Re-Ranking Dense Retrieval

Jake Norton
University of Otago
Dunedin, Otago, New Zealand
norja159@student.otago.ac.nz

ABSTRACT

This paper explores advanced retrieval techniques aimed at enhancing the efficiency and effectiveness of information retrieval systems through use of graph-based and latent feature methodologies. We analyze three distinct approaches: Latent Approximate Document Retrieval (LADR), which utilizes latent features to optimize document retrieval; adaptive re-ranking with a corpus graph (Gar), which dynamically adjusts document rankings within a corpus graph to improve relevance; and adhoc retrieval through traversal of a query-document graph, which employs direct traversal methods to optimize query-specific document retrieval. I propose to build ontop of these by trying the query-document mapping but using the LADR methodologies. Try other LADR with smarter exploration ideas.

CCS CONCEPTS

• **Information systems → Novelty in information retrieval**.

KEYWORDS

dense retrieval, approximate k nearest neighbour, adaptive re-ranking, neural re-ranking, clustering hypothesis

1 INTRODUCTION

2 LEXICALLY-ACCELERATED DENSE RETRIEVAL RESEARCH

2.1 Research Questions

- **R1:** How does LADR compare to other approximation techniques in terms of effectiveness and efficiency, and what computational overheads does it entail?
- **R2:** Is LADR applicable across different single-representation dense retrieval models? and how do its parameters, like the number of neighbors k K, influence its performance? Is an exact nearest neighbor graph necessary for LADR's effectiveness, and what are the trade-offs between proactive and adaptive LADR approaches regarding precision, recall, and latency?

2.2 Main Contributions

LADR is a novel retrieval technique that establishes a new pareto frontier for efficiency/effectiveness

Provides a thorough evaluation of LADR's effectiveness and efficiency, demonstrating its superiority or comparability to other well-known approximation techniques in various scenarios.

LADR's applicability across a range of single-representation dense retrieval models, showing its versatility and adaptability in different contexts.

Use of proactive or adaptive LADR in different scenarios and depending on specific system contraints

The paper might explore the practical requirements and trade-offs involved in implementing LADR, such as the necessity (or not) of an exact nearest neighbor graph, and the balance between proactive and adaptive strategies in terms of precision, recall, and latency.

3 ADAPTIVE RE-RANKING WITH A CORPUS GRAPH

3.1 Research Questions

3.1.1 Impact of Gar on Retrieval Effectiveness. What is the impact of Gar on retrieval effectiveness compared to typical re-ranking and state-of-the-art neural IR systems?

3.1.2 Computational Overheads and Parameter Sensitivity. What are the computational overheads introduced by Gar, and how sensitive is Gar to the parameters it introduces, such as the number of neighbors k K and the batch size b B?

3.2 Main Contributions

Innovative Re-ranking Approach: Introducing Gar, a novel adaptive re-ranking method that uses a corpus graph to enhance retrieval effectiveness. This method represents a significant shift from traditional re-ranking techniques by integrating graph-based structures within the retrieval process.

Empirical Validation: Providing empirical evidence of Gar's effectiveness through rigorous testing on standardized datasets such as the TREC Deep Learning 2019 and 2020 test collections. These results highlight the improvement Gar offers over traditional re-ranking methods and state-of-the-art neural information retrieval (IR) systems.

Parameter Sensitivity Analysis: A detailed exploration of how Gar's performance is affected by various parameters, such as the number of neighbors in the corpus graph and batch size. This contribution is crucial for understanding the adaptability and optimization of Gar in different retrieval environments.

Computational Efficiency: Assessing the computational overheads introduced by Gar, which is vital for practical applications, particularly in scenarios where computational resources are a limiting factor.

Comprehensive Evaluation Metrics: Utilizing a variety of metrics, including nDCG, MAP, and Recall, to evaluate Gar's performance comprehensively. This thorough evaluation helps in understanding the strengths and limitations of Gar across different aspects of retrieval performance.

Overall, the main contribution of this paper is the development and validation of a novel adaptive re-ranking method that leverages a corpus graph for enhanced document retrieval, providing insights into its efficiency, effectiveness, and parameter dependencies, which could influence future research and applications in the field of information retrieval.

4 EFFECTIVE ADHOC RETRIEVAL THROUGH TRAVERSAL OF A QUERY-DOCUMENT GRAPH

4.1 Research Question

Traversal Mechanisms: What specific traversal mechanisms in the query-document graph yield the best adhoc retrieval performance?

Impact on Retrieval Metrics: How does traversing a query-document graph affect key retrieval metrics such as precision, recall, and nDCG compared to standard retrieval models?

Scalability and Efficiency: How scalable and efficient is the traversal process, especially when handling large-scale datasets?

4.2 Main Contributions

Novel Retrieval Framework: Introduction of a novel graph-based retrieval framework that leverages query-document relationships through dynamic traversal methods. This framework provides a new way to conceptualize and implement adhoc retrieval tasks, which could potentially offer improvements over linear or list-based retrieval models.

Performance Enhancement: Demonstrating through empirical evidence that the graph traversal method can enhance retrieval effectiveness by dynamically adjusting to the relevance signals found within the graph structure. This could be particularly advantageous for complex queries where traditional methods struggle.

Comprehensive Evaluation: Conducting a comprehensive evaluation using standard IR benchmarks to validate the effectiveness and efficiency of the proposed method. This might include comparisons with baseline models such as BM25 or newer neural retrieval models to contextualize the performance improvements.

Scalability Analysis: Analysis of the scalability and computational efficiency of the graph traversal method, providing insights into its applicability in real-world scenarios where large document collections are common.

Theoretical Insights and Practical Implications: Offering theoretical insights into the behavior of retrieval systems when modeled as graphs, and discussing practical implications for designing more robust and responsive retrieval systems.

5 RELATIONS BETWEEN THE PAPERS

Graph-Based and Advanced Retrieval Techniques Each paper explores different facets of improving retrieval effectiveness and efficiency by introducing novel structures or frameworks:

LADR explores the use of latent features and approximation techniques to improve the retrieval process.

Adaptive Re-ranking with a Corpus Graph (Gar) introduces a graph-based method to dynamically adjust rankings based on a corpus graph, enhancing the relevance of retrieved documents.

Adhoc Retrieval through Traversal of a Query-Document Graph investigates the use of a graph that includes both queries and documents to potentially improve retrieval outcomes through strategic traversal methods. Methodological Innovations

All three papers push the boundaries of conventional retrieval methods:

LADR challenges traditional approximation techniques with a potentially more effective and efficient method.

Gar employs a corpus graph, which adapatively re-ranks documents based on additional contextual relationships within the corpus.

Traversal of a Query-Document Graph suggests that direct interaction between queries and documents via graph traversal could yield better retrieval results than isolated query-document evaluations.

Enhanced Retrieval Metrics Each study focuses on evaluating and improving key retrieval metrics like precision, recall, and nDCG:

LADR is assessed on its effectiveness and efficiency, likely impacting metrics such as nDCG and precision.

Gar is evaluated through its impact on retrieval effectiveness, comparing its performance with traditional re-ranking and advanced neural IR systems.

Traversal of a Query-Document Graph addresses the impact on retrieval metrics through unique graph traversal techniques, offering potential improvements in precision and recall.

Potential for Future Research Integration The insights and methods from each paper could be integrated into a comprehensive retrieval system:

Techniques from LADR could be combined with the graph-based approaches of Gar and Traversal of a Query-Document Graph to develop a highly sophisticated IR system that uses both latent features and graph structures. The adaptive re-ranking methods of Gar could benefit from the dynamic traversal methods explored in the Traversal of a Query-Document Graph, potentially enhancing the adaptiveness and contextual awareness of the re-ranking process.

6 RESEARCH Q1

6.1 Experiment

6.2 Hypothesis

7 RESEARCH Q2

7.1 Experiment

7.2 Hypothesis

8 CONCLUSION

These papers contribute to the evolving landscape of information retrieval by demonstrating how advanced data structures like graphs and innovative algorithms can be leveraged to solve complex retrieval challenges. By exploring different aspects of IR enhancements—whether through latent features, adaptive re-ranking, or traversal mechanisms—they collectively push the field toward more nuanced and effective retrieval solutions, offering a foundation for future research that could integrate these varied approaches into a unified framework.

REFERENCES