

令和 6 年 6 月 9 日現在

機関番号：11301

研究種目：研究活動スタート支援

研究期間：2022～2023

課題番号：22K20182

研究課題名（和文）Investigation of Opinion Polarization in Online Communication: Towards an Integration of Explanation and Prediction

研究課題名（英文）Investigation of Opinion Polarization in Online Communication: Towards an Integration of Explanation and Prediction

研究代表者

呂 沢宇 (LYU, ZEYU)

東北大学・文学研究科・准教授

研究者番号：30966312

交付決定額（研究期間全体）：（直接経費） 2,200,000円

研究成果の概要（和文）：本研究は、ビッグデータと計算的手法を用いて、ソーシャルメディアにおける意見の形成メカニズムに関する実証的な分析を行った。これにより、個人が自らの党派的な同一性を維持しようとする傾向が、対立する党派に対する敵意につながり、それが分極化の一因となることを明らかにした。その他、意見の分極化を究明するために、大規模言語モデルの応用可能性を提示した。具体的には、特定の党派的傾向や認知バイアスを持つ個人を模倣することで、社会シミュレーションに大規模言語モデルを取り入れる方法を提案し、意見の分極化のメカニズムについてより深い洞察を得る可能性を示した。

研究成果の学術的意義や社会的意義

ソーシャルメディアをはじめとするオンライン場における意見の分極化は、集団間の意見の隔たりを深め、社会的な分断を助長することが問題視されている。本研究が明らかにするオンライン場における意見形成の知見は、意見の分極化を軽減するための適切な政策確立において重要な基礎となる。また、本研究で提案した大規模言語モデルによる社会シミュレーションの手法は、新たな研究の方向性を開拓した点で学術的意義が高いと考えられる。

研究成果の概要（英文）：(1) This project uses big data data and computational methods to investigate the mechanism of opinion expression on social media. We find the individuals' tendency to maintain their partisan identification can lead to hostility toward individuals of opposing partisans, serving as a potential mechanism that contributes to polarization.

(2) This project explores the possibilities of LLMs (Large Language Models) in research related to opinion polarization. We have validated and demonstrated that LLMs can mimic individuals with specific partisan leanings and cognitive biases. This implies that LLMs can be used in social simulations to enrich the fidelity and complexity of simulations, potentially providing deeper insights into the mechanisms of opinion polarization.

研究分野：計算社会科学

キーワード：計算社会科学 意見分極化 ソーシャルメディア ビッグデータ 自然言語処理

1. 研究開始当初の背景

Recent years have seen an increase in attention to the issue of opinion polarization in online communication. Previous studies have shown that online communication can increase the exposure to cross-cutting information and empower people to express diverse views (Brundidge, 2010), while high-choice online discourse may also function as an echo chamber that accentuates connections with like-minded others and causes fragmentation (Stroud, 2010). Since previous studies have reported conflicting findings regarding the mechanism of opinion polarization in online discourse, further research are needed to enhance the generality and robustness of the conclusions.

Moreover, research should not only describe general explanations but also make high-quality predictions about changes in opinion polarization in online communication. More efforts are needed to develop a more sophisticated and high-fidelity simulation model of opinion dynamics and enhance the interpretability and explainability of simulation.

2. 研究の目的

Using a computational social science framework, the research aims to explore the mechanism of opinion polarization and explore how to incorporate the implications from empirical studies into the simulation models.

3. 研究の方法

To achieve these aims, this research mainly conducted two projects.

(1) Empirical analysis on online communication based on Twitter data

This research employs Twitter data and computational method including network analysis and natural language processing techniques to explore the nature and mechanism of online communication. Specifically, we collected large-scale data including Twitter users' following network, profile information, and published content. Then, based on the latent space model assumes that Twitter users are predisposed to follow politicians with whom they share ideological similarities and avoid those with whom they disagree ideologically (Barberá, 2015), this study estimates the ideological positions of users through the computations of the follower–followee relationship on Twitter. Also, we use deep-learning-based text classification method to detect the expressed hostility in their published content. In this term, this study empirically examines how individuals interact with opposing partisans in political discussions.

(2) Investigation on the role-play ability of Large Language Models (LLMs)

Simulating opinion polarization requires that the agents in the simulation models are capable of role-playing characters with various profiles and mimic human behaviors. This research explore whether agents can be endowed with beliefs, experiences, and characteristics through LLMs, discussing its potential applications in the simulation of opinion polarization. We craft prompts to create several LLM-powered agents with specific partisan alignments and cognitive characteristics. Then, agents are exposed to congenial or uncongenial persuasive message and required to report and reasoning their opinion changes. In this term, we examine how partisanship and confirmation bias in the prompts can systematically affect the agents' opinions across various policy issues and exposed information.

4. 研究成果

(1) Partisan identity and inter-party hostility in cross-cutting interactions

This study utilizes social media data and deep-learning-based text classification methods to investigate cross-cutting interactions on social media. Specifically, we attempt to investigate whether the prevalence of offensive speech can be explained by the partisan identity of the message sender and target in an interaction, as well as how such a mechanism of political incivility varies between politicians and ordinary people. Our findings reveal that people are more likely to use offensive speech in response to content published by opposing partisans. Furthermore, we demonstrate how inter-party hostility is associated with the partisan identity of both the message sender and the target in the interaction. On one hand, the findings indicate that strong partisans and people who publicly assert their partisan identities tend to attack opposing partisans, suggesting a relationship between the salience

of partisan identity and value defense mechanisms. On the other hand, strong partisans, especially politicians, are more likely to be the target of offensive speech from opposing partisans. The disparity in the extent of received offensive speech is argued to result from individuals' tendency to maintain their partisan identification by expressing hostility toward representative individuals of opposing partisans. This study has been published in *New Median and Society*.

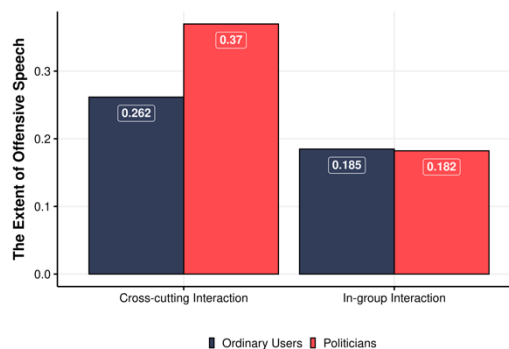


Figure 1. The extent of offensive speech in different types of interactions (Lyu, 2023).

(2) Leveraging Large Language Models to Understand the Opinion of Polarization

This research demonstrates that including descriptions of partisan identity and cognitive bias into prompts can equip agents with in-group favoritism, out-group hostility preferences, and confirmation bias, leading to significantly different responses to congenial and uncongenial information (Figure 2). This experiment indicates that LLMs can help us to create agents with build-in knowledge and ability to mimic complex and nuanced human behaviors and opinion preferences, thus have potential to set a new precedent in the realm of social simulation with their abilities in processing, generating, and interacting with human language in a contextually informed and semantically accurate manner

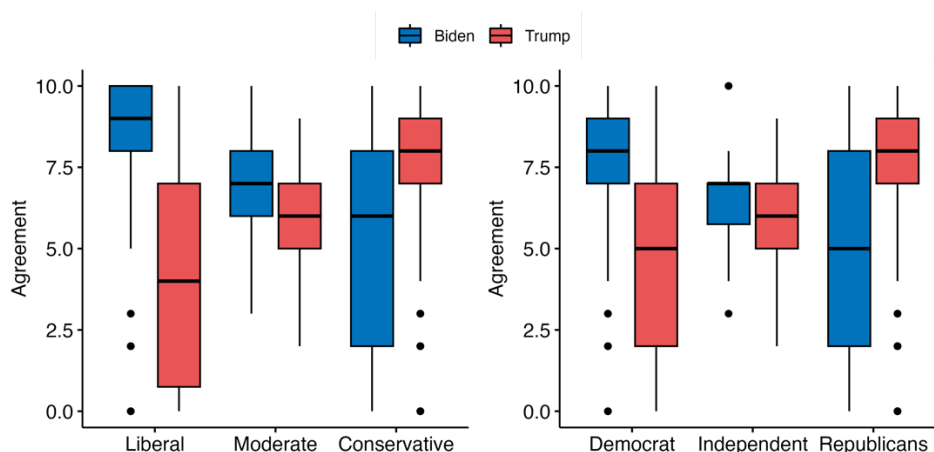


Figure 2. Agreement with policies supported by *Biden* and *Trump* across various ideologies and party affiliations

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 瀧川 裕貴、永吉 希久子、呂 沢宇、下窪 拓也、渡辺 誓司、中村 美子	4. 巻 73
2. 論文標題 ソーシャルメディア言論分析の方法 2020年1月から辞任までの安倍首相に対するTwitter上の投稿分析を事例として	5. 発行年 2023年
3. 雑誌名 放送研究と調査	6. 最初と最後の頁 70~85
掲載論文のDOI（デジタルオブジェクト識別子） 10.24634/bunken.73.3_70	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Lyu Zeyu	4. 巻 なし
2. 論文標題 Cross-cutting interaction, inter-party hostility, and partisan identity: Analysis of offensive speech in social media	5. 発行年 2023年
3. 雑誌名 New Media and Society	6. 最初と最後の頁 なし
掲載論文のDOI（デジタルオブジェクト識別子） 10.1177/14614448231180654	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 瀧川 裕貴、永吉 希久子、呂 沢宇、下窪 拓也、渡辺 誓司、中村 美子	4. 巻 73
2. 論文標題 ソーシャルメディア言論分析の方法2-安倍首相に対するTwitter上の投稿分析を事例としたトピックモデルの適用	5. 発行年 2024年
3. 雑誌名 放送研究と調査	6. 最初と最後の頁 26-43
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計3件（うち招待講演 2件/うち国際学会 2件）

1. 発表者名 Zeyu Lyu
2. 発表標題 Cross-Cutting Interaction, Out-Group Hostility, and Partisan Identity: Investigation of Affective Polarization on Social Media
3. 学会等名 2022 International Conference on Computational Social Science（国際学会）
4. 発表年 2023年

1. 発表者名 呂沢宇
2. 発表標題 大規模言語モデルの心理学における応用
3. 学会等名 日本理論心理学会第69回大会（招待講演）
4. 発表年 2023年

1. 発表者名 Zeyu Lyu
2. 発表標題 Leveraging Large Language Models to Understand the Opinion of Polarization
3. 学会等名 国際ワークショップ「計算社会科学の社会的基盤」（招待講演）（国際学会）
4. 発表年 2024年

〔図書〕 計1件

1. 著者名 Yoshimichi Sato, Hiroki Takikawa	4. 発行年 2024年
2. 出版社 Springer	5. 総ページ数 250
3. 書名 Sociological Foundations of Computational Social Science	

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------