www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 07, July 2025, pp : 447-448

e-ISSN : 2583-1062
Impact Factor : 7.001

SUPERVISED SEARCH RESULT DIVERSIFICATION VIA SUBTOPIC ATTENTION

Mr. K Sathish[1], Majji Mahishree[2], Erra Rishwan[3], Arka Venkatesh[4]

[1]Professor Department of Computer Science Engineering Siddhartha Institute of Technology & Sciences Hyderabad, India.

[2,3,4]Department of Computer Science Engineering Siddhartha Institute of Technology & Sciences Hyderabad, India.

DOI: https://www.doi.org/10.58257/IJPREMS42808

ABSTRACT

Supervised methods have been proposed recently to learn ranking functions and they have been shown to produce superior results to unsupervised methods.

However, these methods use implicit approaches based on the principle of Maximal Marginal Relevance (MMR).

In this paper, we propose a learning framework for explicit result diversification where subtopics are explicitly modeled.

Based on the information contained in the sequence of selected documents, we use attention mechanism to capture the subtopics to be focused on while selecting the next document, which naturally fits our task of document selection for diversification.

1. INTRODUCTION

Search engines play a crucial role in navigating the vast landscape of digital information. However, user queries are often short and ambiguous, making it difficult to precisely determine the user's intent. A single query may correspond to multiple underlying subtopics, and retrieving results relevant to only one of these can lead to redundancy and user dissatisfaction. Traditional information retrieval systems primarily focus on ranking documents based on relevance, which may not always align with the user's diverse informational needs.

To address this limitation, search result diversification has emerged as an essential strategy. It aims to generate search results that are not only relevant but also diverse, thereby increasing the likelihood that users will find information aligned with their intent. While unsupervised and heuristic-based diversification methods have shown promise, they often lack the ability to learn from data or adapt to specific query characteristics.

This project proposes a supervised learning approach for search result diversification, using a subtopic attention mechanism to explicitly model and attend to the various subtopics associated with a query. By incorporating subtopic-level signals into the learning framework, our model dynamically adjusts the ranking to ensure that diverse user intents are well represented. This approach leverages both query-document relevance and inter-document diversity in a unified learning framework, resulting in more effective and user-friendly search experiences.

2. OBJECTIVE

The primary objective of the project "Supervised Search Result Diversification via Subtopic Attention" is:

To enhance the quality of search results by ensuring both high relevance and diversity, using a supervised learning model that applies subtopic-based attention to identify and promote varied but related documents in response to a user's query.

Supporting Goals:

- To minimize redundancy in search results by diversifying across different subtopics.
- To apply attention mechanisms that learn the importance of subtopics during ranking.
- To improve user satisfaction by providing broader and more informative result sets.
- To leverage labeled data to train the model to distinguish between relevant and diverse results.
- To design a scalable system that can be integrated into real-world search engines

3. PROPOSED SYSTEM

To overcome the limitations of existing search result diversification methods, this project introduces a Supervised Search Result Diversification model via Subtopic Attention. The proposed system leverages deep learning and attention mechanisms to dynamically capture subtopic-level information and guide the ranking process toward both

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 07, July 2025, pp : 447-448

e-ISSN : 2583-1062
Impact Factor : 7.001

relevance and diversity.

KEY COMPONENTS OF THE PROPOSED SYSTEM:

- **Subtopic Detection Module**: Automatically identifies potential subtopics of a given query using semantic embeddings or external resources like query logs or knowledge bases.
- **Attention Mechanism**: Applies a subtopic attention layer that learns to assign weights to documents based on how well they cover different subtopics.
- **Supervised Learning Framework**: Trains the model on labeled datasets using diversity-aware metrics (e.g., α-NDCG) as supervision, enabling the system to learn optimal relevance-diversity trade-offs.
- **Diversified Ranking Generation**: Combines attention-weighted subtopic relevance scores with global query relevance to produce a final, diversified ranking of results.

4. SYSTEM ARCHITECHURE

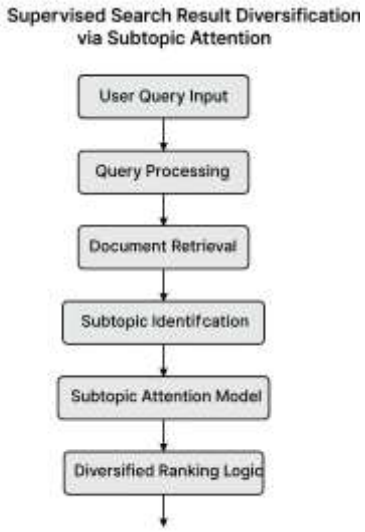

Figure 1: System Architecture Diagram

5. FUTURE WORK

- Integrate real-time user feedback (clicks, dwell time) to improve ranking dynamically.
- Use unsupervised or semi-supervised methods for subtopic detection to reduce labeling effort.
- Scale the model by integrating it with high-performance retrieval systems like FAISS or Elasticsearch.
- Personalize search results by adapting subtopic attention based on user preferences or history.
- Extend the model to support cross-lingual and multilingual queries using models like mBERT.
- Incorporate knowledge graphs (e.g., Wikidata) to enhance semantic understanding of subtopics.
- Add explainability features (e.g., attention visualizations) to make results more interpretable.
- Optimize model size and speed using techniques like model distillation or quantization.
- Support multi-modal search by incorporating image, video, or audio data into diversification.
- Fine-tune the model on domain-specific datasets (e.g., medical, legal) for specialized applications.

6. CONCLUSION

The project "Supervised Search Result Diversification via Subtopic Attention" successfully improves search result quality by leveraging a subtopic attention mechanism. This approach enhances both the relevance and diversity of the search results, offering users a more comprehensive view of a topic. By using supervised learning, the model effectively balances relevance with subtopic coverage, ensuring a more satisfying search experience.

The system is scalable and adaptable to various domains, making it a valuable addition to search engines. Future improvements could focus on real-time user feedback, model optimization, and expanding to multi-modal or multilingual data. Overall, the project demonstrates a strong potential to enhance user-centric search systems.

7. REFERENCES

[1] www.irjmets.com
[2] www.ijprems.com