Video Recommendation Ranking System Review

Mony Chhen
monyrc2@illinois.edu

ABSTRACT

This paper is a tech review of a technology that is used for videos recommendation for users. This is a topic that is very sensitive to many videos systems such as YouTube. The recommendation ranking system is expected to perform flawlessly to keep the users engage to the videos content that they are watching, and properly suggests the next set of videos. This might seem like a simple task from users' perspective, however, the system faces many real-world challenges, such as multiple completing ranking and implicit selection biases in the user feedback loop. This paper will go through details of the system implementation to address these challenges as well as my thoughts on whether I agree or disagree with the implementation.

Keywords

Video, Recommendation, Ranking, Ranking System, Multitask, Video Recommendation, Video Ranking.

1. INTRODUCTION

The problem that the authors are tackling is how to give a user that is currently watching, a recommended next video that the user might watch and enjoy. To design and develop this system in a real-world large scale is very difficult. Some challenges include conflicting objective, and often implicit bias in the system. For example, a user might have clicked and watched a video because it is ranked at the top, instead of the video being relevant to the user. To address this issue the authors, propose an efficient multitask neutral network ranking system by using Multi-gate Mixture-of-Expert (MMoE) and a shallow tower model to remove selection bias. I strongly agree with how the authors are thinking about how the video can be misleading by having high ranking or sharing among friends. Over time this could be biasing the ranking system toward the top-ranking videos. In additional, they evaluate their proposed ranking system by conducting offline live experiment to confirm the effectiveness of multi-task learning and removing common type of selection bias. After their evaluation, their ranking system shows significant improvements. The remaining of this paper will go through the details understanding related work, explaining the problem the authors are trying to solve, as well as sharing my thoughts and opinions on improvements to the existing system that the authors have described.

2. UNDERSTANDING RELATED WORK

The authors did a great job breaking down the different area of relate work items before tackling the problem. I believe understanding the related work such as cost, objective, and model bias prior to tackling the problem is the correct approach. In this recommendation system, they split the related work into three categories: industrial case studies recommendation system, multi-objective learning for recommendation system, and understanding biases in the training data. The following subsections will describe each of the categories in details.

2.1 Industrial Recommendation Systems

To design and develop a successfully ranking system, a large quantities of training data are needed. Due to cost, the ranking system utilize implicit feedback such as clicks and engagement from the users on the recommended items. The main problem with this approach is that the system is that there is a misalignment between user implicit feedback and true user utility on recommended item. The authors' approach uses a different technique, by introducing a deep neutral network-cased ranking model, which use multi-task learning techniques to support the ranking system. This approach allows the feedback to quickly train the algorithm and output most relevant videos back to the user.

2.2 Multi-objective Learning for Recommendation Systems

The most challenging part of the whole system is multi-objective learning. The idea behind this is to constantly update the weight of the ranking for the users base on different type of user behaviors or engagements while they are browsing such as click and likes. The system then uses these engagement and behaviors to estimate and compute a final score for ranking.

2.3 Understanding and modeling Biases in Training Data

Other think to keep in mind is the bias in the training data. Users' logs are used as the training data, which captures the user behavior and responses to recommendation from currently suggested video. The interaction between the users and the current system can generate bias in the feedback base. For example, a user may click on the item because it is rank higher than other. It does not necessarily mean that the video is relevant to the user and that the user will enjoy watching it. This issue will cause a feedback loop effect. I agree with his scenarios as this is something that I do as a user during video browsing. I often find that the top video might not be the video I want to watch and end up using the search bar to look for a new video to watch.

3. PROBLEM DESCRIPTION

I found that the most difficult part of this ranking system is to make the system work in a real-world large scale. The first problem is how to recommend a video to watch next that is relevant to the users without having a bias system. To do this, the author introduce two stage setup of candidate generation and ranking. On top of this challenging, how to build a ranking system with implicit feedback for a real-world large scale video recommendation problem. The authors mention two factors: Multimodal feature space and scalability.

3.1 Candidate generation

Candidate generation is interesting concept and makes sense to me from a user perspective. The idea behind this is to generate candidates by matching topics of query videos, as well as user video browsing history. In addition, a technique is used to generate context-aware high relevant candidates. Then follow by scored by the ranking system. This idea will work if the users allow history caches. How about for users who browse with private session? How would you handle auto generating relevant videos to watch next during the private start to end?

3.2 Ranking System

The ranking system generate a few hundred candidates. This is different from candidate generation system, the aim here is to provide a rank list so that the highest rank item will show at the top for the users. To achieve this, the most advance machine learning techniques using neutral network architecture in the ranking system.

The ranking system learns from two types of user feedback 1) engagement behaviors, such as clicks or watches; 2) satisfaction behavior, such as likes. By using this additional information feedback, the ranking system significantly improves in the query results for videos that are relevant to the users. I agree that this is a good approach to improve the ranking system and limiting bias in the results display on the users' end. However, could we apply the same concept toward negative satisfaction? For example, dislike, or bad comment on the video content. How do we trade the ranking algorithm to take this into account? Another scenario I can think of is when users watch part of the video and leave to another video, this should fall under engagement behavior and the ranking system should be aware of such action and adjust the ranking algorithm parameters accordingly.

4. EVALUATION

After coming with a great idea and design, you need to proof the concept and apply to real world scenario and ensure that your system is efficient and improving compared to existing system. In this section, I will cover how the authors evaluate their system and ensure that the system is working per their expectations.

The authors use YouTube user implicit feedback to train the ranking model and conduct both offline and live experiments. The experiment measure two main metrics: engagement and satisfaction. Both areas show significant improvement. Additionally, they evaluate modeling and reducing position bias by first analyze user implicit feedback by verifying that position bias exists in the train data. Then compared to the baseline model, followed by live experiment to ensure that the bias is reduced. Lastly, learn the position bias and apply to the algorithm.

5. CONCLUSION

In this paper, I've discussed the importance of understanding the problem that you are tackling and the training data that are being used can largely impact the result of the system you are designing. The main reason why this recommendation system has significantly improve compared to existing ranking system is because the authors thought carefully through the different scenarios that a user would perform using video browsing and consider the impact of the behavior and engagement with the system to ensure that it reflects into the ranking system. In addition, the authors did a great job to evaluate their system and ensuring that bias is reduce from the ranking system as well as introducing multi-task learning into the recommendation system. I believe to improve this system further, negative behavior or engagement towards a video should also be captures to adjust the ranking score in the opposite direction. For example, if a user dislike a video or provide negative comment toward a video, the ranking should adjust the score accordingly.

6. REFERENCES

[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 - Proceedings Of the 13th Acm Conference on Recommender Systems – 2019.