Application of Machine Learning on Netflix and it's impact on user acquisition

Mr. Chayan Bhattacharjee
Department of Information Technology
Chikitsak Samuha's Patkar Varde College
Mumbai, India

Abstract — Netflix is a popular online streaming service that allows users to watch movies, TV shows, anime, and dramas. Data science is a critical component of Netflix's success because it enables the company to benefit from big data. Netflix has a variety of data sources, including data from different places and users, as well as data about multiple aspects of its content and users. Personalization, suggestions, enhancing user experience and other functionality are some of the most common examples of how big data is being used by Netflix.

Keywords—Netflix, big data, recommendations system, personalization

I. INTRODUCTION

Netflix is online streaming service and default location for viewers who want to watch movies, shows, anime, dramas, etc. It is very much evident that it has the highest share of online users who use it to binge-watch content. Netflix was DVD rental platform before entering into the business of Over-The-Top (OTT) media provider [1]. Netflix has humongous amount content on its platform to watch along with other data about the users and related metadata about content which it uses for better user experience. Netflix does not provide content free of cost, it uses subscription plans for the content provision and does not show any kind of advertisement to its users. It also provides its content on all the different platforms like computer, smartphone, tablet, iPad, TV, etc. while giving proper enjoyable user experience.

Data science and machine learning are very important aspect of Netflix's success as they help Netflix to harness this big data to achieve profit by staying ahead of other competitors like Hulu, Amazon Prime Video, Disney+ Hotstar, etc. Data that Netflix has is of multiple types, coming from multiple locations and users and about multiple aspects related to its content as well as its users. This is nothing but big data that is in possession of Netflix which is used for multiple different concepts like data analysis, statistics, machine learning and deep learning.

In this paper, different useful features that Netflix has and why they incorporate to Netflix's success are discussed with respect to machine learning and big data that it uses.

II. PERSONALIZATION POWERED BY DATA

Netflix has more than millions of users who use Netflix on daily basis in some way which actually contributes to the data that it gathers about them. Data that Netflix collects about the users include following [1]: Age, gender, location, specific flavour of media that they like, day and time when content is watched, nature of the content, searches on the platform, content that gets re-watched, ratings given by the users. Netflix uses this data in their algorithms and mechanisms which incorporates this data in a form which can generate vital insights that aid in the proper direction of the business.

Netflix's platform employs a recommendation system, which assists 75 percent of its users in watching content on the platform [2]. Given the number of people that utilize the site, this is a pretty amazing figure. Data regarding the user's location is typically crucial in presenting consumers a portion of the huge content collection that Netflix possesses. The content displayed to the user varies substantially depending on the location from which the user accesses the site. For example, Erased, an anime is available in Argentina but not in India. This location wise personalization provides better exposure to content from the same location first than other content. After this first form of filtration, recommendations provided by the Netflix's algorithm are also highly personalized. Users are provided basic recommendations based on the content they view, the ratings given to the content, and the amount of people watching the content. Based on data analysis, several types of recommendations are made, and these suggestions can be user-user or item-item, but the most commonly used pattern is matrix factorization, which is also utilized by Netflix and YouTube for their recommendation engines. Using these recommendations, many parts of the UI are customized, such as the top-10 section of Netflix, which is particular to the location from where the user is viewing the Netflix. Based on the genre of content that user watches these recommendations are also altered to suit the taste of the individual users. For example, if user watches anime and Korean drama more frequently than any other content then for these two categories separate recommendations are created which is shown in figure 1. The easier way to access recommendations is to look in 'more like this' section of the Netflix app.

Another form personalization can be seen in custom artworks generated for content that is shown to users. Netflix uses tool called Aesthetic Visual Analysis (AVA) to select the artwork and image that are shown to user when the browse the media library provided by Netflix. AVA [1]. Users of Netflix are divided into multiple taste group so this personalization needs to consider these groups while finalizing the artwork. Along with these groups, actor facial expressions, placement of the actors on the screen

while showing the artwork, scene lighting, etc. are also taken into consideration until final picture is selected.

With the help of huge amount of data that Netflix has it uses different algorithms to analyze the data to find interesting content that can be shown to users which helps Netflix to gain profit as users will continue to use the platform as long as they have good content to watch with the power of personalization with the help of big data Netflix enhances user experience which helps them to return their customers.

Figure 1: Content specific personalization

III. CONTINUOUS TESTING

Netflix is known for its continuous testing where it rolls out specific feature update for specific userbase and observes the response that it receives to generate useful insights. Netflix perform A/B testing on batch of users along with monitoring of other users who are not part of the test. A/B testing done by Netflix may include more than 1 test as long as they are not conflicting with each other. After rolling out these test specific features the observe the useful metrics such as user retention and hours of streaming. Such metrics are helpful to draw conclusion about the features which might be retained. A basic example of this kind of testing is custom artwork generated for specific movie. Based on the number users who have watched the movie and given positive rating it can be further used more frequently so that a greater number of users will use the platform and in general more related content will be discovered by users. In case of artwork, it is really important for Netflix to customize the experience as it leads to more watch time on the platform which essentially helps Netflix to draw profit from the same. Example under this context is shown in figure 2. The example given shows two different artworks for the same movie Good Will Hunting. This movie is combination of genres consisting of Drama and Romance. As shown in the figure the first set of images contain romantic movies which might be watched by users who like movies in this genre so to those members if the artwork of the Good Will Hunting is shown with Matt Damon and Minnie Driver, it will be more appealing and users might be interested in watching the movie. In the second set of images, movies having genre of drama and comedy are shown at the left side of the arrow which depicts users who generally watch titles having these genres. To them if movie artwork contains actor like Robin Williams, it will be more appealing to watch. This example essentially shows simple but effective analysis of the artwork images to draw more users who will be willing to watch a particular title. It helps Netflix to make the movie, show or in general the content more appealing to everyone who likes any one genre or theme that the content features.

A good case study of this testing and experimenting behavior of Netflix is discussed in [2] with the example of show House of Cards. This case study discusses about how Netflix applied knowledge big data to make House of Cards successful and blockbuster hit. Netflix has big data in the form of different metrics such as viewer behavior, binge watching behavior, ratings given by users, movies that were liked depending on the actor, etc. In case of House of Cards, Netflix had definite data about significant number of people out of its subscribers who have like the film, "The Social Network", directed by David Fincher. Subscribers have streamed the movie from start to finish. This was one part of the analysis. Another part was actor Kevin Spacey. Movies featuring Kevin Spacey were well-received by its audience.

Figure 2: Artwork customization based on data collection [3]

Third part of this analysis was that the British version of the show House of Cards was also well-received in the States and to add more to this people who have watched British version had also watched movies directed by David Fincher as well as acted by Kevin Spacey. From this data it was not difficult for Netflix to conclude that American version of House of Cards featuring Kevin Spacey and directed by David Fincher will be a big hit. This is true because it was in reality successful show along with other

shows like Breaking Bad and Game of Thrones. Other Netflix original shows like Stranger Things, Orange Is the New Black and The Crown are also part of the same process of utilizing the big data.

IV. RECOMMENDER SYSTEM

As mentioned, Netflix has seen that its recommendations are favourable to 3/4th of its total number of users who actually follow the suggestion given by Netflix's recommender algorithm. Netflix's recommender system is essentially a data product which takes data as input and produces data as output. Multiple people involved in Netflix's monitoring playground act as important pillars of giving this data as input to the recommender system. For the recommendation system being used to draw profit it is essential to show the appropriate recommendation in certain category or genre in which users are interested so that users will continue to use the platform in future. Recommendation system of Netflix consists of different parts which are as follows [4]:

A. Personalized Video Ranking (PVR)

This is the most general algorithm that Netflix uses based on the content user has viewed in the past. Data used for this algorithm uses multiple different categories to provide recommendation such as past viewing history of the user, genre that user is heavily watching, popularity of that genre and content that belongs to this genre along with implicit parameters like ratings. For example, if user watches anime and Korean drama more frequently than other type of content then specific rows with titles of those

genres will be generated for user will always be ranked on top on the welcome screen of Netflix. This is shown in figure 1.

B. Trending now

This row uses algorithm to show content which generally follow recent trend in watching pattern of users while taking into consideration the time frame of the trend. In a nutshell this simply means that titles which have received more attention from users because of some third party phenomenon such as coronavirus outbreak, popularity of anime, sitcom, etc. This follows changing behavior of users and targets users so that they might follow the same trend but it generally is not accurate and more supporting to the users' taste. These recommendations are temporary and may change after a week, day or even an hour. This section is sometimes named as "Popular on Netflix" as shown in figure 3.

C. Top-N video ranker

This ranker uses simple ranking of content or titles which are being watched in particular region. This is also similar to trending now section but rather than following trend it simply ranks the shows being watched in ascending order and mostly shown in the count of ten. An example of such ranker is shown in figure 4 where top 10 title in India are shown. This is also temporary ranking and changes daily as the data about watching pattern of users' changes.
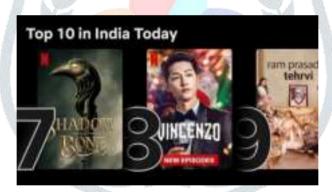

Figure 3: Top 10 video ranker *D. Video-video similarity*

Figure 4: Trending now section of Netflix

Figure 5: Video-video similarity-based recommendation

This is another form of recommendation produced by Netflix which closely follows the data about similarity between videos. This uses object-item similarity matrix where algorithm predicts related videos based on the video that user and other users have watched. This section is shown as "Because you watched" section and multiple rows are shown on homepage with different titles that user has watched in the past. There is less personalization in this category of recommendation but it still follows the behavior of multiple users who have watched the content that user has watched and then suggests other content which might be interesting to the user. This section is also sometimes named as "Because you liked" and shows the recommended content based on not just watching behavior but also the ratings given to the specific title. This section has also less accuracy as it does not take into consideration the total watch time of the content. It shows recommendations even if the user has watched only 1/4th of the content.

E. Top picks

This section is very similar to PVR recommendations but takes into consideration all the genres and category of content that user is interested. These are not ranked at the top unlike PVR and are not accurate. Still, the suggestions are somewhat personalized based on user's taste.

F. Page generation

Homepage shown to each user is very personalized depending on the user's behavior. This page generation takes into account all the different algorithms mentioned above. Netflix has been using A/B testing for page generation to improve the user experience and thus uses multiple rows of different forms of recommendations to keep user on the platform. Algorithm used for page

generation takes into consideration the data about user's interaction with these rows and then orders these rows to generate page for user. This algorithm utilizes millions of videos or video corpus that Netflix has and then filters out the content to show only the most relevant recommendations in each row of the page.

V. CONCLUSION

Netflix is successful company and it is well-known fact that Netflix uses big data on large scale to make profit and gain large number of users every year. Harnessing big data to provide different features like personalization, recommendations, improving user interface, streaming, etc. are most popular examples. Netflix uses the data it collects to not only present content to users more efficiently but also uses the same data to produce the quality content to gain more profit. It is pretty evident that big data is the essential part behind Netflix's success.

REFERENCES

[1] Costa, C. D. (2020, April 19). *How Data Science is Boosting Netflix - Towards Data Science*. Medium. https://towardsdatascience.com/how-data-science-is-boosting-netflix 785a1cba7e45

[2] Bikker, Y. (2020, July 19). *How Netflix Uses Big Data to Build Mountains of Money | The Startup*. Medium. https://medium.com/swlh/how-netflix-uses-big-data-to-build mountains-of-money-829364caefa7

[3] Blog, N. T. (2018a, June 21). *Artwork Personalization at Netflix - Netflix TechBlog*. Medium. https://netflixtechblog.com/artwork personalization-c589f074ad76

[4] D. Chong, "Deep Dive into Netflix's Recommender System," Medium, 02-May-2020. [Online]. Available: https://towardsdatascience.com/deep-dive-into-netflixsrecommender system-341806ae3b48/.