WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n2.11
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Cognitive Bias in Online Public Opinion

Jingqi Chen

Department of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China

2021010801023@std.uestc.edu.cn

ABSTRACT

In the era of big data, there is a tremendous amount of information on the Internet, but people can only access part of it. Information obtained through only limited channels can easily cause cognitive bias and expand the influence of related online public opinions. This article explores the origins of bias, from the overwhelming influx of information to the personalized recommendation algorithms and homogeneity within social media groups. To combat these issues, the paper proposes solutions such as data preprocessing techniques, information dissemination methods, and measures to ensure algorithmic fairness. Beyond this, truly solving the problem requires interdisciplinary collaboration and multi-stakeholder engagement to effectively address these challenges, emphasizing efforts in public education, platform governance, and regulatory frameworks.

KEYWORDS

Internet Public Opinion; Cognitive Bias; Big Data; User Influence; Algorithm Fairness

1. INTRODUCTION

With the advent of the big data era, online public opinion research has become a focus in public opinion communication. Online public opinion can quickly spread information and influence public attitudes and behaviors. Especially in today's fragmented information age, people are usually dominated by emotion and evaluate issues from the perspective of emotion rather than rationality. This triggers public opinion and even reverses it multiple times, affecting social behavior and decision-making. Therefore, network public opinion has essential research significance and practical application value.

Cognitive bias in online public opinion refers to the phenomenon that there is a gap between the facts and the cognitive subject of online public opinion based on the acquisition of network information. This gap is mainly reflected in cognitive subjects' one-sided and unreal understanding of public opinion objects. This includes, but is not limited to, the impact of false information and misleading content on user ideologies, the personalization tendency of information filtering, the recommendation preferences of social media algorithms, and the stance tendencies of individuals or groups in information dissemination.

This work aims to identify the current status of cognitive biases in online public opinion research and provide reference ideas for subsequent research. Cognitive bias is a relatively vague concept. Online public opinion can be divided into at least the social and technical levels. To truly understand the problem of cognitive bias, we must first clarify its definition. In addition, as a dataset to deal with this problem, the data source must ensure its authenticity, so we need to eliminate false news. In the process of dissemination of information, there will inevitably be misinformation. However, through the propagation model, nodes with significant influence can be found to monitor or restrict them and

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

reduce the probability of this phenomenon occurring. As for the algorithm, the general machine learning algorithm is usually a black box, which only obtains specific characteristics of the input data based on requirements. Failure to constrain it can lead to biased results. Due to the unknowability of the machine learning process, it is not easy for people to find that the results are problematic. This article will discuss these problems and try to provide solutions.

2. ANALYSIS

For cognition, bias refers to an understanding close to the standard but with some errors. The reason may be a problem with the information a person obtains or a deviation in their understanding of the information due to their knowledge level and depth of experience (short-term or long-term accumulation). However, prejudice is more likely to be a long-term stereotype of a particular group in society, such as skin color and race.

The leading cause of cognitive bias is information explosion in the big data era. People are facing an influx of massive amounts of information. However, individuals' cognitive resources are limited. People may selectively focus on specific details while ignoring or missing other vital information. This selective focus may lead to a one-sided understanding of facts and cognitive biases. To solve the problem of large amounts of information, various platforms have now developed recommendation algorithms to make personalized recommendations. This recommendation algorithm may allow individuals to enter a state of information cocoon, making it easier for them to access information consistent with their own opinions, thereby increasing the possibility of cognitive bias.

Similarly, in social media, people tend to interact with groups with similar views, called the column aggregation phenomenon. This homogeneity may lead to consensus thinking within the group, making group members more susceptible to similar views and ignoring other information and perspectives. There are also rumors and false information on the Internet, which can be easily deceived by young people with limited knowledge reserves and social experience or older adults who are relatively out of touch with the development of the times.

The above-mentioned cognitive biases are all classified at the social level, while the cognitive biases at the technical level are more related to data and algorithms. If the training data is large enough, the data set will inevitably be biased when training a model. Under the training of these data, the resulting model will take it for granted that some existing social phenomena and social remarks, such as skin color and race, are reasonable. But in fact, this will deepen and solidify these social problems. Ultimately, machine learning is a statistical and predictive algorithm based on data. Without constraints, the characteristics of the data themselves will be reflected in the results intentionally or unintentionally.

3. SOLUTION

3.1. Data preprocessing

In online public opinion research, the data sets used generally select various indicators that can express the popularity of a specific topic, such as the risk assessment system constructed by Jin Dongxue [1]. Based on the public opinion life cycle theory and the evolution mechanism of public opinion risk, the four dimensions of public opinion—participating subjects, public opinion objects, public opinion ontology, and public opinion environment—are used to construct the public opinion risk assessment index system. The same is true of the emergency evaluation system of risk of hot event created by Wang Hua, which establishes a secondary index evaluation system that is more detailed than the ordinary first-level index evaluation system [2].

The former quantifies the indicators by combining the analytic hierarchy process based on the Delphi method. At the same time, the latter uses the optimized DEMATEL-ANP weight determination method to obtain the correlation between the secondary indicators in the evaluation system to determine the indicator weight. Scholars usually add NLP technology for sentiment analysis for the model to have better training effects. By crawling the text data of related topics and using word vector technology to vectorize the text, scholars can obtain the corresponding emotional labels. Scholars usually attach various labels to text data to improve accuracy and facilitate machine learning. However, manual annotation work is time-consuming and labor-intensive. Raza Shaina et al. proposed a semi-autonomous annotation method to reduce the workload and successfully applied it in the NBias model [3].

Rumors and disinformation in the data also need to be distinguished and flagged. In other studies, these data are not used. However, in online public opinion, rumors and false information are some of the major causes of public opinion, so the model needs to be able to identify them. This also requires NLP technology, such as transformer-BERT technology or text feature and graph neural network technology (FakeGFN), and can improve the effect through contextual embedding, semantic enhancement, and word order information weighting [4,5]. Generative adversarial networks (GANs) can also be used to generate fake reviews, providing the dataset required for training.

3.2. Information dissemination

The spread of online public opinion is the online flow and transmission of information. Abstract it into a graph: the nodes are the users on each social platform, and the edges are the social interactions between users. American psychologist Malcolm Gladwell proposed the tipping point theory, which means that people's ideas will sometimes spread rapidly, thus affecting social trends, and the starting point of the trend is the tipping point. Based on this theory, Jiang Wu et al. proposed an analysis of public opinion communication networks based on the user dimension [6]. Taking Weibo as an example, the weight on the top of the graph is the number of Weibo reposts between users, and users are divided into node types according to Weibo user authentication type and number of fans for research. Similar indicators are all social media indicators.

Wen Tao et al. chose network analysis indicators for influence analysis and proposed a more specific evidential reasoning-based influential user evaluation (ERIUE) model [7]. He summarized the commonly used node centrality measurement methods and integrated them into the model, such as classic centrality, closeness centrality, and iterative refinement centrality. Select the threshold model (nodes will only start spreading on their own after being exposed to a certain number of other nodes spreading information) and the independent cascade model (nodes in the set will activate their inactive neighbor nodes with a certain probability) as diffusion The model is used to simulate the flow of information, and after normalizing all the centralities obtained, the overlap, similarity, and conflict relationships between centralities are processed to obtain three weights and averaged. Finally, the influence of the node is brought.

The network structure limits the above methods and needs to consider other factors such as user behavior, content quality, and topic popularity. In large-scale networks, for relatively large amounts of calculations, scholars can use big data distributed storage technologies and distributed computing technologies such as Hadoop and Spark to decompose and refine the number of calculations to achieve better results. Through the obtained user influence, scholars can stratify the monitoring of online public opinion, continuously monitor users with strong impact, and issue early warnings or restrictions if necessary.

3.3. Algorithmic fairness

"Bias" is often used in philosophy, psychology, and sociology. After the advent of artificial intelligence, this concept has also begun to exist in machine learning, like a famous law in computer science: Garbage In, Garbage Out. The abbreviation is "GIGO law," which means that if the input is

garbage data, the output will also be garbage data. If the training data contains bias, there is a high probability that the output results will also be biased. This phenomenon is widespread; platforms rely on algorithms to create price discrimination, and image recognition systems cannot identify black people. This problem also exists in the field of online public opinion. For example, there are many comments about discrimination against minority groups on the Internet in the training data. Through learning, the machine concludes that this part of the speech is normal, ultimately ignoring it in public opinion monitoring.

Equity should be a principle that people can accept even if they are in the most disadvantaged position in society. But this is difficult to achieve because such fairness is also suspected of being unfair to other social status groups. Usually, scholars simplify it to not making decisions directly because of some sensitive or protected attributes. Barry Laurence proposed three types of bias [8]. Type I bias refers to those categories that do not reflect the reality of risk but are purely based on bias; Type II bias refers to those categories that reflect proven statistical reality but not causal relationships; Type III bias refers to those categories that reflect statistical and causal relationships. The fact is that this reality itself is caused by upstream social discrimination or may trigger new discrimination.

For example, the skin color attribute belongs to type I prejudice, which is purely caused by racial discrimination. The gender attribute belongs to type II bias. It correlates with the result you want to predict, but it is not a causal relationship. Using such an attribute is suspected of being arbitrary. Genetic attributes belong to type III bias. Many related characteristics of people are determined by genes, such as a person's appearance and intelligence. In theory, including everyone's genetic attributes in the data will improve the model. But the result is new discrimination, showing what genes are good and harmful. This classification is terrible and must be resisted.

Current research on algorithmic fairness is more about restricting fairness conditions. For example, Wenqiong Zhang et al. have developed seven types of fairness indicators based on probabilistic and non-probabilistic statistics [9]. Different fairness indicators are applicable in different scenarios, and the fairness indicator is chosen based on the conditions of the original distribution. Favier Marco proposed a fair world framework [10]. When selecting training data, use unbiased data, inject different biased data, combine other appropriate conditions for learning, and finally transplant it to large-scale data. Different fairness indicators may conflict. It is only possible to meet some fairness indicators and scholars can only prescribe the right medicine according to the actual situation.

What these group fairness techniques fail to take into account is social norm bias [11]. Norms are not precisely the same as stereotypes. Usually, the norms recognized by society are correct, while the prejudices caused by stereotypes are wrong. However, machine learning cannot easily distinguish the two. The existing method is to select a small number of attributes and constrain the algorithm based on the observed social characteristics. This method has limitations. As more attributes are set, the constraints will become more and more complex. A stereotype is an oversimplified and stereotyped perception of a particular group that is not necessarily based on a code of conduct. Therefore, in the era of big data, it may be possible to observe the behavior of specific groups and identify norms and stereotypes based on the facts of user behavior.

4. CONCLUSION

Artificial intelligence has been developing rapidly in recent years and is cross-integrating with other fields. At the same time, artificial intelligence is also receiving attention from society. Regarding large models, ChatGPT, New Bing, and ERNIE Bot have sprung up like mushrooms after rain and are widely used by people. Various intelligent algorithms have changed people's production and lifestyle and have a disruptive effect. Therefore, some say the fourth industrial revolution will be the artificial intelligence revolution.

After social groups accept artificial intelligence, whether it is fair also arises. This article describes the sources of cognitive biases and classifies and discusses the issues from a technical perspective. However, algorithmic fairness cannot be achieved only by improving technology; other forces are needed to solve these problems. It requires collaborative research in philosophy, law, sociology, journalism, and other disciplines and the assistance of relevant platforms and departments. Educate the public and improve their media literacy and information literacy; strengthen the management of social media platforms, strengthen the real-name system, and establish an effective content review mechanism; formulate relevant laws and regulations to hold accountable for improper algorithms; establish a community participation mechanism to allow interests Stakeholders provide feedback. Of course, how to better promote the implementation of this series of measures requires further research.

REFERENCE

[1] Jin, D. X., Xia, Y. X., Bai, P. Y. et al. (2023) Research on risk assessment of emergency network public opinion for multi-dimensional risk scenarios. Information Research, (05), 17-24.

[2] Wang, H., Luo, L., Qiao, J. (2023) Research on fuzzy comprehensive evaluation of risk of hot events on internet public opinion. Information Studies: Theory & Application, (11), 126-132.

[3] Raza, S., Garg, M., Reji, D. J. et al. (2024) Nbias: A natural language processing framework for BIAS identification in text. Expert Systems with Applications, 237, 121542.

[4] Pelrine, K., Danovitch, J., Rabbany, R. (2021) The surprising performance of simple baselines for misinformation detection. In: Proceedings of the Web Conference 2021. pp. 3432-3441.

[5] Zhang, H.R. (2022) Research on fake information detection methods based on text feature and graph neural network. (Master's Degree, Central University of Finance and Economics)

[6] Wu, J., Huang, X., He, C. C. et al. (2023) Propagation and evolution of public opinion in the outbreak of AIGC based on the theory of tipping point. Journal of Modern Information, 43, 145-161.

[7] Wen, T., Chen, Y.W., abbas Syed, T. et al. (2024) ERIUE: Evidential reasoning-based influential users evaluation in social networks. Omega, 122, 102945.

[8] Barry, L., Charpentier, A. (2023) Melting contestation: insurance fairness and machine learning. Ethics and Information Technology, 25(4), 49.

[9] Zhang, W. Q., Li, Y. (2024) Fairness metrics of machine learning: review of status, challenges and future directions. Computer Science, (01), 266-272

[10] Favier, M., Calders, T., Pinxteren, S. et al. (2023) How to be fair? a study of label and selection bias. Machine Learning, 112(12): 5081-5104.

[11] Cheng, M., De-Arteaga, M., Mackey, L. et al. (2023) Social norm bias: residual harms of fairness-aware algorithms. Data Mining and Knowledge Discovery, 37(5), 1858-1884.