Optimized Use of Multi-Document SummaryNeural Networks and Allied Techniques

D.Krishna[1], Dr.S.Jessica Saritha[2]

[1]Research Scholar, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Ananthapur, Andhra Pradesh - 515002.
[2]Assistant Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Ananthapur, Andhra Pradesh - 515002.

Email: [1]kittu514@gmail.com, [2]sjsaritha.cse@jntua.ac.in

Abstract

Automatic multi document summarization entails involvement of extraction of essential information from various inputs. The chief obstacle is the redundancy of information abstractive process is based on information combination sentence compression and reformulation whereas extractive process covers assignment of salient features including paragraphs sentence various techniques of MDS have been reviewed with emphasis on deep learning networks in which sentence are used as input to the visible layer. The abstractive neural model has been stressed as the preferred technique

Keywords: **Multi document summarization, DLN, NN**

DOI Number: 10.48047/NQ.2022.20.22.NQ10526 **NeuroQuantology 2022; 20(22):5128-5134**

1. Introduction

The information should be automatically extracted for multi-document summaries. As redundant information is available, it is difficult to gather several resources from where the data is obtained. Finally, it is a complicated procedure to get information from the extraction. It is possible to make either extractive or abstractive resumes. The abstractive summary usually includes the collection, compression and rephrasing of materials. The summary is based on additional analysis and examination of incoming texts. The saliency assignment procedure, by comparison, is called an extractive summary. The balancing of information coverage and redundancy is one of the hardest parts of the many automated document summary methods. Several techniques have been used, but the results have yet to be released. A deep learning method may be constructed of words as a contribution to the visible layer of a Deep Neural Network. Before using the layer, the lower-dimensional simple sentence space can be fine-tuned. This article examines various automated content resume techniques with the goal of ensuring that important phrases are accurately removed in a shorter period of time. The focus is on a profound neural network summary to achieve the optimum time compared to current techniques. It is relatively easy to summarize; the abstractive neural model of the single text was proportionally emphasized of neural networks has been covered in the following section. The following are 50 articles from the MDS survey. This notion was followed by the conclusion.

2. Literature Survey

Much research on the contribution of neural networks to MDS has been conducted. The summary task was assisted by two phases of grading and phrase selection [1-4]. This classification is explained in the parts that follow, followed by a summary of the papers reviewed. Methods for extracting query-oriented single document summarization that utilise a deep auto encoder to generate word frequency feature space are included. Ziqiang Cao used recursive neural networks to develop a multi-document synchronisation system (R2N2) [5-9].

Shibhansh Dohare proposed a new full-length text summary pipeline with mid stages representing (AMR) graphic abstraction, but the AMR summary graph is generated from this summary graph for the production of summary phrases. The quality of the summary generated by combining sentence scoring algorithms varies depending on the subject and has been evaluated by Ferreria using a variety of content, including news, blocks, and articles. In a DU C 2004 shared task, this model, known as GIGA Corpus, outperforms the most sophisticated method. Corpus proposes the automated HMDS Hasan building, which uses automatic generation to create a large, heterogeneous, multi-distributed summary of copper. The technique for extracting this method's sentence applications was created in order to construct a summary experiment on the DUC data sets from 2006-2007. This structure was effective [10-14].

Aakash Sinha and Ben King utilised neural networks to create a single, deep NLP architecture in their automated text summary research. Toha Lino and Amacio provided an attractive multi-Document Text Simulation based on grammar independent sets before evaluating the global route effects of this comment. The updated Sukriti Verma & Nidhi utilised neural networks as a basis for text extraction, using the data-based neural network subsequent transmission method. The proposed approach is scaleable, and the arbitrary papers are summarised. Researchers have developed a heterogeneous graph-based neural network for extractive resume analysis [15-21].

Danqing Wange and Yang Liu propose Transformers for the MDS hierarchy. Sarkar et al enhanced MDS by adding better sentence similarity and upgrading the method for computing sentences to increase similarity. Machine learning and neural networks have been utilised in automatic text resumption. The Chunk graph organises all the information inside a phrase cluster using a one-word structure [22].

Performance metrics

1) Inter sentences similarity
2) Readability metrics to validate non-Redundancy.
3) Cohesiveness
4) Readability of summary

Inter sentences similarity

Inter-sentence similarity refers to the measure of similarity between two or more sentences. This concept is used in various fields such as Natural Language Processing (NLP), Information Retrieval (IR), and Text Mining. The similarity between sentences can be determined based on various factors such as the words used, their semantic meaning, and their relationships with each other. The goal of measuring inter-sentence similarity is often to identify redundant or similar information in a text corpus or to cluster text data into semantically similar groups.

There are various ways to determine the inter-sentence similarity, including:

- **Jaccard Similarity:** This measures the similarity between two sentences by calculating the ratio of the number of common words to the number of unique words between the two sentences.
- **Cosine Similarity:** This method calculates the cosine of the angle between two vectors representation of the sentences, where the vectors are created using term frequency- inverse document frequency (TF-IDF) weighting of the words in the sentences.
- **Euclidean Distance:** This method calculates the Euclidean distance between two sentences based on the difference in the values of their vector representations.
- **Semantic Similarity:** This method takes into account the semantic meaning of the words in the sentences and uses techniques such as word embeddings, ontologies, and knowledge graphs to determine the similarity between two sentences.

The goal of measuring inter-sentence similarity is to identify redundant or similar information in a text corpus and to group similar sentences together. This can be useful in applications such as text summarization, where redundant information is eliminated, or in information retrieval, where similar documents are grouped together for easier retrieval. Additionally, inter-sentence similarity can also be used to determine the relatedness between different pieces of information and to improve the accuracy of text classification and information extraction systems.

Readability metrics to validate non-Redundancy.

Readability metrics are measures of the ease with which a text can be understood by a reader. They are used to evaluate the writing style and the level of difficulty of a text and are widely used in fields such as education, publishing, and content creation.

In the context of validating non-redundancy, readability metrics can be used to determine if a text contains repetitive or redundant information. If a text has a low readability score, it may indicate that the text is too complex or that it contains redundant information that can be eliminated.

Some common readability metrics used to validate non-redundancy are:

- Flesch-Kincaid Readability Score: This measures the reading ease of a text based on the average number of syllables per word and the average number of words per sentence. A higher score indicates that the text is easier to read.
- Gunning Fog Index: This measures the readability of a text based on the average number of words per sentence and the number of complex words (long words and words with multiple syllables). A lower score indicates that the text is easier to read.
- SMOG (Simple Measure of Gobbledygook) Index: This measures the readability of a text based on the number of complex words in a text. A lower score indicates that the text is easier to read.
- Coleman-Liau Index: This measures the readability of a text based on the average number of characters per word and the average number of words per sentence. A higher score indicates that the text is easier to read.

Readability metrics can be useful in identifying redundant or complex information in a text and can be used to improve the clarity and simplicity of writing. However, it's important to note that these metrics are not a definitive measure of the quality or value of a text and should be used in conjunction with other methods, such as inter-sentence similarity analysis, to validate non-redundancy in a text.

Cohesiveness

Cohesiveness refers to the quality of a text that makes it logically and semantically connected, so that the text is easily understood and coherent as a whole. It is a measure of the unity and flow of a text and the relationship between its parts.

In writing, cohesiveness is achieved through the use of transitional words and phrases, repetition of key words and concepts, and the establishment of clear relationships between ideas and sentences. A cohesive text is one where all the sentences are logically and semantically connected and where the flow of ideas is easy to follow.

There are various factors that contribute to cohesiveness, including:

- **Connective devices:** These are words or phrases that connect ideas and sentences, such as "and," "however," "in addition," etc.
- **Reference:** The use of personal pronouns, demonstrative pronouns, and other forms of reference that connect sentences and ideas.
- **Repetition:** The repetition of key words, phrases, and concepts helps to create cohesiveness by establishing connections between sentences and ideas.
- **Cohesive themes:** A cohesive text has themes that run throughout the text, connecting the different parts and making the text feel unified.

Cohesiveness is important in written communication because it makes the text easier to understand, improves the readability of the text, and enhances the overall quality of the writing. A lack of cohesiveness can make a text difficult to follow, causing the reader to lose interest or become confused.

In summary, cohesiveness refers to the quality of a text that makes it logically and semantically connected, resulting in a unified and coherent whole that is easy to understand and follow.

Feasibility schemes

Extractive text summarization which uses DNN [23] to obtain a representative subset of the input device by selecting the sentences which contribute the most to the entire content of document. Major advantage is its ability to find out intrinsic semantic space which enables the extraction of semantically relevant sentences and the information average can increased without increasing the redundancy in the summary.

Encoder -Decoder RNN is used to overcome all the limitations faced by the NLP [24] text summarization such as getting a short and accurate summary.

CNN can learn the Abstract representation of N-grams [25] & tackle sentences of various lengths word assembling can be done by CNN with cooling for capturing important features clustering can be done for similar background with reference summaries.

Sentence fusion for abstractive MDS can have:

- Sentence clustering
- Neural sentence fusion

Sentence clustering is the process of grouping a set of sentences into semantically similar clusters. The goal of sentence clustering is to identify and organize related sentences based on their content and meaning. This technique can be applied to various natural language processing tasks, such as document summarization, text classification, and information retrieval.

Sentence clustering can be achieved using various techniques such as hierarchical clustering, k-means clustering, and graph-based clustering. In each method, the sentences are first represented as numerical vectors, typically using techniques such as word embeddings or TF- IDF representations. Then, similarity measures are used to compare the sentences and group them into clusters. The performance of sentence clustering can be evaluated based on metrics such as accuracy, recall, and precision.

Neural sentence fusion refers to the task of combining multiple sentences into a single, coherent, and fluent sentence. This is usually achieved using neural networks, which are trained on large datasets to generate the fused sentences. The models are trained to capture the semantic meaning and the relationships between the sentences and generate a coherent representation. The models can be trained with different objectives, such as maximizing fluency, coherence, and relevance of the fused sentences. Neural sentence fusion is widely used in various applications such as text summarization, machine translation, and answering systems.

Encoder input can be used on related set of sentences complementary methods for different tasks even insured processing input. Hybrid auto text summarization using DNN and physiology can use restricted Boltzmann machine as well as the Fuzzy inference system.

Objectives

a. To Study DNN based methods and design suitable feed-forward network for processing extractive to MDS.
b. To develop a sentence fusion for abstractive to MDS based on sentence clustering and in neural sentence fusion.
c. To hybridize auto text summarization using DNN as well as FIS Restricted Boltzmann machine.
d. To apply nature heuristic techniques for Optimization as such space in MDS so that it is difficult to extract important sentences.

3. Conclusion

Important techniques of MDS have been reviewed with emphasis on those related to DNNS. The main stress is an sentence fusion for abstractive methods based on sentence clustering in neural sentence fusion .Furthermore hybridization of auto text summarization has been considered along with application of nature based heuristic techniques of optimization.

4. References

[1] Abdi, Asad, Shafaatunnur Hasan, Siti Mariyam Shamsuddin, Norisma Idris, and Jalil Piran. "A hybrid deep learning architecture for opinion-oriented multi-document summarization based on multi-feature fusion." Knowledge-Based Systems 213 (2021): 106658.

[2] Lamsiyah, Salima, Abdelkader El Mahdaouy, Bernard Espinasse, and Saïd El Alaoui Ouatik. "An unsupervised method for extractive multi-document summarization based on centroid approach and sentence embeddings." Expert Systems with Applications 167 (2021): 114152.

[3] Dong, Luobing, Meghana N. Satpute, Weili Wu, and Ding-Zhu Du. "Two-Phase Multidocument Summarization Through Content Attention-Based Subtopic Detection." IEEE Transactions on Computational Social Systems (2021).

[4] Can, Duy-Cat, Quoc-An Nguyen,

Quoc-Hung Duong, Minh-Quang Nguyen, Huy-Son Nguyen, Linh Nguyen Tran Ngoc, Quang Thuy Ha, and Mai-Vu Tran. "UETrice at MEDIQA 2021: A Prosper-thy-neighbour Extractive Multi-document Summarization Model." In Proceedings of the 20th Workshop on Biomedical Language Processing, pp. 311-319. 2021.

[5] Khaleghi, Zeynab, Mohammad Fakhredanesh, and Maryam Hourali. "MSCSO: Extractive Multi-document Summarization Based on a New Criterion of Sentences Overlapping." Iranian Journal of Science and Technology, Transactions of Electrical Engineering 45, no. 1 (2021): 195-205.

[6] Shah, Darsh J., Lili Yu, Tao Lei, and Regina Barzilay. "Nutribullets hybrid: Multi- document health summarization." arXiv preprint arXiv:2104.03465 (2021).

[7] Roul, Rajendra Kumar. "Topic modeling combined with classification technique for extractive multi-document text summarization." Soft Computing 25, no. 2 (2021): 1113-1127.

[8] Keerthana, P. "Automatic Text Summarization Using Deep Learning." EPRA International Journal of Multidisciplinary Research (IJMR): 5.

[9] Chowdhury, Radia Rayan, Mir Tafseer Nayeem, Tahsin Tasnim Mim, Md Chowdhury, Saifur Rahman, and Taufiqul Jannat. "Unsupervised Abstractive Summarization of Bengali Text Documents." arXiv preprint arXiv:2102.04490 (2021).

[10] Yadav, Shweta, Deepak Gupta, Asma Ben Abacha, and Dina Demner-Fushman. "Reinforcement Learning for Abstractive Question Summarization with Question-aware Semantic Rewards." arXiv preprint arXiv:2107.00176 (2021).

[11] Nanba, Hidetsugu, Tsutomu Hirao, Takahiro Fukushima, and Manabu Okumura. "Text Summarization Challenge: An Evaluation Program for Text Summarization." In Evaluating Information Retrieval and Access Tasks, pp. 39-48. Springer, Singapore, 2021.

[12] Zhong, Ming, Da Yin, Tao Yu, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah et al. "QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization." arXiv preprint arXiv:2104.05938 (2021).

[13] Bafna, Prafulla B., and Jatinderkumar R. Saini. "Scaled Document Clustering and Word Cloud-Based Summarization on Hindi Corpus." In Progress in Advanced Computing and Intelligent Engineering, pp. 398-408. Springer, Singapore, 2021.

[14] Xu, Hongyan, Hongtao Liu, Wang Zhang, Pengfei Jiao, and Wenjun Wang. "Rating- boosted abstractive review summarization with neural personalized generation." Knowledge-Based Systems 218 (2021): 106858.

[15] Prudhvi, Kota, A. Bharath Chowdary, P. Subba Rami Reddy, and P. Lakshmi Prasanna. "Text Summarization Using Natural Language Processing." In Intelligent System Design, pp. 535-547. Springer, Singapore, 2021.

[16] Ghalandari, Demian Gholipour, Chris Hokamp, Nghia The Pham, John Glover, and Georgiana Ifrim. "A Large-Scale Multi-Document Summarization Dataset from the Wikipedia Current Events Portal." arXiv preprint arXiv:2005.10070 (2020).

[17] Li, Wei, Xinyan Xiao, Jiachen Liu, Hua Wu, Haifeng Wang, and

Junping Du. "Leveraging graph to improve abstractive multi-document summarization." arXiv preprint arXiv:2005.10043 (2020).

[18] Su, Dan, Yan Xu, Tiezheng Yu, Farhad Bin Siddique, Elham J. Barezi, and Pascale Fung. "CAiRE-COVID: a question answering and multi-document summarization system for COVID-19 research." arXiv e-prints (2020):arXiv-2005.

[19] Ghodratnama, Samira, Amin Beheshti, Mehrdad Zakershahrak, and Fariborz Sobhanmanesh. "Extractive document summarization based on dynamic feature space mapping." IEEE Access 8 (2020): 139084-139095.

[20] Mohd, Mudasir, Rafiya Jan, and Muzaffar Shah. "Text document summarization using word embedding." Expert Systems with Applications 143 (2020): 112958.

[21] Tsai, Chih-Fong, Kuanchin Chen, Ya-Han Hu, and Wei-Kai Chen. "Improving text summarization of online hotel reviews with review helpfulness and sentiment." Tourism Management 80 (2020): 104122.

[22] Abualigah, Laith, Mohammad Qassem Bashabsheh, Hamzeh Alabool, and Mohammad Shehab. "Text summarization: a brief review." Recent Advances in NLP: The Case of Arabic Language (2020): 1-15.

[23] Zhu, Junnan, Yu Zhou, Jiajun Zhang, Haoran Li, Chengqing Zong, and Changliang Li. "Multimodal summarization with guidance of multimodal reference." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 05, pp. 9749-9756. 2020.

[24] Haider, Mofiz Mojib, Md Arman Hossin, Hasibur Rashid Mahi, and Hossain Arif. "Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm." In 2020 IEEE Region 10 Symposium (TENSYMP), pp. 283-286. IEEE, 2020.

[25] Wilson, Praveen K., and J. R. Jeba. "Combinational Semantic Approach for Efficient Data Extraction from Multiple Documents." International Journal 9, no. 4 (2020).