



ISSN: 0067-2904
GIF: 0.851

Extractive Multi-Document Text Summarization Using Multi-Objective Evolutionary Algorithm Based Model

Hilal H. Saleh^{1*}, Nasreen J. Kadhim²

¹Department of Computer Science, University of Technology, Baghdad, Iraq

²Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Abstract

Automatic document summarization technology is evolving and may offer a solution to the problem of information overload. Multi-document summarization is an optimization problem demanding optimizing more than one objective function concurrently. The proposed work considers a balance of two significant objectives: content coverage and diversity while generating a summary from a collection of text documents. Despite the large efforts introduced from several researchers for designing and evaluating 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*. The design of generic text summarization model based on sentence extraction is 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 proposed two models are then coupled and defined as a *multi-objective optimization (MOO)* problem. Up to the best of our knowledge, this is the first attempt to address text summarization problem as a MOO model. Moreover, heuristic perturbation and heuristic local repair operators are proposed and injected into the adopted evolutionary algorithm to harness its strength. Assessment of the proposed model is performed using document sets supplied by Document Understanding Conference 2002 (DUC 2002) and a comparison is made with other state-of-the-art methods using Recall-Oriented Understudy for Gisting Evaluation (ROUGE) toolkit. Results obtained support strong proof for the effectiveness of the proposed model based on MOO over other state-of-the-art models.

Keywords: Multi-objective optimization; multi-objective multi-document text summarization problem; MOP; multi-objective evolutionary algorithm; MOEA/D; non-dominated solution.

التلخيص الأقتطاعي للنصوص متعددة المستندات باستخدام نموذج مستند على الخوارزمية التطورية

متعددة الاهداف

هلال هادي صالح^{1*}، نسرين جواد كاظم²

¹قسم علوم الحاسبات، الجامعة التكنولوجية، بغداد، العراق

²قسم علوم الحاسبات، كلية العلوم، جامعة بغداد، بغداد، العراق

الخلاصة

تقنية التلخيص الأوتوماتيكي 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 الاعتبار تحقيق التوازن بين هدفين مهمين هما: تغطية المحتوى

*Email: hhsrq888@yahoo.com

لمجموعة المستندات والتنوع عند توليد ملخص من مجموعة من المستندات النصية. 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 دالتان متناقضتان في أي ملخص. أن تصميم نموذج يهدف الى تلخيص نص عام قائم على أقتطاع الجمل تمت إعادة توجيهه الى تبيير ذات دلالة اكبر يعكس بصورة مستقلة كلا من تغطية وتنوع المحتوى كنموذجي أمثلية صريحين. بعد ذلك تمت عملية اقتران النموذجين المقترحين وتعريفهما كمشكلة أمثلية متعدد الاهداف. حسب علمنا ، هذه هي المحاولة الأولى لمعالجة مشكلة تلخيص النصوص كنموذج أمثلية متعدد الأهداف. وعلاوة على ذلك ، تم اقتراح عامل توجيه اضطراب وعامل توجيه أصلح محلي وحقنهما في الخوارزمية التطورية المعتمدة لتسخير قوتها . عملية تقييم النموذج المقترح تمت باستخدام مجموعة المستندات المجهزة من قبل مجموعة البيانات العالمية (Document Understanding Conference *DUC* 2002) وقد تمت مقارنة النتائج المتحصلة مع مجموعة من الانظمة الحديثة. قياس وتقييم الأداء للنموذج المقترح تم باستخدام أدوات (*ROUGE*). النتائج المتحصلة دعمت العمل بدليل قوي على فعالية النموذج المقترح المستند على أمثلية تعدد الاهداف نسبة الى النماذج الحديثة التي تمت المقارنة بها.

1. Introduction

Availability of massive amount of data in the Internet nowadays has reached such a vast volume that gradually becomes unfeasible for humans to efficiently filter out valuable information from it. Accordingly, a massive demand for innovative technologies that introduce an effective processing of documents is required. A vital technology to overwhelmed this obstacle in technological environments is *automatic text summarization*. 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 [1, 2]. This interprets the growing importance of the area of automatic text summarization which has triggered the race for developing many algorithmic models. This race is not necessary only for professionals who aim to find the information in a short time but also for large search engines like Google, Yahoo, AltaVista, and others. Text summarization problem attracts several disciplines from computer science, multimedia, and statistics to formulate and develop powerful techniques aiming 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 into two classes [3–6]: *generic summary* where a whole sense of document content is presented without any prior knowledge or *query-relevant summary* where the information presented in it 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 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 [5], [7-17]. 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 are not explicitly exist in the original text might be involved in such a summary [18]. Moreover, considering number of simultaneous analyzed documents, summary creation may be performed either from a single or multiple documents [5, 19]. 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.

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 [20]. Maximal Marginal Relevance [21] 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 [22], document summarization is formalized as a multi objective optimization problem. In particular, four objective functions, namely information coverage, significance, redundancy and text coherence are involved. These four objective functions measure the generated summaries according to the cluster of semantically or statistically related core terms.

In [23], an optimization-based method for opinion summarization based on the p -median clustering problem from facility location theory is proposed. Content selection is viewed as selection of clusters of related information.

Formulation for the widely used greedy maximum marginal relevance (MMR) algorithm as an integer linear programming is introduced in [24].

In [25], extractive multi-document text summarization is formalized as a discrete optimization problem and solved using an adaptive differential evolution algorithm. The approach is presented towards all of the three aspects of summarization: content coverage, redundancy and length.

In [26], unsupervised model formulated as an integer linear programming problem for multi-document summarization is proposed. The proposed model demonstrates that the summarization result depends on the similarity measure. A combination of the NGD-based and cosine similarity measures conducts to better result than their use separately.

In [27], document summarization is modeled as a nonlinear 0–1 programming problem where an objective function is defined as Heronian mean of the objective functions defining content coverage maximization and redundancy minimization. The optimization problem is solved using discrete particle swarm algorithm which is based on estimation of distribution algorithm.

The work in [28] formulated text summarization as a modified p -median problem taking in consideration four objectives: relevance, content coverage, redundancy minimization, and bounded length that are of great necessity to generate good summaries. A self-adaptive differential evolution algorithm is created to solve the proposed model.

Multiple document summarization is modeled in [29] as a Quadratic Boolean Programming problem which is a weighted combination of two objectives that are important to generate a good summary: content coverage and redundancy reduction. The optimization problem is solved using a modified differential evolution algorithm.

Extractive multi document summarization is modeled in [30] as the linear and nonlinear optimization problems. Coverage and diversity which are the most important factors to be satisfied in the generated summary are attempted to be simultaneously balanced in these models. Optimization problem is solved via developing a novel particle swarm optimization algorithm.

Work in [31] proposed a constraint-driven models formulated as a quadratic integer programming problem for multi-document summarization that emphasize diversity in summarization and sufficient coverage. It is observed that the proposed models together with alteration of the constraint parameters can drive coverage and diversity in a summary. The optimization problem is solved using a discrete particle swarm optimization algorithm.

Modeling of generic document summarization is performed in [32] as a discrete optimization problem. This model uses sentence-to-document relations, summary-to-document relations and sentence-to-sentence relations for extracting significant sentences from the given set of documents

whereas satisfying redundancy reduction in the final summaries. A self-adaptive differential evolution algorithm is generated for solving discrete optimization problem.

In [33], A mathematical model involving two stages is modeled as a discrete optimization problem. The first stage introduces topics detection via clustering of the sentences in document collection using k-means algorithm. At the second stage, the relevant sentences from each cluster are extracted and redundancy is avoided using sentence and cluster and sentence-to-sentence relations. The discrete optimization problem is solved using a differential evolution algorithm with the development of an adaptive mutation strategy.

In [34], text summarization is modeled as a quadratic integer-programming problem. Relevance, redundancy and summary length are optimized using this model. A novel differential evolution algorithm is produced for solving the optimization problem.

Extractive multi document text summarization is modeled in [35] as a modified p-median problem regarding relevance, information coverage, diversity and length limit which are basic requirements for satisfying summaries. The proposed model expresses summary-to-document relationships and summary-to-subtopics relationships in addition to sentence-to-sentence relationship. The optimization problem is solved using a modified differential evolution algorithm composed mainly on self-adaptive mutation and crossover parameters.

In [36], extractive summarization is modeled as a nonlinear 0–1 programming problem that take in its consideration the coverage, redundancy reduction, and limited length which are the basic requirements to satisfy the summary. The model satisfies a balance between coverage and diversity in the generated extractive summary.

An unsupervised approach based on optimization to automatic extractive document summarization is proposed in [37]. Text summarization is modeled in this work as a Boolean programming problem. Three properties represented by relevance, redundancy reduction, and length are attempted to be optimized via this model. Results clarified that symmetric and asymmetric similarity measures as a combination produces better result than their use individually.

Work in [38] takes in its consideration when generating a multi-document summary, performing a balance between the two important objectives in text summarization, content coverage maximization and minimization of information redundancy. Multi-document summarization is modeled in this work as a Quadratic Boolean Programming problem where the weighted combination of the two objectives produces the objective function. A binary differential evolution algorithm is used to solve the optimization problem.

A fuzzy evolutionary optimization model (FEOM) is proposed in [39]. FEOM is applied to document clustering and extractive summarization. For extractive summarization, categorization to the document sentences in terms of their content is performed initially. Next, from each cluster, a representative sentence for the overall document content which is the most important sentence is selected. For document similarity measures, a normalized Google distance is used.

Simultaneous optimization of many objectives is involved in many real world problems in engineering, industry, and in many other fields. A multi-objective optimization problem (MOP) 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 multi-objective optimization field in order to model and solve multi-objective optimization 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. The main contribution of the proposed work is to redirect modeling text summarization into more semantic measure reflecting individually both content coverage and content diversity as two explicit optimization models. The proposed two models are then coupled and defined as a *multi-objective optimization (MOO)* problem. Up to the best of our knowledge, this is the first attempt to address text summarization problem as a MOO model. The proposed model attempt to rigorously cast on the two 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 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 [40]. 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 metric can be formulated, according to *term-frequency inverse-sentence-frequency* scheme (*tf_isf*), as [40]:

$$w_{ik} = tf_{ik} \times isf, \quad (1)$$

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):

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

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 [19]:

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

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 $\mathbb{D} = \{s_i | 1 \leq i \leq n\}$, where n is the number of distinct sentences from the documents in \mathbb{D} . The aim of this paper is to generate a summary $\mathbb{D} \subset \mathbb{D}$ 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 proposed MOO based model proposed for modeling multi-document text summarization problem.

Definition 1 (Coverage objective function $\Phi_{coverage}$). Let $s_i \in \mathbb{D}$ be a sentence to be included in the summary $\bar{\mathbb{D}}$, then the *content coverage*, expressed by the similarity $sim(s_i, \mathbb{O})$ between s_i and the set of sentences in the document collection \mathbb{D} (represented by its mean vector \mathbb{O}) should be *maximized*.

$$\Phi_{coverage} = \sum_{i=1}^n sim(s_i, \mathbb{O}) x_i \tag{4}$$

Definition 2 (Redundancy reduction objective function Φ_{redund_red}). Let $s_i, s_j \in \mathbb{D}$ be two sentences to be included in the summary $\bar{\mathbb{D}}$. The similarity $sim(s_i, s_j)$ between them should be *minimized*, or quantitatively, *redundancy reduction* should be *maximized*.

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

Now, to formalize our suggestion, the *text summarization problem* will be modeled using the following definition:

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 $\bar{\mathbb{D}}$ (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 $\bar{\mathbb{D}}$ (see Eq. 7). Now, let $X = \{x_i | 1 \leq i \leq n\}$ 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) is a constrained maximization problem taking a combination of maximizing the two objective functions representing content coverage and information redundancy reduction $\Phi_{coverage}$ and Φ_{redund_red} .

$$x_i = \begin{cases} 1 & \text{if } s_i \in \bar{\mathbb{D}} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

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

$$\text{Maximize } MOEA_{\Phi}(x) = \{\Phi_{coverage}, \Phi_{redund_red}\} \tag{8}$$

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

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 have the ability to find simultaneously multiple optima. Formulation of multi-objective optimization problem consisting of m objective functions can be stated as follows:

$$\text{minimize or maximize } F(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))^T \tag{11}$$

subject to $\mathbf{x} \in \Omega$

Formally speaking, a general multi-objective optimization problem aim to find the vector $\mathbf{x}^* = [x_1^*, x_2^*, \dots, x_m^*]^T$ which is the result of optimizing the objective function vector in Eq. 11, where \mathbf{x} denotes the decision variable vector, $F: \Omega \rightarrow \mathbb{R}^m$ 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: [41–44]). Decomposition Based Multi-objective Evolutionary Algorithm (MOEA/D) [44] offered by Zhang and Li is one of the dominant algorithm 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 [44]. 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:

$$g^{te}(\mathbb{x}|\lambda_i, z^*) = \max\{\lambda_{i,j}|f_j(\mathbb{x}) - z_j^*\} \quad (12)$$

$$1 \leq j \leq m$$

Where $\lambda_i = (\lambda_{i,1}, \dots, \lambda_{i,m})^T$ is the weight vector, i.e., $\forall j = 1, \dots, m: \lambda_{i,j} \geq 0$ 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 i th subproblem has the current solution \mathbb{P}_i ,
- FV_1, \dots, FV_N where $FV_i = F(\mathbb{P}_i) = (f_1(\mathbb{P}_i), \dots, f_m(\mathbb{P}_i))^T \forall i = 1, \dots, N$,
- $z^* = (z_1^*, \dots, z_m^*)^T$ 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 [44] 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 $MOEA_\phi i = MOEA_\phi(\mathbb{P}_i) = (MOEA_\phi^1(\mathbb{P}_i), MOEA_\phi^2(\mathbb{P}_i))^T \forall i = 1, \dots, N$; 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 $\mathbb{P}_{1 \leq i \leq N} \in \mathbb{IP}$ is considered as a vector with fixed-length having $size = n$, where each gene determine the existence or absence of the equivalent sentence. $\Phi: \mathbb{P}_i \rightarrow \mathbb{R}^2$ 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:

$$\gamma = \{s_{\theta_s}, c_{\theta_c}, m_{\theta_m} | s_{\theta_s}, c_{\theta_c}, m_{\theta_m}: \mathbb{P} \rightarrow \mathbb{IP}\} \quad (13)$$

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:

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

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., $r_i \leq p_p$) 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 \mathbb{D} . Otherwise, it can be removed from the summary. Formally speaking,

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

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

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.

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

However, the phenotype of the best solution \mathbb{P}^* may still suffer from violating the length constraint i.e. $\sum_{i=1}^n l_i x_i > L$

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:

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

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

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

$$score_{s_i} = sim(s_i, \mathbb{O}) + ((sim(\mathbb{O}^{sum}, \mathbb{O}) - sim(\mathbb{O}^{sum-s_i}, \mathbb{O})) * 10) \tag{20}$$

Where $sim(\mathbb{O}^{sum}, \mathbb{O})$ 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, $sim(\mathbb{O}^{sum-s_i}, \mathbb{O})$ denotes the similarity between the generated summary (excluding sentence s_i) and the centre of document collection \mathbb{O} . The right term of the proposed formula in (Eq. 20) is multiplied by 10 in order to unify the scale of the two terms. The basic idea behind the right 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. Framework of the proposed MOEA/D algorithm is presented in Algorithm 1.

Algorithm 1: Framework of The Proposed MOEA/D	
Input:	
•	<i>Multi-objective Text Summarization</i> problem: Maximize $MOEA_{\Phi}(\mathbf{x}) = \{MOEA_{\Phi}^1(\mathbf{x}), MOEA_{\Phi}^2(\mathbf{x})\}$ Where: $MOEA_{\Phi}^1(\mathbf{x}) = \Phi_{coverage}(\mathbf{x}), MOEA_{\Phi}^2(\mathbf{x}) = \Phi_{redund_red}(\mathbf{x})$
•	Number of sub-problems to be evolved (size of population), N
•	Uniform spread of N weight vectors: $\lambda_1, \dots, \lambda_N$ such that $\lambda_i = (\lambda_{i,1}, \lambda_{i,2})$
•	Neighborhood size of each weight vector, T
•	Maximum number of generations, gen_{max}
•	Probability of crossover, p_c

- Probability of heuristic mutation, p_p

Output:

- EP involves set of non-dominated solutions

Step Zero – Setting step:

- $EP = \emptyset$
- $gen_no = 0$

Step One – Initializing step

- Initial internal population is generated randomly, $\mathbb{P}_0 = \{\mathbb{P}_1, \dots, \mathbb{P}_N\}$
- Set $MOEA_\phi i = MOEA_\phi(\mathbb{P}_i)$.
- Initialize $z^* = (z_1^*, z_2^*)^T$ randomly.
- Between any two weight vectors, Euclidean distance is computed and then the T closest weight vectors are worked out for each weight vector. $\forall i = 1, \dots, N$, set $B(i) = \{i^1, \dots, i^T\}$, where $\lambda_i^1, \dots, \lambda_i^T$ are the T closest weight vectors to λ_i

Step Two – Updating step: For $i = 1, \dots, N$

- Genetic operators: Two indices $k, l \in B(i)$ are selected randomly, and a new solution \mathbb{Y} is generated from \mathbb{P}_k and \mathbb{P}_l thru applying crossover and *Heuristic Perturbation*.
- Updating z^* , $\forall j = \{1, 2\}$, set $z_j^* = MOEA_\phi^j(\mathbb{Y})$ if $z_j^* > MOEA_\phi^j(\mathbb{Y})$.
- Neighbors solutions updating: For every index $j \in B(i)$, set $\mathbb{P}_j = \mathbb{Y}$ and $MOEA_\phi j = MOEA_\phi(\mathbb{Y})$ if $g^{te}(\mathbb{Y}|\lambda_j, z^*) \leq g^{te}(\mathbb{P}_j|\lambda_j, z^*)$.
- EP updating: All vectors dominated by $MOEA_\phi(\mathbb{Y})$ are removed from EP . If $MOEA_\phi(\mathbb{Y})$ is not dominated by any vector in EP , add $MOEA_\phi(\mathbb{Y})$ to EP .

Step Three – Stopping criteria

- Terminate and output EP if $gen_no = gen_max$, else $gen_no = gen_no + 1$, go and perform from Step Two.

Step Four – Apply *Local Repair Heuristic* on EP

7. Simulation results and discussion

7.1. Dataset and parameters setting

Qualitative evaluations of the proposed model was made quantitatively based on the multi-document summarization datasets provided by Document Understanding Conference (*DUC*), particularly using *DUC 2002* dataset [45]. A brief statistics of the dataset is 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,
- Sentences are tokenized,
- Stop words are removed and
- Finally, the remaining words are stemmed using Porter stemming algorithm [46].

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 the DUC2002 dataset

Description	<i>DUC 2002</i> 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 d 400 words

7.2. Evaluation metrics

The proposed work is quantitatively measured using Recall-Oriented Understudy for Gisting Evaluation *ROUGE* evaluation metric [47]. *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 [47]:

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

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 [47]:

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

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

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

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 [47]:

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

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

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

7.3. Models Performance

Table-2 together with Figure-1 present the comparison results on *DUC2002* 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 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 of the proposed models $MOEA_{\Phi}$ with other state of the art models in terms of average *Rouge – 2* and *Rouge – L* scores

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
$\Phi_{[25]}$	0.25184	0.46631
$MOEA_{\Phi}$	0.46578	0.60105

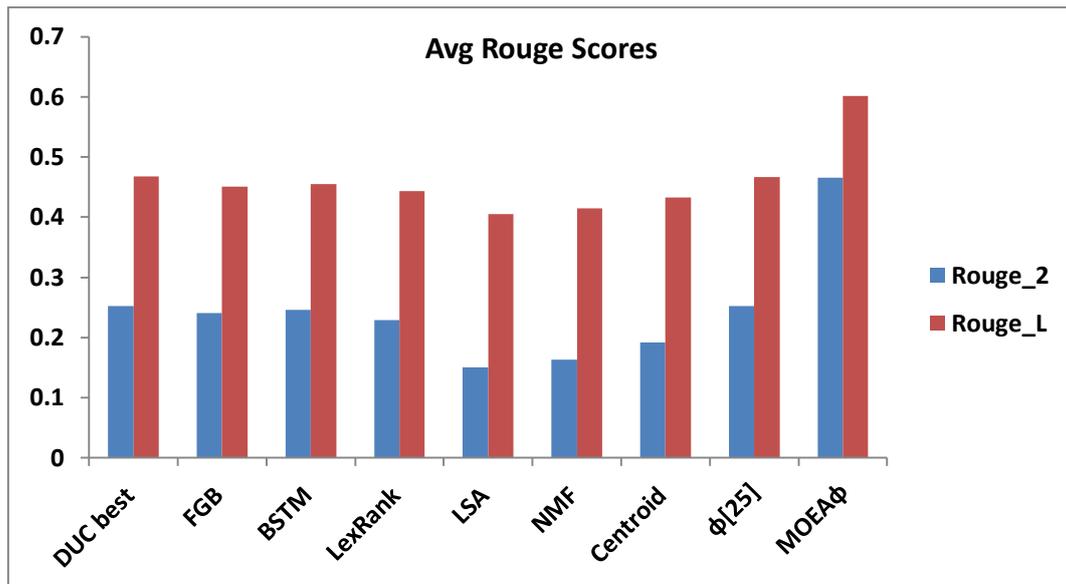


Figure 1-Comparison results between the proposed models $MOEA_{\phi}$ and the other state of the art methods.

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.

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

Table 3- Relative improvement of the proposed model $MOEA_{\phi}$ over other state of the art methods on the DUC2002 dataset

Methods	$MOEA_{\phi}$ Improvement	
	ROUGE – 2	ROUGE – L
DUC best	+0.8462087	+0.2842126
FGB	+0.9324565	+0.3332964
BSTM	+0.8956493	+0.3205247
LexRank	+1.0296309	+0.3557927
LSA	+2.1006524	+0.4838176
NMF	+1.8610565	+0.4478597
Centroid	+1.4283405	+0.3901288
Φ _[25]	+0.8495076	+0.2889494

8. Conclusions and future directions

Multi-document summarization is an optimization problem requiring synchronized optimization of more than one objective function. The need for effective multi-document summarization techniques to extract the important information from a document collection becomes of necessity. A good summary should have the ability to keep the key sentences representing the main topic of the document collection while simultaneously reducing irrelevant and redundant ones from the whole collection. 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 paper, the design of generic text summarization model based on sentence extraction is redirected into more semantic measure reflecting individually the two significant objectives: content coverage and diversity when generating summaries from multiple documents as two explicit optimization models. The proposed two models are then coupled and defined as a *multi-objective optimization (MOO)* problem. Up to the best of our knowledge, this is the first attempt to address text summarization problem as a MOO model. Moreover, heuristic perturbation and heuristic local repair operators are proposed and injected into the adopted evolutionary algorithm to harness its strength. Results obtained

clarify that the proposed MOO based model significantly outperforms other state-of-the-art models. Moreover, extra improvement may be added to the proposed work via a number of ways, such as:

- Additional objectives can be added to the proposed MOO model, for instance, coherence and cohesion objectives to be optimized simultaneously with content coverage and redundancy reduction.
- Applying different MOO algorithms to solve the optimization model.
- The set of the non-dominated solutions found in the external archive can be further improved by adopting one of the well-known local search operators.

References

1. Jones, K. S. **2007**. Automatic summarising: The state of the art. *Information Processing & Management*, 43(6), pp: 1449-1481.
2. Kazantseva, A. and Szpakowicz, S. **2010**. Summarizing short stories. *Computational Linguistics*, 36(1), pp: 71-109.
3. Shen, D., Sun, J. T., Li, H., Yang, Q. and Chen, Z. **2007**. Document Summarization Using Conditional Random Fields. *IJCAI*, 7, pp: 2862-2867.
4. Tao, Y., Zhou, S., Lam, W. and Guan, J. **2008**. Towards more effective text summarization based on textual association networks. *Semantics, Knowledge and Grid*, 2008. SKG'08. Fourth International Conference on IEEE. pp: 235-240.
5. Fattah, M. A., and Ren, F. **2009**. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech and Language*, 23(1), pp: 126-144.
6. Dong, H., Yu, S., and Jiang, Y. **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 IEEE. pp:11-14.
7. Aliguliyev, R. M. **2009**. A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications*, 36(4), pp: 7764-7772.
8. Yeh, J. Y., Ke, H. R., Yang, W. P. and Meng, I. H. **2005**. Text summarization using a trainable summarizer and latent semantic analysis. *Information Processing & Management*, 41(1), pp: 75-95.
9. Alguliev, R. M., and Aliguliyev, R. M. **2005**. Effective summarization method of text documents. In *Web Intelligence*, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on IEEE. pp: 264-271.
10. Alguliev, R. M., and Alguliev, R. M. **2008**. Automatic text documents summarization through sentences clustering. *Journal of Automation and Information Sciences*, 40(9).
11. Alguliev, R. M., Alyguliev, R. M., and Bagirov, A. M. **2005**. Global optimization in the summarization of text documents. *Automatic Control and Computer Sciences*, 39(6), pp:42-47.
12. Aliguliyev, R. M. **2006**. A novel partitioning-based clustering method and generic document summarization. In *Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology* IEEE Computer Society. pp:626-629.
13. Aliguliyev, R. M. **2010**. Clustering Techniques and Discrete Particle Swarm Optimization Algorithm for Multi-Document Summarization. *Computational Intelligence*, 26(4), pp:420-448.
14. Gong, Y., and Liu, X. **2001**. Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* ACM. pp:19-25.
15. Radev, D. R., Jing, H., Styś, M., and Tam, D. **2004**. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40(6), pp:919-938.
16. Salton, G., Singhal, A., Mitra, M., and Buckley, C. **1997**. Automatic text structuring and summarization. *Information Processing and Management*, 33(2), pp:193-207.
17. Svore, K. M., Vanderwende, L. and Burges, C. J. **2007**. Enhancing Single-Document Summarization by Combining Rank Net and Third-Party Sources. *EMNLP-CoNLL*. pp: 448-457.
18. Al-Hashemi, R. **2010**. Text Summarization Extraction System (TSES) Using Extracted Keywords. *Int. Arab J. e-Technol.*, 1(4), pp:164-168.
19. Zajic, D. M., Dorr, B. J. and Lin, J. **2008**. Single-document and multi-document summarization techniques for email threads using sentence compression. *Information Processing and Management*, 44(4), pp:1600-1610.

20. 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 Association for Computational Linguistics. pp:912-920.
21. 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.
22. Huang, L., He, Y., Wei, F. and Li, W. **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.
23. Cheung, J. C. K., Carenini, G. and Ng, R. T. **2009**. Optimization-based content selection for opinion summarization. In Proceedings of the 2009 Workshop on Language Generation and Summarisation Association for Computational Linguistics. pp:7-14.
24. Riedhammer, K., Favre, B. and Hakkani-Tür, D. **2010**. Long story short—Global unsupervised models for keyphrase based meeting summarization. *Speech Communication*, 52(10), pp:801-815.
25. Alguliev, R. M., Aliguliyev, R. M., and Mehdiyev, C. A. **2011**. Sentence selection for generic document summarization using an adaptive differential evolution algorithm. *Swarm and Evolutionary Computation*, 1(4), pp: 213-222.
26. Alguliev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., and Mehdiyev, C. A. **2011**. MCMR: Maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications*, 38(12), pp:14514-14522.
27. Alguliev, R. M., Aliguliyev, R. M. and Mehdiyev, C. A. **2011**. An optimization model and DPSO-EDA for document summarization. *International Journal of Information Technology and Computer Science (IJITCS)*, 3(5), p:59.
28. Alguliev, R. M., Aliguliyev, R. M., and Mehdiyev, C. A. **2011**. pSum-Sade: a modified p-median problem and self-adaptive differential evolution algorithm for text summarization. *Applied Computational Intelligence and Soft Computing*, 2011, p:11.
29. Alguliev, R. M., Aliguliyev, R. M., and Hajirahimova, M. S. **2012**. Quadratic Boolean programming model and binary differential evolution algorithm for text summarization. *Problems of Information Technology*, (2), pp:20-29.
30. Alguliev, R. M., Aliguliyev, R. M., and Mehdiyev, C. A. **2013**. An optimization approach to automatic generic document summarization. *Computational Intelligence*, 29(1), pp:129-155.
31. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2013**. CDDS: Constraint-driven document summarization models. *Expert Systems with Applications*, 40(2), pp:458-465.
32. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2013**. Multiple documents summarization based on evolutionary optimization algorithm. *Expert Systems with Applications*, 40(5), pp: 1675-1689.
33. Aliguliyev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2015**. A New Similarity Measure and Mathematical Model for Text Summarization. *Problems of Information Technology*, 6(1), pp:42-53.
34. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2013**. MrandMr-Sum: Maximum Relevance And Minimum Redundancy Document Summarization Model. *International Journal of Information Technology and Decision Making*, 12(3), pp:361-393.
35. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2012**. DESAMC+ DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization. *Knowledge-Based Systems*, 36, pp:21-38.
36. Alguliev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2013**. Formulation of document summarization as a 0–1 nonlinear programming problem. *Computers and Industrial Engineering*, 64(1), pp: 94-102.
37. Aliguliyev, R. M., Aliguliyev, R. M., and Isazade, N. R. **2015**. An Unsupervised Approach to Generating Generic Summaries of Documents. *Applied Soft Computing*.
38. Alguliev, R. M., Aliguliyev, R. M., and Hajirahimova, M. S. **2012**. GenDocSum+ MCLR: Generic document summarization based on maximum coverage and less redundancy. *Expert Systems with Applications*, 39(16), pp: 12460-12473.

39. Song, W., Choi, L. C., Park, S. C., and Ding, X. F. **2011**. Fuzzy evolutionary optimization modeling and its applications to unsupervised categorization and extractive summarization. *Expert Systems with Applications*, 38(8), pp: 9112-9121.
40. Salton, G., and Buckley, C. **1988**. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), pp:513-523.
41. Srinivas, N., and Deb, K. **1994**. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3), pp:221-248.
42. Zitzler, E., and Thiele, L. **1999**. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *Evolutionary Computation, IEEE Transactions on*, 3(4), pp: 257-271.
43. Coello, C. A. C., Van Veldhuizen, D. A., and Lamont, G. B. **2002**. *Evolutionary algorithms for solving multi-objective problems*. 242. New York: Kluwer Academic.
44. Zhang, Q., and Li, H. **2007**. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *Evolutionary Computation, IEEE Transactions on*, 11(6), pp: 712-731.
45. Document understanding conference: <http://duc.nist.gov>.
46. Porter stemming algorithm: <http://www.tartarus.org/martin/PorterStemmer/>.
47. Lin, C. Y. **2004**. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*. 8.