Behavior-based Popularity Ranking on Amazon Video

With the growth in the number of video streaming services, providers have to strive hard to make relevant content available and keep customers engaged. A good experience would help customers discover new and popular videos to stream with ease. Customer streaming behavior tends to be a strong indicator of whether they found a video engaging. Aggregate customer behavior serves as a useful predictor of popularity. We discuss the use of past streaming behavior to learn patterns and predict a video's popularity using tree ensembles.

CCS Concepts: • **Information systems** → **Recommender systems**; **Content ranking**.

Additional Key Words and Phrases: video streaming; popularity; recommendations

ACM Reference Format:
. 2020. Behavior-based Popularity Ranking on Amazon Video. In *Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3383313.3411555

1 INTRODUCTION

On streaming platforms customers use two main approaches to find videos—search and discovery. Customers can enter a query to search for a video or browse the catalog to discover content. Streaming platforms use customer preferences, genre of videos (or other criteria) to curate personalized rows (or carousels) of videos to recommend to customers. According to Gomez-Uribe et al. the mode of discovery accounts for 80% of hours streamed on Netflix [4].

We refer to models serving the discovery mode as browse models. The purpose of browse models is to introduce new and trending videos to customers. In the discovery mode customers do not express a specific intent in the form of a search query and are likely to be exploring the catalog. When a popular movie like 'Joker' is available to stream, customers should be able to find it in an effortless manner. Popularity can be determined by the number of clicks, ratings or streams that videos earn. While heuristics may be effective on small datasets, they often do not scale and worsen the cold-start problem, where new videos may not have sufficient clicks or ratings to be ranked alongside others [8].

Traditional recommendation systems have used collaborative filtering (matrix factorization) or techniques such as Markov Decision Process (MDP) to improve the quality of results [5]. In the video streaming domain, the Netflix Prize competition paved the way for improving movie rating predictions. The Netflix competition showed the effectiveness of ensembles in improving model predictions [2]. Linden et al. discuss the use of an item-to-item collaborative filtering approach to recommend products to customers on Amazon [7]. They discuss the limitations of the user-item based collaborative filtering approach, which may be intractable and subject to the cold-start issue. Guy et al. use MDPs on an Israeli online book store to make recommendations to customers as they added books to their cart [9].

Personalization tends to be a big part of product and video recommendations. Video recommendations can be personalized by introducing customer preferences for specific genres of videos and past streaming behavior. While personalization is important it poses several challenges. When working with a large catalog containing various types of

Manuscript submitted to ACM

video content (e.g. movies, TV shows, channel subscriptions, free or paid content), scale and performance restrict the degree of personalization. It is further restricted when little is known about the customer. Furthermore, there may also be legal limitations to personalizing results. In this work we focus on improving the ranking of videos that are popular across a wide spectrum of customers on the platform.

The goal of browse models should be to delight customers by helping them serendipitously discover products. While click-through rates are useful, they may be noisy and subject to spam. Individual customer streaming behavior may not be a strong indicator of a video's popularity, but aggregate behaviors from over millions of customers may be indicative of a video's popularity.

Predicting popularity of videos based on observable signals such as comments on videos, review ratings and exchanges on social media platforms has been widely studied [6, 10]. In this work we learn to predict popularity of titles using prior engagement and video metadata signals. Amatriain discussed the effectiveness of using customer streaming behavioral features in improving the ranking of recommendation algorithms [1]. However, they do not discuss how different types of videos are treated. Covington et al. use deep and wide neural models to train recommendation rankers [3]. They pose the problem of recommendations as a supervised classification problem. Latency is a critical metric for models that are served online. Deep neural networks require large computational capabilities and may increase serving time beyond acceptable thresholds.

In this work we use customer streaming behavior as a proxy for a video's popularity, since popular videos tend to get streamed more than unpopular ones. We developed a popularity ranking model that uses aggregate customer streaming behavior to predict engagement. We developed models to predict the popularity of a video. Models were trained with three kinds of feature inputs: (1) aggregate customer streaming and purchase signals, (2) video metadata and (3) video newness. These three classes of features manage to capture the key aspects of videos, allowing learners to identify the importance of each of these features. Aggregate customer signals include the number of times a video was streamed, percentage of the video streamed (minutes watched divided by the total runtime), purchases etc. Video metadata includes information such as runtime of a video, price etc. and video newness includes date features such as the days since the video was available for streaming on the platform. We use a tree ensemble model to develop our solution and tune different targets to cater to the differences between customers' engagement with different types of content. Existing work discuss the use of content-based recommendations but do not factor the differences between video types in their ranking. We tested our models on millions of customer browsing sessions. To the best of our knowledge ours is the first work addressing the effectiveness of developing popularity models catering to different video types.

PRESENTER INFORMATION

Lakshmi Ramachandran is an Applied Scientist at Amazon Search. Lakshmi uses machine learning including deep neural models to solve problems pertaining to search and ranking. She previously worked as a Sr. Research Scientist at Pearson, where she used natural language processing techniques to improve essay and short answer scoring models. Lakshmi has a a PhD in Computer Science from North Carolina State University.

REFERENCES

[1] Xavier Amatriain. 2013. Big & Personal: Data and Models behind Netflix Recommendations. In *Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '13)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/2501221.2501222

[2] Robert M. Bell and Yehuda Koren. 2007. Lessons from the Netflix Prize Challenge. *SIGKDD Explor. Newsl.* 9, 2 (Dec. 2007), 75–79. https://doi.org/10.1145/1345448.1345465

[3] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 191–198. https://doi.org/10.1145/2959100.2959190

[4] Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.* 6, 4, Article Article 13 (Dec. 2016), 19 pages. https://doi.org/10.1145/2843948

[5] Yehuda Koren, Robert Bell, and Chris Volinsky. 2009. Matrix factorization techniques for recommender systems. *Computer* 42, 8 (2009), 30–37.

[6] J. G. Lee, S. Moon, and K. Salamatian. 2010. An Approach to Model and Predict the Popularity of Online Contents with Explanatory Factors. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1. 623–630. https://doi.org/10.1109/WI-IAT.2010.209

[7] G. Linden, B. Smith, and J. York. 2003. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7, 1 (Jan 2003), 76–80. https://doi.org/10.1109/MIC.2003.1167344

[8] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock. 2002. Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*. Association for Computing Machinery, New York, NY, USA, 253–260. https://doi.org/10.1145/564376.564421

[9] Guy Shani, Ronen I. Brafman, and David Heckerman. 2002. An MDP-based Recommender System. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI'02)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 453–460. http://dl.acm.org/citation.cfm?id=2073876.2073930

[10] Zhi Wang, Lifeng Sun, Chuan Wu, and Shiqiang Yang. 2012. Guiding internet-scale video service deployment using microblog-based prediction. In *2012 Proceedings IEEE INFOCOM*. 2901–2905. https://doi.org/10.1109/INFCOM.2012.6195726