Ge Yu
CS410
Technology Review
11/14/2020

Comparison Between Existing Ranking System and A Multitask Ranking System for Video Recommendation

Introduction

In this technology review paper, the algorithm and development for recommending videos to users used by Youtube platform will be introduced and discussed. Current recommendation system used by Youtube be compared with a new system based on multitask ranking system. There have been some existing challenges for current video related websites such as insufficient metadata, a small amount of interaction time(Most videos are under 10 minutes) and short life cycle of videos requiring recommendation's freshness. With the challenges and the background, the recommendation systems are required to focus on the freshness and limitation period, to guarantee the diversity and relevance upon user's action. To illustrate this, the current system uses the algorithms as the following steps: Generating the user's personal activities as seeds and then ranking the system using different signals for relevance and diversity. The problem of the current system is that there are multiple competing ranking objectives, and user feedback indicates that there are implicit selections. To solve this problem, an optimization for multiple ranking objectives improving the current ranking algorithms are introduced. Additionally, to mitigate the biased selection, a new framework is also introduced.

Comparison

Before actually comparing the differences, the main idea of building a recommendation system will be introduced. Generally, the recommendation systems follow the two-stage design strategy, with the candidate generation stage and ranking stage. The system first takes the input data from the user, which could be noisy, then a mapping will be constructed from the video to a set of related videos. The mapping step will generate a directed graph, that each video pair has a weighted dedge. Next step is to generate the recommendation candidates, the related association rules will be combined with the user's personal activity, and generate a union as the seed set for further calculation. After the candidates are generated. The set candidate will be scored and ranked based on a variety of signals. The ranking is categorized into three stages including video quality, user specificity and diversification. The video quality includes the signals about the video's rating, commenting and favoriting. The user specificity signal works to focus on the specific user's preferences, and will be used to boost videos that match the unique interests. To generate a ranked list, a linear combination of the signals will be used. To avoid bias, videos that are balanced between relevancy and diversity across categories will be chosen. A simple constraint on the single seed video's recommendation number will be put to protect the diversity of recommended videos. With this straightforward recommendation system without sophisticated optimization, the results indicate that the co-visitation based recommendation performs at 207% of the baseline Most Viewed page method(Davidson et al. 2010), showing that the three-stage recommendation system works as proposed.

Despite that the results of the architecture mentioned above optimize the baseline method, there are still obstacles and challenges for a large-scale video platform like Youtube. Thus, an improved method using efficient multi task neural network architecture was introduced to enhance the ranking system. The basic idea is to extend the Wide & Deep model architecture by adopting Multi-gate Mixture-of-Experts(MMoE)(Zhao et al. 2019). Since there are multiple ranking objectives, the enhanced method divides the multiple objectives into two categories including engagement categories: a category that collects engagement activities such as user clicking, and a category that collects user's satisfaction such as users' likes about the video. Then the MMoE architecture generates layers that focus different aspects of input. With the architecture, each objective is able to decide experts to share or not with each other. Also, to reduce the selection bias that would affect the ranking results, a shallow tower will be added. It takes the input related to the selection bias, and will return a scalar serving to the prediction for the final main model.

For the ranking system, there are more features to take into consideration. Multimodal feature space is generated from the multiple modalities to provide the user utility information. The major issue for this feature space is that there are semantic gaps from low-level content features, so bridging would be important for content filtering, and the scalability due to the real-time and real-world requirements due to the properties of the video platforms.

The first stage for the ranking system is generating candidates. Since there are multiple aspects of similarities, the each generation algorithm will capture one aspect of those. After the generating step, the candidates will be pooled into a set, then score by the ranking system subsequently. The ranking system then generates a ranked list from the candidates. Ranking system will provide a ranked list that the items with highest utility will appear at the top of the list. To keep the efficiency, neural network architecture will be introduced in this step.

The enhanced ranking system is modeled as a combination of classification and regression problems. Unlike the basic method introduced using par-wise approaches to make predictions on a pair of candidates before, the approach to make prediction is a pointwise approach. The reason is that despite the pair-wise approaches potentially improving the diversity of the recommendations, point-wise approaches are more efficient to scale a large number of candidates.

The exact ranking system user's different types of behaviors as training labels. Like mentioned before, the objectives are categorized into engagement objectives and satisfaction objectives. Since engagement objectives include behaviors like clicks, and the time spent on the video, there are two types of tasks to finish: binary classification and regression. After computing the multiple ranking objectives, a multitask ranking model will be trained to do the predictions. The specific step is to take the input of the predictions, and then generate a score using a combination function, with a weighted multiplication. Note that weights will be adjusted manually to improve the performance of the model.

In order to mitigate the conflicts of multiple objectives, MMoE model architecture will be introduced. The main idea of MMoE is to multitask the learning process with the experts shared across all tasks. The MMoE layer is able to capture the differences between the tasks with substituting the shared layer and adding a separate network for each task. This way of adding a Mixture-of-Experts layer is helpful to learn the information from the input as they are modularized before. There is also a problem that applying the layer will increase the cost for both training and serving, due to the high dimensionality of the input layer.

In this way of multitask ranking using MMoE, the model is trained more efficiently and comes with a more accurate way to provide the ranking results. The results of the experiment shows that multitask ranking with MMoE has increased significantly compared to baseline model with Shared-Bottom method with about double the performance. Although there are still obstacles, with MMoE, recommendation systems could work more efficiently and precisely.

Conclusion

Traditional recommendation systems have been using the ranking system with a linear combination. However, with the existing problems for the input collected by the candidate generation systems face the problem of the presence of multiple competing ranking objectives. To solve this problem, a new technique to optimize the multiple ranking objectives using Multi-gate Mixture-of-Experts has been introduced to solve the problem by mitigating the conflicts of multiple objectives, which has been proven as a new efficient way to finish the ranking stage. Instead of regarding the ranking stage as a linear combination problem, the new model deals with the problem by modeling it into classification and regression steps. The new techniques have made apparent improvements on the engagement and satisfaction metrics, leading to higher accuracy and efficiency.

Reference

[Zhao et al. 2019]Zhao, Zhe & Chi, Ed & Hong, Lichan & Wei, Li & Chen, Jilin & Nath, Aniruddh & Andrews, Shawn & Kumthekar, Aditee & Sathiamoorthy, Maheswaran & Yi, Xinyang. (2019). Recommending what video to watch next: a multitask ranking system. 43-51. 10.1145/3298689.3346997.

[Davidson et al. 2010]Davidson, James & Liebald, Benjamin & Liu, Junning & Nandy, Palash & Vleet, Taylor & Gargi, Ullas & Gupta, Sujoy & He, Yu & Lambert, Michel & Livingston, Blake & Sampath, Dasarathi. (2010). The YouTube video recommendation system. 293-296. 10.1145/1864708.1864770.