DATA SCIENCE AT NETFLIX

KRISHNAVENI RAJAN

201902960

Table of Content

INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Use case of Data Science is very vast and it has been helping out the world to get better day on day in various field like Health Care, Education, Entertainment, Retail & Banking.

This Case study will be focused on the use and importance of Data Science in the Netflix. Netflix was started as a DVD rental service in 1998. It mostly relied on a third party postal and delivery services to deliver its DVDs to the customers. Due to dependency on the delivery and various other factors, Netflix faced heavy losses. To overcome the loss and win, Netflix soon mitigated with the introduction of their online streaming service in 2007.

When the online streaming service was provided; customers had a lot of expectations like the quality of the audio and video. Customer also expect vast variety and huge amount of movies and programs to choose from. Netflix invested in a lot of algorithms to provide a flawless movie experience to its customers.

To make Netflix more successful and impressive, they had to came up with the recommendation system. The recommendation system was expected to provide suggestions to the viewers of the various cinematographic products. Netflix used the Data Science to come up with the best Recommendation systems. Let us look into it in detail in the following sections how Data Science had helped the Netflix to build the Recommendation system and achieve great success.

RECOMMENDATION SYSTEM

A recommendation system is a platform that provides the users with various contents based on their preferences and likings. A recommendation system takes the information about the user as an input. This information can be in the form of the past usage of product or the ratings that were provided for the product by various users. These information are then processed to predict, how much the user would rate or prefer the product.

Other important role that a recommendation system plays today is to search for similarity and for comparison between different products. A recommendation system makes use of a variety of machine learning algorithms.

There are different types of Recommendation Systems namely

- Content-based recommendation systems
- Collaborative filtering recommendation systems
- Hybrid Recommendation System.
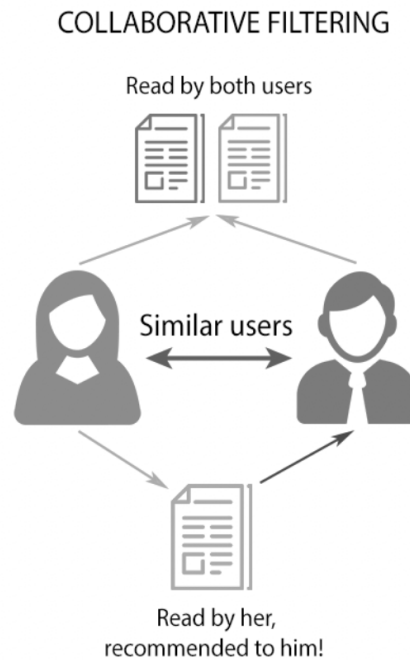
TYPES OF RECOMMENDATION SYSTEM

1. Content-based Recommendation Systems

CONTENT-BASED FILTERING
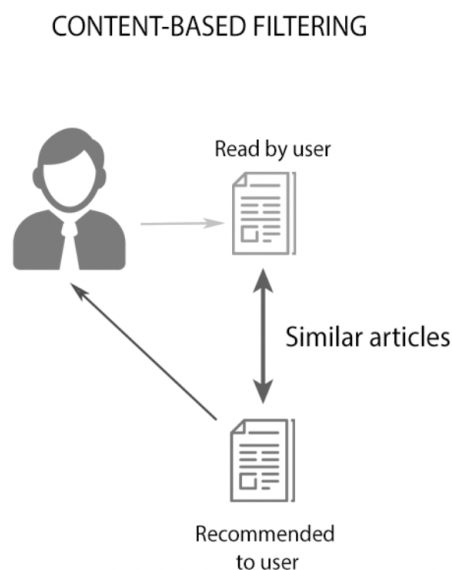
In a content-based recommendation system, the background knowledge of the products and customer information are taken into consideration. Based on the content that you have viewed on Netflix, it provides you with similar suggestions.

For example, if you have watched a film that has a sci-fi genre, the content-based recommendation system will provide you with suggestions for similar films that have the same genre.

2. Collaborative Filtering Recommendation systems

COLLABORATIVE FILTERING

Read by both users

Similar users

Read by her,
recommended to him!

Unlike the content based filtering that provided recommendations of similar products, Collaborative Filtering provides recommendations based on the similar profiles of its users. One key advantage of collaborative filtering is that it is independent of the product knowledge. Rather, it relies on the users with a basic assumption that what the users liked in the past will also like in the future.

For example, if a person A watches crime, sci-fi and thriller genres and B watches sci-fi, thriller and action genres then A will also like action and B will like crime genre.

3. Hybrid Recommendation Systems

There is also a third type of recommendation system that combines both Content and Collaborative techniques. This form of recommendation system is known as Hybrid Recommendation System.

WHY NETFLIX NEEDS RECOMMENDATION SYSTEM?

In 2006, Netflix wanted to tap into the streaming market, it started off with a competition for movie rating prediction. It provided a prize of $ 1 million to

whoever increased the accuracy of their then existing platform 'Cinematch' by 10%.

At the end of competition, the BellKor team presented their solution that increased the accuracy of prediction by 10.06%. With over 200 work hours and an ensemble of 107 algorithms provided them with this result. Their final model gave an **Root Mean Square Error (RMSE)** of 0.8712. For their solution, they made use of K-nearest neighbour algorithm for post-processing of the data.

Then they implemented a factorization model which is popularly known as **Singular Value Decomposition (SVD)** for providing an optimal dimensional embedding to its users. They also made use of **Restricted Boltzmann Machines (RBM)** for enhancing the capability of the collaborative filtering model.

These two algorithms in the ensemble, *SVD and RBM* provided them with the best results. A linear combination of these two algorithms reduced the *RMSE to 0.88.*

However, even after reduction of RMSE and increase in accuracy, Netflix suffered from two major challenges –Firstly, the data that provided during the competition comprised of 100 million movie ratings, as opposed to more than 5 billion ratings that Netflix constituted of. Furthermore, the algorithms were static, meaning that they only dealt with historical data and did not take into account the dynamicity of users adding reviews in real-time.

Netflix to overcame these challenges, it made the winning algorithms a part of its recommendation system.

BUILDING RECOMMENDATION SYSTEM USING DATA SCIENCE

Netflix makes the primary of use Hybrid Recommendation System for suggesting content to its users. In the case of Netflix, the recommendation system searches for movies that are similar to the ones you have watched or have liked previously. This is an important method for scenarios that involve cold start. In cold start, the company does not have much of the user data available to generate

recommendations. Therefore, based on the movies that are watched, Netflix provides recommendations of the films that share a degree of similarity.

Using Interleaving to Improve Personalization

Netflix uses Ranking Algorithms to provide a ranked list of movies and TV Shows that appeal the most to its users. However, with the presence of various ranking algorithms, it is often difficult to accommodate all of them and test their performance simultaneously.

While the traditional A/B testing on a reduced set of algorithms could not identify the best algorithms with smaller sample size and also consumed a lot of time, Netflix decided to innovate its algorithmic process.

In order to speed up its experimentation process of its ranking algorithms, Netflix implemented the interleaving technique that allowed it to identify best algorithms. This technique is applied in two stages to provide the best page ranking algorithm to provide personalized recommendations to its users.

In the first stage, experimentations to determine the member preference between the two ranking algorithms is carried out. Unlike the A/B testing where the two groups of viewers are exposed to the two ranking algorithms, Netflix makes use of interleaving to blend the rankings of algorithm A and B.

Netflix provides its users with enriched content based on this interleaving technique that is highly sensitive towards ranking the algorithm quality.

Importance of Context Awareness in Recommendations

Contextual Awareness is one of the key elements in personalizing recommendations for its users. This not only improves the performance of the recommendation system but also prompts users to provide better feedback that would result in a quality recommendation. There are two categories of contextual classes:

Explicit

- Location

- Language
- Time of the Day
- Device

Inferred

- Binging Patterns
- Companion

In order to predict contexts, we make use of representation learning. It is a ***Deep Learning Technique*** that performs feature engineering that discovers features without explicit programming. Based on the time and periods of watching, Netflix bases its data on various parameters like Day, Week, Season and even longer periods like Olympics, FIFA, and elections.

For performing contextual predictions, Netflix treats recommendations as a sequence classification problem. It takes the input as a sequence of user-actions and performs predictions that output the next set of actions.
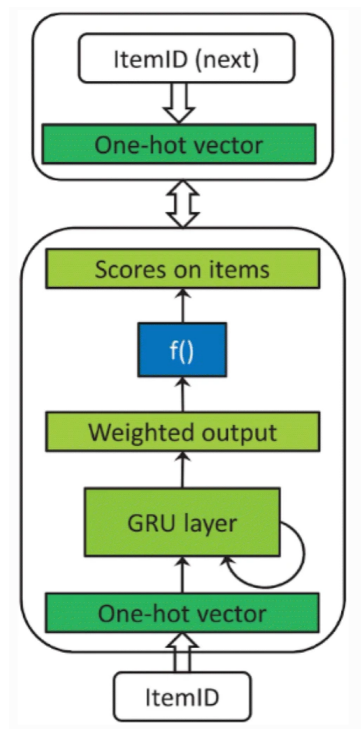
An example of a sequence problem is **Gru4Rec**. And in the case of contextual sequence prediction, the input consists of the contextual user actions as well as the current context of the user.

This helps the recommendation engine solve the question: *"Based on all the historical actions that are taken by the user what is the most probable video that they will play right now?"*

So, this is how Netflix is using Data Science for providing recommendations.

SUMMARY

We studied how Netflix makes use of a recommendation system to provide movie suggestions to its users in this case study. We also studied how Netflix is heavily relying on various techniques in Data Science for providing recommendations to the user.

We went through the **Netflix Prize competition** and how it used the algorithms of the winning team to improve its accuracy. In the end, we discussed contextual prediction and how Netflix leverages it to provide personalized recommendations to its users.