Synonym Substitution Effects on Recall@k in Hybrid Sparse-Dense Retrievers Versus Pure Dense Models on TREC Deep Learning

Assignee Research

June 12, 2026

Abstract

Large-scale text retrieval technology has been widely used in various practical business scenarios. This paper presents our systems for the TREC 2022 Deep Learning Track. We explain the hybrid text retrieval and multi-stage text ranking method adopted in our solution. The retrieval stage combined the two structures of traditional sparse retrieval and neural dense retrieval. In the ranking stage, in addition to the full interaction-based ranking model built on large pre-trained language model, we also proposes a lightweight sub-ranking module to further enhance the final text ranking performanc

1 Introduction

This paper examines: Hybrid Retrieval and Multi-stage Text Ranking Solution at TREC 2022 Deep Learning Track. Research question: How does synonym substitution impact the recall@k metrics of hybrid sparse-dense retrieval systems on TREC Deep Learning datasets compared to pure dense retrievers?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.0/10.

3 Results

16 papers retrieved. 13 claims extracted; 9 independently verified. Quality review score: 7.0/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The submission 'pass3' achieved the best results on the test set for the Passage ranking task among the three submitted	✓	0.17
The submission 'doc3' obtained the highest score for the Document ranking task by training the model based on passage da	✓	0.26
The 'doc1' run achieved an NDCG@10 of 0.7643 on the Dev set and 0.4908 on the Eval set for the TREC 2022 document datase	✓	0.17
The 'doc2' run achieved an NDCG@10 of 0.7631 on the Dev set and 0.4555 on the Eval set for the TREC 2022 document datase	✓	0.17
The 'doc3' run achieved an NDCG@10 of 0.7533 on the Eval set for the TREC 2022 document dataset.	✓	0.17
BM25 serves as a strong baseline for both Passage ranking and Document ranking tasks.	×	0.15
Doc2query improves upon the BM25 baseline performance.	×	0.05
The proposed method consists of three stages: retrieval, ranking, and HLATR re-ranking.	✓	0.17
For the Document ranking task, documents are first divided into passages for retrieval and ranking, with the highest-sco	×	0.12
The retrieval stage utilizes a weighted ensemble of BM25, Doc2query, and SPLADE.	✓	0.18
BM25 is limited by text matching between query and document and lacks semantic expansion ability.	✓	0.17
Doc2query and SPLADE extend semantics from different perspectives to improve sparse retrieval.	✓	0.17
The dense retrieval stage employs a two-tower structure due to computational efficiency constraints.	×	0.15

References

- `http://arxiv.org/abs/2205.11245v5`

- http://arxiv.org/abs/2205.02870v2
- http://arxiv.org/abs/2308.12039v1