

Collaborative filtering: Applications & Challenges

CS 410 Fall 2022

Ashna Anil (aanil2)

In this paper, I will be providing an overview of the concept of Collaborative filtering which is one of the most popular and effective methods used for recommendation systems. I will be double clicking on the applications of this in different fields such as Netflix & Amazon and on the common challenges faced when working with this concept such as Cold Start and Grey Sheep.

Collaborative filtering is a concept that has been widely used for a long time. Even before it was formalized, businesses were using this concept of grouping similar users and predicting preferences on a small scale. In recent times, with growing competition in the market, many businesses are looking for something to put them over the edge and collaborative filtering is one such treasure. In simple words, this includes making predictions about a user's interest by compiling preferences from several other similar users. This assumes that there is a relationship between products and users' interests and uses one of the two techniques: "Memory Based Approaches", or "Model based Approaches" to find these similar user preferences. This similarity is measured by simple mathematical functions such as scalar dot product between these user item vectors/matrices.

Common applications include entertainment platforms such as Netflix movie recommendations, Spotify song recommendations and YouTube and online commercial platforms such as Amazon. Even new articles or anything you search for on the internet has collaborative filtering applied to recommend items based on the preferences of other people who are similar to you. Even small-scale businesses such as hardware manufacturer companies and grocery stores are using these concepts to define upsell or cross-sell to increase sales and revenue. A common extended application of this is Market Basket Analysis which gets its name from the concept of someone going to a grocery store and predicting what other products will be added to their basket based on what they already have/their purchase history and other customer history. This can be

used in situations as simple as keeping milk next to eggs in a grocery store to warehouse slotting and sending promotions pamphlets/emails to customers.

Netflix movie recommendations is probably one of the most popular applications of this, about 80% of the watched content is based on these algorithmic recommendations. This basically uses matrices to connect the users to movies and then predict the missing values (latent features) using the existing patterns.

One of the many challenges with this is the fact that if the accuracy of the model must increase, it needs to consider a lot more data which indeed creates very big user item matrices which tend to be sparse. This significantly increases the runtime of the model at the cost of higher accuracy and scalability also becomes a concern. Another challenge which is at the other end of the amount of data available is called “Cold Start”. The whole concept of collaborative filtering relies on having data of similar users to predict the preferences of a certain user and in the beginning of any process/ for a new user, there might not be a lot of data as the users haven’t started ranking the items yet. Other specific challenges include “Synonyms” where the items have similar names but are treated separately in the system. (Word similarity text mining issues) and “Grey Sheep” which occurs when a particular user has special tastes and does not have any strong preferences. Even spam users (shilling attacks) can skew the data and there is also a lack of diversity because popular items will get more ratings than others causing a long tail effect. Another concern is flux, as peoples’ tastes and preferences change over time so we can never fully product accurate collaborative recommendations.

Lastly, another challenge that I see in the concept is privacy concerns. A lot of user data is required to make these predictions and when they involve demographics this starts to draw a thing line for the privacy concerns. It is very important to keep this in mind while using the concept to make predictions.

Overall, this concept is something that we have always been subconsciously performing and scaling this to businesses and other uses has been very helpful. This helps businesses with keeping their uses for a longer time and gives a better user experience by providing broader exposure to product. There are many challenges but as more

people deep dive into this method and solve these challenges, this technology/concept will have a huge impact in future years!

References:

Chhabra, S. (2017, August 21). *Netflix says 80 percent of watched content is based on algorithmic recommendations*. MobileSyrup. Retrieved November 5, 2022, from <https://mobilesyrup.com/2017/08/22/80-percent-netflix-shows-discovered-recommendation/>

YouTube. (2017). *YouTube*. Retrieved November 5, 2022, from <https://www.youtube.com/watch?v=wTZUHiLUY94&t=7s>.

Rangaiah, M. (n.d.). *Using data handling and digital marketing to maximise customer experience: A Netflix case study: Analytics steps*. Using Data Handling and Digital Marketing to maximise customer experience: A Netflix Case Study | Analytics Steps. Retrieved November 5, 2022, from <https://www.analyticssteps.com/blogs/using-data-handling-and-digital-marketing-maximise-customer-experience-netflix-case-study>