



Feature based cluster ranking approach for single document summarization

Aakanksha Sharaff¹ · Mohit Jain¹ · Geethika Modugula¹

Received: 18 July 2021 / Accepted: 27 December 2021 / Published online: 14 January 2022
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2022

Abstract Text Summarization is a process of creating gist of large set of documents. It creates a summary which depicts the overall information contained in large text documents in a short and accurate way. A model for generating single document text summarization is presented in this paper. This model is based on extractive summarization. The proposed work extracts the informative features and generates the scoring of sentences by using similarity measure technique. Once the score of sentences is generated then clusters of sentences are formed. Clusters and sentences in each cluster are ranked and highly ranked sentences from each cluster of relative importance are included in the final summary. Summary of text document is created by identifying the important sentences from the document.

Keywords Extractive text summarization · Feature extraction · K-means clustering · Hierarchical clustering

1 Introduction

With the help of text summarization a large text document can be shortened in order to create a summary that is user readable and understandable. The summary generated

should provide important information of the document in the shortest way possible. Due to enormous amount of data present, text summarization has become an inevitable task for search engines to provide best search results. Text summarization consists of mainly three phases: interpretation, transformation and generation [1]. During interpretation, the representation of the text document to be summarized is produced. Transformation phase transforms the outcome of interpretation phase to summary and final summary is generated in generation phase. Text summarization methods [2] can be classified into following types based on the type of information and application. Based on the type of approach of summary generation, there are two types of summarization approaches [2, 3]. Extractive and Abstractive Summarization. In extractive summarization, sentences with high score are extracted from the input document and are added to the final summary without any changes. This is the easiest way to implement. This approach may result in lengthy and inconsistent summary because it does not deal with the semantics. In abstractive summarization, the sentences generated in the summary will be semantically related. Natural Language Processing is used to perform abstractive summarization. It is difficult to implement as it needs the complete understanding of the document as human. But it produces more readable summary than the summary produced by extractive summarization. Based on the type of details present in the final summary, it is divided into two types [4] Indicative summaries and Informative summaries. Based on the number of documents given as input to the summarizer, summarization is classified into two types [5] Single Document Summarization and Multi-document Summarization. In Single document Summarization, the automatic summarizer generates final summary from a single document. It takes only one document as input. It involves less

✉ Aakanksha Sharaff
asharaff.cs@nitrr.ac.in; asharaff08@gmail.com

Mohit Jain
mohitj7143@gmail.com

Geethika Modugula
geethikamodugula30@gmail.com

¹ Department of Computer Science and Engineering, National Institute of Technology, Raipur, Chhattisgarh, India

overhead. In multi-document summarization, the automatic summarizer generates final summary from multiple documents. It takes more than one document as input. By clustering the documents or by some other mechanism, summary is generated. It is difficult to implement. Care must be taken to avoid redundancy in the final summary [6]. In Single document extractive summarization, single document is taken as input. The document is segmented into sentences. Feature extraction is performed on the tokenized sentences. Sentence scoring is done as the linear combination of weights of the features extracted. To generate more readable and cohesive summary, sentences of the document are clustered followed by ranking. Sentences with relatively high score are chosen among others to generate the final summary. A good summary is that which gives complete information of the document in few sentences. It is characterized by high compression ratio and retention ratio. Retention ratio gives the amount of information retained in the generated summary. It should be much larger than the compression ratio for an ideal summary.

2 Related work

Summarizing single text document using sentence scoring is not a new idea. Several researchers worked on text summarization and its applications. As an example Nandhini and Balasundaram [7] presented a model in which for each sentence, the score is calculated using weighted combination of learner dependent and text dependent features, where the weights are calculated using genetic algorithm. To train the classifier to extract important and readable sentences based on feature vectors, a supervised learning algorithm is used. Later on Ferreira et al. [8] proposed a model in which sentences are clustered to categorize them into specified topics. The sentences are scored and the ones with highest score are selected to form the cluster. The number of selected sentences depends upon the summarization rate provided by the user. The same method was again used by Ferreira et al. [9] to experiment on news, blogs and articles by combining fifteen sentence scoring methods in different ways. Several researchers also worked on extractive summarization. An extractive model of single document text summarization based on Agglomerative nested clustering approach is proposed by Sharaff et al. [10].

In text summarization, evolutionary approach has also been explored like genetic algorithm, particle swarm optimization, cat swarm optimization etc. Benjumea and Leon [11] developed genetic clustering algorithm named SENCLUS in which sentences are clustered to represent the closeness among text topics using a peculiar fitness

function. The function is based on coverage and redundancy, and then most relevant sentences are selected to be part of the extractive summary from each topic. The multi-value or two value logic has the problem of imprecise values and ambiguity. To overcome this problem, Fuzzy Logic and WorldNet synonyms are used in the model proposed by Yadav and Meena [12]. They also overcome the issues involved in the semantics of the text.

Niu et al. [13] proposed OnSeS, which uses word2vec and neural network. It has three phases: clustering, ranking by building a graph using BM25 and generating important point of cluster using neural machine translation. It is a novel text summarization method for short texts. Jeong et al. [14] proposed an integrated framework which learns using category information and summary. To combine feature distributions, it uses a language model. It performs better than individual text summarization and classification. It is based on POS tagger and approaches using simple statistics. This makes it easy to implement and language independent. A multi variant email classification model using clustering techniques has been developed to generate categorical terms [15]; and presented a model for extractive summarization using fuzzy logic based approach [16]. Rahman and Borah [17] reviewed various extractive based approaches for query-based text summarization. Unsupervised learning methods are approaches based on document graphs, features, etc. Supervised Learning methods are SVM, KNN, Naïve Bayes Classifier, Neural Networks, etc. Yang et al. [18] used Bayesian hierarchical topic model. They distinguishes specific and general topics and indicates their relationship by examining the topic hierarchies. Instead of using term frequency or other traditional methods for keyword extraction, keywords are extracted automatically by training probability distribution of data is proposed by Thomas et al. [19].

Babar and Patil [20] used Latent Semantic Analysis to capture semantic contents in sentences. LSA includes Input Matrix Creation, Singular Value Decomposition and Sentence Selection. A metaheuristics based approach using extra tree classifier has been presented by Sharaff and Gupta to detect spam messages in email [21]. Wang et al. [22] presented a two-phase approach for long text summarization, namely, EA-LTS. It has two phases: extraction phase and abstraction phase. A query based summarization created intentionally for considering a part of input data has been proposed by Jiang et al. [23] to generate summary of only one aspect of document or conversation or tweet called as “targeted summarization”.

Text summarization has several applications in different real world problems. Generic summarization, one of the summarization technique in user intervention was used to create hierarchical summary for large heterogeneous data [24]. The authors have provided open access to their

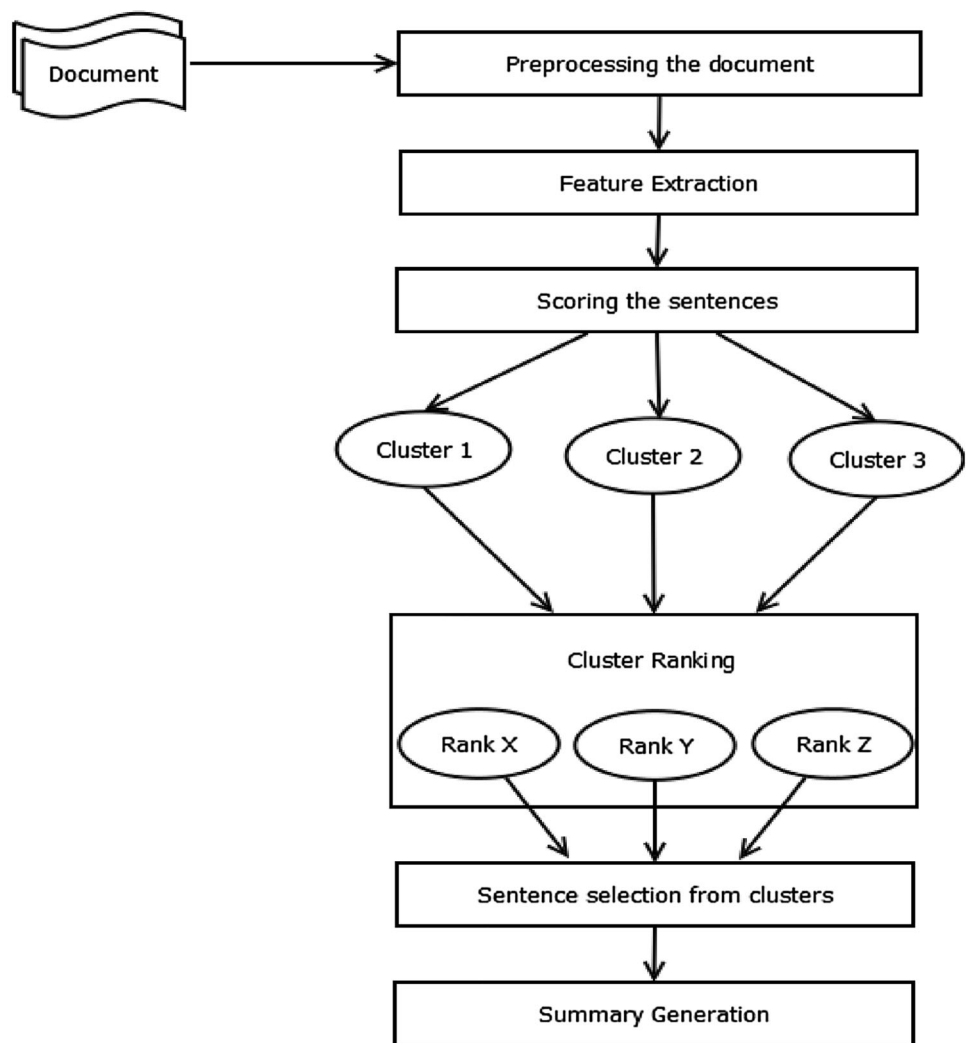
Table 1 Overview of related works

Previous Reviews	Summary generation	Comparative Analysis	Dataset	Architecture	Single/Multi Document Summarization	Performance parameters used	Techniques used
[1]	tick			tick	Single	Tick	Survey
[3]	Tick	Tick	tick		Multi	Tick	Swarm optimization techniques
[5]	Tick	Tick	Tick	tick	Both	Tick	Clustering techniques
[6]	Tick	Tick	Tick		Multi	Tick	Query based
[8]	Tick	Tick	Tick		Multi	Tick	Statistical
[27]	tick	tick	Tick	tick	Both	tick	Genetic algorithms

dataset to perform hierarchical summarization on a particular given topic. Dernoncourt et al. [25] introduced a repository to perform experiments related to creation of summary based on abstractive summarization and perform several experiments on the introduced dataset using

artificial neural network. In order to measure the aboutness of textual documents, a graph based approach has been developed by using TexRank technique [26]. Several evolutionary algorithms have also been explored to generate summary [27]. One such document summarization

Fig. 1 Overall Methodology of summary generation



has been done by using cat swarm based optimization technique. The main contribution of this paper is to develop a sentence scoring based mechanism to generate summary of most informative sentences within a document. Few recent related works and their comparison have been shown in Table 1.

3 Proposed methodology

The proposed method of text summarization walks through the following stages and presented in Fig. 1: pre-processing, feature extraction, sentence scoring, clustering, cluster ranking and summary generation. The document is pre-processed (stop words removal, stemming, etc.) and features are extracted from the sentences. Sentences are scored based on the features extracted. Then these sentences are grouped into clusters using K-means clustering or Hierarchical clustering algorithm. Clusters and sentences in each cluster are ranked and highly ranked sentences from each cluster of relative importance are included in the final summary. Algorithm 1 presents the main components in the proposed model and are discussed below:

The pseudocode of the proposed method is outlined in [Algorithm 1](#).

3.1 Pre-processing the input document

The text document may contain many words which occur very frequently but of no importance, symbols and white spaces. First of all, all the unnecessary white spaces are removed. Then tokenize the given document into sentences and sentences to words. Then stop words removal and stemming is performed.

3.1.1 Sentence segmentation

In Sentence segmentation, the given document is parsed to extract sentences from the input text. This is done by identifying sentence boundaries between words. In English, punctuation marks like full stop (.), question mark (?), exclamation (!), etc., that act as sentence boundaries. As it involves boundary recognition, sentence segmentation is also known as sentence boundary detection or sentence boundary recognition.

3.1.2 Tokenization

In Tokenization, the given sentences are broken down into words. This task is done by identifying the spaces between words.

Algorithm 1

Feature based Cluster Ranking Approach for Single Document Summarization

Input: Set of documents D_1, D_2, \dots, D_n .

Output: Top selected sentences to generate summary of document.

1. Preprocessing of documents
For each document
 - 1a) Parse and perform sentence segmentation
 - 1b) Perform Stopping and Stemming to remove irrelevant words/tokens.
 2. Feature extraction
Different features based on sentence length, sentence position, Similarity with the title, Proper nouns, Term weight, and Dates and numerical data are extracted.
 3. Sentence has been scored based on the weights assigned to each feature.
 4. Clustering of Sentences: Sentences are grouped together based on similar properties using K-means and Hierarchical Clustering
 5. Cluster Ranking: After forming the group of similar sentences, clusters are ranked based on position of sentences in document.
 6. Sentence Selection from Clusters: Sentences were selected based on the centroid calculation and generation of scores of each sentences.
 7. Summary Generation: The sentences selected from all clusters will comprise to generate summary of the document.
-

3.1.3 Stop words removal

Stop words are natural language words that occur very frequently but do not add any importance to the meaning of a sentence/document. They are mostly used to keep a sentence grammatically and syntactically correct. Stop words have to be removed because their presence may mislead the results. For example, if term frequency is calculated with stop words present, stop words will be getting more score than relevant words.

3.1.4 Stemming

Stemming is the process of transforming derived words to root words. A word can be represented in different forms in the same document like amaze, amazed and amazing. All the three words came from the same root word 'amaze'. Stemming is performed to normalize words in the document.

3.2 Feature extraction

In feature extraction, different features of a sentence are explored by its properties. Features used in this model are:

3.2.1 Sentence length

The length of a sentence should be of medium size. A longer sentence may contain unnecessary information. It occupies more space in the summary providing less important information. In an efficient summary, no of words should be as minimum as possible. Short sentences may not provide much of information. Sentence length is computed as the proportion of number of terms (words) in the sentence to the number of terms in the longest distance.

3.2.2 Sentence position

The sentences which are at the start and end of the document are relatively more important. Starting sentence in the paragraph provide more important information. First and last sentence are given the highest score. Second sentence from starting and second sentence from ending are given next highest score and so on.

3.2.3 Similarity with the title

The sentences which contain the words in the title provide more relevant information of the document. The score for similarity with the title is computed as the proportion of

number of terms that match with the title with the total number of terms in the title.

3.2.4 Proper nouns

The presence of proper nouns in a sentence makes it more important and relevant in the document. The score is computed as the proportion of number of proper nouns in a sentence to the number of terms in that sentence.

3.2.5 Term weight

Term weight is calculated based on term frequencies of words. It is the ratio of term frequencies of all words in a sentence to maximum of sum of term frequencies of a sentence in the given document.

3.2.6 Dates and numerical data

Sentences which contain dates and numerical data provide important information of a document. The ratio of number of numerical data/ dates in a sentence to the number of terms in a sentence is computed as score of the sentence.

3.3 Scoring the sentences

The scoring of sentences is performed using the features extracted above. Few or all among the above features are selected and weights are assigned to each feature to calculate the score of sentences. It is calculated as weighted linear combination of the features extracted (Nandhini and Balasundaram, 2013). For example, if sentence length, sentence position and term weight are considered then score of the sentence is computed by the formula in Eq. 1 and 2:

$$S(i) = \alpha_1.F_1 + \alpha_2.F_2 + \alpha_3.F_3 \quad (1)$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \quad (2)$$

where α_1 = Weight of sentence length feature, F_1 = Sentence length feature score, α_2 = Weight of sentence position feature, F_2 = Sentence position feature score, α_3 = Weight of term weight feature, F_3 = Term weight feature score.

If n features are considered, the generalized formula would be presented in Eqs. 3 and 4:

$$S(i) = \sum_{k=1}^n \alpha_k.F_k \quad (3)$$

$$\sum_{k=1}^n \alpha_k = 1 \quad (4)$$

where α_k = Weight of feature k, F_k = Score of feature k.

3.4 Clustering of sentences

To generate an efficient summary, cohesion between sentences is as important as selecting important sentences in the document. In order to achieve cohesion between sentences, clustering of sentences is performed. For this, calculate cosine similarity between two sentences. It measures angular distance between two sentences. It is calculated using the following formula in Eq. 5:

$$\text{Cos}(S_i, S_j) = \frac{2 \times \text{number of words that are both in } S_i \text{ and } S_j}{\text{number of words in } S_i \times \text{number of words in } S_j} \quad (5)$$

Two clustering techniques used in the proposed model are:

3.4.1 K-means clustering

Firstly the number of clusters (n) to be formed is decided in K-means clustering. Select n sentences with highest score and make them as initial centroids for each cluster. Now for each sentence, select the cluster to which it belongs with the help of cosine similarity measure. The sentence to which cluster it belongs is decided local optimally. Now, calculate the closeness of each cluster by measuring distance between each sentence and its centroid. If it is below the threshold value, then the cluster formed is good enough. Otherwise, choose different centroids and continue the entire process for fixed number of iterations.

3.4.2 Hierarchical clustering

In hierarchical clustering, title or the sentence with highest score is chosen as the root. Then n sentences are selected which are similar to the root as children of the root. Now, for each of these n children, select n sentences which are similar to them as children and so on. After forming the tree of the entire document, the path from root to leaf gives the total information of the document and sentences in each level provides cohesion between distances. The sentence (child of a node) at each level is selected in a greedy manner.

3.5 Cluster ranking

The clusters formed above are ranked to select sentences proportionally for the final summary based on its rank. Ranking is done on the basis of sentence position in the document and its score. The sentence order of the original document is maintained in the final summary.

3.6 Sentence selection from clusters

The sentences in each cluster are ranked based on its similarity with centroid sentence. The centroid sentence is given the highest rank in the cluster. Then find the next sentence which is most similar to the centroid and so on. Sentence selection from clusters is based on the ranking of sentences. More sentences are selected from the cluster which contains more sentences with highest score. It is calculated using the following formula in Eq. 6:

$$R(C_k) = \frac{\text{sum of scores of sentences in a cluster}}{\text{No.of sentences in the cluster}} \quad (6)$$

where C_k is a cluster.

3.7 Summary generation

After selecting the sentences for final summary, place them in the order of rank of the cluster they belong to. This will generate a summary with high compression ratio without any information loss.

4 Experimental analysis

The proposed model is tested using DUC (Document Understanding Conference) dataset (<https://duc.nist.gov/>) [28]. The dataset consist of text documents to be summarized and reference summaries written by human (abstractive summary). Randomly 20 documents are taken from DUC dataset and summaries are generated using the proposed model. ROUGE-1 and ROUGE-2 techniques have been used for evaluation purpose. ROUGE-N (Meena and Gopalani, 2014) is computed as follows in Eq. 7:

$$\text{ROUGE - N} = \frac{\sum_{s \in (\text{ref summary})} \sum_{\text{gram}_n \in (s)} \text{count}_{\text{match}}(\text{gram}_n)}{\sum_{s \in (\text{ref summary})} \sum_{\text{gram}_n \in (s)} \text{count}(\text{gram}_n)} \quad (7)$$

The average precision, recall and F-measure of Hierarchical clustering technique using Rouge-1 are 0.67, 0.49 and 0.56 respectively. Figure 2 shows the graphical representation of precision, recall and F-measure of summary generated by Hierarchical clustering technique using Rouge-1.

The average precision, recall and F-measure of Hierarchical clustering technique using Rouge-2 are 0.56, 0.40 and 0.47 respectively. Figure 3 shows the graphical representation of precision, recall and F-measure of summary generated by Hierarchical clustering technique using Rouge-2.

Fig. 2 Precision, Recall and F-measure of summary generated using Hierarchical clustering technique (Rouge-1)

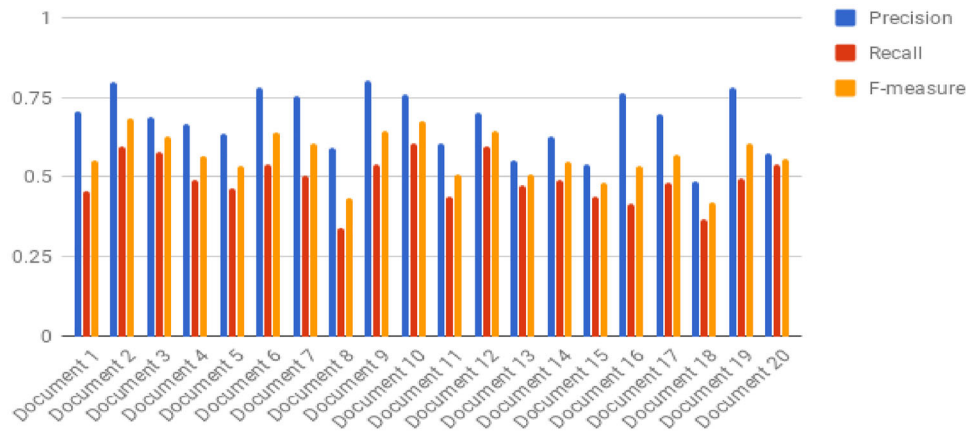


Fig. 3 Precision, Recall and F-measure of summary generated using Hierarchical clustering technique (Rouge-2)

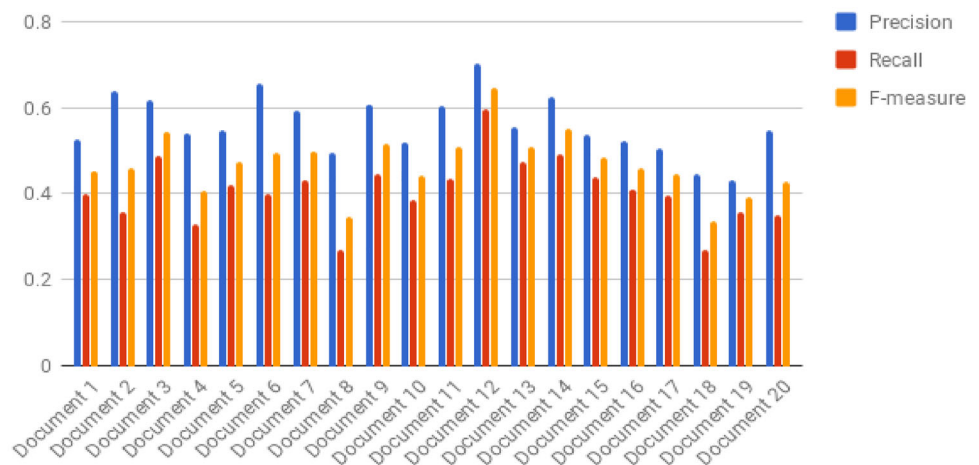
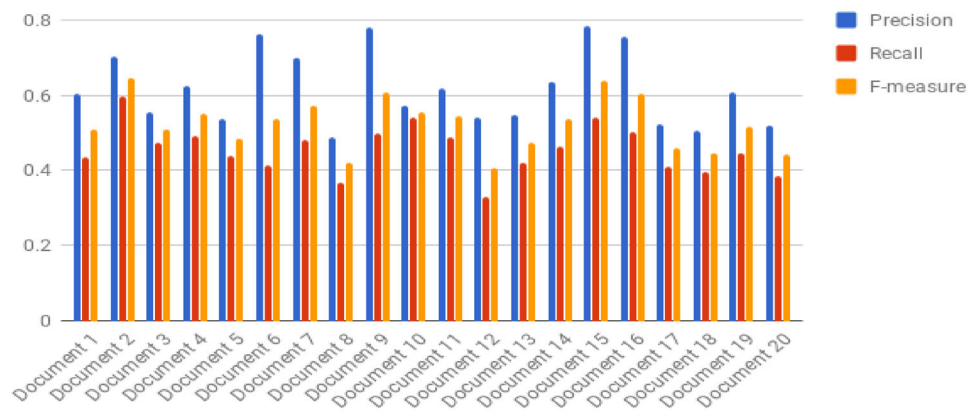


Fig. 4 Precision, Recall and F-measure of summary generated using K-means clustering technique (Rouge-1)



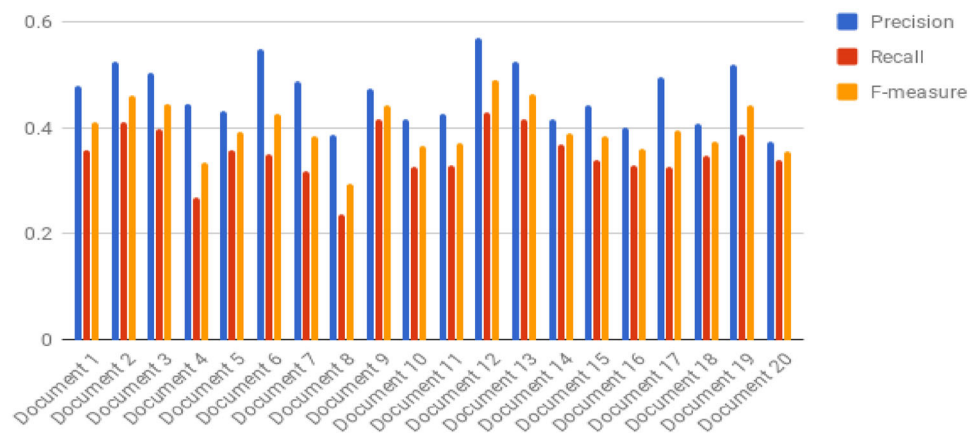
The average precision, recall and F-measure of K-means clustering technique using Rouge-1 are 0.61, 0.45 and 0.52 respectively and presented in Fig. 4.

The average precision, recall and F-measure of K-means clustering technique using Rouge-2 are 0.46, 0.35 and 0.40 respectively and presented in Fig. 5.

5 Discussions on results

It has been observed from the experimental analysis that feature extraction with hierarchical clustering technique gives better summary as compared to feature extraction with k-means clustering technique. The average precision of Hierarchical clustering model is 0.67 and k-means clustering model is 0.61 using Rouge-1 metric while that of

Fig. 5 Precision, Recall and F-measure of summary generated using K-means clustering technique (Rouge-2)



Rouge-2 metric is 0.56 and 0.46 respectively. The precision of the model decreases as n value in the n-gram increases as extractive summaries are compared with abstractive summaries. The weights assigned to feature scores of a sentence play an important role in generating summary. In the proposed model, more weightage is given to term weight, number of proper nouns in the sentence and sentence similarity with the title as compared to other features. Proper assignment of weights to features and clustering techniques used in the proposed model made the resultant summary comparable with the human generated abstractive summary.

6 Conclusion

Summarization using feature extraction of sentences and clustering techniques has been proposed. Feature extraction is used in calculating the score of sentences. The score of a sentence is computed as linear combination of weights of the features. Clustering is performed to improve continuity among the sentences in the final summary. The quality of the generated summary is highly affected by weights assigned to features. In the two clustering techniques used, hierarchical clustering performs better than k-means clustering.

7 Open research

In future, the proposed model could be extended for multi document summarization and query focused summarization. Sentence Reduction or trimming techniques can be employed to generate summary similar to human written summaries.

Acknowledgements The authors would like to thanks National Institute of Technology Raipur for providing necessary infrastructure and facilities to carry research work.

Authors' contributions AS conceived and designed the analysis, collected the data, designed the model and the computational framework, perform experiments and analyzed the obtained results. MJ, and GM contributed data or analysis tools, performed the calculations, conceived, planned, and carried out the experiments. AS wrote the manuscript and oversaw overall direction and planning of research.

Availability of data and material It is not possible to share research data publicly but depending upon the reader's interest, data will be shared privately.

Code availability It is not possible to share research code publicly but depending upon the reader's interest, data will be shared privately.

Declarations

Conflicts of interest The authors declare that they have no competing interests.

References

1. Altmami NI, Menai MEB (2020) Automatic summarization of scientific articles: a survey. *J King Saud Univ Comput Inf Sci* 2:2
2. Rahimi SR, Mozhdehi AT, Abdolahi M (2018) An overview on extractive text summarization. *In 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pp. 0054–0062, IEEE
3. Mosa MA, Anwar AS, Hamouda A (2019) A survey of multiple types of text summarization with their satellite contents based on swarm intelligence optimization algorithms. *Knowl-Based Syst* 163:518–532
4. Kumar A, Sharma A, Sharma S, Kashyap S (2017) Performance analysis of keyword extraction algorithms assessing extractive text summarization. *In Computer, Communications and Electronics (Comptelix), 2017 International Conference on*, pp. 408–414, IEEE, July 2017.

5. Bewoor MS, Patil SH (2018) Empirical analysis of single and multi document summarization using clustering algorithms. *Eng Technol Appl Sci Res* 8(1):2562–2567
6. Abdi A, Shamsuddin SM, Aliguliyev RM (2018) QMOS: query-based multi-documents opinion-oriented summarization. *Inf Process Manage* 54(2):318–338
7. Nandhini K, Balasundaram SR (2013) Improving readability through extractive summarization for learners with reading difficulties. *Egypt Inf J* 14(3):195–204
8. Ferreira R, de Souza Cabral L, Freitas F, Lins RD, de França Silva G, Simske SJ, Favaro L (2014) A multi-document summarization system based on statistics and linguistic treatment. *Exp Syst Appl* 41(13):5780–5787
9. Ferreira R, Freitas F, de Souza Cabral L, Lins RD et al (2014) A context based text summarization system. *In Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on*, pp. 66–70, IEEE, April 2014
10. Sharaff A, Shrawgi H, Arora P, Verma A (2016) Document summarization by agglomerative nested clustering approach. *In Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016 IEEE International Conference on*, pp. 187–191, IEEE, December 2016.
11. Benjumea SS, León E (2015) Genetic clustering algorithm for extractive text summarization. *In IEEE Computational Intelligence, 2015 IEEE Symposium Series on* pp. 949–956, December 2015.
12. Yadav J, Meena YK (2016) Use of fuzzy logic and wordnet for improving performance of extractive automatic text summarization. *In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*, pp. 2071–2077, IEEE, September, 2016
13. Niu J, Zhao Q, Wang L, Chen H, Atiquzzaman M, Peng F (2016) OnSeS: A novel online short text summarization based on BM25 and neural network. *In Global Communications Conference (GLOBECOM), 2016 IEEE* (pp. 1–6)
14. Jeong H, Ko Y, Seo J (2016) How to improve text summarization and classification by mutual cooperation on an integrated framework. *Expert Syst Appl* 60:222–233
15. Sharaff A, Nagwani NK (2020) ML-EC2: an algorithm for multi-label email classification using clustering. *Int J Web-Based Learn Teach Technol (IJWLTT)* 15(2):19–33
16. Sharaff A, Khaire AS, Sharma D (2019) Analysing fuzzy based approach for extractive text summarization. *In 2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pp. 906–910
17. Rahman N, Borah B (2015) A survey on existing extractive techniques for query-based text summarization. *In Advanced Computing and Communication (ISACC), 2015 International Symposium on*, pp. 98–102
18. Yang G, Wen D, Chen NS, Sutinen E (2015) A novel contextual topic model for multi-document summarization. *Expert Syst Appl* 42(3):1340–1352
19. Thomas JR, Bharti SK, Babu KS (2016) Automatic keyword extraction for text summarization in e-newspapers. *In Proceedings of the International Conference on Informatics and Analytics (p. 86)*. ACM, August 2016
20. Babar SA, Patil PD (2015) Improving performance of text summarization. *Proc Comput Sci* 46:354–363
21. Sharaff A, Gupta H (2019) Extra-tree classifier with meta-heuristics approach for email classification. *Advances in computer communication and computational sciences*. Springer, Singapore, pp 189–197
22. Wang S, Zhao X, Li B, Ge B, Tang D (2017) Integrating extractive and abstractive models for long text summarization. *In Big Data (BigData Congress), 2017 IEEE International Congress on* (pp. 305–312). IEEE, June 2017.
23. Jiang Y, Finegan-Dollak C, Kummerfeld JK, Lasecki W (2018) Effective crowdsourcing for a new type of summarization task. *In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol 2 (Short Papers)*, pp. 628–633
24. Tauchmann C, Arnold T, Hanselowski A, Meyer CM, Mieskes M (2018) Beyond generic summarization: A multi-faceted hierarchical summarization corpus of large heterogeneous data. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, May, 2018.
25. Dernoncourt F, Ghassemi M, Chang W (2018) A repository of corpora for summarization. *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*
26. Mallick C, Das AK, Dutta M, Das AK, Sarkar A (2019) Graph-based text summarization using modified TextRank. *Soft computing in data analytics*. Springer, Singapore, pp 137–146
27. Vázquez E, Arnulfo Garcia-Hernandez R, Ledeneva Y (2018) Sentence features relevance for extractive text summarization using genetic algorithms. *J Intell Fuzzy Syst* 35(1):353–365
28. Meena YK, Gopalani D (2014) Analysis of sentence scoring methods for extractive automatic text summarization. *In Proceedings of the 2014 International Conference on Information and Communication Technology for Competitive Strategies (p. 53)*. ACM, November, 2014