically and technologically. Considering the increase of high-publicity international topics and non-structural data, this paper puts forward methods to analyze cross-media online data of public opinion, to apply technologies to big-data-based intelligent discovery of public opinion, monitor and track international trending topics and intelligent prediction of the popularity of a trending online topic and intelligent analysis of emotional valence based on big data. Under the background of increasing non-structural data, analysis of cross-media public opinion on the Internet is conducive to the accuracy and timeliness of the evaluation of public opinion and the development of China’s international communication.

References
[1] Tang Tao 2014 Research on Method of Internet Public Opinion Analysis Based on Big Data. J. Journal of Modern Information. 34 pp 3-11
[2] Duan Fengfeng, Chen Miao 2020 Pluralism and Smartness: Analysis and Processing of Cross-Media Online Public Opinion Data. J. Media Observer. 433 pp 33-38
[3] Chen Lin 2017 Design of Intelligent Information Retrieval System under Big Data Environment. J. China Computer&Communication. 6 pp 128-132
[4] Cao Haijun, LI Ming 2018 Rethinking and Path Expansion of China’s Network Public Opinion Governance in the Big Data Era: Based on the “Technical Governance Path” Embedded Perspective. J. Administrative Tribune. 155 pp 65-73
[5] Zhou Qi, Yin Jian, Zhang Liangjun 2020 Analysis and Design of Public Sentiment Data Center Based on Big Data Structure. J. Modern Computer. 28 pp 104-108