KDD-Cup’22

Amazon Product Search Competition

Background

1. In the competition, we build product query search model and it is a rank task.
2. There are large-scale multilingual query-product dataset. (781744 pair)

Open source code : https://github.com/ChengHSUHSU/KDD_Cup2022_AmazonProductSearchESCI

Data Introduction

us, es, jp

Exact (E),
Substitute (S),
Complement (C),
Irrelevant (I)

	query_id	query	query_locale	product_id	esci_label	product_new_id
0	0	!awnmower tires without rims	us	B00004RA3F	irrelevant	B00004RA3F@us
1	0	!awnmower tires without rims	us	B0018TWDOI	exact	B0018TWDOI@us
2	0	!awnmower tires without rims	us	B005O5Y3QI	exact	B005O5Y3QI@us

	product_id	product_title	product_description	product_bullet_point	product_brand	product_color_name	product_locale	product_new_id
0	B0188A3QRM	Amazon Basics Woodcased #2 Pencils, Unsharpene...	Empty	144 woodcase #2 HB pencils made from high-qual...	Amazon Basics	Yellow	us	B0188A3QRM@us
1	B075VXJ9VG	BAZIC Pencil #2 HB Pencils, Latex Free Eraser,...	<p>BACK TO BAZIC</p><p>Our go...	⭐ UN-SHARPENED #2 PREMIUM PENCILS. Each...	BAZIC Products	12-count	us	B075VXJ9VG@us
2	B07G7F6JZ6	Emraw Pre Sharpened Round Primary Size No 2 Ju...	<p><b>Emraw Pre-Sharpened #2 HB Wood Pencils -...	✓ PACK OF 8 NUMBER 2 PRESHARPENED BEGINNERS PE...	Emraw	Yellow	us	B07G7F6JZ6@us

Problem Statement

Input:

query_id	query	query_locale	product_id
Query_1	"Query_1"	us	product_23
Query_2	"Query_2"	us	product_234

	Exact (E)	Substitute (S)	Complement (C)	Irrelevant (I)
gain	1.0	0.1	0.01	0.0

Output:

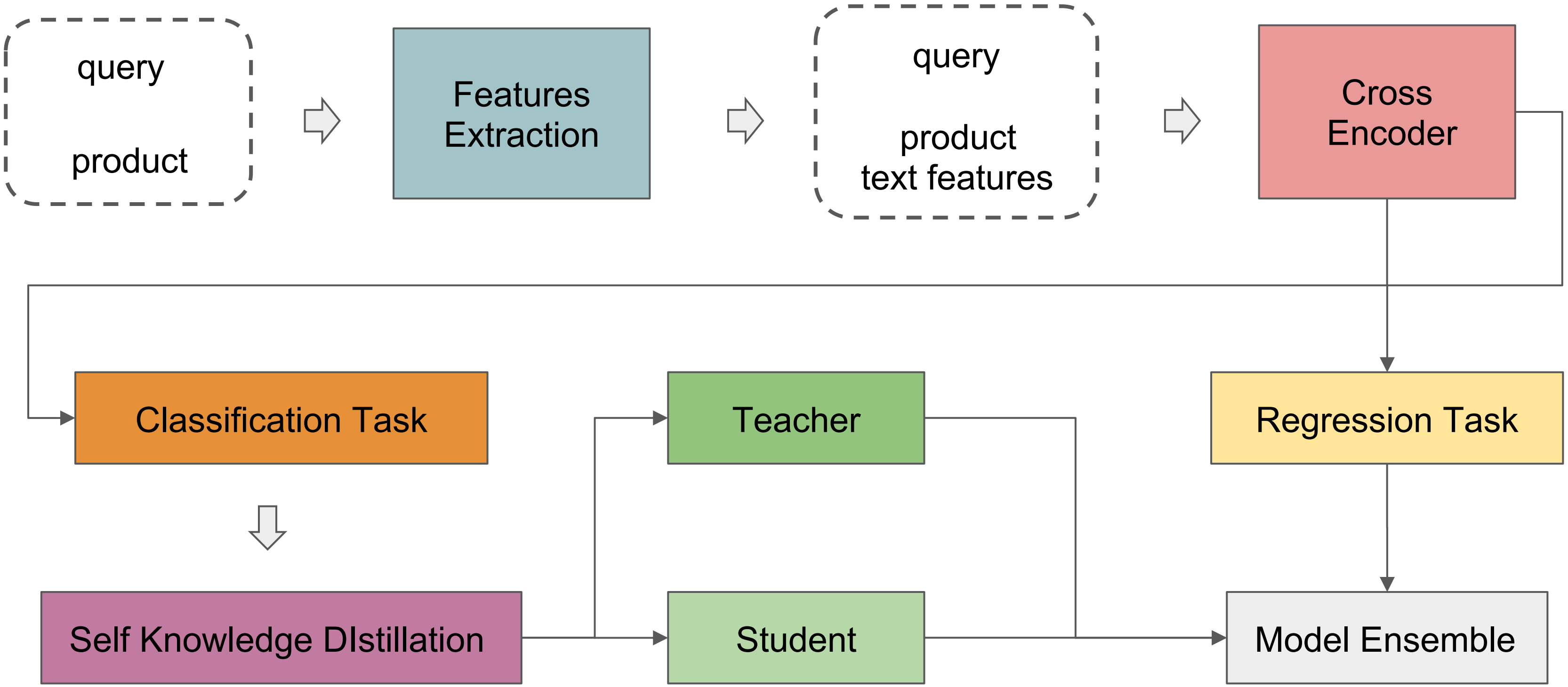
query_id	product_id
Query_1	product_50
Query_1	product_900
Query_1	product_80
Query_2	product_32

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log_2(i+1)}$$

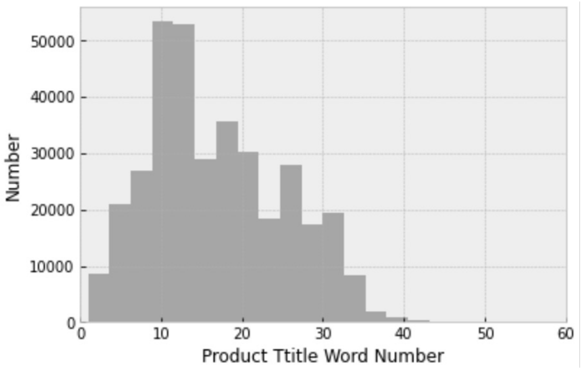
$$IDCG_p = \sum_{i=1}^{|REL_p|} \frac{2^{rel_i} - 1}{log_2(i+1)}$$

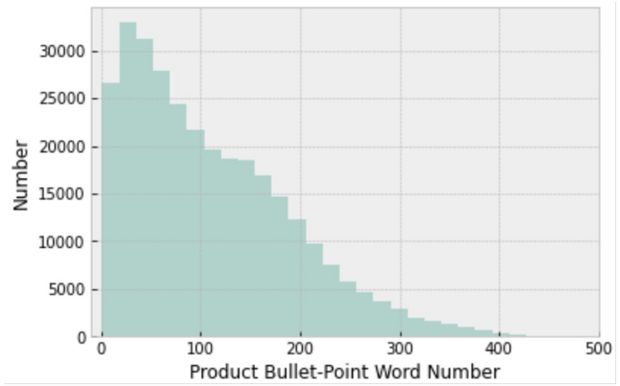
$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

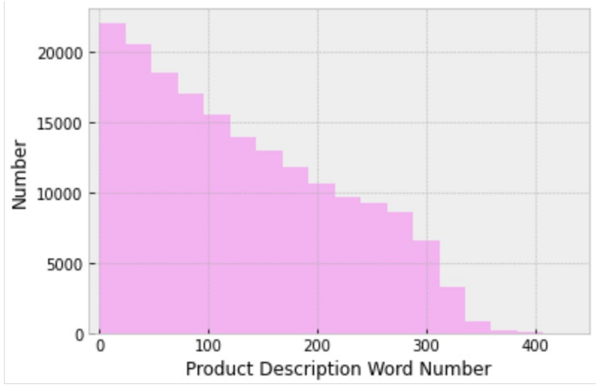
Model Framework



Product Documents Features Extraction (1)

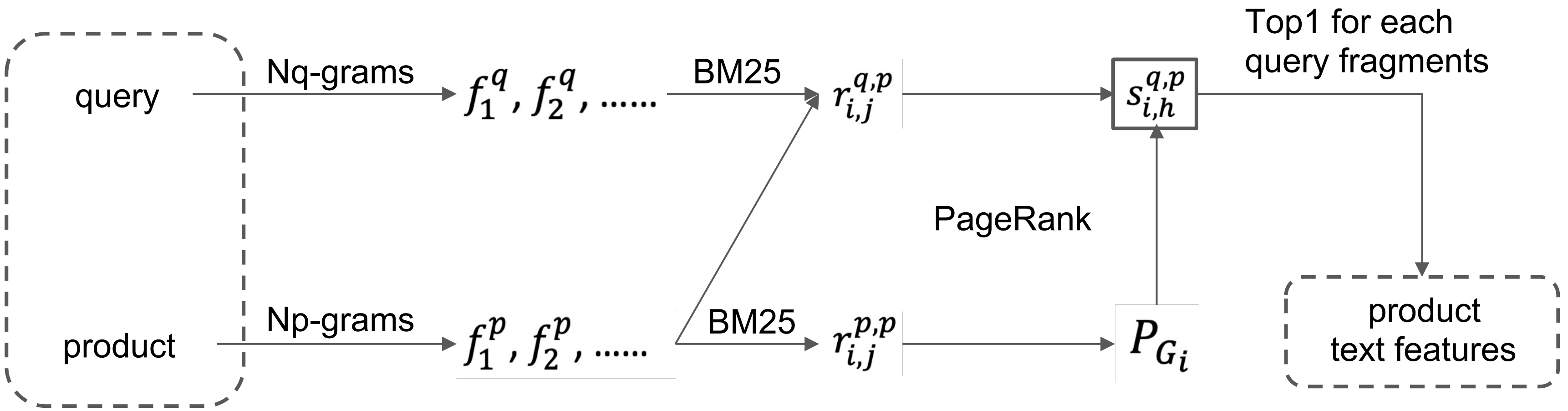




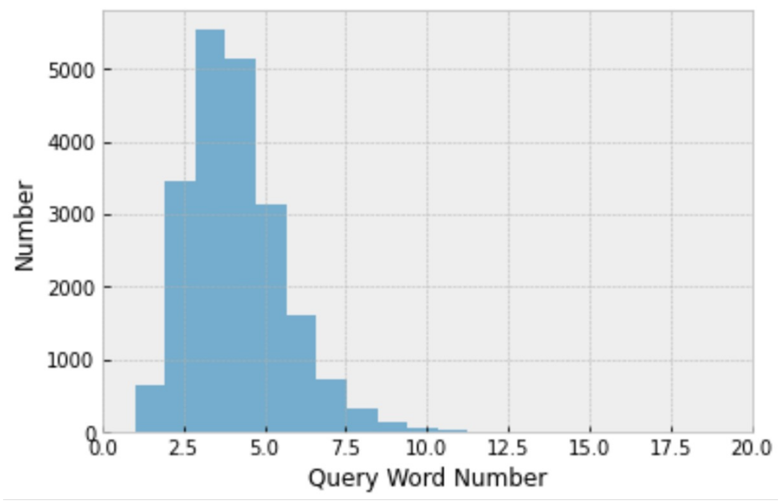


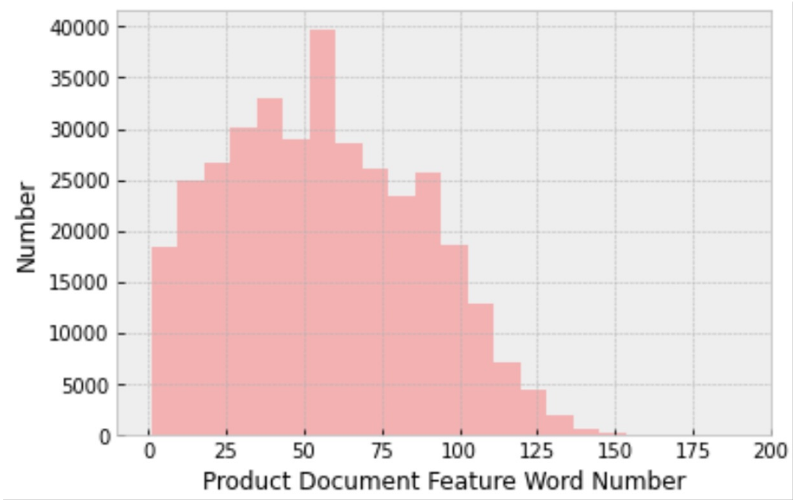
Locale : US

Product Documents Features Extraction (2)



Product Documents Features Extraction (3)





Locale : US

Cross Encoder

[CLS] + query + [SEP] + product text features + [SEP]

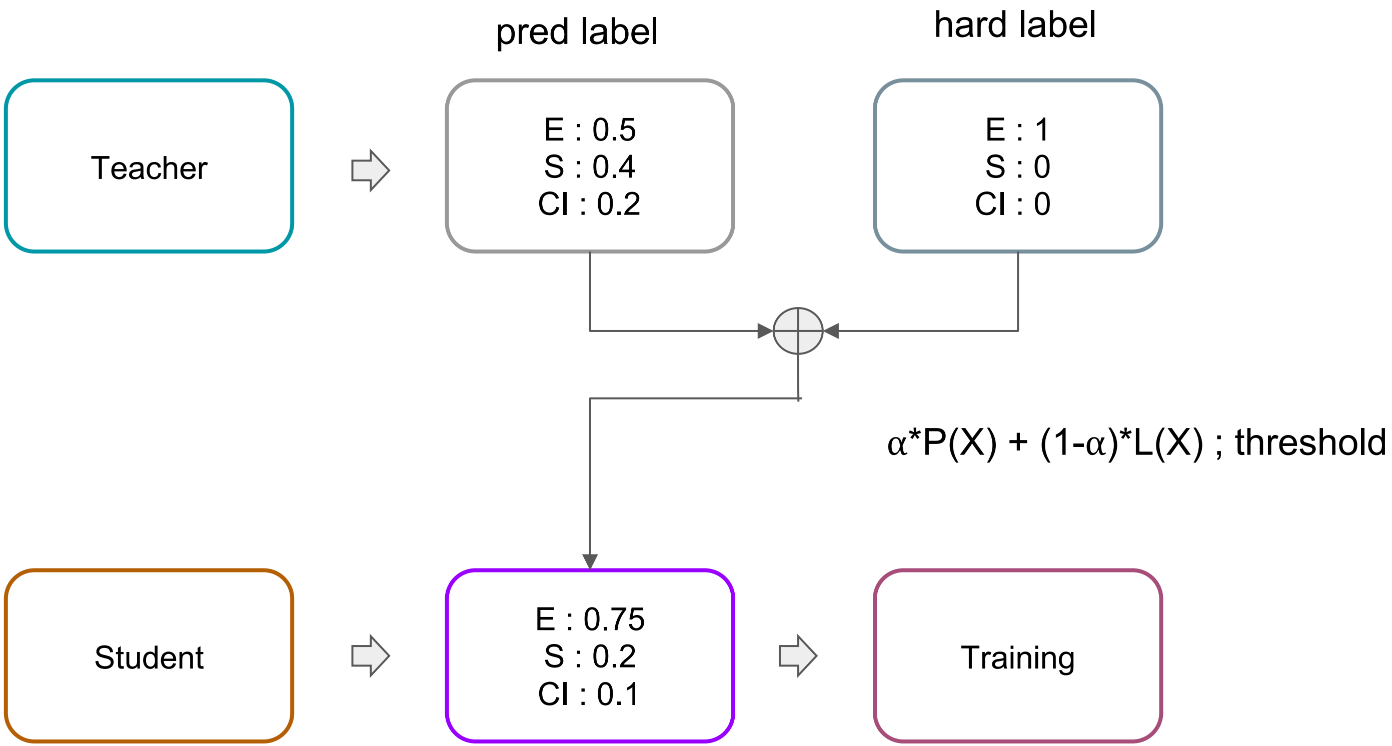
1. Regression : ranking order by predicted value

2. Classification : $P_E * \, gain_E + P_S * \, gain_S + P_C * \, gain_C + P_I * \, gain_I$

Note : After some experiments, we found that Regression task can balancely show well performance and Classification task can achieve better performance in Exact(E) label

α = 0.5 ; threshold=0.8

Self-Knowledge Distillation



Training Strategies

1. Regression task : We design two training stage. First, we set Exact (E) label as 1.0 and other as 0.0. Second, we set Exact (E) label as1.0, Substitute (S) label as 0.4 and other as 0.0.
2. Classification task : After training large model (teacher), we use self-knowledge distillation and label smoothing to train student model.
3. Model Ensemble : Regression-Second-Stage + Classification-Teacher + Classification-Student

Experiment Setting

1. All modeling are one-epoch
2. Metric : nDCG (gain : official setting)
3. Optimizor : AdamW
4. Loss function : MSE, Cross Entropy
5. Learning Rate : 7e-5 (warm-down mechanism)

Experiment Performance

	English(US)				Spanish(ES)				Japanese(JP)			
Label	**TFKD**	**CLF**	**REG**	**CLF+REG**	**TFKD**	**CLF**	**REG**	**CLF+REG**	**TFKD**	**CLF**	**REG**	**CLF+REG**
E-S	0.835	0.829	0.824	0.829	0.831	0.819	0.817	0.823	0.839	0.831	0.829	0.831
E-CI	0.891	0.884	0.882	0.886	0.893	0.887	0.884	0.889	0.891	0.887	0.886	0.888
S-CI	0.823	0.798	0.816	0.819	0.821	0.784	0.812	0.815	0.821	0.768	0.819	0.817
ESCI	0.894	0.885	0.884	0.888	0.896	0.888	0.889	0.890	0.897	0.891	0.892	0.892

Experiment (Feature Extraction)

Component	E-S	E-CI	S-CI	ESCI
None	0.825	0.880	0.820	0.884
N_q=2,N_p=8	0.834	0.891	0.823	0.891
N_q=3,N_p=8	0.836	0.891	0.821	0.896
N_q=4,N_p=8	0.836	0.890	0.823	0.895
N_q=2,N_p=9	0.833	0.889	0.822	0.890
N_q=3,N_p=9	0.835	0.892	0.820	0.896
N_q=4,N_p=9	0.833	0.891	0.822	0.893

Experiment (Self-Knowledge Distillation)

Table 4: Self-knowledge distillation experiment. we set the range of knowledge distillation threshold t is in (0.0, 0.7, 0.8). We don't use teacher model output as soft label when nDCG score of given query is lower than t and we use label smoothing as soft label. When t is 0.0, it means we only use teacher model output as self label.

Component	E-S	E-CI	S-CI	ESCI
teacher	0.831	0.889	0.823	0.892
student(t=0.0)	0.826	0.884	0.817	0.886
student(t=0.7)	0.827	0.885	0.817	0.889
student(t=0.8)	0.828	0.885	0.816	0.890
ensemble(t=0.0)	0.833	0.889	0.819	0.893
ensemble(t=0.7)	0.835	0.891	0.821	0.895
ensemble(t=0.8)	0.836	0.891	0.821	0.896

Conclusion

1. In the paper, we use cross encoder to predict the relevance of query, product. When user query intent is ambiguous, the performance is not well.
2. In the future, we build query-product word-level graph and use GNN to represent more semantic embedding. We believe that it can improve more performance.