Fuzzy Systems and Data Mining VIII A.J. Tallón-Ballesteros (Ed.) © 2022 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/FAIA220401

KNN Algorithm and BP Neural Network Model Applied to Network News and Public Opinion Governance System

Jianghui LIU¹, Fengxi YUN, Runlong WU and Xiaoqi CHEN Guangdong University of Foreign Studies, Guangzhou, China.

Abstract. Online public opinion events occur frequently. The effective governance of online public opinion is in an increasingly important position. This study creatively constructs an Opinion Lexicon to categorize and assign scores to words in major news commentaries. The model utilized in the article is based on KNN algorithm and BP neural network. The usefulness of the model is to screen and score samples of news comments from three representative media platforms, namely "People's Daily Online", "IFENG.com" and "Paper". Through data processing and result comparison of the samples, the study can correctly predict the value of news opinion guidance and the scores assigned to comments. On the one hand, the study wants to provide an effective way for public opinion governance of Internet News. On the other hand, it promotes other news media to think about the key and initiatives of online opinion guidance from the perspective of social opinion orientation.

Keywords. Online opinion governance; opinion guidance; Opinion Lexicon; KNN algorithm; BP neural network

1. Introduction

With the in-depth development of media convergence, more and more Internet users are free to express their comments and opinions on the Internet. The Internet has the characteristics of a wide audience and fast dissemination. The various remarks of netizens will spread quickly as soon as they are posted. Because of the low threshold of the Internet, people behind the network are mixed. Therefore, there is often a mix of good and bad in online discourse. On the other hand, due to the explosion of information on the Internet, many Internet users are often unable to distinguish the correct information. Netizens are extremely susceptible to being blinded by false news or harmful news, or even misguided by certain biased media or remarks. So they may approach public events blindly and make radical, inflammatory statements [1]. This can lead to a crisis of public opinion and social instability. In today's situation, new technologies such as artificial intelligence and big data analysis provide a new paradigm and solution for online public opinion governance. They open a new digital era of online public opinion governance [2]. Therefore, this study will explore the public opinion governance of Internet News

¹ Corresponding Author, Liu Jianghui Guangdong University of Foreign Studies, Guangzhou, China; Email: 247031690@qq.com

This study was financially supported by the Undergraduate Innovation Training Project of Guangdong University of Foreign Studies in 2022.

based on the background of big data. Through lexicon construction and algorithmic analysis, the author will explore a path that is in line with the current online ecology. By exploring the right way to guide public opinion, the article hopes to provide suggestions for the media to be more objective and positive in their news reporting. The article also hopes to help media practitioners to step out of the current dilemma of public opinion governance on the Internet and improve their current approach to public opinion governance.

2. Current situation of Internet public opinion governance

In their book, Bracken and Gordon explained and summarized the various views of crisis management experts and related worker in Europe and the US. They further analyzed how to deploy relevant public opinion crisis management strategies [3]. Ceron A and Negri F pointed out that social media and various social platforms can play a positive role in facilitating communication between the government and citizens. The government can use big data technology to extract valid information from the opinions and remarks of netizens to inform opinions for policy making [4]. Andrew Keen argued that the influence of online bloggers will interfere with the government's online public opinion work to some extent. So it needs to reconstruct and standardize the ecological order of we-media [5].

As the study of online public opinion continues to grow, many Western scholars have built models to study it. Katarzyna Sznajd-Weron developed the Sznajd model. The model states that public opinion is transmitted from the inside out. From the perspective of the technology required for online opinion, Jeong Smith found that the control and management of online opinion requires a variety of highly specialized Internet technologies such as neural networks, genetic algorithms, plain Bayesian and combinatorial classifiers [6]. The relevant research in this area has provided theoretical and technical guidance for online public opinion governance.

At present, most of the studies on online public opinion governance in China are the detailed description of the current situation of online public opinion and lack of summary. In addition, all kinds of studies are fragmented in terms of research disciplines, the study of news communication on online public opinion is the main focus. There are relatively few comprehensive studies and analyses on online public opinion governance. As for foreign studies, they are applicable to China in terms of the universality of public opinion generation and dissemination. But their specific governance models and approaches are difficult to apply to China because of the differences in national conditions and systems [7].

Secondly, the network public opinion governance often appears in the form of countermeasures to solve the problem of the network public opinion. Most of the research focuses on the generation and development of online public opinion, while the content of public opinion governance is only supplemented by suggestions, and less appears as a whole research object. Furthermore, most of the research subjects are the government. Most studies lack of research and in-depth discussion on news media. What's more, most of the research content is biased towards political news or events, and few study life or high mass participation events.

To address the gaps in the current research proposed above, this study will take the news media platform as the research object and focus on major social hot events, to explore the effective path of public opinion governance from a new perspective.

3. Design of public opinion governance system

3.1. Agenda-setting theory

Agenda setting is an important way for mass communication media to influence society. Mass communication may not affect what people think, but it can affect what they think [8], as shown in Figure 1.



Figure 1. Network agenda setting.

The theory is that mass communication cannot determine what people think about an event or opinion, but it can provide information or arrange relevant agenda to influence people's choice of facts and opinions they want to focus on and the order in which they talk about them [9]. The theory reveals the great utility of media in the process of mass communication. It also suggests the possible function of media setting in public opinion governance, which provides ideas and direction for the research of this paper.

3.2. Opinion-oriented analysis of comments on media platforms using dictionary

By calling the trained model in Python, that is, the Opinion Lexicon, we can analyze and predict the comments of various media platforms. Each comment can be divided into several words, and each word has a corresponding score in the Opinion Lexicon. The total score of each comment is obtained by adding up the scores and the public opinion of each comment is judged based on the total score. Finally, the overall amount of public opinion scores of each news media is observed through a histogram comparison. The scores are used as a basis to judge which media's news output brings more positive public opinion comments. The feedback from the positive public opinion comments reflects the public's thinking and understanding of the journalism, it also shows that the news on the platform has played a positive role in the public Opinion Lexicon to judge which media can achieve better public opinion guidance, analyze the advantages, and facilitate the industry to learn from.

In addition, the study adopts a five-layer logical architecture to carry out the research work. The system architecture, from top to bottom are the performance layer, the network layer, the business layer, the database layer and the raw data layer.

3.3. Application of KNN algorithm and BP neural network algorithm

KNN algorithm

KNN algorithm (K nearest neighbor) is a basic machine learning algorithm proposed by Cover and Hart in 1968. The basic idea of the KNN algorithm is that the genus class of a sample in a feature space depends on the class of the majority of the K nearest (or most similar) samples in the space. For the unlabeled new data, the K data closest to the data are found in a given labeled training data set. Most of these K data categories are used to mark the new data.

The study uses the KNN algorithm to perform a preliminary screening of the selected comments. Firstly, the comments such as "oh", "huh", "@" and other comments that are difficult to be rated by Opinion Lexicon are filtered. Although the filtered comment samples are not guaranteed to be free of worthless comments, fewer worthless comments enable the neural network mentioned below to calculate more valuable comments and thus have a higher computational efficiency.

According to the needs of the research object of this study, the value of K can be defined as a certain value that needs to be reached by the test sample L to be labeled by the classification.

The important formula in KNN algorithm is, Euclidean distance formula.

$$D(P, X) = \sqrt{(P_1 - X_1)^2 + (P_2 - X_2)^2 + \dots + (P_n - X_n)^2} = \sqrt{\sum_{i=1}^n (P_i - X_i)^2}$$
(1-1)

Sample Category Center:

$$0 = \frac{1}{N_k} \sum_{x \in O_k} x \tag{1-2}$$

Distance weights:

$$U = e^{[-d(x_{i,r}x_j)]}$$
(1-3)

• BP neural network model

The simple algorithm flowchart of BP neural network is shown in the figure 2.



Figure 2. BP neural network simple algorithm flow chart.

BP (back propagation) neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986 [10]. It is a multi-layer feed-forward neural network

trained in the error reverse propagation algorithm. This model enables computer systems to process the transmitted data information as efficiently as the nervous system of the human body. The BP algorithm is implemented by propagating the data signal forward and the error backward. By iterating the two processes, the final system achieves the smallest mean squared error difference between the actual and expected output values after data calculation. Based on the core of BP neural network, it can be used in this study to investigate the governance of online news opinion and can be expressed in the following words.

The flow of the BP neural network algorithm is as follows.

(1) The data input value of the neuron nodes in the hidden layer is calculated with the formula.

$$I_j = \sum_{i=1}^n W_{ij} X_i + \theta_j \tag{2-1}$$

 I_j is the implicit layer data input at the jth neural node, W_{ij} is the weight between the ith neuron in the input layer and the jth neuron in the implicit layer, and θ_j is the threshold of the jth neuron in the implicit layer.

(2) Calculate the actual output value of the data at the jth neural node in the hidden layer. Here is the formula for it.

$$O_j = f(I_j) = \frac{1}{1+e^{I_j}}$$
 (2-2)

 O_j is the output of the jth neural node in the hidden layer, the function f is the activation function, here the Sigmoid function is chosen as the activation function.

(3) Calculate the actual output value of the data at the kth neural node in the output layer.

$$Y_{k} = \mu \left[\sum_{i=1}^{q} W_{ki} f\left(\sum_{j=1}^{n} W_{ij} X_{j} + \theta_{j} \right) + a_{k} \right]$$
(2-3)

q is the number of neuron nodes in the hidden layer, $\mu(x)$ is the transfer function in the output layer, a_k is the threshold value of the kth neural node in the output layer, k from 1 to s, s is the number of neuron nodes in the output layer, and Y_k denotes the output value of the kth neural node in the output layer.

(4) Calculate the error between the predicted and actual values of the output layer, and here is the formula.

$$Edd_k = Y_k(1 - Y_k)(R_k - Y_k)$$
 (2-4)

Rk denotes the actual value corresponding to the kth neural node in the output layer.

(5) Calculate the error of neuron nodes in the hidden layer with the formula.

$$E_{dd_{i}} = O_{j} \left(1 - O_{j} \right) \sum_{k} E dd_{k} W_{jk}$$

$$(2-5)$$

 E_{dd_j} is the error of the output value in the output layer, and the weights of the hidden layer and the output layer between the jth neural node and the kth neural node are denoted by W_{jk} .

(6) Adjust the neural network weights, which include weights and thresholds, here is the formula.

$$W_{ij} = W_{ij} + H(Edd_jY_j) \tag{2-6}$$

$$W_{ij} = W_{ij} + H(Edd_jO_j)$$
 (2-7)

$$\theta_j = \theta_j + H(Edd_j) \tag{2-8}$$

H is the learning rate, and *H* takes values between 0 and 1.

The purpose of adjusting the weights is to minimize the error, for which the expression is defined as the following loss function.

$$L(e) = \frac{1}{2}SSE = \frac{1}{2}\sum_{j=0}^{k} e_{j}^{2} = \frac{1}{2}\sum_{j=0}^{k} (\overline{y_{j}} - y_{j})^{2}$$
(2-9)

After improving the above BP neural network algorithm, the following formula based on L2 regularization is obtained.

$$C = C_0 + \frac{\varphi}{n} \sum_{w} w^2 \qquad (2-10)$$

The original consumption function is denoted by C_0 and the regularization factor is φ . The number of training comment samples is n.

4. Experimental results

In the selection of the data sample, we take into account the large number of users and the wide influence of the "TikTok" online platform. Several media platforms have been stationed on TikTok to report news in the form of short videos and live broadcasts. Therefore, our study selected the three major news media "People's Daily Online", "IFENG.com" and "Paper" that have been stationed in the "TikTok" platform, took data samples from the short video comments related to the "Tangshan barbecue restaurant beating incident" that was widely concerned by people nationwide in June this year. First filtered the comments that were obviously worthless using the KNN algorithm. Then the BP neural network model mentioned in this study was used to score the Opinion Lexicon.

Since the influence of the three major media platforms varies, the sample of comments is taken according to the number of comments and the number of followers according to a certain weighting ratio. In "Paper", 1892 comments were extracted from a related short video comment. In "IFENG.com", 6119 comments were extracted from a related short video comment. In "People's Daily Online", 23917 comments were extracted from a related from a related short video comment. After filtering by KNN algorithm, the number of comments of "Paper" was 1264, the number of comments of "IFENG.com" was 4128, and the number of comments from "People's Daily" was 14,982.

The data show that among the comments on the short news video of "Tangshan barbecue restaurant beating incident" in the three selected media platforms, comments from the high grouping of People's Daily Online account for a larger proportion. Compared with the other two media platforms, the opinion orientation of the comments under People's Daily is more positive and its efficiency on public opinion governance should be higher. Meanwhile, according to the prediction of the algorithm model, the weight ratio of different grouped comments of the three major media platforms differs from the predicted value by less than 4%. The number of iterations in the model is generally lower for the high grouped comments, higher for the low grouped and unfiltered worthless comments. Although the weight of comments in different subgroups are different, the weight of comments in the middle and high subgroups of the three major media platforms basically tends to be the same. It means that the three major media platforms have positive opinion guiding effects in general, indicating that their own online news opinion environment is also relatively good. Then, based on the model's prediction of the public opinion orientation of such news comments and the overall comment scores, the media can selectively screen the comments in advance for the next similar news. They can top the comments that are more likely to meet the high grouping, so that the public who learn about the news later will be positively guided by the high grouping opinion and thus the overall online public opinion environment can be governed and improved.

5. Conclusion

This study combines public opinion governance and public opinion orientation of online news, innovatively uses public opinion dictionaries, big data and other tools to analyze the effect of online public opinion governance of different media. Our study also summarizes the impact of current public opinion governance on people's opinion orientation and gives suggestions for better governance in the future. Although there are shortcomings in the research process, such as small sample size, the selected research subjects are not representative of all media, and the research design algorithm still needs to be improved, we still hope to provide meaningful lessons for future research. We also hope that future research can expand the scope of the study subjects and increase the sample size, so that the research results can be more widely applied to the current online public opinion governance with more perfect big data technology and algorithms, and solve the difficult problem of online opinion governance for media workers.

References

- [1] Yuan Hong, Li Jia. A study on the strategy of online opinion management of social hotspots in the context of actor network theory. Intelligence Data Work,2021,42(06):31-44.
- [2] Xiao Zhonghua, Zou Xiongzhi. Risks, causes and prevention strategies for social governance in the era of big data. Enterprise Economics, 2020, (05):155-160.
- [3] Paul Bracken, Ian Bremmer, David Gordon. Managing strategic surprise: lessons from risk management and risk assessment. Central Compilation & Translation Press, 2014.
- [4] CERON A,NEGRI F. The social side of public policy: monitoring online public opinion and its mobilization during the policy cycle. Policy&Internet,2016;8(2):131-147.
- [5] Andrew Keen. The cult of amateur: How today's Internet is killing our culture. Nan Hai Publishing Co; 2010.
- [6] Jeong Smith. Public Responses to an oil Spill Accident: A Test of the Attribution Theory and situational Crisis communication Theory. Public Relations Review, 2009, 35(3): 307-309.
- [7] Yue Xiaoli. Research on the Governance of Online Public Opinion in China. Hebei Normal University,2021.
- [8] McCombs M E&Shaw D L., The Agenda-Setting Function of Mass Media, Public Opinion Quarterly, vol.36, no.2, 1972, pp.176-187.
- [9] Zheng Guangliang, Wang Yuhao. Paradigm Shift and Countermeasures of Online Public Opinion Governance in the Perspective of Big Data. Journal of the Party School of the CPC Tianjin Municipal Committee, 2022,24(04):64-74.
- [10] Li Biao. Research on the dissemination stages and thresholds of online events: An example of 34 hot online opinion events in 2010. International Journalism, 2011, 33(10):22-27.