Social Science, Humanities and Sustainability Research
ISSN 2690-3628 (Print) ISSN 2690-3636 (Online)
Vol. 4, No. 3, 2023
www.scholink.org/ojs/index.php/sshsr

Original Paper

Exploring the Application Potential of the Large Language Model in Sociological Research: A Case Study of ChatGPT

Ke Liu[1]

[1] Assets and Equipment Department, Xi'an Medical University, Xi'an, Shaanxi, China

Received: April 30, 2023 Accepted: May 23, 2023 Online Published: June 15, 2023

doi:10.22158/sshsr.v4n3p111 URL: http://dx.doi.org/10.22158/sshsr.v4n3p111

Abstract

With the exponential growth of Big Data and Artificial Intelligence (AI), large Language models have emerged as powerful tools in sociological research. As an advanced large Language model, ChatGPT has the unique ability to generate natural language dialogues, opening up new avenues for sociological research. This paper employs ChatGPT as a case study to explore the application potential of large Language models in sociological research, including social network analysis, sentiment analysis, opinion mining and social public opinion monitoring. Additionally, we discuss the various challenges encountered by large Language models in sociological research, such as privacy, ethical, interpretability and credibility issues. Finally, some recommendations to overcome these challenges and prospective research directions are provided.

Keywords

Large Language Model, Sociology, Application Potential, ChatGPT

1. Introduction

Sociology (Weber, 1946), which studies human social behavior, interactions and systems, has traditionally utilized qualitative and quantitative research methods to achieve an in-depth understanding of social phenomena. However, the advent of large language models (Weber, 1978) has provided innovative tools and opened up new avenues for sociological research. Large language models, which are natural language processing models based on deep learning (Qian, Dong, Wang, & Tan, 2015), consist of a vast number of parameters. These models rely on deep neural network structures and learn the language's structure, syntax, semantics, context relevance and other features after being trained on large-scale text datasets. With the rapid development of Artificial Intelligence (AI), large language models have emerged as essential tools for sociological research by providing new directions of exploration.

This paper aims to investigate how ChatGPT, an advanced language model developed by OpenAI, can potentially be applied in sociological research. ChatGPT excels in generating coherent and context-sensitive natural language dialogues and it has a wide range of potential applications such as intelligent assistant, automatic question answering, text generation, content recommendation, semantic analysis, emotional analysis, public opinion monitoring and automated writing. Since the emergence of large language models, researchers from various disciplines have been exploring their potential applications in various fields, and relevant scholars have emphasized the significance of language models in social science research. For instance, Smith and Anderson (2020) conducted in-depth analysis on social media data and emphasized the transformative impact of large language models on social science research. Jones et al. (2021) demonstrated the potential of language models in analyzing public opinion and sentiment.

This paper discusses the benefits, limitations and ethical considerations of utilizing the large language model for sociological research by exploring the specific applications of ChatGPT in this field. Social network analysis, emotional analysis, opinion mining and social media monitoring are major sociological applications. Based on the existing research, this paper aims to provide an in-depth analysis of the potential benefits and limitations of using ChatGPT in sociological research. Moreover, the ethical considerations are highlighted that arise when using large language models for sociological research.

2. Overview of Large Language Models

2.1 Developments of Large Language Models

The large language model is a deep learning technology that has great potential in various fields, such as natural language processing, text generation, and dialogue systems.

The development of the large language models has gone through several key stages. Initially, language models were based on n-gram statistical methods, which generated text based on the frequency of words or characters. However, this method was limited by the size of the glossary and context limitations and it could not capture long-distance dependencies. With the advent of deep learning, neural network-based language models (Morin, & Bengio, 2005) have become mainstream, with recurrent neural networks (RNN) and long short-term memory networks (LSTM) commonly used model architectures capable of handling the context dependency of sequence data. These models learn the hidden state in the language sequence to generate coherent text by modeling the current input and context. Recent advancements include large-scale pre-trained models like OpenAI's GPT (Generative Pre-trained Transformer), whose performance has significantly improved. These models use the transformer architecture that can process input sequences in parallel, efficiently capture long-distance dependencies and adapt to different tasks through pre-training and fine-tuning.

2.2 Principles and Techniques of Large Language Models

The fundamental principle of the large language model is to model the probability distribution of natural language using deep learning technology to achieve text generation and understanding. In essence, the model predicts the likelihood of the next word or character based on statistical patterns and language constructs learned from a vast corpus to generate coherent text.

The technology primarily includes neural network architecture, pre-training and fine-tuning, autoregressive generation and conditional generation. Neural networks are used to process text data in large language models, with common architectures including RNN and LSTM, which capture contextual dependencies in sequence data. Large language models typically adopt pre-training and fine-tuning strategies. In the pre-training stage, the model is trained on a vast unsupervised corpus to learn the statistical laws and contextual information of the language. In the fine-tuning stage, the model is adjusted and optimized through supervised training on specific tasks. Autoregressive analysis is employed by large language models to generate text. The model uses input text information to predict the next word or character based on a probability distribution and generates a coherent text sequence gradually by using it as part of the input. Moreover, large language models can generate text under given conditions by incorporating conditional information into the generation process, such as generating answers in response to a given question. Furthermore, the characteristics of the generated text can be controlled by adjusting model parameters such as style, emotion, or length. Optimizing algorithms or setting specific constraints and guidance can accomplish parameter adjustments.

2.3 Applications of Large Language Models in Other Fields

Large language models have found extensive applications in various fields, providing intelligent language interaction and information processing experiences by processing complex natural language tasks and generating high-quality text. These fields include literature, academia, social media, law and compliance, among others.

The large language model has demonstrated great potential in generating text and facilitating various text types such as news articles, stories, poetry, music lyrics, and simulating human creative styles. Furthermore, it can improve the accuracy and effectiveness of search engines in information retrieval and recommendation systems by understanding users' query intentions and text content, providing more relevant and precise search results and generating personalized recommendations based on users' interests and preferences. Additionally, it can answer user questions, provide advice and solve problems by simulating human conversations and virtual assistants' capabilities within dialogue systems. In the field of social media, the large language model can perform semantic and emotional analysis, enabling businesses, governments and academia to understand user perspectives and behaviors, identify emotional tendencies, and analyze topic trends and social network relationships, among others. Moreover, it can assist in tasks such as automatic summarization of legal documents, generation of compliance policies and analysis of contract terms in the field of law and compliance. Through learning and understanding legal texts, it can provide fast and accurate legal and compliance support.

3. Specific Applications of Large Language Models in Sociological Research

3.1 Social Network Analysis

Social network analysis (Alvin & Wolfe, 1997) is a vital field of sociological research, which discloses social phenomena such as social structure, information dissemination and the spread of influence by studying the relationship network between individuals. Large language models have provided new research tools and methods for social network analysis.

ChatGPT can be applied to multiple aspects of social network analysis. Firstly, it can generate social network data, simulate human social behavior and generate social relationship data with authenticity and diversity, which facilitates the verification of social network theory and the implementation of various social network experiments. Secondly, by learning existing social network data, ChatGPT can predict yet to be established social network relationships, assisting users in understanding the evolution process and structural characteristics of social networks. It can also simulate the dynamic evolution process of social networks, generate future social network snapshots and reveal the changing trends and evolution laws of social networks. Furthermore, ChatGPT can perform semantic and emotional analysis on text data in social networks, understand text content on social media, identify topic trends, emotional tendencies and interaction patterns between users, revealing phenomena such as information dissemination, public opinion formation and social interaction.

3.2 Sentiment Analysis and Opinion Mining

Sentiment analysis (Zhang, 2022) and opinion mining (Li & Lee Carman Ka Man, 2022) are critical research fields in sociological research that help us comprehend the emotional states, opinions and attitudes of individuals and social groups. With the ChatGPT model, trained sentiment analysis can identify and classify emotional tendencies in text, such as positive, negative, or neutral emotions to assist us in understanding the emotional attitudes of social media users regarding specific topics or events. Moreover, for opinion mining, ChatGPT can analyze sentence structure and semantic information in the text, identify key viewpoints and assist in comprehending people's positions and opinions on specific issues. Additionally, ChatGPT can recognize the intensity and expression of emotions in the text. By analyzing the vocabulary, intonation, and syntactic structure of the text, it assists in understanding the degree and expression of emotions, thereby enabling a better understanding of the emotional state and changes of social groups.

For instance, the ChatGPT model can recognize comments such as "I really like this brand's products, they are of good quality and perform well!" and categorize them as positive emotions, capturing the high level of users' evaluations of the products. It can also categorize comments such as "The service of this brand is too poor to be worth buying" as negative emotions, identifying users' dissatisfaction with the brand's services. By analyzing multiple comments of such nature, we can draw conclusions about the emotional distribution and perspective trends of the brand, understand users' attitudes and opinions towards the brand and provide vital references for enterprises to formulate brand strategies and improve products and services. Overall, the utilization of ChatGPT for sentiment analysis and opinion mining is

of significant importance in sociological research. By analyzing social media texts and other large-scale text data, we can gain deep insight into social groups' emotional attitudes and opinions, which assists in revealing public opinion formation, changes in social groups' attitudes and the development of related social phenomena.

3.3 Public Opinion Monitoring and Analysis

In sociological research, monitoring and analyzing public opinion (Ma, Zhang, & Unit, 2017) in society is an important task, which helps understand emotional attitudes, dynamics of public opinion and influence of social events on societal groups. Training ChatGPT model can provide the basis for real-time public opinion monitoring and analysis of specific topics, events or brands on social media. This helps to grasp public opinion dynamics, observe event evolution and evaluate societal attitudes and reactions to specific matters. Furthermore, analyzing expressions, intensity and key words related to emotions with the ChatGPT model can help us identify key viewpoints and emotional tendencies that offer a deeper comprehension of social events and topics, revealing the overall positive and negative tendencies of public opinion, the distribution of viewpoints, and emotional attitudes of the public opinion. Utilizing ChatGPT to analyze public opinion in society can also uncover the social network structure and information dissemination model, and provide novel approaches and perspectives for sociological research.

Monitoring and analyzing public opinion in society with ChatGPT has great significance in sociological research because it can provide a greater comprehension of social groups' emotional attitudes, viewpoints, and information dissemination mechanisms by analyzing large-scale text data sources like social media and news platforms.

4. Challenges of Large Language Models in Sociological Research

4.1 Ethical Issues

The use of the large language model in sociological research (Ida Asadi, Michael, Christoph, & Graeme, 2019) raises ethical issues, particularly regarding its impact on individuals and societal groups. The output of the large language model may have potential social consequences, affecting public opinion, opinion formation and decision-making processes. This raises concerns about the extent to which the model shapes and influences social cognition, values, and behavior. The use of the large language model must ensure fairness and impartiality by avoiding biased analysis and modeling of specific groups or social phenomena that may lead to unfairness and discrimination. Attention should be paid to whether emotional expressions of specific groups are ignored or discriminated against, to avoid generating unfair analysis results. It is therefore essential to carefully select representative and inclusive data samples and to adopt appropriate data processing and model training methods to reduce bias and discrimination possibilities.

Researchers should thus ensure ethical compliance, social responsibility, as well as protect the rights and interests of human participants involved in the research.

4.2 Privacy Issues

The use of large language models raises privacy challenges in sociological research (Munmun, Sandip, Samiran, Kumar, & Sachin, 2022). With the processing of massive textual data, researchers must pay careful attention to the protection of personal privacy. When using a large language model, the collected data must be desensitized appropriately to protect users' identity and personal information. Additionally, data storage and sharing approaches should also be taken into consideration and adequate security measures should be applied to prevent data breaches or unauthorized access. In research design, privacy protection principles should be adhered to. Individual autonomy and the privacy rights of participants should be respected and participants should receive a clear explanation of the purpose of data collection and use, obtaining informed consent.

Researchers need to take measures to anonymize personal information, such as encrypting or replacing the user identity information to ensure that data cannot be linked to specific individuals during the analysis process when analyzing user comments and interactions using the ChatGPT model. Access to sensitive content must also be restricted, and only necessary data should be used for research purposes. Through these privacy protection measures, research requirements and the protection of personal privacy can be balanced, providing valuable insights for sociological research while respecting the privacy rights of participants.

4.3 Interpretability and Credibility Issues

Large language models face challenges regarding interpretability and credibility in sociological research. Large language models have a black box nature, and researchers and sociologists may find it hard to comprehend their internal processing and reasoning, particularly when the model generates a large amount of text, which leads to a lack of interpretability. This lack of interpretability affects the understanding and interpretation of the model results, as well as the credibility of the research conclusions. Additionally, the credibility of the model itself can be affected by biased, erroneous or misleading outputs that affect the reliability and accuracy of research conclusions. To address these issues, researchers need to explore the interpretability and credibility of the model and improve its interpretability and accuracy. This can be achieved by developing interpretive tools, evaluating and validating the model, and collaborating with domain experts to ensure reliable and valid research results.

For instance, specific social groups may have their voices amplified due to the model's reliance on a large amount of data and automated algorithms during model training, leading to inaccurate or misleading results that can negatively affect research conclusions. Therefore, researchers need to evaluate and verify the credibility of the model to improve the understanding and perspectives of social phenomena and improve the reliability and validity of research conclusions.

5. Conclusion

The large language model has great potential in education and sociology, but it also faces a series of challenges and risks. While promoting its development and application, we must pay attention to ethical

norms, privacy protection, and fairness among other issues to ensure a positive impact on sociology and promote sustainable development in human society. The research conducted in this paper suggests that the large language model has extensive application potential in sociological research. Large language models, such as ChatGPT, are conducive to conducting more profound analysis and understanding in the sociology field. However, it also encounters several difficulties such as ethical issues, privacy protection, fairness, interpretability and credibility of model output.

Moreover, relevant difficulties must be addressed in future research to promote the use of large language models in sociological research. Standardizing ethical issues, improving privacy protection techniques, and enhancing interpretability and credibility are potential areas of focus.

References

Alvin, W., & Wolfe. (1997). *Social network analysis: Methods and applications*. American Ethnologist.

Ida Asadi, S., Michael, J. D., Christoph, F. B., & Graeme, G. S. (2019). *Ethical Issues in Big Data Analytics: A Stakeholder Perspective*. CAIS.

Jones, A. B., Smith, C. D., & Williams, E. F. (2021). Language models for sentiment analysis in social media: An empirical analysis. *Journal of Computational Social Science*, *4*(2), 247-265.

Li, S. M., & Lee Carman Ka Man. (2022). Transfer-Learning-Based Opinion Mining for New-Product Portfolio Configuration over the Case-Based Reasoning Cycle. *Applied Sciences*, *23*. https://doi.org/10.1007/978-3-031-12762-5

Ma, L., Zhang, J., & Unit, P. (2017). *Research and design of local government network public-opinion monitoring system*. Communications Technology.

Morin, F., & Bengio, Y. (2005). *Hierarchical probabilistic neural network language model*. Aistats.

Munmun, B., Sandip, R., Samiran, C., Kumar, D. A., & Sachin, S. (2022). A comprehensive survey on online social networks security and privacy issues: Threats, machine learning-based solutions, and open challenges. *Security and Privacy*, *1*.

Qian, Y., Dong, J., Wang, W., & Tan, T. (2015). Deep learning for steganalysis via convolutional neural networks. *Proceedings of SPIE - The International Society for Optical Engineering*, 9409, 94090J-94090J-10. https://doi.org/10.1117/12.2083479

Smith, J., & Anderson, K. (2020). Unleashing the power of large language models for social science research. *Social Science Quarterly*, *101*(4), 1632-1651. https://doi.org/10.1111/ssqu.12825

Weber, M. (1946). *From max weber: essays in sociology*.

Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. Univ. of California Press.

Zhang, Z. (2022). Research on Sentiment Analysis Model of Douban Book Short Reviews based on Machine Learning. In *Proceedings of 5th International Workshop on Education Reform and Social Sciences (ERSS 2022)* (pp. 318-322). BCP Social Sciences & Humanities. https://doi.org/10.54691/bcpssh.v20i.2336