DOI:10.24996/ijs.2018.59.4B.19

ISSN: 0067-2904

Improving Extractive Multi-Document Text Summarization Through Multi-Objective Optimization

Nasreen J. Kadhim[1*], Hilal H. Saleh[2], Bara'a Attea[1]

[1]Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq

[2]Computer Science Department, University of Technology, Baghdad, Iraq

Abstract

Multi-document summarization is an optimization problem demanding optimization of more than one objective function simultaneously. The proposed work regards balancing of the two significant objectives: content coverage and diversity when generating summaries from a collection of text documents.

Any automatic text summarization system has the challenge of producing high quality summary. Despite the existing efforts on designing and evaluating the performance of many text summarization techniques, their formulations lack the introduction of any model that can give an *explicit representation of – coverage and diversity – the two contradictory semantics of any summary*. In this work, the design of generic text summarization model based on sentence extraction is redirected into more semantic measure reflecting individually both content coverage and content diversity as two explicit optimization models. The problem is defined by projecting the first criterion, i.e. content coverage in the light of text similarity. The proposed model hypothesizes a possible decomposition of text similarity into three different levels of optimization formula. First, aspire to *global* optimization, the candidate summary should cover the summary of the document collection. Then, to attain, less *global* optimization, the sentences of the candidate summary should cover the summary of the document collection. The third level of optimization is content with *local* optimization, where the difference between the magnitude of terms covered by the candidate summary and those of the document collection should be small. This coverage model is coupled with a proposed diversity model and defined as a *Multi-Objective Optimization* (*MOO*) problem. Moreover, heuristic perturbation and heuristic local repair operators have been proposed and injected into the adopted evolutionary algorithm to harness its strength. Assessment of the proposed model has been performed using document sets supplied by Document Understanding Conference 2002 ($DUC2002$) and a comparison has been made with other state-of-the-art methods. Metric used to measure performance of the proposed work is Recall-Oriented Understudy for Gisting Evaluation ($ROUGE$) toolkit. Results obtained support strong proof for the effectiveness and the significant performance awarded to the proposed MOO model over other state-of-the-art models.

Keywords: Multi-objective optimization; multi-objective multi-document text summarization problem; MOO; multi-objective evolutionary algorithm; non-dominated solution, EP.

*Email: nasreen_jawad@yahoo.com

تحسين التلخيص الاقتطاعي للمستندات النصية المتعددة من خلال أمثلية تعدد الأهداف

نسرين جواد كاظم[1*]، هلال هادي صالح[2]، براء علي عطيه[1]
[1]قسم علوم الحاسبات، كلية العلوم، جامعة بغداد، بغداد، العراق
[2]قسم علوم الحاسبات، الجامعه التكنولوجيه، بغداد، العراق

الخلاصة
تقنية التلخيص الأوتوماتيكي في تطور وربما تقدم حل الى مشكلة الحمل الزائد للمعلومات. عملية التلخيص للنصوص متعددة المستندات تصنف على انها مشكلة أمثلية تتطلب الاستفادة المثلى من اكثر من دالة هدف في وقت واحد.

العمل المقترح يأخذ بنظر الأعتبار تحقيق التوازن بين هدفين مهمين هما: تغطية المحتوى لمجموعة المستندات والتنوع عند توليد ملخص من مجموعة من المستندات النصية. أي نظام أوتوماتيكي لتلخيص النصوص يملك التحدي المتمثل في إنتاج ملخص عالي الجودة. على الرغم من الجهود القائمة على تصميم و تقييم أداء العديد من تقنيات تلخيص النصوص، تفتقر صياغات هذه التقنيات الى تقديم أي نموذج يمكن أن يعطي التمثيل الصريح – تغطية المحتوى والتنوع – وهما دلالتان متناقضتان في أي ملخص. أن تصميم نموذج يهدف الى تلخيص نص عام قائم على أقتطاع الجمل تمت أعادة توجيهه الى تدبير ذات دلالة اكبر يعكس بصورة مستقلة كلا من تغطية وتنوع المحتوى كنموذجي أمثلية صريحة. . يتم تعريف المشكلة من خلال عرض المعيار الأول ، أي تغطية المحتوى في ضوء التشابه النصي. يفترض النموذج المقترح تحليلًا محتملًا لتشابه النص إلى ثلاثة مستويات مختلفة من صيغة التحسين. أولاً ، التطلع إلى التحسين العام ، يجب أن يغطي الملخص المرشح ملخص مجموعة المستندات. بعد ذلك ، لتحقيق الحد الأمثل من التحسين العام، يجب أن تغطي جمل الملخص المرشح ملخص مجموعة المستندات. المستوى الثالث من التحسين هو المحتوى ذو التحسين المحلي ، حيث يجب أن يكون الفرق بين قيم المصطلحات المغطاة من قبل الملخص المرشح وتلك الخاصة بملخص مجموعة المستندات صغيرًا. بعد ذلك تتم عملية اقتران النموذج المقترح مع نموذج التنوع المقترح وتعريفهما كمشكلة أمثلية تعدد الاهداف. وعلاوة على ذلك ، تم أقتراح عامل توجيه اضطراب وعامل توجيه أصلاح محلي وحقنهما في الخوارزمية التطورية المعتمدة لتسخير قوتها .

عملية تقييم النماذج المقترحة تمت باستخدام مجموعة المستندات المجهزة من قبل مجموعة البيانات العالمية (Document Understanding Conference DUC2002) وقد تمت مقارنة النتائج المتحصلة مع مجموعة من الانظمة الحديثة. قياس وتقييم الأداء للنماذج المقترحة تم باستخدام أدوات (ROUGE). النتائج المتحصلة دعمت العمل بدليل قوي على فعالية النموذج المقترح على النماذج الحديثة التي تمت المقارنة بها.

1. Introduction

Identification of relevant information that meets user needs becomes very difficult as a result of exponential growth of Internet and the continuous size increase of affordable media storage devices which have allowed for the availability of huge amount of online information. With the richness of data, a massive demand for innovative technologies that introduce an effective processing of documents to reach the user's specific needs is required.

The two fields which have focused on the effective processing of large amounts of data and extraction of information with high quality and relevance to the user's needs are Data Mining (DM) and Information Retrieval (IR). With Data Mining, patterns and trends are typically detected within text to help identify and form interesting and important information. Information Retrieval field searches documents, data within documents and their metadata to help find information that is significant and relevant. An overlap between the two fields is obvious, especially when looking at the sub areas which are covered by both. *Automatic Document Summarization (ADS)* which tend to be a vital technology to overwhelmed this obstacle in technological environments is among these areas which targets the production of summaries meeting the user's needs. The documents to be summarized can be of multimedia type, text, or both.

Systems that emphasis on Text Document Summarization (TDS) often comprise subtasks borrowed from the field of Natural Language Processing such as Text Parsing, Natural Language Understanding, Coreference Resolution, and Anaphor Resolution. Thus, TDS can be regarded as a subfield of Natural Language Processing, which also overlaps with DM and IR [1].

Automatic text summarization technology is maturing and may offer together with the conventional information search engines a solution to the problem of information overload to satisfy accessing the relevance of retrieved documents efficiently [2, 3]. This interprets the growing importance of the area of automatic text summarization which has triggered the race for developing many algorithmic models.

Text summarization problem attracts several disciplines from computer science to formulate and develop powerful techniques. The main goal of these techniques is to introduce the most important information of the original detailed text in a condensed version whilst discarding irrelevant and redundant information. By this, the user can quickly understand the large volume of required information that targets his intent.

Text summarization techniques can be classified according to the task of summarization as generic or query-relevant summary [4–6]. A whole sense of document content is presented without any prior knowledge in a generic summary. On the other hand, the information presented in a query-relevant summary should have some relevance with a given query or topic [7].

Text summarization approaches can, also, be either extractive or abstractive according to the function to be performed. Extractive text summarization systems tend to select a subgroup of words, phrases, or sentences that exist in the original text and are highly significant for generating summary. These approaches are typically based on some rules for extracting sentences, and effort to recognize the combination of most important sentences matching the overall understanding of a particular document.

Sentence extraction methods are normally performed using some kind of similarity or centrality metric [8–11]. In contrast, an internal semantic representation is built by abstractive methods and then a summary that is closer to a human made summary is created via some natural language generation techniques. Novel words that do not explicitly exist in the original text might be involved in such a summary [12].

Moreover, considering number of simultaneous analyzed documents, summary creation may be performed either from a single or multiple documents [5, 13]. Thus, a condensed representation of one document can only be produced via single-document summarization, whereas a summary from multiple documents can be produced thru multi-document summarization.

Depending on the usage (i.e. type of information that the summary involves), summary can be critical summary, indicative summary, informative summary or extract summary. Indicative and informative summaries are the most important summary types. Informative summaries offer a shortening for whole document, retrieving its significant details, whereas decreasing volume of information. The typical length of this type of summary ranges from 20–30% of the original text [2, 5, 14]. An indicative summary is a shortened version of main topics of a document where presentation of the content details is avoided to attract the user into getting the complete document. This type of summary often used as the end part of the information retrieval systems, being retrieved by search system as a substitute of complete document. These summaries have typical lengths that vary from 5% to 10% of the whole text [14].

2. Related work

Extractive document summarization obviously involves selecting the most relevant information and generating a coherent summary from them. The generated summary comprises multiple disjointedly extracted sentences from document(s). Clearly, each of the chosen sentences should separately be important. By including many of the competing sentences in the summary, the problem of information overlap between portions of the generated summary comes up, and this demands a mechanism for addressing redundancy. Consequently, when many of the competing sentences are presented, assumed summary length limit, the scheme of choosing best summary instead of choosing best sentences becomes obviously important. The problem of choosing the best summary is a global optimization problem compared with the process of picking the best sentences. Furthermore, the quality of summary is defined by two main criteria which are coverage and diversity.

In extractive document summarization, generation of the optimal summary can be regarded as a combinatorial optimization problem in which finding a solution to it is NP-hard [15]. Maximal Marginal Relevance (MMR) [16] is one of the standard methods for text summarization problem, where the most relevant sentences are selected by a greedy algorithm, and simultaneously the redundancy is avoided by removing too similar sentences to the already selected sentences. One key problem of MMR is that the decision using it is made based on the scores at the present iteration which make it non-optimal. The following are a review of optimization based works which are most related to the approach proposed in this paper.

In [17], document summarization was formulated as a multi-objective optimization problem. In particular, four objective functions, namely information coverage, significance, redundancy and text coherence were involved. These four objective functions measure the generated summaries according to the cluster of semantically or statistically related core terms. To solve the optimization problem, this work converted the multi-objective optimization problem into single objective optimization problem to produce as a final result one optimal solution.

Multi-document summarization that considers two concerns represented by minimizing redundant information while extracting sentences that are representative was modeled as a discrete optimization problem in [18]. The optimization problem was solved through creation of an adaptive differential evolution algorithm. Testing the performance of the method proposed in this work was performed on the standard DUC2002 and DUC2004 datasets and a comparison was made against baseline systems. The experimental results provided an evidence that the proposed method is a viable method for document summarization.

The work presented in [19] proposed modeling for generic extractive text summarization as an integer linear programming problem that considers covering the main content of document(s) and guarantees including sentences that convey diverse ideas in the generated extractive summary. Generic text summarization model was represented as an optimization problem and the problem was solved globally. The model was compared against several existing methods using DUC2005 and DUC2007 datasets and its performance was evaluated using ROUGE-2 and ROUGE-SU4 metrics. The model proposed in this work demonstrated that the summarization result depends on the similarity measure. Results showed significant improvements to the summarization results and showed that better results could be obtained from a combination of the Normalized Google Distance based NGD-based and cosine similarity measures than their use separately.

Extraction-based generic text summarization was modeled in [20] as linear and nonlinear optimization problems. A simultaneous balancing of summary objectives represented by coverage and diversity was attempted to be performed at these models. The optimization problem was solved through developing an adaptive particle swarm optimization algorithm. Experimental results showed that the proposed models outperformed the best-reported results on DUC2005 dataset, and also compared well on DUC2006 data set using ROUGE-1, ROUGE-2, and ROUGE-SU4 metrics.

The work introduced in [21] modeled text summarization as a quadratic integer-programming problem while attempting to optimize relevance, redundancy and length. A novel differential evolution algorithm was created to solve the optimization problem. The methods were implemented on DUC2005 and DUC2007 data sets and evaluated using ROUGE-1, ROUGE-2 and ROUGE-SU4 metrics. Significant improvement were noticed on the summarization results through applying the method proposed in this work. Experimental Results showed that combining symmetric and asymmetric similarity measures conducted to better result than their use separately.

The work introduced in [22] Modeled text summarization as a Boolean programming problem. The model attempted to optimize three concerns represented by: relevance, redundancy and length. The optimization problem was solved through creating differential evolution algorithm with self-adaptive crossover and mutation strategies. The model was implemented on multi-document summarization task. The proposed model when compared to several existing summarization methods on DUC2005 and DUC2007 datasets, it was found that a significant improvement occurred for the summarization results. Method evaluation was performed using ROUGE-1, ROUGE-2 and ROUGE-SU4 metrics. This Work demonstrated that results of summarization depend on the similarity measure and clarified that symmetric and asymmetric similarity measures as a combination produce better result than their use individually.

An extractive multi-document text summarization model based on genetic algorithm (GA) was proposed in our previous work [8]. First, the problem was modeled as a discrete optimization problem and a specific fitness function was designed to effectively cope with the proposed model. Then, a binary-encoded representation together with a heuristic mutation and a local repair operators were proposed to characterize the adopted GA. Experiments were applied to DUC2002 datasets. Results clarified the effectiveness of the proposed model when compared with another state-of-the-art model .

Simultaneous optimization of many objectives is involved in many real world problems in engineering, industry, and in many other fields. A MOO problem has, in its nature, several objectives that contradict each other (i.e., improvement of one objective cannot be satisfied without deterioration of at least any other objective) and need to be optimized simultaneously in order to solve the problem. Attraction of several researchers recently by MOO field in order to model and solve MOO problems belongs to its large success. In single objective optimization, the goodness of one solution over the other is possible to be determined which results in obtaining a single optimal solution whereas in multi-objective optimization, a straightforward method to determine optimality of a particular solution does not exist. In our previous work introduced in [9], the design of generic text summarization model based on sentence extraction was modeled as an optimization problem redirected into more semantic measure reflecting individually both content coverage and content diversity as an explicit individual optimization models. The two proposed models were then coupled and defined as a *multi-objective optimization* (*MOO*) problem. Up to the best of our knowledge, this was the first attempt to address text summarization problem as a MOO model. Moreover, heuristic perturbation and heuristic local repair operators were proposed and injected into the adopted evolutionary algorithm to harness its strength. Assessment of the proposed model was performed using document sets supplied by DUC 2002 and a comparison was made with other state-of-the-art methods using $ROUGE$ toolkit. Results obtained supported strong proof for the effectiveness of the proposed model based on MOO over other state-of-the-art models.

In the work proposed in this paper, the problem is defined by projecting the first criterion, i.e. content coverage in the light of text similarity. The proposed model hypothesizes a possible decomposition of text similarity into three different levels of optimization formula. First, aspire to *global* optimization, the candidate summary should cover the summary of the document collection. Then, to attain, less *global* optimization, the sentences of the candidate summary should cover the summary of the document collection. The third level of optimization is content with *local* optimization, where the difference between the magnitude of terms covered by the candidate summary and those of the document collection should be small. This coverage model is coupled with a proposed diversity model and defined as a *multi-objective optimization* (*MOO*) problem. The proposed model attempt to rigorously cast on the contradictory nature of text summary by quantitatively controls selection of document's sentences. The selection will emphasize centrality (selection of the sentences having a wider coverage of the document set) and diversity (inclusion of diverse ideas in the final summary). The diverse ideas having a wider coverage of the document set can guarantee, in a reasonable degree, that the generated summary covers the most significant portions of the original document. Multi-objective evolutionary algorithm is adopted in this paper to tackle the text summarization problem. Moreover, heuristic perturbation and heuristic local repair operators are proposed and injected into the adopted evolutionary algorithm to harness its strength.

Organization of this paper is as follows: Section 3 introduces preliminaries of the text summarization problem. The problem of extractive multi-document text summarization is stated in section 4 together with the presentation of the details of the proposed mathematical formulation and modeling. Multi-objective evolutionary algorithms are presented in section 5 in terms of their basic concepts in addition to the introduction of one of the most common multi-objective evolutionary algorithms, Multi-Objective Evolutionary algorithm with Decomposition (MOEA/D). Section 6 presents the proposed multi-objective evolutionary algorithm for multi document text summarization problem. Simulation results and their related discussions are presented in Section 7. Finally, conclusions and some possible extensions to the current work are given in Section 8.

3. Preliminaries

In text summarization, vector-based methods are commonly used [23]. Let $T = \{t_1, t_2, t_3, \dots, t_m\}$ represents m distinct terms in a document collection. *Cosine similarity* is the most popular measure that evaluates text similarity between any pair of sentences being represented as vectors of terms. For a set of m different terms composing n sentences of a document collection $\mathbb{D}$, cosine similarity associates weight w_{ik} to term t_k according to its magnitude in sentence s_i. Cosine similarity measure can then be formulated according to *term-frequency inverse-sentence-frequency* scheme (tf_isf) [23]:

$$w_{ik} = tf_{ik} \times isf, \tag{1}$$

where:

tf_{ik}: is the measure of how *frequently* a term t_k occurs in a sentence s_i, and

$isf = \log(n/n_k)$ is the measure of how *few* sentences n_k contain the term t_k.

Intuitively, if a term t_k does not exist in sentence s_i, w_{ik} should be zero. Now, given two sentences $s_i = [w_{i1}, w_{i2}, \dots, w_{im}]$ and $s_j = [w_{j1}, w_{j2}, \dots, w_{jm}]$, the cosine similarity between these two sentences can be calculated as in $Eq.(2)$ [23]:

$$sim(s_i, s_j) = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_{jk}^2}}, \qquad i, j = 1,2,3, \dots, n \tag{2}$$

Quantitatively, the *main content* of a document collection $\mathbb{D}$ being represented in $T = \{t_1, t_2, t_3, \dots, t_m\}$ space, can be reflected by the *mean* weights of the m terms in T. Thus, for $T = \{t_1, t_2, t_3, \dots, t_m\}$ vector, a mean vector $\mathbb{O} = [o_1, o_1, \dots, o_m]$ can be computed. The k^{th} coordinate o_k of the mean vector $\mathbb{O}$ can be calculated as:

$$o_k = \frac{1}{n} \sum_{i=1}^{n} w_{ik}, \qquad k = 1,2,3, \dots, m \tag{3}$$

4. Problem statement and formulations

The proposed text summarization problem is expressed here while considering three challenges:

- *Content Coverage*: the main topic of the document collection $\mathbb{D}$ *should be covered* by the generated summary.
- *Redundancy Reduction*: similar sentences in the document collection $\mathbb{D}$ *should not be duplicated* in the generated summary.
- *Length*: summary *should be of a bounded length.*

Let $\mathbb{D}$ be a document collection of N documents, i.e. $\mathbb{D} = \{d_1, \dots, d_N\}$. By the language of sentences, $\mathbb{D}$ can be noted by $\mathbb{D} = \{s_i | 1 \le i \le n\}$, where n is the number of distinct sentences from the documents in $\mathbb{D}$. The aim of this paper is to generate a summary $\overline{\mathbb{D}} \subset \mathbb{D}$ that can satisfy the above three criteria.

Multi-document summarization in its nature involves simultaneous optimization of more than one objective function that contradict each other. To this end, Multi-document summarization based on MOO model is proposed. A simultaneous optimization of two objectives: *content coverage* and *redundancy reduction* is suggested. An MOO model is introduced, the model has two objective functions: the first objective $\Phi_{coverage}(x)$ concerns coverage criteria, while the second objective concerns information redundancy criteria $\Phi_{redund\_red}(x)$. Following are definitions of the MOO based model proposed for modeling multi-document text summarization problem.

The problem of multi-document text summarization is defined by projecting content coverage in the light of text similarity. The proposed model hypothesizes a possible decomposition of text similarity into three different levels of optimization formula. First, aspire to *global* optimization, the candidate summary should cover the summary of the document collection. Then, to attain, more or *less global* optimization, the sentences of the candidate summary should cover the summary of the document collection. The third level of optimization is content with *local* optimization, where the difference between the magnitude of terms covered by the candidate summary and those of the document collection should be small. The objective functions $\Phi_{coverage}$ and Φ_{redund_red} and text summarization problem $MOEA_{\Phi}$ can then be formulated as in definitions 1, 2 and 3, respectively.

Definition 1 (*Content coverage objective function $\Phi_{coverage}$*). Let $s_i \in \mathbb{D}$ be a sentence to be included in the generated summary $\overline{\mathbb{D}}$, then three different semantics of coverage (*summary level, sentence level, and term level*) can be cooperated together to define *content coverage* criterion. *Summary level* to be expressed by the degree of summary similarity, $sim(O, \mathbb{O})$, between the mean vector O of the candidate summary $\overline{\mathbb{D}}$ and the center $\mathbb{O}$ of the document collection $\mathbb{D}$. *Sentence level* to be defined by the degree of sentence similarity, $sim(s_i, \mathbb{O})$, between sentence s_i and the mean vector $\mathbb{O}$ of the document collection $\mathbb{D}$. *Term level* to be defined by the degree of $term\_closeness(O_k, \mathbb{O}_k)$ between the impact of term k in the mean vector O of the candidate summary $\overline{\mathbb{D}}$ and its corresponding term in the center $\mathbb{O}$ of the document collection $\mathbb{D}$. Content coverage is expressed by maximizing both $sim(O, \mathbb{O})$ and $sim(s_i, \mathbb{O})$ while simultaneously minimizing sum of $Diff(O_k, \mathbb{O}_k)$ for every k term in sentences included in the automated summary.

$$\Phi_{coverage} = sim(O, \mathbb{O}) + \sum_{i=1}^{n} sim(s_i, \mathbb{O}) x_i - \sum_{k=1}^{m} |O_k - \mathbb{O}_k| \tag{4}$$

As can be seen in Eq. 4, the magnitude O_k of term k in the candidate summary $\overline{\mathbb{D}}$ can be expressed by its impact, i.e. average of total weights of k occurring in the sentences of $\overline{\mathbb{D}}$. Likewise, the magnitude $\mathbb{O}_k$ of term k in $\mathbb{D}$ can be computed by the average of total weights of k occurring in the sentences of $\mathbb{D}$. Intuitively, the difference between these two magnitudes should be small over all terms of $\mathbb{D}$ and $\overline{\mathbb{D}}$. In order to unify the impact of the three terms in Eq. 4, their values are normalized to be in the range [0, 1].

Definition 2 (*Redundancy reduction objective function Φ_{redund_red}*). On the other hand, the *redundancy reduction* should be maximized, or quantitatively, the similarity, $sim(s_i, s_j)$, between any two sentences belong to $\overline{\mathbb{D}}$ should be *minimized*.

$$\Phi_{redund\_red} = \frac{1}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} sim(s_i,s_j)x_{ij}*\sum_{i=1}^{n} x_i} \tag{5}$$

Definition 3 (*multi-objective multi-document text summarization problem $MOEA_{\Phi}$*). Let $x_i \in \{0,1\}$ be a binary decision variable denoting the existence (1) or absence (0) of the sentence s_i in $\overline{\mathbb{D}}$ (see Eq. 6). Also, let $x_{ij} \in \{0,1\}$ be another binary decision variable relating to the existence of both sentences s_i and s_j in $\overline{\mathbb{D}}$ (see Eq. 7). Now, let $X = \{x_i | 1 \le i \le n\}$ be a vector of n such decision variables corresponding to n sentences. Then for the vector X, text summarization problem (see Eq. 8 & Eq. 9) can be expressed as a constrained optimization problem taking a combination of maximizing the two objective functions representing content coverage and information redundancy reduction $\Phi_{coverage}$ and Φ_{redund_red} respectively.

$$x_i = \begin{cases} 1 \; if \; s_i \in \overline{\mathbb{D}} \\ 0 \; otherwise \end{cases}, \tag{6}$$

$$x_{ij} = \begin{cases} 1 \; if \; s_i and \; s_j \in \overline{\mathbb{D}} \\ 0 \; otherwise \end{cases} \tag{7}$$

$$MOEA_{\Phi}(x) = Maximize \; \{\Phi_{coverage}(x), \Phi_{redund\_red}(x)\} \tag{8}$$

$$subject \; to \quad L - \varepsilon \le \sum_{i=1}^{n} l_i x_i \le L + \varepsilon, \tag{9}$$

where:

L: Summary length constraint,

l_i: Length of sentence s_i,

$\mathbb{O}$: Center of the document collection $\mathbb{D} = \{s_1, s_2, \dots, s_n\}$.

ε: A length tolerance introduced in this model as:

$$\varepsilon = max_{i=1,\dots,n}\,(l_i) - min_{i=1,\dots,n}\,(l_i) \tag{10}$$

5. Multi-Objective Evolutionary Algorithms

Evolutionary algorithms (EAs) in their nature are population-based meta-heuristics that have the ability to find simultaneously multiple optima. Formulation of multi-objective optimization problem consisting of m objective functions can be stated as follows:

$$minimize \; or \; maximize \; F(\mathbb{x}) = \left(f_1(\mathbb{x}), \dots, f_m(\mathbb{x})\right)^T \tag{11}$$
$$subject \; to \; \mathbb{x} \in \Omega$$

Formally speaking, general multi-objective optimization problem aims to find the vector $\mathbb{x}^* = [x_1^*, x_2^*, \dots, x_m^*]^T$ which is the result of optimizing the objective function vector in $Eq. 11$, wherein $\mathbb{x}$ denotes the decision variable vector, $F: \Omega \longrightarrow \mathbb{R}^m$ is composed of m real-valued objective functions, $\mathbb{R}^m$ denotes the objective space and the search space is denoted by Ω. In general, the functions $f_1, \dots, f_m$ contradict each other, so a balance between them has to be done and the optimum can be explored by finding a good trade-off between all the functions $f_1, \dots, f_m$ because there is no point in Ω that optimizes all the objective functions of F simultaneously.

A hard work has been dedicated in the last few years in order to apply evolutionary algorithms to the improvement of multi-objective optimization algorithms (see for instance: [24‒27]). Decomposition Based Multi-objective Evolutionary Algorithm (MOEA/D) [27] offered by Zhang and Li is one of the dominant algorithms for multi-objective optimization problems.

In MOEA/D, the MOP is decomposed explicitly into N scalar optimization subproblems that are optimized simultaneously by evolving a population of N solutions. Population at every generation consists of the best solution established thus far for each scalar optimization subproblem. Definition of neighborhood relations among sub-problems takes in consideration the distances calculated between their associated aggregated coefficient vectors. Two neighboring sub-problems should have very similar optimal solutions. Optimization of each subproblem in MOEA/D takes in consideration the information from its neighboring subproblems.

Several methods exist for the construction of aggregation functions. Weighted sum approach and the Tchebycheff approach are the most popular ones among them. Tchebycheff approach will be presented and adopted in this paper. The general framework of MOEA/D can be presented in [27]. Let $\lambda_1, \dots, \lambda_N$ be a set of N even spread weight vectors and $z^* = (z_1^*, \dots, z_m^*)$ be a reference point to the m objective functions $f_1, \dots, f_m$. The approximation problem of the PF of the MOO can be decomposed into scalar optimization sub-problems using the Tchebycheff approach and the objective function of the i^{th} subproblem is:

$$g^{te}(\mathbb{x}|\lambda_i, z^*) = \max_{1 \le j \le m}\{\lambda_{i,j}|f_j(\mathbb{x}) - z_j^*\} \tag{12}$$

where $\lambda_i = (\lambda_{i,1}, \dots, \lambda_{i,m})^T$ is the weight vector, i.e., $\forall j = 1, \dots, m$: $\lambda_{i,j} \ge 0$ and $\sum_{j=1}^{m} \lambda_{i,j} = 1$. All these N objective functions are optimized simultaneously by MOEA/D in a single run. MOEA/D maintains at each generation gen_no with the Tchebycheff approach:

- A population of N points $\mathbb{P}_1, \dots, \mathbb{P}_N \in \Omega$, where the ith subproblem has the current solution $\mathbb{P}_i$,
- $FV_1, \dots, FV_N$ where $FV_i = F(\mathbb{P}_i) = \left(f_1(\mathbb{P}_i), \dots, f_m(\mathbb{P}_i)\right)^T$ $\forall i = 1, \dots, N$,
- $z^* = (z_1^*, \dots, z_m^*)^T$ where z_j^* be the best value occurred so far for objective f_j, and
- An external population (EP) to store non-dominated solutions.

6. Proposed Multi-Objective Evolutionary Algorithm For Multi-Document Text Summarization

The popular multi-objective evolutionary algorithm of Zhang and Li called multi-objective evolutionary algorithm with Tchebycheff decomposition [27] is projected in the light of multi-document summarization problem. A formulation to the representative components of the algorithm is performed to be suitable for the given problem.

MOEA/D is adopted in the proposed work in order to solve the optimization problem of multi document summarization. Considering $N = 50$ which denotes the number of sub-problems, and $m = 2$ which represents the number of contradictory objective functions. Let $\lambda_1, \dots, \lambda_N$ be a set of even spread weight vectors associated with each sub-problem and $z^* = (z_1^*, z_2^*)$ be a reference point to the two objective functions. The problem of approximation of the Pareto Front of the multi objective optimization can be decomposed into scalar optimization sub-problems using the Tchebycheff approach. MOEA/D makes simultaneous optimization of all these N objective functions in a single run. At each generation gen_no, MOEA/D with the aid of Tchebycheff approach preserves: a population of N points $\mathbb{P}_1, \dots, \mathbb{P}_N \in \Omega$, where $\mathbb{P}_i$ is the current solution to the i^{th} subproblem, $MOEA_\Phi 1, \dots, MOEA_\Phi N$ where $MOEA_\Phi i = MOEA_\Phi(\mathbb{P}_i) = \left(MOEA_\Phi^1(\mathbb{P}_i), MOEA_\Phi^2(\mathbb{P}_i)\right)^T$ $\forall i = 1, \dots, N$; where $MOEA_\Phi^1(\mathbb{P}_i) = \Phi_{coverage}$ and $MOEA_\Phi^2(\mathbb{P}_i) = \Phi_{redund\_red}$; and $z^* = (z_1^*, z_2^*)^T$ where z_j^* be the best value found thus far for objective $MOEA_\Phi^j$. Moreover, an external population (EP) is preserved by MOEA/D, which is used as an archive scheme for accumulation of non-dominated solutions discovered throughout the search.

Representation of each individual $\mathbb{P}_{1 \le i \le N} \in \mathbb{IP}$ is considered as a vector with fixed-length having $size = n$, where each gene determine the existence or absence of the equivalent sentence. $\Phi: \mathbb{P}_i \to \mathbb{R}^2$ indicates the objective function vector allotting content coverage, $\Phi_{coverage}$, and redundancy reduction, Φ_{redund_red}, to individual $\mathbb{P}_i \in \mathbb{IP}$. Set of genetic operators is denoted by γ each of them is controlled by a particular parameter:

$$\gamma = \left\{s_{\Theta_s}, c_{\Theta_c}, m_{\Theta_m} \middle| s_{\Theta_s}, c_{\Theta_c}, m_{\Theta_m} : \mathbb{P} \to \mathbb{IP}\right\} \tag{13}$$

For selection operator, two parents $\mathbb{P}_k$ *and* $\mathbb{P}_l$ are selected randomly from the neighbors of the determined individual in the population. Then uniform crossover is applied to these parents according to the probability p_c. A heuristic mutation operator is applied to each allele in the new individuals and it is controlled by two parameters. The first parameter is the well-known mutation probability, p_m, controlling the probability of mutation on each gene. The second parameter is *mutation action*, which controls the role of mutation on each *mutated* gene (See Eq. 14). Mutation action can be projected by the following similarity condition:

$$sim(s_i, \mathbb{O}) \ge \frac{1}{n}\sum_{j=1}^{n} sim(s_j, \mathbb{O}) \tag{14}$$

For a given gene i and for a random uniform variable $r_i \sim [0,1]$, if the sentence corresponds to the i^{th} gene exists, and if p_p is satisfied (i.e., $r_i \le p_p$) then the similarity condition should be checked. The condition checks whether the similarity between the i^{th} sentence and mean vector $\mathbb{O}$ is more or less than the average similarity of sentences in the document collection $\mathbb{D}$. If it is satisfied, then the corresponding sentence, s_i can be selected in the generated summary $\overline{\mathbb{D}}$. Otherwise, it can be removed from the summary. Formally speaking,

$$\forall i \in \{1,\dots,n\} \wedge x_i = 1 \wedge r_i \leq p_p \tag{15}$$

$$x_i' = \begin{cases} 1 \; iff \; sim(s_i, \mathbb{O}) \geq \frac{1}{n}\sum_{j=1}^{n} sim(s_j, \mathbb{O}) \\ 0 \; otherwise \end{cases} \tag{16}$$

Then an update function $\Psi: \mathbb{EP} \rightarrow \mathbb{EP}'$ is applied to update the current $\mathbb{EP}$ by exclusion of dominated solutions and/or inclusion of the new child solutions while implementing γ to the current $\mathbb{P}$. Then one non-dominated solution $\mathbb{P}^*$ is selected from the archive $\mathbb{EP}$ by a decision maker function $\varphi: \mathbb{EP} \rightarrow \mathbb{P}^*$.

The best solution, $\mathbb{P}^*$, of the final generation of the algorithm can be selected as the result to the maximization problem.

$$\mathbb{P}^*: \Leftrightarrow \nexists \mathbb{P} \in \mathbb{P}_{gen_{max}} | \; \Phi(X_{\mathbb{P}}) > \Phi(X_{\mathbb{P}_{best}}) \tag{17}$$

However, the phenotype of the best solution $\mathbb{P}^*$ may still suffer from violating the length constraint i.e.

$$\sum_{i=1}^{n} l_i x_i > L \tag{18}$$

To this end, a *local repair* operator is proposed to handle the existence of more than constraint needs. Firstly, this repair operator removes from $\mathbb{P}^*$ those redundant sentences which have a high degree of similarity between them. Considering a *similarity threshold* $\delta = 0.9$ and two sentences x_i and x_j in $\mathbb{P}^*$, one of them will be excluded from the final generated summary if their similarity is more than or equal to δ (see Eq. 19). Secondly, this operator will only handle the selection of high importance sentences in $\mathbb{P}^*$. Each sentence belongs to $\mathbb{P}^*$ is ranked according to the formula in Eq. 20 to gain a corresponding score:

$$\forall i, j \in \{1,\dots,n\} \wedge x_i, x_j \in \mathbb{P}^* = 1$$

$$sim(s_i, s_j) \geq \delta \tag{19}$$

$$\forall i \in \{1,\dots,n\} \wedge x_i \in \mathbb{P}^* = 1$$

$$score_{s_i} = sim(s_i, \mathbb{O}) + \left(\left(sim(O^{sum}, \mathbb{O}) - sim(O^{sum-s_i}, \mathbb{O})\right)\right) \tag{20}$$

Where $sim(O^{sum}, \mathbb{O})$ refers to the similarity of the centre of the generated summary (including sentence s_i) and the centre of document collection $\mathbb{O}$. On the other hand, $sim(O^{sum-s_i}, \mathbb{O})$ denotes the similarity between the generated summary (excluding sentence s_i) and the centre of document collection $\mathbb{O}$. In order to unify the impact of the two terms, their values are normalized to be in the range [0,1]. The basic idea behind the second term of the formula is to measure the impact of each sentence exist in the best phenotype summary. The sentence with the highest score has a great impact on the summary and it is of high importance whereas the sentence with the lowest score has a little impact on the final summary. The sentences are sorted in descending order and the high scored sentences are selected to be included in the final summary until the required length L is reached. Perturbation Heuristic and Local Repair Heuristic, are presented in Algorithms 1 and 2 respectively.

Algorithm 1: Proposed Heuristic Perturbation Operator
Input: For a given gene i and for each sentence exist in the solution and for a random uniform variable $r_i \sim [0,1]$ $\forall i \in \{1,\dots,n\} \wedge \; x_i = 1 \wedge r_i \leq p_p$ **Process:** Set $x_i' = 1$ if the average similarity of all sentences s_j in the collection to the centre of document collection exceeds s_i similarity, otherwise turn it to zero Set $x_i' = 1 \; if \; sim(s_i, \mathbb{O}) \geq \frac{1}{n}\sum_{j=1}^{n} sim(s_j, \mathbb{O})$, otherwise set $x_i' = 0$
Algorithm 2: Proposed Local Repair Heuristic
Input: • $\forall i, j \in \{1,\dots,n\} \wedge x_i, x_j \in \mathbb{P}^* = 1$: Turn to zero each sentence that does not satisfy the condition: $sim(s_i, s_j) \leq \delta$ • For every sentence s_i included in the best solution $\mathbb{P}^*$ $\forall i \in \{1,\dots,n\} \wedge x_i \in \mathbb{P}^* = 1$ perform the following steps:

Process:

- Find the center of the document collection $\mathbb{O}$ according to the formula: $o_k = \frac{1}{n}\sum_{i=1}^{n} w_{ik}$, $k = 1,2,3, \dots, m$
- Calculate similarity of sentence s_i to the mean vector of the document collection $\mathbb{O}$
- Find the center of the created summary O^{sum}
- Compute similarity of mean vector of system generated summary O^{sum} to the center of document collection $\mathbb{O}$
- Compute similarity of mean vector of system generated summary O^{sum} excluded from it the specified sentence s_i to the center of document collection $\mathbb{O}$
- Apply the formula:

$$score_{s_i} = sim(s_i, \mathbb{O}) + \left(\left(sim(O^{sum}, \mathbb{O}) - sim(O^{sum-s_i}, \mathbb{O})\right)\right)$$

- Attach the calculated score to sentence s_i in order to decide according to this score about the inclusion of s_i in the final summary

7. Simulation results and discussion

7.1Dataset and parameters setting

Qualitative evaluations of the proposed model were made quantitatively based on the multi-document summarization datasets provided by Document Understanding Conference (DUC), particularly using DUC 2002 dataset [28]. A brief statistics of the dataset are given in Table-1. Like all other related works, the documents in DUC 2002 dataset are, first, preprocessed as follows:

- Segmentation of the documents into individual sentences,
- Identical sentences are removed,
- Sentences are tokenized,
- Stop words are removed and
- Finally, the remaining words are stemmed using Porter stemming algorithm [29].

Parameters for the proposed algorithm applied to solve multi objective based model $MOEA_{\Phi}$ are set as follows: a population of $pop_{size} = 50$ individuals is used and evolved over a sequence of $gen_{max} = 100$. For the tournament selection, a tournament size, $tour_{size} = 2$ has been chosen. Crossover probability and mutation probability are set to $p_c = 0.7$ and $p_p = 0.1$, respectively.

Table 1-Description of DUC2002 dataset.

Description	**DUC 2002 dataset**
Number of topics	59 (d061j through d120i)
Number of documents in each topic	~ 10
Total number of documents	567
Data source	TREC
Summary length	200 and 400 words

7.2 Evaluation metrics

The proposed work is quantitatively measured using Recall-Oriented Understudy for Gisting Evaluation $ROUGE$ evaluation metric [30]. $ROUGE$ is considered as the official evaluation metric for text summarization by DUC. It includes measures that automatically determine the quality of a summary generated by computer through comparison made between it and human generated summaries. The comparison is satisfied by counting the number of overlapping units, such as $N-grams$, word sequences, and word pairs between the summary generated by a machine and a set of *reference* summaries generated by humans.

$ROUGE-N$ is an $N-gram$ Recall counting the number of $N-grams$ matches of two summaries, and it is calculated as follows [30]:

$$ROUGE-N = \frac{\sum_{S\in\{reference\ Summaries\}}\sum_{N-gram\,\in\,S} Count_{match}(N-gram)}{\sum_{S\in\{reference\ Summaries\}}\sum_{N-gram\,\in\,S} Count(N-gram)} \quad (21)$$

where N stands for the length of the $N-gram$, $Count_{match}(N-gram)$ is the maximum number of $N-grams$ co-occurring in candidate summary and the set of reference summaries. $Count(N-gram)$ is the number of $N-grams$ in the reference summaries.
The similarity between reference summary sentence X of length m and candidate summary sentence Y of length n is calculated using $ROUGE-L$ measure (also called $f_{measure}$ which is denoted by F_{lcs}). $ROUGE-L$ evaluates the ratio between the length of the longest common subsequence of the two summaries $LCS(X,Y)$ and the length of the reference summary as follows [30]:

$$R_{lcs}=\frac{LCS(X,Y)}{m} \quad (22)$$

$$P_{lcs}=\frac{LCS(X,Y)}{n} \quad (23)$$

$$F_{lcs}=\frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs}+\beta^2 P_{lcs}} \quad (24)$$

Where recall and precision of the $LCS(X,Y)$ is denoted by R_{lcs} and P_{lcs}, respectively and $\beta=\frac{P_{lcs}}{R_{lcs}}$.
If the definition of $ROUGE-L$ is applied to summary-level, the union LCS matches between a reference summary sentence, r_i, and sentences of the candidate summary, C which is denoted by $LCS_{\cup}(r_i,C)$ is taken. Given a reference summary of u sentences containing a total of m words and a candidate summary of v sentences containing a total of n words, then summary-level $ROUGE-L$ is calculated as follows [30]:

$$R_{lcs}=\frac{\sum_{i=1}^{u} LCS_{\cup}(r_i,C)}{m} \quad (25)$$

$$P_{lcs}=\frac{\sum_{i=1}^{v} LCS_{\cup}(r_i,C)}{n}, \quad (26)$$

$$F_{lcs}=\frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs}+\beta^2 P_{lcs}}, \quad (27)$$

7.3 Model Performance

Table-2 together with Figure-1 present the comparison results based on $DUC2002$ dataset for average $Rouge-2$ and $Rouge-L$ scores for 20 runs of the proposed model $MOEA_{\Phi}$ with other baseline methods. The recorded results clarify that the proposed MOO based model significantly outperforms Single Objective Optimization (SOO) version of $MOEA_{\Phi}$ termed as $SOEA_{\Phi}$ and other baseline methods for modeling multi-document summarization despite that the proposed work works on $gen_{max}=100$ and the baseline systems work on $gen_{max}=1000$.

Table 2-Comparison results regarding average $Rouge-2$ and $Rouge-L$ scores of the proposed model $MOEA_{\Phi}$ against $SOEA_{\Phi 1}$, $SOEA_{\Phi}$, $MOEA_{\Phi 1}$ models and other state of the art models.

Method	$\overline{ROUGE-2}$	$\overline{ROUGE-L}$
DUC best	0.25229	0.46803
FGB	0.24103	0.4508
BSTM	0.24571	0.45516
LexRank	0.22949	0.44332
LSA	0.15022	0.40507
NMF	0.16280	0.41513
Centroid	0.19181	0.43237
$\Phi_{[18]}$	0.25184	0.46631
$SOEA_{\Phi 1[8]}$	0.25437	0.48314

$SOEA_{\Phi}$	0.27889	0.49412
$MOEA_{\Phi 1}$[9]	0.46578	0.60105
$MOEA_{\Phi}$	0.47412	0.61742

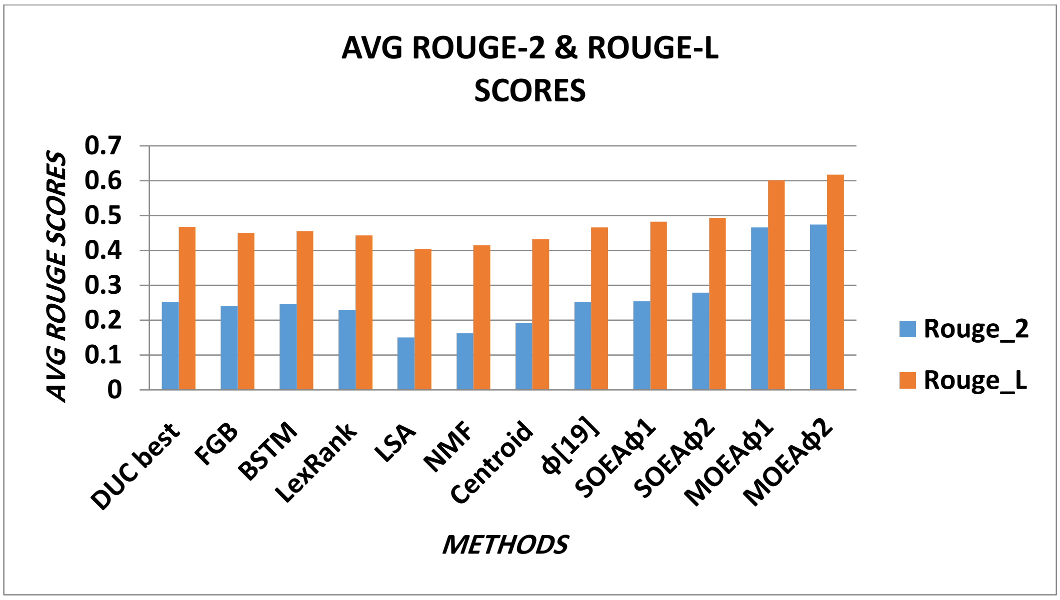

Figure 1-Comparison results of the proposed MOO model $MOEA_{\Phi}$ against SOO models $SOEA_{\Phi 1}$ and $SOEA_{\Phi}$, MOO model $MOEA_{\Phi 1}$ and other state of the art methods.

Table-2 and Figure-1 clearly point out that $MOEA_{\Phi}$ significantly outperforms the other state of the art methods. The reason for this improvement can be turned back to the positive collaboration among the three participants: MOEA/D, MOO based model, and the proposed heuristics. Apart from the common evolutionary operators included in both single EA and MOEA/D, the additional components possessed by MOEA/D represented by the external archive of non-dominated solutions and neighborhood-relative evolutionary operations can further harness its strength against the counterpart single EAs. Vital chromosome solutions can propagate their generated summaries to their neighbors and to the external archive gradually generation by generation. An important point should be regarded here, despite that the number of generations for applying SOO based models are greater than the maximum number of generation where MOO based models implemented, it is observed that MOO based models outperform SOO based models at all $Rouge$ scores.

Finally, for the MOO based models, it is clarified, in general, that both $MOEA_{\Phi 1}$ and $MOEA_{\Phi}$ have nearby behavior in their performance, with additional improvement to the $MOEA_{\Phi}$ model. From our investigation, it is noticed that non-dominated solutions contained in the archive of $MOEA_{\Phi}$ produce summaries with quality higher than the quality of summaries generated from non-dominated solutions contained in $MOEA_{\Phi 1}$ external archive despite the smaller size of $MOEA_{\Phi}$ archive compared to the size of $MOEA_{\Phi 1}$ external archive.

Results recorded in Table-3 summarize the positive impact of adopting MOO to the field of text summarization with the aid of both the proposed model and heuristics in terms of Relative Improvement (RI) of the proposed model $MOEA_{\Phi}$ over all the other state of the art methods at all $Rouge$ scores.

$$RI = \frac{Proposed\ method - Other\ method}{Other\ method} \quad (28)$$

Table 3-Improvement of the proposed $MOEA_{\Phi}$ over MOO model $MOEA_{\Phi1}$, SOO model $SOEA_{\Phi1}$ and $SOEA_{\Phi}$ and other state of the art methods on DUC2002 dataset.

Methods	$\boldsymbol{MOEA_{\Phi2}}$ ***Improvement***	
	$\boldsymbol{ROUGE-2}$	$\boldsymbol{ROUGE-L}$
DUC best	+0.8792659	+0.3191889
FGB	+0.9670580	+0.3696096
BSTM	+0.9295918	+0.3564900
LexRank	+1.0659724	+0.3927186
LSA	+2.1561709	+0.5242304
NMF	+1.9122850	+0.4872931
Centroid	+1.4718211	+0.4279899
$\boldsymbol{\Phi}$[18]	+0.8826239	+0.3240548
$\boldsymbol{SOEA_{\Phi1}}$[8]	+0.8638990	+0.2779319
$\boldsymbol{SOEA_{\Phi}}$	+0.7000251	+0.2495345
$\boldsymbol{MOEA_{\Phi1}}$[9]	+0.0179054	+0.0272357

8. Conclusions and future directions

In the work proposed in this paper, the problem of generic extractive multi document text summarization has been defined by projecting the content coverage in the light of text similarity. The proposed model hypothesizes a possible decomposition of text similarity into three different levels of optimization formula. First, aspire to *global* optimization, the candidate summary should cover the summary of the document collection. Then, to attain, less *global* optimization, the sentences of the candidate summary should cover the summary of the document collection. The third level of optimization is content with *local* optimization, where the difference between the magnitude of terms covered by the candidate summary and those of the document collection should be small. This coverage model has been coupled with a proposed diversity model and defined as a *multi-objective optimization* (*MOO*) problem.

Positive impact of adopting MOO to the field of text summarization with the aid of both proposing a model for coverage criteria that depends on the decomposition of text similarity into three different levels of optimization formula and heuristics has been recorded against all the other state of the art methods.

The proposed work may be Extended or extra improvements may be added to the it through a number of ways represented by the following directions

- Improving the tasks of preprocessing phase that have a positive impact on the improvement of the overall text summarization system and will produce summaries with high quality. The focus may be on adding further rules to the stemmer to improve stems quality, or on dealing with punctuation marks via some effective schemes.
- Applying the proposed system for the summarization of Arabic texts via working on preprocessing phase through considering the rules dedicated for segmentation, tokenization and stemming of texts in Arabic.
- Additional objectives can be taken in consideration by the proposed model. For instance, coherence and cohesion objectives are examples of such objectives to be optimized simultaneously in addition to the content coverage and redundancy reduction objectives.

References

1. Bawakid, Abdullah. **2011**. Automatic documents summarization using ontology based methodologies. PhD diss., University of Birmingham.
2. Gupta, Vishal, and Gurpreet Singh Lehal. **2010**. A survey of text summarization extractive techniques. *Journal of Emerging Technologies in Web Intelligence*, **2**(3): 258-268.
3. Kazantseva, Anna, and Stan Szpakowicz **2010**. Summarizing short stories. *Computational Linguistics*, **36**(1): 71-109.
4. Tao, Yuhui, Shuigeng Zhou, Wai Lam, and Jihong Guan. **2008**. Towards more effective text summarization based on textual association networks. In *Semantics, Knowledge and Grid. SKG'08*. Fourth International Conference on, pp. 235-240. IEEE.
5. Fattah, Mohamed Abdel, and Fuji Ren. **2009**. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech & Language*, **23**(1): 126-144.
6. Dong, Hui, Siwei Yu, and Ying Jiang. **2009**. Text mining on semi-structured e-government digital archives of China. In *Web Mining and Web-based Application*, 2009. WMWA'09. Second Pacific-Asia Conference on, pp. 11-14. IEEE.
7. Aliguliyev, Ramiz M. **2009**. A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*, **36**(4): 7764-7772.
8. Saleh, H., H., Kadhim, N., J. and Attea, B., A. **2015**. A Genetic Based Optimization Model for Extractive Multi-Document Text Summarization. *Iraqi Journal of Science*, **56**(2B): 1489-1498.
9. Saleh, H. H., Kadhim, N. J. **2016**. Extractive Multi-Document Text Summarization Using Multi-Objective Evolutionary Algorithm Based Model. *Iraqi Journal of Science, 5***7**(1C): 728-741.
10. Aliguliyev, Ramiz M **2010**. Clustering Techniques And Discrete Particle Swarm Optimization Algorithm For Multi-Document Summarization. *Computational Intelligence*, **26**(4): 420-448.
11. Radev, Dragomir R., Hongyan Jing, Małgorzata Styś, and Daniel Tam. **2004**. Centroid-based summarization of multiple documents. *Information Processing & Management*, **40**(6): 919-938.
12. Al-Hashemi, Rafeeq **2010**. Text Summarization Extraction System (TSES) Using Extracted Keywords. *Int. Arab J. e-Technol.* **1**(4): 164-168.
13. Zajic, David M., Bonnie J. Dorr, and Jimmy Lin. **2008**. Single-document and multi-document summarization techniques for email threads using sentence compression. *Information Processing & Management*, **44**(4): 1600-1610.
14. Mani, Inderjeet, and Mark T. Maybury. **1999**. *Advances in automatic text summarization*. MIT press.
15. Lin, H. and Bilmes, J. **2010.** Multi-document summarization via budgeted maximization of submodular functions. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 912-920). Association for Computational Linguistics.
16. Carbonell, J. and Goldstein, J. **1998.** The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 335-336). ACM.
17. Huang, Lei, Yanxiang He, Furu Wei, and Wenjie Li. **2010**. Modeling document summarization as multi-objective optimization. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on*, pp. 382-386. *IEEE*.
18. Alguliev, Rasim M., Ramiz M. Aliguliyev, and Chingiz A. Mehdiyev **2011**. Sentence selection for generic document summarization using an adaptive differential evolution algorithm. *Swarm and Evolutionary Computation*, **1**(4): 213-222.
19. Alguliev, Rasim M., Ramiz M. Aliguliyev, Makrufa S. Hajirahimova, and Chingiz A. Mehdiyev. **2011**. MCMR: Maximum coverage and minimum redundant text summarization model." *Expert Systems with Applications*, **38**(12): 14514-14522.
20. Alguliev, Rasim M., Ramiz M. Aliguliyev, and Chingiz A. Mehdiyev **2013**. An optimization approach to automatic generic document summarization. *Computational Intelligence*, **29**(1): 129-155.
21. Alguliev, Rasim M., Ramiz M. Aliguliyev, and Nijat R. Isazade **2013**. MR&MR-sum: maximum relevance and minimum redundancy document summarization model. *International Journal of Information Technology & Decision Making*, **12**(03): 361-393.

22. Alguliyev, Rasim M., Ramiz M. Aliguliyev, and Nijat R. Isazade **2015**. An unsupervised approach to generating generic summaries of documents. *Applied Soft Computing*, **34**: 236-250.
23. Salton, G., & Buckley, C. **1988.** Term-weighting approaches in automatic text retrieval. *Information processing & management*, **24**(5), 513-523.
24. Srinivas, N., & Deb, K. **1994.** Muiltiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary computation*, **2**(3), 221-248.
25. Zitzler, E., & Thiele, L. **1999.** Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *evolutionary computation, IEEE transactions on*, **3**(4), 257-271.
26. Coello, C. A. C., Van Veldhuizen, D. A., & Lamont, G. B. **2002.** *Evolutionary algorithms for solving multi-objective problems* (Vol. 242). New York: Kluwer Academic.
27. Zhang, Q., & Li, H. **2007.** MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, **11**(6), 712-731.
28. Document understanding conference: http://duc.nist.gov.
29. Porter stemming algorithm: http://www.tartarus.org/martin/PorterStemmer/.
30. Lin, C. Y. **2004**. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop* (Vol. 8).