

Amy Qian

11/02/2021

Collaborative Filtering Applications and Challenges

Everyday, people rely on the recommendations of people who are similar to them in order to find new topics and information that may interest them. For example, new books are recommended among book club members, new artists and songs are recommended among people with similar music tastes, and so on. Collaborative filtering is a recommender system approach that models this real life exchange of recommendations between users that are similar and share common interests. When determining if a user will like a particular item, it looks at other users who like that item, and checks if the user is similar. It therefore makes the assumption that users with the same interests have similar preferences. Collaborative filtering has many applications and is used in many different industries and companies to recommend items to users. However, it also faces several challenges such as the cold start problem, gray sheep, and scalability.

I will discuss the application of collaborative filtering in the context of the specific case of Netflix. Recommendation algorithms are essential to Netflix because they provide users with personalized suggestions in order to reduce the amount of time and frustration to find something that they will enjoy watching. Collaborative filtering is especially useful to Netflix because they have a huge database of user behaviors, profiles, and ratings that allow them to build a robust recommendation system based on user similarity. A high level overview of how collaborative filtering is used in Netflix is as follows. First, all users are compared with each other, and if highly similar movie rating or watching behavior is detected between two users, then the users

are deemed similar. If one user watched a particular movie and gave it a high review, but the other has not, then the recommender system will recommend that movie to the other user.

Netflix acknowledges the importance of this recommender system to their business, so they continually strive to improve it and overcome the challenges that collaborative filtering approach faces. I will now discuss three main challenges. The first is data sparsity and the cold start problem. Netflix must evaluate a very large item set that covers thousands of tv shows and movies. This therefore results in a sparse user-item matrix as not every item will have enough users to evaluate it. Additionally, when a new user enters the system, there isn't enough information to evaluate user similarity. This is the cold start problem, or the new user or new item problem.

The next challenge in collaborative filtering is the gray sheep problem. A user is a gray sheep if their opinions don't consistently agree or disagree with any group of people. For example, they may sometimes like what other horror loving users like, and other times not. For these users whose interests are difficult to identify, collaborative filtering is often ineffective since it cannot identify a group of users with which that user belongs. The last challenge is scalability. Netflix has millions of users and thousands of items in their system. The collaborative filtering algorithm has a complexity of $O(n)$, which is large and suffers when the data set is as big as Netflix's. As Netflix's user and item base continues to grow, it will become more difficult to compute user similarity and provide recommendations to each user efficiently.

Despite these challenges and shortcomings to collaborative filtering, it still is a heavily used approach to recommender systems across different companies. In addition, research is being conducted in order to improve the collaborative filtering algorithm and help overcome the challenges previously mentioned.

Sources

Netflix. "Recommendations." *Netflix*, 2021,

<https://research.netflix.com/research-area/recommendations>. Accessed 2 11 2021.

Su, Xiaoyuan, and Taghi M. Khoshgoftaar. "A Survey of Collaborative Filtering Techniques."

Advanced in Artificial Intelligence, vol. 2009, no. 421425, 2009, pp. 1-19. *hindawi.com*,

doi:10.1155/2009/421425.