Some Practice for Improving the Search Results of E-commerce

Fanyou Wu[1], Yang Liu[2], Rado Gazo[1], Bedrich Benes[1], Xiaobo Qu[2]
[1] Purdue University [2] Tsinghua University

Abstract

In the Amazon KDD Cup 2022, we aim to apply natural language processing methods to improve the quality of search results that can significantly enhance user experience and engagement with search engines for e-commerce. We discuss our practical solution for this competition, ranking 6th in task one, 2nd in task two, and 2nd in task 3. The code is available at https://github.com/wufanyou/KDD-Cup-2022-Amazon.

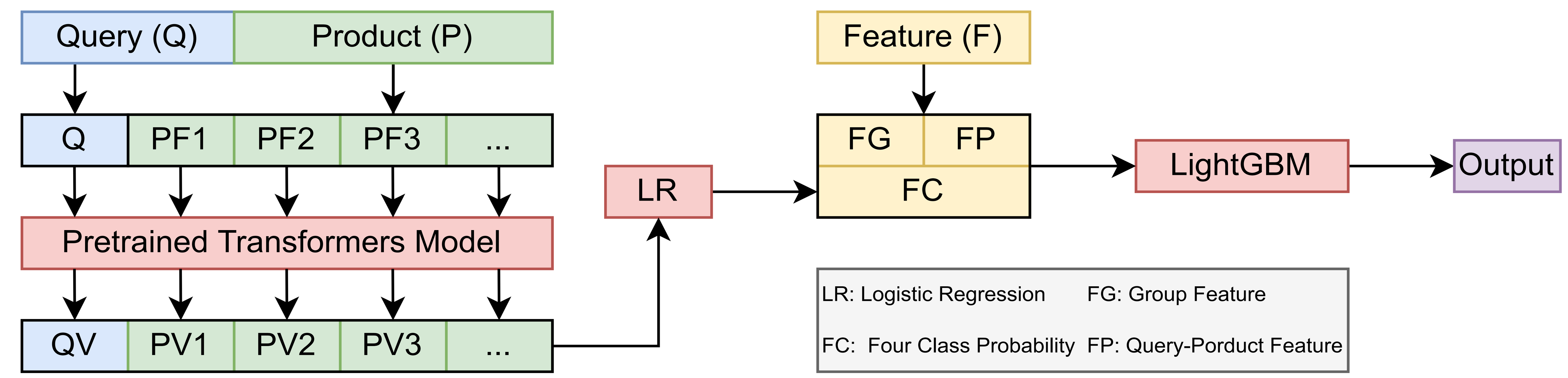

Figure 1. Overall solution

Problem Description

The organizer provides a dataset called the Shopping Queries Dataset. It is a large-scale, manually annotated dataset composed of challenging customer queries. The data is multilingual and includes English, Japanese, and Spanish queries. It comprises query-result pairs with annotated four classes of relevance (ESCI labels):

- **Exact (E):** the item is relevant for the query and satisfies all the query specifications;
- **Substitute (S):** the item is somewhat relevant: it fails to fulfill some aspects of the query, but the item can be used as a functional substitute;
- **Complement (C):** the item does not fulfill the query but could be used in combination with an exact item;
- **Irrelevant (I):** the item is irrelevant, or it fails to fulfill a central aspect of the query.

The primary objective of this competition is to build new ranking strategies and, simultaneously, identify interesting categories of results (i.e., substitutes) that can be used to improve the customer experience when searching for products. The three different tasks for this KDD Cup competition using our Shopping Queries Dataset are:

- **T1.** Query-Product Ranking
- **T2.** Multiclass Product Classification
- **T3.** Product Substitute Identification

T1 use nDCG score, T2 and T3 use Micro-F1 (Accuracy) to. T1 uses a subset of ECSI dataset, while T2 and T3 use the same dataset. It is natural to treat this competition as two different problems.

Key findings

- **E1.** The order of the product catalog is not randomized.
- **E2.** Most products are used once.
- **E3.** The ESCI label proportion is different between T1 and T2T3.
- **E4.** Most queries have 16 or 40 product numbers, and the label distribution of those queries are slightly different.
- **E5.** The product id is called ASIN and will be identical to ISBN (starts with digits) if the product has ISBN.
- **E6.** Most query products group has fewer unique brand numbers than product numbers and the product with the most frequent brand tends to be labeled as Exact.
- **E7.** At least one product in a query-products group will be labeled as Exact, and the label of the query-product pair is affected by other labels in this group as well

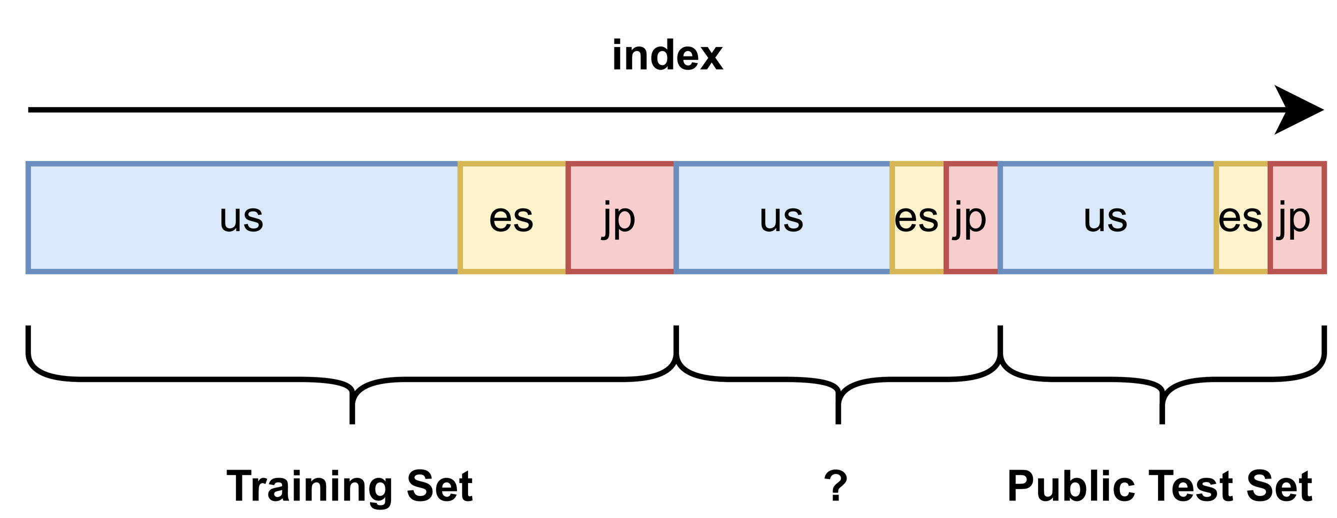

Figure 2. The order of product entries in T2T3.

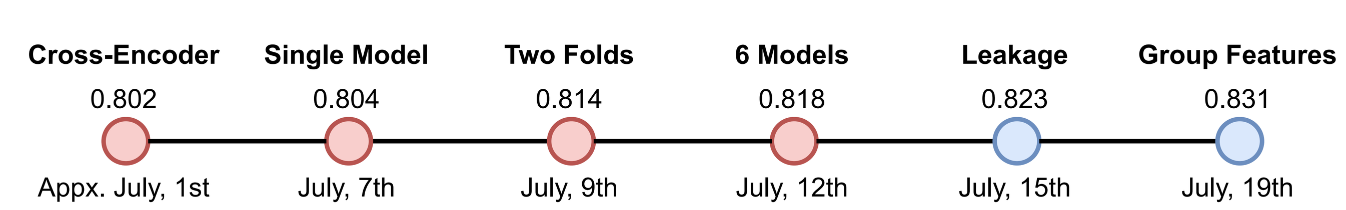

Figure 3. Our milestone of public leaderboard score for T2.

Solution

Figure 1 shows the general schema of our proposed solution for Amazon KDD CUP 2022 for all three tasks. As we planned to attend to all three tasks, for efficiency, we have to train the cross-encoders once and use them for all three tasks. This strategy makes this two-stage solution the only choice. So we trained all cross-encoders with all data from T1 and T2T3 in two folds and then combined the four class probabilities with other essential features, using lightGBM to fuse and calibrate the prediction and adapt results to different tasks. Figure 3 shows the milestone of the public leaderboard score for T2.

- **In the first stage,** we ensembled three cross-encoders [1] for each language that differ from pre-trained models, the training data, or the input fields. For English entries, we used **DeBERTaV3 [2]**, **BigBird [3]** and **COCO-LM [4]**. While for Japanese and Spanish ones, we used a **multi-language version of DeBERTaV3**.
- **In the second stage,** we design several features with cross-encoders predictions to the LightGBM models.

Reference

[1] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019).
[2] Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. arXiv preprint arXiv:2111.09543 (2021).
[3] Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. Advances in Neural Information Processing Systems 33 (2020), 17283–17297.
[4] Yu Meng, Chenyan Xiong, Payal Bajaj, Paul Bennett, Jiawei Han, Xia Song, et al. 2021. Coco-lm: Correcting and contrasting text sequences for language model pretraining. Advances in Neural Information Processing Systems 34 (2021), 23102– 23114.