A Comparative Study of Opinion Summarization Techniques

Surbhi Bhatia

***Abstract*—In the Web 3.0 platforms, enormous amount of information is shared whereby individuals express their thoughts and opinions and learn from others' experiences. Many e-commerce websites provide service of posting opinionated reviews to allow consumers post their opinions using free text. Examples of these e-commerce websites include eBay, Amazon, and Yahoo shopping. Summarizing text is taken as an interesting task of Natural Language Processing (NLP). The proposed work presents a comparative study of different techniques used for opinion summarization. It covers both abstractive and extractive approaches where summary of sentences is achieved by considering aspects. This article highlights the gaps in the previous study by proposing a novel graph-based technique for generating abstractive summary of duplicate sentences. The method discusses the details by constructing graphs, ensuring the sentence correctness using some constraints, and finally scoring the sentences individually by fusing sentiments using SentiWordNet. Extractive approach uses the principle of principal component analysis (PCA). The work includes the application of PCA in summarization of text by reducing the number of dimensions in data (aspects) and relatively finding the summary of the reviews on ranking the most relevant ones, according to the prime aspects without any loss of information respective of a particular domain. The analysis is conducted on the standard Opinosis data set and comparison is made between both of the techniques to discuss which method generates more coherent and complete summary.**

***Index Terms*—Abstractive, comparison, extractive, graphs, Natural Language Processing (NLP), opinion, Principal Component Analysis (PCA), summarization.**

I. Introduction

SINCE the volume of online content is easily accessible, reviews, opinions, and feedbacks are readily available for making decisions. Summarizing text is taken as an interesting task of Natural Language Processing (NLP). The intersection of two fields, computer science and linguistics, leads to the formation of NLP. In other words, thoughts and notions can be swapped over between human and computers by applying NLP on processed data. Lexical analysis, syntax analysis, semantic analysis, and discourse processing are the main components of NLP.

Multiple opinions are required to generate a summary, and it is based on feature selection [1], feature rating [2], and identifying sentence that contain features [3]. Primarily, summarization is categorized into two types but there may be other approaches, such as graph based, general, and hybrid techniques [4].

Manuscript received November 21, 2019; revised March 31, 2020, July 19, 2020, and September 14, 2020; accepted October 17, 2020. Date of publication November 18, 2020; date of current version January 29, 2021.

The author is with the Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Hofuf 31982, Saudi Arabia (e-mail: surbhibhatia1988@yahoo.com).

Digital Object Identifier 10.1109/TCSS.2020.3033810

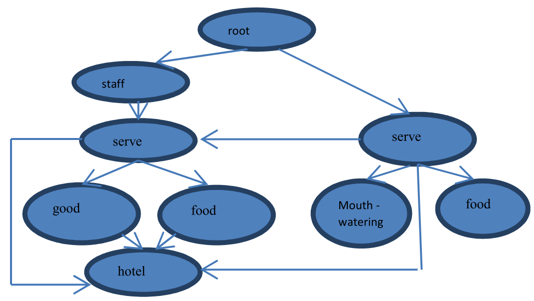


Fig. 1. Semantic representation of the text for abstractive summarization.

Extractive summarization involves combining all the extracts taken from corpus into summary. It helps the user by extracting most important pieces of information from the huge corpus. It acts as verbose, that is nonessential parts of sentence also get included only when data are redundant. It suffers from the dangling problem as it solely lies on the content and extracts sentences. Sometimes, irrelevant sentence gets included and the important parts of sentences such as pronouns are left out, those, otherwise, needs to be conserved [5]. Single-document summarization focuses on extractive techniques like machine learning methods-Naïve Bayes (NB), decision trees (DT), hidden Markov model, log linear, and neural networks (NN).

Abstractive summarization involves generating new cohesive text that may not be present in original information. It requires deep learning over the text to determine the meaning of each word and phrase generate summary. It builds a semantic representation of the text from which summary gets generated. Fig. 1 presents the merging of the text in a cohesive manner. In the semantic representation, graph nodes are concepts and edges are semantic relations.

The example to explain the above semantic representation is given as follows.

Sentence A: Mouth-watering food is served.

Sentence B: Staff serves good food in hotel.

Summary Generated: Staff serves good mouth-watering food in hotel.

Advanced language generation techniques are needed to produce a grammatically oriented summary highlighting the "form." Multidocument summarization follows abstractive approaches. Automatic summarization is involved either using statistical approach, linguistic approach, or hybrid approach. The sentence is ranked on the basis of keywords and determination of probability of key terms through weight by statistical approach. On the other hand, the linguistic approach examines text and finds out the concepts and associations between sentences by looking into semantics using POS tagging, analyzing grammar, and dictionary meaning of relevant sentences [6]. Abstractive techniques make use of linguistic or hybrid approaches but extractive technique makes use of statistical approach. Computation is stronger in extractive approaches but better results on summarization are produced by abstractive approach [7]. There are hybrid approaches that tend to merge both statistical-based and linguistic-based methods. Large numbers of researchers have proposed summarization methods based on NLP. Because of the reasons listed above and the previous other researches, it also demonstrated that abstractive summaries can give much better results when one has to summarize product reviews, blogs, and articles than extractive summaries [8].

The proposed work covers both abstractive and extractive approaches where summary of sentences is achieved by considering aspects. Abstractive approach is based on constructing graphs, ensuring the sentence correctness using some constraints, and finally scoring the sentences individually by fusing sentiments using the previous described technique, that is, using CNN. The abstractive summary generated on the aspect "GPS" with data set containing 100 reviews as Garmin Nuvi 255W GPS is shown in the Appendix. Extractive approach uses the principle of principal component analysis (PCA). It is a statistical technique that helps in transforming an array of data values which are associated or correlated in some form into values that are linearly uncorrelated data sets known as principal components using orthogonal transformation. These are selected as the most relevant sentences since their vector representations are best approximated by projecting them into the boundary of the sparse principal components [9].

The proposed work outlines the novel abstractive summarization algorithm by clearly demonstrating the gaps found in [20] and how the present method uses sentiment analysis to overcome the issue of merging two sentences using a pre-existing connector. The experiments show the improvement in performance by 13%. The contribution of this article lies in leveraging the order of sentence words to produce sentences which are comprehensive. In addition, this article uses the simple approach of generating summaries using PCA and compares it with novel abstractive approach by validating whether abstractive summaries perform well than extractive summaries with document having a lot of duplicate sentences. This clearly states that abstractive summaries present information more clearly and in a more condensed manner. It is definitely a not so easier task to combine sentences using abstractive approach than extractive approach.

The results of both the above two techniques are evaluated and compared using ROUGE tool. The primary targets of this research work under summarization are as follows.

1) Generating well-structured and concise summary on the queried aspects.
2) Redundancy should be eliminated, correctness of sentence is ensured, and maintained accurately.
3) Ranking of sentences is achieved by assigning weights using SentiWordNet score and PCA is applied cleverly to find important sentences which will contribute in summarization of opinions by setting threshold.
4) Comparing the accuracy results of the summaries generated by both the techniques.

A. Article Organization

This article has been structured as follows: Section II explains the related work of the existing text summarization approaches. Section III presents the proposed work which includes novel algorithms based on abstractive and extractive summarization techniques. Section IV lists the experiments performed on the data sets by analyzing the results and Section V summarizes this article and offers future research directions.

II. Related Work

The extensive study has been done on abstractive-based technique rather than extractive technique. The rule-based approach [10], sentence compression [11], [12], and merging sentence based on their semantics [2], [13] have been studied. Sharma and Chitre [14] proposed effective ranking method for summarizing text on using hash algorithms for constructing graphs. Ganesan *et al*. [15] developed the novel method of generating abstractive summary using the directed graphs. The use of connectors by providing input in the form of graphical form helps in the reduction of redundant sentences as opinions. However, the proposed work produces readable, concise, and fairly well-formed summaries but still has a limitation mark. The unavailability of a pre-existing connector might not be able to fuse the sentences which are capable of stacking up together. The complexity of this approach is high as it lays too many emphases on the surface order of words. Bhargava *et al*. [16] proposed the graph-based summary by incorporating the rules of Natural Language Processing. The methodology describes by building the word graph, ensuring the correctness by imposing POS constraints, and the scoring of paths based on the calculation of overlap. Finally, the fusion of sentences is achieved and the results are analyzed with ROUGE tool. The work limits to combining the sentences that are semantically related but not related syntactically. Tayal *et al.* [5] focused on a method for automatic text summarization using soft computing approach. Title and semantic similarity are primarily used for generating summary. The author developed the NLP parser and used human summarization rules; Subject, Verb, and Object (SVO) for achieving task of summarizing text. The proposed algorithm is not able to cope up with complex sentences as the rules stated are not enough to support these errors. Also, the problem of tag-based ambiguity needs to be resolved. Brian *et al*. [9] discussed the summarization process of large corpora using PCA. The author well formulated the low-rank approximation of a Salton matrix by taking text documents and running the procedure over the collection of news articles. Although the author tried to explain well the benefits of sparse PCA over normal PCA, but still the paper did not explain well mathematically how encoding of the Salton matrix is achieved and how dropping low variance features gives no downside in the proposed work. This research will fill this gap. Khan *et al.* [17] proposed dictionary-based approach for classifying sentences using supervised learning techniques. Finally, the sentiments were analyzed and after identifying the semantic priority, the summary is generated. The mathematical models were discussed in detail for computing polarity using SentiWordNet. Comparison is presented by taking different data sets. Sankarasubramaniam *et al.* [18] worked on extractive summaries using graph-based technique on Wikipedia and used iterative ranking algorithm to map sentences specific to the topic. The drawback of the approach decreased the value of the precision than abstractive summaries. Lloret and Palomar [19] used page rank algorithm and the frequently used words to calculate weight of the sentences. Few grammar rules were used to validate correctness of sentences. Ranking of the sentences was taken from highest TF-IDF score from the first node which had first 10 words. A lot of important information from the sentence cannot be retained because of applying highest TF-IDF scores policy and grammar of the sentences obligations. Also it will give a lot of duplicate information which will not be a problem in the proposed abstractive graph-based technique in this research.

Recently, much work on text summarization takes the advantage of deep learning techniques [20] such as convolutional neural networks (CNNs), recurrent neural networks (RNNs) [21], generative adversarial networks (GANs) [22], sequence-to-sequence model [23], and autoencoders. Shi *et al.* [23] proposed a novel model for abstractive summary using deep learning algorithms that produce syntactic and semantic intelligible summaries using the advantage of multiple layers of peephole convolutional long short-term memory (LSTM). The proposed model forms the relevant pattern of training data and its performance has been compared with traditional LSTM giving better results. The research [21] is based on feedforward neural network for extracting extractive summary. Different baselines consisting of supervised and unsupervised summarizations were compared using ROUGE tool with good accuracy results claiming contextual summary with its recurrent nature. The probability and mathematical formulation were developed for locating a sentence. The sequence-to-sequence model [24] has been developed using attentional encoder–decoder for the task of abstractive summarization and various state-of-the-art methods have been evaluated including the famous document of understanding (DUC) data set [25]. The advantage of the proposed model lies in its robustness using pointer mechanism for managing rare words. Masum *et al.* [26] proposed a summarizer taking advantages of the deep learning methods using RNN to generate the summary by machine. The researchers claim to reduce the training loss with a value of 0.036 which was there with the sequence-to-sequence model.

III. Proposed Work

Human tendency is to formulate a summary which is achieved by determining which opinions are important and which are not. Both approaches initially accept the text document in the form of tokens. Given the sets of opinionated sentences which are fed as inputs to the system, it aims to generate a comparative summary. Different steps are followed to compute relevant summaries.

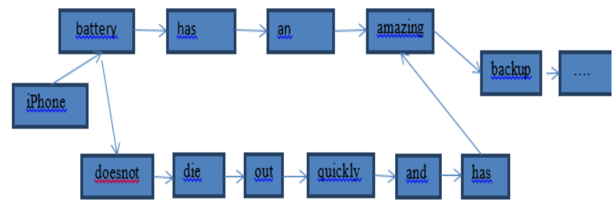


Fig. 2. Construction of graph.

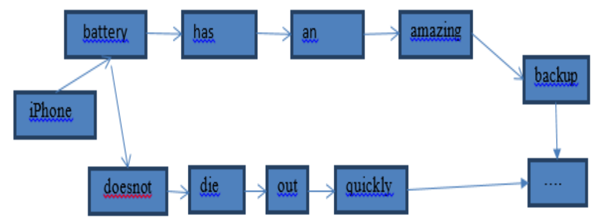


Fig. 3. Sentences fused together.

A. Abstractive-Based Summarization

The proposed work is inspired from the abstractive summarization framework as discussed by Ganesan [27]. The algorithm starts with construction of graphs from text and exploring the properties of graphs such as scoring its several subpaths, ensuring validity, and removing redundancy to finally generate candidate summaries. This differs from Ganesan methodology as it fuses sentiment by employing the classification task achieved by CNN as proposed in previous research and allows the user to view summarization results taking aspects into account instead of using SentiWordNet. Encouraging results are reported for generating summaries using the proposed graph-based techniques. Different researchers have used various techniques by exploiting graphs [28], [29].

The algorithm is given as follows.

Input: I = Opinions with explicitly mentioned aspect.

Output: O = Summaries.

The following are the steps for generating summary.

Step 1 (Constructing Graphs):

The two basic components of graphs are nodes and edges (arrows). This approach relies on constructing graphs where each node is stored which represents a token in the text and the arrows represent the adjacency of the words in the sentence. The graphs constructed which captures redundancy is shown in Fig. 2.

The examples sentences which are fused together are shown in Fig. 3.

An alternative approach is to calculate the sentiment of both the sentences to be fused and to look for a connector that can be accurately used. This sentiment is calculated using SentiWordNet 3.0 [30]. The preexisting list is checked to select the connector after the sentiment has been calculated. Generally for the contradictory sentiments, “but” is used.

TABLE I
EXAMPLE OF SENTENCE AND ITS CORRESPONDING SENTIMENT

Main Sentence		Sentiment
Iphone	Battery has an amazing backup	+ve
	Doesnot die out quickly	+ve
	connector	and

For the positive sentences, "and," "or" may be used. The illustration is shown in Table I.

Every node is a correct node represented as N $(t,\ p_t,\ p_s)$, where t denotes the part of speech of the specified token, p_t is the token's place in the sentence, and p_s is the point where it is found in the document.

Step 2 (Ensure the Accuracy of the Sentence):

In the sentence "iPhone battery has an amazing backup" and "iPhone battery does not die out quickly and has amazing backup" contain the main aspect "battery." All the paths constructed using graphs are correct if it satisfies the following set of rules.

All the nodes are linked together with directed edges; correct start node (CSN) ensures that it is the first node and correct end node (CEN) ensures the sentences are completed.

The accuracy of the sentence is ensured by the set of rules explained as follows.

/nn following /vb and an /jj
/jj following /nn and a/vb
/vb following /jj and a/nn
/rb following /jj and a /nn
/rb following /nn

where /nn is a noun, /vb is a verb, /jj is an adjective, /vb is a verb, and /rb is an adverb. Conjunction can appear in between a sentence or at the start of a sentence

Step 3 (Scoring Sentences and Merging Sentiments):

This step can be decided in two separate rounds. The first details out how to score the paths and the second explains the fusion of sentences.

The duplicity of the overlapping sentences follows the scoring of paths. The redundancy is calculated by finding the intersection point of the position of the words in the sentences, considering the difference between the positions should not exceed the threshold, P. The duplicity makes us aware of the common words in the different sentences, at each intersection in the path. The calculated overlap will give us the resultant scores. It will be used to compute the length of the path (sentence).

The nodes are checked and if they can be fused, the sentiments of those sentences are calculated. The connectors are selected accordingly from the stored list of connectors. For example, in the above Fig. 3, two sentences are taken, "iPhone battery has an amazing backup" and "iPhone battery does not die out quickly." These two sentences are connected via connector "and," referring to the common aspect "battery." Individually, the sentiments of two sentences are calculated as both can be fused and the connector "and" is used as it is connecting both the sentences accurately. The whole graph is traversed again to find further nodes. The score is regenerated and repeated sentences are discarded from the fused sentences. The new sentences are added with their scores.

Step 4 (Ranking of Sentences for Summarization):

The rest of the sentences are sorted in descending order and top n sentences are chosen to be the candidate summary (n is chosen by the user).

B. Extractive-Based Summarization

A novel approach has been proposed that generate the extractive summary from the set of reviews. It aims to reduce the number of dimensions by dropping irrelevant thematic words and relatively finding the summary of the reviews by taking the reviews according to their rank and ranking is done according to the prime aspects without any loss of information, respective to a particular domain. This method will safely remove those aspects which are not in the top priority. Instead of traversing and searching the reviews on different sites, the decision can be taken by the user quickly by looking at the summary of the corpus. Vector representation of sentences in the summary is best approximated by projecting them into the boundary of the principal components.

C. Principal Component Analysis

Karl Pearson, 1901 invented PCA. It is a statistical technique and dimension reduction tool that is used to decrease the dimension of the data set and reduce the correlation between variables. PCA identify various patterns present in the data, and represent the data in such manner that their similarities and dissimilarities are highlighted. The purpose of incorporating it for summarization is to safely eliminate many features from the data space [31]. Smith [32] discovered PCA is useful in many areas, such as face recognition and image compression. The major advantage of using PCA is after finding the patterns in data, the compression task becomes easier. Thus, dimensions can be reduced without losing important information [33]. The proposed work is to incorporate PCA in text so that the most important relevant reviews can be extracted from the list of reviews. This technique is purely novel and the analysis of the results after comparing the peer summaries with the gold summaries will prove that the proposed technique is unique and results are more accurate.

D. Algorithm

A novel approach has been proposed that generate the extractive summary from the set of reviews. It aims to determine all relevant sentences and reduces dimensions by dropping irrelevant thematic words and nonessential sentences by applying combination of PCA and SVD.

The following steps are involved in generating summary.

Input: I = Opinions with explicitly mentioned n aspect.

Output: O = Summaries.

Each n aspect will be associated with m number of reviews; these reviews are generally adjectives associated with each aspect.

Step 1: Score individual words using SentiWordNet.

Step 2: Matrix Creation for Summarization.

Step 3: Implementing PCA: Applying PCA on matrix created.

Step 4: After sorting the values of cosine similarity in descending order, the threshold is applied on the sorted values.

Details on the algorithm are explained here:

For each thematic word, sentences in which it is present we extract the adjectives and its phrases associated with it by feeding it into Stanford dependence parser. For each extracted adjective and its phrases, score from the database (SentiWordNet) is retrieved. SentiWordNet is a lexical resource for opinion mining and scores are assigned to each synset of WordNet. This dictionary calculates the scores of the sentiments denoted by positivity, negativity, and objectivity. The scoring for the words is calculated using the connectors "and" and "or" for the same features. If both the words are adjectives and connected by "and," then the scoring is proceeded by adding the scores of adjective 1 and adjective 2. It is explained below.

Adv1 and Adv2: Scoring (Adv1) + Scoring (Adv2) for connector, and (both of two adjectives)

Adv1 or Adv2: Scoring (Adv1) or Scoring (Adv2) for connector, or (Either of the two adjectives).

It has n number of thematic words. Each n thematic word will be associated with m number of reviews.

Matrix T created as follows:

$$T = ab, \ldots\ldots, n[\ ]$$

where a, b, $c \ldots, n$ are the thematic words. Values of matrix $T[ij]$ will be SentiWordNet score of sentiment extracted.

Step 1: Apply SVD (singular value decomposition) to the created matrix. The result of SVD will be three matrices S and U, and VT.

Step 2: U and V matrices will be considered up to rank K (i.e., ignoring lower sparse part of those matrices) called as Rank k approximation. The value of k is a parameter whose choice is of critical importance.

Step 3: Applying the counter for each thematic word to get the value of its frequency count, that is term frequency of each thematic word.

Step 4: The thematic word with the highest counter value is taken as query vector.

Step 5: Calculating cosine similarities such as the measure of similarity between query vector and sentence vector

$$(q, r) = qr/|q| \cdot |r|. \tag{1}$$

Step 5: Sorting the values of cosine similarity in descending order.

Step 6: Applying the threshold on sorted values.

IV. Results and Discussion

This section discusses the data set used for comparing the techniques. The results are analyzed and evaluation is done using charts and graphs.

A. Data Sets

To implement the proposed method, two data sets are used.

Data Set I: Opinosis data set [27] contains reviews, gold standard summaries, script for ROUGE, and documentation. Various topics, altogether 51, include cars, battery, and food consisting 100 reviews of different persons taken from famous sources such as Trip advisor, Edmunds.com, and Amazon.com are present in Opinosis data set.

Data Set II: Document Understanding Conference (DUC) is conducted every year in National Institute of Science and Technology (NIST). It comprises a corpus of total 50 documents, having 500 words on an average. The gold summaries are also listed with each article [25].

Data Set III: The data set taken consists of user-generated data set obtained from crawling on query "iPhone 7s." The set of four seed URL are taken from "Amazon.com, FlipKart.com, Yelpme.com, and sitejabber.com." The crawler developed collected about 55–65 response pages corresponding to each URL. In total, a sample of 200–250 pages have been gathered and analyzed, and have been taken for the experimental evaluation of our work The implementation details are discussed on first 100 opinions extracted from opinion retriever [34].

B. Experiments and Results

For the evaluation of the proposed opinion summarization work, comparison is made between two types of summaries. One of the system-generated summaries is the candidate summary that is referred to as "peer summaries" and one of the human-generated summaries is the reference summary that is referred to as "gold standard summaries" [35]. The comparison is also made between the abstractive and extractive technique. Summary evaluation method used is the N-gram Co-occurrence Statistics–ROUGE.

The quality of summary can be measured with ROUGE. Lin [36] introduced the concept of ROUGE metric which has been adopted by the DUCs and conferences help on NLP

$$\mathrm{ROUGE-N} = \frac{\sum_{S\in\{\mathrm{ReferenceSummaries}\}}\sum_{\mathrm{gram}_n\in S}\mathrm{Count}_{\mathrm{match}}(\mathrm{gram}_n)}{\sum_{S\in\{\mathrm{ReferenceSummaries}\}}\sum_{\mathrm{gram}_n\in S}\mathrm{Count}(\mathrm{gram}_n)} \tag{2}$$

where Countmatch (gramn) counts the overlapping units of n-grams that occur in candidate summary and gold summary, and Count (gramn) is the number of n-grams in the reference summaries. ROUGE-1 refers to the overlap of unigrams and ROUGE-2 refers to the overlap of bigrams between system-generated summary and gold summary.

The evaluation is performed using IR measures stated as below. Purity and entropy can also be considered as performance metrics [37]. Precision is defined as the ratio of recovered words that are appropriate to search. Recall is defined as the ratio of words that are appropriate to that which is successfully recovered. F-Measure is defined as the harmonic mean of the above-defined two ratio. Mathematically,

Precision: Appropriate words ∩ Recovered words/ Recovered words

$$\mathrm{Precision} = \mathrm{RRW}/\mathrm{RRW} + \mathrm{IRW} \tag{3}$$

where RRW is the relevant words retrieved, and IRW is the irrelevant words retrieved.

Recall: Appropriate words ∩ Recovered words/appropriate words

$$\mathrm{Recall} = \mathrm{RRW}/\mathrm{RRW} + \mathrm{RNW} \tag{4}$$

TABLE II
RESULTS WITH DATA SET I

	Abstractive based summarization		Extractive based summarization	
	ROUGE 1	ROUGE 2	ROUGE 1	ROUGE 2
Precision	0.4983	0.2529	0.0826	0.0175
Recall	0.2719	0.0791	0.5045	0.1106
F-Measure	0.3518	0.1205	0.1419	0.0302

TABLE III
RESULTS WITH DATA SET II

	Abstractive based summarization		Extractive based summarization	
	ROUGE 1	ROUGE 2	ROUGE 1	ROUGE 2
Precision	0.46	0.32	0.14	0.07
Recall	0.19	0.15	0.62	0.13
F-Measure	0.26	0.20	0.22	0.09

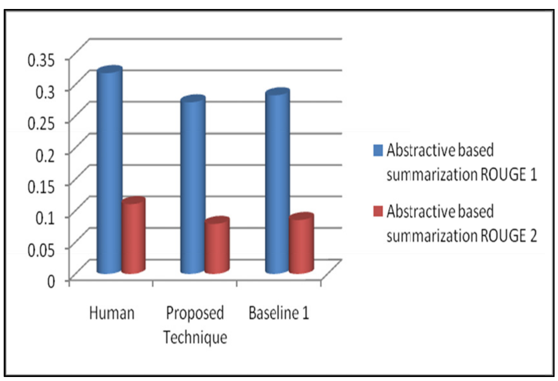


Fig. 4. Results with Recall for Abstractive Summarization on Data set I.

where RRW is the number of relevant words retrieved and RNW are relevant tweets not retrieved.

$$\text{F-Measure} : 2^{*}\text{Precision}^{*}\text{Recall}/(\text{Precision} + \text{Recall}). \quad (5)$$

C. Performance Analysis

The comparison is made between abstractive technique and the extractive technique for both the data sets explained above. It was found that extractive methods have very high recall scores as compared to precision scores. The criteria of "selecting sentences" and not much significant information in summary results in giving higher values for recall but low values for precision, respectively. The results are shown in Tables II and III.

The comparison is done for two different data sets Opinosis data set [27] and second DUC data set [25]. It was observed

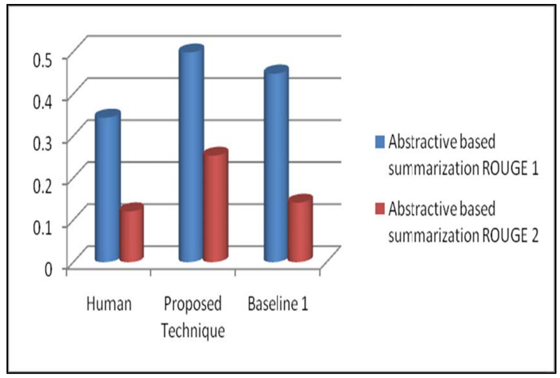


Fig. 5. Results with Precision for Abstractive Summarization on Data set I.

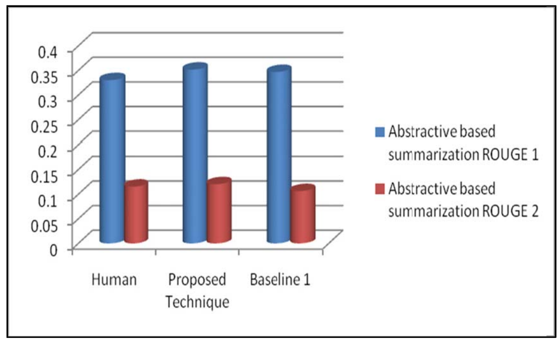


Fig. 6. Results with F-Measure for Abstractive Summarization on Data set I.

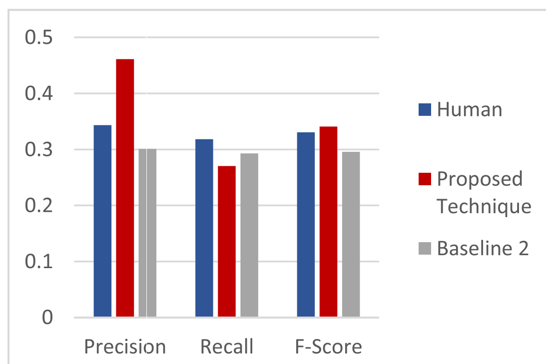


Fig. 7. Results for Abstractive Summarization on Data set II.

that better results were reported with the proposed algorithms. The results are shown from Figs. 4–6.

The above results show that the proposed technique has comparable results with human summaries. Better precision results are reported as compared with baselines taken from researches [27] and [19], as the precious researches claim to connect only those sentences where prior connector exist for fusing sentences. The proposed technique outperforms by a good margin. Fig. 5 illustrates the low results on Recall because of the existence of duplicate information.

D. Sample Summaries

The topic "food" is chosen for producing sample summaries. The data set for the same is given in the Appendix. Fig. 7 shows the summaries generated by the above-discussed

About food and service for the query 'food'

Human Summaries:
[1] Food was excellent with a wide range of choices and good services.
[2] The food is good, the service great. Very good selection of food for breakfast buffet.
Abstractive summary:
The rooms were very nice and cool with great service. The toilets were clean. Very good selection of food for breakfast. It's a chain but the food was cheap, reasonable and delicious. The food was delicious in both the consortia and forum restaurants in the hotel.

Extractive summary:
[1] The local restaurant stood out for having excellent food, Olives located one block away.
[2] Staff was busy clearing tables promptly and fresh food kept arriving all the time.
[3] The nicest surprise was how fantastic the food was which I hadn't expected in a large chain hotel.
[4] I ordered room service one night, and the food was delightful, and the serive was fast and friendly.
[5] The food is very, very good and the staff as very friendly.
[6] The hotel itself was very pleasant, the staff friendly and helpful and I can only comment on the breakfast as this was the only meal that we had there, the food was excellent with a wide variety of choice
[7] The food was excellent and the service great
[8] If you like Indian food we ate at Delhi Brassiere, it was excellent and more reasonably priced.

Fig. 8. Sample summaries.

methods. The "alikeness" is reflected between the proposed abstractive-based summary and human summary. Extractive summary has chosen eight sentences to be included in summary by ranking the sentences.

V. Conclusion

Large corpus of online posted reviews has been summarized using two different methodologies and results are measured by ROUGE-1 and ROUGE-2 scores. The proposed abstractive technique gives summarized sentences in a beautifully structured way including the relevant and important segments of the sentence. The algorithm is proposed by constructing graphs as a preliminary task that denotes the richness of representational structure of sentences. This summarization approach is also compared with the extractive technique which uses PCA by calculating eigen values and scoring the reviews in decreasing order by applying cosine similarity. Results show that extractive summarization does not require much hard work in understanding the content of opinions or understanding text in depth. It can be easily achieved by ranking and choosing thresholds that will filter out the most appropriate sentences to be included in the summary. Despite this, some important parts are left out in this method. Results are verified by calculating ROUGE scores and comparison is done using IR evaluation measures.

Future Research Directions and Open Challenges: In the future, we shall explore the applicability of the present algorithms and any potentially needed extensions in the following main directions.

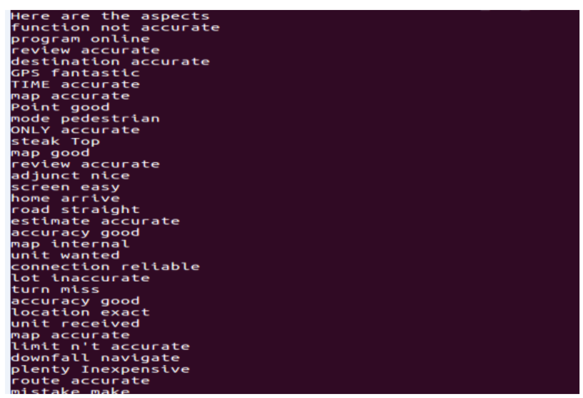
First, semiabstractive extractive techniques shall be employed to carry out the benefits of both the approaches, and the demonstration will be conducted on the fair summary length. Some improvements shall be applied in the recall and precision scores, respectively, of the preexisting extractive and abstractive techniques.

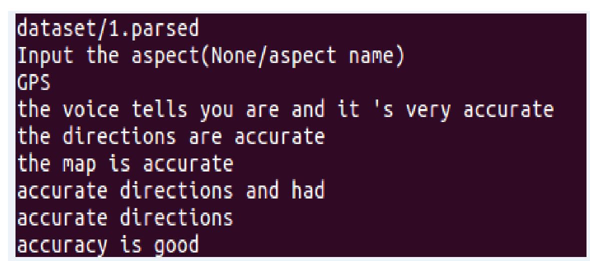
Second, selecting a threshold has a major effect on the performance of the algorithms [38]. It is necessary to study how the selected threshold affects the performance of the summarization systems as they have particular focus on many information retrieval (IR) systems. This requires various challenges, involving the formulation of mathematical equations in calculating scores reducing the redundancy of the overlapping sentences, and similar [5], [39].

Some of these aspects were tackled by the present study; nevertheless, more in-depth optimization and tuning of parameters are necessary to study their effect on the efficiency of IR system components that led to better coverage and relevant summaries.

Appendix

The aspects for Opinosis Data Set (Garmin Nuvi 255W GPS)





References

[1] C. L. Liu, W. H. Hsaio, C. H. Lee, G. C. Lu, and E. Jou, "Movie rating and review summarization in mobile environment," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 3, pp. 397–407, May 2012.
[2] J. Liu, S. Seneff, and V. Zue, "Harvesting and summarizing user-generated content for advanced speech-based HCI," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 8, pp. 982–992, Dec. 2012.
[3] V. B. Raut and D. D. Londhe, "Survey on opinion mining and summarization of user reviews on Web," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 1026–1030, 2014.
[4] J. Steinberger, M. Poesio, M. A. Kabadjov, and K. Ježek, "Two uses of anaphora resolution in summarization," *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1663–1680, 2007.

[5] M. A. Tayal, M. M. Raghuwanshi, and L. G. Malik, "ATSSC: Development of an approach based on soft computing for text summarization," *Comput. Speech Lang.*, vol. 41, pp. 214–235, Jan. 2017.
[6] D. K. Gupta and A. Ekbal, "IITP: Supervised machine learning for aspect based sentiment analysis," in *Proc. SemEval*, vol. 319, 2014, pp. 319–323.
[7] H. G. Silber and K. F. McCoy, "Efficient text summarization using lexical chains," in *Proc. 5th Int. Conf. Intell. User Interfaces*, 2000, pp. 252–255.
[8] G. Carenini and J. C. K. Cheung, "Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controversiality," in *Proc. 5th Int. Natural Lang. Gener. Conf.* Stroudsburg, PA, USA: Association Computational Linguistics, 2008, pp. 33–41.
[9] B. Gawalt, Y. Zhang, and L. El Ghaoui, "Sparse PCA for text corpus summarization and exploration," in *Proc. NIPS Workshop Low-Rank Matrix Approximation*, 2010, pp. 1–4.
[10] P. E. Genest and G. Lapalme, "Fully abstractive approach to guided summarization," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics*, vol. 2. Stroudsburg, PA, USA: Association Computational Linguistics, 2012, pp. 354–358.
[11] K. Knight and D. Marcu, "Summarization beyond sentence extraction: A probabilistic approach to sentence compression," *Artif. Intell.*, vol. 139, no. 1, pp. 91–107, 2002.
[12] D. Zajic, B. J. Dorr, J. Lin, and R. Schwartz, "Multi-candidate reduction: Sentence compression as a tool for document summarization tasks," *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1549–1570, Nov. 2007.
[13] D. Wang, T. Li, S. Zhu, and C. Ding, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2008, pp. 307–314.
[14] N. R. Sharma and V. D. Chitre, "Opinion mining, analysis and its challenges," *Int. J. Innov. Advancement Comput. Sci.*, vol. 3, no. 1, pp. 59–65, 2014.
[15] K. A. Ganesan, C. Zhai, and J. Han, "Opinosis: A graph based approach to abstractive summarization of highly redundant opinions," in *Proc. 23rd Int. Conf. Comput. Linguistics (COLING)*, 2010, pp. 1–9.
[16] R. Bhargava, Y. Sharma, and G. Sharma, "ATSSI: Abstractive text summarization using sentiment infusion," *Procedia Comput. Sci.*, vol. 89, pp. 404–411, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S187705091631153X, doi: 10.1016/j.procs.2016.06.088.
[17] F. H. Khan, U. Qamar, and S. Bashir, "A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet," *Knowl. Inf. Syst.*, vol. 51, pp. 851–872, Sep. 2016.
[18] Y. Sankarasubramaniam, K. Ramanathan, and S. Ghosh, "Text summarization using wikipedia," *Inf. Process. Manage.*, vol. 50, no. 3, pp. 443–461, May 2014.
[19] E. Lloret and M. Palomar, "Analyzing the use of word graphs for abstractive text summarization," in *Proc. 1st Int. Conf. Adv. Inf. Mining Manage.*, Barcelona, Spain, 2011, pp. 61–66.
[20] R. Bhargava and Y. Sharma, "Deep extractive text summarization," *Procedia Comput. Sci.*, vol. 167, pp. 138–146, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050920306566, doi: 10.1016/j.procs.2020.03.191.
[21] M. Zhong, P. Liu, Y. Chen, D. Wang, X. Qiu, and X. Huang, "Extractive summarization as text matching," 2020, *arXiv:2004.08795*. [Online]. Available: http://arxiv.org/abs/2004.08795
[22] Y. Zhang, Q. Liu, and L. Song, "Sentence-state LSTM for text representation," 2018, *arXiv:1805.02474*. [Online]. Available: http://arxiv.org/abs/1805.02474
[23] T. Shi, Y. Keneshloo, N. Ramakrishnan, and C. K. Reddy, "Neural abstractive text summarization with sequence-to-sequence models," 2018, *arXiv:1812.02303*. [Online]. Available: http://arxiv.org/abs/1812.02303
[24] M. Rahman and F. H. Siddiqui, "An optimized abstractive text summarization model using peephole convolutional LSTM," *Symmetry*, vol. 11, no. 10, p. 1290, Oct. 2019.
[25] Hyatt Regency Rochester, Rochester, NY, USA. *Document Understanding Conference*. Accessed: Apr. 26–27, 2007. [Online]. Available: http://duc.nist.gov
[26] A. K. M. Masum, S. Abujar, M. A. Islam Talukder, A. K. M. S. A. Rabby, and S. A. Hossain, "Abstractive method of text summarization with sequence to sequence RNNs," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2019, pp. 1–5.
[27] K. A. Ganesan, "Opinion driven decision support system," Dept. Comput. Sci., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, Tech. Rep., 2013. [Online]. Available: http://hdl.handle.net/2142/46628
[28] R. V. Kumar and K. Raghuvee, "Web user opinion analysis for product features extraction and opinion summarization," *Int. J. Web Semantics Technol.*, vol. 3, no. 4, pp. 69–82, Oct. 2012.
[29] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
[30] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," *Lrec*, vol. 10, pp. 2200–2204, May 2010.
[31] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 1999, pp. 50–57.
[32] L. I. Smith, "A tutorial on principal components analysis," Dept. Comput. Sci., Cornell Univ., Ithaca, NY, USA, Univ. Otago, Dunedin, New Zealand, Tech. Rep. OUCS-2002-12, 2002, vol. 51, no. 52, p. 65.
[33] G. Vinodhini and R. M. Chandrasekaran, "Opinion mining using principal component analysis based ensemble model for e-commerce application," *CSI Trans. ICT*, vol. 2, no. 3, pp. 169–179, 2014.
[34] S. Bhatia, M. Sharma, and K. K. Bhatia, "A novel approach for crawling the opinions from world wide Web," *Int. J. Inf. Retr. Res.*, vol. 6, no. 2, pp. 1–23, Apr. 2016.
[35] S. Bhatia, P. Chaudhary, and N. Dey, "Introduction to opinion mining," in *Opinion Mining in Information Retrieval*. Singapore: Springer, 2020, pp. 1–22.
[36] C. Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proc. Text Summarization Branches Out, Proc. ACL-Workshop*, vol. 8, 2004, p. 74–81.
[37] M. Alojail and S. Bhatia, "A novel technique for behavioral analytics using ensemble learning algorithms in E-commerce," *IEEE Access*, vol. 8, pp. 150072–150080, 2020.
[38] S. A. Babar and P. D. Patil, "Improving performance of text summarization," *Procedia Comput. Sci.*, vol. 46, pp. 354–363, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915000952, doi: 10.1016/j.procs.2015.02.031.
[39] W. Xiao and G. Carenini, "Extractive summarization of long documents by combining global and local context," 2019, *arXiv:1909.08089*. [Online]. Available: http://arxiv.org/abs/1909.08089

Surbhi Bhatia Project Management Professional certified, has rich eight years of teaching and academic experience. She is currently a Doctorate in computer science and engineering from Banasthali Vidypaith, Vanasthali, India. She is also an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Hofuf, Saudi Arabia. She has authored over seven national and international patents. She has authored more than 30 articles in reputed journals and conferences in high indexing databases. She has successfully authored two books at Springer and Wiley. She is currently editing three books at CRC Press, Elsevier, and Springer.

Dr. Bhatia is the Editorial Board Member of the *International Journal of Hybrid Intelligence* (Inderscience Publishers), *SN Applied Sciences* (Springer), and also in several IEEE conferences. She is currently serving as a Guest Editor for special issues in reputed journals. She has delivered talks as the Keynote Speaker in the IEEE conferences and in faculty development programs. She is an Active Researcher in the field of data mining, machine learning, and information retrieval.