IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Approaching Text Summarization Using ML And DNN

Prof Priyadarshini M, Pavan R, Punith K M, Naveen Kumar G S and Rajala Chirra Reddy

[,2]Assistant Professor, Department of Computer Science and Engineering, Cambridge Institute of Technology (CITech), Bengaluru, India

[1 ,3,4 5]Student, Department of Computer Science and Engineering, CITech, Bengaluru, India

Abstract: Extractive text summarization using Latent Semantic Analysis (LSA) involves analyzing the underlying structure of a document by creating a matrix of term-document relationships. The TFIDF (Term Frequency-Inverse Document Frequency) vectorizer is employed to highlight important words in the document, assigning weights based on their frequency and uniqueness. Machine learning algorithms leverage these vectorized representations to identify and extract key sentences or phrases, forming the basis of the summary. Additionally, Deep Neural Networks (DNN) come into play, employing intricate layers of interconnected nodes to learn and understand complex patterns within the text. The DNN further refines the summarization process, enhancing the model's ability to capture nuanced relationships and context. This fusion of traditional ML and DNN approaches results in a powerful summarization system capable of distilling large volumes of information into concise, informative abstracts.Keywords—CNN, maximum pooling layers, dropout layers, softmax activation

Keywords: LSA, DNN, TF-IDF Vectorizer

INTRODUCTION`

Extractive text summarization is a fascinating field within natural language processing (NLP) that aims to condense large volumes of text while retaining the essential information. One of the prominent algorithms used for this purpose is Latent Semantic Analysis (LSA). LSA leverages mathematical and statistical techniques to identify patterns and relationships within a corpus, enabling the extraction of key content without losing context.

In conjunction with LSA, another crucial component in extractive summarization is the TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer. TF-IDF assigns weights to words based on their frequency in a document relative to their frequency across the entire corpus. This allows the algorithm to discern the significance of each word, aiding in the extraction of salient information

Moreover, modern approaches incorporate Machine Learning (ML) and Deep Neural Networks (DNN) to enhance the summarization process. These advanced techniques contribute to more accurate and contextually relevant summaries, pushing the boundaries of what traditional methods can achieve.

phrases, forming the summary's foundation. Additionally, Deep Neural Networks (DNN) contribute by employing intricate layers of interconnected nodes to comprehend complex patterns within the text. The DNN further refines the summarization process, augmenting the model's capacity to capture nuanced relationships and context.

Related Work:

In the realm of extractive text summarization, Latent Semantic Analysis (LSA) serves as a foundational technique, enabling the analysis of document structures through the creation of a term-document relationship matrix. The utilization of TF-IDF vectorization further enhances this process by emphasizing important words based on their frequency and uniqueness, thereby facilitating the identification of key sentences or phrases by machine learning algorithms.

Furthermore, Deep Neural Networks (DNNs) play a pivotal role in advancing extractive summarization techniques. By employing sophisticated layers of interconnected nodes, DNNs are adept at learning and comprehending intricate patterns within textual data. This capability enables DNNs to refine the summarization process, improving the model's ability to capture nuanced relationships and contextual information.

The integration of traditional machine learning (ML) approaches with DNNs results in a synergistic effect, leading to the development of robust summarization systems capable of distilling vast amounts of information into concise and informative abstracts.

Related works in this field may include research on alternative summarization techniques such as graph-based methods, reinforcement learning approaches, and hybrid models that combine multiple strategies to achieve optimal summarization performance. Additionally, studies focusing on the evaluation metrics and benchmarks for proposed a superpixel method combining SVM for tumor segmentation, while Md. Abu Bakr Siddique et al. [7] used MRMR, ERT, and SVM for segmentation-based automatic detection.

Methodology:

Extractive text summarization using Latent Semantic Analysis (LSA) involves several key steps:

1. **Matrix Construction**: The first step is to analyze the underlying structure of the document by constructing a matrix of term-document relationships. Each row of the matrix represents a term, and each column represents a document. The entries in the matrix indicate the frequency or importance of each term in each document.

2. **TF-IDF Vectorization**: The Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer is then applied to the matrix to highlight important words in the document. This process assigns weights to terms based on their frequency within the document and their uniqueness across the corpus. Terms that are common in the document but rare in the corpus receive higher weights, indicating their importance.

3. **Machine Learning Algorithms**: Machine learning algorithms leverage these vectorized representations to identify and extract key sentences or phrases. Techniques such as clustering, classification, or ranking algorithms can be applied to identify the most salient information in the document. These algorithms form the basis of the extractive summarization process by selecting sentences or phrases that best represent the main ideas of the document.

4. **Deep Neural Networks (DNN)**: Additionally, Deep Neural Networks (DNNs) are employed to further enhance the summarization process. DNNs utilize intricate layers of interconnected nodes to learn and understand complex patterns within the text. By capturing nuanced relationships and context, DNNs refine the summarization process, improving the model's ability to generate concise and informative summaries.

The fusion of traditional machine learning techniques with DNN approaches results in a powerful summarization system capable of distilling large volumes of information into succinct and informative abstracts. This methodology enables the system to analyze the document's structure, identify key information, and generate coherent summaries that capture the essence of the original text.improve performance norms, particularly in relation to information technology. less education or computer use.

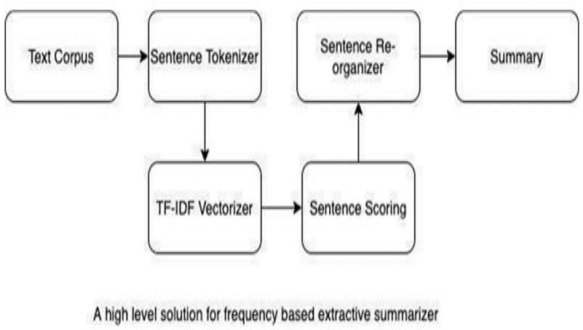

Figure 1: Extractive summarization

The process begins by analyzing the document's structure, creating a matrix that represents the relationships between terms and documents. Each row corresponds to a term, and each column corresponds to a document, with the entries indicating the frequency or importance of each term in each document.

The TF-IDF vectorizer is applied to the matrix to emphasize significant words in the document. This technique assigns weights to terms based on their frequency within the document and their rarity across the corpus. Terms that appear frequently in the document but rarely in the corpus receive higher weights, signifying their importance in the context of the document.

Leveraging these vectorized representations, machine learning algorithms are employed to identify and extract key sentences or phrases. These algorithms, such as clustering or classification methods, utilize the weighted term-document matrix to discern the most salient information. Key sentences or phrases are selected based on their representation in the matrix, forming the foundation of the summary.

In parallel, Deep Neural Networks (DNNs) are integrated into the process to further enhance summarization capabilities. DNNs utilize sophisticated layers of interconnected nodes to learn intricate patterns within the text data. By comprehending nuanced relationships and context, DNNs refine the summarization process, improving the model's ability to generate concise and informative summaries.

The synergy between traditional machine learning techniques and DNN approaches results in a robust summarization system. This fusion allows for the distillation of large volumes of information into concise, informative abstracts by capturing both the statistical relationships present in the data and the complex patterns discernible through deep learning.

Continuous evaluation and refinement of the summarization model are essential. Techniques such as cross-validation, human evaluation, and benchmarking against existing summarization systems ensure the quality and effectiveness of the generated summaries.

By following this methodology, extractive text summarization systems can effectively distill the essence of large documents into coherent and informative summaries, facilitating efficient information retrieval and understanding.

Conclusion

The integration of Latent Semantic Analysis (LSA), TF-IDF vectorization, and Deep Neural Networks (DNN) in extractive text summarization represents a significant advancement in the field. This fusion of traditional machine learning and deep learning techniques has led to the development of a powerful summarization system capable of distilling large volumes of information into concise and informative abstracts.

By analyzing the underlying structure of a document and creating a matrix of term-document relationships, LSA lays the groundwork for understanding the document's semantic context. The TF-IDF vectorizer then highlights important words by assigning weights based on their frequency and uniqueness, providing valuable insights into the document's key terms.

FUTURE WORK

In short, extractive text summarization involves:
- Analyzing document structure with LSA
- Highlighting key terms with TF-IDF
- Using ML to extract key sentences
- Enhancing with DNN for nuanced understanding
- Resulting in concise summaries.

REFERENCES

[1]1. Shashank Bhargav, Abhinav Choudhary, Shruthi Kaushik, Varun Dutt, “A comparison study

[2]abstractive and extractive methods for text summarization.”- Article in Advances in IntelliTumor Detection Using Convolutional Neural Network, 2019

[3]2. Pooja Batra, Sarika Chaudhary, Kavya Bhatt, Saloni Varshney, Srashti Verma. “A Review:

[4]Abstractive Text Summarization Techniques using NLP”- 2020.

[5]3. Hritvik Gupta, Mayank Patel. “Method of Text Summarization using LSA and sentence based

[6]topic modelling with BERT.”- International conference on Artificial Intelligence and Smart

Systems (ICAIS-2021).

[7]4. Jiawen Jiang, Haiyang Zhang, Chenxu Dai, Qingjuan Zhao, Hao Feng1 , Zhanlin ji, (member,

[8]ieee), and Ivan Ganchev. “Enhancements of Attention-Based Bidirectional LSTM for Hybrid

[9]Automatic Text Summarization”- Received July 17, 2021, accepted August 13, 2021, date of

publication September 3, 2021, date of current version September 14, 2021(IEEE access).

[10]5. D. Delen and M. D. Crossland, “Seeding the survey and analysis of research literature with

[11]text mining,” Expert Systems with Applications, vol. 34, no. 3, pp. 1707– 1720, 2008.

[12]6. J.-G. Yao, · Xiaojun Wan, and J. Xiao, “Recent advances in document summarization,” Knowl.

Inf. Syst., vol. 53, pp. 297–336, 2017.

[13]Jain, M.D.Borah,andA.Biswas, Summarization of legal documents: Where are we now and

[14]the way forward, Comput. Sci. Rev., vol. 40, May 2021, Art. no. 100388, doi: