International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 4 Ver. II || April 2018 || PP. 43-52

Text Summarizers for Education News Articles

Vishnu Preethi K, Vijaya MS
[1]*Research Scholar PSGR Krishnammal College for Women Coimbatore, India*
[2]*Research Scholar PSGR Krishnammal College for Women Coimbatore, India*
Corresponding Author: Vishnu Preethi K

***ABSTRACT**: Text summarization is a powerful text mining technique for condensing the contents of the documents without loss of context and information. As the text summarization models are highly adopted in areas like natural language processing, information retrieval, text compression, email thread summarization, library sciences there is necessity for building innovative text summarization models. In this research work, the text summarization models have been built using TextRank algorithm, though algorithm like LexRank, LSA have been used earlier. New dataset has been developed to test the performance of TextRank based text summarization model. The model is compared with LexRank and LSA based text summarization models and opinosis datasets. ROUGE scores such as ROUGE 1, ROUGE 2, ROUGE L which includes precision, recall, F-measure along with cosine similarity and relative utility are used as metrics for performance evaluation. The comparative analysis of all three summarizers shows that Text rank algorithm performs better for education dataset than other two algorithms.*

***KEYWORDS:** Latent Semantic Analysis, LexRank, TextRank Text Summarization*

Date of Submission: 31-03-2018 Date of acceptance: 16-04-2018

I. Introduction

Text summarization is a compression of text into a shorter version protective its info content and overall which means. The terribly tough for kinsmen people in general persons groups of people individual personalities to manually summarize large document of text. Text summarization strategies are often classified into extractive report and theoretic report.

Text summarization is associated with recent challenge in text miningbut in dire need of researcher's attention within the areas of process intelligence, machine data and linguistic communication process. Text summarization is that the method of mechanically making a compressed version of a given text that gives helpful information for the user. In huge organization or company, scientist does not have time to scan all documents so that they summarize document and highlight outline with details. An outline could be a text that created from one or a lot of texts that contains a major portion of the data reduced long and keeps the general which means because it is within the original texts. Text summarization involves varied strategies that use text categorization, like neural networks, call trees, linguistics graphs, regression models, symbolic logic and swarm intelligence. However, all of those strategies have a standard drawback, that is, the standard of the event of classifiers is variable and extremely passionate about the kind of text being summarized.

Automatic Text summarization analysis is incredibly necessary to the event of report systems that generates rational summaries that states the most goal of the given document. With the expansion of quantity of matter data, automatic report of matter data is in pressing want for economical process from large data well-structured, coherent documents.

Automatic summarization is challenging problem in computational linguistics, since text account is a good tool for process massive data resources in Microsoft world. Each of the summaries is totally different options for text outline extraction from given massive documents and studied its end in terms of range of options thought-about for extracting text outline. Extracted outline result naturally stricken by size of documents massive then restricted range of thought-about options might cause to the surprising result prefer to generate poorly connected sentence or incoherent outline.

II. Literature Survey

Various research works have been carried out in text summarization using several techniques. Only less work has been reported on Text Rank algorithm. Some of the existing works are listed here.

Deerwester S.Proposed, et al.[14] approach for text report by compartmentalization by latent linguistics analysis that is tried to beat downside of retrieval techniques supported extraction result by victimizations word queries and word of documents. However in Latent Semantic (LSA) Analysis there is also a probability of selection of unimportant or tangential ideas from document. as a result of one word having several which means and if we have a tendency to area unit didn't offer proof for extracting text by victimizations latent linguistics techniques then users question might not conclude expected output.

Abdel Fattah, Mohamed et al. [15] has proposed a simple approach for text summarization.They have considered features like position, length, name entities, numerical data, bushypaths, vocabulary overlaps etc. to generate summary. In this approach, sentences are modeled as vectors of features. Sentences are marked as correct if they are to be put in summarywhile are marked as in corrected if not. While making the final choice of sentences, each sentenceis given a value in between 0 and 1 and using a machine learning model, the sentences are selected using those scores.

Kamal Sarkar et al. [16] is built for summarization of medical documentsusing machine learning approach. Various features are very command and specific to medical documents. It uses the concept of cue phases such that if a sentence contains n cue phases,it gets n as its score for the feature. It also uses position of cue phases in the document such that if it appears at the beginning or at the end of the sentence, it get an additional 1point. Acronyms are also used as a feature and sentences having these get extra points. In some papers [17], sentences are also broken down by special cue makers and sentences arerepresented as a feature vectors.

Ryang, D Seonggi et al. [18] proposed a method of automatic text summarization with reinforcement learning. Researches have also been done for summarization of Wikipedia articles. Hingu, D et al. [19] implemented various method for summating Wikipedia pages. In one of their methods, sentences containing citations are given higher weightage. In the5her approach, the frequency of words are adjusted based on the root form of the word. The words are stemmed with the objective to assign equal weightage to words with the same rootword.

Edmundson (1969) [20] proposed an approach of extraction-based summarization using features like position, frequency of words, cue words and the skeleton of an article by manuallyassigning weight to each of these features. The system was tested using manually generated summaries of 400 technical documents. The results were good with 44% of summariesgenerated by it were matching the manual summaries.

From literature survey, it was observed that Lexrank algorithm is commonly used for text summarization, but both TextRank and Latent Semantic Analysis algorithms are limited in use. In the existing research, only open source datasets have been used for experiments. In the proposed research work a new dataset has been developed to implement text summarization and two techniques such as TextRank algorithm and LSA algorithm are employed to build models.

III. Proposed Work

Text Summarizers for Education News Articles is built using Lexrank, Textrank and Latent Semantic Analysis approaches. The proposed frameworks include different phases such as data collection, pre-processing, creation of summarizers, and performances evaluation. In the first stage the news articles are collected and datasets are formed. In pre-processing stage the news articles are pre-processed to remove stop words, removing incomplete sentences, elimination of duplicate sentences, finding meaningful words. The final phase produces the summarization models. The performance of the summarization models are evaluated by comparing model summaries against gold summaries using ROUGE scores.

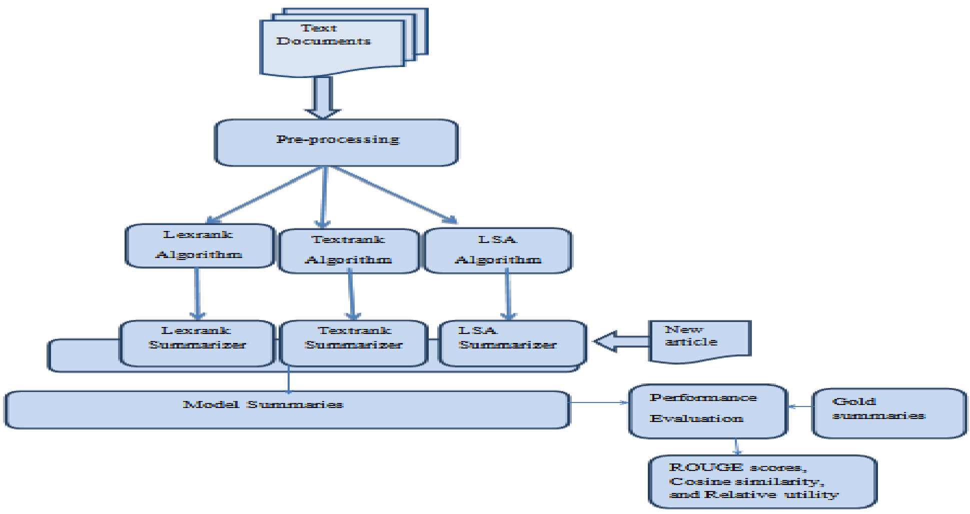

Fig.1 Architecture of the text summarization models.

1. Data Collection and Data Sets

The Opinonsis datasets was collected from online which consists of 51 articles. Each article is about a product's feature, like iPod's Battery Life, etc. and is a collection of reviews by customers who purchased that product.

A new dataset called education dataset is created by collecting 50 news articles related to education. Each article is about preschool, primary school, higher education and secondary education etc. Each article in the dataset possess manually written "gold" summary.

2. Pre Processing

Generally, data pre-processing task in any data mining application includes cleaning, instance selection, normalization, feature extraction and selection etc. In this research work, the text documents are pre-processed using stop word removal, removing incomplete sentences, and elimination of duplicate sentences. A total of 50 education news articles have been pre-processed prepare the corpus.

Tokenization

The first step in the pre-processing is tokenization. Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing in text mining. This is accomplished with the following functions.

Finding meaningful words

The use of very common words like 'the' does not indicate the type of similarity between documents in which one is interested. Single letters or other small sequences are also rarely useful for understanding content. So there is need for removing term which appears very few times in documents, because very rare words tell little about the similarity of documents, and most common words in the corpus, because words that are ubiquitous also tell little about the similarity of documents.

Removing Incomplete Sentences

Some documents in the dataset may be missing or empty because some words are filtered in the tokenization process. These documents can be disregarded by applying the following function.DocumentMinimumLengthFilter(length) functionis used to remove some documents in the dataset that are empty. Length can be specified to identify the shorter documents.

Elimination of Duplicate Sentences

Duplicate elimination () function step, only one copy of exact duplicated records are retained and eliminated other duplicate records. The elimination process is very important to produce a cleaned data. Before the elimination process, the similarity threshold values are calculated for all the records which are available in the data set.The similarity threshold ( ) values are important for the elimination process. In the elimination process, select all possible pairs from each cluster and compare records within the cluster using the selected attributes. Most of the elimination processes compare records within the cluster only. Sometimes other clusters may have duplicate records, same value as of other clusters. This approach can substantially reduce the probability of false mismatches, with a relatively small increase in the running time.Eliminate duplicate record based on data quality, threshold value, number of missing value and range of each field value. Retain only one duplicate record which is having high data quality, high threshold value and high certainty factor.

IV. Algorithms

In this research work the text has been summarized using computational techniques.The various summarizationmodels have been built using three algorithms such as LexRank, TextRank and LSA algorithm were used for summarization. TexRank and LexRank algorithm are graph based techniques whereas LSA is a statistical method. These three algorithms are described below.

LEXRANK ALGORITHM

LexRank is a graph based Lexical Centrality as Salience in Text Summarization. LexRank is a new approach for computing sentence importance based on the concept of eigenvector centrality in a graph representation of sentences. This approach models the document as a graph and uses an algorithm similar to Google's PageRank algorithm to find top-ranked sentences. This TF-IDF formulation is then used as a measurement for similarity between sentences by using it in this idf-modified-cosine formula

$$\text{idf-modified-cosine}(x,y) = \frac{\sum_{w \in x,y} \text{tf}_{w,x}\text{tf}_{w,y}(\text{idf}_w)^2}{\sqrt{\sum_{x_i \in x}(\text{tf}_{x_i,x}\text{idf}_{x_i})^2} \times \sqrt{\sum_{y_i \in y}(\text{tf}_{y_i,y}\text{idf}_{y_i})^2}} \quad (2.1)$$

This formula is basically measuring the 'distance' between two sentences x and y. the more similar two sentences, the more 'closer' they are to each other. This similarity measure is then used to build a similarity matrix, which can be used as a similarity graph between sentences. The LexRank algorithm measure the importance of sentences in the graph by considering its relative importance to its neighbouring sentences, where a positive contribution will raise the importance of a sentence's neighbour, while a negative contribution will lower the importance value of a sentence's neighbour. This idea is basically the same with PageRank, unless it is used in counting the importance of sentence in a given set of sentences.

TEXTRANK ALGORITHM

TextRank is an algorithm based upon PageRank for text summarization. Graph-based ranking algorithms are essentially a way of deciding the importance of a vertex within a graph, based on global information recursively drawn from the entire graph. The basic idea implemented by a graph-based ranking model is that of "voting" or "recommendation".Formally, let G = (V,E) be a directed graph with the set of vertices V and set of edges E, where E is a subset of VXV. For a given vertex $V_{i,}$ the score will be defined as,

$$S(V_i) = (1-d) + d * \sum_{j \in In\, v_i} \frac{1}{|out\ v_j|} S(V_j) \qquad (2.2)$$

where L is a damping factor that can be set between 0 and 1, which has the role of integrating into the model the probability of jumping from a given vertex to another random vertex in the graph.TextRank works well because it does not only rely on the local context of a text unit (vertex), but rather it takes into account information recursively drawn from the entire text (graph). Through the graphs it builds on texts, TextRank identifies connections between various entities in a text, and implements the concept of recommendation.

LSA ALGORITHM

Latent Semantic Analysis (LSA) is an algebraic-statistical method. It is an unsupervised method which extracts hidden semantic structures of words and sentences. LSA uses context of the input document and extracts information such as which words are used together and which common words are seen in different sentences. If the number of common words between sentences is high, it means that the sentences are more semantically related.LSA method has the ability to represent the meaning of words, and meaning of sentences at the same time. Meaning of a sentence is decided using the word it contains, and meaning of words are decided using the sentences that contains the word. The Latent Semantic Analysis (LSA) method can extract the meaning of words and sentences using only the input document, without any external information. It also has the ability of finding out the concepts in the input document. To perform the summarization based on LSA, first input matrix is created, and then LSA related calculations are done, and lastly sentences are selected as a part of summary.

V. Evaluation Metrics.

An evaluation measure that contains the content based evaluation. This used to compare the actual words in sentences, rather than entire sentences. Content evaluation uses two things they are co-selection and content based evaluation. The co-selection works is precision, recall, f-scores and relative utility. The content-based work is cosine similarity and n-gram matching is nothing but the ROUGE Scores.

Rouge scores

The ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measure is based on the Bleu metrics used in machine translation tasks. The idea is to compare the differences between the distribution of words in the candidate summary and the distribution of words in the reference summaries. Given h reference summaries and a candidate summary they are split into n-grams to calculate the intersection of n-grams between the references and the candidate. Precision is measured as:

$$\frac{\text{number\_of\_overlapping\_words}}{\text{total\_words\_in\_system\_summary}} \qquad (2.3)$$

Recall is the fraction of sentences chosen by the person that were also correctly identified by the system.Pyrouge could be a python wrapper for the ROUGE report analysis packages. Obtaining ROUGE to figure will need quite a little bit of time. Pyrouge is intended to create obtaining ROUGE scores easier by mechanically changing your summaries into a format ROUGE understands, and mechanically generating the ROUGE configuration file.

COSINE SIMILARITY

Cosine similarity circular function Similarity is employed to calculate the similarity between completely different documents. Circular function similarity between two documents may be a live that calculates the circular function of the angle between them. This metric may be a measuring of orientation and not magnitude, a comparison between documents on a normalized area as a result of not taking into the thoughtonly the magnitude of each word count (tf-idf) of each document, but the angle between the documents. Using the formula given below the cosine similarity between two documents is computed.

Cosine Similarity

(d1, d2) = Dot product(d1, d2) / ||d1|| * ||d2||

||d1|| = square root(d1[0]2 + d1[1]2 + ... + d1[n]2) (2.4)

||d2|| = square root(d2[0]2 + d2[1]2 + ... + d2[n]2) (2.5)

RELATIVE UTILITY

RU is applicable in both singledocument and multidocument summarization. When the target sentences are given, the judges (manual and automated summarizers) pick different sentences. This is called Summary Sentence Substitutability (SSS). RU agreement is outlined because the relatives core that one decide would get, given his own extract and therefore the alternative judge' sentence judgments. In RU, variety of judge's is asked to assign utility scores to any or all n sentences during a document. The highestsentences per utility score are then used as a sentence extract of size. In things wherever machine-controlled summaries are compared to manual summaries wherever sentences don't seem to be hierarchical, the Relative Utility technique couldn't be used as associate analysis technique.

VI. Gold Summaries Vs Model Summaries

The gold summary is nothing but the actual summary. This unit is generally quite accurate. Opinonsis datasets contain topics and summaries called gold summaries. The gold contains 51 instances and each single instance may contain 4 gold summaries. The Model summary for this datasets is generated by three algorithmsthey are Lexrank, Textrank and Latent semantic analysis. The generated model summary is evaluated with actual summary using rouge scores, cosine similarity and relative utility.

Education datasets that contain topics and summaries called gold summaries. The gold contains 51 instances and each single instance may contain 4 gold summaries. The Model summary for this datasets is generated by three algorithms they are Lexrank, Textrank and Latent semantic analysis. The generated model summary is evaluated with actual summary using rouge scores, cosine similarity and relative utility.

VII. Experiments And Results

Three experiments have been carried out using LexRank, TextRank, and Latent Semantic Analysis to develop the text summarizers. Two datasets, opinonsis an open source dataset, education dataset, a newly created dataset have been used for implementation. The summaries produced by the text summarizers are assessed based on the gold summaries using ROUGE scores. The results of the experiments are compared and the findings are discussed. In this research work, the entire experimental setup is established using Python packages.

The LexRank and TextRank summarizers have been generated using LexRank and TextRank algorithms in Python by importing Gensim package and tested with same datasets using two modules of Gensim 'summarize' and 'Rouge' imported in Python. Latent Semantic Analysis (LSA) is an algebraic-statistical method. It is an unsupervised method which extracts hidden semantic structures of words and sentences. LSA uses context of the input document and extracts information such as which words are used together and which common words are seen in different sentences. If the number of common words between sentence is high it means that the sentences are more semantically related. To implement LSA function, LSA packages is installed in R by loading related modules. The results of LexRank summarizer evaluated using Rouge scores by comparing the model summaries of all the text documents with the predefined gold summaries are shown in Table I(a) and Table I(b) and graphically illustrated in Fig 4.1a, Fig. 4.1b.

Table I(a) –Rouge score measures of Lexrank summarizer for Opinionsis and Education datasets

Datasets	Precision			Recall			F –scores		
	Rouge 1	Rouge 2	RougeL	Rouge 1	Rouge 2	Rouge L	Rouge 1	Rouge 2	Rouge L
Opinonsis dataset	0.244	0.171	0.237	0.416	0.295	0.394	0.273	0.216	0.299
Education dataset	0.412	0.295	0.594	0.273	0.267	0.499	0.244	0.175	0.433

Table I (b) –Cosine Similarity and Relative Utility measures of Lexrank summarizer for Opinionsis and Education datasets

Datasets	Cosine Similarity	Relative Utility
Opinonsis dataset	0.345679	0.323765
Education dataset	0.154379	0.187632

Note: P- Precision, R- Recall, F- F scores

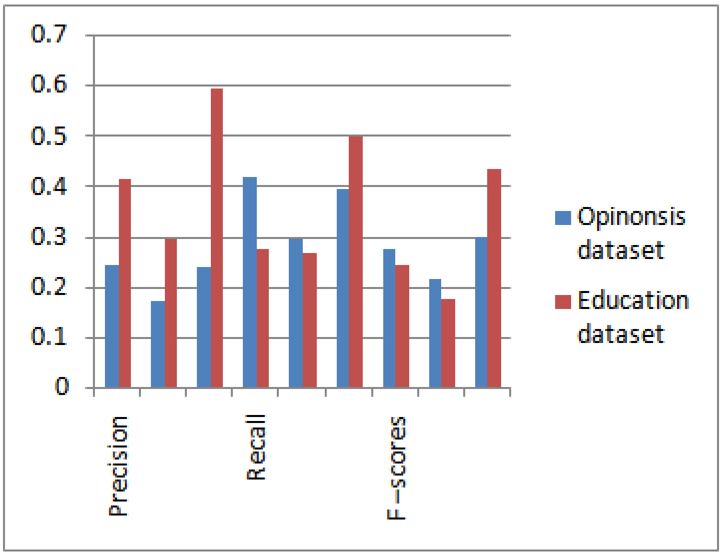

Fig. 4.1a Rouge score measures of LexRank summarizers for both opinonsis datasets and education system

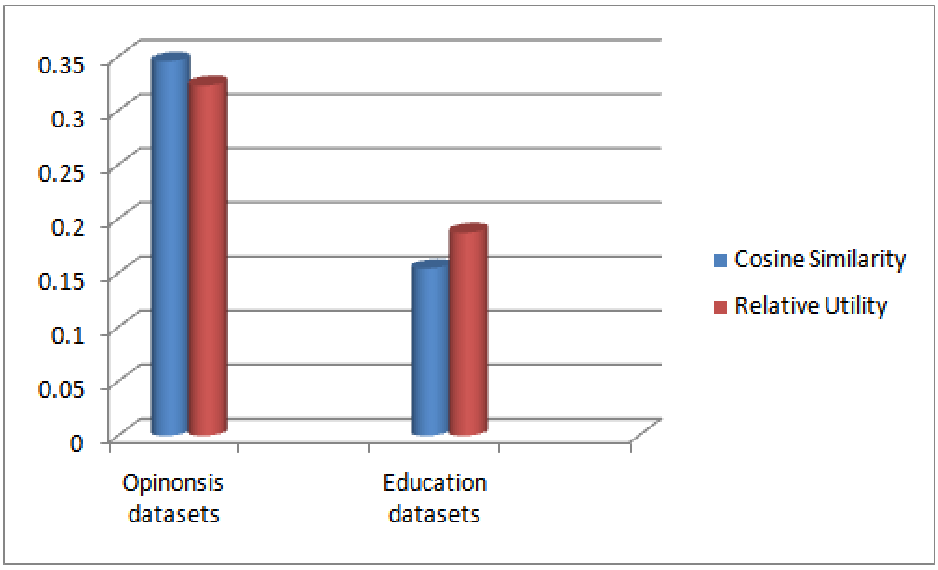

Fig.4.1b Cosine similarityand Relative utility measures of LexRank summarizers for both opinonsis datasets and education system.

in The results of TextRank summarizer evaluated with respect various metrics mentioned above by comparing the model summaries of all the text documents with the predefined gold summaries are shown in Table II (a) and Table II(b) and graphically illustrated Fig 4.2a, Fig. 4.2b.

Table II (a) –Rouge score measures of TextRank summarizer for Opinionsis and Education datasets

Datasets	Precision			Recall			F –scores		
	Rouge 1	Rouge 2	Rouge L	Rouge 1	Rouge 2	Rouge L	Rouge 1	Rouge 2	Rouge L
Opinonsis dataset	0.348	0.275	0.237	0.365	0.495	0.394	0.174	0.216	0.399
Education dataset	0.376	0.495	0.594	0.373	0.516	0.499	0.248	0.175	0.439

Table II(b) –Cosine Similarity and Relative Utility measures of TextRank summarizer for Opinionsis and Education datasets

Datasets	Cosine Similarity	Relative Utility
Opinonsis dataset	0.254379	0.423765
Education dataset	0.223456	0.387632

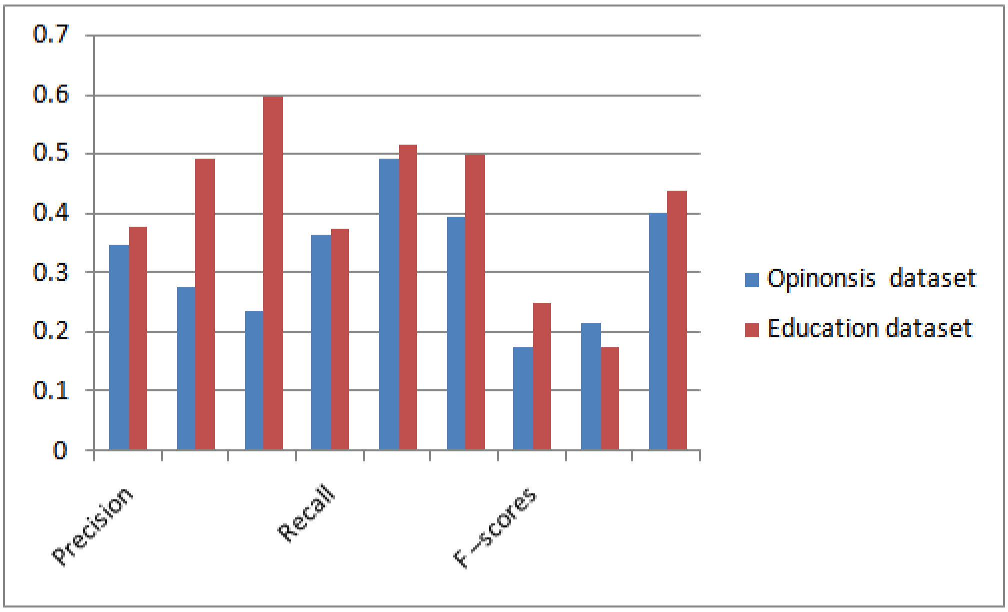

Fig. 4.2a Rouge score measures of TextRank summarizers for both opinonsis datasets and education system

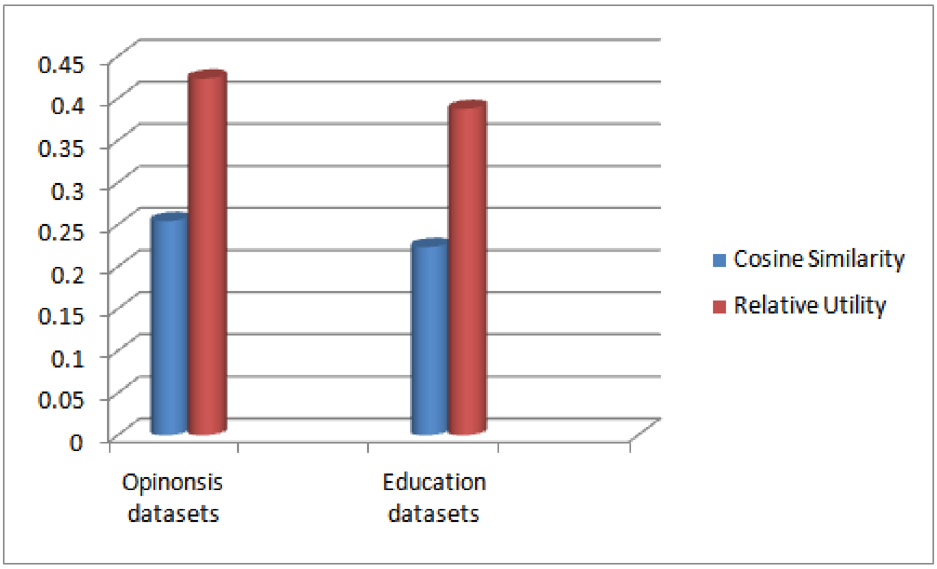

Fig. 4.2b Cosine similarityand Relative utility measures of TextRank summarizers for both opinonsis datasets and education system.

The results of LSA based text summarization model evaluated using Rouge scores, cosine similarity and relative utility by comparing the model summaries of all the test documents with the predefined gold summaries are shown in Table III(a) and Table III(b) and graphically illustrated in Fig 4.3a, Fig 4.3b.

Table III(a) –Rouge score measures of LSA summarizer for Opinionsis and Education datasets

Datasets	Precision			Recall			F –scores		
	Rouge 1	Rouge 2	Rouge L	Rouge 1	Rouge 2	Rouge L	Rouge 1	Rouge 2	Rouge L
Opinonsis dataset	0.248	0.475	0.235	0.417	0.195	0.394	0.173	0.216	0.599
Education dataset	0.416	0.395	0.594	0.173	0.267	0.499	0.344	0.171	0.437

Table III(b) –Cosine Similarity and Relative Utility measures of LSA summarizer for Opinionsis and Education datasets

Datasets	Cosine Similarity	Relative Utility
Opinonsis dataset	0.345679	0.223765
Education dataset	0.354379	0.327632

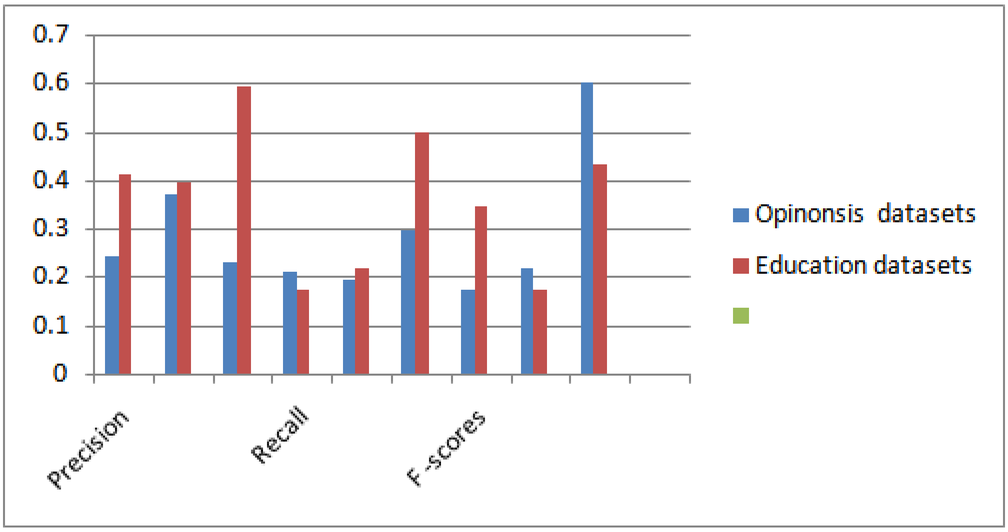

Fig. 4.3a Rouge score measures of LSA summarizers for both opinonsis datasets and education system

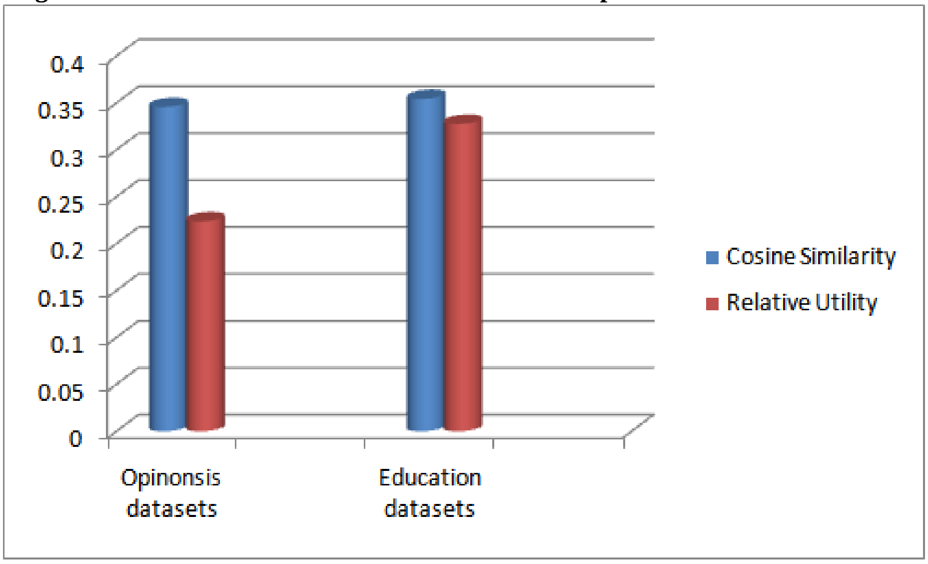

Fig 4.3b Cosine similarityand Relative utility measures of LSA summarizers for both opinonsis datasets and education system

The overall comparison of performances of all three summarizers is carried out with respect to ROUGE scores on two datasets. The comparative analysis with respect to various metrics such as Precision, Recall, F-Measure, Cosine similarity, Relative Utility are shown in Table IV(a) and Table IV(b) and graphically illustrated in Fig 4.4a, Fig. 4.4b.

Table IV (a) Comparative ROUGEscores of all three algorithms for two datasets.

Algorithm	Lexrank									TextRank									LSA								
	Rouge 1			Rouge 2			Rouge L			Rouge 1			Rouge 2			Rouge L			Rouge 1			Rouge 2			Rouge L		
Datasets	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Opinonsis dataset	0.244	0.171	0.237	0.416	0.295	0.394	0.273	0.216	0.299	0.348	0.275	0.237	0.365	0.495	0.394	0.174	0.216	0.399	0.248	0.475	0.235	0.417	0.195	0.394	0.173	0.216	0.599
Education dataset	0.412	0.295	0.594	0.273	0.267	0.499	0.244	0.175	0.433	0.376	0.495	0.594	0.373	0.516	0.499	0.248	0.175	0.437	0.416	0.395	0.594	0.173	0.267	0.499	0.344	0.171	0.437

Table IV(b) Comparative Cosine similarity and Relative utility of all three algorithms for two datasets.

Algorithm	Cosine similarity			Relative utility		
Datasets	Lexrank	Textrank	LSA	Lexrank	Textrank	LSA
Opinonsis dataset	0.345679	0.254379	0.345679	0.323765	0.423765	0.223765
Education dataset	0.245769	0.223456	0.354379	0.223765	0.387632	0.327632

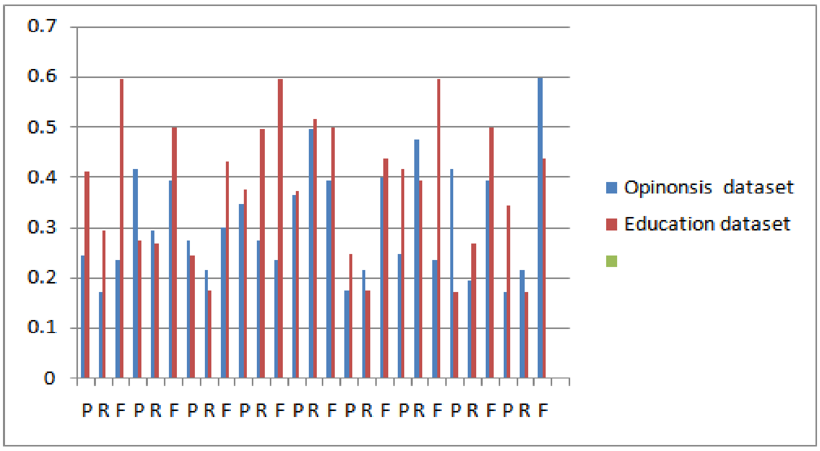

Fig. 4.4a Comparative ROUGE scores of LexRank, TextRank and LSA

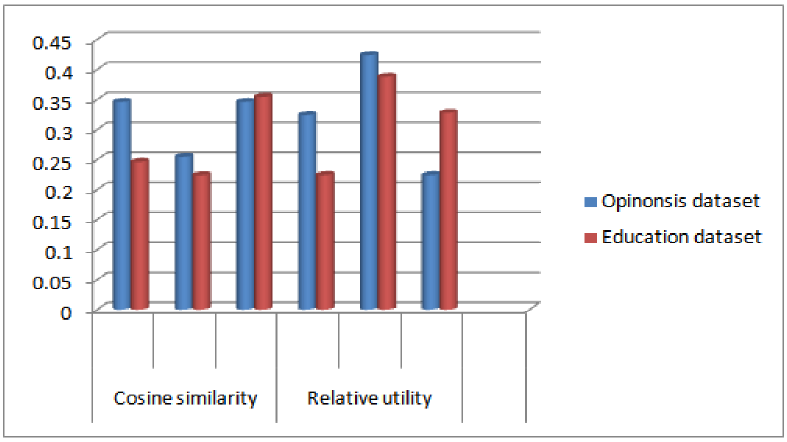

Fig. 4.4b Comparative analysis of both cosine similarity and relative utility of LexRank, TextRank and LSA

From the above comparative analysis it has been found that model summaries of the education datasets closely matches with their respective gold summaries when all three summarizers have been evaluated.The evaluation measures such as ROUGE scores, cosine similarity and relative utility are calculated for all algorithms. TextRank algorithm achieves best result compared with other two algorithms.

VIII. Conclusion

"Text Summarizers for Education News Articles" demonstrates a generative approach for summarizing the text using python libraries.An education dataset has been prepared by collecting and pre-processing news articles related to educational domain such as preschool, primary school, and higher secondary from online resources like BBC, TOI, and Science Daily.Text summarization models have been built using three algorithms such as LexRank, TextRank, Latent Semantic Analysis and the performance of the models have been evaluated by comparing model summaries with predefined gold summaries. The measures called ROUGE scores were used. The comparative performance analysis was made and found that the popular Text rank algorithm which was not used much in text summarization research produce better results for both dataset.

References

[1]. Shweta Kharya," Using data mining techniques for diagnosis and prognosis of cancer disease", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012.
[2]. N. Padhy, D. Mishra, R. Panigrahi et al., "The survey of data mining applications and feature scope,2012
[3]. R. Agrawal and M. Batra, "A detailed study on text mining techniques," International Journal of Soft Computing and Engineering (IJSCE) ISSN,2013.
[4]. Sholom M. Weiss, NitinIndurkhya, Tong Zhang, and Fred Damerau,"Text Mining: Predictive Methods for Analyzing Unstructured Information", Springer, Edition 2005.
[5]. John Elder, Gary Miner, Bob Nisbet, (2012), "Practical Text Mining and Statistical Analysis for Non-structured Text Data applications",Elsevier press.
[6]. V. Gupta and G. S. Lehal, "A survey of text mining techniques and function," Journal of emerging technologies in web intelligence, vol. 1, no. 1, pp. 60–76, 2009.
[7]. Manu Konchady (2006), "Text Mining Application programming", Charles River Media.
[8]. http://www.targetmarketingmag.com/article/text-mining-database-marketing-customer-service-analysis/all/.
[9]. Lin, C.-Y. 2004. Rouge: a package for automatic evaluation of summaries. In Proceedings of the ACL-04 Workshop on Text Summarization Branches Out (WAS 2004)
[10]. Mani, I., and Maybury, M. 1999. Introduction. In Mani, I., and Maybury, M. T., eds., Advances in Automatic Text Summarization. The MIT Press.
[11]. D. S. Dang and P. H. Ahmad, "A review of text mining techniques associated with various application areas," International Journal of Science and Research (IJSR), 2015
[12]. Hongyan Jing; "Sentence Reduction for Automatic Text Summarization"; Proceedings of the sixth conference on Applied natural language processing, Seattle, Washington, 2000.
[13]. Lin, C (2004). ROUGE (Version 1.5.5) [Software]. Available from http://www.berouge.com/Pages/default.aspxFRRF
[14]. Y. Gong, X. Liu: Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New Orleans, Louisiana, United States 2001.
[15]. R. Radev, S. Teufel, H. Saggion, W. Lam, J. Blitzer, H. Qi, A. Celebi, D. Liu, E. Drabek: Evaluation Challenges in Large-scale Document Summarization. Proceeding of the 41th annual meeting of the Association for Computational Linguistics, Sapporo Japan 2003.

International Journal of Engineering Science Invention (IJESI) is UGC approved Journal with Sl. No. 3822, Journal no. 43302.

Vishnu Preethi K "Text Summarizers for Education News Articles" International Journal of Engineering Science Invention (IJESI), vol. 07, no. 04, 2018, pp 43-52