Understanding the Radicalizing Effects of Recommendation Algorithms

***Sasha Rabeno*, Dr. Larry Snyder**

Department of Industrial and Systems Engineering, Lehigh University

Abstract

Algorithms that recommend personalized content to users are a staple of nearly all social media platforms. However, these algorithms often push users towards ideologically extreme content, with the potential to escalate from on-screen hate to off-screen violence. This work explores the radicalizing effects of a scoring-based video recommendation system, modeled off the deep learning systems in place at YouTube. We created a scoring function to create a list of videos for a given user to watch that best align with their preferences. This research finds that with our scoring system in place, users are pushed to watch videos more extreme than they would in the absence of a recommendation system. As more extreme and provoking content increases engagement (and money made) for social media companies, the scoring equation emphasizes a video's extremeness—which we found to further polarize the videos users watched as this emphasis increases.

Discussion

In Fig. 1, users without the influence of the video scoring system watch videos of about average extremeness (0.56, compared to 0.5 being "average," non-partisan videos. However, with the scoring system, users already close to the political extremes (0.0 and 0.1) are pushed even further towards their personal extremes, and watch less mainstream videos (or videos that differ at all in extremeness from their tastes. In Fig. 2, users with the scoring system watch almost double the amount of videos than without. As the scoring system rewards shorter videos, users can fit more videos into a watching session—allowing social media companies to show users a larger variety of inflammatory content to engage with. This further aligns with the rise in "short form" content on social media websites, such as TikTok, YouTube's Shorts, and Instagram's Reels.

Acknowledgements

- David and Lorraine Freed Undergraduate Research Symposium
- The Clare Booth Luce Research Scholarship Program
- Josie & Charlotte Rabeno

Simulation Graphs

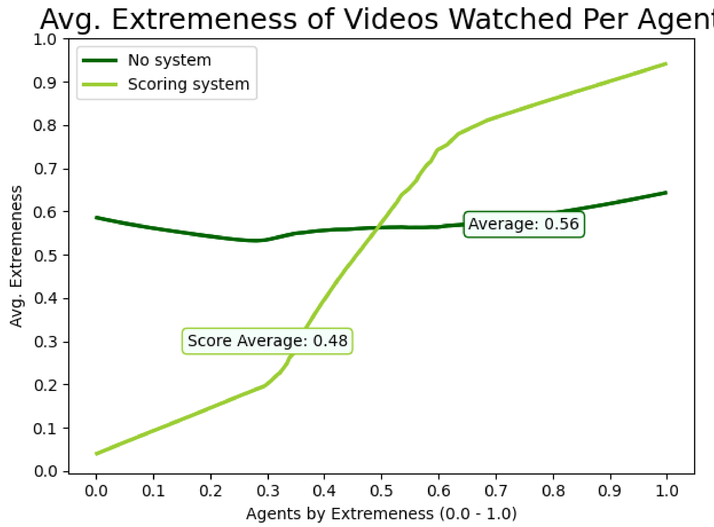

Fig. 1: Graph of the avg. extremeness a user watches, plotted against their own extremeness

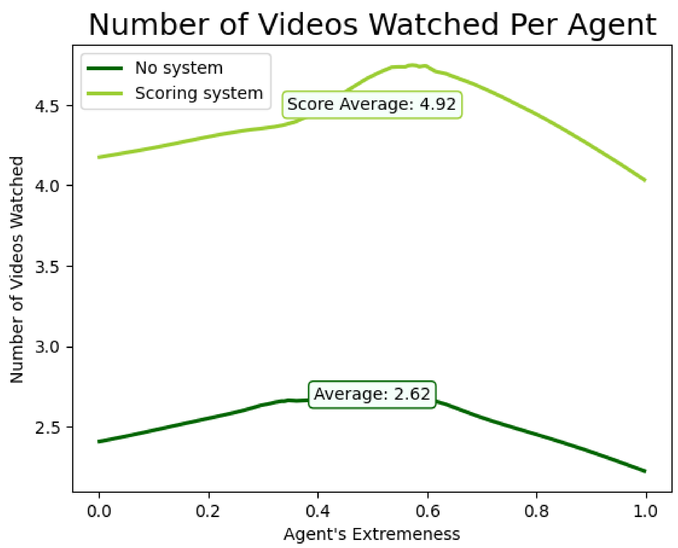

Fig. 2: Graph of the number of videos a user watches, plotted against their own extremeness

The Scoring Equation

Our scoring system uses a weighted-sum equation, described as follows:

$$\text{score} = -(\alpha \cdot \text{l\_return}) - (\beta \cdot \text{p\_return}) + \gamma * \text{abs(ev-ea)} - (\delta \cdot \text{e\_return})$$

Where:
α = weight placed on video length *(0.20)*
β = weight placed on video popularity *(0.15)*
γ = weight placed on video alignment (how similar video extremeness is to the user's) *(0.5)*
δ = weight placed on rewarding extremeness in either direction *(0.15)*

l_return will return 0 if the video length is greater than the user's preference, and 1 if it is less. This rewards shorter videos, as we assume that users do not want to spend lots of time on one video.
p_return will return 0 if the video's number of views is less than the user's preference, and 1 if it is greater. This rewards videos with a larger number of views.
abs(ev-ea) records the difference in magnitude between the user's extremeness (*ea*) and that of the video (*ev*). Videos that minimize this difference are rewarded, as we assume users want to watch videos that align with their ideological beliefs.
e_return will return 0.625 if the video has an extremeness ≤ 0.2 or ≥ 0.8, and 0 otherwise. This rewards videos that are more extreme.

LEHIGH UNIVERSITY | P.C. Rossin College of Engineering and Applied Science

IDEAS INTEGRATED DEGREE IN ENGINEERING, ARTS, AND SCIENCES

DAVID AND LORRAINE FREED UGRS UNDERGRADUATE RESEARCH SYMPOSIUM

INDUSTRIAL AND SYSTEMS ENGINEERING ISE LEHIGH UNIVERSITY