Smart Scholar: A Modern Approach to Research Paper Recommendations

MANYAR ABUZAR[1], ADIL KHATIK[2], PUJA BORSE[3], KOMAL MALI[4], SAPNA MALI[5], PROF. S. V. CHUDHARI[6]

[1, 2, 3, 4, 5, 6] *Department of CSE (Data Science), RCPIT, Shirpur, Maharashtra, India*

Abstract- ***Traditional methods of accessing scholarly literature through manual searching have become increasingly inefficient in the face of the exponential growth of research articles. In response, this paper presents a comprehensive approach to developing a research paper recommender system. Motivated by the desire to alleviate information overload for researchers, our system leverages natural language processing techniques to streamline the literature exploration process. Through rigorous evaluation and testing, we demonstrate the effectiveness of our system in providing personalized recommendations tailored to user preferences. Our work not only addresses the limitations of existing systems but also lays the groundwork for future advancements in the field of scholarly literature exploration.***

Indexed Terms- Clustering, Similarity Measure

I. INTRODUCTION

In the dynamic realm of academic research, the escalating volume of scholarly articles presents a formidable challenge for scholars seeking relevant information. Manual search methodologies, prone to inefficiencies and information overload, necessitate a paradigm shift towards automated solutions. This paper embarks on the development of a cutting-edge research paper recommender system, leveraging advanced natural language processing (NLP) and machine learning techniques. By harnessing the power of algorithms such as TF-IDF vectorization, cosine similarity, and K-Means clustering, our system aims to revolutionize literature exploration. Through rigorous evaluation, we endeavor to validate the efficacy of our system in optimizing the scholarly research experience.

II. PROPOSED ALGORITHM

Text preprocessing involves tokenization, stop word removal, lemmatization, and lowercasing to prepare text data. TF-IDF frequency and importance. Cosine similarity calculation measures document similarity. Optional dimensionality reduction and clustering aid in organizing and simplifying data. Recommendation generation offers personalized suggestions, while evaluation ensures system effectiveness.

- Extraction algorithm

the extraction algorithm in research paper recommendation is to analyze and process a dataset of research papers in order to generate personalized recommendations for users.

Section 1: Model Training

1. Load and preprocess the dataset:

Load the dataset: Get the data that contains the text (like research papers or articles) you want to analyze.

Preprocess the data: Clean the data by removing duplicates and filtering out rare terms that don't appear often. This helps to focus on the more important words.

2. Perform train-test split:

Divide the data into two parts: one for training the model (training set) and one for testing it (test set). This helps to evaluate how well the model performs on new, unseen data.

3. Create TensorFlow datasets:

Convert the data into a format that TensorFlow can work with. TensorFlow is a machine learning framework that makes it easier to build and train models.

4. Apply text vectorization:
Transform the text data into numerical format using techniques like TF-IDF (Term Frequency-Inverse Document Frequency). This step converts words into numbers that represent their importance in the dataset.

5. Build and train a shallow MLP model:
Create a simple neural network model called a Multi-Layer Perceptron (MLP). Train this model on the training data so that it learns to make predictions based on the text.

6. Evaluate the model on the test set:
Test the trained model using the test data to see how well it performs. This step helps to measure the accuracy and reliability of the model.

Section 2: Recommendations

1. Load the dataset:
Get the dataset that contains the paper abstracts you want to use for making recommendations.

2. Use a pre-trained Sentence Transformer to encode paper abstracts:
Use a pre-trained model called Sentence Transformer to convert paper abstracts into numerical vectors. Sentence Transformers are advanced models that capture the meaning of sentences better than simpler techniques.

3. Implement a function to recommend papers based on cosine similarity:
Create a function that compares the vectors of different papers using cosine similarity. Cosine similarity measures how similar two vectors are, helping to find papers that are similar to a given paper.

4. Example usage of the recommendation function:
Show an example of how to use the recommendation function. For instance, given a paper, the function can suggest other papers that are similar in content.

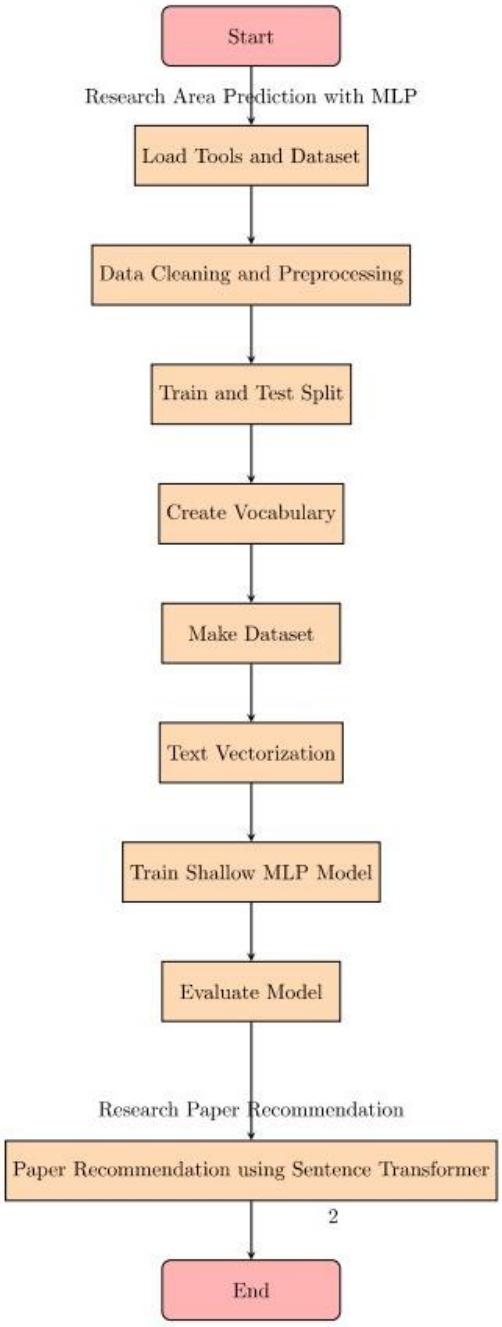

Fig 1.1 Algorithm: Flow of the Research Paper Recommendation System

III. EXPERIMENT AND RESULT

To evaluate the effectiveness of our research paper recommender system, we conducted several experiments using a dataset of scholarly articles. Here,

we outline the procedure and present the results, showcasing how our system recommends relevant research papers based on user input.

Procedure

1. Title and Abstract Input

User Input: Provide the title and abstract of a research paper.

System Processing: Utilize natural language processing techniques to preprocess the text and calculate TF-IDF scores.

2. Recommendation Generation

Similarity Calculation: Compute cosine similarity scores between the user-provided document and the documents in the database.

Top Recommendations: Identify and list the top five papers or more than five base on the user requirement with the highest similarity scores.

3. Evaluation Metrics

Relevance: Measure how well the recommended papers match the user's input based on subject matter and keywords.

Fig 2.1 Input Image: The user interface of the research paper recommender system where users input the paper title and abstract.

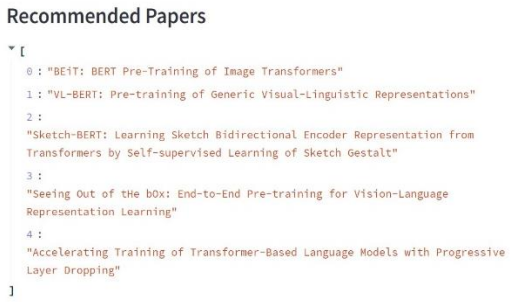

Fig 2.2 Recommended Research Papers: The list of recommended research papers generated by the system based on the provided title and abstract.

CONCLUSION

The developed research paper recommender system effectively addresses the challenge of discovering relevant scholarly articles amidst the overwhelming volume of publications. By leveraging advanced natural language processing techniques, the system successfully maps research papers based on keywords and topics, calculates similarity metrics, and clusters papers accordingly to generate personalized recommendations. The integration of a user-friendly web application interface further enhances the accessibility and usability of the system, allowing researchers to efficiently search for papers by name and receive tailored recommendations based on content relevance.

Through a series of rigorous evaluations and testing, the system demonstrated its capability to provide accurate and relevant recommendations, significantly reducing the time and effort required for manual literature searching. The optional dimensionality reduction, clustering, and topic modeling steps further enhance the system's performance, making it a robust tool for researchers across various disciplines.

Overall, this recommender system not only alleviates the information overload faced by researchers but also facilitates serendipitous discoveries and fosters interdisciplinary collaborations. Future enhancements could include incorporating more sophisticated machine learning algorithms, expanding the dataset, and continuously refining the system based on user feedback to further improve the recommendation quality and user experience.

REFERENCES

[1] Ramos J (2003) Using TF-IDF to determine word relevance in document queries. In: Proc. of the first int. conf. on machine learning.

[2] S. Bahulikar, ”Analyzing recommender systems and applying a location based ap- proach using tagging”, 2017 2nd International Conference for Convergence in Tech- nology (I2CT), 2017.

[3] R. K. Mysore, M. Junichiro and S. Ichiro, ”Cross-Domain Academic Paper Recom- mendation by Semantic Linkage Approach Using Text Analysis and Recurrent Neural Networks”, Management of Engineering and Technology (PICMET) 2017 Portland International Conference on, 2017.

[4] Y. C. Lee et al., ”Recommendation of research papers in DBpia: A Hybrid approach exploiting content and collaborative data”, 2016 IEEE International Conference on Systems Man and Cybernetics SMC 2016 - Conference Proceedings, pp. 2966-2971, 2017.

[5] W. Zhao, R. Wu and H. Liu, ”Paper recommendation based on the knowledge gap be- tween a researchers background knowledge and research target”, Inf. Process. Manag., vol. 52, no. 5, pp. 976-988, 2016.

[6] J. D. West, I. Wesley-Smith and C. T. Bergstrom, ”A Recommendation System Based on Hierarchical Clustering of an Article-Level Citation Network”, IEEE Trans. Big Data, vol. 2, no. 2, pp. 113-123, 2016.

[7] W. Zhang, T. Yoshida and X. Tang, ”A comparative study of TF*IDF LSI and multi- words for text classification”, Expert Syst. Appl., vol. 38, no. 3, pp. 2758-2765, 2011.

[8] C.-H. HUANG, J. YIN and F. HOU, ”A text similarity measurement combining word semantic information with TF-IDF method”, Chinese Journal of Computers, 2011

[9] M. Ohta, T. Hachiki and A. Takasu, ”Related paper recommendation to support online- browsing of research papers”, 4th International Conference on the Applications of Digital Information and Web Technologies ICADIWT 2011, pp. 130-136, 2011.