Recommendation AI in Social Media: Consequences and Mitigations

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Raj Asthana

Fall 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Rosanne Vrugtman, Department of Computer Science

ABSTRACT
Social media recommender system algorithms are large sources of information to many people, but their content curation leads to content polarization which can form echo chambers. To stop these side effects, I propose the use of natural language processing to curate sources that would instead promote a more diverse understanding of events. Natural language processing (NLP) can be used in many ways to try and diversify feed recommendations. Sentiment analysis can be used to analyze the sentiment towards a topic and avoid recommending content that skews towards extreme sides. Models can also be trained to detect to recognize words that are indicative of bias and flag them so recommender systems present alternate viewpoints. Additionally, clustering could be used with these methods, and recommender systems could be required to recommend media from different clusters. In using these methods, the content curated for users can be more diverse. This can prevent heavily one-sided media from repeatedly appearing on feeds and a subsequent feedback loop of polarization that leads to echo chambers forming. However, this will have to be done carefully to maintain the strong user engagement that recommender systems present, so the platform can remain engaging.

1. INTRODUCTION

The usage of social media has been rapidly increasing in recent times. A study by Keipos suggests that there are approximately 4.76 billion users on social media platforms, though this figure may be somewhat inflated due to the presence of duplicate accounts. A key factor in the sustained engagement of users on these platforms is the role of recommendation algorithms. These algorithms are designed to present content tailored to the individual preferences of users, often without the need for active searching. This tailored content delivery is a primary reason many users find themselves continuously exposed to materials they prefer. Given the extensive reach of social media, these algorithms play a significant role in shaping the information that a large segment of the population encounters, thereby influencing their perceptions and beliefs.

The impact of these algorithms is particularly noticeable when users engage with political or opinion-based content. The more a user interacts with such content, the more the social media system is inclined to present similar materials. This pattern can lead to a gradual increase in the polarization of the content being shown. Moreover, this personalized content may also be suggested to the user's network, amplifying the effect, and leading to the creation of 'echo chambers.' These echo chambers tend to circulate and reinforce one-sided perspectives within a social network. Given the widespread usage of social media, the presence of these echo chambers poses a challenge to the public's ability to understand and interpret events from a balanced perspective.

2. RELATED WORKS

With polarization and echo chambers being formed by social media algorithms, it becomes easy for misinformation to propagate through social networks. Murugasen (2019) proposes using NLP to detect fake news. Murugasen was able to get strong results that had few false positives. This approach leverages advanced machine learning algorithms to analyze and classify news content, distinguishing between genuine and fake news. By integrating such NLP techniques into social media algorithms, the recommendation AI can be significantly

improved. Instead of merely promoting content based on user engagement or similarity of viewpoints, these enhanced algorithms could also evaluate the credibility and factual accuracy of content. This would help in reducing the spread of misinformation, as the recommendation systems would be less likely to amplify fake news.

To address the challenges posed by recommender systems in social media, particularly in forming echo chambers and polarization, Suhaim and Berri (2021) conducted a comprehensive review of context-aware recommender systems specifically developed for social networks. Their research highlighted the significance of integrating context into recommender systems, which traditionally rely on user feedback but often overlook the user's current context. This review systematically analyzed various studies between 2015 and 2020, focusing on the approaches, techniques, and context elements used in developing these systems. The key takeaway from their work is the realization that context-aware recommender systems can significantly improve the personalization and relevance of recommendations by considering the dynamic and rich data ecosystem of social networks. By accounting for factors like the user's location, time, activities and social connections, these systems can adapt recommendations to be more aligned with the user's current situation and interests. This shows how recommendation algorithms can be much more flexible with data, and information about events can be incorporated as context to help properly filter media that gets curated.

3. PROPOSED DESIGN

To mitigate the issues related to polarization that are caused by social media algorithms, a new modified recommender system is proposed. This system leverages natural language processing and other techniques to handle and prevent content polarization and echo chambers, aiming to provide the users with a more balanced feed that provides accurate information while maintaining the user engagement that social media provides.

3.1 Design

This proposal will feature three main steps to try and combat polarization and misinformation spread. They will be to use process data on social media to determine if polarized content is present, to recommend more variety in content, and to detect echo chambers and try and adjust recommendations when they are present.

3.1.1 Detecting Polar Content

To detect polar content, NLP models for sentiment analysis can be implemented. These models will evaluate the emotional tone and bias in user-generated content. The primary objective is to identify highly polarized content by assessing extreme sentiments in posts, comments, and shared articles. This data will then be used to flag and monitor polarizing trends within the platform, which will inform the recommendation system to provide a more balanced content distribution.

3.1.2 Enforcing Diversity in Content

Content clustering will be utilized to categorize content into diverse thematic and ideological groups. This will be achieved by employing clustering algorithms that analyze semantic content and user interaction patterns. The aim here is to ensure that users are exposed to a variety of perspectives, thus reducing the likelihood of only encountering reinforcing viewpoints. This clustering mechanism will be integrated into the platform's feed algorithm to diversify the content visible to users, effectively counteracting the echo chamber effect.

3.1.3 Echo Chamber Detection

The detection of echo chambers will involve the analysis of user interactions, content sharing patterns, and network dynamics. An NLP system will be developed for this purpose. The goal is to identify networks or user groups where information loops are homogeneous and lack diverse content. The strength of echo chambers will be continuously assessed, and this information will be used to adjust the recommendation algorithms, thereby introducing more varied content and perspectives.

3.1.4 Algorithm Adjustment and Integration

A dynamic recommendation system will be created by seamlessly incorporating the outputs from sentiment analysis, content clustering, and echo chamber detection into the existing recommendation engines. This system will aim to maintain a balance between personalized content and diverse, contrasting viewpoints. The recommendation algorithms will be continuously refined based on the insights gained from these analyses, ensuring a more holistic and less polarized user experience on the platform.

4. ANTICIPATED RESULTS

The implementation of this modified recommender system is anticipated to yield several positive outcomes in the social media environment. By detecting and mitigating the spread of polar content through sentiment analysis, we expect a significant reduction in the spread of extremist viewpoints and divisive content. This should contribute to a healthier, more constructive online discourse.

The enforcement of content diversity using clustering algorithms is likely to broaden users' exposure to a wider range of perspectives, which will foster a more informed and open-minded user base. The active detection and modulation of echo chambers will likely reduce the prevalence and impact of these phenomena on social media platforms. By identifying and adjusting the strength of echo chambers, the recommender system will provide a more balanced and less insular content feed to users.

Overall, integrating these approaches into social media algorithms is expected to promote a more balanced, informative, and less polarized online environment. It should enhance user experience by providing exposure to a broader range of content. The long-term effect could be a more informed, engaged, and tolerant social media community.

5. CONCLUSION

My proposal presents a novel approach to addressing the challenges of polarization and echo chambers created by social media recommender systems. By incorporating natural language processing (NLP) techniques, such as sentiment analysis, bias detection, and clustering algorithms, the proposed design aims to curate a more diverse and balanced content feed for users.

This approach is not just about mitigating misinformation or extreme views but also about enhancing the overall quality of discourse on social media platforms. The integration of NLP tools for detecting polar content, enforcing content diversity, and identifying echo chambers marks a significant shift toward more responsible content curation in social media. The proposed system seeks to maintain user engagement while simultaneously promoting a broader understanding of events and viewpoints. The expected results indicate a potential reduction in content polarization and a decrease in the prevalence of echo

chambers. This shift could lead to a more informed and diverse online community.

6. FUTURE WORK

The next steps will be to implement the proposed recommander system. In order to do this, extensive user data will need to be collected to train machine learning models on. This will include interactions with posts, user posts, and social network data. With this data, the NLP models for sentiment analysis, echo chamber detection, and also content indexing for clustering can be trained. After training these models, they can individually be tweaked and evaluated. Once ready, they can all be used to design the overall recommander system for a social media platform.

Once functioning, the algorithm will need to be monitored and updated. The user experience of the model can be compared to a control. The data can be used to update and further advance the algorithm. With this user data, a feedback loop of refining the algorithm and then seeing the user's response can be made to improve the algorithm, which should help the social media platform decrease the amount of polar content and echo chambers present.

REFERENCES

[1] Keipos (2023). Global Social Media Statistics. https://datareportal.com/social-media-users

[2] Suhaim, S., & Berri, J. (2021). Context-aware recommender systems for social networks: A survey. In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 25-30). IEEE. https://ieeexplore.ieee.org/document/9399450

[3] Murugesan, M. K. (2019). Natural language processing approach for fake news detection. [Master's thesis, National College of Ireland]. NORMA. https://norma.ncirl.ie/4157/1/manojkumarmurugesan.pdf