

Sampling from Bounded Clusters for Diverse Relevant Argument Retrieval

XFACT-team0331 at PerspectiveArg2024 – Wan Ju Kang, Jiyoung Han, Jaemin Jung, and James Thorne

soarhigh@kaist.ac.kr Korea Advanced Institute of Science and Technology

Problem Statement

Perspective Argument Retrieval:

On top of the argument retrieval problem, also incorporate the demographic properties of the argument authors when retrieving the arguments

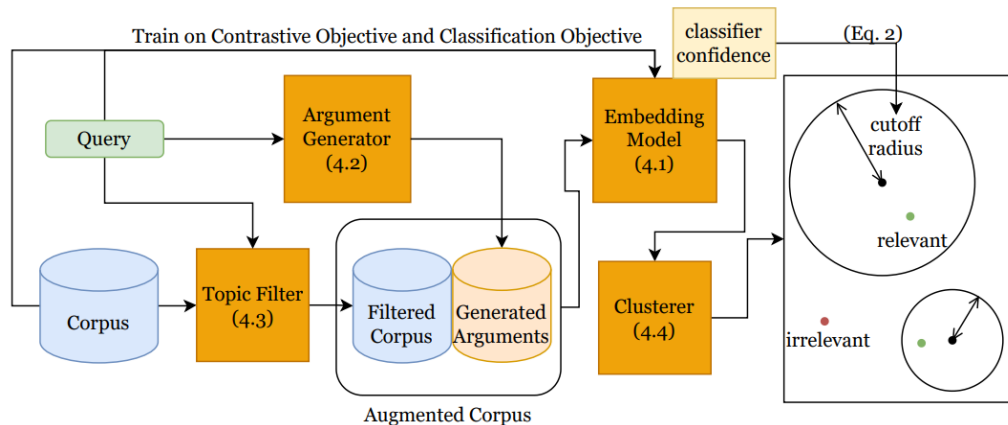
Capture the relevance of the arguments with respect to the query, as well as the diversity of the argument authors' profiles

Dataset

2019 Swiss general elections interview data, with social issues and interviewee response and socio-cultural variables

Methodology

1. LLM generates synthetic arguments, whose embeddings act as centroids later at the clustering stage
2. A smaller LM is trained to produce embeddings that aid (i) a favor/against classification objective and (ii) a relevant/irrelevant contrastive objective



3. Based on the classifier confidence in the final layer, bound each cluster to discard irrelevant arguments
4. Sample from the resulting clusters

Results

k	ndcg@k	precision@k
4	0.694	0.692
8	0.679	0.671
16	0.670	0.660
20	0.677	0.673

Table 1: Relevance scores averaged across the three scenarios

k	alpha_ndcg@k	kl_divergence@k
4	0.625	0.151
8	0.618	0.134
16	0.626	0.100
20	0.638	0.091

Table 2: Diversity scores averaged across the three scenarios

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.RS-2019-II190075 Artificial Intelligence Graduate School Program (KAIST)) and by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS2023-00252535).