Mitigating the Harm of Recommender Systems
7/7/2019

In the article *Up Next: A Better Recommendation System* [1] the author makes a strong case that recommender systems technology cannot be indifferent to the harm that it can cause, even when such harm is unintentional. She offers specific suggestions for mitigating that harm.

The author's main argument is that systems that frequently recommend extremist content cannot be considered to be operating "hands-off", since they are already deciding what we see and furthermore, "there is no right to amplification". The notion of a neutral ground for the technology is an illusion.

I find this argument persuasive, and her recommendation that "Platforms need to transparently, thoughtfully, and deliberately take ownership of this issue" to be worthy of serious consideration.

In support of the author's viewpoint, consider again the original motivation behind UBCF. As *Social Information Filtering: Algorithms for automating "Word of Mouth"* [2] states:

> User-user collaborative filtering is based on automating the process of "word-of-mouth" recommendations: items are recommended to a user based on values assigned by other people with similar taste. The system determines which users have similar taste via standard formulas for computing statistical correlations.

In a real-world scenario, if an acquaintance offers a “word-of-mouth” recommendation that I then find too extreme or radicalized for my tastes, then my future interaction with that acquaintance is likely to change: I may devalue recommendations from that source, or register disapproval with their viewpoints in a way likely to discourage similar recommendations. If my disapproval does not change the behavior I may take more extreme measures to block the offending content, such as putting the sender on a Spam list, for example. In traditional automated recommendation systems, however, this process of incorporating the high cost of recommending content deemed distasteful by the user is rarely emphasized or fully implemented. One could, therefore, argue that current algorithms capture only part of the word-of-mouth process, and ignore the feedback costs for dubious, distasteful or otherwise objectionable content. For a more complete simulation of the word-of-mouth process, they would need to take into greater account the business costs and risks of fringe and extreme content.

This situation may be changing with more emphasis being placed on the ethics and social costs of recommender systems and research on technology for implementing them. The article describes several current approaches for handling this situation that are promising.

References:

1. Up Next: A Better Recommendation System. https://www.wired.com/story/creating-ethical-recommendation-engines/
2. Shardanand, Upendra, and Pattie Maes. "Social Information Filtering: Algorithms for Automating "Word-of-Mouth". http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.30.6583&rep=rep1&type=pdf