Amazon KDD Cup 2022 Workshop Presentation

A Winning Solution for All Three Tasks of KDD CUP 2022 ESCI Challenge for Improving Product Search

Team: day-day-up

Jinzhen Lin, Lanqing Xue, Zhenzhe Ying,

Changhua Meng, Weiqiang Wang, Haotian Wang, Xiaofeng Wu

Ant Group (Hangzhou, China)

Background

Amazon published a large scale product search dataset and hosted KDD CUP 2022 ESCI challenge.

- Four classes of query and product relevance: Exact (E), Substitute (S), Complement (C) and Irrelevant (**I**).
- The dataset contains queries in English, Japanese and Spanish.
- The task2 and the task3 use the same dataset, while the task1 use a smaller dataset with different ESCI distribution.

Task1 (Ranking): Query-Product Ranking

Rank all of the products given a query and a set of products.

Metric: NDCG

Task2 (Classification): Multiclass Product Classification

Find the relevance class (E, S, C, I) of each (query, product) pair.

Metric: Micro-F1 (=Accuracy for classification problem)

Task3 (Classification): Product Substitute Identification

Find whether the product is a substitute for a given query

Metric: Micro-F1 (=Accuracy for classification problem)

Dataset ESCI distribution (in %)

Dataset	E	S	C	I
Task1 (Small Dataset)	43.72	34.33	5.13	16.82
Task2 & Task3 (Large Dataset)	65.20	21.91	2.89	10.00

Overall Architecture

Training Dataset: Concatenate the training set of all three tasks and remove duplicates

Model Structure: Cross-Encoder based on InfoXLM-large

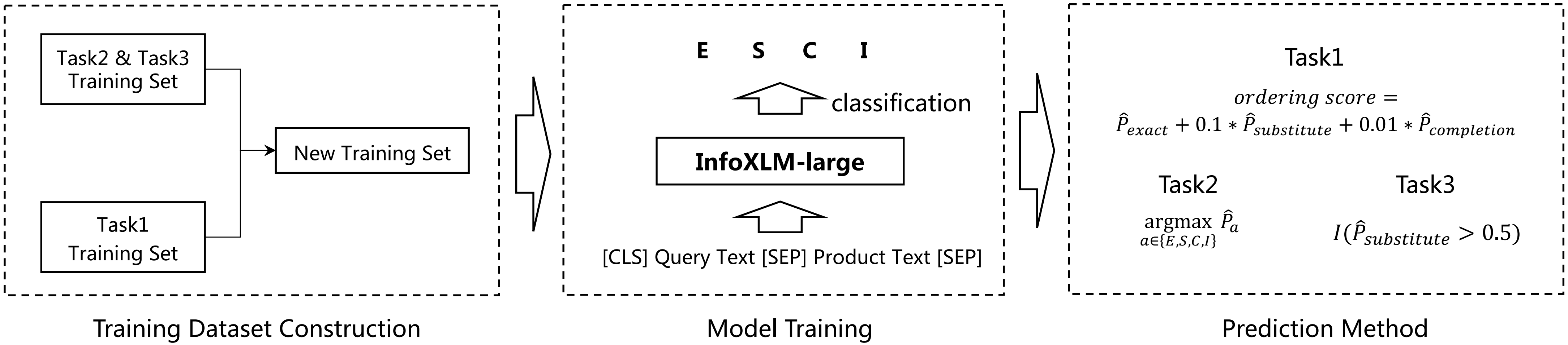
Training Objective: Use the objective function of the task2 (multiclass classification)

Prediction Method:

Task1 (Ranking): Order the product by $\hat{P}_{exact} + 0.1 * \hat{P}_{substitute} + 0.01 * \hat{P}_{completion}$

Task2 (Classification): Take the label with the highest prediction probability as the prediction result

Task3 (Classification): Check whether $\hat{P}_{substitute}$ is greater than 0.5



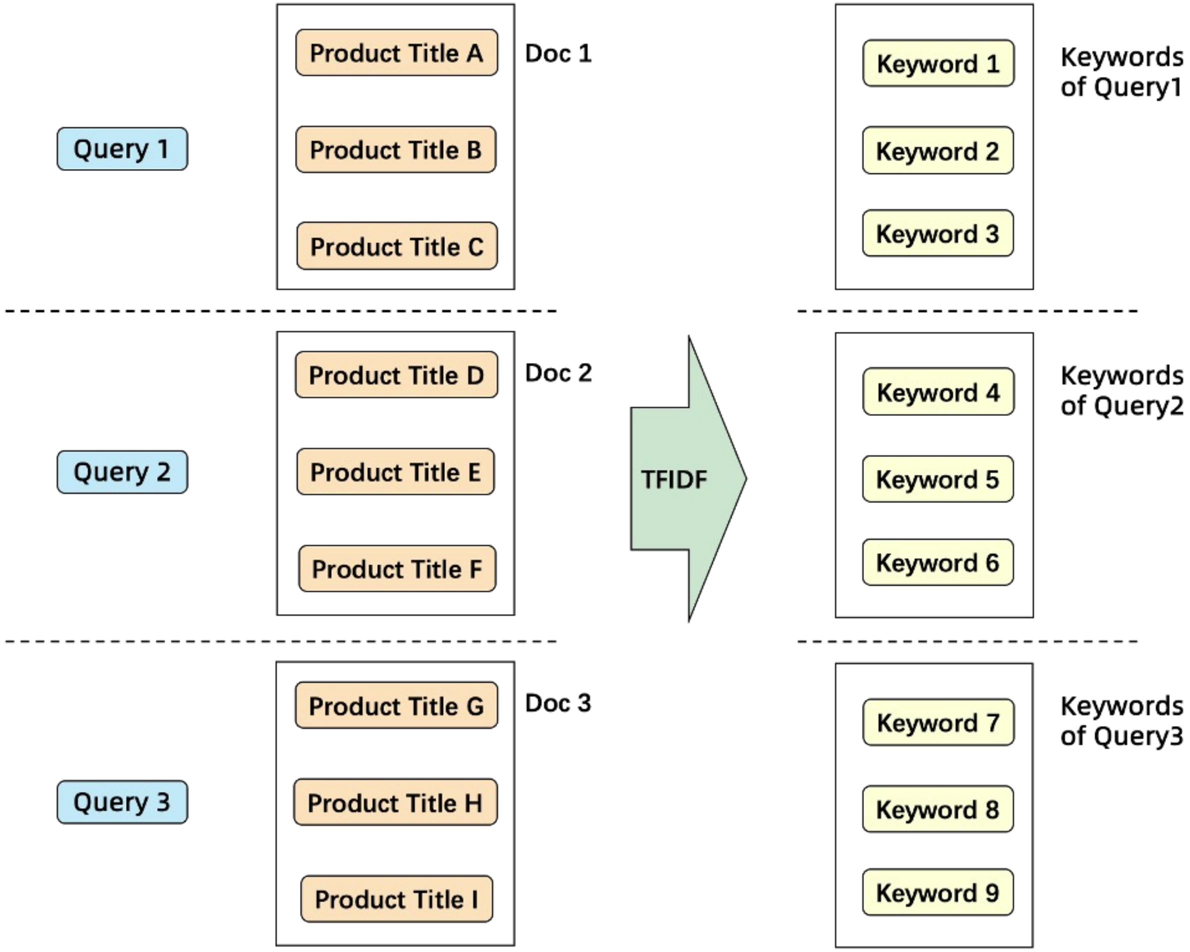
Query-Based TF-IDF Procedure

Query	Query Locale
seagate 250gb	English
v-shaped pillow	English
muelle disfraz	Spanish
炭マスク	Japanese

The queries are short text, it is difficult for the model to understand the queries accurately. **How to solve this?**

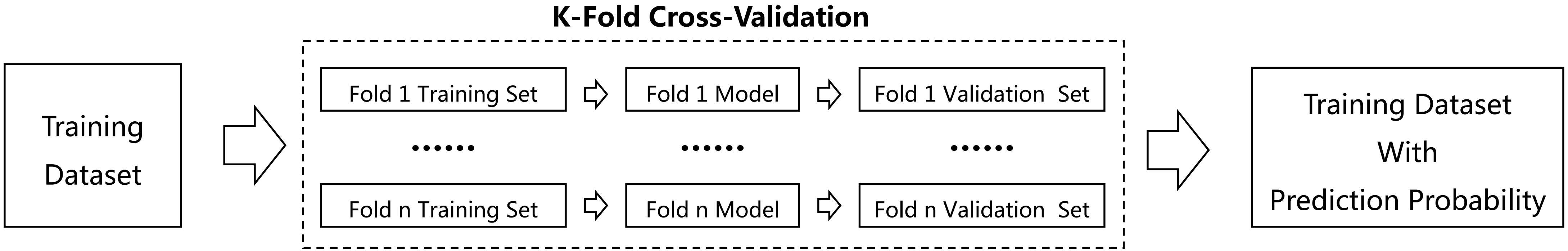
Our Solution:

- Extract the keywords of the query with query-based TF-IDF procedure and treat them as features of the query.
- Add the brand and color names of all products with the same query to the model.



Self Distillation

Use **self distillation** to overcome the impact of **noise data** and improve **model robustness**



ID	E	S	C	I
1	1	0	0	0
2	0	0	1	0
3	0	1	0	0
4	1	0	0	0
5	0	0	0	1

True Label

×0.5

ID	E	S	C	I
1	0.80	0.10	0.05	0.05
2	0.20	0.05	0.75	0.0
3	0.00	0.95	0.05	0.00
4	0.65	0.20	0.10	0.05
5	0.05	0.05	0.05	0.85

Soft Label
For Training

×0.5

ID	E	S	C	I
1	0.6	0.2	0.1	0.1
2	0.4	0.1	0.5	0.0
3	0.0	0.9	0.1	0.0
4	0.3	0.4	0.2	0.1
5	0.1	0.1	0.1	0.7

Prediction Probability

ANT GROUP

Other Methods

Some piecemeal methods :

- **Ensemble**: simple probability averaging
  - 8 models for task1
  - 4 models for task2 and task3
  - use some methods to improve the difference between models
- **Post Processing**: post process the prediction probability with rule and LightGBM
- **External Dataset**: some additional text of products, contribute little to the final score
- **Model Acceleration**: speed up model inference

See our paper for more details.

Results and Conclusion

With our solution, our team **day-day-up** won 1st place in the task2 and the task3, and won 3rd place in task1.

Task1 Final Leaderboard

Rank	Team Name	Score (Private)	Score (Public)
1	www	0.9043	0.9057
2	qinpersevere	0.9036	0.9047
3	**day-day-up**	**0.9035**	**0.9056**
4	GraphMIRAcles	0.9028	0.9036
5	ZhichunRoad	0.9025	0.9035
6	ETS-Lab	0.9014	0.9025
7	ALONG	0.9014	0.8999
8	ljr333	0.9008	0.9012
9	NeuralMind	0.9007	0.9012
10	zackchen	0.8998	0.9030

Task3 Final Leaderboard

Rank	Team Name	Score (Private)	Score (Public)
1	**day-day-up**	**0.8790**	**0.8766**
2	ETS-Lab	0.8771	0.8749
3	Uni	0.8754	0.8744
4	cmb-ai	0.8734	0.8708
5	LYZD-fintech	0.8708	0.8688
6	qinpersevere	0.8701	0.8684
7	wookiebort	0.8687	0.8673
8	ZhichunRoad	0.8686	0.8678
9	NTT-DOCOMO-LABS-GREEN	0.8677	0.8655
10	rein20	0.8668	0.8652

Task2 Final Leaderboard

Rank	Team Name	Score (Private)	Score (Public)
1	day-day-up	0.8326	0.8320
2	ETS-Lab	0.8325	0.8303
3	Uni	0.8273	0.8281
4	cmb-ai	0.8251	0.8234
5	MetaSoul	0.8207	0.8182
6	www	0.8204	0.8209
7	ZhichunRoad	0.8194	0.8176
8	qinpersevere	0.8191	0.8181
9	zackchen	0.8189	0.8212
10	LYZD-fintech	0.8183	0.8177

Thank You!

Contact us: **jinzhen.ljz@antgroup.com (Jinzhen Lin)**