MMI Diversity Based Text Summarization

Mohammed Salem Binwahlan moham2007med@yahoo.com
Faculty of Computer Science and Information Systems
University Teknologi Malaysia
Skudai, Johor, 81310, Malaysia

Naomie Salim naomie@utm.my
Faculty of Computer Science and Information Systems
University Teknologi Malaysia
Skudai, Johor, 81310, Malaysia

Ladda Suanmali ladda_sua@dusit.ac.th
Faculty of Science and Technology,
Suan Dusit Rajabhat University
295 Rajasrima Rd, Dusit, Bangkok, 10300, Thailand

Abstract

The search for interesting information in a huge data collection is a tough job frustrating the seekers for that information. The automatic text summarization has come to facilitate such searching process. The selection of distinct ideas "diversity" from the original document can produce an appropriate summary. Incorporating of multiple means can help to find the diversity in the text. In this paper, we propose approach for text summarization, in which three evidences are employed (clustering, binary tree and diversity based method) to help in finding the document distinct ideas. The emphasis of our approach is on controlling the redundancy in the summarized text. The role of clustering is very important, where some clustering algorithms perform better than others. Therefore we conducted an experiment for comparing two clustering algorithms (K-means and complete linkage clustering algorithms) based on the performance of our method, the results shown that k-means performs better than complete linkage. In general, the experimental results shown that our method performs well for text summarization comparing with the benchmark methods used in this study.

Keywords: Binary tree, Diversity, MMR, Summarization, Similarity threshold.

1. INTRODUCTION

The searching for interesting information in a huge data collection is a tough job frustrating the seekers for that information. The automatic text summarization has come to facilitate such searching process. It works on producing a short form for original document in form of summary. The summary performs a function of informing the user about the relevant documents to his or her need. The summary can reduce the reading time and providing quick guide to the interesting information.

In automatic text summarization, the selection process of the distinct ideas included in the document is called diversity. The diversity is very important evidence serving to control the redundancy in the summarized text and produce more appropriate summary. Many approaches have been proposed for text summarization based on the diversity. The pioneer work for diversity based text summarization is MMR (maximal marginal relevance), it was introduced by Carbonell and Goldstein [2], MMR maximizes marginal relevance in retrieval and summarization. The sentence with high maximal relevance means it is highly relevant to the query and less similar to the already selected sentences. Our modified version of MMR maximizes the marginal importance and minimizes the relevance. This approach treats sentence with high maximal importance as one that has high importance in the document and less relevance to already selected sentences.

MMR has been modified by many researchers [4; 10; 12; 13; 16; 21; 23]. Our modification for MMR formula is similar to Mori et al.'s modification [16] and Liu et al.'s modification [13] where the importance of the sentence and the sentence relevance are added to the MMR formulation. Ribeiro and Matos [19] proved that the summary generated by MMR method is closed to the human summary, motivating us to choose MMR and modify it by including some documents features. The proposed approach employs two evidences (clustering algorithm and a binary tree) to exploit the diversity among the document sentences. Neto et al. [17] presented a procedure for creating approximate structure for document sentences in the form of a binary tree, in our study, we build a binary tree for each cluster of document sentences, where the document sentences are clustered using a clustering algorithm into a number of clusters equal to the summary length. An objective of using the binary tree for diversity analysis is to optimize and minimize the text representation; this is achieved by selecting the most representative sentence of each sentences cluster. The redundant sentences are prevented from getting the chance to be candidate sentences for inclusion in the summary, serving as penalty for the most similar sentences. Our idea is similar to Zhu et al.'s idea [25] in terms of improving the diversity where they used absorbing Markov chain walks.

The rest of this paper is described as follows: section 2 presents the features used in this study, section 3 discusses the importance and relevance of the sentence, section 4 discusses the sentences clustering, section 5 introduces the document-sentence tree building process using k-means clustering algorithm, section 6 gives full description of the proposed method, section 7 discusses the experimental design, section 8 presents the experimental results, section 9 shows a comparison between two clustering algorithms based on the proposed method performance. Section 10 concludes our work and draws the future study plan.

2. SENTENCE FEATURES

The proposed method makes use of eight different surface level features; these features are identified after the preprocessing of the original document is done, like stemming using porter's stemmer[1] and removing stop words. The features are as follows.

a. Word sentence score (WSS): it is calculated using the summation of terms weights (TF-ISF, calculated using eq. 1, [18]) of those terms synthesizing the sentence and occur in at least in a number of sentences equal to half of the summary length (LS) divided by highest term weights (TF-ISF) summation of a sentence in the document (HTFS) as shown in eq. 2, the idea of making the calculation of word sentence score under the condition of occurrence of its term in specific number of sentences is supported by two factors: excluding the unimportant terms and applying the mutual reinforcement principle [24]. MAN`A-LO`PEZ *et al*. [15] calculated the sentence score as proportion of the square of the query-word number of a cluster and the total number of words in that cluster.

[1] http://www.tartarus.org/martin/PorterStemmer/

Term frequency-inverse sentence frequency (TF-ISF) [18], term frequency is very important feature; its first use dates back to fifties [14] and still used.

$$W_{ij} = tf_{ij} \times isf = tf\ (t_{ij}, s_i)\left[1 - \frac{\log(sf\ (t_{ij}) + 1)}{\log(n+1)}\right] \quad (1)$$

Where W_{ij} is the term weight (TF-ISF) of the term t_{ij} in the sentence s_i.

$$WSS(S_i) = 0.1 + \frac{\sum_{t_j \in S_i} W_{ij}}{HTFS} \quad | no.of\ \ sentences\ containing\ t_j >= \frac{1}{2}LS \quad (2)$$

Where 0.1 is minimum score the sentence gets in the case its terms are not important.

b. Key word feature: the top 10 words whose high TF-ISF (eq. 1) score are chosen as key words [8; 9]. Based on this feature, any sentence in the document is scored by the number of key words it contains, where the sentence receives 0.1 score for each key word.

c. Nfriends feature: the nfriends feature measures the relevance degree between each pair of sentences by the number of sentences both are similar to. The friends of any sentence are selected based on the similarity degree and similarity threshold [3].

$$Nfriends(s_i, s_j) = \frac{s_i\,(friends) \cap s_j\,(friends)}{|\,s_i\,(friends) \cup s_j\,(friends)\,|} \,|\, i \neq j \quad (3)$$

d. Ngrams feature: this feature determines the relevance degree between each pair of sentences based on the number of n-grams they share. The skipped bigrams [11] used for this feature.

$$Ngrams(s_i, s_j) = \frac{s_i\,(ngrams) \cap s_j\,(ngrams)}{|\,s_i\,(ngrams) \cup s_j\,(ngrams)\,|} \,|\, i \neq j \quad (4)$$

e. The similarity to first sentence (sim_fsd): This feature is to score the sentence based on its similarity to the first sentence in the document, where in news article, the first sentence in the article is very important sentence [5]. The similarity is calculated using eq. 11.

f. Sentence centrality (SC): the sentence has broad coverage of the sentence set (document) will get high score. Sentence centrality widely used as a feature [3; 22]. We calculate the sentence centrality based on three factors: the similarity, the shared friends and the shared ngrams between the sentence in hand and all other the document sentences, normalized by n-1, n is the number of sentences in the document.

$$SC(S_i) = \frac{\sum_{j=1}^{n-1} sim(S_i, S_j) + \sum_{j=1}^{n-1} nfriends(S_i, S_j) + \sum_{j=1}^{n-1} ngrams(S_i, S_j)}{n-1} \,|\, i \neq j\ and\ sim(S_i, S_j) >= \theta \quad (5)$$

Where S_j is a document sentence except S_i, n is the number of sentences in the document. θ is the similarity threshold which is determined empirically, in an experiment was run to determine the best similarity threshold value, we have found that the similarity threshold can take two values, 0.03 and 0.16.

The following features are for those sentences containing ngrams [20] (consecutive terms) of title where n=1 in the case of the title contains only one term, n=2 otherwise:

g. Title-help sentence (THS): the sentence containing n-gram terms of title.

$$THS(s_i) = \frac{s_i(ngrams) \cap T(ngrams)}{|s_i(ngrams) \cup T(ngrams)|} \quad (6)$$

h. Title-help sentence relevance sentence (THSRS): the sentence containing ngram terms of any title-help sentence.

$$THSRS(s_j) = \frac{s_j(ngrams) \cap THS(s_i(ngrams))}{|s_j(ngrams) \cup THS(s_i(ngrams))|} \quad (7)$$

The sentence score based on THS and THSRS is calculated as average of those two features:

$$SS\_NG = \frac{THS(s_i) + THSRS(s_i)}{2} \quad (8)$$

3. THE SENTENCE IMPORTANCE (IMPR) AND SENTENCE RELEVANCE (REL)

The sentence importance is the main score in our study; it is calculated as linear combination of the document features. Liu *et al*. [13] computed the sentence importance also as linear combination of some different features.

$$IMPR(S_i) = avg(WSS(S_i) + SC(S_i) + SS\_NG(S_i) + sim\_fsd(S_i) + kwrd(S_i)) \quad (9)$$

Where *WSS*: word sentence score, SC: sentence centrality, *SS_NG*: average of THS and THSRS features, *Sim_fsd*: the similarity of the sentence s_i with the first document sentence and $kwrd(S_i)$ is the key word feature.

The sentence relevance between two sentences is calculated in [13] based on degree of the semantic relevance between their concepts, but in our study the sentence relevance between two sentences is calculated based on the shared friends, the shared ngrams and the similarity between those two sentences:

$$Rel(s_i, s_j) = avg(nfriends(s_i, s_j) + ngrams(s_i, s_j) + sim(s_i, s_j)) \quad (10)$$

4. SENTENCES CLUSTERING

The clustering process plays an important role in our method; it is used for grouping the similar sentences each in one group. The clustering is employed as an evidence for finding the diversity among the sentences. The selection of clustering algorithm is more sensitive needing to experiment with more than one clustering algorithm. There are two famous categories of the clustering methods: partitional clustering and hierarchical clustering. The difference between those two categories is that hierarchical clustering tries to build a tree-like nested structure partition of clustered data while partitional clustering requires receiving the number of clusters

then separating the data into isolated groups [7]. Example of the hierarchical clustering methods is agglomerative clustering methods like Single linkage, complete linkage, and group average linkage. We have tested our method using k-means clustering algorithm and complete linkage clustering algorithm.

5. DOCUMENT-SENTENCE TREE BUILDING (DST) USING K-MEANS CLUSTERING ALGORITHM

The first stage for building the document-sentence tree is to cluster the document sentences into a number of clusters. The clustering is done using k-means clustering algorithm. The clusters number is determined automatically by the summary length (number of sentences in the final summary). The initial centroids are selected as the following:

- Pick up one sentence which has higher number of similar sentences (sentence friends).
- Form a group for the picked up sentence and its friends, the maximum number of sentences in that group is equal to the total number of document sentences divided by the number of clusters.
- From the created group of sentences, the highest important sentence is selected as initial centroid.
- Remove the appearance of each sentence in the created group from the main group of document sentences.
- Repeat the same procedure until the number of initial centroids selected is equal to the number of clusters.

To calculate the sentence similarity between two sentences $s_i and\ s_j$, we use *TF-ISF* and *cosine* similarity measure as in eq. 11 [3]:

$$sim(s_i, s_j) = \frac{\sum_{w_i \in s_i, s_j} tf(w_i, s_i) tf(w_i, s_j) \left[1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}\right]^2}{\sqrt{\sum_{w_i \in s_i} \left(tf(w_i, s_i) \left[1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}\right]\right)^2} \times \sqrt{\sum_{w_i \in s_j} \left(tf(w_i, s_j) \left[1 - \frac{\log(sf(w_i)+1)}{\log(n+1)}\right]\right)^2}} \qquad (11)$$

Where *tf* is term frequency of term w_i in the sentence $s_i or\ s_j$, *sf* is number of sentences containing the term w_i in the document, *n* is number of sentences in the document.

Each sentences cluster is represented as one binary tree or more. The first sentence which is presented in the binary tree is that sentence with higher number of friends (higher number of similar sentences), then the sentences which are most similar to already presented sentence are selected and presented in the same binary tree. The sentences in the binary tree are ordered based on their scores. The score of the sentence in the binary tree building process is calculated based on the importance of the sentence and the number of its friends using eq. 12. The goal of incorporating the importance of sentence and number of its friends together to calculate its score is to balance between the importance and the centrality (a number of high important friends).

$$Score_{BT}(s_i) = impr(s_i) + (1 - (1 - impr(s_i) \times friendsNo(s_i))) \qquad (12)$$

Where $Score_{BT}(s_i)$ is the score of the s_i sentence in the binary tree building process, $impr(s_i)$ is importance of the sentence s_i and $friendsNo(s_i)$ is the number of sentence friends.

Each level in the binary tree contains 2^{ln} of the higher score sentences, where ln is the level number, ln=0, 1, 2, ….., n, the top level contains one sentence which is a sentence with highest

score. In case, there are sentences remaining in the same cluster, a new binary tree is built for them by the same procedure.

6. METHODOLOGY

The proposed method for summary generation depends on the extraction of the highest important sentences from the original text, we introduce a modified version of MMR, and we called it MMI (maximal marginal importance). MMR approach depends on the relevance of the document to the query, and it is for query based summary. In our modification we have tried to release this restriction by replacing the query relevance with sentence importance for presenting the MMI as generic summarization approach.

Most features used in this method are accumulated together to show the importance of the sentence, the reason for including the importance of the sentence in the method is to emphasize on the high information richness in the sentence as well as high information novelty. We use the tree for grouping the most similar sentences together in easy way, and we assume that the tree structure can take part in finding the diversity.

MMI is used to select one sentence from the binary tree of each sentence cluster to be included in the final summary. In the binary tree, a level penalty is imposed on each level of sentences which is 0.01 times the level number. The purpose of the level penalty is to reduce the noisy sentences score. The sentences which are in the lower levels are considered as noisy sentences because they are carrying low scores. Therefore the level penalty in the low levels is higher while it is low in the high levels. We assume that this kind of scoring will allow to the sentence with high importance and high centrality to get the chance to be a summary sentence, this idea is supported by the idea of PageRank used in Google [1] where the citation (link) graph of the web page or backlinks to that page is used to determine the rank of that page. The summary sentence is selected from the binary tree by traversing all levels and applying MMI on each level sentences.

$$MMI(S_i) = Arg \max_{S_i \in CS \setminus SS} \left[(Score_{BT}(S_i) - \beta(S_i)) - \max_{S_j \in SS} (\mathrm{Rel}(S_i, S_j)) \right] \qquad (13)$$

Where $\mathrm{Rel}(S_i, S_j)$ is the relevance between the two competitive sentences, S_i is the unselected sentence in the current binary tree, S_j is the already selected sentence, SS is the list of already selected sentences, CS is the competitive sentences of the current binary tree and β is the penalty level.

In MMR, the parameter λ is very important, it controls the similarity between already selected sentences and unselected sentences, and where setting it to incorrect value may cause creation of low quality summary. Our method pays more attention for the redundancy removing by applying MMI in the binary tree structure. The binary tree is used for grouping the most similar sentences in one cluster, so we didn't use the parameter λ because we just select one sentence from each binary tree and leave the other sentences.

Our method is intended to be used for single document summarization as well as multi-documents summarization, where it has the ability to get rid of the problem of some information stored in single document or multi-documents which inevitably overlap with each other, and can extract globally important information. In addition to that advantage of the proposed method, it maximizes the coverage of each sentence by taking into account the sentence relatedness to all other document sentences. The best sentence based on our method policy is the sentence that has higher importance in the document, higher relatedness to most document sentences and less similar to the sentences already selected as candidates for inclusion in the summary.

7. EXPERIMENTAL DESIGN

The Document Understanding Conference (DUC) [6] data collection became as standard data set for testing any summarization method; it is used by most researchers in text summarization. We have used DUC 2002 data to evaluate our method for creating a generic 100-word summary, the task 1 in DUC 2001 and 2002, for that task, the training set comprised 30 sets of approximately 10 documents each, together with their 100-word human written summaries. The test set comprised 30 unseen documents. The document sets D061j, D064j, D065j, D068j, D073b, D075b and D077b were used in our experiment.

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) toolkit [11] is used for evaluating the proposed method, where ROUGE compares a system generated summary against a human generated summary to measure the quality. ROUGE is the main metric in the DUC text summarization evaluations. It has different variants, in our experiment, we use ROUGE-N (N=1 and 2) and ROUGE-L, the reason for selecting these measures is what reported by same study [11] that those measures work well for single document summarization.

The ROUGE evaluation measure (version 1.5.5[2]) generates three scores for each summary: recall, precision and F-measure (weighted harmonic mean, eq. 14), in the literature, we found that the recall is the most important measure to be used for comparison purpose, so we will concentrate more on the recall in this evaluation.

$$F=\frac{1}{\left(alpha\times\left(\frac{1}{P}\right)+\left(1-alpha\right)\times\left(\frac{1}{R}\right)\right)} \qquad (14)$$

Where P and R are precision and recall, respectively. Alpha is the parameter to balance between precision and recall; we set this parameter to 0.5.

8. EXPERIMENTAL RESULTS

The similarity threshold plays very important role in our study where the most score of any sentence depends on its relation with other document sentences. Therefore we must pay more attention to this factor by determining its appropriate value through a separate experiment, which was run for this purpose. The data set used in this experiment is document set D01a (one document set in DUC 2001 document sets), the document set D01a containing eleven documents, each document accompanied with its model or human generated summary. We have experimented with 21 different similarity threshold values ranging from 0.01 to 0.2, 0.3 by stepping 0.01. We found that the best average recall score can be gotten using the similarity threshold value 0.16. However, this value doesn't do well with each document separately. Thus, we have examined each similarity threshold value with each document and found that the similarity threshold value that can perform well with all documents is 0.03. Therefore, we decided to run our summarization experiment using the similarity threshold 0.03.

We have run our summarization experiment using DUC 2002 document sets D061j, D064j, D065j, D068j, D073b, D075b and D077b where each document set contains two model or human generated summaries for each document. We gave the names H1 and H2 for those two model summaries. The human summary H2 is used as benchmark to measure the quality of our method summary, while the human summary H1 is used as reference summary. Beside the human with human benchmark (H2 against H1), we also use another benchmark which is MS word summarizer

[2] http://haydn.isi.edu/ROUGE/latest.html

The proposed method and the two benchmarks are used to create a summary for each document in the document set used in this study. Each system created good summary compared with the reference (human) summary. The results using the ROUGE variants (ROUGE-1, ROUGE-2 and ROUGE-L) demonstrate that our method performs better than MS word summarizer and closer to the human with human benchmark. For some document sets (D061j, D073b and D075b), our method could perform better than the human with human benchmark. Although the recall score is the main score used for comparing the text summarization methods when the summary length is limited[3], we found that our method performs well for all average ROUGE variants scores. The overall analysis for the results is concluded and shown in figures 1, 2 and 3 for the three rouge evaluation measures. MMI average recall at the 95%-confidence interval is shown in Table-1.

Metric	95%-Confidence interval
ROUGE-1	0.43017 - 0.49658
ROUGE-2	0.18583 - 0.26001
ROUGE-L	0.39615 - 0.46276

Table 1: MMI average recall at the 95%-confidence interval.

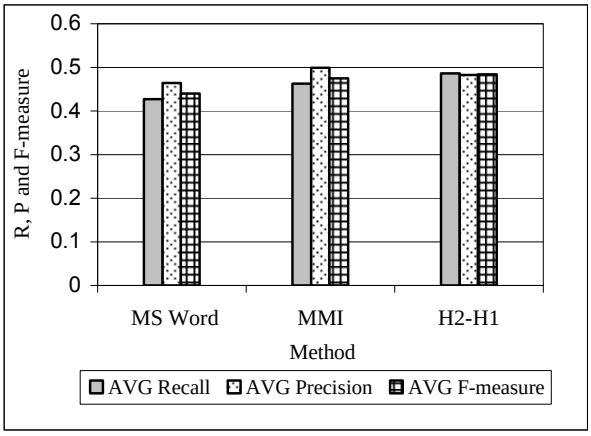

FIGURE 1: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-1.

[3] http://haydn.isi.edu/ROUGE/latest.html

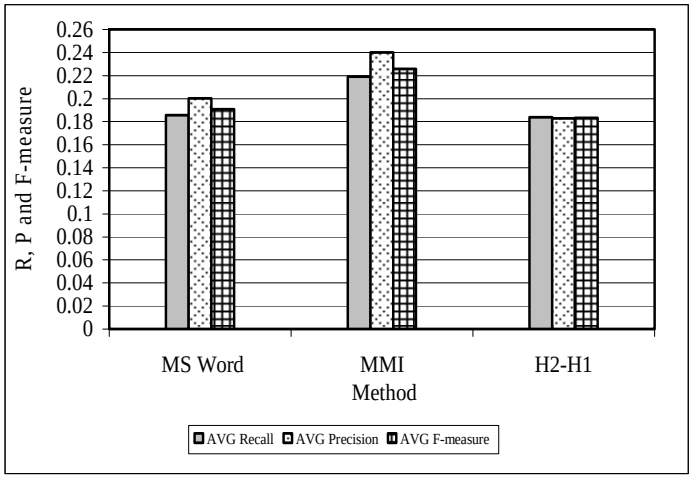

FIGURE 2: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-2.

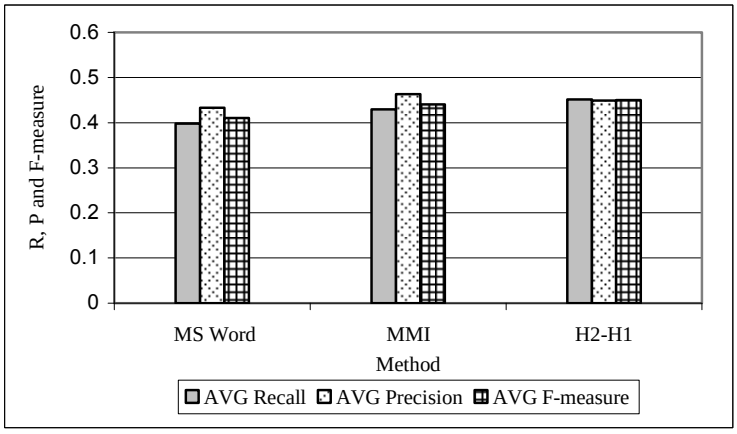

FIGURE 3: MMI, MS Word Summarizer and H2-H1 comparison: Recall, Precision and F-measure using ROUGE-L.

For ROUGE-2 average recall score, our method performance is better than the two benchmarks by: 0.03313 and 0.03519 for MS word summarizer and human with human benchmark (H2-H1) respectively, this proves that the summary created by our method is not just scatter terms extracted from the original document but it has meaningful. For ROUGE-1 and ROUGE-L average recall scores, our method performance is better than MS word summarizer and closer to human with human benchmark.

9. COMPARISON BETWEEN K-MEANS AND C-LINKAGE CLUSTERING ALGORITHMS BASED ON MMI PERFORMANCE

The previous experiment was run using k-means as clustering algorithm for clustering the sentences, we also run the same experiment using c-linkage (complete linkage) clustering algorithm instead of k-means to find out the clustering method which performs well with our method. The results show that c-linkage clustering algorithm performs less than k-means clustering algorithm. Table 3 shows the comparison between those clustering algorithms.

ROUGE	Method	R	P	F-measure
1	MMI-K-means	0.46293	0.49915	0.47521
	MMI-C-linkage	0.44803	0.48961	0.46177
2	MMI-K-means	0.21885	0.23984	0.22557
	MMI-C-linkage	0.20816	0.23349	0.21627
L	MMI-K-means	0.42914	0.46316	0.44056
	MMI-C-linkage	0.4132	0.45203	0.42594

Table 2: comparison between k-means and c-linkage clustering algorithms.

10. CONCLUSION AND FUTURE WORK

In this paper we have presented an effective diversity based method for single document summarization. Two ways were used for finding the diversity: the first one is as preliminary way where the document sentences are clustered based on the similarity - similarity threshold is 0.03 determined empirically - and all resulting clusters are presented as a tree containing a binary tree for each group of similar sentences. The second way is to apply the proposed method on each branch in the tree to select one sentence as summary sentence. The clustering algorithm and binary tree were used as helping factor with the proposed method for finding the most distinct ideas in the text. Two clustering algorithms (K-mean and C-linkage) were compared to find out which of them performs well with the proposed diversity method. We found that K-means algorithm has better performance than C-linkage algorithm. The results of our method supports that employing of multiple factors can help to find the diversity in the text because the isolation of all similar sentences in one group can solve a part of the redundancy problem among the document sentences and the other part of that problem is solved by the diversity based method which tries to select the most diverse sentence from each group of sentences. The advantages of our introduced method are: it does not use external resource except the original document given to be summarized and deep natural language processing is not required. Our method has shown good performance when comparing with the benchmark methods used in this study. For future work, we plan to incorporate artificial intelligence technique with the proposed method and extend the proposed method for multi document summarization.

References

1. S. Brin, and L. Page. *"The anatomy of a large-scale hypertextual Web search engine"*. Computer Networks and ISDN System. 30(1–7): 107–117. 1998.
2. J. Carbonell, and J. Goldstein. "*The use of MMR, diversity-based reranking for reordering documents and producing summaries*". SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 24-28 August. Melbourne, Australia, 335-336. 1998
3. G. Erkan, and D. R. Radev. *"LexRank: Graph-based Lexical Centrality as Salience in Text Summarization"*. Journal of Artificial Intelligence Research (JAIR), 22, 457-479. AI Access Foundation. 2004.
4. K Filippova, M. Mieskes, V. Nastase, S. P. Ponzetto and M. Strube. *"Cascaded Filtering for Topic-Driven Multi-Document Summarization"*. Proceedings of the Document Understanding Conference. 26-27 April. Rochester, N.Y., 30-35. 2007.
5. M. K. Ganapathiraju. *"Relevance of Cluster size in MMR based Summarizer: A Report 11-742: Self-paced lab in Information Retrieval*". November 26, 2002.
6. *"The Document Understanding Conference (DUC)"*. http://duc.nist.gov.

7. A. Jain, M. N. Murty and P. J. Flynn. *"Data Clustering: A Review"*. ACM Computing Surveys. 31 (3), 264-323, 1999.
8. C. Jaruskulchai and C. Kruengkrai. *"Generic Text Summarization Using Local and Global Properties"*. Proceedings of the IEEE/WIC international Conference on Web Intelligence. 13-17 October. Halifax, Canada: IEEE Computer Society, 201-206, 2003.
9. A. Kiani –B and M. R. Akbarzadeh –T. *"Automatic Text Summarization Using: Hybrid Fuzzy GA-GP"*. IEEE International Conference on Fuzzy Systems. 16-21 July. Vancouver, BC, Canada, 977 -983, 2006.
10. W. Kraaij, M. Spitters and M. v. d. Heijden. *"Combining a mixture language model and naive bayes for multi-document summarization"*. Proceedings of Document Understanding Conference. 13-14 September. New Orleans, LA, 109-116, 2001.
11. C. Y. Lin. *"Rouge: A package for automatic evaluation of summaries"*. . Proceedings of the Workshop on Text Summarization Branches Out, 42nd Annual Meeting of the Association for Computational Linguistics. 25–26 July. Barcelona, Spain, 74-81, 2004b.
12. Z. Lin, T. Chua, M. Kan, W. Lee, Q. L. Sun and S. Ye. *"NUS at DUC 2007: Using Evolutionary Models of Text*". Proceedings of Document Understanding Conference. 26-27 April. Rochester, NY, USA, 2007.
13. D. Liu, Y. Wang, C. Liu and Z. Wang. *"Multiple Documents Summarization Based on Genetic Algorithm"*. In Wang L. et al. (Eds.) Fuzzy Systems and Knowledge Discovery. (355–364). Berlin Heidelberg: Springer-Verlag, 2006.
14. H. P. Luhn. *"The Automatic Creation of Literature Abstracts"*. IBM Journal of Research and Development. 2(92), 159-165, 1958.
15. M. J. MAN`A-LO`PEZ, M. D. BUENAGA, and J. M. GO´ MEZ-HIDALGO. *"Multi-document Summarization: An Added Value to Clustering in Interactive Retrieval"*. ACM Transactions on Information Systems. 22(2), 215–241, 2004.
16. T. Mori, M. Nozawa and Y. Asada. *"Multi-Answer-Focused Multi-document Summarization Using a Question-Answering Engine"*. ACM Transactions on Asian Language Information Processing. 4 (3), 305–320 , 2005.
17. J. L. Neto, A. A. Freitas and C. A. A. Kaestner. *"Automatic Text Summarization using a Machine Learning Approach"*. In Bittencourt, G. and Ramalho, G. (Eds.). Proceedings of the 16th Brazilian Symposium on Artificial intelligence: Advances in Artificial intelligence. (pp. 386-396). London: Springer-Verlag ,2002.
18. J. L. Neto, A. D. Santos, C. A. A. Kaestner and A. A Freitas. *"Document Clustering and Text Summarization"*. Proc. of the 4th International Conference on Practical Applications of Knowledge Discovery and Data Mining. April. London, 41-55, 2000.
19. R. Ribeiro and D. M. d. Matos. *"Extractive Summarization of Broadcast News: Comparing Strategies for European Portuguese"*. In V. M. sek, and P. Mautner, (Eds.). Text, Speech and Dialogue. (pp. 115–122). Berlin Heidelberg: Springer-Verlag, 2007.
20. E. Villatoro-Tello, L. Villaseñor-Pineda and M. Montes-y-Gómez. *"Using Word Sequences for Text Summarization"*. In Sojka, P., Kopeček, I., Pala, K. (eds.). Text, Speech and Dialogue. vol. 4188: 293–300. Berlin Heidelberg: Springer-Verlag, 2006.
21. S. Ye, L. Qiu, T. Chua and M. Kan. *"NUS at DUC 2005: Understanding documents via concept links"*. Proceedings of Document Understanding Conference. 9-10 October. Vancouver, Canada, 2005.
22. D. M. Zajic. *"Multiple Alternative Sentence Compressions As A Tool For Automatic Summarization Tasks"*. PhD theses. University of Maryland, 2007.
23. D. M. Zajic, B. J. Dorr, R. Schwartz, and J. Lin. *"Sentence Compression as a Component of a Multi-Document Summarization System*". Proceedings of the 2006 Document Understanding Workshop. 8-9 June. New York, 2006.
24. H. Zha. *"Generic summarization and key phrase extraction using mutual reinforcement principle and sentence clustering*". In proceedings of 25th ACM SIGIR. 11-15 August. Tampere, Finland, 113-120, 2002.
25. X. Zhu, A. B. Goldberg, J. V. Gael and D. Andrzejewski. *"Improving diversity in ranking using absorbing random walks"*. HLT/NAACL. 22-27 April. Rochester, NY, 2007.