Heuristic Initialization And Similarity Integration Based Model for Improving Extractive Multi-Document Summarization

Nasreen J. Kadhim\(^1\), Dheyaa Abdulameer Mohammed\(^2\).

\(^1\)Computer Science Department, College of Science, University of Baghdad, Baghdad, Iraq.

\(^2\)Imam Ja'afar Al-Sadiq University, Baghdad, Iraq.

Email : \(^1\)nasreen_jawad@yahoo.com, \(^2\)thyaee@gmail.com

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 high quality summary. In this paper, the design of generic text summarization model based on sentence extraction has been redirected into more semantic measure reflecting 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 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 measure that bases on using Cosine, Dice and Jaccard similarity measures for measuring text similarity involving integrating double similarity measures and triple similarity measures. The proposed optimization model has been solved using Genetic Algorithm. Moreover, heuristic initialization has been proposed and injected into the adopted evolutionary algorithm to harness its strength. Document sets supplied by Document Understanding Conference 2002 (DUC2002) have been used for the proposed system as an evaluation dataset and as an evaluation metric, Recall-Oriented Understudy for Gisting Evaluation (ROUGE) toolkit has been used for performance evaluation of the proposed method and for performance comparison against other baseline systems. Comparison results for the proposed optimization based model against other baselines verified that the proposed system outperforms other baseline approaches in terms of Rouge – 2 and Rouge – 1 scores wherein it has recorded a score of 0.4542 for Rouge – 1 and 0.1623 for Rouge – 2.

Keywords : Heuristic Initialization, integrations of similarity measures, Gisting Evaluation (ROUGE), optimization based model
I. 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 method for shortening a huge amount of information into a summarizing form by the process of selecting significant information and discarding insignificant and redundant information. The area of automatic text summarization is becoming very important in the field of Information Retrieval with the amount of textual information existing in the world wide web.

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 definite time frame and to read through all the documents is simply difficult. Availability of the core of the document makes the process speeds up considerably. When dealing with problems like that, the technology of automatic text summarization becomes critical.

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 [VII].

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 [I].

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, describe 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 complete document, retaining significant information, while decreasing it's volume. An informative summary is normally 20–30% of the original text [VIII].

The chief contribution of this paper is modeling multi-document text summarization task as an optimization problem. The proposed model emphasizes the detection of important sentences that satisfies coverage for the main topic of the document collection while minimizing the occurrence of redundant sentences. Different integrations of double and triple similarity measures have been 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. Moreover, heuristic initialization has been proposed and injected into the adopted evolutionary algorithm to harness its strength.

This paper is organized as follows. Section 2 presents the works related to the work proposed in this paper. Elementary concepts for extractive multi-document text summarization together with the statement of the problem are introduced in section 3. Section 4 presents the proposed mathematical formulation and modeling in detail. The proposed genetic algorithm for solving the optimization problem is introduced in section 5. The results of the performed experiments illustrating the performance evaluation for the proposed method are presented in Section 6 in addition to the performance comparison with other baselines. Finally, conclusions and some possible extensions to the current work are given in Section 7.

II. Related works

In literature, multi-document summarization approaches varies in their essence. Various extraction-based techniques have been proposed for generic text summarization [IX]. 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 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 DE 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 [X].

ALGULIEV et al. (2011) proposed an unsupervised model for TS 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 [XI].

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 [XII].

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 used for selecting significant sentences from 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[XIII].

ALGULIEV et al (2015) presented an unsupervised optimization based method for automatically summarizing text. They modeled text summarization 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[XIV].

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 [II].

Saleh et. Al. model the multi document text summarization task as a discrete optimization problem. The proposed model emphasized the discovery of essential sentences that cover the main topic of the document collection while transcending the occurrence of redundant sentences. A binary encoded genetic algorithm together with heuristic mutation and local repair operators were proposed to handle the modeled
optimization problem excrement were applied to DUC2002 dataset result clarified the effectiveness of the proposed model when compared with other baselines.

In Rasmita Rautray, Rakesh Chandra 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 based summarizer such as Harmony Search (HS) based summarizer and Particle Swarm Optimization (PSO) based summarizer [XV].

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 new mutation approach [XVI].

III. Extractive generic multi-document text summarization

III.i. Preliminaries

For measuring similarity between texts, several approaches have been explored, however, they are concentrated around four major categories. These are corpus-based methods, word co-occurrence/vector-based methods, descriptive feature-based methods and hybrid methods [V].

The commonly used approaches in text summarization are vector-based methods [14]. Assume $T = \{t_1, t_2, t_3, ..., t_m\}$ to be a vector that represents the m distinct terms in a document collection $\mathbb{D}$. 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 [14]. 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 thing, used in the same context [XVII].

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 [III]:

$$cosine_similarity(s_i, s_j) = \frac{\sum_{k=1}^{m} w_{ik} \cdot w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2} \cdot \sqrt{\sum_{k=1}^{m} w_{jk}^2}}$$
where:

\( w_{jk} = t_{jk} \times \text{isf}, \)  

\( t_{jk} \) measures how frequently a term \( t_k \) occurs in a sentence \( s_i \), and

\( \text{isf} = \log(n/n_k) \) measures how few sentences \( n_k \) comprise the term \( t_k \).

Automatically, if a term \( t_k \) does not exist in sentence \( s_i \), \( w_{jk} \) should be zero.

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 angle between them. So, if the inner product is used for finding the distance between two vectors, the cosine is used for finding the angle between these vectors. Using cosine similarity is a good technique for ranking document through discovering the closest document to the user query [IV].

\[
\text{sim} (S_i, S_j) = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_{jk}^2}} i, j = 1,2,3 \ldots, n
\]  

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

\[
J(S_i, S_j) = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} = \frac{|S_i| - |S_i - S_j|}{|S_i| + |S_j| - |S_i \cap S_j|} i, j = 1,2 \ldots, n
\]  

Dice Similarity is similar to Jaccard used for finding the similarity between two vectors but “gives twice the weight to agreements” [VI].

\[
D(S_i, S_j) = \frac{2|S_i \cap S_j|}{|S_i| + |S_j|} i, j = 1,2,3 \ldots, n
\]

III.ii. 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 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 \( \mathbb{D} \subset \mathbb{D} \) while tackling three challenges:

- **Covering Contents**: the generated summary \( \mathbb{D} \) should cover the main topic of the collection \( \mathbb{D} \).
- **Reducing Redundancy**: the created summary \( \mathbb{D} \) should not involve similar sentences contained in \( \mathbb{D} \).
- **Bounded length**: length of the summary \( \mathbb{D} \) should be restricted.
IV. The proposed model: definitions and formulations

In this paper, text summarization problem is addressed as a single objective optimization problem. The intended summary \( D \) is projected in the light of the defined problem as in the definitions of the proposed SOO based model GA\( _D \), introduced in what follows.

Definition 1 (Summary \( D \)). Let \( s_\ell \in D \) be a sentence to be involved in \( D \), then the content coverage, stated by the summation of similarity for each pair of sentences: \( \text{sim}(s_\ell, O) \) between \( s_\ell \) and the set of sentences in the document collection \( D \) (represented by its mean vector \( O \)) and \( \text{sim}(s_\ell, O) \) between \( s_\ell \) and the set of sentences in the document collection \( D \) should be maximized. Alternatively, reduction of redundancy, or quantitatively, the similarity \( \text{sim}(s_i, s_j) \) between the same pair of sentences belong to \( D \) 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 GA\( _D \)). Let \( x_\ell \in \{0,1\} \) be a binary decision variable that denotes the absence (0) or presence (1) of the sentence \( s_\ell \) in \( D \) (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 \( D \) (Equation 6). Currently, let \( X = \{x_\ell|1 \leq i \leq n\} \) 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)

\[
x_\ell = \begin{cases} 1 & \text{if } s_\ell \in D \\ 0 & \text{otherwise} \end{cases}, \quad (5)
\]

\[
x_{ij} = \begin{cases} 1 & \text{if } s_i \text{ and } s_j \in D \\ 0 & \text{otherwise} \end{cases} \quad (6)
\]

Maximize \( G_{\text{GA}_D}(X) = \sum_{\ell=1}^{n-1} \sum_{j=i+1}^{n} \left( \frac{\text{sim}(s_\ell, O) + \text{sim}(s_j, O)}{\text{sim}(s_i, s_j)} \right) x_{ij} \quad (7) \)

subject to \( \sum_{i=1}^{n_{\text{candidate}}} l_i x_\ell \leq L \quad (8) \)

where:

- \( L \): Summary length constraint,
- \( l_i \): Length of sentence \( s_i \),
- \( n_{\text{candidate}} \): Total number of sentences in candidate summary,
- \( O \): Center of the document collection \( D = \{s_1, s_2, ..., s_n\} \).

Copyright reserved © J. Mech. Cont.& Math. Sci.
Nasreen J. Kadhim et al
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.

IV.i. The proposed similarity integrations

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

- **Single metric similarity measures**: These metrics measure the similarity between pair of sentences and between sentence and center of document collection through implementing individually the Eqs. (2, 3, 4) For Cosine, Jaccard and Dice similarity measures respectively.

$$\text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_1} = \sigma \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_2} + (1 - \sigma) \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_3}$$ (9)

- **Double metrics similarity measures**: These metrics measure the similarity between pair of sentences and between sentence and 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).

$$\text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_1 + \text{metric}_2} = \sigma \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_1} + (1 - \sigma) \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_2}$$ (10)

- **Triple metrics similarity measures**: These metrics measure the similarity between pair of sentences and between sentence and center of document collection through implementing formulas that are considered as weighted sum equations of the best score recorded for double similarity measures (Cosine and Jaccard), (Cosine and Dice) and (Jaccard and Dice) and the remaining similarity measure under consideration:

$$\text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_1 + \text{metric}_2 + \text{metric}_3} = \sigma \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_1} + (1 - \sigma) \times \text{sim}(\text{text}_i, \text{text}_j)_{\text{metric}_2}$$ (11)

V. 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 $\delta$ for the proposed GA can be calculated by the Cartesian product of existence/nonexistence of all $n$ sentences:

$$\delta = \prod_{i=1}^{n} (\{0,1\}) = 2^n$$ (11)

Consider a population $\rho$ of $\delta$ genotype solutions, $\mathbb{P}_{1 \leq k \leq K} \in \rho$. Then, $\forall k \in \{1, ..., K\}$ and $\forall j \in \{1, ..., n\}; \mathbb{P}_k = (\mathbb{P}_{k1}, \mathbb{P}_{k2}, ..., \mathbb{P}_{kn})$ 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 $\Psi: \rho \rightarrow \rho'$ with $\Psi(\rho_i) = \rho_{i+1}$, where $\rho_i$ is the population at iteration $i$. The population begins with an initial population $\rho_0$ formed as in the proposed heuristic.
initialization illustrated in Eq.(11) and carries on until a maximum number of iterations $i_{\text{max}}$ is got. Formally speaking:

$$\forall i \in \{1, ..., n\}: \rho_0 = (\rho_{01}, \rho_{02}, ..., \rho_{0n})$$

s.t. $\rho_0$ is initiated through applying the formula:

$$\rho_{0i} = \sigma \times \text{initialization}_{\text{random}} + (1 - \sigma) \times \text{initialization}_{\text{heuristic}}$$

Wherein for gene $j$ that corresponds to $s_j \in \rho_{0i}$,

$$j \in \{0, 1\} \text{ initialisation}_{\text{random}}$$

$$\text{sim}(s_j, \Theta) \text{ initialisation}_{\text{heuristic}}$$

The evolution function $\Psi$ 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 noted as:

$$\Psi = s_{\Theta_s} \circ x_{\Theta_x} \circ p_{\Theta_p}$$  \hspace{1cm} (12)$$

Through the application of the selection operator, $s_{\Theta_s}$, 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 $\Theta_s$ 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 $\Theta_x$ determines the crossover rate.

The best solution (in terms of maximum $\Phi$), $P^*$, of the final generation of GA can be selected as the result to the maximization problem. Formally specified as:

$$P^*: \iff \not\exists P \in \rho_{i_{\text{max}}} \mid \Phi(X_P) > \Phi(X_{P^*})$$  \hspace{1cm} (13)$$

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

$$\sum_{i=1}^{n_{\text{candidate}}} |X_i| > L$$  \hspace{1cm} (14)$$
VI. Experimental results

Vi.i. Requirements and parameter setting

The proposed system has been coded in C# and the environment is Microsoft visual studio ultimate 2013. The experiments have been achieved 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 \( \text{pop}_{\text{size}} = 50 \) individuals is used and evolved over a sequence of \( \text{iter}_{\text{max}} = 100 \). For the tournament selection, a tournament size of 2 has been selected. Mutation probability and crossover probability are set to \( p_m = 0.1 \) and \( p_c = 0.7 \), respectively. The overlapping parameter \( k \) used for applying Dice and Jaccard similarity has been set to 3.

VI.ii. DUC 2002 dataset

- Qualitative evaluation of the proposed model has been performed quantitatively based on the multi-document summarization datasets supplied by Document Understanding Conference DUC, mainly by DUC2002 dataset [19]. A brief statistic of the dataset is given in Table-1. Firstly, as in all works working on text summarization, the preprocessing step is performed on document collection as in what 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
  - Lastly, Porter stemming algorithm [20] is applied for the remaining words also the duplicate stems are removed.

| Description                          | DUC2002 dataset                  |
|--------------------------------------|----------------------------------|
| Number of topics                     | 59 (d061j through d120i)        |
| Number of documents in each topic    | \( \sim 10 \)                   |
| Total number of documents            | 567                              |
| Data source                          | TREC                             |
| Summary length                       | 200 words                        |
VI.iii. Evaluation metrics

The performance of the model proposed in this paper is quantitatively evaluated using Recall-Oriented Understudy for Gisting Evaluation ROUGE [21]. ROUGE is reflected as the official evaluation metric for text summarization by DUC. It comprises measures wherein through them the quality of a summary generated by computer is automatically determined by a comparison performed between it and the summaries formed by human. The comparison is performed through calculating the number of overlapping units, such as word pairs, word sequences, and N grams between the summary produced by a machine and a set of reference summaries created by humans. ROUGE – N is an N gram Recall that counts the number of N grams matches between two summaries, and it is considered as follows [21]:

\[
\text{ROUGE} – N = \frac{\sum_{\text{reference summaries}} \sum_{N\text{-gram} \in s} \text{Count}_{\text{match}}(N\text{-gram})}{\sum_{\text{reference summaries}} \sum_{N\text{-gram} \in s} \text{Count}(N\text{-gram})} \tag{13}
\]

where N stands for the length of the N gram, Count_{\text{match}}(N\text{-gram}) is the maximum number of N grams co-occurring in candidate summary and the set of reference summaries. Count(N\text{-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 state of the art methods.

VI.iv. System performance

Table 2 records the average ROUGE scores of the proposed model GA\_\Phi wherein the similarity has been calculated using single metricsimilarity measures: Cosine, Jaccard and Dice similarity and the performance has been evaluated using DUC2002 dataset and represented by an average of 20 different runs with the same parameters.

| Similarity measure | Rouge2 | Rouge1 |
|--------------------|--------|--------|
|                    | R | P | F | R | P | F |
| Cosine             | 0.1406 | 0.1460 | 0.1386 | 0.4045 | 0.3956 | 0.3997 |
| Dice               | 0.1240 | 0.1089 | 0.1147 | 0.4196 | 0.3928 | 0.4032 |
| Jaccard            | 0.1266 | 0.1174 | 0.1273 | 0.4163 | 0.4072 | 0.4090 |

Considering Table 2, 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 as double and triple similarity metric measures and applying them for the proposed model in order to measure similarity.

**Table 3** records the average ROUGE scores of the proposed model $\text{GA}_{\Phi}$ wherein the similarity has been calculated using double metric similarity measures generated from introducing different combinations regarding Cosine, Jaccard and Dice similarity and the performance has been evaluated using DUC2002 dataset and represented by an average of 20 different runs with the same parameters and taking in consideration the value of $\sigma = 0.1$ through 0.9 using step of 0.1.

**Table 3** Average Rouge – 1 and Rouge – 2 scores resulted from applying $\text{GA}_{\Phi}$ using the integration of Cosine and Dice similarity measures and implemented on DUC2002 dataset.

| Sim integration equation | $\sigma$ | Rouge2 | Rouge1 |
|--------------------------|---------|--------|--------|
|                          |         | $R$    | $F$    | $F$    | $R$    | $F$    | $F$    |
| $\sigma \times \text{Cos} + (1 - \sigma) \times \text{Dice}$ | 0.2     | 0.3834 | 0.3888 | 0.3833 | 0.1285 | 0.1236 | 0.1252 |
| $\sigma \times \text{Cos} + (1 - \sigma) \times \text{Jac}$ | 0.2     | 0.3990 | 0.4210 | 0.4075 | 0.1344 | 0.1381 | 0.1354 |
| $\sigma \times \text{Dice} + (1 - \sigma) \times \text{Jac}$ | 0.1     | 0.4325 | 0.4199 | 0.4244 | 0.1364 | 0.1216 | 0.1280 |

**Tables 4** through 6 are the average ROUGE scores recorded for the proposed model $\text{GA}_{\Phi}$ wherein the similarity has been calculated using triple metric similarity measures generated from introducing different combinations based on Cosine, Jaccard and Dice similarity and the performance has been evaluated using DUC2002 dataset and represented by an average of 20 different runs with the same parameters. The value of $\sigma$ has been considered equal to 0.1 through 0.9 with a step of 0.1.

**Table 4** Average Rouge – 1 and Rouge – 2 scores resulted from applying $\text{GA}_{\Phi}$ using the integration: $\sigma \times \text{Jac} + (1 - \sigma) \times (\text{Cos} + \text{Dice})$ and implemented on DUC2002 dataset. Values of $\sigma$ have been considered 0.1 through 0.9 with a step of 0.1.
Table 5 Average Rouge − 1 and Rouge − 2 scores resulted from applying GAs using the integration: \( \sigma \times \text{Dice} + (1 - \sigma) \times (\text{Cos} + \text{Jac}) \) and implemented on DUC2002 dataset. Values of \( \sigma \) have been considered 0.1 through 0.9 with a step of 0.1.

| \( \sigma \) | Rouge1 | Rouge2 |
|------------|--------|--------|
| 0.1        | 0.4064 | 0.1270 |
| 0.2        | 0.4038 | 0.1269 |
| 0.3        | 0.4013 | 0.1267 |
| 0.4        | 0.3987 | 0.1264 |
| 0.5        | 0.3961 | 0.1262 |
| 0.6        | 0.3936 | 0.1260 |
| 0.7        | 0.3910 | 0.1258 |
| 0.8        | 0.3884 | 0.1256 |
| 0.9        | 0.3858 | 0.1254 |

Table 6 Average Rouge − 1 and Rouge − 2 scores resulted from applying GAs using the integration: \( \sigma \times \text{Cos} + (1 - \sigma) \times (\text{Jac} + \text{Dice}) \) and implemented on DUC2002 dataset. Values of \( \sigma \) have been considered 0.1 through 0.9 with a step of 0.1.

| \( \sigma \) | Rouge1 | Rouge2 |
|------------|--------|--------|
| 0.1        | 0.4036 | 0.1167 |
| 0.2        | 0.4040 | 0.1188 |
| 0.3        | 0.4045 | 0.1209 |
| 0.4        | 0.4049 | 0.1229 |
| 0.5        | 0.4053 | 0.1250 |
| 0.6        | 0.4058 | 0.1270 |
| 0.7        | 0.4062 | 0.1292 |
| 0.8        | 0.4066 | 0.1312 |
| 0.9        | 0.4070 | 0.1332 |
Tables 7 through 9 and their related figures are the average ROUGE scores recorded for the proposed model GA$_{\Phi}$ wherein the similarity has been calculated using triple metric similarity measures and injecting the heuristic initialization into the initialization step of the GA from 10% to wholly heuristic initialization and the performance has been evaluated using DUC2002 dataset and represented by an average of 20 different runs with the same parameters. The value of $\sigma$ has been considered equal to 0.1 through 0.9 with a step of 0.1.

Table 7 Average Rouge – 1 and Rouge – 2 scores resulted from applying GA$_{\Phi}$ using the integration: $\sigma \times \text{Cos} + (1 - \sigma) \times (\text{Jac} + \text{Dice})$ and implemented on DUC2002 dataset.

| Heuristic weight% | Random weight% | Rouge1 | Rouge2 |
|-------------------|----------------|--------|--------|
|                   |                | R      | F      | R      | F      | F      |
| 0                 | 0              | 0.15   | 1343   | 0.0    | 1417   | 0.0    | 470    | 0.4    | 266    | 0.4    | 356    | 0.4    |
| 10                | 90             | 0.14   | 1303   | 0.0    | 1367   | 0.0    | 429    | 0.4    | 188    | 0.4    | 304    | 0.4    |
| 20                | 80             | 0.16   | 1561   | 0.0    | 1623   | 0.0    | 664    | 0.4    | 372    | 0.4    | 542    | 0.4    |
| 30                | 70             | 0.15   | 1405   | 0.0    | 1443   | 0.0    | 524    | 0.4    | 322    | 0.4    | 415    | 0.4    |
| 40                | 60             | 0.14   | 1330   | 0.0    | 1382   | 0.0    | 412    | 0.4    | 203    | 0.4    | 299    | 0.4    |
| 50                | 50             | 0.13   | 1252   | 0.0    | 1314   | 0.0    | 387    | 0.4    | 239    | 0.4    | 280    | 0.4    |
| 60                | 40             | 0.14   | 1271   | 0.0    | 1343   | 0.0    | 411    | 0.4    | 166    | 0.4    | 284    | 0.4    |
| 70                | 30             | 0.13   | 1294   | 0.0    | 1324   | 0.0    | 344    | 0.4    | 233    | 0.4    | 293    | 0.4    |
| 80                | 20             | 0.14   | 1347   | 0.0    | 1425   | 0.0    | 489    | 0.4    | 238    | 0.4    | 359    | 0.4    |
| 90                | 10             | 0.14   | 1321   | 0.0    | 1365   | 0.0    | 433    | 0.4    | 248    | 0.4    | 329    | 0.4    |
| 0                 | 10             | 0.13   | 1213   | 0.0    | 1251   | 0.0    | 292    | 0.4    | 121    | 0.4    | 203    | 0.4    |
Table 8 Average Rouge − 1 and Rouge − 2 scores resulted from applying GA using the integration: $\sigma \times \text{Jac} + (1 - \sigma) \times (\text{Cos} + \text{Dice})$ and implemented on DUC2002 dataset.

| Heuristic weight% | Random weight% | Rouge 2 | Rouge 1 |
|-------------------|----------------|---------|---------|
|                   |                | $\bar{R}$ | $\bar{P}$ | $\bar{F}$ | $\bar{R}$ | $\bar{P}$ | $\bar{F}$ |
| 0                 | 100            | 0.1474  | 0.1311  | 0.1385  | 0.4422  | 0.4224  | 0.4312  |
| 10                | 90             | 0.1460  | 0.1324  | 0.1387  | 0.4449  | 0.4210  | 0.4340  |
| 20                | 80             | 0.1630  | 0.1499  | 0.1561  | 0.4620  | 0.4404  | 0.4498  |
| 30                | 70             | 0.1501  | 0.1380  | 0.1417  | 0.4475  | 0.4286  | 0.4378  |
| 40                | 60             | 0.1474  | 0.1357  | 0.1410  | 0.4429  | 0.4223  | 0.4316  |
| 50                | 50             | 0.1373  | 0.1248  | 0.1308  | 0.4397  | 0.4249  | 0.4310  |
| 60                | 40             | 0.1477  | 0.1332  | 0.1403  | 0.4462  | 0.4227  | 0.4344  |
| 70                | 30             | 0.1383  | 0.1312  | 0.1343  | 0.4359  | 0.4256  | 0.4314  |
| 80                | 20             | 0.1462  | 0.1315  | 0.1393  | 0.4462  | 0.4209  | 0.4330  |
| 90                | 10             | 0.1385  | 0.1265  | 0.1313  | 0.4368  | 0.4182  | 0.4263  |
| 100               | 0              | 0.1333  | 0.1243  | 0.1281  | 0.4314  | 0.4143  | 0.4227  |

Table 9 Average Rouge − 1 and Rouge − 2 scores resulted from applying GA using the integration: $\sigma \times \text{Dice} + (1 - \sigma) \times (\text{Cos} + \text{Jac})$ and implemented on DUC2002 dataset.

| Heuristic weight% | Random weight% | Rouge 2 | Rouge 1 |
|-------------------|----------------|---------|---------|
|                   |                | $\bar{R}$ | $\bar{P}$ | $\bar{F}$ | $\bar{R}$ | $\bar{P}$ | $\bar{F}$ |
| 0                 | 100            | 0.1472  | 0.1310  | 0.1383  | 0.4426  | 0.4232  | 0.4318  |
| 10                | 90             | 0.1476  | 0.1340  | 0.1404  | 0.4455  | 0.4216  | 0.4337  |
| 20                | 80             | 0.1620  | 0.1490  | 0.1551  | 0.4614  | 0.4391  | 0.4490  |
| 30                | 70             | 0.1512  | 0.1390  | 0.1431  | 0.4488  | 0.4306  | 0.4392  |
| 40                | 60             | 0.1506  | 0.1389  | 0.1442  | 0.4450  | 0.4249  | 0.4339  |
The detailed results recorded in Table 3 for the performance of the proposed model using different integrations for double similarity measures in addition to the Tables (4 through 6) for the performance of the proposed model using different integrations of the triple similarity measures with totally random initialization 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\textsuperscript{DoubleSim}\textsubscript{\textPhi} compared to GA\textsuperscript{SingleSim}\textsubscript{\textPhi} at all Rouge scores. Also, applying GA\textsuperscript{TripleSim}\textsubscript{\textPhi} to the proposed model for measuring similarity has recorded higher system performance than GA\textsuperscript{DoubleSim}\textsubscript{\textPhi} and GA\textsuperscript{SingleSim}\textsubscript{\textPhi}.

The study introduced in this thesis that takes in consideration the triple similarity measures for measuring similarity in the proposed model that studies the process of injecting the proposed heuristic initialization in the initialization step of the GA from 10% through totally initializing the population by heuristic (Tables 7 through 9) for solving the proposed model GA\textsubscript{\textPhi} clarifies that the injection of the proposed heuristic initialization to initialize 20% of the initial population and initializing 80% of the initial population randomly recorded the highest system performance. It is clarified that the proposed system has recorded the best performance considering for measuring similarity the integration of Cosine similarity with the combination of Jaccard and Dice similarity.

VI.v. Performance comparison

As mentioned earlier, evaluation metrics represented by Rouge − 1 and Rouge − 2 have been used for comparison purposes for comparing the performance of the proposed system and the baseline systems performance. The model proposed in [X] has been implemented and its performance has been evaluated to be compared with the method proposed in this paper. Comparison results reported in Tables 9 and 10 and their related figures in addition to the relative improvement percentage illustrated in Table 11 clarify that the proposed model outperforms model\textsuperscript{[5]} for modeling multi-document summarization in terms of Rouge − 1 and Rouge − 2 on DUC2002.
Table (10): Comparison results of the proposed system against model\(_5\) in terms of average Rouge-1 and Rouge-2 scores using triple similarity measure integration for measuring similarity without injecting heuristic initialization and applied on DUC2002 dataset.

| Similarity integration | Rouge # | \(\text{model}_5\) | Our model |
|------------------------|---------|-------------------|-----------|
| \(\sigma \times \cos + (1 - \sigma) \times (\text{Jac} + \text{Dice})\) | 2       | 0.1266 0.1301 0.1230 | 0.1506 0.1344 0.1417 |
|                        | 1       | 0.4055 0.3966 0.4001 | 0.4470 0.4266 0.4356 |
| \(\sigma \times \text{Jac} + (1 - \sigma) \times (\text{Cos} + \text{Dice})\) | 2       | 0.0966 0.1013 0.0986 | 0.14743 0.1312 0.1385 |
|                        | 1       | 0.3842 0.3954 0.3893 | 0.4422 0.4224 0.4312 |
| \(\sigma \times \text{Dise} + (1 - \sigma) \times (\text{Cos} + \text{Jac})\) | 2       | 0.1159 0.1323 0.1216 | 0.1472 0.1309 0.1383 |
|                        | 1       | 0.3473 0.4279 0.3757 | 0.4426 0.4232 0.4318 |

Fig.1 Performance comparison of the proposed system against model\(_5\) in terms of average Rouge-1 and Rouge-2 scores using triple similarity measure integration for measuring similarity without the injection of the heuristic initialization and applied on DUC2002 dataset.
Table (11): Comparison results of the proposed system against model\(^{[5]}\) in terms of average Rouge-1 and Rouge-2 scores using triple similarity measure integration for measuring similarity with the injection of the heuristic initialization and applied on DUC2002 dataset.

| Similarity integration | Rouge # | \(\text{model}^{[5]}\) | Our model |
|------------------------|--------|----------------|-----------|
| \(\sigma \times \text{Cos} + (1-\sigma) \times (\text{Jac} + \text{Dice})\) | 2      | \[
\begin{array}{c}
\text{R} \\
0.112 \\
0.100 \\
0.128
\end{array}
\] | \[
\begin{array}{c}
\text{R} \\
0.1257 \\
0.1103 \\
0.1300
\end{array}
\] | \[
\begin{array}{c}
\text{R} \\
0.1692 \\
0.1630 \\
0.1620
\end{array}
\] |
|                        | 1      | \[
\begin{array}{c}
\text{P} \\
0.1215 \\
0.1154 \\
0.1338
\end{array}
\] | \[
\begin{array}{c}
\text{P} \\
0.4023 \\
0.4088 \\
0.4248
\end{array}
\] | \[
\begin{array}{c}
\text{P} \\
0.1562 \\
0.1499 \\
0.4142
\end{array}
\] |
| \(\sigma \times \text{Jac} + (1-\sigma) \times (\text{Cos} + \text{Dice})\) | 2      | \[
\begin{array}{c}
\text{F} \\
0.1257 \\
0.1300 \\
0.1342
\end{array}
\] | \[
\begin{array}{c}
\text{F} \\
0.4023 \\
0.3992 \\
0.4142
\end{array}
\] | \[
\begin{array}{c}
\text{F} \\
0.1623 \\
0.1561 \\
0.1551
\end{array}
\] |
|                        | 1      | \[
\begin{array}{c}
\text{P} \\
0.1257 \\
0.1300 \\
0.1342
\end{array}
\] | \[
\begin{array}{c}
\text{P} \\
0.4023 \\
0.3992 \\
0.4142
\end{array}
\] | \[
\begin{array}{c}
\text{P} \\
0.1623 \\
0.1561 \\
0.1551
\end{array}
\] |
| \(\sigma \times \text{Dice} + (1-\sigma) \times (\text{Cos} + \text{Jac})\) | 2      | \[
\begin{array}{c}
\text{R} \\
0.05 \\
0.15
\end{array}
\] | \[
\begin{array}{c}
\text{R} \\
0.1215 \\
0.1257
\end{array}
\] | \[
\begin{array}{c}
\text{R} \\
0.1692 \\
0.1692
\end{array}
\] |
|                        | 1      | \[
\begin{array}{c}
\text{F} \\
0.1257 \\
0.1300 \\
0.1342
\end{array}
\] | \[
\begin{array}{c}
\text{F} \\
0.4023 \\
0.3992 \\
0.4142
\end{array}
\] | \[
\begin{array}{c}
\text{F} \\
0.1623 \\
0.1561 \\
0.1551
\end{array}
\] |

Fig.2. Performance comparison of the proposed system against model\(^{[5]}\) in terms of average Rouge-1 and Rouge-2 scores using triple similarity measure integration for measuring similarity with the injection of the heuristic initialization and applied on DUC2002 dataset.
Table (12): Relative improvement percentage satisfied through implementing the proposed system against model 5 in terms of average Rouge-1 and Rouge-2 scores using triple similarity measure integration for measuring similarity with and without injecting heuristic initialization and applied on DUC2002 dataset.

| Similarity measure integration | Relative improvement% with heuristic | Relative improvement% without heuristic |
|-------------------------------|--------------------------------------|----------------------------------------|
|                               | Rouge1 | Rouge2 | Rouge1 | Rouge2 |
| \(\sigma \times \cos + (1 - \sigma)\) \(\times (Jac + Dice)\) | +10.99 | +29.11 | +8.87 | +15.2 |
| \(\sigma \times Jac + (1 - \sigma)\) \(\times (Cos + Dice)\) | +12.67 | +41.52 | +10.76 | +40.46 |
| \(\sigma \times Dise + (1 - \sigma)\) \(\times (Cos + Jac)\) | +8.40 | +19.30 | +14.93 | +13.73 |

Considering the performance comparison results recorded in table 4.11 and its related figure, it is observed that the proposed system using triple metric similarity with fully random initialization for initializing the population of potential solutions outperforms in terms of Rouge-1 metric. For the proposed system that cores on using an integration of triple metric similarity measure and injecting the heuristic initialization in the initialization step, it is clarified that it outperforms the system and the Proposed triple similarity without heuristic in terms of Rouge-1 and Rouge-2 metric.

VII. Conclusions

In this paper, the design of generic text summarization model based on sentence extraction has been 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 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 and triple similarity measures. Moreover, heuristic initialization has been proposed and injected into the adopted evolutionary algorithm to harness its strength.

Performance evaluation an performance comparison for the proposed system in terms of Rouge – 1 and Rouge – 2 scores reveal that the proposed method has supported strong proof for the effectiveness of the proposed SOEA-based model over other state-of-the-art models. Also, positive impact has been shown through applying
different integrations of similarity measures for measuring similarity in the proposed SOEA-based model. Finally, positive impact of injecting heuristic initialization to aid the random initialization for initializing the population the evolutionary algorithm that has led to a reduction in the amount of time to reach to the required summary while producing a summary with high quality through preserving the significant sentences in it.

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 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.

- Applying the proposed system for the summarization of Arabic texts via working on preprocessing phase through considering the rules dedicated for segmentation, tokenization and stemming of texts in Arabic.

- Additional objectives can be taken in consideration by the proposed model. For instance, coherence and cohesion objectives are examples of such objectives to be optimized simultaneously in addition to the content coverage and redundancy reduction objectives.

References

I. Asad Abdi, Norisma Idris, Rasim M. Aliguliev, Ramiz M. Aliguliyev. (2015), Automatic summarization assessment through a combination of semantic and syntactic information for intelligent educational systems.

II. Asad Abdi, Norisma Idris, Rasim M Aliguliev, Ramiz M Aliguliyev. (2015), Asad Abdi, Norisma Idris, Rasim M Aliguliev, Ramiz M Aliguliyev

III. Anna Huang. (2008), Similarity Measures for Text Document Clustering.

IV. Amit Singhal. (2001), Modern Information Retrieval: A Brief Overview

V. 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.

VI. Pang-Ning; Steinbach, Michael; Kumar, Vipin (2005), Introduction to Data Mining.
VII. RASIM M. ALGULIEV, RAMIZ M. ALIGULIYEV, AND CHINGIZ A. MEHDIYEV. (2013), AN OPTIMIZATION APPROACH TO AUTOMATIC GENERIC DOCUMENT SUMMARIZATION.

VIII. Rasim M. Aliguliev, Ramiz M. Aliguliyev, Chingiz A. Mehdiyev. (2011), An Optimization Model and DPSO-EDA for Document Summarization

IX. Radev, D., Jing, H., Stys, M. and Tam, D. 2004. Centroid-based summarization of multiple documents, Information Processing & Management 40 (6) 919–938.

X. Rasim M Aliguliev, Ramiz M Aliguliyev, Chingiz A Mehdiyev. (2011), An optimization model and DPSO-EDA for document summarization.

XI. Rasim M Aliguliev, Ramiz M Aliguliyev, Makrufa S Hajirahimova, Chingiz A Mehdiyev. (2011), MCMR: maximum coverage and minimum redundant text summarization model

XII. Rasim M Aliguliev, Ramiz M Aliguliyev, Nijat R Isazade. (2013), Formulation of document summarization as a 0-1 nonlinear programming problem

XIII. Rasim M Aliguliev, Ramiz M Aliguliyev, Chingiz A Mehdiyev. (2013), An optimization approach to automatic generic document summarization

XIV. Rasim M Aliguliev, Ramiz M Aliguliyev, Nijat R Isazade. (2015), An unsupervised approach to generating generic summaries of documents

XV. Rasmita Rautray, Rakesh Chandra Balabantaray. (2017), Cat swarm optimization based evolutionary framework for multi document summarization

XVI. Rasim M Aliguliev, Ramiz M Aliguliyev, Nijat R Isazade, Asad Abdi, Norisma

XVII. Rada Mihalcea, Courtney Corley, Carlo Strapparava. (2006), Corpus-based and Knowledge-based Measures of Text Semantic Similarity.

XVIII. saleh et. Al. (2015), A genetic based optimization model for extractive multi dormant text summarization. Iraqi Journal of Science. 2015;56(2B):1489-98.