International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Semantic Similarity Based Automatic Document Summarization Method

K. Srinivasa Rao, D.S. R. Murthy, Gangadhara Rao Kancherla

Abstract: ***Document summarization is the process of generating the summary of the documents gathered from the web sources. It reduces the burden of web readers by reducing the necessity of reading the entire document contents by generating the short summary. In our previous research work this is performed by introducing the method namely Noun weight based Automated Multi-Document Summarization method (NW-AMDSM). However the previous research work doesn't concentrate on the semantic similarity which might reduce the accuracy of the summarization outcome. This is resolved in the proposed research method by introducing the method namely Semantic Similarity based Automatic Document Summarization Method (SS-ADSM). In this research work, multi document grouping is done is based on semantic similarity computation, thus the document with similar contents can be grouped more accurately. Here the semantic similarity computation is performed with the help of word net analyzer. The document grouping is done by introducing the modified FCM clustering algorithm. Finally hybrid neuro fuzzy genetic algorithm is introduced to perform the automatic summarization. The numerical analysis of the proposed research method is conducted in the matlab simulation environment and compared with other research methods in terms various performance metrics. The simulation analysis proved proposed method tends to have better performance in terms of increased accuracy of document summarization outcome.***

Keywords: Document summarization, semantic similarity, fuzzy classifier, genetic algorithm, similarity grouping

I. INTRODUCTION

Increased web services leads to generation of mass volume of documents in the internet servers which might be required by various industries and individuals to gain the usable knowledge [1]. Finding the required documents from the millions of documents will be more difficult task which would require more computational overhead and computational cost[2]. This can be avoiding by creating the short summary of every documents, so that web users can identify their required documents more easily instead of reading the entire document contents [3]. Text summarization is the process of generating the short overview of the documents in one or two pages with essential information about the corresponding document [4]. Summarization provides flexible environment for the different applications by reducing their computation overhead and processing time with shorter description about the entire documents [5].

Summarization involves many experts to read and predict the essential information about the corresponding documents in order to generate the most valuable summarized outcome [6]. As the millions of documents are present in the web servers, it requires millions of years for the experts to generate the summarized content [7]. This can be resolved by integrating the suitable artificial techniques with the web servers to generate the proper and accurate summarized result for every document. This process of generating the summarized content through machine learning algorithm is called as automatic summarization [8]. This automatic summarization process doesn't require human involvement and ensured to generate the summarization outcome in the shorter time period [9].

There are various automatic summarization techniques are proposed earlier to enhance the summarized outcome [10]. Each technique tends to follow different procedures and techniques to generate the summarized outcome. Summarization of the documents are done with the concern of the varying factors such as more frequent phrases in the document, most repeated sentences, sentence measurement in terms of presence of nouns and verbs and so on [11]. The common goal behind all these technique is to weigh the sentences present in the documents and selects the most valuable sentence with high weight values to generate the summarized outcome. The automatic summarization leads to ensure the most efficient summarized outcome.Automatic summarization outcome can be enhanced more by considering the semantic meaning of the sentences present within the document [12]. Semantic similarity based summarization would generate more meaningful summarization outcome than by considering only the number of repetitions. Semantic similarity of sentences present within the document would lead to learn the exact meaning of the documents. From the meaning, authors can weigh the more relevant sentences as high. Thus the accurate and reliable summarization outcome can be attained. In this research work, Semantic Similarity based Automatic Document Summarization Method is introduced. The main goal of the research work is to introduce the technique that can lead to automated and enhanced summarized outcome. Consideration of semantic meaning of the sentences present in the document tends to have better outcome than the existing techniques. The detailed discussion of proposed research technique is given in the following sub sections.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
K Sriniva sa Rao*, Research Scholar, JNTUK Asst professor Dept of IT, RVR & JC College of engineering, Guntur, Andhra Pradesh, India. Email: kongarasmail@gmail.com
D S R Murthy, Professor, Dept. of CSE, Geethanjali College of Engineering and Technology. Hyderabad, Telangana, India Email: dsrmurthy.1406@gmail.com
Gangadhara Rao Kancherla, Professor, Dept. of CSE, ANU, , Guntur, Andhra Pradesh, India. Email: kancherla123@gmail.com

II. RELATED WORKS

Liu et al [13] introduced the incremental clustering approach for the summarization. This research technique enhances the summarization outcome by clustering the most similar documents together with the help of visualization interfaces. This method supports the incremental update of document summarization outcome. This performance assessment of this work tends to have better performance in terms of increased scalability and accuracy.

Chen et al [14] introduced the method namely topic anatomy whose intention is to generate the summarized content based on topic interrelationship. The summary values are extracted from the given input documents and their topic similarity is measured to know the final outcome. The overall evaluation of the research technique proved that this method tends to have better performance over the TDT4 database.

Sharifi et al [15] attempted to perform multi document summarization by combining the summaries of multiple single documents. This research technique tends to consume more computational time for the summary generation. The performance evaluation of the research method is tends to prove that the proposed research technique guarantees the optimal outcome than the existing research techniques.

Shimada et al [16] introduced the research oriented summarization technique to generate the short description for the lecture notes and studies. This technique concentrates on both text and images to generate the summarized copies. The main goal of this research work is to provide the support to the readers by finding the most important pages with essential points based on summarized contents. The performance assessment of this works proved that the summarization process leads to better outcome with increased preview achievement ratio.

Sun et al [17] attempted to generate the summaries for the scientific documents. This research work focus on the semantic meaning of the documents to ensure the accurate and correct summary creation. The optimal summarization outcome is ensured in this research work by introducing the iterative algorithm where dynamic summary generation support will be given. This research method tends to provide the optimal outcome of summary generation by creating the most accurate and meaning summary generation.

Dilawari et al [18] focused on video summary generation to provide the proper description labelling for the online stored videos. This is done by introducing the method namely innovative joint end to end solution. The main goal of this research work is to introduce the method that can ensure the optimal video summary creation. This method enables users to find the most relevant videos. This is achieved by finding the discrimination between the relevant and irrelevant contents of the videos that are uploaded into the system.

Chen et al [19] adapted the recurrent neural network technique to generate the summarization of the news broadcasted in the multiple locations of world. This is done by measuring the word cues of broadcasted new documents by focusing on the key words present within the news content. The performance analysis of the research work is carried out on the various domains in terms of multiple contents present in the news. The assessment of the research work proved proposed method leads to increased accuracy in news summary creation.

Chen et al [20] introduced the information distillation framework to generate the summaries of larger documents. This research work mainly focuses on the essential models that tend to have better importance spoken words about the summarization contents. This work utilizes the various back ground information to generate the summarization outcome. The research analysis tends to prove that the proposed D-E-V model leads to better outcome than the existing techniques in terms of increased accuracy.

Liu et al [21] attempted to predict the summarization for the movie rating and reviews. This is done by adapting the sentiment classification framework. This is done by introducing the method namely latency semantic analysis whose main goal is to predict the optimal features to ensure the accurate summarized outcome.

Zhang et al [22] introduced the summarization framework for the tweet model. This method attempts to generate the summarized content for the tweets to learn about the most important prospects of the network. This method provides support for the both text based tweets and the audio based tweets. The overall evaluation is carried out n the 100 topic datasets from which it is proved that the proposed research technique leads to enhance the optimal outcome for the 100 topic datasets.

Wang et al [23] introduced the method namely sumblr to handle the continuous generated tweets for the different time period. The main goal of the research work is to introduce the framework that can generate the summarized outcome for the posted tweets. From this summarized outcome, optimal result can be obtained in terms of increased accuracy and the performance terms.

III. SEMANTIC SIMILARITY BASED AUTOMATIC DOCUMENT SUMMARIZATION

In this research work, multi document grouping is done is based on semantic similarity computation, thus the document with similar contents can be grouped more accurately. Here the semantic similarity computation is performed with the help of word net analyzer. The document grouping is done by introducing the modified FCM clustering algorithm. Finally hybrid neuro fuzzy genetic algorithm is introduced to perform the automatic summarization.

3.1. SEMANTIC SIMILARITY COMPUTATION

Semantic similarity of documents provides the conceptual meaning by using which accurate decision making can be made. Semantic similarity of documents helps us finding the most important information about the document, thus the accurate and effective summarization can be created. In this work automatic summarization is performed based on semantic similarity of documents. Here initially semantic meaning of the each document will be learned based on which document clustering would be performed. The document clustering will cluster the multiple documents that speak about the similar topic together. However finding the semantic meaning is more difficult task which can be predicted with the help of word net analyser. In this work, features of wordnet and the Wikipedia is utilized to learn the semantic meaning that is context of the particular document.

The steps covered in the semantic based document grouping process are given below:

- Verb, noun identification process
- Semantic similarity prediction based on ontology

Here initially documents will be processed to predict the verbs and nouns involved with that document. The predicted words will then be given as input to the wordnet analyser which will resultant with the related terms such as meanings, opposites and so on. If those similar contextual words are found in the other documents then those documents will be considered as similar document to the corresponding processed document. The overall flow of the proposed research work is given in the following figure 1.

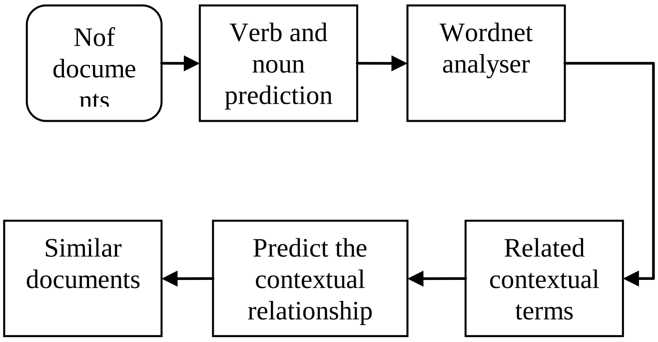

Figure 1. Semantic similarity computation process

The detailed explanation processing of each component in the figure 1 is discussed in the following subsections.

3.1. DATASET CONSIDERED

In this work, two news article datasets, originating from BBC News, provided for use as benchmarks for summarization purpose. All rights, including copyright, in the content of the original articles are owned by the BBC.

- Consists of 2225 documents from the BBC news website corresponding to stories in five topical areas from 2004-2005.
- Class Labels: 5 (business, entertainment, politics, sport, tech)

The datasets have been pre-processed as follows: stemming (Porter algorithm), stop-word removal (stop word list) and low term frequency filtering (count < 3) have already been applied to the data. The files contained in the archives given above have the following formats:

- *.mtx: Original term frequencies stored in a sparse data matrix in Matrix Market format.
- *.terms: List of content-bearing terms in the corpus, with each line corresponding to a row of the sparse data matrix.
- *.docs: List of document identifiers, with each line corresponding to a column of the sparse data matrix.
- *.classes: Assignment of documents to natural classes, with each line corresponding to a document.
- *.urls: Links to original articles, where appropriate.

3.1.1. VERB, NOUN EXTRACTION

To extract the semantic meaning of the documents, initially content in the documents are segmented into sentences and then nouns and verbs present in the sentences will be extracted. Based on these extracted nouns and verbs, semantic similarity of the document will be identified. In this work, Hidden Markov Model is utilized for finding the verbs and nouns present in the document. Here HMM is utilized to predict the hidden factors from the derived factors. HMM representation of the input sentences are given as follows:

Hidden term: T sequence term

Observed term: word sequence

Transition probability:

$$a_{j-1,j} = P(ts_j \mid ts_{j-1})$$

Result probability:

$$b_j = P(w_j \mid ts_j)$$

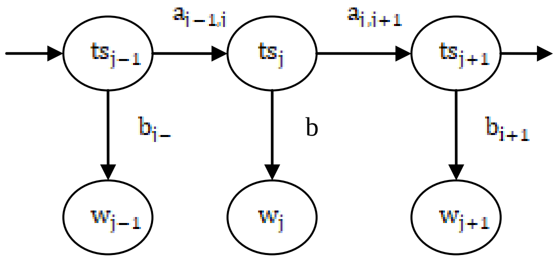

Figure 2. Noun, Verb extraction using HMM model

Term tagged sequence can be represented as

$$\hat{T} = ts_1, ts_2, \ldots, ts_n$$

This is extended as

$$\underset{T}{\text{argmax}}\; P(w_1, w_2, \ldots, w_n \mid ts_1, ts_2, \ldots, ts_n) * P(ts_1, ts_2, \ldots, ts_n)$$

The verbs and nouns present in the sentences would lead to easy prediction of the contextual words. The probability of finding the word is given as follows:

$$P(W|T) = \prod_{j=1}^{n} P(W_j|ts_j)$$

ased on this probability value P(T) can be computed as like given below:

$$P(T) = P(ts_1) * P(ts_2|ts_1) * \ldots * P(ts_n|ts_1, ts_2, \ldots, ts_n)$$

These probability values will be compared with the wordnet analyser to extract the contextual meaning words. This will retrieve the similar words, based on which document grouping will be performed.

3.1.2. MULTIDOCUMENT GROUPING BASED ON SEMANTIC SIMILARITY

The contextual meaning of the verbs and nouns of corresponding document will be compared with the other documents to find the similarity level. In this work modified FCM algorithm is introduced for the clustering multiple documents based similarity. The similarities of the documents are identified based semantic similarity. The mutual similarity calculation procedure is given in the following equation

$$\text{Mutual Similarity} = \frac{\log(p(c_j)p(c_k))}{p(c_j, c_k)}$$

$$P(c_j) = \frac{w_c}{w}$$

Where c → concept

P() → probability function

P(cj, ck) → joint probability distribution

P(cj) → probability of specific concept in the document
To find the exact similar documents, it is required to implement the fuzzy relationship values for the different relationships and conceptual vectors. The fuzzy membership evaluation is given in the following equation

$$\mu_i(c_k) = \alpha * P(c_j, c_k) \log_2 \left(\frac{P(c_j)P(c_k)}{P(c_j, c_k)} \right)$$

Based on these similarity modified FCM algorithm will group the documents that similar together. The working process of modified SVM is given below:

1. Initialize the clustering parameters
   Initialize number of clusters c
   Initialize the parameter m
   Initialize the centroid vector
   $V = [v_1, v_2, \ldots, v_c]$
   Initialize $\varepsilon = 0$
2. Calculate the fuzzy membership function u_{ij}

$$u_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d(x_j, v_i)}{d(x_j, v_k)} \right)^{2/(m-1)} \right)^{-1}$$

3. Calculate the centroid v_i

$$v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} x_j}{\sum_{j=1}^{n} \mu_{ij}^{m}}$$

4. Update the membership function

$$\mu_{ij} = \frac{\mu_{ij}^{m} S_{ij}^{n}}{\sum_{k=1}^{c} \mu_{kj}^{m} s_{kj}^{n}}$$

5. Update centroids

$$v_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{m} x_j}{\sum_{j=1}^{n} \mu_{ij}^{m}}$$

The above steps will be repeated until the end criterion met. End criterion is given in the following equation

$$|v_{new} - v_{old}| < \varepsilon$$

The above algorithm leads to clustered documents which shares similar contextual meaning with each other. For these group of documents summarization will be generated which is explained in the following section.

3.3. AUTOMATIC SUMMARY CREATION USING NEURO FUZZY GENETIC ALGORITHM

In this work neuro fuzzy genetic algorithm is introduced for the automatic summarization process. The Neuro fuzzy genetic algorithm leads to optimal summarization creation than the conventional algorithm. In this work, genetic algorithm is hybridized with the ANFIS to ensure the optimal weight update for hidden layers. Thus the accurate prediction summarization can be made. In this work, six layers of neurons in GANFIS is considered for the automatic summarization.

Neuron layer 1 → This layer depicts the input given to the system. That is set of sentences are given as input to the GANFIS to predict the optimal sentences that can make summarization outcome accurate. The layer will resultant with the linguistic corresponding to each input neuron.

Neuron layer 2 → This layer attempts to generate the membership value for the each input which is given as linguistic variable. The membership calculation is done as follows:

$$F_2(x_i) = \mu_{Ai}(x)$$

Neuron layer 3 → In this layer, weight values of each combination of sentences is generated as firing strength. The weight values assigned in this layer will decide the summarization outcome. The weight association is given as follows:

$$f_3(x_i) = \mu_{Ai}(x) * \mu_{Bi}(x) * \mu_{Ci}(x)$$

Neuron layer 4 → In this layers, weight values calculated in the previous layer will be normalized to increase the firing strength. The normalized weight values is calculated as like given below:

$$f_4(x_i) = \frac{w_1}{w_1 + w_2 + w_3}$$

This equation can be summarized as like given below:

$$f_4(x_i) = \frac{w_1}{\sum_{j=1}^{3} w_j}$$

Neuron layer 5 → In this layer output computation will be done based on normalized weight values. The output computation is given as follows:

$$f_5(x_i) = f_4(x_i) * R_{out}(x_i)$$

Neuron layer 6 → This layer represents the final outcome of summarization decision. This is computed by calculating the aggregation value of all outcomes obtained from previous layer.

$$Y = \sum_{i=1}^{n} f_5(x_i) = \sum_{i=1}^{n} (f_4(x_i) * R_{out}(x_i))$$

The result obtained from the above equation will be in the crisp form which will represent the final summarized outcome. In our work crisp outcome can segmented into three type's namely high importance, medium important and low importance based on their importance level for considering in the summarization outcome.

The prepared GANFIS model is then used to order new sentences to one of its group, that is, summary or non-summary sentence. The GANFIS model yield, which is the anticipated sentence score is utilized to set the grouping rule for GANFIS to characterize the sentence into paired worth (1 or 0). Sentences which are ordered to class '1' speaks to outline sentence, while sentences which are arranged to class '0' speaks to non-rundown sentence. The limit worth used to characterize the anticipated yield to one of these two classes were chosen dependent on exploratory perception which gave us the least root mean square error (RMSE).

IV. EXPERIMENTAL RESULTS

The implementation and numerical assessment of the proposed research work is done in the matlab simulation environment.

The performance of proposed Semantic Similarity based Automatic Document Summarization Method (SS-ADSM) is compared to our previous work Noun weight based Automated Multi-Document Summarization method (NW-AMDSM) [24] and Multi-document Automated Summarization Method (MDASM) and existing Fuzzy Rule based Automated Summarization Method (FRASM), Automatic text structuring and summarization (ATSS), Machine Learning based Automatic Text Summarization (MLATS) to assess the performance improvement. The performance metrics that are considered in this research work for the assessment are precision, recall, f-measure and accuracy.

$$\text{Precision} = \frac{T_p}{(T_p + F_p)}$$

Recall is calculated based on the formula

$$\text{Recall} = \frac{T_p}{(T_p + F_n)}$$

Accuracy is calculated based on the formula

$$\text{Accuracy} = \frac{T_p}{(T_p + F_p + F_n)}$$

where

T_P – True Positive (Correct result),
T_N – True Negative (Correct absence of result),
F_P– False Positive (Unexpected Result),
F_N– False Negative (Missing result).
F – Measure is calculated based on the formula

$$F=2.\frac{\text{precision.recall}}{\text{prrecision+recall}}$$

The simulation results for the evaluation of the proposed approach against various performance measures like Precision, Recall, Accuracy and F-Measure. The simulation values are shown in the following table 1.

Table 1. Performance Metric values

Performance Metrics	Methods					
	ATSS	MLATS	FRASM	MDASM	NW-AMDSM	SS-ADSM
Accuracy	83.91	86.67	93	94.5	97	98.2
Precision	76.8	77.53	85.6	91	93	95.6
Recall	77.19	78.96	86	92	95	96
F-Measure	76.9	78.2	87	94	96.3	97.3

Accuracy: The proposed SS-ADSM produced better accuracy rate which is shown in Fig 3. When the number of concepts increases the accuracy of the result is increases.

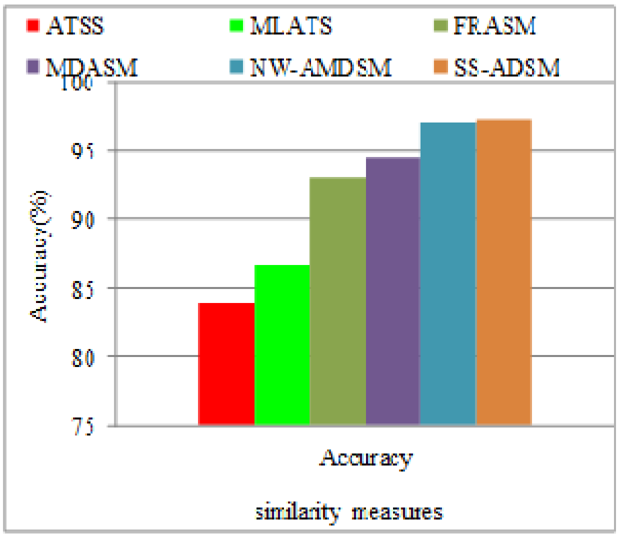

Fig 3: Accuracy comparison of similarity measures

The automatic summarization outcome of the proposed research work SS-ADSM is done in terms of accuracy parameter which will be compared with the existing methods. This comparison is shown in the above figure 3 where proposed and existing methodology is shown in the x axis and the accuracy value is shown in the y axis. From this comparison it is proved that the proposed SS-ADSM shows 1.2% increased accuracy than NW-AMDSM, 3.7% better than the MDASM, 5.2% better than the FRASM, 11.53% better than MLATS and 14.29% better than the ATSS.

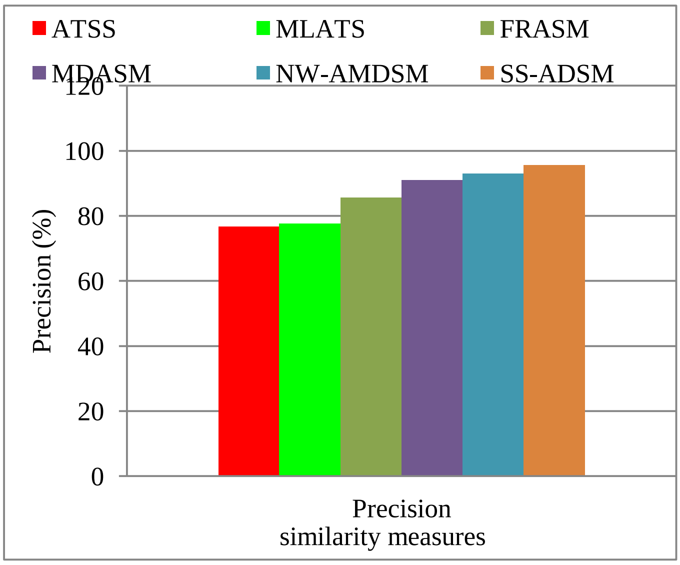

Fig 4: Precision comparison of similarity measures

The automatic summarization outcome of the proposed research work SS-ADSM is done in terms of precision parameter which will be compared with the existing methods. This comparison is shown in the above figure 4 where proposed and existing methodology is shown in the x axis and the precision value is shown in the y axis. From this comparison it is proved that the proposed SS-ADSM shows 2.6% increased precision than NW-AMDSM, 4.6% better than the MDASM, 10% better than the FRASM, 18.07% better than MLATS and 18.8% better than the ATSS.

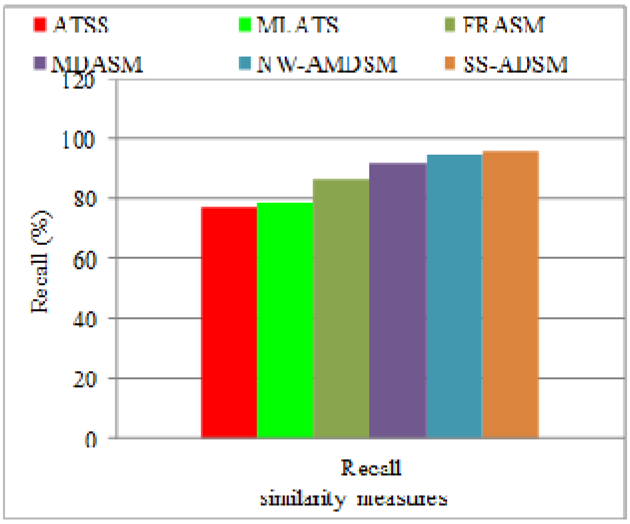

Fig 5: Recall comparison of similarity measures

The automatic summarization outcome of the proposed research work SS-ADSM is done in terms of recall parameter which will be compared with the existing methods. This comparison is shown in the above figure 5 where proposed and existing methodology is shown in the x axis and the recall value is shown in the y axis. From this comparison it is proved that the proposed SS-ADSM shows 1% increased recall than NW-AMDSM, 4% better than the MDASM, 10% better than the FRASM, 17.04% better than MLATS and 18.81% better than the ATSS.

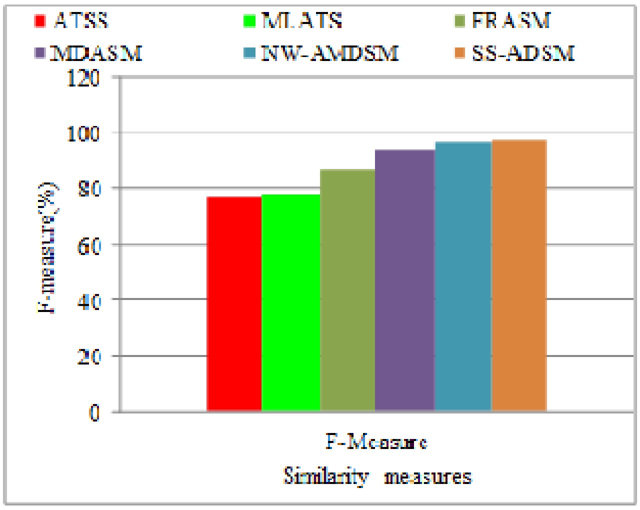

Fig 6: F-measure comparison of similarity measures

76.9	78.2	87	94	96.3	97.3

The automatic summarization outcome of the proposed research work SS-ADSM is done in terms of F-Measure parameter which will be compared with the existing methods. This comparison is shown in the above figure 6 where proposed and existing methodology is shown in the x axis and the F-Measure value is shown in the y axis. From this comparison it is proved that the proposed SS-ADSM shows 1% increased f-measure than NW-AMDSM, 3.3% better than the MDASM, 10.3% better than the FRASM, 19.1% better than MLATS and 20.4% better than the ATSS.

V. CONCLUSION

In this research work, multi document grouping is done is based on semantic similarity computation, thus the document with similar contents can be grouped more accurately. Here the semantic similarity computation is performed with the help of word net analyzer. The document grouping is done by introducing the modified FCM clustering algorithm. Finally hybrid neuro fuzzy genetic algorithm is introduced to perform the automatic summarization. The numerical analysis of the proposed research method is conducted in the matlab simulation environment and compared with other research methods in terms various performance metrics. The simulation analysis proved proposed method tends to have better performance in terms of increased accuracy of document summarization outcome.

REFERENCE

1. Wei, F., Qin, H., Ye, S., & Zhao, H. (2018, December). Empirical study of deep learning for text classification in legal document review. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 3317-3320). IEEE.
2. Fagan, J. L. (2017, August). Automatic P hrase indexing for document retrieval: an examination of syntactic and non-syntactic methods. In ACM SIGIR Forum (Vol. 51, No. 2, pp. 51-61). ACM.
3. Khan, A., Salim, N., & Kumar, Y. J. (2015). A framework for multi-document abstractive summarization based on semantic role labelling. Applied Soft Computing, 30, 737-747.
4. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.
5. Chopra, S., Auli, M., & Rush, A. M. (2016, June). Abstractive sentence summarization with attentive recurrent neural networks. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 93-98).
6. Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, L., & Shazeer, N. (2018). Generating wikipedia by summarizing long sequences. arXiv preprint arXiv:1801.10198.
7. Pérez-Pérez, M., Pérez-Rodríguez, G., Blanco-Míguez, A., Fdez-Riverola, F., Valencia, A., Krallinger, M., & Lourenço, A. (2017, April). Benchmarking biomedical text mining web servers at BioCreative V. 5: the technical Interoperability and Performance of annotation Servers-TIPS track. In Proceedings of the BioCreative V. 5 Challenge Evaluation Workshop (pp. 19-27).
8. Hassan, M., & Hill, E. (2018, September). Toward automatic summarization of arbitrary java statements for novice programmers. In 2018 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 539-543). IEEE.
9. Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. Artificial Intelligence Review, 47(1), 1-66.
10. Campr, M., & Ježek, K. (2015, September). Comparing semantic models for evaluating automatic document summarization. In International Conference on Text, Speech, and Dialogue (pp. 252-260). Springer, Cham.
11. Nallapati, R., Zhai, F., & Zhou, B. (2017, February). Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In Thirty-First AAAI Conference on Artificial Intelligence.
12. Liu, F., Flanigan, J., Thomson, S., Sadeh, N., & Smith, N. A. (2018). Toward abstractive summarization using semantic representations. arXiv preprint arXiv:1805.10399.
13. Liu, C. Y., Chen, M. S., & Tseng, C. Y. (2015). Incrests: Towards real-time incremental short text summarization on comment streams from social network services. IEEE Transactions on Knowledge and Data Engineering, 27(11), 2986-3000.

14. Chen, C. C., & Chen, M. C. (2010). TSCAN: A content anatomy approach to temporal topic summarization. IEEE transactions on Knowledge and Data Engineering, 24(1), 170-183.
15. Sharifi, B. P., Inouye, D. I., & Kalita, J. K. (2013). Summarization of twitter microblogs. The computer journal, 57(3), 378-402.
16. Shimada, A., Okubo, F., Yin, C., & Ogata, H. (2017). Automatic Summarization of Lecture Slides for Enhanced Student PreviewTechnical Report and User Study. IEEE Transactions on Learning Technologies, 11(2), 165-178.
17. Sun, X., & Zhuge, H. (2018). Summarization of scientific paper through reinforcement ranking on Semantic Link Network. IEEE Access, 6, 40611-40625.
18. Dilawari, A., & Khan, M. U. G. (2019). ASoVS: Abstractive Summarization of Video Sequences. IEEE Access, 7, 29253-29263.
19. Chen, K. Y., Liu, S. H., Chen, B., Wang, H. M., Jan, E. E., Hsu, W. L., & Chen, H. H. (2015). Extractive broadcast news summarization leveraging recurrent neural network language modeling techniques. IEEE Transactions on Audio, Speech, and Language Processing, 23(8), 1322-1334.
20. Chen, K. Y., Liu, S. H., Chen, B., & Wang, H. M. (2017). An information distillation framework for extractive summarization. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(1), 161-170.
21. Liu, C. L., Hsaio, W. H., Lee, C. H., Lu, G. C., & Jou, E. (2011). Movie rating and review summarization in mobile environment. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(3), 397-407.
22. Zhang, R., Li, W., Gao, D., & Ouyang, Y. (2012). Automatic twitter topic summarization with speech acts. IEEE transactions on audio, speech, and language processing, 21(3), 649-658.
23. Wang, Z., Shou, L., Chen, K., Chen, G., & Mehrotra, S. (2014). On summarization and timeline generation for evolutionary tweet streams. IEEE Transactions on Knowledge and Data Engineering, 27(5), 1301-1315.
24. K. Srinivasa Rao, D. S. R. Murthy,Gangadhara Rao Kancherla(2018). Automated Multi Document Summarization Framework to Enhance the Readers Knowledge, Jour of Adv Research in Dynamical & Control Systems, 10(3),422-433.

AUTHORS PROFILE

Mr. Srinivasa Rao Kongara currently with RVR&JC College Of Engineering, Guntur, A.P, INDIA. Working as Asst. Professor in the department of Information Technology. He has 12 years of experience in Teaching. He is pursuing his Ph.D in Computer Science & Engineering from JNTUK, Kakinada, ,B.Tech (CSE) from JNTU affiliated college in 2004 and M.Tech (CSE)from SRM University, Chennai, Tamilnadu in the year 2006..His areas of Interests are Information Retrieval Systems, Database Management Systems, Distributed Databases, Natural Language Processing, Machine Learning , Data Mining and Computer Networks.

Dr. D. S. R. Murthy did, M. Tech. (CSE, Osmania University) and Ph.D. (CSE, JNTUH). He is presently working as Professor in Computer Science and Engineering and Controller of Examinations, Geethanjali College of Engineering and Technology, Hyderabad and worked in various positions in REC, Warangal; JNTUCE, Anantapur; Icfaian Foundation, Hyderabad; SNIST, Hyderabad. He is a Member of IEEE, Member of IDES, Fellow of IE (I), Fellow of IETE, Senior Life Member of Computer Society of India, Life Member of ISTE, Life Member of SSI, Member of IACSIT, DOEACC Expert member and Chartered Engineer (IE (I) & IETE). He published a text book on C Programming & Data Structures. He published no. of research papers in international journals. Currently He is guiding 6 Research Scholars for their PhD. He is Active Reviewer of various international journals and Sai Conference, London, UK. His research interests are Image Processing, Image Cryptography and Network Security, Data Mining and Machine Learning.

Dr.K.Gangadhara Rao, received the Ph.D. degree from ANU, Guntur in 2011. He has 27 years of teaching experience and he is presently working as a Professor in CSE department, ANU, Guntur. He has more than 20 Publications in various International Journals and Conferences. He is guiding 5 Research Scholars for their PhD and 3 Scholars are awarded with PhD. His research interest includes Cloud Computing, Computer Networks, Software Engineering, Operating Systems data mining, Machine Learning and Text Mining.