ANALSYING THE PARALLEL FUSION LOGIC AND FUSION K-MEANS (FKM) WITH LANGUAGE DOCUMENT

Dr.P.Senthil Pandian[1], Dr.C.Seelammal[2], Mrs.S.Valli Mayil[3] , Mr.K.A Mohammed Faiz[4]

Associate Professor, Department of CSE, Solamalai College of Engineering, Madurai, Tamilnadu.
Associate Professor & Head, Department of CSE, Solamalai College of Engineering, Madurai, Tamilnadu.
Assistant Professor, Department of CSE, Solamalai College of Engineering, Madurai, Tamilnadu.
Assistant Professor, Department of CSE, Solamalai College of Engineering, Madurai, Tamilnadu.

---***---

Abstract - Text summarization is a method of retrieving information, and it is very effective. Text summarization approach techniques fall into two categories: abstract and abstract. This white paper focuses on extraction techniques so that the output of the application is the key sentence taken verbatim from the original text without verbal modification. There are four main processes in this research. These are the preprocessing phase, the scoring function, the optimization of summarization by two methods (logic and fused K-Means), and the extraction of summarization results. In this study, seven features are used to calculate a score for each sentence. The core of this study is to compare the reliability of fused logic and fused k-means methods in optimizing the aggregation results process. The results showed that the Fusion Logic method outperformed the Fusion K-Means method and had the highest similarity to human summarization, with a percentage accuracy of 48.25% for Fusion Logic and 44% for Fusion K-Means. It was %.33%. Accuracy improves with the additional compression of 0.67% for Fusion Logic Method and 3.5% for Fusion K-Means.

Keywords: *Language, Text Features, Scores Sentences, Fusion Logic, Fusion K–Means*

1.INTRODUCTION

Text summarization is a very important and useful technique for quickly and accurately identifying the subject of a text. Summarization is the process of condensing the source text into a shorter version while preserving the information content. The goal is to help potential readers save time and effort in finding useful information in your text. In search, automatic text summarization uses different methods and approaches. Previous studies have used genetic algorithms as a method to select the optimal text set and use different combinatorial feature sets. The result is the percentage of accuracy for each feature combination [3]. Another study uses pseudo-genetic and probabilistic methods of text summarization. Text summarization techniques fall into two categories: extraction and abstraction. Researchers can use either method or both methods at the same time to create automatic text summaries. Extraction summarization is a technique of sentence selection that presents the subject of a text and inserts it into a new text that is shorter than the original without altering the original text, while abstraction methods summarize with phrases that may not exist. Create a. The text included is the original text. This study uses three types of summarization: deductive and inductive. A deduction is a type of thought topic found in the first sentence of each paragraph. Therefore, the first sentence has the highest score in the entire paragraph below the following text. The topic is called inductive at the end of each paragraph, and the topic is in the middle of the paragraph.

The summarization process consists of her four phases:

1. Text preprocessing (Phase separation process includes sentence, case, sentence filtering process, tokenization, and stemming)
2. Feature scoring
3. Sentence selection (This phase includes Fusion Logic and Fusion (optimized using K–Means)
4. Selected preparation process sentences

The research in this work focuses on comparing summarization results using Fusion Logic and Fusion C-Means, thus demonstrating which method yields better summarization results. Fusion Logic is a knowledge representation built by if-then rules. An essential feature of a system based on fusion rules is their comprehensibility, since each fusion rule can be interpreted linguistically.

The characteristics of the Fusion Logic method are:

1. The problem is solved by describing the system in terms of language or variables containing uncertainty/ambiguity rather than numbers
2. Use if-then rules to clarify relationships between one variable and another
3. Describe the system using the Fusion algorithm.

Fusion K-Means is a data clustering technique in which the presence of each data point in a cluster is determined by its degree of membership. The first basic concept of FCM is to determine the cluster center, which indicates the position of the mean of each cluster. Initially, the cluster centers are not yet accurate. Each data point has a degree of membership in each cluster. Thus, by iteratively improving the cluster centers and degree of membership for each data point, it becomes clear that the cluster centers are moving to the right place. The Fusion K-Means clustering algorithm is based on minimizing an objective function called the K-Means function [8]. Both methods are the focus of testing in this document.

2. SEGMENTATION APPROACH

Preprocessing is the initial stage of preparation for assembling the input text into data for ready to use. The steps involved in text preprocessing are paragraph splitting, sentence splitting, case sensitivity, sentence filtering, tokenization, and stemming. Paragraph segmentation is the division and separation of text into paragraphs. Sentence splitting is splitting a paragraph into sentences. Characters other than "a-z" characters are removed from the text and uppercase letters are converted to lowercase. This process is called case sensitive. Each sentence should be filtered from stop words to text content.

Tokenization is the stage of breaking a sentence into its constituent parts using the "space" character as a delimiter. Finally, the result of the tokenization process is processed again down to the stem phase level to obtain the base word for each word. After preprocessing, each sentence in the text is represented by a weight derived from evaluating several features. Features have been carefully selected to ensure that the system provides nearly correct results. This paper focuses on seven characteristics of each sentence. Each feature has a value between '0' and '1'. The position of a sentence can determine whether the sentence is important for the display of textual content. Generally, the first sentence of each paragraph scores the highest and is called a deductive summary. However, it is possible that the main subject of the text comes at the end of the paragraph and is called an inductive summary.

This study uses three types of summaries. The feature position value is determined from the position of that sentence within each paragraph (X) divided by the number of sentences in each paragraph contained in the paragraph (N). The sentence length function is used to filter short sentences containing only names and other authors. This short sentence is probably not part of the summary. The calculation of the value of this characteristic is the result of dividing the sentence length by the number of words in the sentence (X) and the total number of texts (S). Words in sentences, including titles, get high scores. This phase calculates the relevance of the words in the title. Titles usually consist of a few words, so we need to preprocess the text of the title text.

Tokens in the body are then matched against the token title. You can use the number of occurrences of terms in a sentence to identify key phrases. The score of each sentence can be determined by summing the score of each constituent word. Because the term weight method calculates the weight of a sentence, the method is called term frequency. The frequency of occurrence of terms or entities appearing in the source text can be a factor in calculating key terms. The term entity is a name, a job title, a country name, and so on. The score for this feature can be calculated as the ratio of the number of entities/terms in the sentence to the length of the sentence. Sentences containing numerical data are assumed to be highly significant and the most likely words that should be included in the results summary. Numerical feature data scores were calculated as the ratio of the number of numeric data contained in a sentence to the length of the sentence. This feature is important because words that contain the subject term can be displayed as key sentences. The 10 most frequent terms are used as subject terms for this feature. The subject term feature score calculation is the quotient of the number of words in a sentence and the thematic maximum of the subject terms in the entire sentence.

3. THE METHODOLOGHY

This section describes how to extract optimized aggregate results using Fusion Logic and the Fusion K-Means method. As explained at the beginning, the essence of this paper is to compare the reliability of his two methods used in this study. The results of the optimization phase using Fusion Logic and Fusion K-Means are produced after the preprocessing phase and merit function. The success of automatic text summarization is the sentence selection at this stage of the optimization results. During the optimization phase, summarization results in the generation of high-scoring sentences. The fusion logic optimization stage has three processes: Fuzzifier, Inference Engine, and Defuzzifier. Inputs to the Fusion Logic method are scores for the seven traits shown in Table 1 for each sentence. We enter the crisp fuzzification phase that fusion should have.

Very low (SR), Low (L), Medium (M), High (H) and Very High (ST). Membership functions which are used in here is "Curve Triangle"

SR	R	M	T	ST
a	0.1	0.3	0.6	0.9
b	0	0.25	0.5	0.75 2
c	0.2	0.4	0.7	0.7

Table 1. Fusion Logic method

In many studies, antecedent Fusion sets were generated and tuned by numerical input data for rule base construction to improve the classification accuracy of FRBCSs. The most important part of inference engine is Fusion IF Then rules. The experience of the human controller is usually expressed as some linguistic "IF-THEN" rules that state in what situation(s) which action(s) should be taken. One example of IF-THEN rule is as follows: IF (Position sentence is Very High) and (Long sentence is Low) and (Title Feature is High) and (Term Weight is Medium) and (Noun Entities is Low) and (Data Numerical is Very High) and (Thematic Word is Very High) THEN sentence is important. Tsukamoto is a settlement that will be used in the Inference Engine. In the end of the stage, defuzzifier crisp sentences are classified into categories important, average and not important, then the next set is converted into crisp numbers of sentence weight.

Optimization with Fusion K-Means method will classify the input sentence into three clusters include: cluster of important sentences, sentence of the average cluster and cluster sentences are not important. The input which is used in the form of sentences and sentence attributes in form of features. There are 7 Attributes that are used namely: Position Sentence (x1), length of sentence (x2), Title Feature (x3), Term Weight (x4), Entities word (x5), Numerical Data (x6), Thematic Words (x7). FCM output is not Fusion inference system, but a row of cluster centers and several degree of affiliation for each data point. This information can be used to construct

Fusion inference system. Fusion K-Means algorithm is as follows in Table 2.

Step 1: Input data that will be clustered, X , in the form of matrix n x m (n = number of data samples , m = attributes of each data = 70, where Xij = data sample i (i = 1,2,…, n), attribute j (j = 1,2,…,7)
Step 2: Set the number of clusters = c = 3, Rank = w = 2, Maximum iterations = 100, Error smallest expected = 0.00001, initial objective function (P0) = 0, the initial iteration (t) = 1.
Step 3: Generate random numbers I, k, where i = 1,2,…,n; k = 1,2,3; as elements initial partition matrix U. Count the number of each column: with i = 1,2,…,n
Step 4: Calculate cluster k: Vkj, with k = 1,2,3; and j = 1,2,…,m.
Step 5: Calculate the objective function at iteration t, Pt
Step 6: Calculate the change in the partition matrix With: i = 1,2,…,n; and k = 1,2,…,m
Step 7: Check the condition to stop.
If the value of
(|Pt – Pt-1| < 0.00001) or (t > 100) then
stop and
then specify the cluster affiliation.
If not: t = t + 1, repeat step 4.

Table 2. Fusion K-Means

Fusion Logic and Fusion K-Means extraction generate sentences with the highest scores in descending order as a result of summarization. Sentences were extracted and 25% and 20% of the text was taken from the original text. There are various extraction possibilities to determine the best summary composition. The Fusion Logic and Fusion K-Means results were compared to ideal human summarization. The sentences are sorted according to the actual order of the original text. In addition, this study compares the summarization results to be deductive and inductive with the optimal summary construction of the extraction results.

4. RESULT AND ANALYSIS

The main purpose of this study was to compare the reliability of Fusion Logic and Fusion K-Means methods in terms of optimizing results for Indic text summaries and it is shown in the Table.3. So you get two results. One is summarization methods (deductive, inductive) and the other is extraction-synthesis. It was also investigated by summarizing 22 Indic text samples collected from several news sites. A result evaluation system using the formulas Precision, Recall, and F-Measure called built-in methods. Evaluation involves creating a set of ideal summaries, one for each text, testing the texts, and comparing the results to the ideal system summaries. Duplicate measurements. Reproducibility and accuracy of sentences and phrases are often mentioned, but sometimes single words are duplicated.

Recall = Relevant system φ Retrieved system / Total system

Precision =Relevant system Ω Retrieved system / ∑ Total system

Optimizing Method	Average K Measure	
	25%	20%
Fusion Logic	0.037652	0.037736
Fusion K Means	0.036543	0.036892

Table 3. Optimizing Method

5. CONCLUSION

In this work, we compared Fusion Logic and Fusion K-Means as methods of optimizing sentence extraction for automatic text summarization. In this study, sentence position, sentence position, Sentence length, feature titles, term weights, entity words or terms, numerical data, and topic words. Using the deductive form for summarization, he applied this research to two sentences. Compression ratios were 20% and 25%. The best combination turned out to be a summary compression ratio that reached 20%. Study derived the accuracy of 48.25% for the Fusion Logic method and 45.5% for the Fusion K Means method. An additional compression ratio of 0.67% for the Fusion Logic method and 3.5% for Fusion K-Means improves accuracy. For automatic text summarization, the Fusion Logic method was able to optimize text summarization results better than the Fusion K-Means method.

REFRENCES:

[1] Michael "Using lexical chains for text summarization," Proceedings of Intelligent Scalable Text Summarization Workshop, 2020.
[2] Ali "Pseudo Genetic And Probabilistic-Based Feature Selection Method For Extractive Single Document Summarization", 2021.
[3] O.Dehzangi "Efficient Fusion rule generation: A new approach using data mining principles and rule weighting," IEEE FSKD, 2017.
[4] Jamesl., "The Fusion K-Means Clustering Algorithm," Computers & Geosciences, Vol. 20, No. 23 pp. 291-303, 2014.
[5] Seema Shukla, "Fusion Database Using Extend Fusion K-Means Clustering," Issue 13, 2021 pp.1467-1478.
[6] Ladda Suanmali "Fusion Logic Based Method for Improving Text Summarization," IJCSIS, Vol. 12, No. 19, 2019.
[7] Talla,Z. "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," Master of Logic Project. Institute for Logic, Language and Computation 2017.
[8] Thawonmas, "A Fusion classifier with ellipsoidal regions," IEEE Trans. On Fusion Systems, pp. 558-668, 2018.
[9] Li-Xin Wang "Generating Fusion Rules by Learning from Examples," IEEE Transaction on Systems, Man, and Cybernetics, Vol.26, No. 6, 2020.

BIOGRAPHIES

I completed my Doctorate in the field of Computer Science and Engineering from Anna University, Chennai in 2016. I also have published 5 books, 22 Journal papers in the reputed Journals and 8 papers in conferences.

Seelammal C is currently a Associate Professor in Solamalai College of Engineering, Tamil Nadu, India. She Completed Ph.D in Information and Communication Engineering in Anna University. Her research focuses include Network Security, Intrusion detection system and Machine Learning. She is a member of CSI and ISTE, India.

I completed my Master of Engineering in the field of Computer Science and Engineering from Anna University, Chennai in 2014. I also have published 2 books, and 4 papers in conferences.

I completed my Master of Engineering in the field of Computer Science and Engineering from Anna University, Chennai in 2013. I also have published 2 and 1 papers in conferences.