A simple but effective solution for Task 1 of KDD Cup 2022 Challenge on improving product search

Jinrui Liang, Yali Shangguan, Zhaohao Liang
Netease Games Operation Department

Outline

- Datasets and Features
- Models
    - Problem Definition
    - Dataset Split
    - Model Selections
- Training Settings
- Experimental Results
    - Local Results
    - Online Results
    - Unworked Attempts

Datasets and Features

- Contains a list of query-result paired with annotated E/S/C/I labels, and it includes queries from English, Japanese and Spanish [2].
- Contains many fields from the query and product
- Relationship between Product type and Gain

Table 1: Relationship between Product type and Gain

Product type	Gain
Exact(E)	1.0
Substitute(S)	0.1
Complement(C)	0.01
Irrelevant(I)	0.0

[2] Chandan K. Reddy, Lluís Màrquez, Fran Valero, Nikhil Rao, Hugo Zaragoza, Sambaran Bandyopadhyay, Arnab Biswas, Anlu Xing, and Karthik Subbian. 2022.Shopping Queries Dataset: A Large-Scale ESCI Benchmark for Improving Product Search. arXiv:2206.06588

Models

- Problem Definition
  - Inputs
    - query + <SEP> + product title + ‘ ’ + product description + ‘ ’ + product bullet point + ‘ ’ + product brand + ‘ ’ + product color name
  - Outputs
    - A value, which is regarded as the predicted gain for the input <query, product>
- Dataset Split
  - the proportion of the validation set: 0.1
  - use the data that appear in Task 1 from Task 2 as supplementary data

Models

- Model Selections
  - BERT Encoder + Regression Layer
  - supplementary data(data-sup) + k-fold

Table 2: Model Selections

Locale	Selected Model
us	DeBERTaV3-large [1]
es	mDeBERTaV3-base [1]
jp	mDeBERTaV3-base [1]

[1] Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. CoRR abs/2111.09543 (2021). arXiv:2111.09543 https://arxiv.org/abs/2111.09543

Training Settings

- strictly follow the implementation described in the original paper [1] for the settings of the hyper-parameters

Table 3: Training settings for DeBERTaV3-large

Hyper-parameters	Value
epochs	3
learning rate	8e-6
batch size	32
random state	42
learning rate scheduler	linear scheduler
weight decay	0.01
Adam betas	(0.9, 0.999)
Adam eps	1e-6
grad max norm	1.0
warmup steps	100
dropout rate	0.15
validation steps	500

Table 4: Training settings for mDeBERTaV3-base

Hyper-parameters	Value
epochs	3
learning rate	2e-5
batch size	32
random state	42
learning rate scheduler	linear scheduler
weight decay	0.01
Adam betas	(0.9, 0.999)
Adam eps	1e-6
grad max norm	1.0
warmup steps	100
dropout rate	0.1
validation steps	500

[1] Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing. CoRR abs/2111.09543 (2021). arXiv:2111.09543 https://arxiv.org/abs/2111.09543

Experimental Results

- Local Results
  - Both supplementary data and k-fold training consistently improve the model results
  - Combining supplementary data and k-fold training gives the best performance

Table 5: Local results of us locale

Model	Fold	Val nDCG
baseline	-	0.8972
baseline-sup	-	0.8998
baseline-2fold	0	0.8975
baseline-2fold	1	0.8967
baseline-2fold	ensemble	0.9005
baseline-sup-2fold	0	0.9013
baseline-sup-2fold	1	0.9007
baseline-sup-2fold	**ensemble**	**0.9021**

Table 6: Local results of es locale

Model	Fold	Val nDCG
baseline	-	0.8997
baseline-sup	-	0.9011
baseline-2fold	0	0.8995
baseline-2fold	1	0.8989
baseline-2fold	ensemble	0.9010
baseline-sup-2fold	0	0.9051
baseline-sup-2fold	1	0.9043
baseline-sup-2fold	**ensemble**	**0.9053**

Table 7: Local results of jp locale

Model	Fold	Val nDCG
baseline	-	0.8971
baseline-sup	-	0.8974
baseline-2fold	0	0.8984
baseline-2fold	1	0.8969
baseline-2fold	ensemble	0.8985
baseline-sup-2fold	0	0.8993
baseline-sup-2fold	1	0.8989
baseline-sup-2fold	**ensemble**	**0.9001**

Experimental Results

- Online Results
  - baseline-sup-2fold-all gives the best result

Table 8: Online(Final) results

Model	Public Test nDCG	Private Test nDCG	
baseline(official)	0.8503	-	
baseline(ours)	0.8968	-	+0.0465
baseline-sup-us	0.8981	-	
baseline-sup-all	0.8983	-	+0.0480
baseline-2fold-us	0.8983	-	
baseline-2fold-all	0.8988	-	+0.0485
baseline-sup-2fold-us	0.9009	-	
baseline-sup-2fold-es-jp	0.9008	-	
baseline-sup-2fold-all	**0.9012**	**0.9008**	+0.0509

Experimental Results

- Unworked Attempts
  - Pseudo labels
  - 4-fold
  - InfoXLM