Volume 65, Issue 1, 2021

Journal of Scientific Research

Institute of Science,
Banaras Hindu University, Varanasi, India.

Comparative Assessment of Extractive Summarization: TextRank, TF-IDF and LDA

Ujjwal Rani[*] and Karambir Bidhan

Dept. of Computer Science Engineering, University Institute of Engineering &Technology, Kurukshetra University, Kurukshetra, Haryana.
ujjwalpradhan16@gmail.com*, karambir2015@kuk.ac.in

Abstract: **Automatically generating a shorter version of text documents referred to as text summarization. It is an effective method of finding important details from the documents. There is a massive increment in the data worldwide because of rapid growth rate of the internet. It becomes difficult to manually summarize large documents by human beings. Automatic Text Summarization is an approach of NLP which reduces the time and efforts of the human being to produce a summary. There are various approaches to summarize the data. This paper provides a comparative study over the three approaches namely TF-IDF, TextRank, and Latent Dirichlet Allocation (LDA). The comparison is made by using three different types of datasets like reviews of documents, news articles, legal text, etc. The result shows the best-suited approach for the complexity oriented text inputs. Also, the results are evaluated using ROUGE measures.**

Index Terms: **Text Summarization, Extractive, Term Frequency-Inverse Document Frequency (TF-IDF), TextRank, Latent Dirichlet Allocation (LDA).**

I. Introduction

The increase in the amount and variety of data leads to problems in the data handling and determining the relevancy of the data for the users. Whenever a human need to lessen the data collected by him, the data is cut short without making any compromise with the text details. But manual summarization of a huge amount of is not feasible and time-consuming. Also, every human being has its kind of understanding and learning power, so for a particular document, there is a possibility of generation of more than one human-produced summaries. As the complexity of the data in the document increases it also becomes difficult for the human being to understand the text and generate a precise summary. Sometimes there is a need for a prior knowledge base for the summarization of documents. A person of average knowledge might not be able to make an efficient summary of text data which is complex like medical records, scientific papers, biological details, legal documents, etc. Due to different domains of text inputs it also becomes difficult to determine the best method to summarize the articles. Therefore automated system tools for text summarization are needed to generate concise and fluent summary containing important details from the source document. Broad classification of text summarization can be done in the following two ways: Extract and Abstract summarization (Hidayat, Firdausillah, Hastuti, Dewi, & Azhari, 2015*).* Extraction based summary choose the relevant words from the sentences and combine them to generate a meaningful summary while in an abstract form of summarization interpretation of the source document is presented in the form of shorter text by using rephrased words (Manalu, 2017).

This paper attempts to address the best-suited summarization approach amongst the TF-IDF, TextRank, and LDA respective for the three different input domains of text according to their degree of complexity. All the three approaches used to summarize the data are extraction-based summarization approaches. After precisely generating the summary from all the three approaches, evaluation of the generated summary is done by using ROUGE metrics.

II. Related Study

Natural Language Processing is an emerging field for investigation and research under which text summarization is applied. Various approaches have been designed to address the problems associated with the summarization task. The earliest approaches are based on the content feature as introduced in (Luhn,1958). These features are mostly based on frequency measures of the text document. The very first approach emerged is TF-IDF and various advancements have been introduced for better results. Summary generated from the various summarizers available online is compared by using the TF-IDF approach with

[*] Corresponding Author

DOI: 10.37398/JSR.2021.650140

better accuracy as in (Christian, Agus, & Suhartono, 2016). In (Ramos,2003) a query-based relevant word retrieval method is presented using TF-IDF. Hybrid use of approaches is also used like TF-IDF clustering in (Bafna, Pramod, & Vaidya,2016) and fusion of TextRank and TF-IDF in (Yao, Pengzhou, & Chi, 2019). Also, certain modifications are applied to traditional TF-IDF to check the performance of the approach (Roul, Sahoo, & Arora, 2017). TextRank algorithm is a graph-based approach introduced in the early stage of the 21st century (Mihalcea & Tarau, 2004) inspired by the methodology of PageRank introduced in (Brin & Page,1998). Application of TextRank is also used for the keywords extraction and uses the graph-based approach for the relevant information retrieval (Mallick, Das Dutta & Sarkar, 2019). Text Rank is a variation of the PageRank algorithm of Google which is a ranking algorithm for web pages available online based on the search results (Mallick, Das, Dutta, & Sarkar, 2019). TextRank uses certain ways to calculate the relation between sentences, cosine similarity is one of them as described in (Barrios, Lopez, Argerich, & Wachenchauzer, 2016). The meaningless words generally called stop words need to be removed for better summary production as in (Manalu, 2017), (Qaiser & Ali, 2018). TextRank also helps in determining the review assessment and credibility assessment as in (Manalu & Sundjaja, 2017) and (Balcerzak, Jaworski, & Wierzbicki, 2014) respectively. A statistical model for determining the abstract topic from the collection of documents gives the best result by using LDA introduced in (Blei, Ng, & Jordan, 2003) which is the first LDA implementation in machine learning. LDA has been used in various ways like topic modeling (Nagwani, 2015) and clustering (Hidayat *et al*, 2015*)*. In the LDA approach, text data is modelled at the word and document level (Chang & Chien, 2009). Multi-document summarization using LDA (Arora & Ravindran, 2008) and it is also applied to the complex input text data like legal documents as in (Kumar & Raghuveer,2012) proved to be a very efficient approach for summarization. ROUGE is an evaluation metric that is introduced in (Lin, 2004). ROUGE evaluates each approach by giving input of three different datasets one by one.

III. Proposed Methodology

Extraction based summarization process finds the most relevant words from the input document. Basic extractive text summarization includes the text pre-processing, extraction of words and sentences based on features then selecting the sentences and assembling them to produce a summary. There are various approaches to generate extraction oriented summaries. The methods which are implemented in this paper are described below:

A. TextRank

TextRank is a widely used method as no prior requirements of linguistic or domain knowledge. TextRank is an unsupervised approach for text summarization to generate extraction based summaries. The steps followed in the system used in this paper are shown in Figure 1

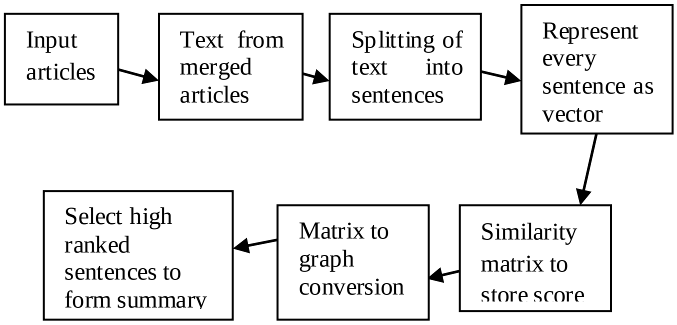
Input articles
Text from merged articles
Splitting of text into sentences
Represent every sentence as vector
Similarity matrix to store score
Matrix to graph conversion
Select high ranked sentences to form summary

Fig. 1: Flow Chart for TextRank Approach

Firstly, the input articles to the system are combined. The text obtained from this stage is then split into sentences. For each and every sentence obtained, vector representation is used after the stopword removal at the text preprocessing stage. There are certain similarity measures which are used to determine the similarity relation between the sentences based on the overlapping content between them (Barrios et al, 2016). Cosine Similarity measure is used in the proposed approach of TextRank as given by equation (1).

$$\text{Cosine Similarity } (S_1, S_2) = \frac{\vec{S}_1.\vec{S}_2}{|| S_1|.| S_2||} \quad (1)$$

S_1, S_2 are vectors used to represent sentences.

Similarity scores as stores in the similarity matrix. This approach models the similarity matrix into graphs, where nodes of the graph represent the sentences present in the documents and the edges represent the semantic relation through which the sentences are connected (Manalu, 2017). The similarity between the nodes is equivalent to the weighted edges of the graph (Balcerzak *et al*, 2014). After the similarity scores computation sentence ranking is done and the final summary includes the top ranked sentences

B. TF-IDF

This approach is termed as Term frequency-inverse document frequency is a statistical extraction approach that works by the comparison of the frequency of words in a particular document with the inverse proportion frequency of that word in other documents. It means if a word appears frequently in a document then it might be assumed by the user that it is important for the document (Yao et al, 2019). But if the same word appears frequently in other documents also then that word is not significant at all.

$$\text{tf}(w) = \left(\frac{\text{Total count of appearance of a term } w \text{ in D}}{\text{Total count of terms in D}} \right) \quad (2)$$

$$\text{idf}(w) = \log_e \left(\frac{\text{Total count of } D_n}{\text{Total Number of } D_n \text{ with term } w \text{ in it.}} \right) \quad (3)$$

D here represents a particular document

D_n here represents collection of documents

Hence TF-IDF is calculated in the following way for *w* which is a word in the document:

tf-idf (*w*) = (2) * (3).

The system proposed for TF-IDF is shown in Figure 2, where a sequential manner is adopted for the steps.

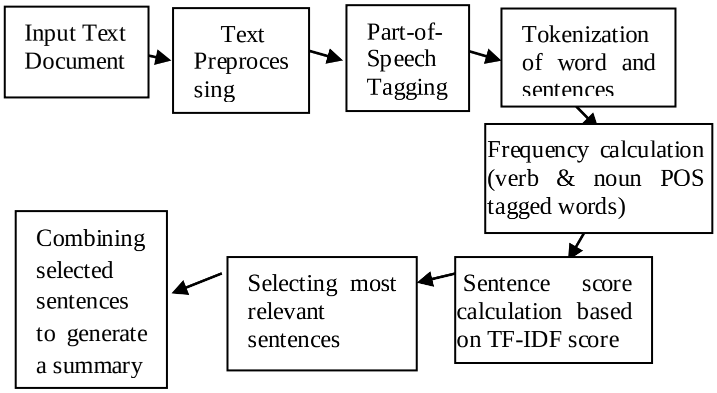
Input Text Document
Text Preprocessing
Part-of-Speech Tagging
Tokenization of word and sentences
Frequency calculation (verb & noun POS tagged words)
Sentence score calculation based on TF-IDF score
Selecting most relevant sentences
Combining selected sentences to generate a summary

Fig. 2: Flow Chart for TF-IDF Approach

The first step of the system consists of an input text document. After which Text Preprocessing is applied to clean the text by removing special characters, digits by the use of regular expressions. POS Tagging used in this approach labels nouns and verbs only. Tokenization splits the text into a collection of tokens Words work as a token for the sentences and the sentences work as tokens for the paragraph. The frequency score is calculated for the verbs and the nouns. TF-IDF calculation is applied at this stage to determine the scores of sentences. Based on those scores the sentences are selected which are most relevant for the summary generation. Selected sentences are then combined to produce the precise summary.

C. Latent Dirichlet Allocation

Latent Dirichlet Allocation is an unsupervised approach based on probabilistic algorithm extensively used for topic modeling. Topic modeling comes up with the ways to understand, organize, and generate summaries of large documents (Kumar & Raghuveer, 2012). The sequenced manner for this approach is shown in Figure 3. The first step towards this approach is the extraction of text data from the documents and then preprocessing the text. Preprocessing includes cleaning of text data, stop word removal and next step includes application of LDA for topic modeling. LDA represents the documents as the combination of topics as it breaks down the text document into topic clusters, which are based on probability distribution to mark the importance of the topic with regards to document (Arora & Ravindran, 2008).

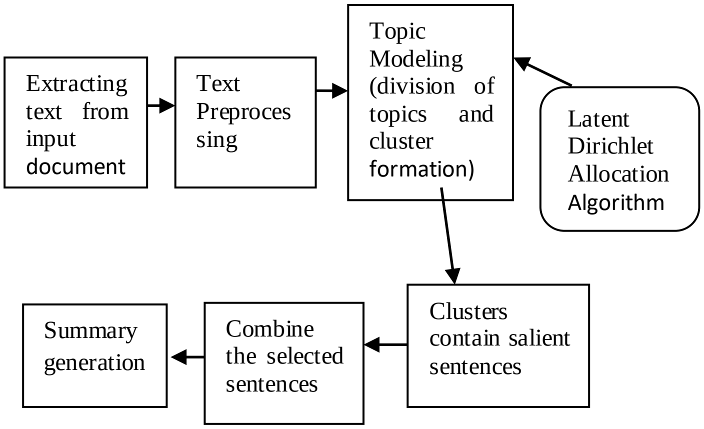
Extracting text from input document
Text Preprocessing
Topic Modeling (division of topics and cluster formation)
Latent Dirichlet Allocation Algorithm
Clusters contain salient sentences
Combine the selected sentences
Summary generation

Fig. 3: Flow Chart for the LDA approach

The clusters generated, contains the salient sentences from the original text document. The relevant topics which are identified would be linked to each cluster. The sentences of the source document are assigned to different recognized topics, in order to increase coverage. The property of topic selection from the document enhances the summarization process. LDA approach used in this paper isolates the five topics which are distinct from each other to generate a summary.

IV. Analysis of Experimental Results

A. Experimental Data

The research conducted in this paper used text documents from 3 different datasets. A collection of 50 documents from each dataset is taken as an input for the 3 different summarization approaches. The input to the system is the dataset on Reviews, News Articles, and Legal Information used in (Ganesan, Zhai, & Han, 2010), (Greene, Zhai, & Han, 2006), and (Galgani & Hoffmann, 2010) respectively.

As each group of 50 documents is fed to three different approaches which are in the form of the text file, the outputs generated are also in the form of text files. The outputs from a single approach count to 150 summary texts. The total summary files generated from the above described approaches are 450. The input text files are variable in the length, but their reference summaries are near of equal length within a dataset. There is a variation in the length of reference summaries when compared among three datasets. In order to generate summaries of an equal length corresponding to their reference summaries from all the three approaches, a manual setting is done for the compression of the output summary. The compression ratio for the legal dataset, when given as input to TF-IDF, is set to 25%, for other datasets it is set to 50% when given as input to TF-IDF. Also for the TextRank and LDA, the number of output sentences for the legal document is set to 10 but for the other two approaches range for output sentences is set to 15-20. Therefore, after analyzing these generated files, we can provide a conclusion by applying the comparative evaluation study on them.

B. *Evaluation*

Traditional ways of evaluation of the summaries are done on the basis of certain quality metrics which includes the human judgments. There are two ways for evaluation of methods of text summarization to check the performance: *Intrinsic* and *Extrinsic* (Mani,2001), (Saziyabegum & Sajja, 2017). In the first method, an intrinsic evaluation checks the output quality of summarization with respect to certain measures like accuracy, relevancy, readability, comprehensiveness, and informativeness (Moradi & Ghadiri, 2017). The extrinsic method checks the performance of summarization on the basis of effects it put on the completion of other tasks (Mani, 2001). The impacts tested on tasks like relevance assessment, reading comprehension, etc. When a summary helps other tasks then it is considered as a good summary. ROUGE is used for the evaluation of summaries generated by the system.

ROUGE is termed as Recall Oriented Understudy for Gisting Evaluation (Lin, 2004). This is a way of doing evaluation that computes the similarity between system produced summaries and the human-generated summary. There are five measures of ROUGE which are used to evaluate namely ROUGE-N, ROUGE-W, ROUGE-S, ROUGE-SU and ROUGE-L (Lin, 2004), [20]. Following measures of ROUGE 2.0 version are used for evaluation of system generated summary in this paper:

- ROUGE-1 & ROUGE-2: Both of them are the variants of the ROUGE -N, that computes N-gram common units in the reference summaries collection and the candidate summary (Saziyabegum & Sajja, 2017). N-gram's length is depicted by N in ROUGE-N. 1 in ROUGE-1 variant depicts overlapping of unigram and 2 depict overlapping of bigram in ROUGE-2.
- ROUGE-SU4: It is a variant of ROUGE-SU, which is an enhanced version of ROUGE-S (Skip bigram) (Saziyabegum & Sajja,2017). ROUGE-SU4 skips a maximum distance of 4 between the bigrams used.
- ROUGE-L: L here represents the Longest common subsequence (LCS), the measure used in this evaluationvariant. LCS determines the maximum length of common when the two summaries are compared namely system generated and reference summaries which are human generated.

The criteria for the evaluation of summaries are Recall (R), Precision (P), and F-measure (F). Recall(R) is computed as count of common sentences which overlap between the system generated summaries and the reference summaries dividing it by the count of sentences present in the reference summary generated by human. Precision (P) is measures as a count of common sentences between the system summaries and reference summaries dividing it by the sentence count present in the summary produced by the system. The higher the values of Precision and Recall, the better is the accuracy of results. The F-measure (F) is a combined measure of Precision(P) and Recall (R). It is computed as harmonic average of P and R value.

C. *Results and Observations*

In this research paper, the comparative study on 3 different approaches applied to 3 different datasets is done by using Recall(R), Precision(P) and F-Measure(F) criteria. The average scores of R, P and F are used to check the system performance corresponding to the datasets. The approaches are TF-IDF, TextRank and LDA.

Tables 1-3 present the average values R_a, P_a and F_a corresponding to the Reviews, News and Legal dataset applied on all the three approaches used respectively.

0.12345 Colored values represent the maximum value of Average of Recall (R), Precision (P) and F- Measure (F) as R_a, P_a and $_a$ respectively according to the ROUGE type.

It is observed from Tables 1-3 that the system summarization capability achieves maturity by using the TextRank approach. As in the review dataset the maximum average value of Recall for ROUGE-1, 2, L and SU4 are obtained from the TF-IDF, but the Precision and F-measure for this dataset is obtained from the TextRank approach. In the news dataset TextRank dominates as the highest Recall, Precision and F-Measure average scores are obtained from it for all the ROUGE types. Also, LDA approach gives the second highest values of ROUGE scores for each ROUGE type for the news dataset. In the legal dataset ROUGE-L and ROUGE-1 values for Recall are obtained from the TF-IDF approach, but for ROUGE-2 and ROUGE-SU4 highest Recall values are given by TextRank. Also, high Precision and F-measure value for the Legal dataset for most of ROUGE types are obtained from TextRank. So, if only the evaluation is done only on the basis of ROUGE scores then TextRank dominates. The range of values for Review dataset for Recall value is 0-0.54, for Precision value is 0-0.156 and for F-Measure is 0-0.23. The range of values for the News dataset for Recall value is 0-0.80, for Precision value is 0-0.56 and for F-Measure is 0-0.65. The range of values for the Legal dataset for Recall value is 0-0.32, for Precision value is 0-0.34 and for F-Measure is 0-0.26. Hence, the News dataset relates to the widest range of Recall, Precision and F-measure values. Also, the value of ROUGE-1 for R, P and F are greater than ROUGE –L, ROUGE-SU4 and ROUGE-2 in the Review and Legal dataset. But, for the News dataset most ROUGE-L values of each measure are greater than ROUGE-1 values. Graphs 1-3 helps in visualizing the performance for the three approaches according to the dataset provided.

Summarization Approach	→ Dataset Type	Review Dataset			
	→ ROUGE Type	ROUGE-L	ROUGE-SU4	ROUGE-1	ROUGE-2
TextRank	R_a *(Avg_Recall)*	*0.379889804*	*0.246103922*	*0.486273922*	*0.126542549*
	P_a*(Avg_Precision)*	*0.12113*	*0.057842745*	*0.155989608*	*0.035849216*
	F_a *(Avg_F-Measure)*	*0.179807843*	*0.091985882*	*0.233059412*	*0.055018824*
TF IDF	R_a *(Avg_Recall)*	*0.453446667*	*0.279472353*	*0.543041373*	*0.154938824*
	P_a*(Avg_Precision)*	*0.060070392*	*0.028067647*	*0.07563902*	*0.017999412*
	F_a *(Avg_F-Measure)*	*0.103046863*	*0.049262353*	*0.128096667*	*0.030910588*
LDA	R_a *(Avg_Recall)*	*0.361572549*	*0.205583529*	*0.44298098*	*0.090805686*
	P_a*(Avg_Precision)*	*0.0690992*	*0.03194451*	*0.09712549*	*0.017959412*
	F_a *(Avg_F-Measure)*	*0.11256667*	*0.054235098*	*0.15685*	*0.02941*

Table 1: The ROUGE scores for the Review Dataset

Summarization Approach	→ Dataset Type	News Dataset			
	→ ROUGE Type	ROUGE-L	ROUGE-SU4	ROUGE-1	ROUGE-2
TextRank	R_a *(Avg_Recall)*	*0.803823*	*0.7707474*	*0.7902034*	*0.7603536*
	P_a*(Avg_Precision)*	*0.5686992*	*0.5494134*	*0.5607558*	*0.5405252*
	F_a *(Avg_F-Measure)*	*0.6520132*	*0.6232262*	*0.636767*	*0.6139468*
TF IDF	R_a *(Avg_Recall)*	*0.5352098*	*0.4290038*	*0.5030586*	*0.4150356*
	P_a*(Avg_Precision)*	*0.4541242*	*0.4151762*	*0.4675862*	*0.3959512*
	F_a *(Avg_F-Measure)*	*0.4868588*	*0.4161304*	*0.479677*	*0.4003722*
LDA	R_a *(Avg_Recall)*	*0.7611762*	*0.652136*	*0.7144186*	*0.6328212*
	P_a*(Avg_Precision)*	*0.507353*	*0.4563576*	*0.4697372*	*0.4517186*
	F_a *(Avg_F-Measure)*	*0.602426*	*0.528521*	*0.5572332*	*0.5188444*

Table 2: The ROUGE scores for the News Dataset

Summarization Approach	→ Dataset Type	Legal Dataset			
	→ ROUGE Type	ROUGE-L	ROUGE-SU4	ROUGE-1	ROUGE-2
TextRank	R_a *(Avg_Recall)*	*0.2756578*	*0.1688628*	*0.3103956*	*0.1326522*
	P_a*(Avg_Precision)*	*0.2660594*	*0.1585424*	*0.343344*	*0.1151544*
	F_a *(Avg_F-Measure)*	*0.2346422*	*0.1282904*	*0.2624326*	*0.0960334*
TF IDF	R_a *(Avg_Recall)*	*0.2995956*	*0.1601368*	*0.32738*	*0.1175512*
	P_a*(Avg_Precision)*	*0.2254726*	*0.1388356*	*0.3034898*	*0.0953482*
	F_a *(Avg_F-Measure)*	*0.2330002*	*0.1275294*	*0.2691028*	*0.0915976*
LDA	R_a *(Avg_Recall)*	*0.2458288*	*0.093973*	*0.2158992*	*0.0703488*
	P_a*(Avg_Precision)*	*0.2454166*	*0.1228304*	*0.3191844*	*0.0824316*
	F_a *(Avg_F-Measure)*	*0.210304*	*0.0822142*	*0.2045426*	*0.057496*

Table 3: The ROUGE scores for the Legal Dataset

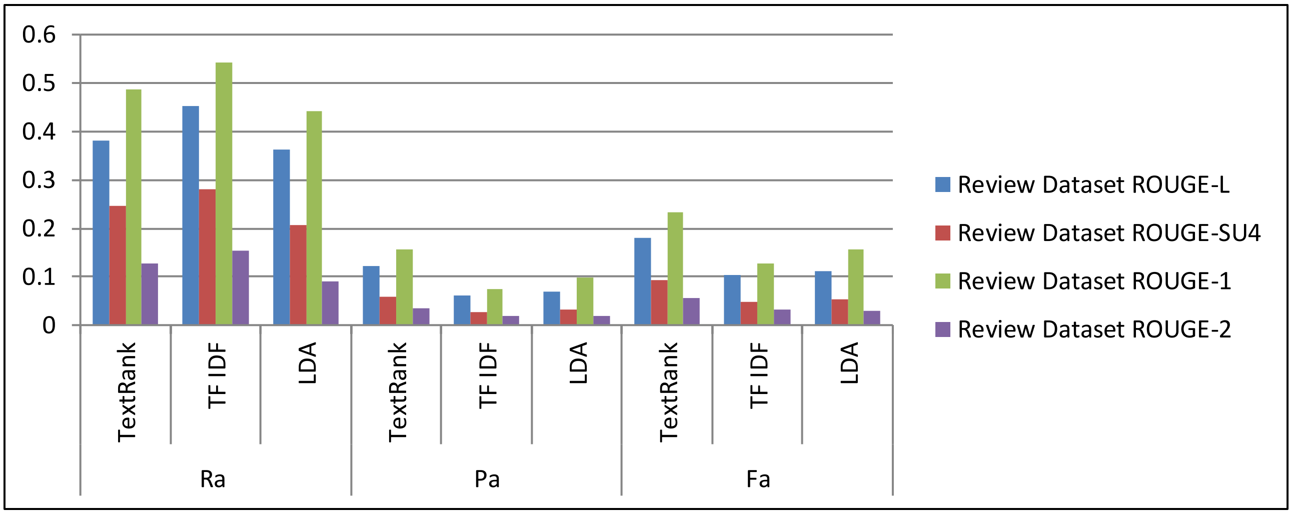

Fig. 4: Graphical presentation of ROUGE scores of Review Dataset

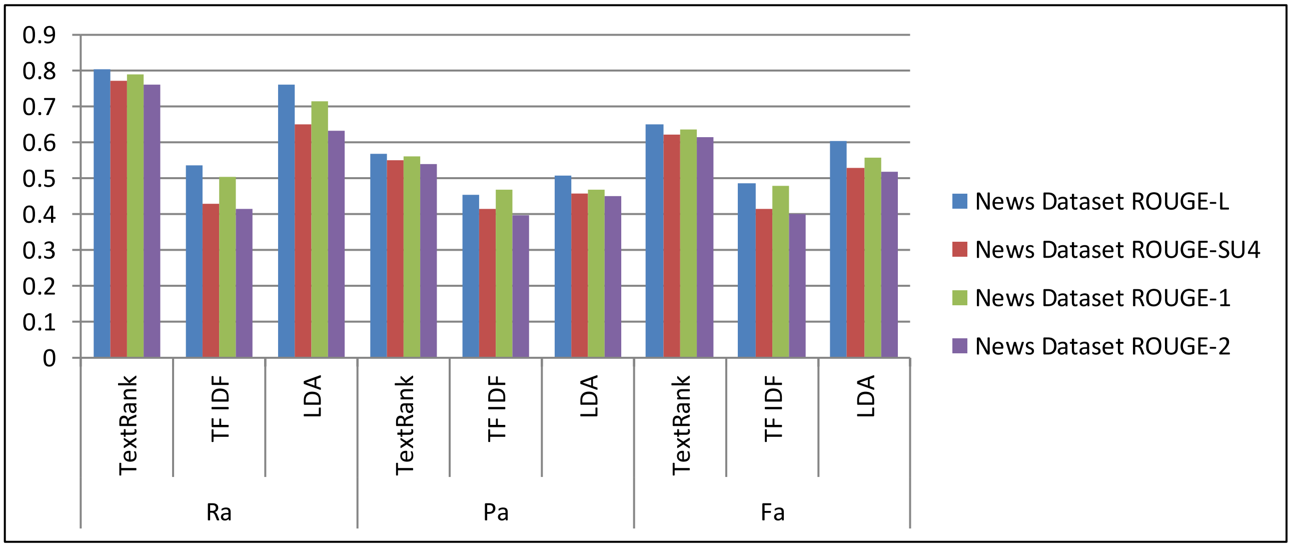

Fig. 5: Graphical presentation of ROUGE scores of News Dataset

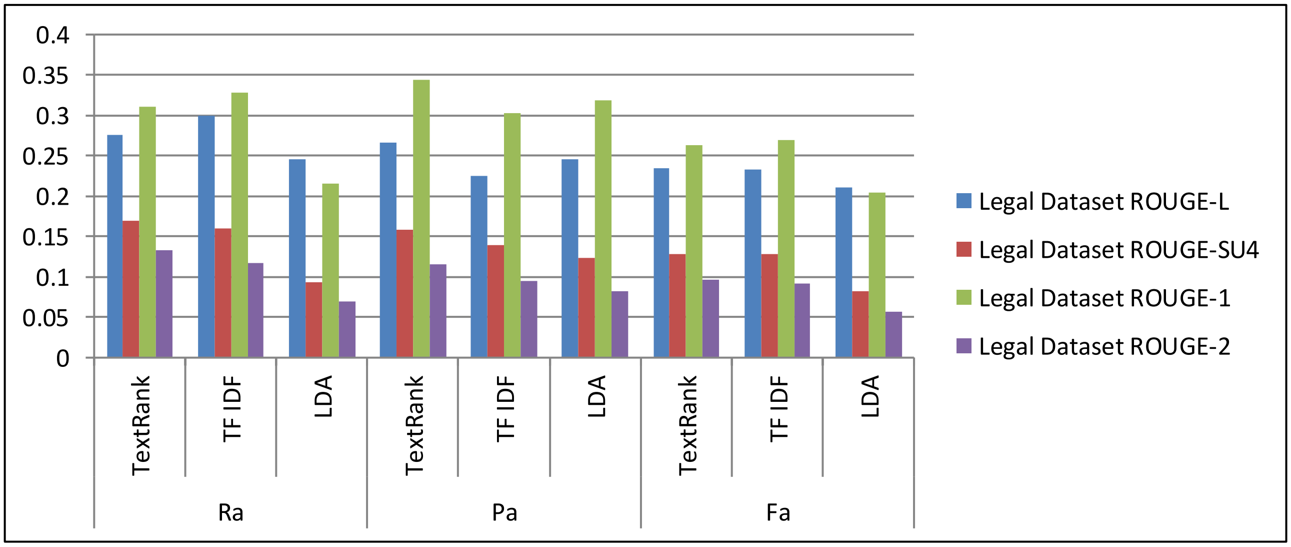

Fig. 6: Graphical presentation of ROUGE scores of Legal Dataset

Conclusion

Text Summarization is the process of compressing of the original documents into the summaries in such a way that the generated summaries can be substituted with the original document without making any compromise with the information delivered from the document. Abstractive and Extractive are two widely used approaches for summarization. In this paper, the extraction based methodologies are implemented.TF-IDF, TextRank, and LDA are widespread techniques that are used for document summarization. This paper presented a comparative analysis of the above said approaches by applying them on three different datasets to determine the better approach as compared to others. The used datasets are having different text domains. The evaluation is done using the ROUGE metrics using Recall, Precision and F-measure criteria. TextRank performed better as compare to TF-IDF and LDA approaches. But, for the dataset of News Articles LDA gives the ROUGE values higher than the TF-IDF but lower than TextRank. So, there is possibility of better performance of LDA for the input based text documents. Also, the highest values of ROUGE are obtained for the News Datasets by all the three approaches. Clearly, this study shows that TextRank is better than TF-IDF and LDA when used in an unsupervised way and according to the properties of each approach used in the implementation.

Future Scopes

Since there are numerous other approaches for summarizing the text documents, which can also be used for the comparative analysis to determine the best performing methodology. TF-IDF approach used the noun and verb for the POS tagging. The variants like adjectives etc. can be used to check the performance of TF-IDF with or without stopword removal. TextRank used in this paper applies the cosine similarity function. Other similarity measures can be used in the future for performance evaluation. LDA is a novel approach for text summarization as compare to TF-IDF and TextRank. Unsupervised LDA is used in this paper. So, there is a possibility of better performance by using the guided LDA. Comparative Analysis of approaches is also gaining importance these days.

References

Arora, R., & Ravindran, B. (2008, July). Latent dirichlet allocation based multi-document summarization. In *Proceedings of the second workshop on Analytics for noisy unstructured text data* (pp. 91-97).

Bafna, P., Pramod, D., & Vaidya, A. (2016, March). Document clustering: TF-IDF approach. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 61-66). IEEE.

Balcerzak, B., Jaworski, W., & Wierzbicki, A. (2014, August). Application of TextRank algorithm for credibility assessment. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)* (Vol. 1, pp. 451-454). IEEE.

Barrios, F., López, F., Argerich, L., & Wachenchauzer, R. (2016). Variations of the similarity function of textrank for automated summarization. *arXiv preprint arXiv:1602.03606*.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine.

Chang, Y. L., & Chien, J. T. (2009, April). Latent Dirichlet learning for document summarization. In *2009 IEEE international conference on acoustics, speech and signal processing* (pp. 1689-1692). IEEE.

Christian, H., Agus, M. P., & Suhartono, D. (2016). Single document automatic text summarization using term frequency-inverse document frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, *7*(4), 285-294.

Galgani, F., & Hoffmann, A. (2010, December). Lexa: Towards automatic legal citation classification. In *Australasian Joint Conference on Artificial Intelligence* (pp. 445-454). Springer, Berlin, Heidelberg.

Ganesan, K., Zhai, C., & Han, J. (2010). Opinosis: A graph based approach to abstractive summarization of highly redundant opinions.

Greene, D., & Cunningham, P. (2006, June). Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning* (pp. 377-384).

Hidayat, E. Y., Firdausillah, F., Hastuti, K., Dewi, I. N., & Azhari, A. (2015). Automatic text summarization using latent Drichlet allocation (lda) for document clustering. *International Journal of Advances in Intelligent Informatics*, *1*(3), 132-139.

Kumar, R., & Raghuveer, K. (2012). Legal document summarization using latent dirichlet allocation. *International Journal of Computer Science Telecommunications*, *3*, 114-117.

Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out* (pp. 74-81).

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, *2*(2), 159-165.

Mallick, C., Das, A. K., Dutta, M., & Sarkar, A. (2019). Graph-based text summarization using modified TextRank. In *Soft Computing in Data Analytics* (pp. 137-146). Springer, Singapore.

Manalu, S. R. (2017, June). Stop words in review summarization using TextRank. In *2017 14th International Conference on Electrical Engineering/Electronics, Computer,* *Telecommunications and Information Technology (ECTI-CON)* (pp. 846-849). IEEE.

Manalu, S. R., & Sundjaja, A. M. (2017). Review assessment support in Open Journal System using TextRank. *JPhCS*, *801*(1), 012074.

Mani, I. (2001). Summarization evaluation: An overview.

Mihalcea, R., & Tarau, P. (2004, July). Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411).

Moradi, M., & Ghadiri, N. (2017). Quantifying the informativeness for biomedical literature summarization: An itemset mining method. *Computer methods and programs in biomedicine*, *146*, 77-89..

Nagwani, N. K. (2015). Summarizing large text collection using topic modeling and clustering based on MapReduce framework. *Journal of Big Data*, *2*(1), 1-18.

Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, *181*(1), 25-29.

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).

Roul, R. K., Sahoo, J. K., & Arora, K. (2017, December). Modified TF-IDF term weighting strategies for text categorization. In *2017 14th IEEE India Council International Conference (INDICON)* (pp. 1-6). IEEE.

Saziyabegum, S., & Sajja, P. S. (2017). Review on Text Summarization Evaluation Methods. *Indian Journal of Computer Science Engineering*, *8*(4), 497500.

Yao, L., Pengzhou, Z., & Chi, Z. (2019, June). Research on News Keyword Extraction Technology Based on TF-IDF and TextRank. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)* (pp. 452-455).IEEE Computer Society.

***