

Feasibility Analysis of Internet Rumor Recognition Research based on Basic Theme Crawlers and Text Tendencies

Weiling Huang, Zixin Yu, Baihui Li, Zichen Wang

Jiangxi University of Finance and Economics, Nanchang, Jiangxi, China 330013, China

Abstract

During the global outbreak of COVID-19 in 2020, a series of online rumors broke out on social media. These online public opinions are uneven and endless, which have a serious impact on social public opinion and epidemic prevention and control. Because this public opinion has the characteristics of high concentration, wide influence, short fermentation time, and rapid spread. Based on this, this article conducts an in-depth analysis of the text of Internet rumors (including rumors, facts and to be determined) on the theme of the new crown pneumonia epidemic, mainly including two aspects: one is based on the current rumor-defying platform and Internet public opinion system that can be used for reference, and a brief analysis of topic crawlers Regarding the feasibility of this study; the second is based on text mining technology and NLP theory to explain the topic mapping and sentiment analysis of the Internet gossip text data related to the epidemic, and thus analyze the feasibility of identifying rumor texts with text orientation.

Keywords

Internet Rumors; Topic Crawlers; Text Mining.

1. Introduction

The Internet is increasingly becoming an important source of health information. Web2.0 has promoted the development of virtual communities on the Internet, and people are increasingly inclined to express opinions and opinions on the Internet. Coupled with the low barriers to use of various websites, anonymity of user identities, strong relationships, and high connectivity, various speeches are extremely easy to spread on the Internet. Therefore, while the Internet promotes communication, it also creates a hotbed for malicious rumors to spread. Rumors are prone to large-scale outbreaks through the Internet, especially during the outbreak of major epidemics, which will have a negative impact on the real society. Today, online rumors have become a new public health problem on the Internet. Therefore, controlling the spread of rumors on the Internet has become a major social issue that stabilizes the society and correctly guides the people's thinking. Our country has established a number of rumors identification and public opinion monitoring platforms, through the use of big data to analyze the attributes of rumors, algorithm models to obtain training sets, classification algorithm analysis and other methods to release professional anti-rumor information, so as to curb the spread of rumors. Therefore, research on the text characteristics of online rumors, transmission models, influence evaluation algorithms, and rumor recognition platforms will help to more quickly and accurately grasp the rules of rumors spreading on the Internet and provide a powerful way to effectively identify online rumors and stop the spread of rumors. According to [1][2].

2. Current Research Status at Home and Abroad

2.1. Research on the Governance Mechanism for the Spread of Epidemic Rumors

Luan Yushu [3] mentioned in the study that during the special period of the epidemic, the relevant information involves the vital interests of the public, and government agencies must strengthen crisis response capabilities, improve information transparency, and establish an emergency disclosure mechanism for information on major emergencies. At the same time, the mainstream media and the general public cannot avoid or be absent. They should stick to their social responsibilities, quickly and accurately release news and information about the epidemic that society is generally concerned about, and promptly dispel rumors.

2.2. Social and Psychological Analysis of Internet Rumors

(1) Yang Shuai, Zhong Hancheng[4][5] and others have systematically analyzed the time series and social psychological characteristics of online rumors during the epidemic and found that: Internet rumors can be divided into five categories during the epidemic, and the number of rumors in descending order is as follows Health protection category, government policy category, social life category, celebrity category, technical research category. Among them, the number of the first two types of rumors is significantly greater than the last three types, indicating that the public has the greatest demand for health protection and government information disclosure; medical care, scientific and technological research and other rumors also occupy a certain proportion, reflecting the public's anti-epidemic medical and technological anti-epidemic needs. Pray and approve. Secondly, the research conducted a comprehensive analysis based on the relevant theoretical assumptions and the above content, and divided the people's psychological development into four stages, which are mainly manifested as fear and worry in the panic period, and the multiple mentalities in the balance-seeking period are intertwined to seek psychological balance, and The recovery period gradually calms down psychologically[6][7].

(2) Coincidentally, Lu Tu[8] and others pointed out that when facing rumors, netizens tend to lower their personal cognition and imitate others, and then tend to spread online rumors in a herd manner. This is because the ambiguity and uncertainty of the content of the rumors disturb the individual's rational judgment of information and reduce the individual's confidence in their own judgment. Individuals in the context of rumors tend to refer to the opinions of the majority and imitate the actions of the majority to keep themselves consistent with the group in cognition and action, so as to ensure that they pay less due to misjudgments and avoid "no The stress and anxiety brought about by gregariousness. In addition, the inhibitory effect of the perception of rumors on conformity lies in "thinking." Zhou Jianlin[8] mentioned that the public's perception of the rumor-defying mechanism will prevent people from not lowering their personal cognition when facing online rumors, but it will not reduce their imitating others in their actions, that is, they will only achieve "rational thinking and blind behavior". the result of.

2.3. Relevant Support Platforms for Network Rumors Identification

(1) There is currently no support platform that can directly identify online rumors in real time, but there have been research and applications on the characteristics of online rumors and recognition models. In addition to the above research, Lan Yuexin [9] used numerical simulation and phase trajectory analysis to study the network rumors propagation law model of emergencies in addition to the above-mentioned research on the characteristics of network rumors. In the study of recognition models, for example, Zeng Ziming [10] used the LDA topic model to deeply dig the topic distribution characteristics of the microblog rumor text, and based on the above feature variables, the random forest algorithm was used to conduct the

model training for rumor recognition. He pointed out that the random forest based on the LDA model The model can effectively improve the accuracy of network rumors recognition. In the online rumor identification theory application platform, there is currently no effective identification platform before or during the large-scale spread of online rumors, but there are already online rumor-defying platforms after the large-scale spread of Internet rumors (including Tencent's more authentic rumor-defying platforms, China Internet Joint rumor rejection platforms, etc.), these platforms will list the content of the rumors, the main points of the verification and the original text of the verification when dispelling the rumors, in order to eliminate the impact afterwards [11].

3. Feasibility Analysis

From the above, there is a certain foundation for the research on Internet rumors, but there is still a large research on the targeted recognition and text analysis of Internet rumors (rumors) on such a large-scale and widespread topic (new crown pneumonia epidemic). space. Therefore, this article will take the online rumors data of the new crown pneumonia period on the Tencent News Verification Platform and Weibo as the research object, analyze the necessity and feasibility of the topic crawler to intercept the Internet rumors related to this topic in real time, and elaborate on such Internet rumors. The research value and significance of text data, summarizing the general dissemination and text characteristics of online rumors during the entire major outbreak, pave the way for the training of a suitable online rumor recognition platform for application.

3.1. The Feasibility of Subject Crawlers to Intercept Information Flow in Real Time

Online rumors about the new crown epidemic are spread on various social platforms on the Internet. The first thing to do for targeted research is to sort out a large number of rumors and related accompanying information. The realization of this technology is generally implemented quickly and automatically by web crawling technology. Since the epidemic is not over yet, the rumor text on this topic is still in a real-time update state. Therefore, the choice of crawler technology needs to be fully considered in the system design stage. Existing web crawlers are generally divided into general crawlers and topic crawlers. The search strategy of the former is generally breadth. Prioritize the search strategy, while the latter adopts the best search strategy. This method is subject-oriented and highly pertinent. It crawls as much web page information related to the subject as possible to avoid data redundancy. In addition, the real-time stream of the theme crawler helps the automatic update data of the recognition platform, so as to achieve higher accuracy of recognition and analysis. In summary, the simplicity and efficiency of the topic crawler has great feasibility for the establishment of a database of epidemic network gossip texts [11].

3.2. The Textual Characteristics and Psychological Mapping of Rumors

Regarding the textual content of rumors, it is not accidental that rumors are favored by the public, but they have complicated social background and psychological roots. Starting from the psychosocial elements, Allport once proposed: R (Rumor, rumors) = A (Ambiguous, situational ambiguity) \times I (Important, importance of the event). Peng Xiaozhe[14] et al. summarized the influencing factors of rumors as: situational characteristics, which refer to the ambiguity of the situation and the openness of the public opinion field, etc.; content characteristics, which refer to the importance and urgency, directivity and emotionality of rumors; Audience characteristics, including the reliability of the communicator, the individual and psychological characteristics of the audience, etc. Combining the above theories, this study uses the text of COVID-19 rumors for text mining: First, a detailed text analysis of network rumors can be

performed based on NLP and LIWC. For the subject, professional nouns, numbers, 7 types of emotional words (happiness, good, anger, sorrow, fright, fear, evil), degree range words, special names, etc. of the text content, use LIWC theory to perform word frequency statistics and word feature value statistics. In order to discover the text characteristics of rumors in different time periods and the text differences between true and false Internet rumors. Secondly, the emotional and psychological words can be classified through the psychology dictionary, and then the text emotion can be judged or the psychological concept can be captured by its use. This text tendency feature is used as an index item for network rumors recognition [8] [12][13].

4. Conclusion

The spread of online rumors on the theme of the new crown pneumonia epidemic has seriously affected the production and life of the general public and has become an important social phenomenon. Therefore, it is very urgent to scientifically understand the dissemination characteristics of rumors and to study the status quo of rumors. Especially during major outbreaks, the unknown new virus outbreak has brought people more panic and other negative emotions. Therefore, research on online rumors during the epidemic is of great significance to the prevention and control of the epidemic in China and the world as well as major public crisis events in the future. The topic crawler real-time monitoring and text tendency analysis involved in the focus of this article are extremely feasible and important for building a new crown pneumonia epidemic rumors recognition platform.

Acknowledgments

Fund Project: Jiangxi University of Finance and Economics 2021 Undergraduate Innovation and Entrepreneurship Training Program, Project Number: 202110421114.

References

- [1] Xiang Zhuoyuan, Chen Yuling. Research on Weibo Rumor Spread Model and Impact Evaluation[J]. Scientific Research Management 2016.1: 39-47.
- [2] Du Hongtao, Teng Lin, Zhao Zhiyun. The effect of public management communication in the public opinion of sudden infectious diseases[J]. Information Magazine. 2017.2.
- [3] Luan Yushu. Research on the characteristics of the spread of rumors about the new crown pneumonia epidemic and its governance mechanism [J]. News Research Guide. 2020 (6): 30-34.
- [4] Shuai Yang, Yi Zhao. Time series characteristics and social psychological analysis of Internet rumors during the novel coronavirus epidemic [J]. Journal of Yibin College. 2020. No. 4.
- [5] Zhong Hancheng. Analysis of the current situation and countermeasures of rumor refutation during the prevention and control of the new crown pneumonia epidemic --Take the Tencent News Verification Platform "Real-time Rumor Refutation of New Coronavirus Pneumonia" as an example [J]. Popular Science Research, 2020, 15(02): 43- 51+105.
- [6] Yang Chun. Classification and response to rumors about the new crown pneumonia epidemic[J]. News Collection, 2020 (02): 20-21.
- [7] Wang Ruizhe. Research on Weibo content in the early stage of H7N9 avian influenza epidemic [D]. Central China Normal University, 2014.
- [8] Lu Tu, Chen Hao, Lin Huan, Zhou Jianlin. Research on the Influence of Internet Rumor Governance Strategies on the Willingness to Disseminate Rumors in Public Emergencies[J]. Journal of Information, 2020, 39(07): 87-93.
- [9] Lan Yuexin. Research on the Spread Law Model and Countermeasures of Internet Public Opinions and Rumors in Emergencies[J]. Information Science, 2012, 30(09): 1334-1338.

- [10] Zeng Ziming, Wang Jing. Research on Weibo Rumor Recognition Based on LDA and Random Forest-Taking 2016 Haze Rumors as an Example [J]. Journal of Information, 2019, 38(01): 89-96.
- [11] Zeng Dewei. Research and implementation of network public opinion system based on topic crawler [D]. Chongqing University of Technology, 2020.
- [12] Wang Qingjia. Research and Implementation of University Public Opinion System Based on Textual Tendency Analysis [D]. University of Jinan, 2019.
- [13] Dang Jingyuan. Research and design of the text tendency analysis system of network public opinion based on deep learning [D]. Xidian University, 2019.
- [14] Peng Xiaozhe, Cui Fang, Jiao Can, Li Hong. The context, content, communicator and audience characteristics of rumors and related cognitive neuroscience research[J]. Psychological Science, 2018, 41(04): 916-921.