End-to-end ML Pipelines in Big Retail: Showcases of Recommendation & Search Systems at TOPS Online

Witchapong Daroontham, Tittasin Sukcharoen, Chonlapat Patanajirasit, Benyapa Matupumanon
Central Technology Organization

Abstract: The basic application of ML in E-commerce platforms is a product discovery system including a recommendation system and a search engine. However, the implementation of the two systems could be done by various techniques and is open for discovery for any E-commerce platform. Furthermore, even though the development in ML is on the rise, the focus on the ML operation and pipelines is still in an immature phase.

We aim to demonstrate how we implement the recommendation system which is a collection of several recommendation models and the search ranking model at TOPs Online, the online platform of a well-known groceries chain in Thailand. We would also address the current state of our ML pipelines which include the planning, implementation, and maintenance of our ML projects.

Introduction:

The application of ML in Recommendation & Search Systems for E-commerce is never a straightforward task as the problems could be formulated in several different ways. The variation of techniques includes collaborative filtering, content-based, or hybrid method. On the other hand, creating an end-to-end ML Pipeline starting from defining a business problem to model delivery, evaluation, and maintenance is still not really a well-defined process. Apart from ML leader companies like Google, Facebook, Amazon, Uber, or Airbnb, not many of the detailed end-to-end ML pipelines are published.

Recommendation System & Search Ranking Model:

Based on our ML Pipelines delivery for TOPs Online, we would demonstrate how we efficiently train a pairwise model for personalised recommendation which outperforms vanilla collaborative filtering model by 1.9x using @20 recommendation accuracy. We also utilize the learnt representation across several applications including our *also-bought-together* product recommendation, basket recommendation, and search ranking model. For the search ranking, our ranking model outperformed the baseline ranking using CTR/CR by 10.9% using @20 click-NDCG.

End-to-End ML Pipeline:

Additionally, we would also explain how we built an end-to-end ML pipeline from defining a business problem up until A/B testing a ML model on production. We use Airflow to trigger weekly model training and daily inference in batch operation manner. We use AWS for cloud services. Training and inference codes are containerised using Docker Image registered on ECR. The Docker Image is used in the compute, store, and destroy process on either ECS or EC2 instances.

References

Steffen Rendle Et al., BPR: Bayesian Personalized Ranking from Implicit Feedback, UAI, 2009
Christopher J.C. Burges,From RankNet to LambdaRank to LambdaMART: An Overview, Microsoft Research Technical Report MSR-TR-2010-82