

# Recommending What Video to Watch Next: A Multitask Ranking System Tech Reviews

ZongYun,Li

November 4, 2022

## Intro

This paper is about using a large scale multi-objective ranking system for recommending what video to watch next on Youtube. It uses two kind of features. One is enengagement behaviors, such as clicks and watches. The other is satisfaction behaviors, such as likes and dismissals. The way they produce recommending list is using candidate genration to generate a list, then use ranking system to rank the list. It uses Multi-gate Mixture-of-Experts (MMoE) as its model architecture, and introduces a shallow tower to reduce selection bias.

## Body

It had multiple problems solved in the paper. The two things I learn most are multimodal feature space and reducing selection bias. For multimodal feature space, the model need to learn user utility of candidate videos with feature space generated from multiple modalities. This leads two problems. First is bridging the semantic gap from low-level content features for content fltering. Second is learning from sparse distribution of items for collaborative fltering. To solve multimodal feature

space, they extract features such as video meta-data and video content signals as its multimedia feature. For context, they use features such as user demographics, device, time, and location. Selection bias is that the user may click the item that isn't that useful just because the item is selected by recommendation system. The most famous selected bias is position bias. They remove selection bias by adding a shallow tower to learn selection bias. Its experiment shows that by combining three methods of reducing selection bias: input feature, adversarial loss and shallow tower, shallow tower is the most efficient. I think that there may be a new way to reduce position bias. By setting new position to the old position, it will be an effective way to reduce position bias. For instance, if old position is [1,2,3,4,5,6,7,8,9], then new position will be [1,1,1,2,2,2,3,3,3]. By grouping three positions to same position number, it will reduce position bias, because the distance between each position is shortened and old recommendation is precise that the first three recommendations may be really useful for user. The benefit of this method is the model will be more lighter than using shallow tower, although it may not have better effect than shallow tower. I think let model lighten is an important issue, because the model should output recommendation list timely in this kind of real-time problem..

## Conclusion

How to reduce selection bias and solve multimodal feature space is really important to the ranking system for recommending what video to watch next on Youtube. Efficiency is important for this kind of real-time problem. I think that the way I propose to reduce selection bias can make recommendation system faster. For future, I will experiment my idea to let this

model more faster. Then, it will be a light-weight and efficient model.

#### Reference

Recommending What Video to Watch Next: A Multitask Ranking System(<https://daiwk.github.io/assets/youtube-multitask.pdf>)