JBNU at TREC 2023 Product Search Track

Gi-taek An, Woo-Seok Choi, Jun-Yong Park, Kyung-Soon Lee

Department of Computer Science & Artificial Intelligence,
CAIIT, Jeonbuk National University, Republic of Korea
{gt, ccwwsss, pjy010608, selfsolee}@jbnu.ac.kr

Abstract

This paper describes the participation of the JBNU team for TREC 2023 Product Search Track. Our primary focus revolves around tackling the issue of performance degradation in queries. We categorize queries into specific and abstract types, leveraging the power of the DeBERTa deep learning model for reranking. This enhancement involves the incorporation of nine specialized tokens, such as brand, material, category, and others, and is specifically applied to queries of the specific type.

1. Introduction

The TREC 2023 Product Search Track [1] centers on information retrieval within the domain of product search, aiming to assist users in locating the products they desire by aligning with their objectives and intentions. Our team, JBNU, has participated in the Product Ranking Task and Product Retrieval Task.

In the context of product search, we have observed the frequent occurrence of common errors in queries. Traditional typo correction methods often led to incorrect corrections for words that are not commonly found in dictionaries, such as product names, brand names, and author names in product search queries. To tackle this challenge, we have created a specialized dictionary designed to refine and correct product search queries.

For all tasks, the queries undergo the following preprocessing steps:

- Translation of multilingual queries to English utilizing Googletrans [4].
- Typo correction in queries using a dedicated product search dictionary for Pyspellchecker [5].
- Replacement of product codes (ASIN) in queries with the corresponding product titles.

Furthermore, we observed a common occurrence of product attributes within queries. We pinpointed attributes from the product information that held notable relevance to the queries and integrated nine special tokens within our deep learning methodology to ensure the attribute information substantially influences the learning and inference processes.

We categorize queries into specific and abstract types, and our reranking process with a deep learning model is specifically targeted at the specific query types. These specific query types are characterized by the inclusion of one of nine special tokens, such as brand name, color, material, and more.

In our experiments for Product Ranking Task, we employed the DeBERTa (Disentangled Attention and Enhanced Mask Decoder) model [2] with the incorporation of special tokens. For Product Retrieval Task, we employed the BM25 model [3] for the initial retrieval and employed the DeBERTa model for reranking, with the specific approach varying depending on the query type.

2. Submitted Runs

We have submitted two runs for the Product Ranking Task.

- JBNU-1: We utilized a model trained with DeBERTa, incorporating special tokens.
- JBNU-2: We extended the categories of the JBNU-1 model into a single token implementing fixed character tokenization derived from category information.

Three runs are submitted for the Product Retrieval Task.

- JBNU-A: Reranking with the JBNU-1 model for the specific query type for initial search through query preprocessing.
- JBNU-B: Reranking with the JBNU-2 model for the specific query type for initial search through query preprocessing.
- JBNU-C: Searching by expanding the query with the top n product titles through Pseudo-relevance feedback, and enhancing the relevance of titles and categories.
- JBNU-D (Off-Run): Reranking for candidate retrieval from JBNU-C using the JBNU-1 model for the specific query type.

3. Experimental Results

The results of our performance in the Product Ranking Task and Product Retrieval Task can be found in Table 1 and Table 2, respectively.

Run Name	NDCG	NDCG@10	P@10	Recip_rank
JBNU-1	0.5901	0.6531	0.6070	0.8930
JBNU-2	**0.5927**	**0.6583**	**0.6075**	**0.9252**

Table 1. Results of Product Ranking Task.

Run Name	NDCG	NDCG@10	P@10	Recip_rank
JBNU-A	0.6764	0.5989	0.5489	0.8735
JBNU-B	0.6639	0.5763	0.5177	0.8919
JBNU-C	<u>0.7716</u>	<u>0.7251</u>	<u>0.6737</u>	<u>0.9180</u>
JBNU-D(Off-Run)	**0.781**	**0.7502**	**0.6892**	**0.9258**

Table 2. Results of Product Retrieval Task.

4. Conclusion

Expanding queries with product titles has shown its effectiveness in product retrieval. Furthermore, our experiments for product ranking have affirmed the advantages of employing query-specific deep learning models.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW) supervised by the IITP(Institute of Information & communications Technology Planing & Evaluation) in 2023”(2022-0-01067)

References

[1] TREC2023 Product Search Track(2023). https://trec-product-search.github.io/

[2] He, P., Liu, X., Gao, J., & Chen, W. (2020). Deberta: Decoding-enhanced bert with disentangled attention. arXiv preprint arXiv:2006.03654.

[3] Elasticsearch(2023). Elasticsearch Platform. https://www.elastic.co/

[4] Googletrans(2023). Googletrans: Free and Unlimited Google translate API for Python. https://py-googletrans.readthedocs.io/en/latest/

[5] Pyspellchecker(2023). Quickstart - pyspellchecker 0.7.2 documentation. https://pyspellchecker.readthedocs.io/en/latest/quickstart.html