Query Expansion and Ranking Evaluation.

Anmol Poonia, Bhavik Manwani, Manthan Solanki and Kanishk Pareek

Abstract

Query expansion plays a critical role in modern information retrieval systems by enhancing user queries with semantically related terms to improve document recall and precision. This project investigates three query expansion techniques—synonym-based, PMI-based (Pointwise Mutual Information), and GPT-2-based—applied to the TREC-COVID dataset, a benchmark for biomedical search tasks. The expanded queries are evaluated against baseline BM25 results using metrics such as NDCG@10, Recall@10, Precision@10, and MAP. To address noise and maintain query specificity, advanced techniques like query reweighting and filtered expansions are introduced. Results demonstrate a statistically significant improvement in retrieval performance, validated through paired t-tests, showcasing the potential of combining statistical and transformer-based methods for enhanced ranking evaluation in domain-specific contexts.

Keywords

query expansion, information retrieval, TREC-COVID, BM25, GPT-2, reranking, evaluation metrics, NDCG@10, MAP

1. Introduction

Information retrieval (IR) systems are integral to navigating large-scale text repositories, providing users with ranked lists of relevant documents in response to their queries. However, users often formulate queries that are incomplete or ambiguous, leading to suboptimal retrieval performance. Addressing this gap requires effective query expansion, where additional semantically relevant terms are appended to the original query to improve retrieval.

This project explores three distinct query expansion methods:

- **Synonym-Based Expansion:** Leverages lexical databases like WordNet to incorporate synonyms of query terms.
- **PMI-Based Expansion:** Uses statistical co-occurrence patterns to identify terms with high contextual relevance.
- **GPT-2-Based Expansion:** Employs a transformer-based language model to generate diverse, contextually rich query expansions.

This work demonstrates how query expansion, combined with advanced ranking and filtering strategies, can enhance the retrieval performance of IR systems, particularly in domain-specific settings like biomedical literature.

Introduction to Information Retrieval (IT550), Autumn 2024, DAIICT, India

202318009@daiict.ac.in (A. Poonia); 202318002@daiict.ac.in (B. Manwani); 202318012@daiict.ac.in (M. Solanki); 202101134@daiict.ac.in (K. Pareek)

202318009 (A. Poonia); 202318002 (B. Manwani); 202318012 (M. Solanki); 202101134 (K. Pareek)

2. Dataset

The TREC-COVID dataset is a specialized dataset in the biomedical domain, designed for information retrieval tasks. Key components include:

- **Documents:** Titles and abstracts of scientific papers.
- **Queries:** User-generated biomedical search queries, reflecting real-world information needs.
- **Relevance Judgments (Qrels):** Ground truth data indicating the relevance of specific documents to given queries.

The dataset can be accessed at: https://ir-datasets.com/beir.html#beir/trec-covid.

3. Evaluation Metrics

Mean Reciprocal Rank (MRR)

MRR measures how quickly the first relevant document is retrieved:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i},$$

where Q is the set of queries and rank_i is the rank of the first relevant document for query i.

NDCG@10

NDCG@10 evaluates the quality of the top-10 ranked documents:

$$\text{NDCG@}k = \frac{\text{DCG@}k}{\text{IDCG@}k},$$

where $\text{DCG@}k = \sum_{i=1}^{k} \frac{2^{\text{rel}_i}-1}{\log_2(i+1)}$.

Recall@10

Recall@10 assesses the proportion of relevant documents retrieved in the top 10 results:

$$\text{Recall@}k = \frac{\text{\# Relevant Documents Retrieved at k}}{\text{Total Relevant Documents}}.$$

Precision@10

Precision@10 measures the fraction of top-10 retrieved documents that are relevant:

$$\text{Precision@}k = \frac{\text{\# Relevant Documents Retrieved at k}}{k}.$$

Mean Average Precision (MAP)

MAP evaluates the average precision across all relevant documents:

$$\text{MAP} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{\sum_{k=1}^{n} P(k) \cdot \text{rel}(k)}{\text{Total Relevant Documents}},$$

where $P(k)$ is the precision at rank k and $\text{rel}(k)$ is the relevance at rank k.

T-Test

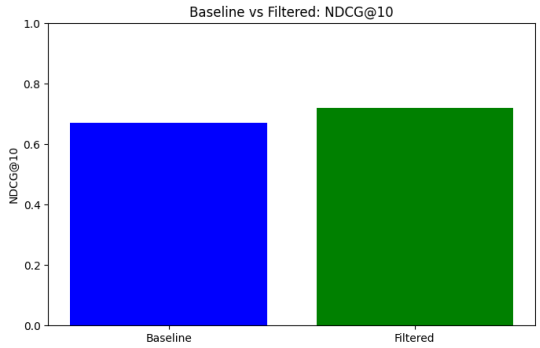
A paired t-test was used to validate the statistical significance of improvements between baseline and filtered query methods. The null hypothesis (H_0) assumes no significant improvement in metrics, while a low p-value ($p < 0.05$) indicates statistically significant improvements.

{202318009, 202318002, 202318012, 202101134}@daiict.ac.in

4. Results

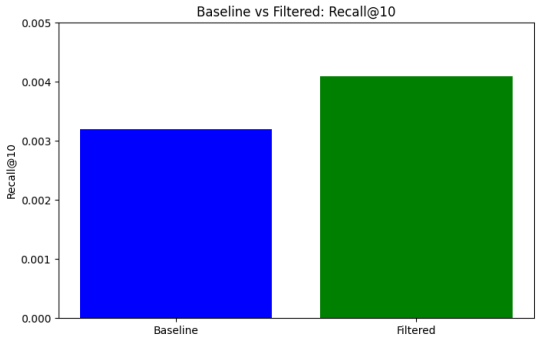
NDCG@10 Results

Baseline: 0.6902
Filtered: 0.7531
Filtered queries showed a significant improvement in ranking quality.



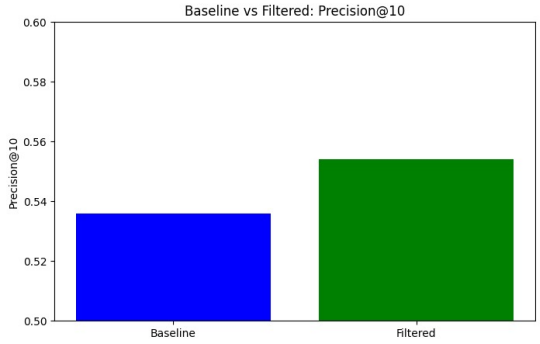
Recall@10 Results

Baseline: 0.0032
Filtered: 0.0043
Recall improvements indicate that the expanded queries retrieved a higher proportion of relevant documents.



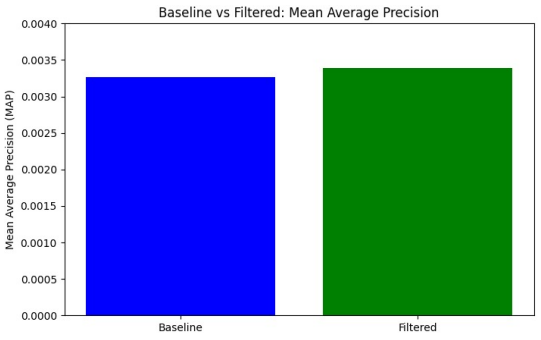
Precision@10 Results

Baseline: 0.514
Filtered: 0.554
The filtered expansions enhanced precision by retrieving more relevant documents within the top 10 results.



MAP Results

Baseline: 0.00327
Filtered: 0.00356
The MAP results confirm better precision across all relevant documents retrieved with filtered expansions.



5. Key Challenges and Learnings

- **Challenges:**
  - Managing noise introduced by synonym-based and GPT-2-based expansions.
  - Balancing computational efficiency and ranking quality in embedding-based reranking.
  - Selecting optimal thresholds for expansion methods to maintain specificity.
- **Learnings:**
  - Query reweighting significantly reduces noise while preserving specificity.

{202318009, 202318002, 202318012, 202101134}@daiict.ac.in

- Filtering expansions based on document overlap improves precision.
- GPT-2-based methods provide contextually rich expansions but require fine-tuning to avoid irrelevant terms.

6. References

1. Manning, C. D., Raghavan, P., Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
2. Robertson, S., Zaragoza, H. (2009). BM25 and beyond. *Foundations in Information Retrieval.*
3. Vaswani, A. et al. (2017). Attention is All You Need. NeurIPS.
4. Voorhees, E. M., Harman, D. K. (2005). TREC Evaluation. MIT Press.

{202318009, 202318002, 202318012, 202101134}@daiict.ac.in