DOI: 10.24996/ijs.2020.61.6.30

ISSN: 0067-2904

Extractive Multi-Document Summarization Model Based On Different Integrations of Double Similarity Measures

Dheyaa Abdulameer Mohammed[1], Nasreen J. Kadhim[2]*

[1]Imam Ja'afar Al-Sadiq University, Baghdad, Iraq

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

Received: 8/8/2019 Accepted: 22/10/2019

ABSTRACT

Currently, the prominence of automatic multi document summarization task belongs to the information rapid increasing on the Internet. Automatic document summarization technology is progressing and may offer a solution to the problem of information overload.

Automatic text summarization system has the challenge of producing a high quality summary. In this study, the design of generic text summarization model based on sentence extraction has been redirected into a more semantic measure reflecting individually the two significant objectives: *content coverage* and *diversity* when generating summaries from multiple documents as an explicit optimization model. The proposed two models have been then coupled and defined as a *single-objective optimization* problem. Also, for improving the performance of the proposed model, different integrations concerning two similarity measures have been introduced and applied to the proposed model along with the single similarity measures that are based on using *Cosine*, *Dice* and $Jaccard$ similarity measures for measuring text similarity. For solving the proposed model, Genetic Algorithm (GA) has been used. Document sets supplied by Document Understanding Conference 2002 ($DUC2002$) have been used for the proposed system as an evaluation dataset. Also, as an evaluation metric, Recall-Oriented Understudy for Gisting Evaluation ($ROUGE$) toolkit has been used for performance evaluation of the proposed method. Experimental results have illustrated the positive impact of measuring text similarity using double integration of similarity measures against single similarity measure when applied to the proposed model wherein the best performance in terms of $Rouge-2\ (0.1354)$ and $Rouge-1\ (0.4210)$ has been recorded for the integration of Cosine similarity and $Jaccard$ similarity.

Keywords: Text summarization, genetic algorithm, single metric similarity measure, double metric similarity measure.

نموذج تلخيص أقتطاعي للمستندات المتعددة مستند إلى تكاملات مختلفة لمقياس التشابه المزدوج

ضياء عبد الأمير محمد[1], نسرين جواد كاظم[2]*

[1]جامعة الأمام جعفر الصادق, بغداد, العراق.

[2]قسم علوم الحاسبات, كلية العلوم, جامعة بغداد, بغداد, العراق.

الخلاصه

في الوقت الحالي ، تعود أهمية تلخيص المستندات المتعددة الى تزايد المعلومات المتسارع على الإنترنت.

[*] Email: nasreen_jawad@yahoo.com

تقنية تلخيص المستندات في تقدم وقد توفر حلاً لمشكلة التحميل الزائد للمعلومات. نظام تلخيص النصوص التلقائي يمتلك التحدي لإنتاج ملخص عالي الجودة. في هذه البحث، تمت إعادة توجيه تصميم نموذج تلخيص النص العام استنادًا لاستخراج الجملة إلى مقياس أكثر دلالة يعكس بشكل فردي الهدفين المهمين: تغطية المحتوى وتنوعه عند إنشاء ملخصات من مستندات متعددة كنموذج تحسين صريح. تم بعد ذلك اقتران النموذجين المقترحين وتعريفه على أنه مشكلة تحسين أحادية الهدف. أيضًا، تم تقديم تكاملات مختلفة تستند على تقديم تدابير التشابه المزدوج وتطبيقها في النموذج المقترح بالإضافة إلى مقاييس التشابه الفردية لقياس تشابه النصوص. أوضح تقييم أداء النموذج المقترح التأثير الإيجابي لحساب تشابه النصوص من خلال تطبيق تكاملات مختلفة تتضمن مقياس التشابه المزدوج على إجراء هذه الحسابات باستخدام مقياس تشابه منفرد. تم استخدام الخوارزمية الجينية لحل الموديل المقترح. عملية تقييم النموذج المقترح تمت باستخدام مجموعة المستندات المجهزة من قبل مجموعة البيانات العالمية (Document Understanding Conference DUC2002). قياس وتقييم الأداء للنماذج المقترحة تم باستخدام أدوات (ROUGE). نتائج التجارب أوضحت التأثير الايجابي لقياس تشابه النصوص عبر استخدام التكامل المزدوج لمقاييس التشابه ضد استخدام مقياس التشابه المنفرد عندما تم استخدامه في النموذج المقترح حيث تم تسجي افضل اداء للنظام حيث سجل $Rouge-2 = 0.13542$ كما سجل قيمة $Rouge-1 = 0.4210$ عبر تكامل التشابه المكون من Cosine مع Jaccard

1 Introduction

One of the most important challenges facing humans today is the rapid increase in the amount of data generated by users, especially those on the Internet. Also, one of the most important types of data facing such a large increase is textual data, which made it very difficult for humans to take advantage of this data in its natural state. This has made the need for an automated summary system for those data more important. Although research on a system to automatically summarize documents began at the end of the 20th century, so far there is no satisfactory outcome, and all researches have relatively modest progress.

Text summarization is a way to condense the large amount of information into a concise form by the process of selection of important information and discarding unimportant and redundant information. With the amount of textual information present in the world wide web, the area of automatic text summarization is becoming very important in the field of Information Retrieval.

The search engines do a remarkable job in searching through a mass of information to dish out the most related information the user is searching for. Even the information picked by search engines with a great precision is of a daunting amount. Reading through whole length of the document is very time consuming. Always a certain task demands a decision to be made in a definite time frame, and to read through all the documents is simply difficult. Availability of the core of the document makes the process speed up considerably. When dealing with problems like that, the technology of automatic text summarization becomes critical.

The document summarization system can be classified as follows: Document summarization methodologies can be generally divided into extractive and abstractive methodologies. *Abstractive summarization* can be defined as producing a summary that involves concepts/ideas reserved from the source, which are then “reinterpreted” and offered in a dissimilar form. An ***extractive summarization*** is an approach for constructing a summary that consists of units of text reserved from the source and offered verbatim [1].

Taking in consideration the number of documents under summarization, the summary can be a condensed form of multiple documents or one document. Multiple document summarization aims at extracting information relevant to an **implicit** or **explicit** subject from different documents written about that subject or topic [2].

The approaches of extraction-based summarization can be categorized as **supervised** or **unsupervised**. **Supervised** approaches are constructed on algorithms that use a large number of summaries generated by human, and as an outcome, are most convenient for documents related to the summarizer model. Accordingly, they do not necessarily yield an adequate summary for documents that are dissimilar to the model. Furthermore, when the summarization purpose or documents' features are modified by the users, it becomes essential for reeducating the model or rebuilding the training data. **Unsupervised** approaches do not necessitate training data for training the summarizer.

Automatic summary can either involves the most significant information overall **(generic summarization)** or the most relevant information considering an information need of the user **(query-based summarization)**. **Generic summarization** approaches focus on covering diversity of the summary for delivering broader content coverage. Usually, they are described in terms of certain key features which relate to the concepts of intent, focus, and coverage.

Considering the usage, the summary can be **indicative** or **informative**. A condensed information on the key topics of a document can be provided through an **Indicative** summary. Document's most important passages should be preserved in this summary type and often used as the end part of the information retrieval systems, being retrieved by search system rather than full document. Their target should be to aid the user for deciding whether the reading for the original document is valuable or not. The typical length of an indicative summary ranges from 5% to 10% of the whole text. Dissimilarly, **informative** summaries deliver a condensation for a complete document, retaining significant information, while decreasing its volume. An informative summary is normally 20–30% of the original text [3].

The main contribution of this paper is to model the *multi-document text summarization* task as an optimization problem. The proposed model emphasizes the discovery of essential sentences that cover the main topic of the document collection while transcending the occurrence of redundant sentences. Different integrations of double metric similarity measure are introduced to the proposed model for measuring similarity to improve system performance. A binary-encoded genetic algorithm has been adopted to solve the modeled optimization problem. The organization of this paper is as follows. Section 2 presents the related works on extractive summarization. Elementary concepts for extractive multi-document text summarization together with the statement of the problem are introduced in section 3. Section 4 introduces the details of the proposed mathematical formulation and modeling. The proposed genetic algorithm for solving the optimization problem is introduced in section 5. The experiments performed and results are presented in Section 6. Finally, conclusions and some possible extensions to the current work are given in Section 7.

2 Related works

In literature, multi-document summarization approaches vary in their essence. Various extraction-based techniques have been proposed for generic text summarization [4]. In extraction based document summarization, generation of the optimal summary can be regarded as a combinatorial optimization problem wherein finding a solution to the problem is NP-hard. A review of the works based on optimization and are the most related to the method proposed in this paper is illustrated in what follows.

Alguliev et al. (2011) presented a document summarization model aimed at extracting significant sentences from a given collection of documents while performing reduction of information redundancy in the summary. An inventive aspect of their model lies in its capability to eliminate redundant information while choosing representative sentences. The representation of the model was performed as a discrete optimization problem. For solving the discrete optimization problem in their work, they created an adaptive Differential Evolution algorithm. They implemented their model on the task of multi-document summarization. Their experimental results showed that their proposed optimization approach was competitive on the DUC2004 and DUC2002 datasets [5].

ALGULIEV et al. (2011) proposed an unsupervised model for text summarization which performs generation to a summary by means of an extraction to the significant sentences in given document(s). They modeled TS as an integer linear programming problem. Their model has the ability for covering the core content of the collection through discovering the important sentences in it. This model also guaranteed that the summary cannot involve several sentences conveying similar information [6].

ALGULIEV et al. (2013) achieved a modeling to document summarization as nonlinear and linear optimization problems. These models attempted balancing diversity and coverage in the summary. The optimization problem was solved through developing a new particle swarm optimization (PSO) algorithm. Their experiments revealed that their proposed models produced very competitive results, which considerably outperformed the NIST baselines [7].

In ALGULIEV et al (2013), a model based on optimization for generic text summarization has been proposed. Their proposed model generated a summary through performing an extraction of significant sentences from documents. This method has been used for selecting significant sentences from a given collection of documents and reducing summary redundancy; the sentence-to-sentence, the summary-to-collection and the sentence to document collection relations. An improved differential evolution algorithm has been created for solving the optimization problem. For their proposed work, an adaptive adjustment could be performed on the crossover rate by the algorithm in accordance to individual fitness [1].

ALGULIEV et al (2015) presented an unsupervised optimization based method for automatically summarizing text. They modeled text summarization is a Boolean programming problem. In their model, three properties were attempted to be optimized, namely relevance, reducing redundancy and creating a summary with bounded length. Their proposed method was applicable to multiple and single-document summarization[8].

Asad Abdi et. Al. (2015) proposed a specialized method that works well in assessing short summaries. Their proposed method integrated the semantic relations between words and their syntactic composition. As a result, the proposed method was able to obtain high accuracy and improve the performance compared with the current techniques. Experiments showed that their work was preferred over the existing techniques [2].

In Rautrayand Balabantaray (2017), a novel Cat Swarm Optimization (CSO)-based multi document summarizer was proposed to address the problem of multi document summarization. The proposed CSO-based model was also compared with two other nature-inspired summarizers, namely the Harmony Search (HS)-based summarizer and Particle Swarm Optimization (PSO)-based summarizer [9].

Text summarization was modeled by ALGULIEV *et al*. (2019) as a two-stage sentence selection model constructed on optimization and clustering methods. Firstly, for discovering all topics in the text, they clustered the set of sentences through applying k-means method. Secondly, to select significant sentences from clusters, they proposed a model based on optimization. An objective function expressed as a harmonic mean of the objectives enforcing the coverage and diversity of the selected sentences in the summary was optimized in their optimization model. For providing the summary readability, their model also controlled the length of the chosen sentences. The optimization problem was solved through developing an adaptive differential evolution algorithm with a new mutation approach [10].

3 Extractive generic multi-document text summarization

3.1Preliminaries

Several methodologies have been explored for text similarity, however, they are centered around four major categories. These are word co-occurrence/vector-based methods, corpus-based methods, hybrid methods, and descriptive feature-based methods [11].

In text summarization, vector-based methods are commonly used [12]. 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 follows [12]:

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

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.

Measuring the similarity between words, sentences, paragraphs and documents is an important component in text-associated research and applications in several tasks, including text classification, text summarization, IR, document clustering and others. Calculating similarity between words is an essential part of measuring similarity between texts, which is used later as a primary stage for calculating similarities between sentences, paragraphs and documents [11].

Similarity between words can be satisfied lexically and semantically. Lexical similarity between words can be occurred if they have a similar character sequence. Whereas semantic similarity can be occurred if the words have the same meaning used in the same context [13].

For the model proposed in this paper, similarity between two texts has been measured using Cosine, Jaccard and Dice similarity. Cosine similarity is a measure used for computing the similarity between two vectors. This is achieved through calculating the cosine of the angel between them. Hence, if the inner product is used for finding the distance between two vectors, the cosine is used for finding the angel between these vectors. Using cosine similarity is a good technique for ranking documents through discovering the closest document to the user query [14].

$$sim\left(S_i, S_j\right) = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_{jk}^2}} \qquad i, j = 1,2,3 \ldots., n \tag{2}$$

Jaccard Similarity is a statistical similarity measure between sample sets. It performs a comparison between members for two sets to discover the shared and distinct members. Although its interpretation is easy and it is very sensitive to small samples sizes, it might provide incorrect results, particularly with very small data sets with missing observations [15].

$$J(Si, Sj) = \frac{|Si \cap Sj|}{|Si \cup Sj|} = \frac{|Si \cap Sj|}{|Si| + |Sj| - |Si \cap Sj|} \qquad i, j = 1,2 \ldots, n \tag{3}$$

Dice Similarity is similar to Jaccard and used for finding the similarity between two vectors, but " gives twice the weight to agreements" [16, 17, 18].

$$D(Si, Sj) = \frac{2|Si \cap Sj|}{|Si| + |Sj|} \qquad i, j = 1,2,3 \ldots, n \tag{4}$$

3.2 Problem statement

Consider a collection of documents $\mathbb{D}$ comprising N documents, i.e. $\mathbb{D} = \{d_1, \ldots, d_N\}$. Also, consider that $\mathbb{D}$ is totally composed of n sentences. In the language of sentences, $\mathbb{D}$ can be then denoted by $\mathbb{D} = \{s_i | 1 \leq i \leq n\}$, wherein n refers to the number of different sentences contained in all documents in $\mathbb{D}$. The objective of the proposed work is to generate a summary $\overline{\mathbb{D}} \subset \mathbb{D}$ while tackling three challenges:

- ***Covering Contents:*** the generated summary $\overline{\mathbb{D}}$ should cover the main topic of the collection $\mathbb{D}$.
- ***Reducing Redundancy:*** the created summary $\overline{\mathbb{D}}$ should not involve similar sentences contained in $\mathbb{D}$.
- ***Bounded length*:** length of the summary $\overline{\mathbb{D}}$ *should be restricted.*

4 The proposed model: definitions and formulations

In this paper, the text summarization problem is addressed as a *single objective optimization* problem. The intended summary $\overline{\mathbb{D}}$ is projected in the light of the defined problem as in the definitions of the proposed SOO based model $\boldsymbol{GA_\Phi}$ introduced in what follows.

Definition 1 (*Summary* $\overline{\mathbb{D}}$). Let $s_i \in \mathbb{D}$ be a sentence to be involved in $\overline{\mathbb{D}}$, then the *content coverage*, stated by the summation of similarity for each pair of sentences: $sim(s_i, \mathbb{O})$ between s_i and the set of sentences in the document collection $\mathbb{D}$ (represented by its mean vector $\mathbb{O}$) and $sim\left(s_j, \mathbb{O}\right)$ between s_j and the set of sentences in the document collection $\mathbb{D}$ should be *maximized*. Alternatively, *reduction of* redundancy, or quantitatively, the similarity $sim\left(s_i, s_j\right)$ between the same pair of sentences that belong to $\overline{\mathbb{D}}$ should be *minimized*. Now, to formulate our proposal, the problem of *text summarization* will be modeled through the definition introduced in what follows:

Definition 2 (*text summarization problem* $\boldsymbol{GA_\Phi}$). Let $x_i \in \{0,1\}$ be a binary decision variable that denotes the absence (0) or presence (1) of the sentence s_i in $\overline{\mathbb{D}}$ (Equation 5). Moreover, let $x_{ij} \in \{0,1\}$ be an additional binary decision variable related to the presence of both s_i and s_j in $\overline{\mathbb{D}}$ (Equation 6). Currently, let $X = \{x_i | 1 \leq i \leq n\}$ be a vector involving n such decision variables related to the n sentences. At that point, for a vector X, the problem of text summarization (see Eq. 7 & Eq. 8) is a constrained maximization problem considering maximization of the content coverage (numerator) and minimization of redundancy (denominator)

$$x_i = \begin{cases} 1 \; if \; s_i \in \overline{\mathbb{D}} \\ 0 \; otherwise \end{cases}, \tag{5}$$

$$x_{ij} = \begin{cases} 1 \; if \; s_i \; and \; s_j \in \overline{\mathbb{D}} \\ 0 \; otherwise \end{cases} \tag{6}$$

$$Maximize \quad \boldsymbol{GA_{\Phi}}(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{sim(s_i, \mathbb{O}) + sim(s_j, \mathbb{O})}{sim(s_i, s_j)} \right) x_{ij} \tag{7}$$

$$subject \; to \quad \sum_{i=1}^{n_{candidate}} l_i x_i \leq L, \tag{8}$$

where:
L: Summary length constraint,
l_i: Length of sentence s_i,
$n_{candidate}$: Total number of sentences in candidate summary,
$\mathbb{O}$: Center of the document collection $\mathbb{D} = \{s_1, s_2, \dots, s_n\}$.
Wherein it can be calculated as the average of weights w_{ik} of the term t_k at all n sentences contained in $\mathbb{D}$ as:

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

The SOO based model aims to include in the candidate summary the pair of sentences that gain high similarity to the main contents of the document collection in order to satisfy content coverage and, simultaneously, achieve low similarity between each other in order to introduce diverse ideas to the candidate summary.

4.1 The proposed similarity integrations

Different integrations of similarity measures are introduced and applied to the proposed model for measuring similarity, including:

- ***Single similarity measures integration:*** These metrics measure the similarity between a pair of sentences and between a sentence and the center of document collection through implementing individually the $Eqs.$ (2, 3, 4) for *Cosine*, $Jaccard$ and *Dice* similarity measures, respectively.
- ***Double similarity measures integration*:** These metrics measure the similarity between a pair of sentences and between a sentence and the center of document collection through implementing formulas that are considered as weighted sum equations of two similarity measures under consideration: *(Cosine* and $Jaccard$), *(Cosine* and *Dice)* and ($Jaccard$ and *Dice*).

$$sim(text_i, text_j)_{SimM_1 + SimM_2} = \sigma \times sim(text_i, text_j)_{SimM_1} + (1 - \sigma) \times sim(text_i, text_j)_{SimM_2} \tag{9}$$

5 The proposed genetic algorithm

Each genotype solution in the proposed GA is encoded using binary encoding and characterized by a fixed-length vector of size n, wherein each gene value is an indicator to the existence or nonexistence of its related sentence. Then, the entire search space δ for the proposed GA can be calculated by the Cartesian product of existence/nonexistence of all n sentences:

$$\delta = \prod_{i=1}^{n}(\{0,1\}) = 2^n \tag{10}$$

Consider a population ρ of $K \ll \delta$ genotype solutions, $\mathbb{P}_{1 \leq k \leq K} \in \rho$. Then, $\forall k \in \{1, \dots, K\} \; and \; \forall j \in \{1, \dots, n\}$: $\mathbb{P}_k = (\mathbb{P}_{k1}, \mathbb{P}_{k2}, \dots, \mathbb{P}_{kn}) \; s.t. \mathbb{P}_{kj} \in \{0,1\}$. The description of the proposed GA can be stated as a process expressed in an iterative function $\Psi: \rho \to \rho'$ with $\Psi(\rho_i) = \rho_{i+1}$, where ρ_i is the population at iteration i. The evolution function Ψ at every iteration i will be composed of three key operators: selection, crossover, and mutation operator, wherein their corresponding control parameters control each of them. Formally, this is noted as:

$$\Psi = s_{\Theta_s} \circ x_{\Theta_x} \circ p_{\Theta_p} \tag{11}$$

Through the application of the selection operator, s_{Θ_s}, copying the good quality chromosomes that are the fittest to the next generation is performed for improving the average quality of the population, whereas elimination of bad chromosomes is performed. The proposed work adopts the tournament selection wherein a selection is made to only one individual for the next generation if it is the fittest from several randomly chosen individuals. The control parameter Θ_s determines the number of randomly chosen individuals, i.e. *tournament size*.

The proposed algorithm adopts the Uniform Crossover. In accordance to this type of crossover, the creation of each gene of the child chromosome is performed through randomly selecting the corresponding gene from one of its parents. Both parents have an equal chance for contributing in the creation of the chromosomes that are produced from them. The control parameter Θ_x determines the crossover rate.

The best solution (in terms of maximum Φ), $\mathbb{P}^*$ of the final generation of GA can be selected as the result to the maximization problem, which is formally specified as:

$$\mathbb{P}^*: \Leftrightarrow \nexists \mathbb{P} \in \rho_{i_{max}} | \Phi(X_{\mathbb{P}}) > \Phi(X_{\mathbb{P}^*}) \quad (12)$$

Though, the phenotype of the best solution $\mathbb{P}^*$ may still suffer from violating the length constraint:

$$\sum_{i=1}^{n_{candidate}} l_i x_i > L \quad (13)$$

6 Experimental results

6.1 Requirements and parameter setting

The proposed system has been coded in C# and the environment is Microsoft visual studio ultimate 2013. The experiments were executed on a THINK-PC Lenovo z5170 with Intel core i7-5500 CPU 2.4GHz and a Memory of 8 GB RAM, HDD: 1TB and Video card: AMD Radeon 4GB. GA's parameters have been set as follows: a population of pop_{size}=50 individuals is used and evolved over a sequence of $iter_{max}$=100. For the tournament selection, a tournament size equals to 2 has been chosen. Crossover probability and mutation probability are set to p_c=0.7 and p_m=0.1, respectively. The overlapping parameter k used for applying Dice and $Jaccard$ similarity has been set to 3.

Qualitative evaluations of the proposed two models were made quantitatively based on the multi-document summarization datasets provided by Document Understanding Conference DUC , particularly using $DUC2002$ dataset . A brief statistics of the dataset is given in Table-1. Like all other related works, the documents in DUC2002 dataset are, first, preprocessed as follows:

- Documents are segmented into individual sentences considering '.', '?', and '!' as delimiters. Identical sentences and sentences with 3 words or less are removed,
- Sentences are tokenized, tokens are lowercased and duplicate tokens are excluded.
- Punctuation marks are removed,
- Stop words are excluded and
- Finally, the remaining words are stemmed using Porter stemming algorithm [17] and the duplicate stems are removed.

Table 1-Description of the $DUC2002$ dataset.

Description	DUC2002 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 words

6.2 Evaluation metrics

The proposed work is quantitatively measured using Recall-Oriented Understudy for Gisting Evaluation $ROUGE$ evaluation metric. $ROUGE$ 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 a 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:

$$ROUGE - N = \frac{\sum_{S \in \{reference\ Summaries\}} \sum_{N\text{-}gram \in S} Count_{match}(N\text{-}gram)}{\sum_{S \in \{reference\ Summaries\}} \sum_{N\text{-}gram \in S} Count(N\text{-}gram)} \quad (14)$$

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, and $Count(N - gram)$ is the number of $N - grams$ in the reference summaries. For the work proposed in this paper, ROUGE-1 and ROUGE-2 have been used for evaluating the performance of the proposed system and for performance comparison with other states of the art methods.

6.3 System performance

Table-2 and its related figures record the average $ROUGE$ scores of the proposed model $\boldsymbol{GA_{\Phi}}$ wherein the similarity has been calculated using single metric similarity measures: Cosine, $Jaccard$ and Dice similarity, while the performance has been evaluated using $DUC2002$ dataset and represented by an average of 20 different runs with the same parameters.

Table 2 Average $Rouge-1$ and $Rouge-2$ scores resulted from applying GA_{Φ} using single similarity measures: Cosine, $Jaccard$ and Dice similarity and implemented on $DUC2002$ dataset .

Similarity measure	$\overline{Rouge2}$			$\overline{Rouge1}$		
	$\overline{R}$	$\overline{P}$	$\overline{F}$	$\overline{R}$	$\overline{P}$	$\overline{F}$
Cosine	0.14062	0.14605	**0.13855**	0.40455	0.39557	0.39968
Dice	0.12399	0.10886	0.11467	0.41960	0.39279	0.40319
Jaccard	0.12660	0.11739	0.12727	0.41626	0.40716	**0.40899**

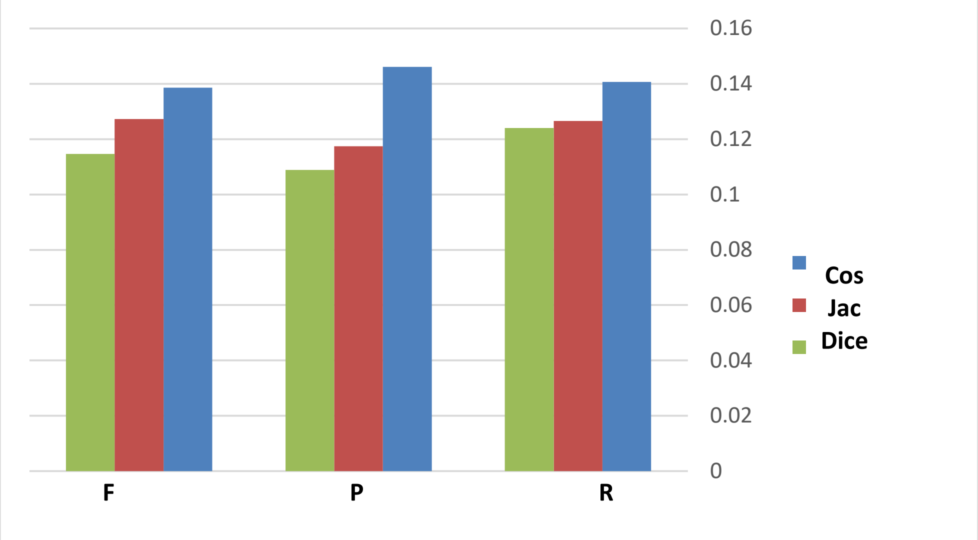

Figure 1a-Average Rouge-2 scores resulted from applying $\boldsymbol{GA_{\Phi}}$ using for measuring text similarity, single similarity measures: Cosine, $Jaccard$ and Dice similarity and implemented on DUC2002 dataset

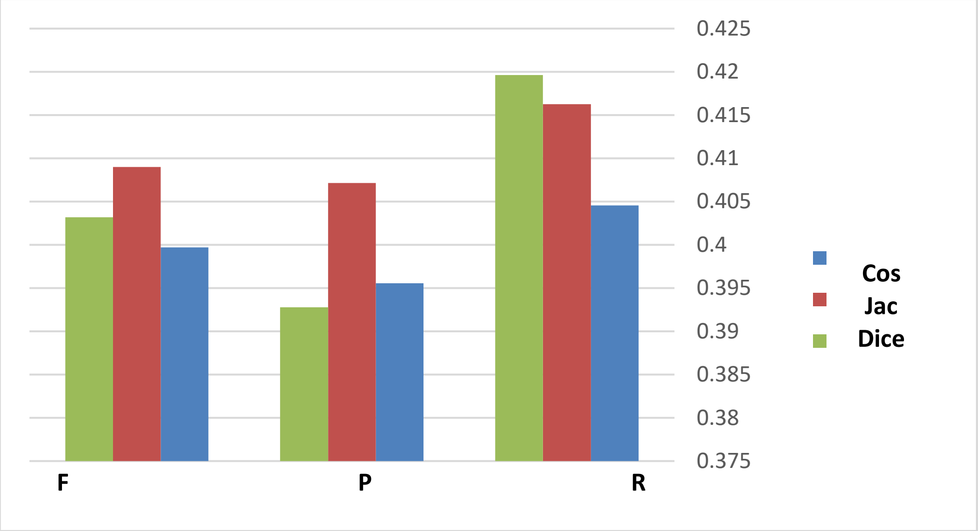

Figure 1b Average Rouge-1 scores resulted from applying $\boldsymbol{GA_{\Phi}}$ using for measuring text similarity, single metric similarity measures: Cosine, Jaccard and Dice similarity and implemented on DUC2002 dataset

Considering Table-2 and its related figures, it is obvious that the proposed system performs better using Cosine similarity for measuring text similarity in terms of Rouge-2, whereas better performance has been recorded in terms of Rouge-1 using $Jaccard$ similarity also for Dice similarity. Thus, these results encouraged us for introducing different integrations of these similarity measures and applying them for the proposed model in order to measure similarity to improve its performance.

Tables-(3, 4 and 5) and their related figures record the average $ROUGE$ scores of the proposed model $\boldsymbol{GA_{\Phi}}$ wherein the similarity has been calculated using double metric similarity measures generated from introducing different combinations regarding Cosine, $Jaccard$ and Dice similarity, while the performance has been evaluated using $DUC2002$ dataset and represented by an average of 20 different runs with the same parameters, taking into consideration the value of $\sigma = 0.1$ through 0.9 using step of 0.1. The summarized results shown in Table 6 are the highest scores recorded from applying the three integrations to the proposed model $\boldsymbol{GA_{\Phi}}$ in terms of Rouge-1 and Rooge-2. Values from 0.1 through 0.9 have been considered for σ.

Table 3-Average $Rouge - 1$ and $Rouge - 2$ scores resulted from applying GA_{Φ} using the integration of Cosine and Dice similarity measures through applying $(\sigma \times \text{CosSim} + (1 - \sigma) \times \text{Dice}Sim)$ and implemented on $DUC2002$ dataset.

σ	$\overline{Rouge2}$			$\overline{Rouge1}$		
	$\overline{R}$	$\overline{P}$	$\overline{F}$	$\overline{R}$	$\overline{P}$	$\overline{F}$
0.1	0.1249	0.1298	0.1215	0.3976	0.4093	**0.4001**
0.2	0.1285	0.1236	**0.1252**	0.3834	0.38878	0.3833
0.3	0.1265	0.1215	0.1232	0.3895	0.3999	0.3922
0.4	0.1061	0.0786	0.1039	0.3760	0.3954	0.3782
0.5	0.1199	0.1061	0.1091	0.3792	0.38321	0.3765
0.6	0.1018	0.1002	0.1001	0.3722	0.3845	0.3757
0.7	0.0902	0.1108	0.0915	0.3477	0.3838	0.3600
0.8	0.0954	0.0952	0.0939	0.3692	0.3823	0.3708
0.9	0.0981	0.1042	0.0997	0.3509	0.3867	0.3631

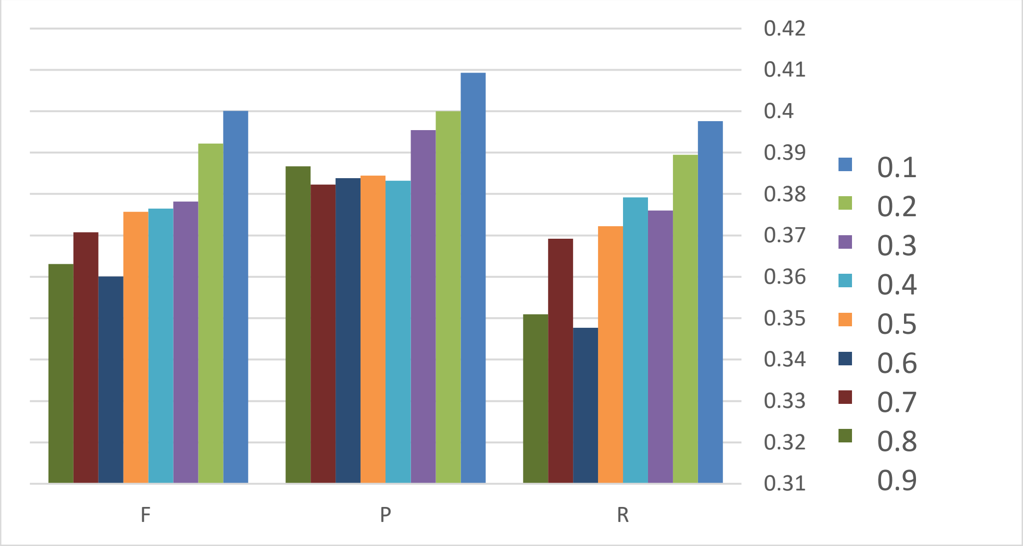

Figure 2a-Average $Rouge - 1$ scores resulted from applying GA_{Φ} using the integration of Cosine and Dice similarity measures and implemented on $DUC2002$ dataset.

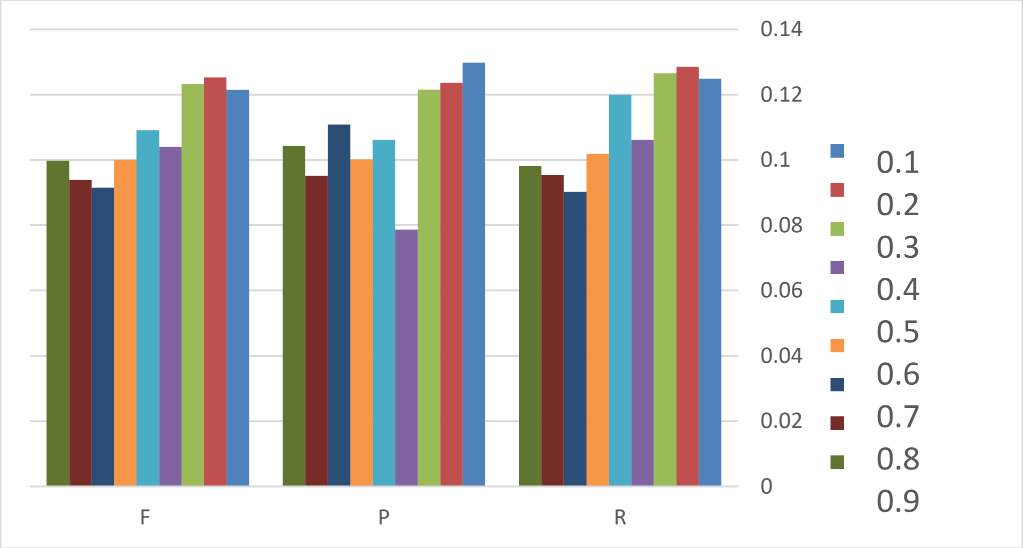

Figure 2b Average $Rouge-2$ scores resulted from applying GA_{Φ} using the integration of Cosine and Dice similarity measures and implemented on $DUC2002$ dataset.

Table 4-Average $Rouge-1$ and $Rouge-2$ scores resulted from applying GA_{Φ} using the integration of Cosine and $Jaccard$ similarity measures through applying (σ×CosSim+(1 − σ) × JacSim) and implemented on $DUC2002$ dataset.

σ	$\overline{\boldsymbol{Rouge2}}$			$\overline{\boldsymbol{Rouge1}}$		
	$\overline{\boldsymbol{R}}$	$\overline{\boldsymbol{P}}$	$\overline{\boldsymbol{F}}$	$\overline{\boldsymbol{R}}$	$\overline{\boldsymbol{P}}$	$\overline{\boldsymbol{F}}$
0.1	0.0978	0.0981	0.0969	0.3570	0.3849	0.3665
0.2	0.1344	0.1381	**0.1354**	0.3991	0.4210	**0.4075**
0.3	0.1123	0.1094	0.1097	0.3783	0.3887	0.3793
0.4	0.0978	0.0795	0.0955	0.3937	0.3915	0.3874
0.5	0.1066	0.1062	0.1053	0.3739	0.3906	0.3779
0.6	0.1142	0.1128	0.1113	0.3857	0.3933	0.3817
0.7	0.1033	0.1036	0.1026	0.3730	0.3919	0.3805
0.8	0.0993	0.1059	0.0986	0.3584	0.3982	0.3653
0.9	0.1112	0.1047	0.1069	0.3755	0.3838	0.3765

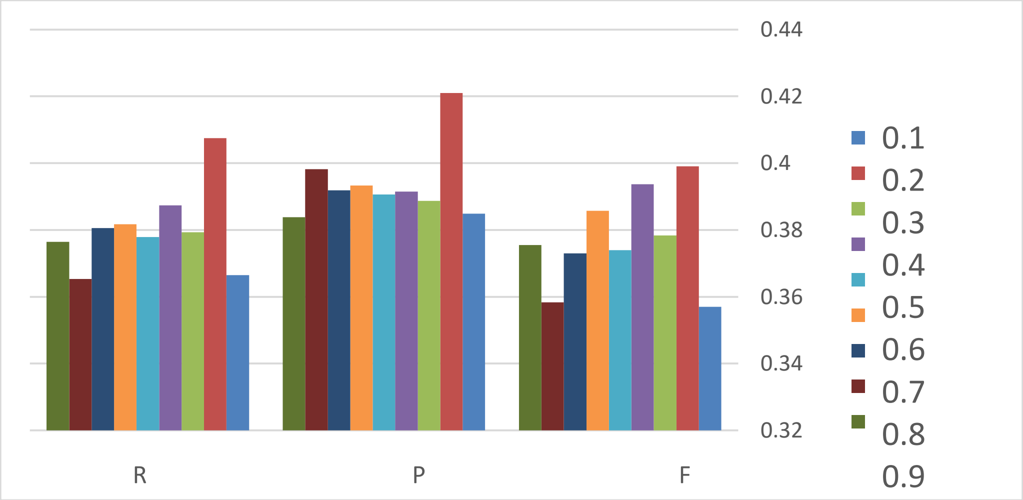

Figure 3a-Average $Rouge-1$ scores resulted from applying GA_{Φ} using the integration of Cosine and $Jaccard$ similarity measures and implemented on $DUC2002$ dataset.

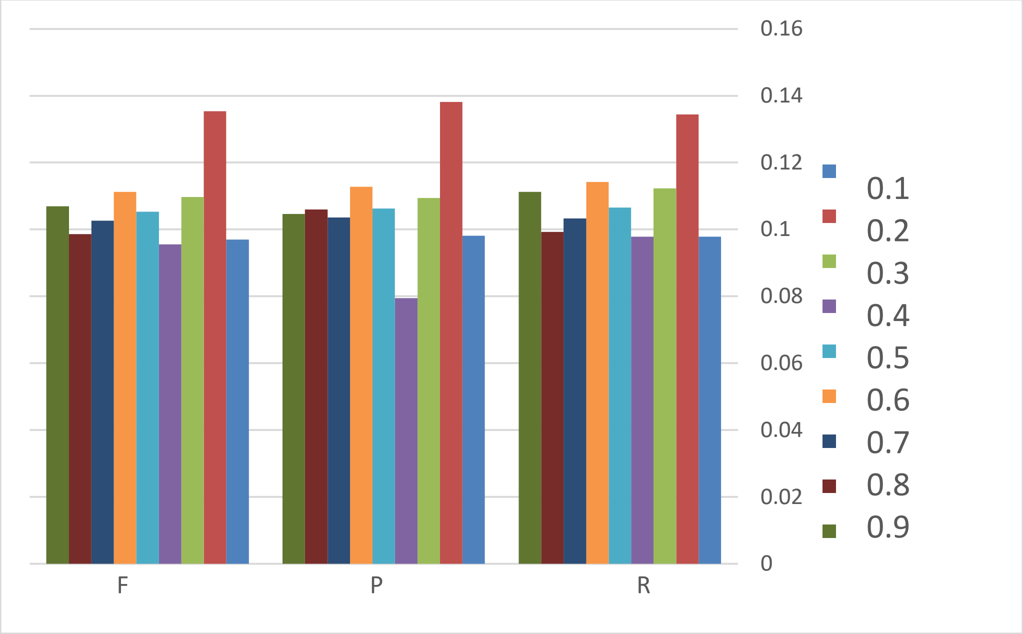

Figure 3b-Average $Rouge-2$ scores resulted from applying GA_{Φ} using the integration of Cosine and $Jaccard$ similarity measures and implemented on $DUC2002$ dataset.

Table 5-Average $Rouge-1$ and $Rouge-2$ scores resulted from applying GA_{Φ} using the integration of $Jaccard$ and Dice similarity measures through applying $(\sigma \times \text{DiceSim} + (1-\sigma) \times \text{Jac}Sim)$ $and\ implemented\ on\ DUC$2002 dataset.

σ	$\overline{Rouge2}$			$\overline{Rouge1}$		
	$\overline{R}$	$\overline{P}$	$\overline{F}$	$\overline{R}$	$\overline{P}$	$\overline{F}$
0.1	0.1364	0.1216	**0.1281**	0.4325	0.4199	**0.4244**
0.2	0.1261	0.1171	0.1206	0.4161	0.4218	0.4130
0.3	0.1327	0.1216	0.1264	0.4168	0.4083	0.4112
0.4	0.1175	0.0871	0.1123	0.4091	0.4047	0.4053
0.5	0.1162	0.1073	0.1110	0.3965	0.3962	0.3941
0.6	0.1279	0.1153	0.1210	0.4184	0.4060	0.4121
0.7	0.1219	0.1148	0.1164	0.4070	0.4020	0.4029
0.8	0.1219	0.1112	0.1159	0.4075	0.4012	0.4026
0.9	0.1153	0.1074	0.1109	0.4070	0.4008	0.4028

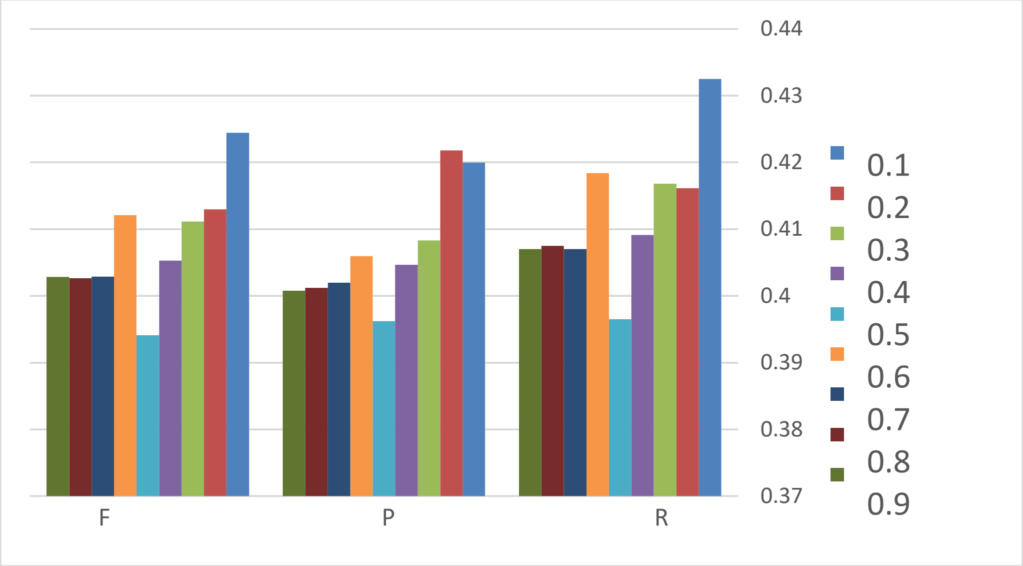

Figure 4a-Average $Rouge - 1$ scores resulted from applying GA_{Φ} using the integration of $Jaccard$ and Dice similarity measures and implemented on $DUC2002$ dataset.

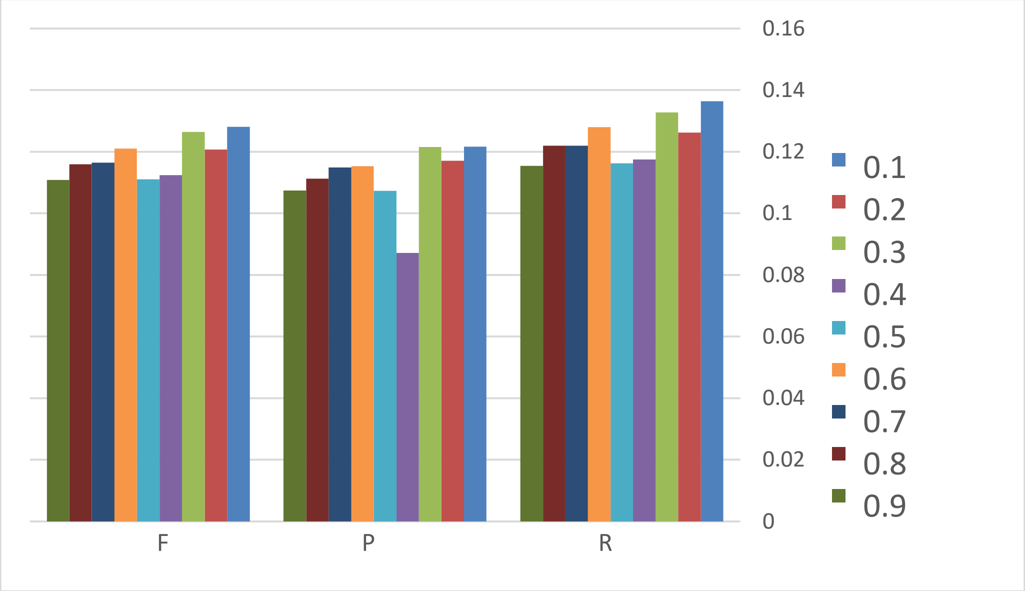

Figure 4b Average $Rouge - 2$ scores resulted from applying GA_{Φ} using the integration of $Jaccard$ and Dice similarity measures and implemented on $DUC2002$ dataset.

Table 6-Summarized results for performance evaluation for the proposed system through applying double similarity integration implemented on DUC2002.

Similarity integration	σ	$\overline{\text{Rouge1}}$			$\overline{\text{Rouge2}}$		
		$\overline{\text{R}}$	$\overline{\text{P}}$	$\overline{\text{F}}$	$\overline{\text{R}}$	$\overline{\text{P}}$	$\overline{\text{F}}$
σ × Cos + (1 − σ) × Dice	0.2	0.3834	0.3888	0.3833	0.1285	0.1236	0.1252
σ × Cos + (1 − σ) × Jac	0.2	0.3990	**0.4210**	0.4075	0.1344	0.1381	**0.1354**
σ × Dice + (1 − σ) × Jac	0.1	0.4325	0.4199	0.4244	0.1364	0.1216	0.1280

Regarding Table-6, it is obvious that the best performance of the proposed system has been achieved through using the integration σ × Cos + (1 − σ) × Dice by assigning the value of (σ = 0.2). Also, through a value of 0.2 for σ, when the integration σ × Cos + (1 − σ) × Jac has been applied for the proposed system, the system has recorded the best performance. Whereas for the integration $\sigma \times \text{Dice} + (1-\sigma) \times \text{Jac}$, when σ is set to 0.1, the best performance has been recorded for the proposed system.

The detailed results recorded in Tables 3 through 5 for evaluating the performance of the proposed model using different integrations of double similarity measures clarify the positive impact of measuring similarity between texts through the integration of more than one similarity measure against single similarity measure, wherein the proposed model recorded higher performance using $GA_{\Phi}^{DoubleSim}$ compared to $GA_{\Phi}^{SingleSim}$ at all $Rouge$ scores.

7 Conclusions

Automatic text summarization system has the challenge of producing high quality summary. In this paper, the design of a generic text summarization model based on sentence extraction was redirected into more semantic measure reflecting individually the two significant objectives: content coverage and diversity when generating summaries from multiple documents as an explicit optimization model. The proposed two models have been then coupled and defined as a single-objective optimization problem. Also, different integrations of similarity measures have been introduced and applied to the proposed model in addition to the single similarity measures for measuring text similarity involving double similarity measures integration.

Positive impact has been shown through applying different integrations of similarity measures for measuring similarity in the proposed SOEA-based model. When a single similarity measure represented by Cosine, $Jaccard$ or Dice similarity was applied for the proposed SOO model to measure text similarity and the performance evaluated, it was noticed that the proposed system has performed well in either Rouge-1 or Rouge-2. Whereas applying an integration of two similarity measures has improved the performance in terms of both Rouge-1 and Rouge-2.

The proposed work may be Extended or extra improvements may be added to it through a number of ways represented by the directions recorded in what follows:
Improving the tasks of the preprocessing phase has 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. Also as a future work, 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. Moreover, 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. Rasim M. Alguliev, Ramiz M. Aliguliyev, And Chingiz A. Mehdiyev. **2013**. An Optimization Approach to Automatic Generic Document Summarization. *Computational Inteligence*, **29**(1):129-155.
2. Asad A., Norisma I., Rasim M., Ramiz M. **2015**. Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems. *Information Processing & Management*, **51**(4): 340-358.
3. Rasim M. Alguliev, Ramiz M. Aliguliyev, Chingiz A. **2011**. An Optimization Model and DPSO-EDA for Document Summarization. *I.J. Information Technology and Computer Science*, **5**: 59-68.
4. Radev, D., Jing, H., Stys, M. and Tam, D. **2004**. Centroid-based summarization of multiple documents, *Information Processing & Management*, **40**(6): 919–938.
5. Rasim M Alguliev, Ramiz M Aliguliyev, Makrufa S Hajirahimova, Chingiz A. **2011**. MCMR: maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications*, **38**(12):14514-14522.
6. Rasim M Alguliev, Ramiz M Aliguliyev, Nijat R **2013**. Formulation of document summarization as a 0-1 nonlinear programming problem *Computers & Industrial Engineering,* **64**(1): 94-102.
7. Rasim M Alguliev, Ramiz M Aliguliyev, Nijat R. **2015**. An unsupervised approach to generating generic summaries of documents. *Applied Soft Computing*, **34**(C): 236-250.

8. Rasmita R. and Rakesh Ch. **2017**. Cat swarm optimization based evolutionary framework for multi document summarization. *Physica A: Statistical Mechanics and its Applications,* 477(1): 174-186.

9. Rasim M., Ramiz M., Nijat R., Asad A. and Norisma I. **2019**. COSUM: Text summarization based on clustering and optimization. *Expert Systems,* **36**(1): e12340.

10. Islam, A. and Inkpen, D. **2008**. Semantic text similarity using corpus-based word similarity and string similarity, *ACM Transactions on Knowledge Discovery from Data*, **2**(2): Article 10, 25 p.

11. Salton, G. and Buckley, C. **1988**. Term-weighting approaches in automatic text retrieval, *Information Processing & Management*, **25**(5): 513–523

12. Rada M., Courtney C. and Carlo S. **2006**. Corpus-based and Knowledge-based Measures of Text Semantic Similarity.

13. Anna H. **2008**. Similarity Measures for Text Document Clustering. *Mathematics,*

14. Amit S. **2001**. Modern Information Retrieval: A Brief Overview.

15. Pang-Ning; Steinbach, Michael; Kumar, Vipin **2005**. Introduction to Data Mining. 1st Ed, Addison Wesley.

16. Document understanding conference: http://duc.nist.gov.

17. Porter stemming algorithm: http://www.tartarus.org/martin/PorterStemmer/.

18. Lin, C.-Y. **2004**. ROUGE: a package for automatic evaluation summaries, in: Proceedings of the Workshop on Text Summarization Branches Out, Barcelona, Spain, July 25–26, pp. 74–81.