## Tech Review

### Ranking Recommendation Systems: Challenges and Solutions

Ranking recommendation systems are extremely important, especially in recent years given the sharp rise in online video services that are used worldwide. Businesses that offer streaming services such as Netflix and Amazon have grown their online video offerings in the past decade. As online content continues to grow in popularity, businesses need to optimize their online services by adapting to make the user experience in navigating content effortless and as appealing as possible. One way of adapting is by implementing effective ranking recommendation systems in order to keep users engaged and make it easy for those users to quickly obtain the content that they enjoy. However, in order to make recommendation systems effective, businesses will need to overcome certain challenges. Challenges for ranking recommendation systems include multiple competing ranking objectives, biases, and scalability [1]. Fortunately, researchers have identified ways to address these challenges as best as possible by experimenting with an online video ranking recommendation system.

What is the challenge that ranking objectives introduce into a recommendation system? Before diving into this, let's first outline what ranking objectives are. Ranking objectives can be considered as criteria that a learning model uses to predict an outcome of user behavior; and the prediction is part of the algorithm that calculates a score for recommendations and rankings [1]. For example, a system could have objectives for ranking video recommendations that are based on the watch time, number of shares, or ratings; and a model may choose certain video candidates to rank based on one or more of those metrics. Now, when there are multiple objectives as mentioned in the earlier example, it is a challenge for recommendation systems to know what to prioritize and weigh the most in the prediction [1]. If it prioritizes one metric over another then the model performance might become hindered. For example, if an objective is weighted more heavily on watch time, then the system's model may observe a decrease in data and learning based on click rate [2]. To address the challenge, researchers in one study by Google Inc. first simplified the multitude of objectives by grouping them into two broad categories when experimenting with video recommendation: engagement objectives, which involve user clicks and watches, and satisfaction objectives, which involve liking or rating a video [1]. Moreover, the researchers used a model called Multi-gate Mixture-of-Experts, a multi-task learning approach that "explicitly learns to model task relationships from data" [3]. By using this approach, the model can learn how to weigh different objectives in its prediction. As a result of grouping the objectives and using the Multi-gate Mixture-of-Experts model, researchers found the approach to improve engagement and satisfaction metrics and address the challenge of multiple competing objectives [1].

The other main challenge in ranking systems is selection bias, particularly position bias [1]. For example, if a system provides the top ten videos for a user query, a user may simply choose the top one, thinking the system has chosen the best match; when in reality the video is not what the user is looking for. In this example, the ranking system will log that the user has clicked on the video, which will indicate a positive relevance judgement; and that feedback will be favored when used in future prediction calculations. As a result, the model will continue learning in an erroneous feedback loop [1]. To address this challenge, researchers from Google Inc. proposed to adopt a Wide & Deep framework in their experiment to help control or mitigate bias in video recommendation ranking; more specifically, they proposed to add a “shallow tower” to the main Multi-gate Mixture-of-Experts model [1]. The approach from the research study appeared to mitigate bias by having a part of the model focus on gathering features of bias, and this part worked together with the main model to continue to learn the selection bias and take it into account when computing the final score. In other words, researchers were able to combine user-utility and bias in their learning model when computing a score for the rank. In the end, the study showed that the researchers observed reduction in position bias as a result of their proposed framework, thus finding that this framework helps prevent the overall model from leaning too much on position for ranking [1].

Lastly, scalability is another challenge in designing an effective ranking recommendation system. Scalability is critical for large platforms, particularly for video platforms such as Netflix and Youtube since they serve large amounts of content to billions of customers around the world. To address scalability, researchers found that industries can typically use infrastructure improvements and machine learning models [1]. Furthermore, the study performed by Google Inc. suggests that simplifying the architecture is best to ensure process efficiency. In the study, researchers adopted a simple structure where it had two stages: candidate generation and ranking -- and during the candidate generation, their system obtained less than a thousand candidates from a much larger body of content. Moreover, as part of the experiment in the study, researchers used the Multi-gate Mixture-of-Experts model which was found to be efficient and effective for large-scale systems [3]. By keeping the results limited, simplifying the architecture, and using an optimal learning model, they were able to achieve an efficient and scalable ranking recommendation system for Youtube, a large-scale online platform. With that in mind, it is clear that complicated models won’t scale up with efficiency [1]. Therefore, a good approach to overcome scalability issues that impact the business and the use experience is to ensure the recommendation systems avoid overly complex architecture.

In conclusion, video ranking recommendation systems have challenges which need to be addressed for success. There are challenges in supporting multiple objectives that compete with each other during learning and prediction, and there are challenges of bias where systems may fall into a trap of relying on relevance data that is not truly representative. All the while, there are concerns of managing at scale, which is a critical component that businesses must keep in mind to keep their services running efficiently to optimize costs and optimize the user experience. Fortunately, researchers from Google Inc. conducted a successful experiment that clearly demonstrated that a ranking system that uses the Wide & Deep model with Multi-gate Mixture-of-Experts for multitask learning addresses the challenge of multiple competing objectives and mitigates selection bias. At the same time, the overall model maintains an adequate level of efficiency and scale without compromising quality.

## Works cited

**[1] Recommending What Video to Watch Next: A Multitask Ranking System.** *Zhe Zhao, Lichan Hong, Li Wei, Jilin Chen, Aniruddh Nath, Shawn Andrews, Aditee Kumthekar, Maheswaran Sathiamoorthy, Xinyang Yi, Ed Chi*.

**[2] Online Article Ranking as a Constrained, Dynamic, Multi-Objective Optimization Problem** *Jeya Balaji Balasubramanian, Akshay Soni, Yashar Mehdad, Nikolay Laptev*

**[3] Modeling Task Relationships in Multi-task Learning with Multi-gate Mixture-of-Experts** *Jiaqi Ma1, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, Ed H. Chi*