Data 607 - DISCUSSION/ASSIGNMENT 11: RECOMMENDER SYSTEMS

Enid Roman

2022-11-06

HOW TO PERFORM A SCENARIO ANALYSIS

Amazon's US retail e-commerce share is expected to increase from 44% last year to half of total e-commerce sales in 2018. Amazon owns the richest world dataset on how consumers consume and how sellers sell. This allows the retail giant to continuously optimize its online shopping experience through data signals like purchases, searches, and reviews in an ongoing data feedback loop, creating a classic network effect.

WHO ARE AMAZON'S TARGET USERS?

Amazon's target users are the new and potential customers and the loyal customers.

In order to help Amazon accomplish their goals we first need to look at Amazon's recommender and their potential problem. While doing my research I found this interesting article, https://www.linkedin.com/pulse/amazons-recommendation-engine-secret-sauce-mario-gavira/

AMAZON'S RECOMMENDER

Amazon´s recommendation algorithm allows to create a personalized shopping experience and increase the amount of revenue generated from each customer. According to a McKinsey report, 35% of all Amazon's transactions come from algorithmic product recommendations.

Amazon's have developed their own recommendation engine capable of handling tens of millions of customers and products in near real time.

Amazon´s recommendation system is based on several data signals collected throughout the shopping experience. What a user has bought in the past, which products they place in their shopping cart, items they've rated and liked, and what other customers have viewed and purchased.

Amazon combined the best of both worlds to create its own algorithm called "item-to-item collaborative filtering", collaborative filtering and content based filtering. Collaborative filtering is looking at the user-product interaction by finding customers with similar transaction history and recommend the top products bought by that similar buyer to the shopper under study. Content based filtering is looking at the product and not the customer, by simply recommending the top items most similar to the product viewed by the user.

HOW THE LIST RECOMMENDATION IS BUILT

1. It begins by looking at the items that are associated with the user and built the recommendation table by computing how similar it is to other items in the collection.

2. To determine how relevant the recommended items are, the algorithm looks at customer ratings for each product and filters out items that have already been bought by the user.

3. Most of the computation is done offline. Once the recommendation table is built it is injected into the engine.

4. This allows to display recommendations almost in real time.

Amazon uses neural networks for their engine. To be capable to compute hundreds of millions of customers and products in real time, they created a so called DSSTNE, "Deep Scalable Sparse Tensor Neural Engine"

ADVANTAGES

This system trains neural networks and powers the different personalized experiences for millions of customer journeys.

The advantages of Amazon's item-based collaborative recommendation algorithm are:

1. The recommendations are highly relevant

2. They are computed in real time

3. The algorithm scales to hundreds of millions of users and tens of millions of items without sampling or other techniques that reduce the quality of the recommendations

4. It updates immediately on new information about shopper's interests.

5. This feedback loop allows to constantly improve and tweak the algorithmic models.

THE DISADVANTAGE

This system is far from being perfect and Amazon only started what AI can offer in terms of recommendation intelligence.

Modelling time correctly in the recommendation algorithm is both an art and a science.

Amazon.com's catalog is continually changing. Thousands of new items arrive and disappear daily, especially in categories such as seasonal clothing fashions and consumer electronics.

The cold-start problem means that new arrivals can be at a disadvantage because they don't have enough data yet to have a strong correlation with other products.

The recommendation engine also faces the cold-start problem for new visitors with no information about their interests and behavior.

Even for loyal customers the algorithm needs to factor in critical timing elements:

1. Older purchases becoming less relevant to is current interests. The speed of the decline is different between products indicating a durable long-term interest, ex. bike helmet, and items that fulfill a short-term need, ex. light bulbs.

2. Some purchases will trigger a change in recommendations over a longer period, ex. from baby diapers to child games.

3. For daily use products (FMCG´s) such as toiletries or packaged food, recommendations can be scheduled in regular time periods based on purchasing patterns.

4. Time-limited external events can massively influence buyers' behavior and need to be factored into the recommendation engine.

Amazon must have in its back office a unique ID for each product from each seller, but in many cases one sellers will offer nearly identical products to other sellers with different ID's. Therefore, the engine is unable to match the recommended item with the previously purchased one.

This is a problem that is likely to increase over time with the number of worldwide Amazon seller and overlapping catalogues exponentially growing.

External events do influence consumer behavior, such as fashion trends, marketing campaigns or even economic or political changes can influence buyers behavior and shopping habits.

Trying to factor in all these elements by adding external data signals into the recommendation algorithm seems to be a potentially biased approach: how do we prioritize external data sources and where do we set the limits?

Certain recuring events where a clear correlation is identified, and a robust data source is available might justify being integrated into the algorithm. An example would be the local weather conditions and their impact in the apparel industry sales.

WHAT IS AMAZON'S KEY GOALS?

Amazon's vision is that recommendation engines will move beyond the current paradigm of searching, clicking, and buying and will become like talking to a friend who knows you, your interests, what happens around you and anticipates your needs.

HOW CAN YOU HELP THEM ACCOMPLISH THOSE GOALS?

Even though we don't have control of recuring events, the most effective systemic solution to this problem will be continuously refreshing the weight of items in the algorithm, based on the aggregated navigation, and buying behavior of other users - without the need of understanding what external factors trigger changes in the purchase behaviors.

For the thousands of new items that arrive and disappear daily, especially in categories such as seasonal clothing fashions and consumer electronics requires an explore/exploit process to give items an opportunity to be shown.

It is critical to collect data for these first-time users on referral sites, what ads they are attracted to, what categories they browse, what items are added to the shopping cart and which ones are abandoned. Computing this browsing behavior on the immediately to generate relevant recommendations to convert first-time visitors to customers.

To solve the problem of the engine unable to match the recommended item with the previously purchased one. Amazon must combine some image recognition programs with text mining algorithms, and it should be straightforward to match different product ID's when images and text descriptions overlap and filter the item out of the recommendation table in case a similar one was already bought.

CONCLUSTION

A recommendation engine boils down to several pipelines (or filter pattern implementations) that allow for a context to be evaluated by several modules applying certain business rules.

Amazon´s leading edge in machine learning technics, computing power and massive wealth of consumer data puts them in a unique position to keep tweaking and optimizing its recommendation algorithms moving forward.