Applied GIS

ISSN: 1832-5505
Vol-11 Issue-04 Nov 2023

A NATURAL LANGUAGE PROCESSING BASED FRAMEWORK FOR AUTOMATIC TEXT SUMMARIZATION

Varun Rramawat, Vishwajeeth Yadav, S. Sravan, G. Praneeth Reddy,
Dr. T. Srikanth(HOD& Associate professor),
Department of CSE,
MALLA REDDY INSTITUTE OF TECHNOLOGY AND SCIENCE, Telangana Hyderabad.

Abstract:

There is no better place to save information than on the internet. The data is saved in several forms, including PDFs, HTML, XML, and Word documents. The lack of a predetermined structure is a problem when dealing with natural language material. Thus, the text mining method becomes relevant in this setting. The idea of written The purpose of mining is to glean valuable insights from texts written in a natural language.

KEYWORDS: Assessment, Summarization, Natural Language Processing, Template

I. INTRODUCTION

Research in the field of Natural Language Processing (NLP) has grown in importance in this age of exponentially increasing digital information. Using a computer to analyze text in order to attain human-like language comprehension is the goal of natural language processing (NLP) [10]. Concepts including text classification, automatic text summarization, events extraction, dialogue management, clustering, and topic identification have been the focus of study in natural language processing.

bring up a couple. Extracting useful documents from a large corpus of papers is the primary focus of the current study. Automatic text summarization, conversation management, and the Naïve-Bayes algorithm. One definition of text categorization is the act of grouping together related texts into a single category. This is A classifier is constructed by training a machine learning algorithm using a dataset. Following that, the classifier is used to foretell the class of fresh information. The goal of automatic text summarizing is to condense a document's extensive material into a manageable size while maintaining quality little writing that is high quality, practical, and significant. The two main approaches to text summarizing are: Abstractive Synopsizing Text and Extractive Synopsis In order to answer a question, one must first identify the question and then find the answer in text in a natural language[2, 16].

II. RELATED WORK

Using an HMM tagger, [6] presents a technique for text summarizing based on frequently used terms. Case folding, tokenization, stemming, POS tagging, feature term identification, noun and verb chunking, term frequency, term weight, and sentence score analysis are all used to accomplish the summary. In reference 3, the SUMMARIST method of summary is described. In addition to the usual steps involved in preparing the text in its original state, it employs supplementary ideas for subject identification via the usage of cue phrases and topic analysis to derive the synopsis. Researchers in [9] used a strategy that focuses on phrase extraction, sentence selection using important terms included in them. You can see how well two potential phrases match up with the title of using the similarity metric. that paperwork. At last, the summary is composed of the sentences with the greatest ratings.

A non-negative matrix factorization (NMF) based technique with generic relevance weight is detailed in [13]. The search results' main and subtopic sentences provide the basis for the summary. In order to summarize, the approach described in [4] relies heavily on fuzzy logic. One may think of it as a value between 0 and 1. generated for every word in the output by combining sentence characteristics with the rules stored in the database. The significance of the statement in the end summary is determined by the value that was achieved in the output. In [5], the authors provide a machine learning technique that combines generic and query-based summaries. The ranking of the sentences is achieved automatically by combining several sentence attributes. Next, terms that appear in the same Features are derived by the grouping of sentences into word clusters. A two-stage process for extracting sentences for the purpose of text summary is being studied by researchers [15]. As a preliminary measure, two connected phrases are combined to form bigram psudo. Title and location-based statistical approaches are used for the purpose of determining the significance of those phrases. Bigram pseudo sentences are then segmented in the subsequent stage. With the aggregating similarity approach, the second sentence is extracted from the original sentences. The A sentence's importance may be determined using the aggregate similarity approach, which involves computing the similarities of all other sentences. in an article. Since it does not make use of linguistic resources like WORDNET, it provides a strong method at a minimal cost. A technique for unsupervised summarization that builds the summary by grouping and removing phrases from the implementation of the original text document is found in [11]. To achieve this goal, novel criteria functions for grouping sentences have has been suggested. Additionally, a discrete differential evolution method has been created by academics to maximize the criteria.

III. OUR APPROACH

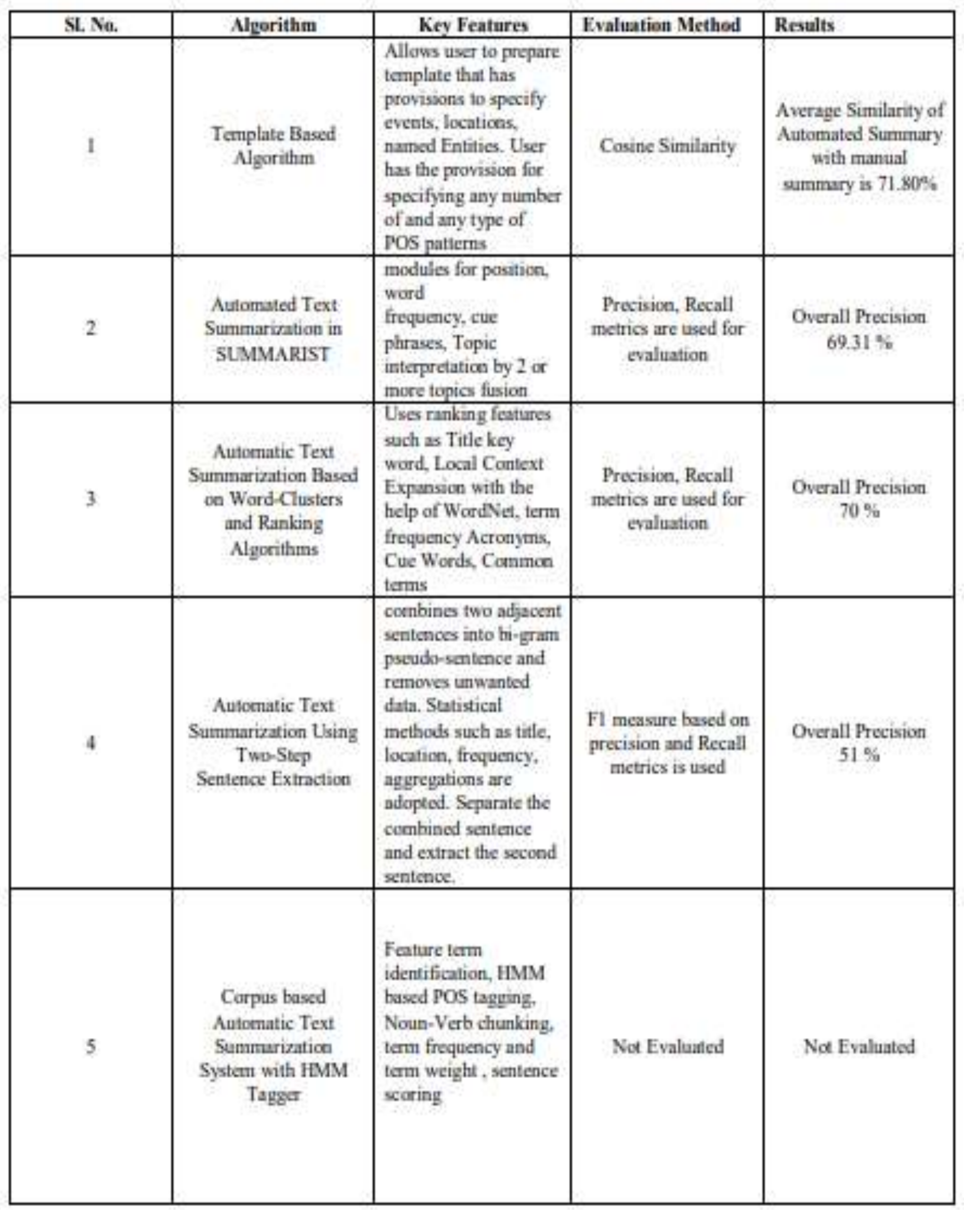
An important part of the document preparation process is part-of-speech (POS) tagging. We focused on developing a POS tagger in the early stages of our study [7]. The original plan was to leverage existing open source software to create a point-of-sale tagger, which could subsequently be used for document preparation. Everything went according to plan. Regardless, however A significant drawback was the limited language used in its construction. Consequently, thereafter it was choose to use Stanford POS tagger [12] instead of other methods for POS tagging. Automatic text summarizing techniques, as described in Prashant G Desai et al. [8], and management of conversation was put into place. An automatic text summarizing algorithm's technique relies on needs of the user prior to the original content being summarized, whereas the algorithm used for conversation management determines relationships between users and their machine. This report's primary contribution is the template-based system for managing conversations and summarizing texts. Users are able to set requirements using the algorithm. Specify the desired summary format. Such a guide is known as a template. Next, the algorithm processes the template in the same way input by the user in order to produce the synopsis. Experiments carried out throughout the execution of tasks have yielded promising outcomes.

IV RESULTS

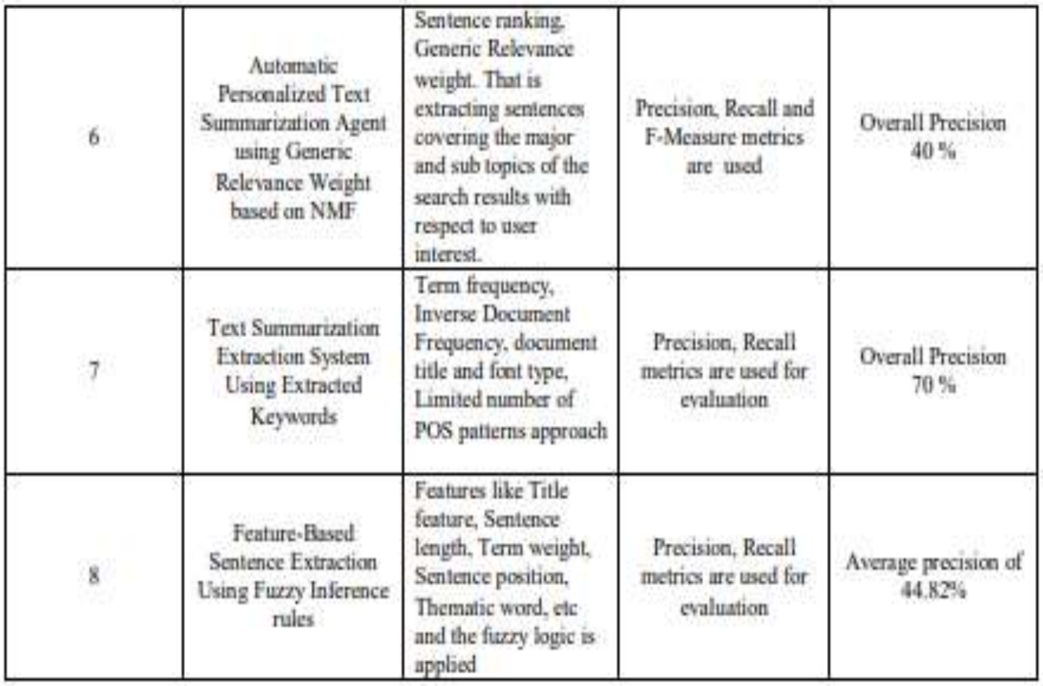
Various techniques may be used to assess the summaries. Intrinsic and extrinsic approaches are presented in [1]. Intrinsic summary assessment makes use of a variety of measures to contrast reference summaries created by humans with those produced automatically. An intrinsic summary assessment takes into account the following: similarity metrics using the cosine and jaccard matrices, as well as the euclidean distance and precision-recall-F-measure, many of others. Acceptability, correctness, and its impact on tasks like determining relevance are the pillars upon which extrinsic assessment rests. reading with understanding [14]. An overwhelming number of studies rely on intrinsic summary assessment methods that are grounded on Finding the F-measure, recall, and precision. The intrinsic assessment based on Cosine, however, is what we've chosen. Congruence, as it contrasts an algorithmically produced summary with one that was crafted by a human (the Reference) [8]. Results from an analysis of the methods suggested for use in automated text summarization are shown in Table -1. and several methods that have been suggested by earlier scholars [Sl.No. 1]. This proves that the research methods used in the

ISSN: 1832-5505
Vol-11 Issue-04 Nov 2023

Table 1: Key Features and Performance of Various Algorithms Used in Text Summarization



Sl. No.	Algorithm	Key Features	Evaluation Method	Results
1	Template Based Algorithm	Allows user to prepare template that has provisions to specify events, locations, named Entities. User has the provision for specifying any number of and any type of POS patterns	Cosine Similarity	Average Similarity of Automated Summary with manual summary is 71.80%
2	Automated Text Summarization in SUMMARIST	modules for position, word frequency, cue phrases, Topic interpretation by 2 or more topics fusion	Precision, Recall metrics are used for evaluation	Overall Precision 69.31 %
3	Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms	Uses ranking features such as Title key word, Local Context Expansion with the help of WordNet, term frequency Acronyms, Cue Words, Common terms	Precision, Recall metrics are used for evaluation	Overall Precision 70 %
4	Automatic Text Summarization Using Two-Step Sentence Extraction	combines two adjacent sentences into bi-gram pseudo-sentence and removes unwanted data. Statistical methods such as title, location, frequency, aggregations are adopted. Separate the combined sentence and extract the second sentence.	F1 measure based on precision and Recall metrics is used	Overall Precision 51 %
5	Corpus based Automatic Text Summarization System with HMM Tagger	Feature term identification, HMM based POS tagging, Noun-Verb chunking, term frequency and term weight , sentence scoring	Not Evaluated	Not Evaluated



6	Automatic Personalized Text Summarization Agent using Generic Relevance Weight based on NMF	Sentence ranking, Generic Relevance weight. That is extracting sentences covering the major and sub topics of the search results with respect to user interest.	Precision, Recall and F-Measure metrics are used	Overall Precision 40 %
7	Text Summarization Extraction System Using Extracted Keywords	Term frequency, Inverse Document Frequency, document title and font type, Limited number of POS patterns approach	Precision, Recall metrics are used for evaluation	Overall Precision 70 %
8	Feature-Based Sentence Extraction Using Fuzzy Inference rules	Features like Title feature, Sentence length, Term weight, Sentence position, Thematic word, etc and the fuzzy logic is applied	Precision, Recall metrics are used for evaluation	Average precision of 44.82%

V. CONCLUSION

For a long time, studies attempting to process and comprehend automated summarization of natural language text have been fraught with difficulty. Many other types of documents are finding uses for it, including research papers, product evaluations, emails, blogs, and educational domains. Having said that, not every summary needs the same one. subject areas. Title, sentence location, and sentence length were the primary characteristics used by the majority of the approaches. What we do prioritized the development of rules based on end-user needs using a template. The tests performed and outcomes achieved are satisfactory.

REFERENCES

[1]. Ahmed A. Mohamed, Sanguthevar Rajasekaran, "A Text Summarizer Based on Meta-Search", Proceedings of IEEE International Symposium on Signal Processing and Information Technology, pp.670-674, 2005.

[2]. Daniel Sonntag, "Distributed NLP and Machine Learning for Question Answering Grid", Proceedings of ECAI, 2004

[3]. Eduard Hovy, Chin-Yew Lin, "Automated Text Summarization in SUMMARIST", Proceedings of Advances in Automated Text Summarization, pp.2-14, 1999

[4]. Ladda Suanmali, Naomie Salim, Mohammed Salem Binwahlan, "Feature-Based Sentence Extraction Using Fuzzy Inference rules", Proceedings of International Conference on Signal Processing Systems, IEEE, pp.511-515, 2009

[5]. Massih R. Amini, Nicolas Usunier, Patrick Gallinari, "Automatic Text Summarization Based on Word-Clusters and Ranking Algorithms", Proceedings of ECIR, pp.142-156, 2005

[6]. M.Suneetha, S. Sameen Fatima, "Corpus based Automatic Text Summarization System with HMM Tagger", Proceedings of International Journal of Soft

Computing and Engineering, ISSN: 2231-2307, Vol. 1, Issue-3, pp.118-123, 2011

[7]. Prashant G. Desai , Niranjan N. Chiplumkar, Saroja Devi, "An Approach for POS Tagging for a Conversational Software System", Proceedings of the First International Conference on Research Trends in Computer Technologies, pp.423-425, 2013

[8]. Prashant G. Desai , Saroja Devi H ,Niranjan N. Chiplumkar, "A Template Based Algorithm for Automatic Summarization and Dialogue Management for Text Documents", Proceedings of International Journal of Research in Engineering and Technology, Vol. 04 Issue: 11, pp334-340, 2015

[9]. Rafeeq Al-Hashemi, "Text Summarization Extraction System (TSES) Using Extracted Keywords", Proceedings of International Arab Journal of e-Technology, Vol. 1, No. 4, pp. 164-168, 2010

[10]. Raju Barskar, Gulfishan Firdose Ahmed, Nepal Barska, "An Approach for Extracting Exact Answers to Question Answering (QA) System for English Sentences", Proceedings of ICCTSD, pp. 1187 – 1194, 2012

[11]. Rasim Alguliev, Ramiz Aliguliyev, "Evolutionary Algorithm for Extractive Text Summarization", Proceedings of Intelligent Information Management, pp. 128-138, 2009

[12]. Stanford University POS Tagger

[13]. Sun Park, "Automatic Personalized Text Summarization Agent using Generic Relevance Weight based on NMF", Proceedings of Honam University Gwangju, pp. 142-144, 2008

[14]. Vishal Gupta, "A Survey of Various Summary Evaluation Techniques", Volume 4, Issue 1, pp159-162, 2014

[15]. Wooncheol Jung, Youngjoong Ko, Jungyun Seo, "Automatic Text Summarization Using Two-Step Sentence Extraction", Proceedings of AIRS, pp. 71-81, 2005

[16]. Zhiguo Gong 1, Mei Pou Chan, "iQA: An Intelligent Question Answering System