BLADE:
The University of Maryland at the TREC 2023 NeuCLIR Track

Suraj Nair[1,2] and Douglas W. Oard[1,3]

[1]UMIACS
[2]Computer Science Department
[3]College of Information Studies
University of Maryland, College Park
suraj.nair04@gmail.com

Abstract

The University of Maryland submitted three runs to the Ad Hoc CLIR Task of the TREC 2023 NeuCLIR track. This paper describes three systems that cross the language barrier using a learned sparse retrieval model using bilingual embeddings.

1 Introduction

The TREC NeuCLIR track focuses on Cross-Language Information Retrieval (CLIR), where the goal is to match English queries with documents that are written in different languages (Chinese, Persian, and Russian).

2 BLADE

A key challenge in building CLIR systems is to match the queries with documents that use a different vocabulary. Traditional lexical-matching systems such as BM25 [6], which relies on term frequencies and inverse document frequencies, would not work straight off the shelf due to the vocabulary mismatch problem. In our TREC 2022 NeuCLIR submission, we used a strong non-neural baseline system called Probabilistic Structured Queries (PSQ) [1], which leverages translation techniques from Statistical Machine Translation (SMT). PSQ maps term frequency vectors from the document language to the query language using a matrix of translation probabilities (normalized to sum to one in the document language to query language direction) as a simple matrix-vector product. These translation probabilities are learned through word-alignment tools (e.g., GIZA++ [5], BerkeleyAligner [3]) that are traditionally part of SMT systems. Any traditional term weighting function that can accept partial term counts (e.g., BM25 or query likelihood) can then be computed on the resulting term frequency vector.

In our submission this year, we chose a learned sparse neural retrieval model called BLADE [4], whose architecture is inspired by its monolingual cousin, SPLADE [2]. BLADE (Bilingual Lexical AnD Expansion model) is a neural CLIR method that creates sparse bilingual vectors for queries and documents in two different languages. BLADE is powered by a bilingual language model[1] created by pruning a multilingual pretrained language model (mBERT). The BLADE model is trained in two stages: i) intermediate

[1]This work was done prior to Suraj Nair joining Amazon.
[1]An example EN-ZH model: `https://huggingface.co/Geotrend/bert-base-en-zh-cased`

pretraining, where the Masked Language Model (MLM) head is pretrained using aligned texts (e.g., comparable passages), and ii) task-specific fine-tuning, where the model is trained on English queries paired with translated mMARCO passages.

For the official TREC submission, we use the BLADE-C variant, which uses comparable passages for intermediate pretraining. The training and inference setup are listed in the BLADE paper [4]. We publicly release the trained models for the three NeuCLIR languages.[2]

3 Results

Table 1 shows the results of the BLADE runs on the three NeuCLIR languages with the full relevance judgment set (the earlier notebook paper were based on partial relevance judgments). In addition, we also report PSQ results and the Reciprocal Rank Fusion results of the two systems, also both on the full relevance judgment set. All runs complied with the NeuCLIR track's definition of an automatic run using title queries.

System	Run	Language	nDCG@20	MAP	Recall@100	Recall@1000
BLADE	zho-umd_hcil-AELN_blade	Chinese	0.3597	0.2849	0.6221	0.8716
BLADE	fas-umd_hcil-AELN_blade	Persian	0.3379	0.2644	0.5476	0.7726
BLADE	rus-umd_hcil-AELN_blade	Russian	0.3409	0.2651	0.5334	0.7835
PSQ	-	Chinese	0.2932	0.2262	0.5544	0.7957
PSQ	-	Persian	0.3834	0.3127	0.5970	0.8115
PSQ	-	Russian	0.3292	0.2632	0.5040	0.7424
Fusion	PSQ + BLADE	Chinese	0.3946	0.3204	0.7155	0.9095
Fusion	PSQ + BLADE	Persian	0.4260	0.3436	0.6559	0.8871
Fusion	PSQ + BLADE	Russian	0.3893	0.3144	0.6445	0.8632

Table 1: Official evaluation measures for PSQ, BLADE, and their reciprocal rank fusion, using title queries.

Comparing the two individual systems, we observe that BLADE has a numerically higher nDCG in Chinese and Russian, whereas PSQ has a numerically higher nDCG in Persian. However, reciprocal rank fusion of the two substantially outperforms the individual systems by all the measures.

4 Conclusion

In this paper, we describe our BLADE runs that were submitted to the TREC 2023 NeuCLIR track. We leverage a learned sparse neural retrieval model using bilingual embeddings, fine-tuned on comparable passages.

5 Acknowledgments

We would like to thank Eugene Yang for providing us with the PSQ runs for NeuCLIR 2023 in order to perform system fusion experiments.

References

[1] K. Darwish and D. W. Oard. Probabilistic structured query methods. In *Proceedings of the 26th Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 338–344, 2003.

[2] An example EN-ZH model: https://huggingface.co/srnair/blade-en-zh

[2] T. Formal, C. Lassance, B. Piwowarski, and S. Clinchant. Splade v2: Sparse lexical and expansion model for information retrieval. *arXiv preprint arXiv:2109.10086*, 2021.

[3] A. Haghighi, J. Blitzer, J. DeNero, and D. Klein. Better word alignments with supervised ITG models. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 923–931, Suntec, Singapore, Aug. 2009. Association for Computational Linguistics.

[4] S. Nair, E. Yang, D. Lawrie, J. Mayfield, and D. W. Oard. Blade: Combining vocabulary pruning and intermediate pretraining for scaleable neural clir. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 1219–1229, New York, NY, USA, 2023. Association for Computing Machinery.

[5] F. J. Och and H. Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51, 2003.

[6] S. E. Robertson, S. Walker, S. Jones, M. M. Hancock-Beaulieu, M. Gatford, et al. Okapi at TREC-3. *NIST Special Publication Sp 109*, 1995.