Student: Sahil Agrawal

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Google's Multitask Ranking System: Video Recommendation on a Large Scale

November 6, 2022

---

1 Introduction

Nowadays, with the advent and huge progress' in the internet, there are a lot of people (millions) who watch videos online, with YouTube being one of the most famous platforms. All video platforms are trying to provide new and innovative features to enhance the user experience. One of these features is recommendations which suggests other videos to watch based on the video currently being watched and other factors like the users preference, background, time of day, etc. This is a very important feature these days and newer and much better methods are coming up to provide better and faster recommendations. Google has just come up with a new method for exactly such a use case.[1]

2 Model

There already exist a lot of recommendation systems, but the need of the hour is to create one that can handle tons of data (large scale) very fast as it has to be real time recommendations. Along with this, there are other challenges like multiple objectives which might conflict with one another. An example of this is when you want to recommend videos that the user might want to watch and will additionally want to like it, or comment on it and share with their circle. Lastly, there is also an implicit bias in the recommendation systems as they train on the results of their previous suggestions which introduces a bias as training data is only available on the videos recommended, and not on the videos not recommended. Additionally, users might watch a recommended video even if they don't like it, which leads to bad feedback, which in turn trains the model on bad data. All these challenges make video recommendation a hard task, but the model suggested overcomes these issues. The model focuses on ranking relevant videos which is the second stage of the process; it doesn't go into detail on the first stage which is collection of relevant videos.

The three main challenges faced by ranking systems and how the authors overcome them are:

2.1 Scalability:

The authors use point-wise ranking which implies making predictions for each candidate based on only itself. This makes the model efficient enough to provide real time recommendations which is a necessity. The other option is to use pair-wise approaches, which make predictions on 2 or more candidates. These are useful to increase the different type of videos included in the topmost ranks, but they are slower and can't provide real time or live suggestions nor can it scale well for huge data-sets and number of users, which is why it wasn't used by the authors.

2.2 Multiple Objectives:

The authors divide the objectives into 2 broad categories: engagement objectives (watching videos, clicking on videos, etc.) and satisfaction objectives (liking a video or marking it as bad, etc.). Both of these are solved as either a classification task (click a video or not, like a video or not) or a regression problem (time spent on a video, rating provided to a video). They use a combination function to combine the outputs of these multiple objectives and use this combined score in future steps.
To accomplish this, they propose an extension of the Wide and Deep model by using Multi-gate Mixture of Experts (MMoE). The idea is to replace a shared layer with an MoE layer. This allows the creation of multiple experts, each of which learns a particular feature from the input. Then there is a gating network that is trained for each task which chooses which of the experts are needed for the particular task. In this way, experts are shared across all tasks which gives a soft sharing of parameters, which is better suited for this use case. The old shared-bottom model had parameters shared in a hard way, which is harmful if the correlation between tasks is low, which is the case here. Thus, the MMoE layer is a better choice here.

2.3 Position Bias:

As it is not efficient and possible to explicitly get feedback from users on the recommended videos, feedback needs to rely on implicit methods like clicks, reviews and time spent on a video. But this results in a bias as feedback is present only for videos which were recommended and also the feedback may not be correct which results in a vicious cycle giving wrong training. So, to solve this issue, the authors propose a shallow tower in addition to the main tower proposed above. The features fed into this shallow tower are such that it can model different types of biases (like position for position bias). The output of this shallow tower is combined with the output of the main tower to give the final result which is free of bias or at least reduced bias. To prevent the model from putting too much weight on this bias removal, a drop out is used during the training stage and it is made zero in the serving stage.

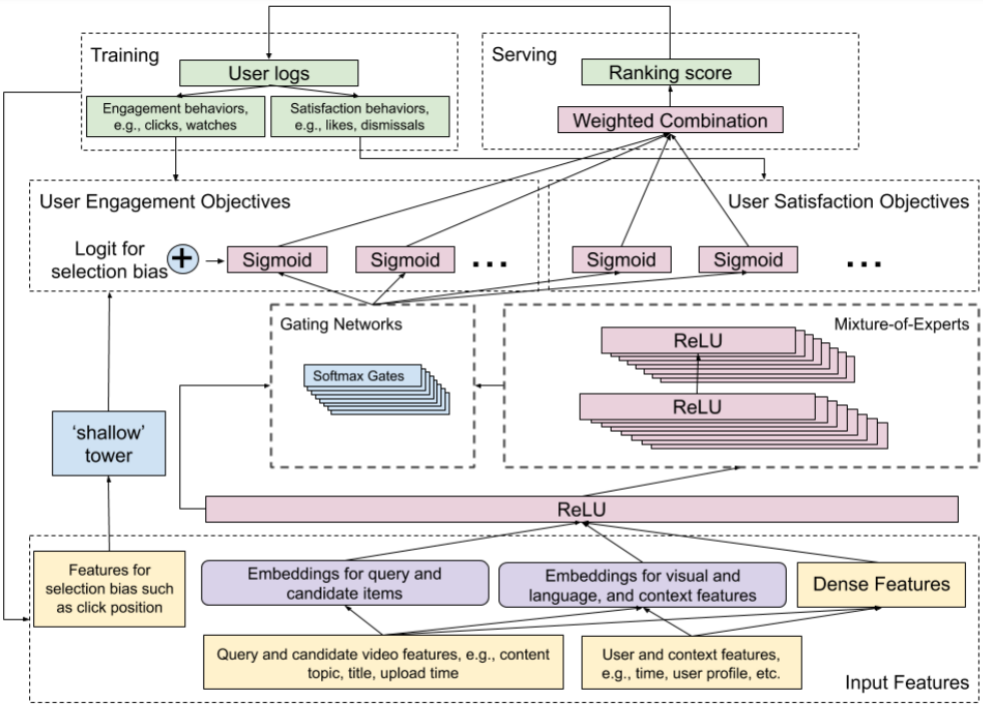

Figure 1: Model architecture of our proposed ranking system. It consumes user logs as training data, builds Multi-gate Mixture-of-Experts layers to predict two categories of user behaviors, i.e., engagement and satisfaction. It corrects ranking selection bias with a side-tower. On top, multiple predictions are combined into a final ranking score.

In this way, the authors cater to all the challenges and have proposed a model that can recommend videos much better than previous state-of-the-art methods.

3 Experiments

The authors conducted experiments on their model using YouTube and compared it with a baseline model (shared-bottom model). The experiments were both static (classification and regression) and live experiments. As can be seen from Table 1 below, the model proposed outperforms the baseline on both the engagement and satisfaction metric when compared with same model size (number of multiplications). Additionally, from Table 2, we see that the shallow tower improves the engagement metric compared to other methods by removing bias the best.

Model Architecture	Number of Multiplications	Engagement Metric	Satisfaction Metric
Shared-Bottom	3.7M	/	/
Shared-Bottom	6.1M	+0.1%	+ 1.89%
MMoE (4 experts)	3.7M	+0.20%	+ 1.22%
MMoE (8 Experts)	6.1M	+0.45%	+ 3.07%

Table 1: YouTube live experiment results for MMoE.

Method	Engagement Metric
Input Feature	-0.07%
Adversarial Loss	+0.01%
Shallow Tower	+0.24%

Table 2: YouTube live experiment results for modeling position bias.

4 Conclusion

The authors have proposed a model that is similar to Wide and Deep architecture which has a wide network (memorisation) and a deep network (generalisation). In the proposed model, the shallow tower is the wide component and the MMoE layer is the deep component.[2] Even though the proposed model is better than state-of-the-art models, there is always scope for improvement. Another better model instead of MMoE could have been used to cater to multiple objectives by making it more efficient for this use case. Models that automatically identify and cater for biases can be developed. More compress models which are faster can be created. But aside from all these potential improvements, the authors have created a model that outperforms all current models and is thus a huge step forward in the field of recommendations.

References

[1] Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., ... Chi, E. (2019, September). Recommending what video to watch next: a multitask ranking system. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 43-51).

[2] Suneet Bhatia (2022, February). A Multitask Ranking System: How YouTube recommends the Next Videos