

Exploring BERT Synonyms and Quality Prediction for Argument Retrieval

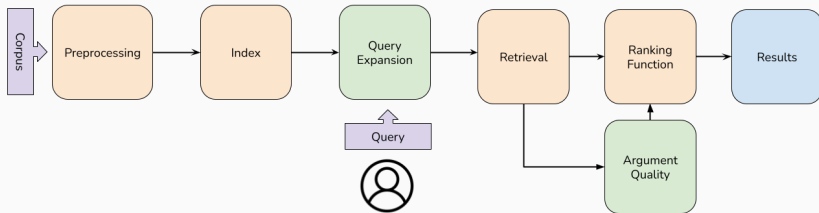
Touché @ CLEF 2021 - Task 1

Team Yeagerists

Tommaso Green¹, Luca Moroldo¹, Alberto Valente¹

¹University of Padua

Overview of our approach



- Preprocessing + Indexing
- Query Expansion
- Retrieval
- Argument Quality Prediction

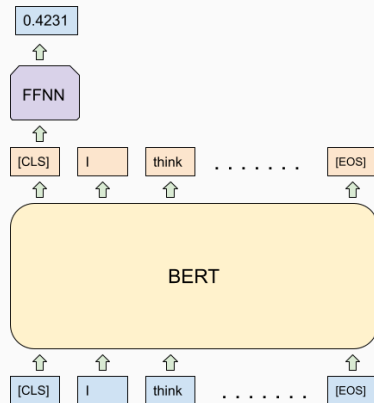
- Cleanup: deleted duplicate arguments or with empty body.
- Indexing: Lucene StandardTokenizer + LowerCaseFilter.
- Stopwords were not removed.

- We used a BERT Synonym substitution approach:
 1. Identify nouns, adjective, past participles via POS-tagging and [MASK] them.
 2. For every masked token, generate at least 10 substitute candidates using BERT.
 3. Keep candidates having a cosine similarity greater than 85% with the embedding of the original token.
 4. Generate new queries by taking a subset of the cartesian product of all possible substitutions for each token.

- Query parsing: Lucene MultiFieldQueryParser.
- Retrieval Model: LMDirichletSimilarity, with $\mu = 2000$, returning at most 100 documents.
- With query expansion, results for each expanded query are merged and duplicates are removed.
- Top-5 arguments are passed to Argument Quality Module.

Argument Quality Prediction

- Argument Quality Predictor using BERT features: [CLS] token passed to a FFNN with MSE loss function.
- Trained and tested BERT, DistilBERT, RoBERTa, Albert on the Webis Argument Quality Corpus 2020 using an 80-10-10 train-val-test split.
- Re-ranking of arguments by combining relevance and quality scores in different ways (normalization or using sigmoid function).



Results

- Slightly below the baseline on Touché 2021 for relevance, improvement in terms of argument quality.
- Query expansion significantly downgraded the performance, providing wrong but interesting substitutions, e.g. abortion \mapsto divorce.
- BERT and DistilBERT were the most promising quality predictors, but improvement over the baseline was not significant.

Results

- Slightly below the baseline on Touché 2021 for relevance, improvement in terms of argument quality.
- Query expansion significantly downgraded the performance, providing wrong but interesting substitutions, e.g. abortion \mapsto divorce.
- BERT and DistilBERT were the most promising quality predictors, but improvement over the baseline was not significant.

Thank you for watching!