FUZZY LOGIC EXTRACTION APPROACH FOR DOCUMENT SUMMARIZATION

[1]Ms D Triveni,[2] Mr Rokesh Kumar,

[1,2] Assistant Professor,Dept. of CSE,

Malla Reddy Engineering College (Autonomous), Secunderabad, Telangana State

ABSTARCT

Today, where the great amount of information is available on the internet, it is difficult to extract the correct and meaningful information quickly and efficiently. There are so many text materials available on the internet. We required a good mechanism to extract the most relevant information from it. Automatic Text Summarization mechanism solves this problem. "Text Summarization" is a process of creating a shorter version of original text which contains the important information. Text summarization can be broadly classified into two types: Extraction and Abstraction. This paper focuses on the Fuzzy logic Extraction approach for text summarization.

INTRODUCTION

Before going to the Text summarization, first we, have to know that what a summary is. A summary is a text that is produced from one or more texts, that conveys important information in the original text, and it is of a shorter form [1]. The goal of automatic text summarization is presenting the source text into a shorter version with semantics [2]. The most important advantage of using a summary is, it reduces the reading time. Text Summarization methods can be classified into extractive and abstractive summarization [1]. An extractive summarization method consists of selecting important sentences, paragraphs etc. from the original document and concatenating them into shorter form. An Abstractive summarization is an understanding of the main concepts in a document and then expresses those concepts in clear natural language. There are two different groups of text summarization [2]: Indicative and Informative. Inductive summarization only represents the main idea of the text to the user. The typical length of this type of summarization is 5 to 10 percent of the main text. On the other hand, the informative summarization systems gives concise information of the main text .The length of informative summary is 20 to 30 percent of

the main text. The automatic summarization means an automatically summarized output is given when an input is applied. Remember that input is well structured document. For this there are initially preprocesses such as Sentence Segmentation, Tokenization, Removing stop words and Word Stemming. Sentence Segmentation is separating document into sentences. Tokenization means separating sentences into words. Removing stop words means removing frequently occurring words such as a, an, the etc. And word stemming means removing suffixes and prefixes. After preprocessing each sentence is represented by attribute of vector of features. For each sentence there are 8 features and each feature has a value between 0 and 1. The 8 features are: Title features, Sentence length, Term weight, Sentence position, Sentence to sentence similarity, Proper noun, thematic word and Numerical data. Our approach is as follows: After extraction of 8 features the result is passed to fuzzifier then to inference engine and finally to defuzzifier. Rules for Inference engine is supplied from Fuzzy rule base. After this each sentence will have score and the sentence is sorted in the decreasing order of the score. Then 20% of these finally sorted sentences will be the summary of the given document.

RELATED WORKS The first Automatic text summarization was created by Luhn in 1958[1] based on term frequency. Automatic text summarization system in 1969, which, in addition to the standard keyword method (i.e., frequency depending weights), also used the following three methods for determining the sentence weights:

    a) Cue Method

    b) Title Method

    c) Location Method.

The Trainable Document Summarizer in 1995 performs sentence extracting task, based on a number of weighting heuristics. Following features were used and evaluated [2]:

1. Sentence Length Cut-O Feature: sentences containing less than a pre-specified number of words are not included in the abstract

2. Fixed-Phrase Feature: sentences containing certain cue words and phrases are included

3. Paragraph Feature: this is basically equivalent to Location Method feature

4. Thematic Word Feature: the most frequent words are defined as thematic words. Sentence scores are functions of the thematic words' frequencies

5. Uppercase Word Feature: upper-case words (with certain obvious exceptions) are treated as thematic words.

In 1990s the machine learning techniques in Natural Language Processing used statistical techniques to produce document summaries. They have used a combination of appropriate features and learning algorithms. Other approaches have used hidden Markov models and log-linear models to improve extractive summarization. Now a day's neural networks are used to generate summary for single documents using extraction. LaddaSuanmali [4] in his work has used sentence weight, a numerical measure assigned to each sentence and then selecting sentences in descending order of their sentence weight for the summary. Recently, neural networks are used to generate summary for single documents using extraction [6]. A lot of work has been done in single document and multi

document summarization using statistical methods. A lot of researchers are trying to apply this technology to a variety of new and challenging areas, including multilingual summarization and multimedia news broadcast.

III. MOTIVATION FOR TEXT SUMMARIZATION

Text Summarization is an active field of research in both the IR and NLP communities.

- ➢ People keep up with the world affairs by listening to news bites.
- ➢ People even go to movies largely on the basis of reviews they've seen.
- ➢ People base investment decisions on stock market updates.
- ➢ With summaries, People can make effective decisions in less time.
- ➢ The motivation here is to build such tool which is computationally efficient and creates summaries automatically

IV.APPROACHES TO SUMMARIZATION

Text summarization approach5 consists of following stages:

1. Preprocessing

2. Feature Extraction

3. Fuzzy logic scoring

4. Sentence selection and Assembly

CONCLUSION Automatic summarization is a complex task that consists of several sub affects the ability to generate high quality summaries. In extraction based summarization the important part of the process is the identification of important relevant sentences of text. Use of fuzzy logic as a summarization sub-task improved the quality of summary by a great amount. The results are clearly visible in the comparison graphs. Our algorithm shows better results as compared to the output produced by two online summarizers. Thus our proposed method improves the quality of summary by in the sentence feature extracted fuzzy logic system to capture the semantic relations between concepts in the The focus of this paper is narrow: summarization of documents, but the ideas are more broadly ap conclude that we need to extend the proposed method for multi document summarization with in a large data set.

REFERENCES

[1] SaeedehGholamrezazadeh ,MohsenAminiSalehi, "A Comprehensive Survey on Text Summarization Systems ", 978-1-4244-4946-0,2009 IEEE.

[2] Vishal Gupta and Gurpreet Singh Lehal "A survey of Text summarization techniques ",Journal of Emerging Technologies in Web Intelligence VOL 2 NO 3 August 2010.

[3] Oi Mean Foong ,Alan Oxley and SuziahSulaiman "Challenges and Trend Summarization ",International Journal of Information and Telecommunication

[4] Archana AB, Sunitha. C ,"An Overview on Document Summarization Techniques" ,International Journal on Advanced Computer Theory and Engineering (IJACTE) ,ISSN (Print) : 2319 ″U 2526, Volume 2013 .

[5] Rafael Ferreira ,Luciano de Souza Cabral ,Rafael DueireLins ,Gabriel Pereira e Silva ,Fred Freitas ,George D.C. Cavalcanti ,Luciano Favaro , "Assessing sentence scoring techniques for extractive text summarization ",Expert Systems with Applications 40 (2013) 5755-5764 ,2013 Elsevier .

[6] L. Suanmali , N. Salim and M.S. Binwahlan,"Fuzzy Logic Based Method for Improving Text Summarization" , International Journal of Computer Science and Information Security, 2009, Vol. 2, No. 1,pp. 4-10.

[7] Mrs.A.R.Kulkarni ,Dr.Mrs.S.S.Apte "A DOMAIN-SPECIFIC AUTOMATIC TEXT SUMMARIZATION USING FUZZY LOGIC ",International Journal of Computer Engineering and Technology (IJCET), ISSN 0976- 6367(Print), ISSN 0976 - 6375(Online) Volume 4, Issue 4, July-August (2013).

[8] FarshadKyoomarsi ,Hamid Khosravi ,EsfandiarEslami ,PooyaKhosravyanDehkordy; "Optimizing Text Summarization Based on Fuzzy Logic ",Seventh IEEE/ACIS International Conference on Computer and Information Science ,978- 0-7695-3131-1 ,2008

[9] LaddaSuanmali ,NaomieSalim and Mohammed Salem Binwahla ,"Feature-Based Sentence Extraction Using Fuzzy Inference rules ",2009 International Conference on Signal Processing Systems ,978-0-7695-3654-5 ,2009 IEEE .

[10] LaddaSuanmali, NaomieSalim and Mohammed Salem Binwahlan "Fuzzy Genetic Semantic Based Text Summarization ", 2011 Ninth Ninth International Conference on Dependable, Autonomic and Secure Computing ,978-0-7695-4612-4 ,2011 IEEE