Domain Adaptive Dense Retrieval with Query Generation

Authors : Rui Yin, Haojie Wang, Xun Li

Abstract : Recently, mainstream dense retrieval methods have obtained state-of-the-art results on some datasets and tasks. However, they require large amounts of training data, which is not available in most domains. The severe performance degradation of dense retrievers on new data domains has limited the use of dense retrieval methods to only a few domains with large training datasets. In this paper, we propose an unsupervised domain-adaptive approach based on query generation. First, a generative model is used to generate relevant queries for each passage in the target corpus, and then, the generated queries are used for mining negative passages. Finally, the query-passage pairs are labeled with a cross-encoder and used to train a domain-adapted dense retriever. We also explore contrastive learning as a method for training domain-adapted dense retrievers and show that it leads to strong performance in various retrieval settings. Experiments show that our approach is more robust than previous methods in target domains that require less unlabeled data.

Keywords : dense retrieval, query generation, contrastive learning, unsupervised training
Conference Title : ICSLP 2023 : International Conference on Speech and Language Processing
Conference Location : San Francisco, United States
Conference Dates : November 06-07, 2023