CS410 Fall2022 Technology Review
Completed by Yueming Pang (NetID: ypang10)

Introduction:

This review studies and compares two techniques for industrial ranking system which are developed by Google, Inc. and Alibaba, Inc. respectively. Before introducing the Google's framework, the Wide&Deep framework is first explained because it is extended by Google's multitasks ranking system.

Typically, recommendation system contains two stages, which are candidate generation and candidate ranking. These two papers both focus on the second stage where candidates are ranked following certain algorithms or model. In Google's multitask ranking system, a variety of soft-parameter sharing techniques (the **Multi-gate Mixture-of-Experts** model) along with a **Wide & Deep framework** are utilized to improve the recommendation quality on YouTube. Alibaba, on the other hand, focuses on improving the post-click conversion rate (CVR) model in ranking system, where they proposed that the conventional CVR model had a sample selection bias problem and a data sparsity problem. To solve these two problems, Alibaba developed the **Entire Space Multi-Task Model** (ESMM) that utilizes sequential pattern of user actions. Alibaba's ESMM was examined to outcompete the current CVR model in Taobao's recommender system.

Body:

The Wide & Deep framework [3]:

There are two key concepts to introduce before the Wide & Deep framework:

1) Memorization: learning the frequent co-occurrence of items or features and exploiting the correlation available in the historical data.
2) Generalization: based on transitivity of correlation, exploring new feature combinations that have never or rarely occurred in the past.

Wide & Deep learning jointly trained wide linear models and deep neural networks to combine the benefits of memorization and generalization for recommender systems. This framework provides a solution to achieving both memorization and generalization in a ranking system. The Google's ranking system is developed based on this framework.

Google's multitask ranking system [1]:

There are mainly two challenges in developing and designing real-world video recommendation system: 1) different and conflicting objectives need to be optimized, and 2) implicit bias in the system. To address such challenges, Google proposed a multitask neural network architecture for ranking system.

This architecture extends the Wide & Deep framework by adopting Multi-gate Mixture-of-Experts (MMoE) for multitask learning. It also introduces a shallow tower to model and remove selection bias. The system first categorizes the multiple objectives into two group: engagement objectives (such as user click) and satisfaction objectives (such as user liking). The MMoE is used to learn parameters across these objectives including the conflicting ones. Then the MMoE further modularizes input layer into experts that each focuses on different aspects of input. Each objective can choose experts (can share or not share with other) by using multiple gating networks. In the shallow tower model, training data is labeled into two parts: the unbiased user utility (from the main framework) and the estimated propensity score (from this shallow tower model).

In most current industrial ranking systems, the misalignment between user implicit feedback and true user utility on recommended items exists. Whereas Google's deep neural network-based ranking system supports multiple ranking objectives by utilizing multitasks learning techniques. This raking system is able to learn from multiple types of user behaviors and combines them to generate a final utility ranking score. This ranking system also addresses the challenges in scalability and multimodal feature spaces. To address scalability, relatively small group of candidates are retrieved from a huge corpus during candidate generation stage and then they are ranked in the final list. Additionally, this system also uses multiple candidate generation algorithms to capture different aspect of similarity between query and candidate. In the end, all candidates are pooled into a set and then they are ranked. For multimodal feature space, features such as meta-data and content signals are extracted and used as its representation. User demographics, device time, and location are used for context.

How the system works in details:

The Google's multitask ranking system learns from types of user feedback: engagement behaviors and satisfaction behaviors. For each candidate, the system uses its features, query and context as input, and learns to predict multiple user behaviors. Such behaviors are later on used as training labels. Each objective supported by the system is to predict one type of user behavior related to user utility. The prediction of these behaviors is then formulated into two types: binary classification and regression. Once the objectives are decided, the ranking model is trained, where each candidate takes input of these predictions, and output the combined score using manually-tuned weighted multiplication. As for the MMoE, the ReLu layer is substituted with the MoE layer, and a separate gating network is added for each task. The shallow tower models and removes position and selection biases. In training, the model does not over rely on the position feature. In serving time, the position feature is treated as missing.

The experiment results demonstrated that the system could lead to substantial improvements on recommendation quality on YouTube platform.

Alibaba's ESMM [2]:

Estimating post-click conversion rate accurately is critical for ranking system in industry. As also mentioned in the Google's multitask ranking system, the user click is a kind of engagement behaviors for training the models.

To begin with, some abbreviations need to be introduced:

SSB: sample selection bias, which will hurt the generalization performance of trained models

DS: data sparsity, which makes CVR model fitting rather difficult

CVR: post-click conversion (rate)

CRT: click-through (rate)

CTCVR: post-view click-through & conversion (rate)

Given recommended items, users may click interested ones and further buy some of them. This can be described as a sequential pattern of user action: impression -> click -> conversion. Therefore, CVR modeling is the task of estimating this conversion rate.

To eliminate the SSB and DS problems, the Entire Space Multi-Task Model (ESMM) was proposed by Alibaba. In this model, two auxiliary tasks of predicting the CTR ad CTCVR were introduced. The CVR model was not trained directly by samples of clicked impressions. Instead, ESMM formulates CVR as an intermediate variable, which calculated by diving the CTCVR with CTR! Both CTCVR and CTR were estimated using the entire space with samples of all impressions. As a result, CVR was also applicable throughout the entire space, which indicates the elimination of SSB. Additionally, the CTR was also trained with much richer samples and this led to the mitigation of the DS problem tremendously.

Conclusion:

The Google's multitask ranking system draw a big picture of how the data was obtained and model was trained. On the other hand, the Alibaba's entire space multi-task model focuses on estimating the post-view click. What is more, Google's model is tested on the video sharing platform YouTube, while Alibaba's model is tested on the e-commerce website Taobao. Therefore, further studies are needed to find the shared attributes between these two domains. For example, can we define the term: conversion rate of a video? Can we consider online purchasing the same as watching videos? We need further abstraction across different domains. In this case, I will propose the term **accept** to cover the meaning of conversion on e-commerce and user click on YouTube. Given that, the ESMM can be integrated into learning the engagement behaviors of the multitask ranking system. Maybe modifications are needed for integration, but it can further increase the reliability of the system.

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] Ma, X., Zhao, L., Huang, G., Wang, Z., Hu, Z., Zhu, X., & Gai, K. (2018, June). Entire space multi-task model: An effective approach for estimating post-click conversion rate. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 1137-1140).

[3] Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... & Shah, H. (2016, September). Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems* (pp. 7-10).