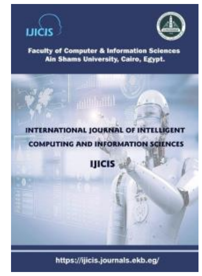
IJICIS, Vol.22, No.4, 62-85
DOI: 10.21608/ijicis.2022.160179.1215

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



USING A SEMANTIC FUZZY SYSTEM TO INTELLIGENT DOCUMENTS SUMMARIZATION

Ahmed E. Amin*
Department of Computer Science,
Mansoura University,
Mansoura 35516, Egypt
Ahmedel_sayed@mans.edu.eg

Received 2022-09-02; Revised 2022-09-25; Accepted 2022-10-02

Abstract: *Due to the information technology revolution, there are many and varied methods of document summarization to obtain specific information from documents. Automated summarization methods rely on identifying important points in all relevant documents to produce a concise summary. Therefore, this paper presents an intelligent classification-based automated summarization system using a semantic neuro-fuzzy approach. The proposed system consists of five integrated phases, which are the Document Pre-processing, the intermediate representation, the Index Matrices Weight Calculation, the Neuro fuzzy system, and the Summary Generation, respectively. The first stage divides paragraphs into sentences and sentences into words, by removing the most frequent words that do not carry any information and stripping the word from suffixes and prefixes to extract the « root » of the words. In the second stage, the Latent Semantic Index was used to produce the words/concepts matrix and concepts/sentences matrix. The third stage used the pointwise mutual information measure that defines particularly informative about the target word, as well as the best weighting of association between words. The knowledge is then extracted using a neuro-fuzzy network learning technique in phase four, which encodes the learned knowledge in its structure as a set of fuzzy rules. In order to build a number of fuzzy models with an increasing number of input variables chosen by the user according to their rankings, a quick clustering technique is then implemented. Then, according to a user-defined confidence level, the summary is generated from the knowledge base by a better understanding of the fuzzy rules. Recall-Oriented Understudy for Gisting Evaluation (ROUGE), which showed improved results in comparison to previous strategies in terms of average accuracy, recall, and F-measure in the document understanding conference (DUC) dataset, was used to assess the performance of the suggested model.*

Keywords: *Latent Semantic Index; Singular Value Decomposition; Pointwise Mutual Information; Neuro-fuzzy Network.*

*Corresponding Author: Ahmed E. Amin
Computer Science Department, Mansoura University, Mansoura 35516, Egypt
Email address: Ahmedel_sayed@mans.edu.eg

1. Introduction

The science of text mining cares about extracting important information from one document or more and aims to discover the structures of information from complex or semi-complex documents; that is called summarization. The definition of a summary is "a text that is created from one or more texts, that conveys substantial information in the original text(s), and that is not longer than half of the original text(s) and typically, much less than that." [1].

Researchers in this area has focused on the use of information technology in the summarization process that contributed to the emergence of an automatic document summarization (ADS). The task of ADS is producing fluent and concise summarization at the same time preserve on key information content and overall meaning.

The initial version of the ADS relied on text properties like word frequency and phrase frequency to extract the important lines from the document. The sentences in the documents are given weighting during the ADS development phases based on a function of high-frequency terms, disregarding the aberrant values that indicate the extremely high frequency of common words. There are three different ways to gauge a sentence's weight, including the Cue Method [2], which gauges a sentence's weight based on whether or not it contains terms from the Cue lexicon, and the Title Method [3]. Here, the Location Method [4] and the total number of times the content words appear in the document's head title and subtitle are used to determine the sentence's weight. This approach is predicated on the idea that the likelihood of a significant sentence is higher when it comes to the start of a text or a paragraph. These techniques are used with sentence-level information to create models for automated summarization [5]. The many models and procedures [6] that were employed in ADS were separated into two categories: abstraction and extraction. The methodology of work for the two ways is different, with the extractive summary approach [7] relying on extracting the phrases from the source text after recognizing and producing them verbatim. While the abstractive summary approach [8] depends on sophisticated natural language processing algorithms to analyze and understand the text in order to produce a brief paragraph that contains the most crucial details from the original text.

Abstractive summaries [9] have been the subject of recent research, which addresses issues including semantic representation, inference, and natural language creation. Machine learning methods [10] have helped to solve these issues, making them significant axes in ADS. At the beginning of the 1990s, natural language programming (NLP) used machine learning techniques in summarization, where statistical models such as the naive-Bayes methods emerged, which initially assumed the independence of the features that used in the summarization. While some research focused on other models such as the Hidden Markov model [11] and log-linear models [12] in improving the summary. Finally, research has begun to use neural networks [13] and adapted them in the process of automated summarization using the third-party features that is represented in common words in search engine queries.

This paper aims to address the issue of the numerous ways that words and sentences may be meaning interpreted in some languages, such as Arabic, where a word's meaning can only be grasped from the context of a sentence or from the context of a paragraph. The proposed system uses a fuzzy system to create probabilities for words and sentences, and from an analysis of the context by using the stage of pre-processing, intermediate representation stage, and index matrices weight calculation the highest likelihood of the right interpretation is chosen.

The rest of this paper is organized as follows in section two shows related work. The third section focuses on the methodology of the proposed system in terms of topology and algorithm of training. After providing the experimental setup in section four, the results and discussion are presented in section five. Section six explains the conclusion. References are placed in the last section.

2. Related Work

There are many techniques used in the process of automatic summarization both extraction and abstraction types. One of the most widely used techniques in this field is the techniques of machine learning. Latent topic models are classified as unsupervised approaches [15] whereas classification and regression are classified as supervised techniques [14]. Among the techniques of machine learning that contributed to the success of the automatic summary are a neural network that NetSum [16] relied on the idea of its work. There are some systems [17] have inspired the idea of the success of shallow neural networks and built its idea of their work by changing the network topology such as feed-forward neural networks and recurrent neural networks (RNNs). RNN is utilized to create the sentence ranking in a hierarchical regression method [18] by using hand-crafted word characteristics as variable-length inputs. Based on the idea of extracting the candidate sentences for the summary, learned representations employ a deep architecture [19] to filter out words from a document that aren't significant in the early layer and find keywords in the latter layer. As supervised models, a convolutional neural network (CNN) [20] is utilized to extract potential words for the summary.

Soft computing's capacity to recognize crucial information in papers has recently increased its prominence. To extract key lines and put them in the summary, several researchers have suggested fuzzy logic reasoning-based summary systems [21]. Fuzzy logic models employed semantic analysis in addition to sentence scoring to provide a text summary. Where two fuzzy logic models were created. In order to extract information from documents, the initial [22] was built on an examination of cross-document interactions between sentences. While the second model [23] improved the quality of the summary by extracting the semantic relationships between the concepts based on the latent semantic analysis in the choice of summary phrases.

Along with both neural networks model and fuzzy logic model [24], the idea of some research has been to use the combination of the two techniques to improve summary results.

From the above, it can be observed that the possibilities and characteristics of the neuro-fuzzy approach can be adapted to extract a knowledge base from documents based on latent semantic analysis that extrapolates the semantic vector of sentences. Through the semantic query of the basic words in the knowledge base, the importance of sentences are determined and arranged to produce a summarization.

3. Methodology of proposed system

An Intelligent Documents Summarization System (IDSS) relies on the neuro-fuzzy approach that belongs to soft computing techniques. The neuro-fuzzy approach is characterized by combining the benefits of both fuzzy logic that based on knowledge-driven reasoning while neural networks are based on data-based approximation. In addition to the vector semantics in which the meaning of the word is calculated by the distribution of words around it. The IDSS architecture consists of five phases are Document Pre-processing, Intermediate Representation, Neuro-Fuzzy, Extract Summarization and Summary Generation as shown in Figure 1.

Phase 1: Document Pre-processing:

The stage of pre-processing is carried out through four sub-processes are:

- Tokenization: Sometimes known as "word segmentation," this technique explores the words in a phrase by segmenting the text's characters along their lengths and their beginning and ending locations.
- Sentence Segmentation (SS): The terms sentence boundary detection, sentence boundary disambiguation, and sentence border recognition are also used to describe this process. By dividing the text into sentences for further processing, including recognizing the boundaries between words in various phrases, SS seeks to discover the lengthier processing units made up of one or more words.
- Stop Word Removal (SWR): SWR aims to remove the most frequent words that do not carry any information (such as ‘a’, ‘and’, ‘the’ …etc.) leading to improve the effectiveness and efficiency of the summary.

- Stemming: This process aims to strip the word from suffix and prefix to extract the « root » of the words based on the dictionary or a set of rules developed by linguists.

Phase 2: Intermediate Representation:

Intermediate representation stage is one of the independent tasks used in the summarization process. Summarization systems have multiplied due to the multiplicity of intermediate representation methods which can be classified into two main types are topic representation [25] and indicator representation [26]. The Latent Semantic Index (LSI) [27] is one of the methods used in the intermediate representation that belongs to the approach of topic representation. The LSI relies on the Singular Value Decomposition (SVD) [28] in the representation of the basic semantics of the document using the matrix calculation as follows:

$$A = U_k S_k V_k^T$$

Where $S \in R^{m\times n}$ is a diagonal matrix with the singular values as its nonnegative diagonal members, $U \in R^{m\times m}$ and $V \in R^{n\times n}$are orthogonal matrices. The columns in V show the relationship between the documents and each of the k ideas, while the rows in U show the presence of the original words that correlate to the k concepts in the new factor space. Therefore, the documents can be represented as vectors where the semantics of these vectors are used to represent the meaning of the words by associating each word with the vector. So; the summarized topic can be classified as two matrixes as:

I. Word / Concept Matrix (U matrix) that represents that represents the relationship between each word and its concept in the document.

II. Concept / Sentence Matrix (C_s) is calculated from the following equation:

$$C_s = S_k V_k^T \quad \ldots\ldots (1)$$

Where (C_s) describes how much a concept represents a sentences.

Phase 3: Index Matrices Weight Calculation:

There are many measures are used to an association between words, the most commonly used are simple frequency [29]. However, this measure has some problems, such as raw frequency is much skewed and not very discriminative. The pointwise mutual information (PMI) [30] measure is the most important measures that define particularly informative about the target word, as well as the best weighting of association between words. PMI is based on the notion of mutual information (MI) where MI between two random variables X and Y can be calculated as:

$$I(X,Y) = \sum_x \sum_y P(x,y)\ \log_2 \frac{P(x,y)}{P(x)P(y)}$$

PMI can apply to co-occurrence independent vectors like a word (*w)* and a concepts (*c*) as:

$$PMI(w,c) = \log_2 \frac{P(w,c)}{P(w)P(c)}$$

The PMI ranges from negative infinity to positive infinity. Negative values imply that occurrences occur less frequently together than would be predicted by chance, which makes them unreliable for summarizing a small number of documents. Moreover, it is not possible to assess the degree of "non-coherence" with human judgments. The positive pointwise mutual information (PPMI) was replaced instead of negative values with zero as:

$$PPMI(w,c) = max\left(\log_2 \frac{P(w,c)}{P(w)P(c)}, 0\right)$$

One of the main drawbacks of PMI is the bias toward low-frequency events to slightly change the computation for $P(c)$, where the tendency of very rare words to get very high PMI values. The $P_\alpha(c)$ function contributed to solving this problem by raises contexts to the power of (α) as:

$$PPMI_\alpha(w,c) = max\left(\log_2 \frac{P(w,c)}{P(w)P_\alpha(c)}, 0\right)$$

$$P_\alpha(c) = \frac{count(c)^\alpha}{\sum_c count(c)^\alpha}$$

According to research [31], 0.75 is the ideal number for (α) to increase embedding performance in comparison to weights comparable to those used in skip-grams.

Phase 4: Neuro fuzzy system:

Neuro-fuzzy system is collected between two techniques from the soft computing namely fuzzy logic and neural network. As a result, it blends learned implicit information with explicit knowledge reasoning that may explain the link between input and output. Auto-summarization systems based on acknowledged fuzzy reasoning and the related neuro-fuzzy architecture may be able to maximize the benefits of both methodologies for producing document summaries.

The summary extraction is accomplished by learning a neuro-fuzzy network that extracts the knowledge by encoding the learned information in the structure of fuzzy rules. The knowledge extraction process is carried out by combining two fundamental steps, as shown in figure 2.

Applying unsupervised learning to define the fuzzy rules' structure and parameters is the first step. In order to increase the precision of the knowledge deduced, step two involves modifying the fuzzy rule parameters through supervised learning.
The size of the network topology, which refers to the number of fuzzy rules and membership functions, as well as network weights, which relate to the initial values of the

parameters of the rules, are all set during an unsupervised learning stage. The competitive learning algorithm (CLA) [32] is employed to attain the aim of this stage. The mean squared error (MSE) is used as a performance indicator in the second learning stage, which is a part of the supervised learning algorithm, in order to optimally adjust the fuzzy rules' parameters.

The proposed system, as depicted in figure 3, combines the two preceding steps by using the CLA to define the structure and parameters of the fuzzy rules, then using supervised learning to fine-tune the fuzzy rule parameters to increase the precision of knowledge derived. Assuming, for example, that there are (N) input-output pairs that describe the documents to be condensed as:

$$T = \{(W_t, S_t)\}_{t=1}^{N}$$

To extract the knowledge as:

$$If\ (w_1\ is\ V_{1f})\ AND \ldots AND\ (w_n is\ V_{nf})\ THEN\ (s\ is\ b_f)$$

Where: $f = 1,2, \ldots, F$, F is the fuzzy rules number of the model. The output variables S define the fuzzy singleton b_f. The vocabulary input variables (v_i) define the sentence fuzzy sets which are $(S_{if})_{i=1}^{n}$. Each (S_{if}) is represented by a Gaussian membership function as:

$$\mu_{ir}(w_i) = e^{-\frac{(w_i - C_{ir})^2}{2a_{ir}^2}}; \quad \ldots\ldots \text{ Eq. I}$$

$$where\ C_{ir}\ and\ a_{ir}\ are\ the\ center\ and\ the\ width\ of\ the\ Gaussian\ function, respectively.$$

A neuro-fuzzy model based on rules that predicts the output values $W = (w_1, w_2, \ldots, w_m)$, given the input values $V = (v_1, v_2, \ldots, v_n)$, is derived. There are two integrated stages are:

- A fuzzy reasoning stage: From each cluster of the input space, a fuzzy rule is to be derived in this step. Based on the maximum rules number (R max), which can be viewed as the network's meta-node and is denoted by a grey circle in Figure 3, the CLA is used to build the neuro-fuzzy network. The appropriate rules number is then automatically chosen to represent the input data during learning, lowering the total number of clusters. For example, if T_0 is an unknown input vector, the following is how the output of this stage is obtained:

I. By using the Larsen product operator meaning, the matching degree is calculated as:

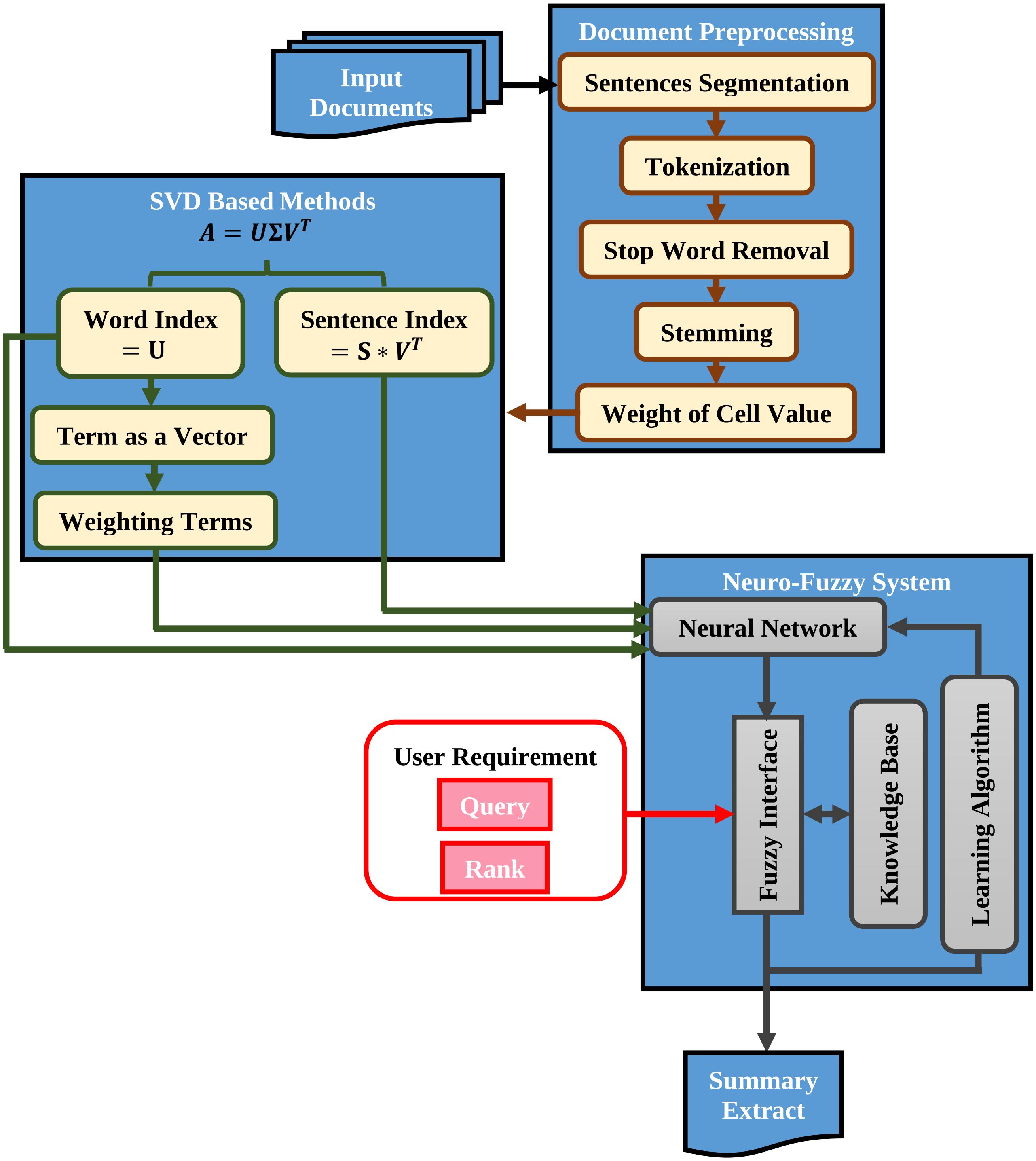


Figure 1: The architecture of proposed system.

$$\mu_r(T_0) = \prod_{i=1}^{n} \mu_{ir}\,(T_{0i}); \quad for\ r = 1,2,\dots,R$$

I. Calculate the deduced output $(\widehat{S_0})$ as:

$$\widehat{S_0} = \frac{\sum_{r=1}^{R} b_r \mu_r(W_0)}{\sum_{r=1}^{R} \mu_r(W_0)}$$

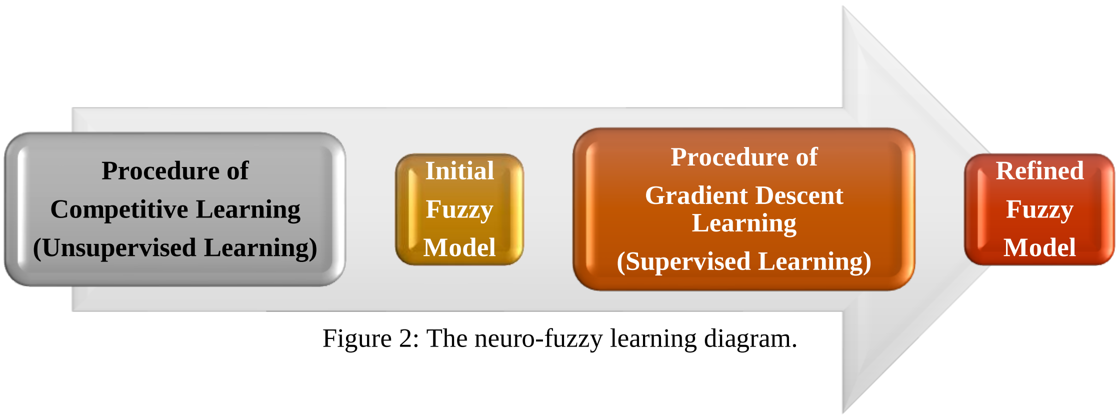


Figure 2: The neuro-fuzzy learning diagram.

The network meta-nodes are involved in the CLA by using the reward/punishment mechanism (RPM) [33]. By competing for the weight vector (C) that is closest to the input vector, the RPM tries to choose the meta-node, whereas the competitor node is the second closest node. The network organizes its own structure by looking for R meta-nodes whose weight vectors $(C_r = C_{1r}, C_{2r}, \dots, C_{nr})$ represent the clusters centers of the Gaussian membership functions (μ_{ir}) for the data space. The first-nearest-neighbor is used to define the width (a_{ir}) as:

$$a_{ir} = \frac{\|C_r - C_s\|}{\gamma}$$

Where $C_s \rightarrow$ the cluster center nearest to C_r, and $\gamma \rightarrow$ an overlap parameter ranging in [1.0, 2.0]. The resulting parameters b_r are obtained as follows:

$$b_r = \frac{\sum_{t=1}^{N} \mu_{rV_t W_t}}{\sum_{t=1}^{N} \mu_{rV_t}}$$

Where $\mu_r(W_t)$ is the matching level of the premise part of the rule calculated in Eq. I.

- A neuro-fuzzy network stage: The neuro-fuzzy network that has been created consists of three layers. Each layer of the network includes the neuron units that have the following specifications:

I. Neurons are organized in the first layer, sometimes referred to as the input layer, according to the number of rules (R), with each neuron standing for a fuzzy rule. Each neuron in this layer receives an input value and then calculates the Gaussian member function value and produces one output by the following function:

$$O_{ir}^{(1)} = e^{-\frac{(v_i - c_{ir})^2}{2a_{ir}^2}}, \quad i = 1,2,\dots,n \ \text{ and } r = 1,2,\dots,R$$

II. The second layer is known as the hidden layer, it contains a number of neurons. The IF-part of the fuzzy rule is represented by the n-fixed connections connecting each neuron to the input layer. Then, the output of these neurons is calculated according to the following equation:

$$O_r^{(2)} = \prod_{i=1}^{n} O_{ir}^{(1)}; \quad r = 1,2,\dots,R$$

III. The third layer - the output layer - which the rules are derived according to:

$$O^{(3)} = \frac{\sum_{r=1}^{R} b_r O_r^{(2)}}{\sum_{r=1}^{R} O_r^{(2)}}$$

Phase 5: Summary Generation:

The process of creating the summary involves two parts, the first of which involves building a series of fuzzy models with an increasing number of input variables that are chosen by the user based on their rankings. The following is an example of a quick clustering algorithm that assigns a rank value:

A Fast Clustering Algorithm

Given a data set $\boldsymbol{W} = \{\boldsymbol{w_1}, \boldsymbol{w_2}, \dots, \boldsymbol{w_n} | \boldsymbol{w_t} = (\boldsymbol{x_t}, \boldsymbol{y_t})\}$

1. **Set the first prototype $\boldsymbol{v_1} = \boldsymbol{w_1}$**
2. **Set k=1, $\boldsymbol{N_{S_1}} = \boldsymbol{1}$**
3. **For each $\boldsymbol{t} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{N}$;**
    a. **Find the nearest prototype $\boldsymbol{v_j}$such that**
    $$\|\boldsymbol{v_j} - \boldsymbol{w_t}\| = \min_{i=1,2,\dots,N} \|\boldsymbol{w_i} - \boldsymbol{w_t}\|$$
    b. **If $\|\boldsymbol{v_j} - \boldsymbol{w_t}\| \le \boldsymbol{\varepsilon}$ (where $\boldsymbol{\varepsilon}$ is a predefined threshold), then**
        i. **Set $\boldsymbol{N_{S_j}} = \boldsymbol{N_{S_j}} + \boldsymbol{1}$**
        ii. **Set $\boldsymbol{v_j} = \boldsymbol{v_j} + \|\boldsymbol{v_j} - \boldsymbol{w_t}\| / \boldsymbol{N_{S_j}}$**
    c. **Else**
        i. **Set $\boldsymbol{k} = \boldsymbol{k} + \boldsymbol{1}$**
        ii. **Create a new prototype $\boldsymbol{v_k} = \boldsymbol{w_t}$**
4. **End for.**

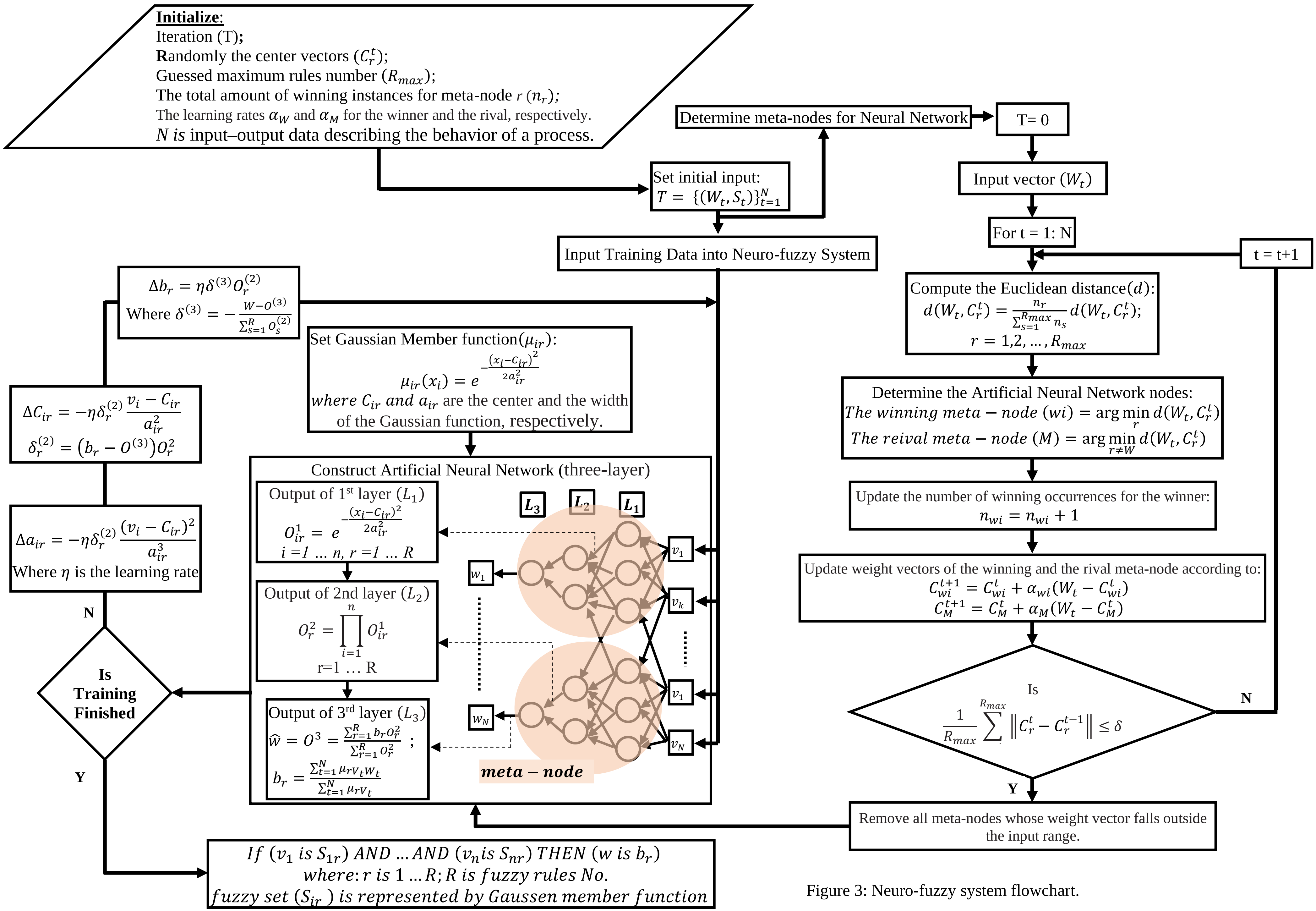


Figure 3: Neuro-fuzzy system flowchart.

To better comprehend the fuzzy rules, the summary is constructed in the second stage, when the knowledge extraction process is provided with an extensive representation of the projected outputs based on a user-defined confidence level. These granules establish the following prediction intervals for various levels of output value variation:

$$e_t = \hat{y}_t - y_t \quad ; \quad t = 1,2,\dots,N$$

Where; e_t is a sample of a random variable e from an unknown distribution that is independent and identically distributed. Therefore, the mean value is:

$$\bar{e} = \frac{\sum_{t=1}^{N} e_t}{N}$$

If the value of (N) is higher, the value of (e) is closer to the normal distribution so, the prediction intervals for (e) can be calculated as:

$$P(e_{new} \in [L_\propto, U_\propto]) \geq 1 - \propto$$

Where; $[L_\propto, U_\propto]$ is lower and upper limits of prediction intervals at confidence level $\propto$ for the error (e_{new}) that corresponding the newly input variable with probability greater than $(1 - \propto)$. This equation can be equivalent to:

$$P(y_{new} \in [\,\bar{y} - U_\propto, \bar{y} - L_\propto]) \geq 1 - \propto$$

The limits of prediction interval are defined as:

$$L_\propto = \bar{e} - t_{\left(\frac{\propto}{2}, N-1\right)} \left(S \sqrt{\frac{1}{N} + 1} \right),$$

$$U_\propto = \bar{e} + t_{\left(\frac{\propto}{2}, N-1\right)} \left(S \sqrt{\frac{1}{N} + 1} \right)$$

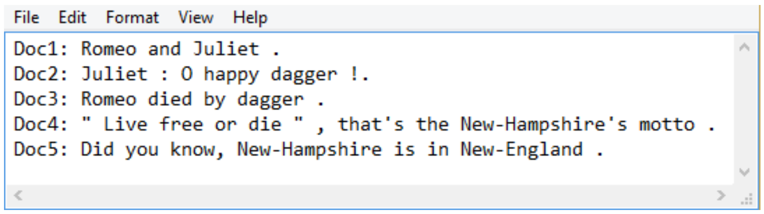
Where, $t_{\left(\frac{\propto}{2}, N-1\right)}$ is Student distribution value with (N-1) degree of freedom that corresponding to the critical value $\left(\frac{\propto}{2}\right)$ and (S) is the sampled standard deviation:

$$S = \sqrt{\frac{1}{N-1} \sum_{t=1}^{N} (e_t - \bar{e})^2}$$

4. Experimental Work

To illustrate the idea of the proposed system, a simulation was carried out to extract a summary of the documents. The steps of the summary process can be summarized in the steps of the following algorithm:

1. Input document: figure 4 shows the input document.



File Edit Format View Help

```
Doc1: Romeo and Juliet .
Doc2: Juliet : O happy dagger !.
Doc3: Romeo died by dagger .
Doc4: " Live free or die " , that's the New-Hampshire's motto .
Doc5: Did you know, New-Hampshire is in New-England .
```

Figure 4: Document example.

2. Preprocessing: the document was segmented as in phase 1.
3. Intermediate Representation by SVD:
   As previously mentioned in phase 2, the document is converted to basic semantics as shown in figure 5. Both words and document have a new representation in terms of hidden concepts as:

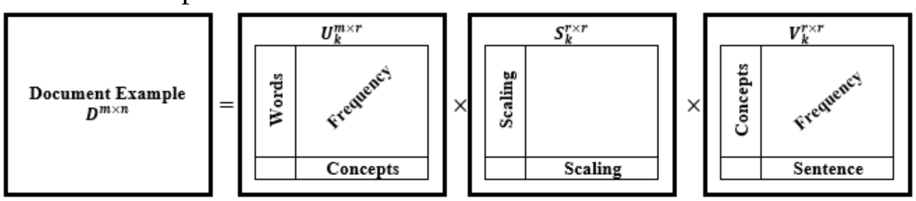


Figure 5: The document basics semantic.

   - The words index ($U_k^{m\times r}$) that represent the words as row vectors of the $(m \times r)$ matrix.
   - The Document index that represents the column vectors the $(r \times n)$ matrix by applied the Eq. 1.
4. The index matrices weight are calculated by applied phase 3 as shown in figures 6 and 7.

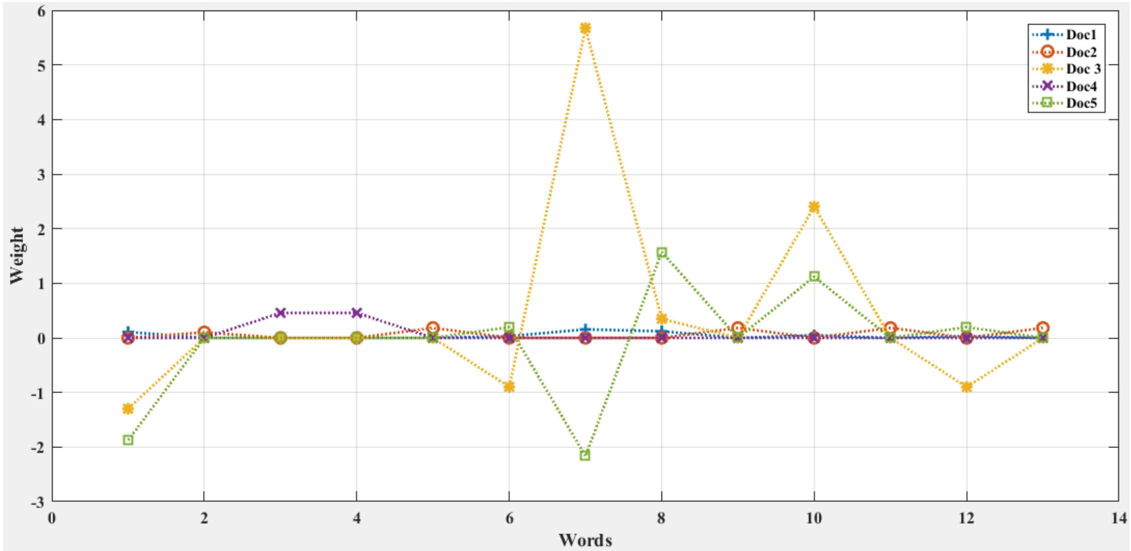


Figure 6: Weight of words index.

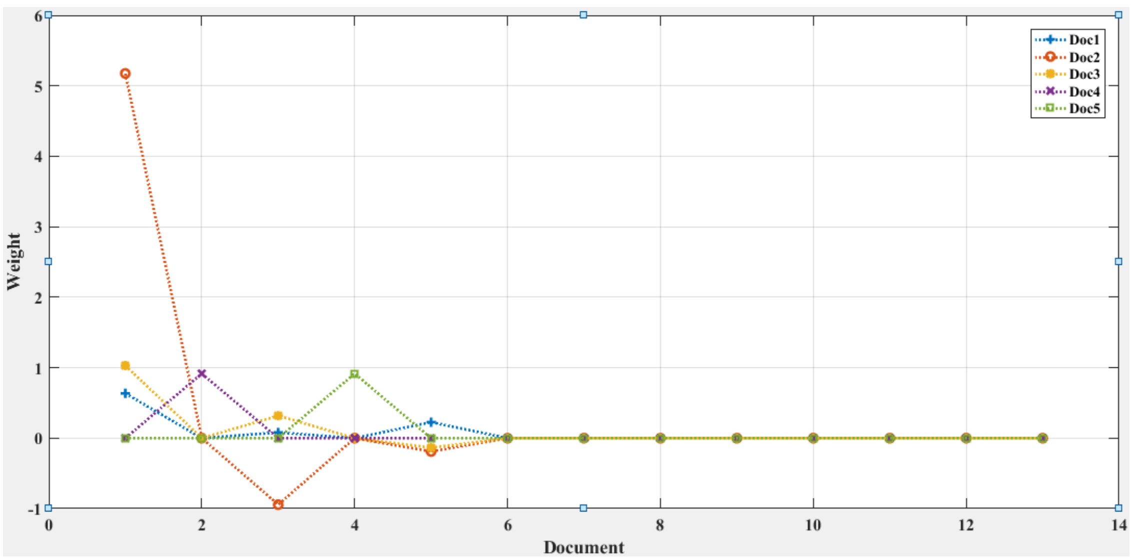


Figure 7: Weight of document index.

5. Extract knowledge base by using Neuro-fuzzy Model: by applied phase 4, there are three main steps used to extract rules:
   A. Fuzzy logic designer: the main parameters of fuzzy logic designer as shown in figure 8 are input fuzzy sets and output fuzzy set that represents as (weight of index words matrix and weight of document index matrix) and (Words Weight) respectively as shown in figure 9, membership function (Gaussian member function) and shape of membership function as shown in figures 10.

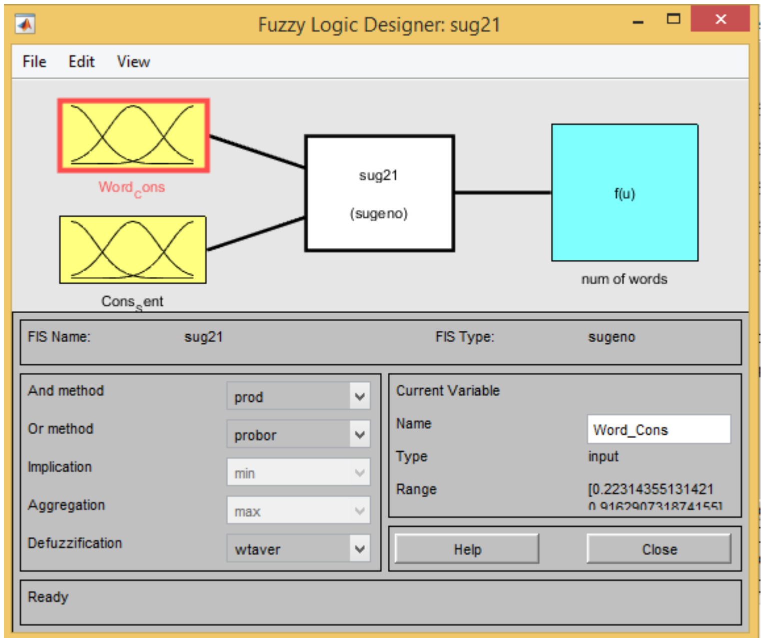


Figure 8: fuzzy logic designer.

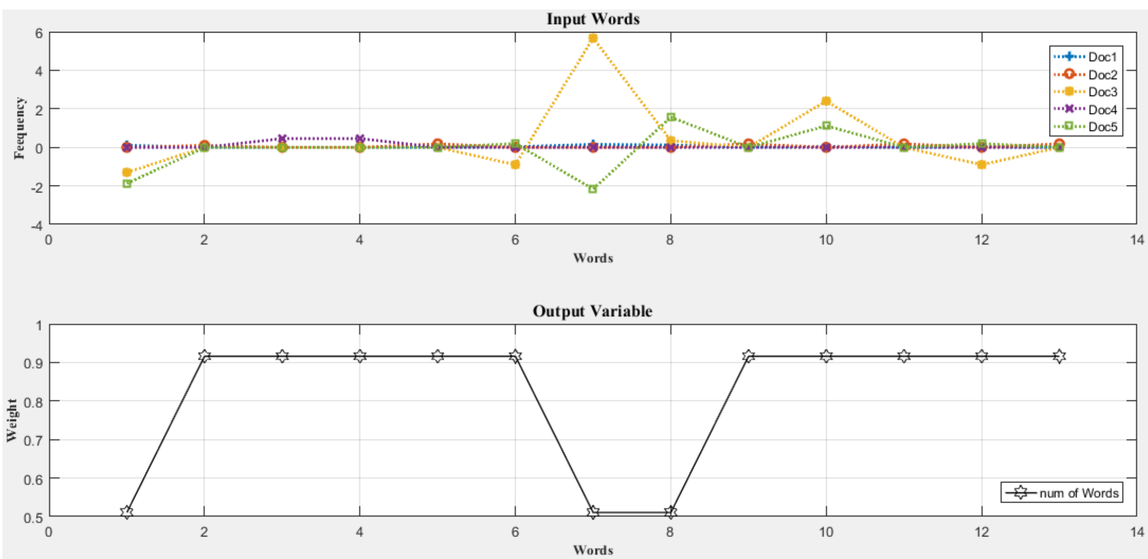


Figure 9: Weight of input and output fuzzy sets.

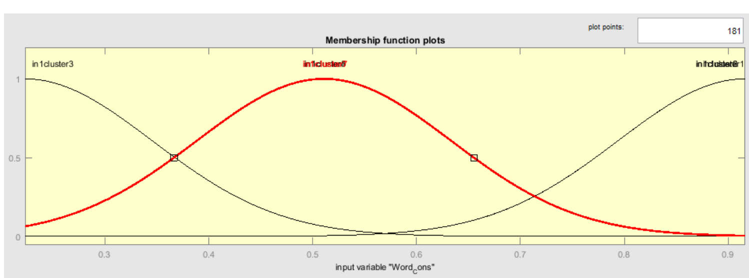


Figure 10: Shape of membership function.

B. Neuro-fuzzy designer: the network is designed automatically So that they are compatible with inputs and outputs through their respective membership functions and its parameters to interpret the input and output map as shown in figure 11.

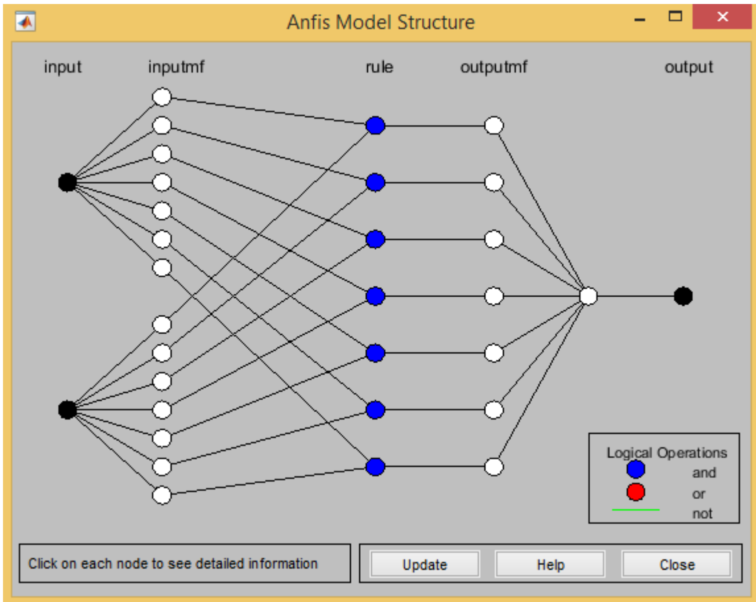


Figure 11: Neural network model structure.

Figures 12 and 13 are shown the procedures of training.

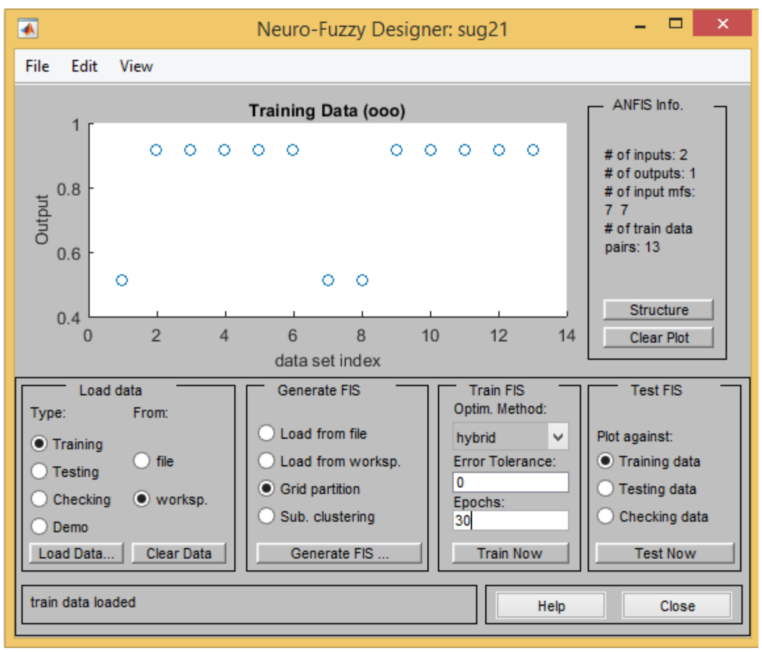


Figure 12: Neuro-fuzzy designer.

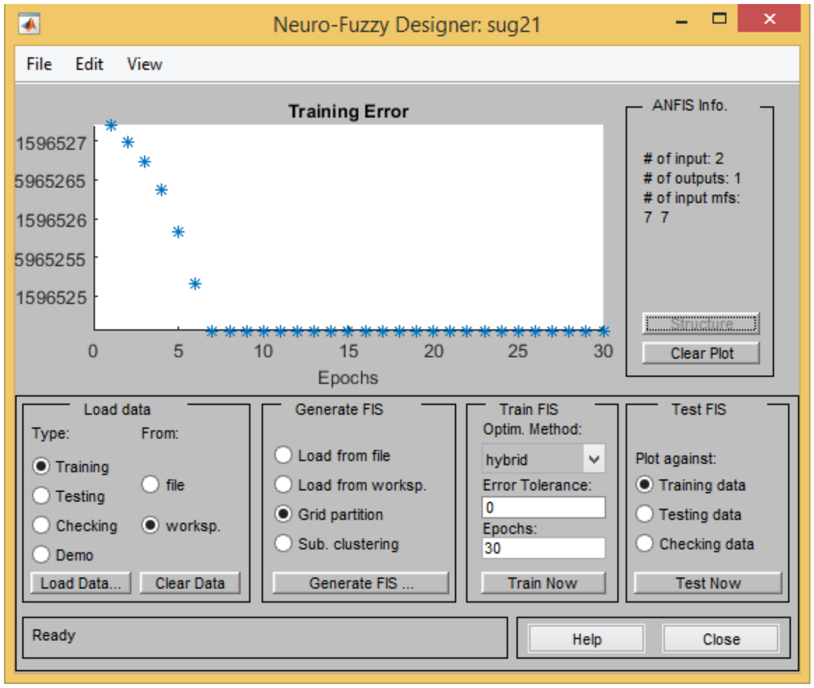


Figure 13: Training processes.

C. Adaptive neuro-fuzzy inference system (ANFIS) model structure is shown in figure 14. Figure 15 shows the corresponding rules extraction.

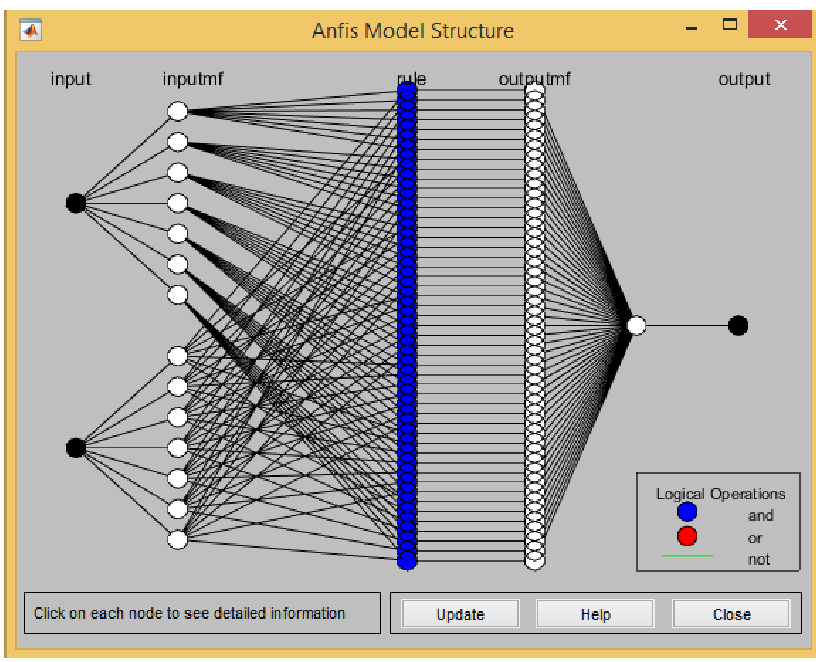


Figure 14: ANFIS model.

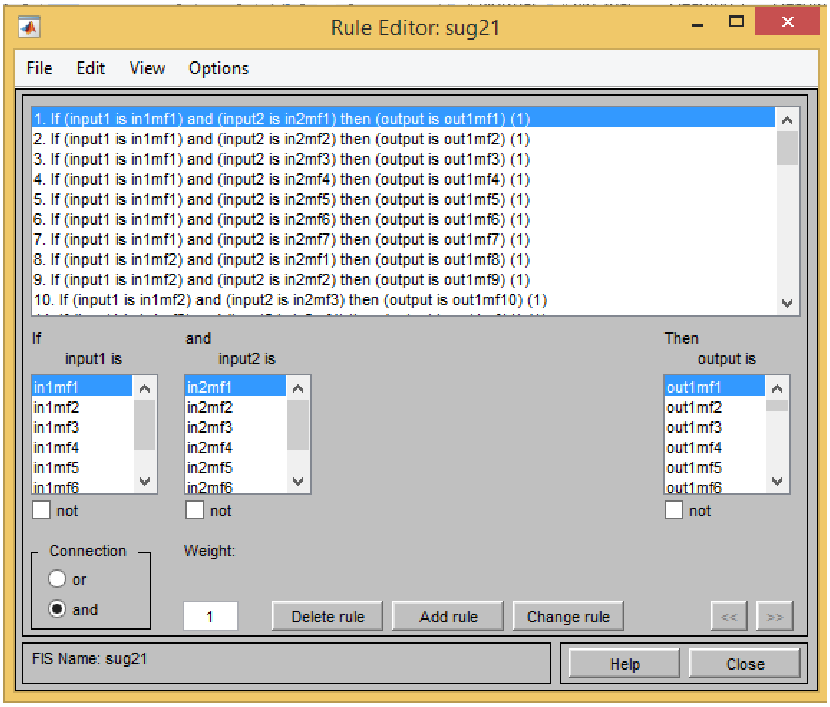


Figure 15: Corresponding rule extraction.

6. Summary generation: by applying the phase 5, after the training of neural network fog and identify the ambiguous rules as shown in figures 16 and 17. The user determines both of the query words and the rank which are used to rank rules according to their importance.

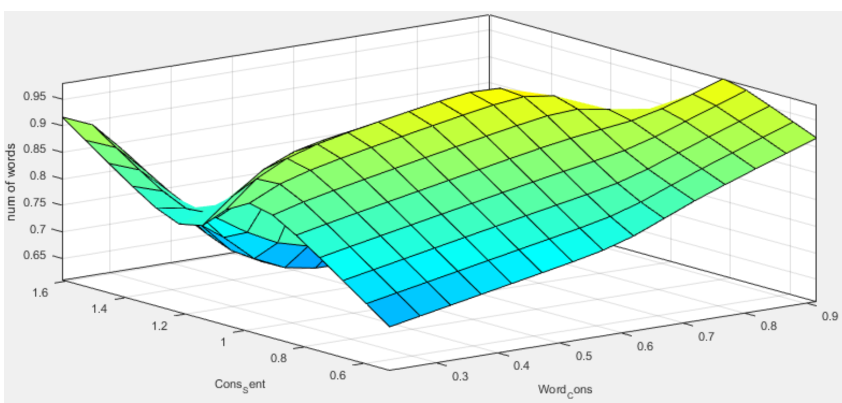


Figure 16: The ANFIS model surface.

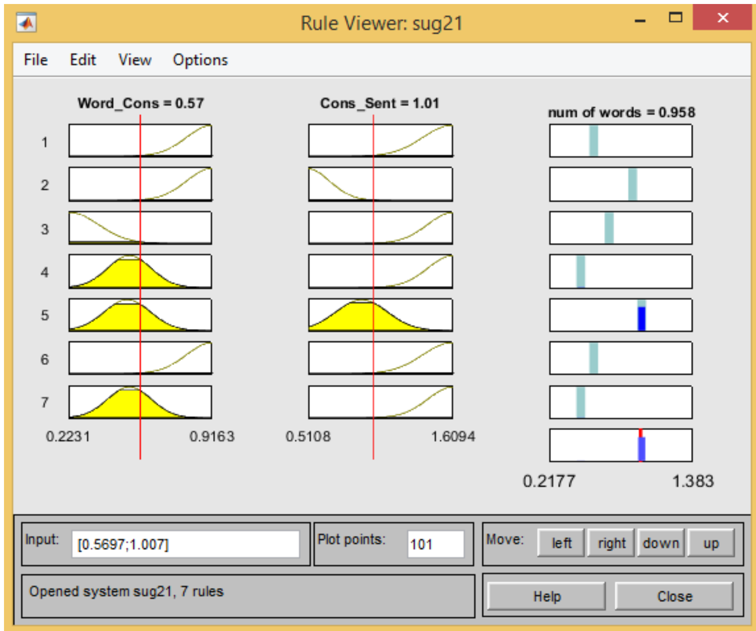


Figure 17: The results of rules.

4.1. The proposed system evaluation

4.1.1. Dataset

The proposed approach is validated using the standard summarization benchmark data sets for summarization, DUC2007 [34]. Table 1 provides a brief summary of the DUC2007 datasets:

Table 1: Description of Dataset

Parameters of Dataset	DUC2007 size
Data Source	AQUAINT
Doc. No. in each clusters	25
Sentence Average No. / Doc.	37.5
Sentence Max. No. / Doc.	125
Sentence Min. No. / Doc.	9
Clusters No.	45
Summary length (in words)	250

4.1.2. Evaluation Matrix

Recall-Oriented Understudy for Gisting Evaluation, or ROUGE [35], is the abbreviation for the method used to assess the suggested system. The ROUGE method counts how many times n-grams appear together in the reference and system summaries. Four alternative measurements are offered by ROUGE: ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S, and ROUGE-SU. The ROUGE-N (N=1 and 2) is selected to evaluate the proposed system because it works well with single documents, ROUGE-N score can be calculated as:

$$ROUGE - N = \frac{\sum_{S \in Summ_{Ref}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in Summ_{Ref}} \sum_{gram_n \in S} Count\ (gram_n)}$$

ROUGE-N assesses the recall of n-grams, where S is the reference sentence, n is the length of the n-gram, $Count_{match}(gram_n)$ is the number of n-grams that occur in the system summary, and $Count\ (gram_n)$ is the number of n-grams that are shared across a set of reference summaries and the system summary.
Additionally, there are three criteria were used to evaluate the summarization known as sensitivity (Sum_{Sen}), positive predictive value (Sum_{PPV}), and accuracy of the summary (Sum_{Acc}), all of them based on the outcomes of proposed summary (Sum_{Prop}), reference summary (Sum_{Ref}), true summary (Sum_{Tru}), true sentences (Sen_{Tru}), and least significant sentences (Sen_{LS}) . The criteria are calculated as:

$$Sum_{Sen} = \frac{|Sen_{Tru}|}{|Sen_{Tru}| + |Sum_{Ref}|}$$

$$Sum_{PPV} = \frac{|Sen_{Tru}|}{|Sen_{Tru}| + |Sum_{Prop}|}$$

$$Sum_{Acc} = \frac{|Sen_{Tru}| + |Sen_{LS}|}{|Sen_{Tru}| + |Sen_{LS}| + |Sum_{Prop}| + |Sum_{Prop}|}$$

5. Results and Discussion

The recommended system was implemented using Matlab version (R2016a) and Windows 8.1. The suggested system's outcomes were compared to those of Particle Swarm Optimization (PSOS)-based summarization and Cat Swarm Optimization (CSOS)-based summarization with regard to two years' worth of DUC datasets [36, 37].Table 2 shows, the controlling parameters of PSOS and CSOS that derived by a number of simulations.

Table 2: The Parameters that Used for PSOS And CSOS Summarization

Parameters	PSOS	CSOS
Population size	50	50
C_1	[0, 2]	---
C_2	[0, 2]	---
V_{min}, V_{max}	[0, 1]	---
W	0.45	---
SMP	---	3
CDC	---	0.2
SRD	---	0.2
Mixture Ratio (MR)	---	0.5
W, C	---	0.5, 4

Based on the summary's content coverage, cohesion, and text readability, the ROUGE-N (N=1, 2 and L) evaluation is conducted. There are three highly correlated with the human judgments types from ROUGE-N are ROUGE-1, which measures the overlap between the unigram between the summary of the proposed system and the human summary and ROUGE-2, which compares the overlap of the bigrams [38].

The more closely the resulting summary resembles the source document sets, the higher the ROUGE measure. The ROUGE-N value is represented by the metrics Precision, Recall, and F-measure. The f-measure is chosen as having higher relevance for the selection of a summary from Table 3, which displays the statistical analysis of F-measure of ROUGE-1, ROUGE-2, and ROUGE-L evaluation metrics for algorithms of PSOS, CSOS, and the proposed system.

Table 3: The Performance Comparison Between PSOS And CSOS F-Measure for DUC2007 Data

Evaluation Metric	**Method**	**F-measure**		
		Worst	**Mean**	**Best**
ROUGE-1	PSOS	0.3916	0.3991	0.40967
	CSOS	0.3908	0.4098	0.4207
	Proposed	0.3989	0.4183	0.4302
ROUGE-2	PSOS	0.0743	0.0758	0.0762
	CSOS	0.0809	0.0881	0.08903
	Proposed	0.08103	0.0890	0.0999
ROUGE-L	PSOS	0.3908	0.4001	0.41127
	CSOS	0.4003	0.4070	0.4229
	Proposed	0.40513	0.4108	0.4311

When comparing the summarization results of the proposed system with the summarization human-generated present in the DUC, it was found that the F-measure for the proposed system is falling within range 0.39 - 0.43 for the ROUGE-1 whereas, it's falling within the range of 0.07-0.1 and 0.39 – 0.43 with respect of ROUGE-2 and ROUGE-L respectively. Further, the other measures (Precision and Recall) for both ROUGE scores are specified in Table 4.

Table: 4. The Recall and Precision of ROUGE-N for DUC2007 data

Evaluation Metric	**Method**	**Recall**	**Precision**
ROUGE-1	PSOS	0.44679	0.37825
	CSOS	0.46158	0.38662
	Proposed	0.46	0.398
ROUGE-2	PSOS	0.0841	0.0697
	CSOS	0.0924	0.0859
	Proposed	0.0987	0.09615
ROUGE-L	PSOS	0.36605	0.45272
	CSOS	0.39261	0.45692
	Proposed	0.39596	0.46479

From the previous table, it's noted that the recall and precision of the proposed system have the best results than other methods in all the standards of the ROUGE.

Similar to how the F-measure of ROUGE-N depends on both recall and precision, the summarization accuracy depends on both sensitivity and PPV score. The F measure value and summary accuracy value are used to validate the proposed system, as shown in table 5.

Table: 5. Sensitivity, PPV and Summary Accuracy of ROUGE-N for DUC2007 Data

Method	Criteria		
	Sum_{Sen}	Sum_{PPV}	Sum_{Acc}
PSOS	0.5	0.3529	0.9808
CSOS	0.5833	0.5	0.9904
Proposed	0.652	0.581	0.9921

6. Conclusion

This paper develops an adaptive neuro-fuzzy inference system, a hybrid intelligent model based on fuzzy logic and neural networks (ANFIS). The advantages of ANFIS were exploited to extract a knowledge base linking the relationship among the important words in documents after clustering process, the semantic concepts of each word and sentences in documents. Hence, the summarization well extracts from the documents based on the query words of the users. The positive results of the proposed system showed the good performance of this system and its importance in the process of automatic summarization. The experimental results are confirmed when comparing the proposed system with both PSOS and CSOS, where the F-measure evaluation appears the proposed system outperforms on PSOS and CSOS summarize by 0.4302, 0.40967 and 0.4207 respectively for ROUGE-1, 0.0999, 0.0762 and 0.08903 respectively for ROUGE-2, and 0.4311, 0.41127 and 0.4229 respectively for ROUGE-L. The main reason why the proposed system performs better than other methods is to clustering the similar concepts of words by using semantic vectors to represent those vague concepts of important words within documents.

References:

1. Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408.
2. Saggion, Horacio, and Thierry Poibeau. "Automatic text summarization: Past, present and future." *Multi-source, multilingual information extraction and summarization*. Springer, Berlin, Heidelberg, 2013. 3-21.
3. Gupta, Vishal, and Gurpreet Singh Lehal. "A survey of text summarization extractive techniques." *Journal of emerging technologies in web intelligence* 2.3 (2010): 258-268.
4. Mendoza, Martha, et al. "Extractive single-document summarization based on genetic operators and guided local search." *Expert Systems with Applications* 41.9 (2014): 4158-4169.
5. Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." *Mining text data*. Springer, Boston, MA, 2012. 43-76.

6. Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis." *IEEE Intelligent Systems* 28.2 (2013): 15-21.
7. Croft, W. Bruce, and John Lafferty, eds. *Language modeling for information retrieval*. Vol. 13. Springer Science & Business Media, 2013.
8. Lewis, Sarah. "Qualitative inquiry and research design: Choosing among five approaches." *Health promotion practice*16.4 (2015): 473-475.
9. Banaszkiewicz, Paul A. "Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty: an end-result study using a new method of result evaluation." *Classic Papers in Orthopaedics*. Springer, London, 2014. 13-17.
10. Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. "Practical bayesian optimization of machine learning algorithms." *Advances in neural information processing systems*. 2012.
11. Tokuda, Keiichi, et al. "Speech synthesis based on hidden Markov models." *Proceedings of the IEEE* 101.5 (2013): 1234-1252.
12. Rush, Alexander M., Sumit Chopra, and Jason Weston. "A neural attention model for abstractive sentence summarization." *arXiv preprint arXiv:1509.00685* (2015).
13. Karpathy, Andrej, et al. "Large-scale video classification with convolutional neural networks." *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.
14. Tsytsarau, Mikalai, and Themis Palpanas. "Survey on mining subjective data on the web." *Data Mining and Knowledge Discovery* 24.3 (2012): 478-514.
15. Hu, Xia, et al. "Unsupervised sentiment analysis with emotional signals." *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013.
16. Mendoza, Martha, et al. "Extractive single-document summarization based on genetic operators and guided local search." *Expert Systems with Applications* 41.9 (2014): 4158-4169.
17. Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." *IEEE Signal Processing Magazine* 29.6 (2012): 82-97.
18. Chopra, Sumit, Michael Auli, and Alexander M. Rush. "Abstractive sentence summarization with attentive recurrent neural networks." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016.
19. Betz, Vaughn, Jonathan Rose, and Alexander Marquardt. *Architecture and CAD for deep-submicron FPGAs*. Vol. 497. Springer Science & Business Media, 2012.
20. Li, Haoxiang, et al. "A convolutional neural network cascade for face detection." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
21. Bezdek, James C., Didier Dubois, and Henri Prade, eds. *Fuzzy sets in approximate reasoning and information systems*. Vol. 5. Springer Science & Business Media, 2012.
22. Kumar, Yogan Jaya, et al. "Multi document summarization based on news components using fuzzy cross-document relations." *Applied Soft Computing* 21 (2014): 265-279.

23. Gandomi, Amir, and Murtaza Haider. "Beyond the hype: Big data concepts, methods, and analytics." *International Journal of Information Management* 35.2 (2015): 137-144.
24. Pedrycz, Witold, ed. *Fuzzy modelling: paradigms and practice*. Vol. 7. Springer Science & Business Media, 2012.
25. Pearl, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.
26. Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." *Mining text data*. Springer, Boston, MA, 2012. 43-76.
27. Abbasi, Ameer Ahmed, and Mohamed Younis. "A survey on clustering algorithms for wireless sensor networks." *Computer communications* 30.14-15 (2007): 2826-2841.
28. Xue, Jian, Jinyu Li, and Yifan Gong. "Restructuring of deep neural network acoustic models with singular value decomposition." *Interspeech*. 2013.
29. Neale, M. C. C. L., and Lon R. Cardon. *Methodology for genetic studies of twins and families*. Vol. 67. Springer Science & Business Media, 2013.
30. Mehrotra, Rishabh, et al. "Improving lda topic models for microblogs via tweet pooling and automatic labeling." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013.
31. Levy, Omer, and Yoav Goldberg. "Neural word embedding as implicit matrix factorization." *Advances in neural information processing systems*. 2014.
32. Gurney, Kevin. *An introduction to neural networks*. CRC press, 2014.
33. Guitart-Masip, Marc, et al. "Go and no-go learning in reward and punishment: interactions between affect and effect." *Neuroimage* 62.1 (2012): 154-166.
34. Y. Chali, S.A. Hasan. Query-Focused Multi-Document Summarization: Automatic Data Annotations and Supervised Learning Approaches. Natural Language Engineering, 18(1): 109-145. (DUC06, DUC07)
35. Campr, Michal, and Karel Ježek. "Comparing semantic models for evaluating automatic document summarization." *International Conference on Text, Speech, and Dialogue*. Springer, Cham, 2015.
36. Javanbarg, Mohammad Bagher, et al. "Fuzzy AHP-based multicriteria decision making systems using particle swarm optimization." *Expert Systems with Applications* 39.1 (2012): 960-966.
37. R. Rautray, R.C. Balabantaray, Cat swarm optimization based evolutionary framework for multi document summarization, Phys. A: Stat. Mech. Appl. 477 (2017) 174–186.
38. Ganesan, Kavita, ChengXiang Zhai, and Evelyne Viegas. "Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions." *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012.