Expert Systems With Applications 171 (2021) 114555

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A novel extractive multi-document text summarization system using quantum-inspired genetic algorithm: MTSQIGA

Mohammad Mojrian, Seyed Abolghasem Mirroshandel *

Department of Computer Engineering, University of Guilan, Rasht, Iran

A R T I C L E I N F O

Keywords:
Extractive summarization
Multi-document summarization
Quantum-inspired genetic algorithm
Objective function
Self-adaptive rotation gate
Quantum measurement

A B S T R A C T

The explosive growth of textual data on the web and the problem of obtaining desired information through this enormous volume of data has led to a dramatic increase in demand for developing automatic text summarization systems. For this reason, this paper presents a novel multi-document text summarization approach, called MTSQIGA, which extracts salient sentences from source document collection to generate the summary. The proposed generic summarizer models extractive summarization as a binary optimization problem that applies a modified quantum-inspired genetic algorithm (QIGA) in its processing stage to find the best solution. Objective function of our approach plays an important role in optimizing linear combination of coverage, relevance, and redundancy factors which consists of six sentence scoring measures. To ensures the generation of a summary with predefined length limit, the presented QIGA employs a modified quantum measurement and a self-adaptive quantum rotation gate based on the quality and length of the summary. Evaluation of the proposed system was performed on DUC 2005 and 2007 benchmark datasets in terms of ROUGE standard measures. Comparison of MTSQIGA with existing state-of-the-art approaches for multi-document summarization shows superior performance of the proposed systems over other methods on both existing benchmark datasets. It also indicates promising efficiency of our proposed algorithm on applying quantum-inspired genetic algorithm to the text summarization tasks.

1. Introduction

With the growth of various fields of science and engineering, massive amount of information has been produced on the Internet in recent years. Accessing the required information through this massive amount of information is one of the most challenging tasks for researchers in this situation. For this reason, the tendency towards automatic summarization systems has grown over the last decade. The great deals of important information are still in the text format on the Internet. Nowadays, with the massive proliferation in the text information that we receive every day, text summarization system could be helpful in finding the most important contents of the text in a short time. Thus, automatic text summarization has become as a main part of many text processor tools such as decision support systems, search engines, and medical information processing systems. It is also an important topic in natural language processing field these days.

Automatic text summarization (ATS) is defined as the process of extracting important information from one or more text documents in order to produce a shorter version of the original document(s) for the needs of the user (or different users) that is performed by a computer program. As a short version of the input documents, a good summary should convey the most important content in the documents while keeping redundancy to the minimum (Mani, 1999). The main purpose of automatic text summarization systems is generating a summary of input text that delivers important content to the user in less time. Although many frameworks and algorithms have achieved improvement in many benchmarks or task-specific applications, it is still the challenging job to summarize texts automatically in an intelligent way.

In overall, automatic text summarization systems are categorized into several groups based on certain criteria (Gambhir & Gupta, 2017). A system can belong to different categories at once (Aries, Zegour, & Hidouci, 2019). Regarding the number of source documents, text summarization can be classified into single-document or multi-document. Text summarization systems can be put into two categories on the basis of their application: generic and query-oriented. Query-oriented summaries only includes contents related to the user need, while the generic one provides a general outline of the information presented in the source document(s). In another assortment, summarization can be

* Corresponding author.
E-mail addresses: mojrian@msc.guilan.ac.ir (M. Mojrian), mirroshandel@guilan.ac.ir (S.A. Mirroshandel).

https://doi.org/10.1016/j.eswa.2020.114555
Received 8 November 2019; Received in revised form 30 December 2020; Accepted 30 December 2020
Available online 5 January 2021
0957-4174/

done through supervised or unsupervised approaches. In a supervised system, training data is needed which helps the system to learn a model to extract important sentences from the source document(s). The need of such enormous amount of labeled data (human-made summaries) is one of the challenges in this kind of text summarization systems. On the contrary, unsupervised approaches are applied in summarization systems that do not require any training dataset. These systems can generate the final summary only by accessing test data. Based on the style of the output summary, there are two categories including indicative and informative summaries (Gambhir & Gupta, 2017). Extractive and abstractive summaries are one of the main taxonomies existing for summarization systems based on the way the final summary is generated. In extractive summarization, the purpose of the system is generating a summary that includes the most salient and relevant contents through selecting a set of important sentences from the original text. The extractive summarizer tries to select the most informative sentences of the source text using different methods and then forms the summary using the selected sentences without making any changes in them. However, the final summary obtained from abstractive approach is the result of changing the words and sentences of the original text. In other words, the concepts provided in the summary are taken from the original text, but the sentences are shown in a different form which are generated by applying some operations such as merging, increasing, and decreasing.

There are different challenges that harden the development of ATS. Alongside content selection, redundancy in text is one of the most important issues in this field which should be handled to generate a perfect summary. This problem becomes more obvious in multi-document summarization since the number of input documents to these kind of systems is much greater than in single-document one. A desirable summarization approach should be able to select those sentences of the text that have the least redundancy while represent the most important and informative content of the document collection.

The coherence and cohesion of the summary is another challenge in the field of automatic text summarization, specially in multi-document one, when it comes to the summary's readability. It means that the applied approach has to be able to organize the selected sentences in such a way that the generated summary has coherence and cohesion in both its structure and semantic which leads to an appropriate readability. Having extracted most important and related sentences from the source text, an extractive text summarization system should take a method into account to put the selected sentences into a proper order as well as not selecting sentences containing dangling anaphora references.

There are some other challenges that are related to the nature of language being processed. Different languages have their own grammatical and morphological structures, semantic rules and lexicon to form a text that makes them difficult to process. For instance, many words in English have several meanings that should be used based on the context. Co-reference units in the sentences such as pronouns are another language-related issues that are difficult to find in many cases. Furthermore, text is an unstructured data so that it can be written in different formats according to the source language. This issue poses lots of difficulties in preprocessing step of summarization. Therefore, developing a text summarization system that could address all of these challenges at the same time is the subject of most researches in this field of study these days.

In this paper, an extractive multi-document text summarization system, called MTSQIGA, is proposed which aims at satisfying the main purpose of summarization approaches, that is, providing the maximum coverage of presented content in source text and minimizing redundancy in the summary. In our system, that is generic and language-independent, multi-document text summarization is implemented as an optimization problem that is in line with the nature of identifying salient sentences of source documents to be included in the summary. The proposed summarizer employs an adapted version of quantum-inspired genetic algorithm (QIGA) to solve the optimization problem in which some operators of the canonical QIGA have been modified in order to be applicable in ATS. The applied QIGA ensures that the length of the selected sentences will not exceed the given summary length limit. To this end, quantum measurement and quantum rotation gate operators, that play an important role in our proposed approach, have been modified in such a way that they would handle the length of the final summary. Given the high impact of the objective function on the performance of evolutionary algorithms, the proposed QIGA has been developed with focus on defining an effective objective function. Thus, the objective function of the proposed algorithm is a combination statistical and cosine similarity based measures that needs to be maximized during the evolution. To the best of our knowledge, this is the first study in which a type of quantum-inspired evolutionary algorithms is applied in ATS field.

We evaluated the performance of MTSQIGA on the prevalent DUC 2005 and DUC 2007 benchmark datasets and compare it with some of the best approaches that have been proposed in recent years. These datesets have been chosen due to their challenging structure of source text in comparison with other DUC datasets. Although the original purpose of these datasets is evaluating query-oriented multi-document summarization methods, the evaluation results of our proposed generic summarizer demonstrates the effectiveness of MTSQIGA as well as promising performance of the adaptive quantum-inspired genetic algorithm in comparison with the canonical evolutionary algorithms in summarization field. As the results indicates, our system could outperform some of the widely used approaches that have recently been presented for multi-document text summarization task.

The remainder of this paper is organized as follows. In Section 2, a number of studies related to the ATS field and specifically extractive summarization are briefly discussed. Also, some of the proposed methods employing quantum-inspired genetic algorithm are reviewed at the end of this section. Section 3 provides a more detailed description of our proposed summarization system including its different stages as well as the operators and features of the adaptive QIGA. Section 4 is dedicated to the representation of the conducted experiments and the results obtained from these evaluations. This section also compares the performance of MTSQIGA with other state-of-the-art methods. Finally, we conclude the paper and present future work according to the research potentials available in ATS field in Section 5.

2. Related works

This section reviews some existing works proposed in the literature of extractive text summarization at first. Following that, a brief discussion about several studies which have applied quantum-inspired genetic algorithm is presented.

2.1. Extractive text summarization

Extractive summarization approaches attempt to generate the final summary by selecting a set of salient sentences from source document(s) that are most informative and relevant. To tackle this challenging task, a variety of techniques and methods have been applied in extractive summarizers so far. One of the first attempts at automatic text summarization field was done in 1985 which extracts important sentences based on word frequency (Luhn, 1958).

Statistical techniques are one of the most widely used methods in extractive summarization as they are independent of any language and do not require any additional linguistic knowledge. In a study, a weighted combination of ten statistical features is used to assign a certain score between "0" and "1" to each sentence of the source document (Fattah & Ren, 2009). Although essential statistical features can lead to adequate performance, applying them alone may not necessarily yield sufficient results in all cases. They are mostly used in combination with other sentence scoring measures. Ferreira et al. proposed an approach in which a combination of statistical and linguistic features including discourse relations is considered (Ferreira et al., 2014). Semantic based sentence scoring methods such as sentiment analysis along with statistical features have been recently applied in many summarization approaches to improve their performance (Mogren, Kågebäck, & Dubhashi, 2015; Yadav, Sharan, Kumar, & Biswas, 2016). A hybrid summarizer has been proposed in a work for Arabic which synthesizes statistical-based similarity measures with a semantic sentence scoring method (Alami, Adlouni, En-nahnahi, & Meknassi, 2018). The semantic method used in this approach is based on information from Arabic WordNet ontology. After that, PageRank algorithm is employed to compute the score of each sentence by aggregating the results obtained from different similarity measures. Wang et al. proposed a multi-document summarization framework which utilize sentence-level semantic analysis (SLSS) and symmetric non-negative matrix factorization (SNFM) techniques (Wang, Li, Zhu, & Ding, 2008). In the first step of this approach, pairwise sentence semantic similarity matrix is constructed based on both semantic role analysis and WordNet word relation. Following that, sentences are fallen into different clusters by applying symmetric non-negative matrix factorization technique. In each cluster, sentences are ranked using within-cluster sentence-to-sentence and sentence-to-topic similarities. Finally, the most important sentence of each cluster is identified to constitute the summary.

Graph-based approaches have been recently shown that they are able to achieve promising results in text summarization field. GRAPHSUM is a graph-based generic summarizer introduced for the purpose of multi-document text summarization (Baralis, Cagliero, Mahoto, & Fiori, 2013). It creates a correlation graph in which the nodes depicts different terms of the source document collection and each edge between a pair of nodes is associated with a weight representing the correlation between them. In order to extract correlation between different terms, GRAPHSUM applies association rule mining technique. After extracting frequent itemsets from transactional representation of the source document collection and generating correlation graph, PageRank algorithm is employed to select salient sentences based on the graph. In another study, an extractive multi-document summarization framework was proposed that applies archetypal analysis to select salient sentences (Canhasi & Kononenko, 2016). Constructing an undirected and weighted sentence similarity graph in the first stage, this approach uses archetypal analysis (AA) to perform sentence soft clustering into archetypes and sentence probabilistic ranking simultaneously. In the similarity graph, each vertex indicates one of the sentences and each edge is a representation of cosine similarity between corresponding vertices. The single-document summarizer of Mallick et al. is a graph-based approach that applied a modified TextRank algorithm to assign a score to each sentence of the source document (Mallick, Das, Dutta, Das, & Sarkar, 2019). In this method, the weighted similarity graph consists of sentences as the nodes and similarity between each pair of sentences as the weight of corresponding edge. Inverse sentence frequency-cosine similarity method is used as similarity measure to assign a weight to each edge. TL-TranSum is an extractive query-oriented summarizer that utilizes the theory of hypergraph transversals (Van Lierde & Chow, 2019). In this approach, topics presented in the source document collection are extracted using a topic model based on the semantic clustering of terms. After each sentence is labeled with related topics, a theme is assigned to each topic. The associated themes are used to form a hypergraph in which nodes represent sentences of the input text, hyperedges are the resulting themes, and the importance of each theme and its relevance to the query is reflected by hyperedge weights. Following that, a sentence selection method based on the minimal hypergraph transversals.

Several works in the context of extractive text summarization have exploited complex networks. In a study, Antiqueira, Oliveira, da Fontoura Costa, and das Graças Volpe Nunes (2009) proposed complex networks-based approach, named CN-Summ, where each node in the model represents a sentence of the source text and an edge makes a link between each pair of nodes if the corresponding sentences have at least one shared word. The model employs a set of 14 network measurements as well as a voting strategy to rank vertices of the network that leads to the generation of the summary. In the work of Amancio, Nunes, and Oliveira (2012), two models, named SUM-RC and SUM-P, based on complex networks have been used for text summarization. In these models, the structure of the network is established based on the adjacency relation between words. Each distinct word in the text represents a node and edges are formed between two nodes whose corresponding words are immediately adjacent. In SUM-P method, syntactic dependency of words are also added to the network as edges between nodes. Diversity, strength, shortest path, betweenness, and vulnerability are the network metrics employed in these models to determine the centrality of each vertex. Tohalino and Amancio (2018) proposed a method for multi-document text summarization based on complex networks that models a collection of documents as a multi-layer network. The model differentiates between similarity of sentences in a document (intra-layer) and similarity of sentences belongs to different documents (inter-layer). In this approach, each node represents a sentence and edges connect the nodes based on the cosine similarity between the corresponding sentences. Each document constitutes one of the layers of the model. Finally it applies both traditional and dynamical network measurements to determine importance of sentences. Ribaldo, Akabane, Rino, and Pardo (2012) used a hybrid graph-based approach for text summarization in Brazilian Portuguese which consists of an adapted version of the classical Relationship Map for multi-document summarization in combination with complex networks metrics. In this method, Cross-document Structure Theory is applied as a linguistic knowledge to enrich the effect of the traditional Relationship Map. Complex networks measurements lead to select most important sentences in this model.

Many studies in ATS have employed supervised, semi-supervised, or unsupervised machine learning-based approaches in recent years. In a study, a generic single-document summarizer has been proposed which utilizes partitional clustering technique to extract important sentences (Aliguliyev, 2009). This approach uses a global objective function based on the normalized google distance (NGD) as dissimilarity measure between sentences. Then, a discrete differential evolution algorithm is applied to optimize the objective function whose optimization results in creating high quality clusters of sentence. Deep neural network based strategies are widely used in extractive summarization. NN-ML-CL is a general framework for extractive single-document summarization based on a multi-task learning approach that exploits recurrent neural network (RNN) encoder-decoder for sentence extraction jointly document classification (Isonuma, Fujino, Mori, Matsuo, & Sakata, 2017). This model uses convolutional neural network to obtain a sentence embedding from word embeddings, before it applies LSTM-RNN encoder-decoder for sentence extraction. Ren et al. proposed a hybrid deep neural network model for extractive summarization, called sentence relation-based summarization (SRSum), which consists of five sub-module (Ren et al., 2018). Each of these deep neural sub-modules is responsible for taking a kind of sentence relation such as contextual sentence-to-sentence relations, topic-to-sentence relation, and query-to-sentence relation into account. EXTRACT is a neural latent variable based summarizer which models extractive summarization as an instance of sequence labeling (Zhang, Lapata, Wei, & Zhou, 2018). This model considers sentences as binary variables and utilizes those with activated latent variables to derive gold summaries. The summarizer consists of three parts including a Bi-LSTM sentence encoder, Bi-LSTM document encoder, and LSTM document decoder in which the output of this deep neural network is the predicted labels of sentences. Yasunaga et al. proposed a multi-document summarization system, called GRU + GCN, based on the integration of graph representation and deep neural network (Yasunaga et al., 2017). In the first stage of this method, a graph relation graph is constructed with sentences as nodes and similarity between each pair of them as edges. Then, it employs a recurrent neural network with gated recurrent units (GRU) to extract sentence embeddings. In this step, a graph convolutional network (GCN) is applied to create the final sentence embeddings based on the relation graph. A second level GRU produces the document cluster embedding, followed by a greedy sentence selection method which forms summary based on the estimated score of sentences. In the work of Xu, Gan, Cheng, and Liu (2020), a discourse relation based neural summarization method named DiscoBERT has been proposed that is built upon BERT. DiscoBERT consists of a document encoder and a graph encoder. This model extracts Elementary Discourse Units (EDU) instead of sentences as the selection unit to generate the summary. In this approach, two discourse-oriented graphs, named RST Graph and Coreference Graph, are constructed based on RST parse tree and coreference mentions respectively to capture long-range dependencies among discourse units. Then, a Graph Convolutional Network is imposed over the two discourse graphs to update the EDU representation and select salient EDUs. SummCoder is a single-document extractive summarizer that leverages deep auto-encoder (Joshi, Fidalgo, Alegre, & Fernández-Robles, 2019). It applies sentence encoder to obtain sentence embeddings. Following that, a deep encoder-decoder network generates latent representation of the original document. Ranking sentences in this model is done based on sentence content relevance, sentence novelty and sentence position metrics. In a study, a Long Short-Term Memory based deep neural network model was proposed for extractive summarization (Mutlu, Sezer, & Akcayol, 2020). This method utilizes three different sentence scoring solutions including syntactic feature space, semantic feature space, and a combination of them as ensembled feature space. Lastly, a LSTM-NN based neural network model with hierarchical structure is exploited to select salient sentences based on these features.

Evolutionary algorithms have commonly shown promising results in solving the problem of extractive summarization. These kinds of approaches cast text summarization task as an optimization problem where different evolutionary algorithms are applied to select a set of relevant sentences which optimizes covering the main content of the source text jointly decreasing redundancy in the summary. Litvak et al. proposed a learning approach, called MUSE, based on genetic algorithm to find an optimal weighted linear combination of 31 statistical sentence scoring methods (Litvak, Last, & Friedman, 2010). This multilingual extractive summarizer obtains a model to rank sentences of the source text after it was trained on a collection of human-generated document summaries. In many researches, GA has been applied either alone or in combination with other methods. In an unsupervised single-document suumarizer, genetic algorithm is used to extract salient sentences directly from the input document collection based on term frequency (García-Hernández & Ledeneva, 2013). In a study, a multi-document summarization framework has been proposed based on the idea of extracting summaries through the use of integer linear programming and submodularity (Peyrard & Eckle-Kohler, 2016). This framework exploits two different summarizers where one of them applies genetic algorithm and the other is based on swarm intelligence optimization techniques. A discrete optimization based summarization method was proposed by Peyrard and Gurevych to learn a summary-level scoring function θ from a pool of human-generated summaries as supervision and automatically generated data as regularization (Peyrard & Gurevych, 2018). This approach applies simple linear regression to learn an objective function before using genetic algorithm in automatic data generation step to create and optimize summaries in an iterative way. Finally, it extract salient sentences based on the objective function by the use of genetic algorithm.

Several approaches have developed extractive summarization systems based on varied memetic algorithms that are a combination of a population-based evolutionary algorithm and a local search. Mendoza et al. proposed a single-document summarization system which models extracting sentences as a binary optimization problem (Mendoza, Bonilla, Noguera, Cobos, & León, 2014). To solve this problem, it employs a memetic algorithm that couples a genetic population-based global search with a guided local search. MA-MultiSumm is an extractive multi-document summarization approach which attempts to optimize the coverage and redundancy factors in output summary by the use of a memetic algorithm (Mendoza et al., 2014). The algorithm in this summarizer is based on CHC (Cross-generational elitist selection, Heterogeneous recombination, Cataclysmic mutation) evolutionary algorithm, and greedy local search.

Differential evolution (DE) algorithm is another evolutionary strategy which has shown applicable results in ATS. Multi-document generic text summarization has been modeled as a linear fractional integer programming problem (LFIPP) in an approach where it applies an adaptive DE with a new trial vector generation strategy and adaptive crossover operator to cover the main content of the source text and minimize redundancy (Alguliev, Aliguliyev, & Mehdiyev, 2011). In DESAMC + DocSum summarizer, a variant of DE algorithm with self-adaptive mutation and crossover parameters is used to generate an extract summary (Alguliev, Aliguliyev, & Isazade, 2012). In another study, an optimization based approach was proposed for extractive multi-document summarization that models sentence selection as a boolean programming problem (Alguliev, Aliguliyev, & Isazade, 2015). This unsupervised strategy employs a DE algorithm, which is a formidable population-based optimizer, to find the best solution.

Some researches in the field of ATS, based on optimization approaches, have utilized particle swarm optimization (PSO) algorithm. Alguliev et al. proposed a text summarization model that maps multi-document summarization as an integer linear programming problem (Alguliev, Aliguliyev, Hajirahimova, & Mehdiyev, 2011). This model uses two evolutionary strategies in sentence selection phase including branch&bound algorithm and binary PSO to optimize the objective function which is based on cosine similarity and normalized Google distance similarity measures. CDDS is another text summarizer that formulates multi-document summarization as a constraint-driven quadratic integer programming problem and applies PSO to select most important and informative sentence of the source document collection (Alguliev, Aliguliyev, & Isazade, 2013). A study has incorporated a optimization based approach with clustering for single-document extractive summarization (Verma & Om, 2019). It employs a variable dimension particle swarm optimization (VDiPSO) based model for sentence clustering where each particle represents a cluster and each cluster center is shown by the dimension of a particle. Finally, extract summary is generated by selecting top 30% high scored sentences of each cluster.

Ant Colony and Bee Colony optimization algorithms are other kind of evolutionary techniques which have been applied for extractive text summarization (Al-Saleh & Menai, 2018). The generation of extract summaries from multiple source documents is addressed with a Multi-Objective Artificial Bee Colony (MOABC) algorithm where it attempts to optimize two objective functions independently at the same time (Sanchez-Gomez, Vega-Rodríguez, & Pérez, 2018). Saini et al. proposed a single-document summarization framework based on the concept of multi-objective clustering that employs three different evolutionary algorithms to automatically figure out the number of clusters (Saini, Saha, Jangra, & Bhattacharyya, 2019), that is, this approach presents three various text summarizers with regard to the applied evolutionary algorithm. The first approach is based on a combination of self-organizing map and multi-objective DE and the other two utilize lesser-known optimization algorithms, namely multi-objective grey wolf optimizer and multi-objective water cycle algorithm. MCRMR is a optimization based multi-document summarization in which a weighted combination of multiple features is used to score sentences (Verma & Om, 2019). To find the optimal set of weights in this sentence scoring function, Shark Smell Optimization (SSO) algorithm is applied, which is a relatively new evolutionary method.

2.2. Quantum-inspired genetic algorithm

Since the first version of QIGA was proposed in 1996 (Narayanan &

Moore, 1996), A lot of progress has been made in this field of evolutionary computation. This algorithm has been successfully employed in many fields to solve their complex optimization problems. Various versions of this algorithm including sequential and parallel ones were applied to solve knapsack problem (Han & Kim, 2000; Han, Park, Lee, & Kim, 2001). Another version of QIGA was introduced by Talbi et al. to solve the travelling salesman problem which is one the most known combinatorial optimization problems (Talbi, Draa, & Batouche, 2004). In a recent study, a quantum-inspired genetic algorithm has been proposed for k-means clustering, called KMQGA, which can obtain the optimal number of clusters as well as providing the optimal cluster centroids (Xiao, Yan, Zhang, & Tang, 2010). This approach allows variable-length quantum individuals in a population during evolution process. Another research utilizes a variant of QIGA with a novel quantum gate, called learning gate, to tackle the antenna positioning problem that is a known issue in quality design of cellular phone networks (Dahi, Mezioud, & Draa, 2016). Zhang et al. proposed a clustering approach based on QIGA with adaptive quantum rotation strategy which aims at determining the optimal number of clusters and the optimal cluster centers of the input dataset (an Zhang, Deng, & Xia Chang, 2014).

Although multi-document text summarization is considered a complex optimization problem with a large search space, specially when it comes to real-world applications, there is a lack of quantum-inspired evolutionary strategies applied in this context. Thus, with regard to the capabilities of quantum-inspired genetic algorithm and also the nature of extractive summarization, we have developed an extractive multi-document summarizer based on QIGA with modified quantum measurement and self-adaptive quantum rotation gate.

3. The proposed approach

In this section, we introduce the proposed system (MTSQIGA) for extractive multi-document text summarization. We first explain the motivation of our proposed approach. Then, we present the search process of the proposed approach including solution representation, objective function, and its main operators.

3.1. Motivation

Genetic Algorithm (GA) is a powerful search technique that has been shown acceptable results solving complex problems in wide range of science and engineering domains (Zhang, Jin, & Hu, 2003). However, it has shown some limitations such as premature convergence, getting stuck in local optima, and large computation storage in some problems (Zhang et al., 2003; Liao, 2012).

Due to this limitations, a new version of evolutionary algorithms called quantum-inspired genetic algorithm was proposed that is based on the principles of quantum computing and genetic algorithm. Since the first QIGA was proposed, it has been applied to a large number of optimization problems. Many of these studies has shown that QIGA provides better capabilities in terms of diversity, convergence, and search capability compared to conventional GA (Junan, Bin, & Zhenquan, 2003; Han & Kim, 2000; Han et al., 2001). Moreover, Malossini et al. showed in their study that a specific version of QIGA called QGOA has lower complexity, $\mathcal{O}(1)$, than conventional GA, $\mathcal{O}(N\log N)$, where N is the population size (Malossini, Blanzieri, & Calarco, 2008).

The QIGA properties for the most part has arisen from the quantum representation of its individuals as well as quantum operators it uses to evolve the population. Quantum representation allows representing the superposition of all possible solutions of given problem at the same time. Based on this feature, a quantum individual can represent 2^k linear state, where k is number of quantum bits (Q-bits) in given individual, while in classical GA, each individual represents only one state at a time (Li & Wang, 2007). Therefore, the quantum population (Q-pop) can potentially be larger than the population in its classic version with the same size. These features make the QIGA able to do better exploration in search space of complex optimization problems, and also reduce the computation time to find optimal solution(s) (Rylander, Soule, Foster, & Alves-Foss, 2001; Zhang et al., 2014).

However, QIGA has better performance in many situations compared to conventional GA, It has some defects. Some studies shown that quantum-inspired genetic algorithm may diverge or get premature convergence to a local optimum. In Xu, Mei, Dai, and Su (2014), the authors showed lower convergence speed as well as premature convergence to local optima of traditional QIGA in optimizing the multimodal function.

As mentioned in section 1, redundancy is one of the most important challenges in multi-document summarization as a result of an increase in the number of documents. In this case, increase in the number of sentences causes the size of search space grows exponentially. On the other hand, QIGA can be a very promising method for exploring large search spaces while preserving the relation between efficiency and performance. Therefore, we use QIGA as our approach to find the optimal solution in this problem.

3.2. MTSQIGA

In extractive summarization, a subset of sentences are chosen from within the original document and then they are put together in an appropriate order to form the summary. The most important issue in extractive summarization is the selection of most important sentences of the input text. This will be more crucial in multi-document summarization since the number of sentences is much more than in single-document summarization. Generally, extractive summarization consists of three main steps: preprocessing, processing, and generation of the summary; for each of them different approaches have been proposed (García-Hernández & Ledeneva, 2013).

In this paper, we proposed a multi-document generic summarizer which uses QIGA to extract important sentences independently from linguistic resources (Fig. 1). The QIGA presented in this study was designed with emphasis on defining an effective objective function and modifying quantum operators of QIGA. In our proposed method, we control the length of the generated summary by modifying the reproduction operators of QIGA. In addition, objective function of the proposed algorithm consists of statistical features along with cosine similarity. In other words, a combination of four statistical measures and two cosine similarity based measures have been used as the objective function to evaluate the quality of each sentence which should be maximized during the evolution process. Thus, the problem of summarization in the proposed approach is a maximization problem and the QIGA tries to select sentences with the highest score from the original text. In the following subsections, we focus on the stages of MTSQIGA.

3.2.1. Preprocessing

Preprocessing of text, which is an unstructured data, is one of the inevitable steps in natural language processing. This causes the input text to be Processable in next phases. Furthermore, the accuracy of the preprocessing phase has a significant effect on the results of applying main algorithm in processing phase. In other word, the higher the accuracy of the preprocessing step, the evaluation results will be closer to their actual value. Thus, before we could use the QIGA to process the original text in the second phase, it should be adapted to the format of the QIGA.

Redundancy is one of the most critical challenges in text summarization. In multi-document summarization, the amount of redundancy in source text is often much more than single-document one. Furthermore, the search space of multi-document summarization is much larger than the search space of single document summarization. For this reason, multi-document summarization is much more difficult than single-document summarization and the proposed approach for this type of summarization should be scalable since the input text to this type of

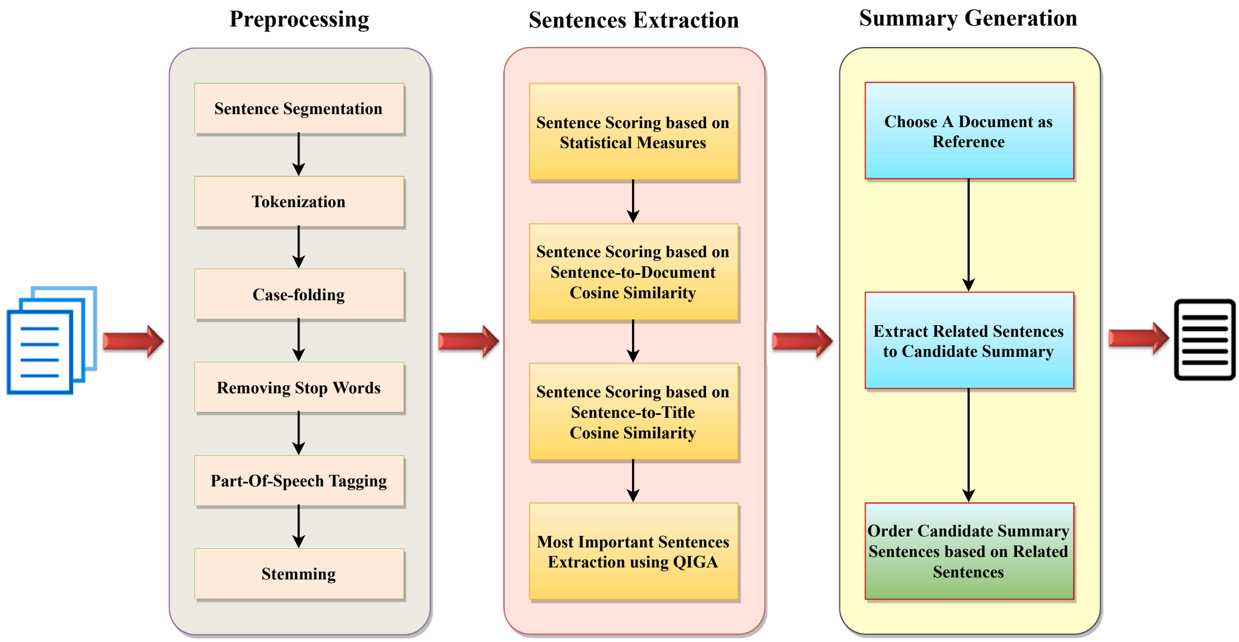

Fig. 1. Overview of MTSQIGA.

summarization systems can be hundreds of text documents. The preprocessing step tries to normalize the original text and adapts it to the format which we need in processing step. First of all, a model should be considered to aggregate the text of all the documents in the collection.

The input to the proposed summarization system is a set of text documents in the form of $D = \{d_1, d_2, \ldots c, d_m\}$, where m is the number of documents. Each document contains a set of sentences that have been considered as $d_i = \{s_1, s_2, \ldots c, s_{|d|}\}$ in our representation model, $|d|$ is the number of sentences in a document d_i and $i = 1, 2, \ldots c, m$. In order to reduce the complexity of the model used in our proposed system, the input document collection is represented as a set of all sentences constituting these documents in the collection as follows:

$$D = \{s_1, s_2, \ldots c, s_n\}$$

where s_i and n represent i^{th} sentence in D and number of all sentences in the document collection, respectively. In this model, the relationship between the sentence s_j and the document collection (D) is as follow condition.

$$s_j \in D \quad \text{iff} \quad s_j \in d_i \in D \tag{1}$$

After the text of all documents in the collection is aggregated, the input text should be separated into it's constituent sentences. The proposed summarization system is looking for a subset of the sentences in the document collection, so that the selected sentences cover the main and important content of the document collection. Let S represents the set of sentences constituting the final summary ($S \in D$), then $sim(\overrightarrow{D}, \overrightarrow{S})$ represents the similarity between the document collection and the generated summary, which all the efforts of the proposed system is to increase the value of this similarity, and reducing the redundancy in the final summary as well.

In the preprocessing of the set of input document collection in the proposed system, the steps are applied in the order indicated below.

- ***Sentence segmentation*:** One of the challenges in some input text documents is that they are written entirely in English uppercase letters. These type of documents make problems while segmenting the text into sentences, since the sentence boundary may be unclear in this situation. In order to solve this problem, we used a statistical language modeling based truecaser which determines the proper capitalization of terms in each sentence. The truecaser uses a greedy strategy to build the modal during training. In this model, the correct casing of each word is determined based on the current word, the previous word, the next word, and their corresponding case content in a sentence. All observed casings for the current word are tested and the casing with the highest score is selected (Lita, Ittycheriah, Roukos, & Kambhatla, 2003). In this method, we calculate the score of each casing of a word using Eq. 2.

$$score(w_0) = p(w_0) \times p(w_0|w_{-1}) \times p(w_0|w_1) \times p(w_0|w_{-1}, w_1) \tag{2}$$

where w_0, w_1, and w_{-1} represent current word, next word, and previous word, respectively. Each of unigram, bigram, and trigram probabilities, p, in Eq. 2 are computed based on a large training corpus.

After truecasing the input text, the sentence segmentation method is applied on the input text of document collection using Natural Language Toolkit (NLTK) (Bird, Klein, & Loper, 2009) which is a module for Python programming language. The output of this step is the set of all sentences from all the documents in the collection, D.

- ***Tokenization*:** In this step, each sentence in D is splitted into its tokens which can be used later in calculating the value of objective function. The output of this step is a set like T which contains all the terms occur in the sentences of D as follows:

$$T = \{t_1, t_2, \ldots c, t_k\} \tag{3}$$

where k is the number of distinct terms used in the input document collection.

- ***Stop words removal*:** Given that one of the sentence scoring measures used in the proposed system is based on term frequency to measure the relevance of each term, then we remove stop words in this step. Stop words are extremely common words such as *because*, *or*, and *from* which would appear to be of little value in helping select important and related sentences of the input documents. In out proposed system, we have used a list of English stop words which includes 214 words in order to remove stop words. By removing these types of words, the accuracy of the cosine similarity based sentence scoring methods increases and thus, this will reduce the redundancy in the final summary.

- *Part-of-speech tagging*: One on the sentence scoring measures used in objective function of our proposed system is the similarity between the sentences of input text and the titles of the document collection based on the number of identical proper nouns in them. To reach this purpose, the summarizer needs a method to tag the words in a sentence based on their functional role within that context in order to identify the proper nouns of the sentence. For this reason, after removing stop words and before stemming step, we attempt to assign POS tags to the words within each sentence of the input text. Given that POS tagging process is case sensitive, so that this phase is applied before reducing all letters of words to lower case in order to obtain more accurate results. In order to assign a POS tag to words, each sentence is given as an input to the POS tagger and a vector of (word-POS tag) pairs is generated as output.
- *Case-folding*: According to the sentence scoring measures used in the objective function of our proposed system, the MTSQIGA requires to have a same behavior with all instances of a word which are written with upper- or lower-case letters in different places of the input text. A common strategy is to do case-folding by reducing all letters to lower case which allows instances of a word at the beginning of a sentence to match with other instances of that word in the middle of the sentence. Hence, converting all words to lower-case letters makes it possible to consider all instances of a word with different case the same. This can be very helpful for the system at the time of calculating the value of term frequency and the cosine similarity based measures which are based on the frequency of distinct terms.
- *Stemming*: One of the important steps in preprocessing of the input text is stemming as it reduces inflectional forms and sometimes derivationally related forms of a word to a common base form of that word. This task has a positive impact on enhancement of the accuracy of objective function, especially similarity-based measures. For example, words such as *argued*, *argues*, and *arguing* are different forms of *argue* which are used in a text based on their grammatical functions. After applying stemming on these words, all of them are reduced to their base or root form, i.e. *argue*. In our proposed system, the Porter stemming algorithm (Porter, 1980) which is the most common algorithm for stemming English is used to reduce inflectional forms and sometimes derived forms of a word to it's base or root form by stripping derivational affixes. Thus, in the processing phase, different forms of a word with the same root are considered the same at the time of calculating the value of sentence scoring measures used in the proposed objective function. Given that redundancy in multi-document summarization is much more than single-document one, thus, stemming plays an important role in enhancement of the performance of the proposed method in the processing phase.

3.2.2. The overview of search process in MTSQIGA

QIGA is a combination of quantum computing and genetic algorithm. It is mainly based on Q-bits and the superposition of states. Unlike the classical representation of chromosomes in GA, here individuals are represented by vectors of Q-bits (quantum register). Thus, an individual can represent the superposition of all potential states at a time.

The overall procedure of the proposed adaptive QIGA for summarization is shown in Fig. 2. In the first step, the proposed algorithm initializes the population of quantum individuals (solutions) with a certain value (initial population step). Each individual in the initial population indicates a potential solution for the summarization problem which is in the form of Q-bit representation. In the next step, each quantum individual is collapsed to a certain state which is a binary string (measurement step) and then evaluated by the objective function in order to be optimized (fitness evaluation step). In this sense, there is not an absolute solution for the problem, but there is a set of possible solutions where some of them are better than others. At this phase, the quantum-inspired genetic algorithm mostly considers the best solutions for the next generation (parents selection step) using roulette wheel and elite selection.

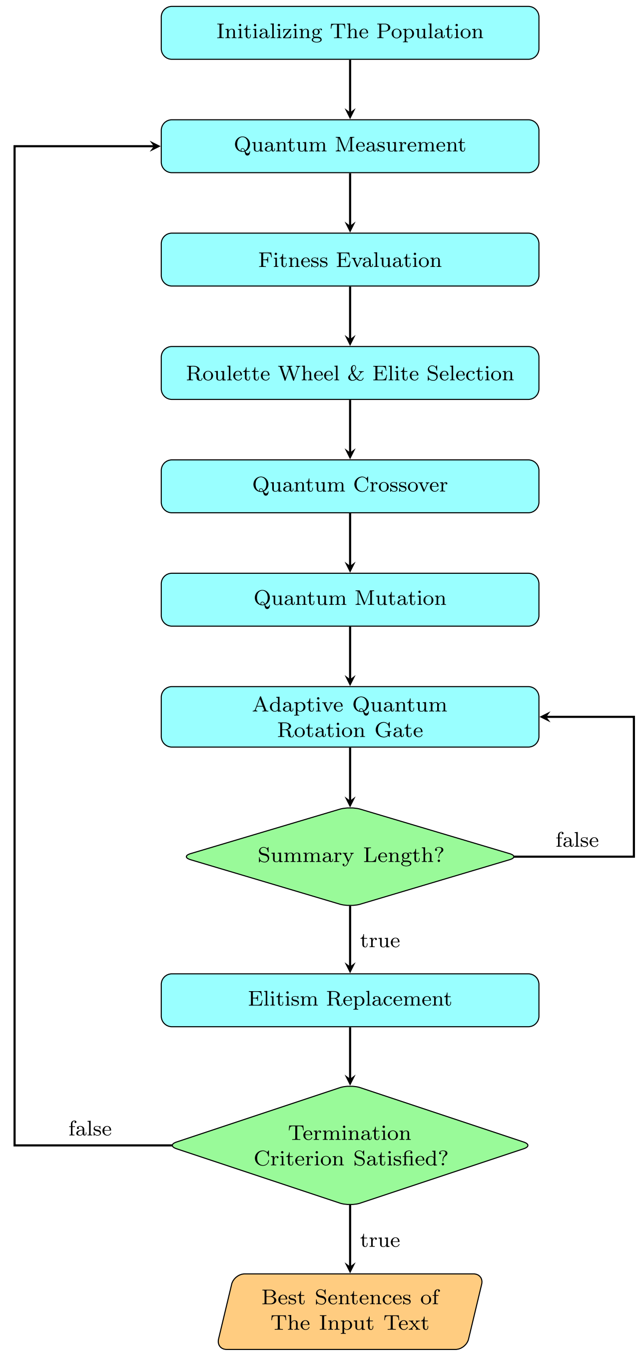

Fig. 2. The overall procedure of search process in MTSQIGA.

Then, MTSQIGA produces a new population by combining (crossover step) some parts from the codification (individual encoding step) of selected solutions in order to obtain better solutions gradually (the evolution principle). After that, the MTSQIGA applies a small variation (the mutation step) to the solutions in the new population in order to explore new areas in the search space. Then, the modified quantum rotation gate is applied to each individual to take advantage of the best solution in the population (quantum rotation gate step). At the final step, elitism replacement is applied to produce new generation from current one (replacement step). The new produced population is evaluated again and this process is repeated until a desired solution is obtained or some termination criterion is reached (stopping criteria).

The proposed quantum-inspired genetic algorithm has been modified so that it can be used for automatic text summarization. In other words, the QIGA has been modified in such a way that it guarantees the length of the final summary cannot be longer than the user specified value. Therefore, We modified some reproduction operators of the algorithm such as measurement and quantum rotation gate in order to achieve this. In the following subsections, each step of the adaptive QIGA will be discussed in more detail.

3.2.3. *Representation of individuals*

The classical GA uses several representations such as binary, discrete, permutation, and real values to encode solutions. Unlike classical GAs, the quantum-inspired ones use probabilistic representation called Q-bit (quantum-bit), which is adopted based on the principles of quantum computing. A Q-bit is the smallest unit of information stored in a two-state quantum computer, which may be in the “1” state, or in the “0” state, or in a linear superposition of these two states (Narayanan & Moore, 1996; Hey, 1999). Each Q-bit is represented as a pair of probabilities α and β, where α and β are complex numbers that specify the probability amplitudes of the corresponding states. The state of a Q-bit can be represented as follows using Bra/Ket notation:

$$|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{4}$$

According to Eq. 4, $|\alpha|^2$ denotes the probability of being in state “0” and $|\beta|^2$ represents the probability of being in state “1”. If the states are normalized to unity, the following should be guaranteed (Narayanan & Moore, 1996):

$$|\alpha|^2 + |\beta|^2 = 1 \tag{5}$$

Each quantum individual in the quantum population is represented as a string of n Q-bits as follows:

$$\overrightarrow{X}_Q = \begin{bmatrix} \alpha_1\alpha_2 & \ldots\alpha_n \\ \beta_1\beta_2 & \ldots\beta_n \end{bmatrix}$$

where n is the number of Q-bits in the individual.

In our proposed approach, each quantum individual represents a superposition of states, which is a potential solution for the summarization problem. Given that the extractive text summarization system should decide on the selection of sentences, each quantum individual potentially represents selected sentences for the final summary. The size of all individuals in the population is the same which is equal to the number of all sentences in the document collection. In other words, each individual of the quantum population contains n, where n represents the number of all sentences in the input text. Also, in each individual, i^{th} Q-bit represents corresponding sentence of the source text. Each Q-bit is encoded as a pair of complementary probabilities α and β which confirms Eq. 5. In this case, The value of $|\alpha|^2$ in a Q-bit determines the probability of having corresponded sentence selected to appear in the final summary and the value $|\beta|^2$ determines the probability of having corresponded sentence discarded.

3.2.4. *Initial quantum population*

Initializing quantum population (Q-pop) plays a crucial role in the convergence of QIGA and its efficiency (Bonissone, Subbu, Eklund, & Kiehl, 2006). Thus, after determining the encoding of the individuals, we can generate initial Q-pop by considering some parameters. One of the important parameters in the generation of initial population is the size of it which influences the performance of QIGA. In our proposed approach, the size of initial population is considered with three different values in order to examine the effect of initial population size on the performance of the proposed system. This values include $k = \frac{n}{4}, k = \frac{n}{10}$, and a value mapped to the range $[50, 100]$ based on the number of sentences used in each set of input document collection, which is calculated using Eq. 6.

$$v_{new} = \frac{n - min_{old}}{max_{old} - min_{old}}\left(max_{new} - min_{new}\right) + min_{new} \tag{6}$$

where n is the number of sentences in the source text. min_{old} and max_{old} represent the size of the shortest and longest document collections in terms of the number of their sentences which are set to 356 and 1451 in our proposed approach, respectively. Also, min_{new} is equal to 50 and max_{new} indicates 100.

The initialization of population in MTSQIGA is the same as the canonical QIGA. In our work, we set the initial value of α and β of each Q-bit in the Q-pop to $\frac{1}{\sqrt{2}}$. Thus, at the beginning of the algorithm, each Q-bit has equal probabilities to be in the state “0” or “1”, that is, all sentences have equal probabilities to be selected or not at the beginning. Fig. 3 demonstrates an example of initial quantum population with k individuals for a summarization problem with n sentences in the proposed QIGA. As it can be seen from the figure, each Q-bit corresponds to a certain sentence of the input text.

3.2.5. *Quantum measurement*

One of the most effective operators in our proposed quantum-inspired genetic algorithm is *quantum measurement*, which has a direct control on the final summary length. QIGA plays a key role in sentence selection phase of our summarization system. Given the purpose of extractive summarization task, this problem has binary nature. In binary optimization problems, an acceptable solution is represented as a vector of 0 and 1. Thus, in automatic extractive text summarization, each vector of binary values such as $\overrightarrow{X}_B = \{x_1, x_2, \ldots c, x_n\}$ represents selected and discarded sentences for appearing in the final summary, where n denotes the number of all sentences of the document collection. In this representation, the rank of each element in the solution represents the rank of the corresponding sentence in the input text, i.e. $i = 1$ in the solution vector denotes the first sentence in the input text, $i = 2$ for the second sentence, and $i = n$ stands for the last sentence. An example of a potential solution for an ATS problem which is composed of 10 sentences is shown in Fig. 4.

Each element in a potential solution for automatic extractive summarization is binary-valued ($x_i \in \{0, 1\}$), where $x_i = 1$ means the i^{th} sentence is selected for appearing in the summary, while $x_i = 0$ means the corresponding sentence is discarded.

Common binary encoding of solutions in classical genetic algorithm is one of the main weaknesses of this algorithm for multi-document summarization, since in multi-document summarization, the number of sentences is much more than single-document summarization. On the other hand, The determinism of the solution encoding in classical GA

$$\overrightarrow{X}_{Q_1} = \begin{array}{c} \left[\begin{array}{c|c|c|c} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \end{array}\right] \\ \begin{array}{cccc} 1^{st}\ sent & 2^{nd}\ sent & & n^{th}\ sent \end{array} \end{array}$$

$$\overrightarrow{X}_{Q_2} = \begin{array}{c} \left[\begin{array}{c|c|c|c} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \end{array}\right] \\ \begin{array}{cccc} 1^{st}\ sent & 2^{nd}\ sent & & n^{th}\ sent \end{array} \end{array}$$

$$\vdots$$

$$\overrightarrow{X}_{Q_k} = \begin{array}{c} \left[\begin{array}{c|c|c|c} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \cdots & \frac{1}{\sqrt{2}} \end{array}\right] \\ \begin{array}{cccc} 1^{st}\ sent & 2^{nd}\ sent & & n^{th}\ sent \end{array} \end{array}$$

Fig. 3. An example of initial quantum population in MTSQIGA.

$\overrightarrow{X}$	0	0	1	0	1	0	0	0	1	1
	1^{st} Sent	2^{nd} Sent	3^{rd} Sent	4^{th} Sent	5^{th} Sent	6^{th} Sent	7^{th} Sent	8^{th} Sent	9^{th} Sent	10^{th} Sent

Fig. 4. Representation of a potential solution for extractive summarization with 10 sentences.

causes an increase in the number of individuals in the population exponentially based on linear increase in the number of the sentences of the input text. Unlike the genetic algorithm, the population in QIGA is based on the Q-bits which use probabilities. In this situation, the QIGA needs an operator to map each quantum solution into a binary one in order to measure the quality of the solutions based on the generated summary. Quantum measurement is responsible for the mapping of each quantum individual to its corresponding binary solution among all possible solutions that it contains. In other word, at this step, each quantum individual $\overrightarrow{X}_Q$ is observed using quantum measurement in order to generate a binary solution $\overrightarrow{X}_B$. The Eq. 7 is applied as a condition to map a Q-bit to its corresponding binary bit. In this equation, *rand* is a random number which is generated from the standard uniform distribution $U(0,1)$.

$$X_B = \begin{cases} 0 & \text{if } rand \leqslant |\alpha|^2 \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

The other importance of measurement in MTSQIGA is that it can control the generated summary length. In fact, one of the mechanisms used to control the summary length in the proposed algorithm is quantum measurement. In order to control the summary length not exceed from the user-specified size using this operator, the genes of each binary solution are assigned 1 in a controlled manner when the algorithm wants to map each quantum individual to its corresponding binary solution. The pseudocode of quantum measurement used in the MTSQIGA is shown in Algorithm 1.

Algorithm 1. Pseudocode of presented quantum measurement in MTSQIGA

```
begin
  summary_length←0
  foreach quantum individual do
    for small each Q-bit (α_i / β_i) in the current quantum individual then
      if summary_length +lengthofsentence_i < desired_length then
        Pickrrandomlyfrom(0,1)
        if r≤|α|^2 of Q−bit_i then
          binary_solution[i]←0
        else
          binary_solution[i]←1
          summary_length += length of sentence_i
        end if
      else
        binary_solution[i]←0
      end if
    end for
    while summary_length is not reached to desired_length do
      State Apply quantum measurement on each Q-bit again
    end while
  end for
end
```

3.2.6. Objective function

One of the main components of evolutionary algorithms is the objective (fitness) function, which is used to evaluate the quality of the solutions. Generally, calculating the value of the objective function for individuals of a population is the most costly part of QIGA. In our proposed approach, the emphasize is on the design of a novel and effective objective function.

The purpose of MTSQIGA is to generate a summary, in which the similarity between the summary and the document collection is maximized (coverage) and at the same time the generated summary has less redundancy. In order to generate such a good quality summary, we apply three different objective functions to assess the results obtained by each of them. In our proposed approach, the third objective function is the main one which generates better quality summaries. The first objective function comprises two sentence scoring measures which are based on term frequency and sentence position. In our second objective function, the combination of three sentence scoring measures including inter-sentence cosine similarity of the input text, the cosine similarity between sentences of the input text and the title of each input document, and the sentence length measure are applied. Finally, the third objective function which is our main proposed objective function, is composed of a combination of six metrics for scoring sentences of the input documents. Four of the sentence scoring measures in this fitness function are based on statistical methods and two other measures are based on the cosine similarity. All the three objective functions applied in MTSQIGA are a maximization function, that is, the algorithm tries to find those solutions during evolution which maximize the value of the objective function. In the rest of this section, we describe the sentence scoring metrics which are used in our main proposed objective function.

- *term frequency*: The relevance of a term (w) can be determined by the frequency of it in the input text, $frequency(w,T)$, where T represents input text. On the other hand, expressiveness of a summary is represented by the number of distinct terms in it. Therefore, a golden summary must contain the most relevant (most frequent) terms of the input text (regardless of stop words) and it must also include distinct terms mostly (García-Hernández & Ledeneva, 2013). In order to normalize the term frequency measure, total frequencies of distinct terms in the summary is divided by total frequencies of the most frequent terms of the input documents, which is shown in Eq. 8.

$$f_{freq} = \frac{\sum_{i=(term \in S)}^{m} frequency\left(i, T\right)}{\sum_{j=(term \in T)}^{m} frequency\left(j, T\right)} \tag{8}$$

where S represents all the distinct terms occurring in generated summary and T denotes all the distinct terms of input text.

- *Sentence position*: Sentence position measure is a heuristic method. In some studies, it has shown that most of the time, the relevant information in a document is found in specific sections such as titles, headings, the first sentences of each paragraphs, and the last sentences of paragraphs (Lin et al., 1997). This metric is based on the idea that the first sentences of the original text can be good candidates for appearing in the final summary. According to the classical version of this selection criteria, the inverse position order of sentences is used to score them, which means that if an input text contains 50 sentences, the first sentence will be 50 times more important than the last sentence. This will almost make the probability of appearing the last sentence in the summary impossible. We used the modified sentence position measure proposed in García-Hernández and Ledeneva (2013) in order to make the difference between the probability of selecting first and last sentences softer using a linear equation with slope t. Based on this feature, if the sentence i is selected from a text with n sentences for appearing in the final summary, the corresponding sentence weight is defined according to Eq. 9:

$$t(i-x)+x \tag{9}$$

where $x = 1+\frac{n-1}{2}$ and t is a heuristic slope. According to our experiments, in our proposed approach, the value of t has set equal to -0.6. Also, in order to normalize the sentence position measure (δ), the value obtained from Eq. 9 is divided by the relevance of the first k sentences, where k is the number of selected sentences for the final summary.

$$f_{pos} = \frac{\sum_{|C_i|=1}^{n} t\left(i-x\right)+x}{\sum_{j=1}^{k} t\left(j-x\right)+x}, \qquad x = 1+\frac{n-1}{2} \tag{10}$$

- ***Sentence length*:** The idea of sentence length measure is that longer sentences usually contain more information than short sentences. thus, their likelihood of appearing in the final summary is more. In the proposed method, the weight of each sentence based on this metric is equal to the number of terms in the sentence, which is represented in Eq. 11. The length of each sentence is normalized using the length of the longest sentence of the input text.

$$f_{len} = \sum_{S\in Summary} \frac{Length(S)}{Length(L)} \tag{11}$$

where L is the longest sentence of original text.

- ***Similarity with titles based on identical proper nouns*:** Different types of words have various effects in sentences according to the part they play in a sentence. In a general categorization of words, they can be placed in each of the five categories called word classes or part of speech which includes proper nouns, common nouns, verbs, adjectives, and other words. In this categorization, proper nouns are more important than other words and sentences containing these types of words usually include more information than other sentences. On the other hand, in most cases, proper nouns which appear in the titles of documents are the most important proper nouns in their corresponding text and sentences containing these words are likely to be among the most important sentences of the input document collection. Hence, in the objective function of the proposed method, this sentence scoring measure is used to increase the similarity between the sentences of the input text and the titles of the documents as well as finding more important sentences. Unlike most previous methods, in which all the proper nouns in each sentence are counted (Suanmali, Salim, & Binwahlan, 2011; Christensen, Mausam, Soderland, & Etzioni, 2013), in the objective function of our proposed quantum-inspired genetic algorithm, only proper nouns, which are identical between titles of documents and sentences of input text, are counted. Based on this method, not only more important sentences of the input document collection obtain higher scores but also these sentences are more similar to the titles of the input documents. In order to compute this measure, the output of part-of-speech tagging phase which are vectors composed of pairs of (word-POS tag) is directly used. Using these vectors, the number of proper nouns which are identical between the sentences of the input text and the sentences constituting the titles of the document collection is counted. Based on this measure, the score of each sentence of the input text is computed using Eq. 12.

$$f_{p-noun} = \sum \frac{Count_{proper-noun}\left(\vec{s_i}, \vec{T}\right)}{Length(s_i)} \tag{12}$$

In Eq. 12, $\vec{s_i}$ is the vector which is composed of (word-POS tag) pairs corresponding to the i^{th} sentence of input document collection, $\vec{T}$ is the vectors composed of (word-POS tag) pairs corresponding to the sentences of titles of input documents, and *Count* function returns number of identical words between each sentence of input text and set of sentences of titles of documents. In order to normalize the output of this measure, the value obtained from the *Count* function is divided by the length of the corresponding sentence, s_i.

- ***Coverage (similarity between the summary and the input text)*:** The main idea of the proposed objective function in this paper is to use cosine similarity in order to maximize the coverage of the input text's main content and reduce redundancy in the final summary as well. Hence, this sentence scoring method consists of two parts. The first part of coverage metric calculates the cosine similarity between the candidate sentences for the summary and all sentences of the document collection. The second part consists of calculating the similarity between the selected sentences for the summary with each other. Assuming that D and S are the set of all sentences of the input document collection and the set of sentences constituting a summary respectively, then the similarity between the document collection and the summary, $sim(\vec{D}, \vec{S})$, should be maximized. $\vec{D}$ and $\vec{S}$ denote the feature vectors of the input text and the summary, respectively. Calculating this sentence scoring method in the proposed objective function reduces redundancy in the generated summary and it also gives higher probability of selection to those sentences of the input documents which cover much important content of the input text. Formally, according to coverage method, we can formalize the summarization problem in the form of the Eq. 13.

$$\begin{aligned} &\text{maximize } sim\left(\vec{D}, \vec{S}\right), \\ &s.t. \qquad len(S) < L \end{aligned} \tag{13}$$

Where $len(S)$ is the length of the summary and L is the user-specified size. Calculating the similarity between textual units requires that each of them be presented as a vector. The vector space model is one of the most known methods for representing textual units. In this model, each textual unit is represented by the number of its constituent terms.

The cosine similarity is calculated using the weighting vector representation of the terms corresponding to each sentence. Each sentence like s_i is denoted as $\vec{s_i} = \{w_{i1}, w_{i2}, \dots c, w_{im}\}$ according to this type of representation, where m is the number of distinct terms in the input text and w_{ik} is the weight of the term t_k in the sentence s_i. In our proposed method, w_{ik} is calculated based on *tf-isf* scheme, which assigns a weight to each term of the s_i sentence based on the Eq. 14.

$$w_{ik} = f_{ik} \times log\left(\frac{n}{n_k}\right), \tag{14}$$

where f_{ik} is the frequency of term t_k in sentence s_i, n_k denotes the number of sentences contains term t_k. Then, the cosine similarity between two feature vectors $\vec{s_i}$ and $\vec{s_j}$ is calculated as follows:

$$sim_{cos}\left(\vec{s_i}, \vec{s_j}\right) = \frac{\sum_{k=1}^{m} w_{ik} \times w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \cdot \sum_{k=1}^{m} w_{jk}^2}}, \quad i,j = 1,2,\dots c,n \tag{15}$$

Coverage measure is one of the key elements of the objective function in our proposed approach. It is absolutely helpful for maximizing the coverage of important content of source text and minimizing redundancy in the summary. This cosine similarity based measure is computed using Eq. 16 which is based on Eq. 15.

$$f_{COV} = \sum_{i=1}^{n-1} \times \sum_{j=i+1}^{n} \left[sim_{cos}\left(\overrightarrow{D}, \overrightarrow{s_i}\right) + sim_{cos}\left(\overrightarrow{D}, \overrightarrow{s_j}\right) - sim_{cos}\left(\overrightarrow{s_i}, \overrightarrow{s_j}\right) \right], \quad s_i, s_j \in S \tag{16}$$

Eq. 16 consists of two important parts that they have a significant effect on the generation of high quality summary. The first part of this equation, $sim_{cos}(\overrightarrow{D}, \overrightarrow{s_i}) + sim_{cos}(\overrightarrow{D}, \overrightarrow{s_j})$, attempts to maximize the coverage of the main content of document collection, while the second part, $-sim_{cos}(\overrightarrow{s_i}, \overrightarrow{s_j})$, has been used to minimize redundancy in the summary. Increasing the value of the first part while reducing the second part in Eq. 16 will result in the selection of the most important sentences of the source text.

- *Cosine similarity to the titles of document collection*: Another important sentence scoring measure in the proposed objective function is the similarity of selected sentences for appearing in the summary with the titles of the documents in the collection. According to this measure, Sentences which are more similar to the titles of the documents, are more appropriate candidates to appear in the final summary. In other words, increasing the value of this measure for a sentence indicates that it has better quality. In the objective function of the proposed method, the cosine similarity has been used to calculate the resemblance of the sentences constituting the summary and the titles of the documents. As we have aggregated titles of all documents of the source cluster as a single unit in the preprocessing phase, this measure computes the similarity between each main body sentence and titles of all documents of the cluster. The value of this measure is computed by Eq. 17.

$$f_{TS} = \sum sim_{cos}\left(\overrightarrow{h}, \overrightarrow{s_i}\right), \quad s_i \in S \tag{17}$$

Where $\overrightarrow{h}$ is the weighting vector of the distinct terms appearing in the titles of the input documents. Note that the value of f_{TS} for documents without any title will be zero.

Based on the explained sentence scoring measures, our proposed objective function is designed as Eq. 18.

$$fitness = \left[0.6 \times \left(W_1 \times f_{COV} + W_2 \times f_{p-noun} \times f_{TS}\right) + 0.4 \times f_{len}\right] \times \left(f_{freq} \times f_{pos}\right) \tag{18}$$

In our proposed objective function, the coefficient of similarity-based sentence scoring measures is larger than sentence length measure, as they have more important role in sentence selection phase. Also, it can be seen from Eq. 18 that the sentence length measure has more effect than term frequency and sentence position measures, because it has better performance in selecting appropriate sentences in our point of view. Term frequency and sentence position are considered as coefficients so that these two measure have an impact on the whole equation, although their effect is lower than other four measures. Furthermore, W_1 and W_2 are considered to be equal to 0.5 based on various experiments on our validation set (i.e., DUC 2002 dataset).

3.2.7. Selection

After evaluating quantum individuals based on the proposed objective function, a given number of quantum parents are selected using a specific strategy to create the mating pool. Selected quantum parents are responsible for creating the quantum individuals of the next generation. In MTSQIGA, the classical roulette wheel selection method is used for this step, whose pseudocode is shown in Algorithm 2. In our proposed system, the number of individuals in the current population is produced. Therefore, the number of the current parent population should be selected. Selected parents randomly pair together, and then quantum reconfigurations are applied to each parent pair for the production of children.

Algorithm 2. Pseudocode of applied roulette wheel selection

```
begin
  μ←population_size/2
  current_pair←1
  while (current_pair⩽μ) do
    Pick r1, r2 randomly from [0, 1]
    i←1
    j←1
    while a[i] < r1 do
      i←i + 1
    end while
    while a[j] < r2 ∨ individual[j] = individual[i] do
      j←j + 1 until j⩾population_size
    end while
    mating_pool[current_pair]←(parents[i], parents[j])
    current_pair←current_pair + 1
  end while
end
```

The roulette wheel sampling is based on the absolute value of the fitness function. In this way, the individuals with greater fitness function, more likely to be chosen as parents. However, in this case there is a possibility of choosing the worst individual, but its probability is small. In addition, the elitism selection is also used. The elitism approach prevents the best solution found in the current generation. In fact, using this strategy, the best solution found in the current generation will be copied in the next generation directly.

3.2.8. Quantum crossover

Having constituted the mating pool, these $\frac{k}{2}$ pairs of selected quantum parents are subjected to quantum crossover to produce k new quantum offspring, where k is the size of population. In theory, all types of crossover operators in canonical GA are applicable to quantum counterpart. In our proposed approach, two-point quantum crossover with probability $P_c = 0.75$ is applied. The quantum version of two-point crossover operator leads to exchange of α and β corresponding to the Q-bits between two randomly chosen points in each pair of quantum parents in order to produce two new quantum offspring (Han & Kim, 2000). An example of how this operator functions has been illustrated in Fig. 5. As shown in the figure, after determining the two intersection points randomly, the Q-bits between these two points are exchanged in their corresponding parents.

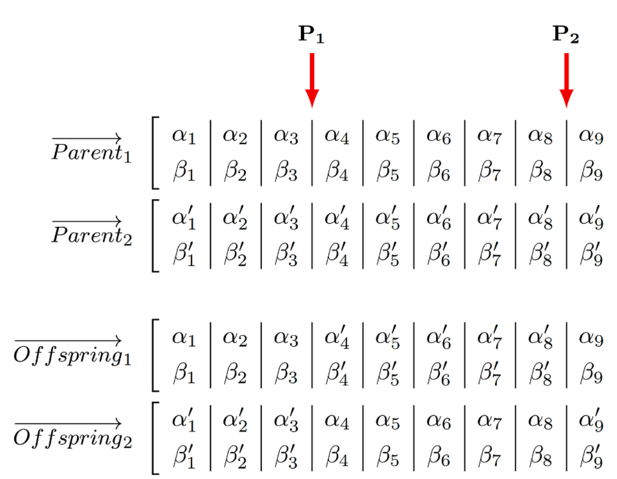

Fig. 5. An example of applied two-point quantum crossover.

3.2.9. Quantum mutation

However QIGA is a powerful search technique that has a high ability to solve optimization problems with big search space, it may get premature convergence with a small probability. So, the proposed method applies a quantum mutation operator in order to prevent premature convergence to local optimum through increasing exploration of QIGA. In the proposed approach, we applied quantum bit-flip mutation which operates on each Q-bit according to pre-specified probability (Sivanandam & Deepa, 2007). Fig. 6 illustrates an example of how the quantum bit-flip mutation works. For most optimization problems, the mutation probability is set to a low value to prevent divergence of the algorithm. On the other hand, QIGA has high exploration capabilities, which stems mainly from its quantum representation. Thus, the value of this control parameter (mutation probability) has been set to $P_m = 0.005$ in MTSQIGA.

3.2.10. Adaptive quantum rotation gate

The most effective step in MTSQIGA is the quantum interference which we used a modified version of the canonical rotation gate operator for this step. The proposed rotation gate is effective not only in guiding the search to a global optimal solution but also control the length of the summary. The quantum rotation gate is applied to each produced quantum offspring based on the binary solution produced by the best quantum individual $\overrightarrow{X}_{Q^*}$, that is, this operator is applied on the Q-bits of quantum individuals in order to reinforce the probability amplitudes of the best quantum individual in other individuals of the population and decrease the probability amplitudes of having other quantum individuals. Therefore, the quantum interference phase increases the utilization of the best individuals which founded in the evolution process and can be considered as an exploitation mechanism in QIGA.

In most versions of QIGA, a fixed lookup table is used to determine the direction and value of the rotating angle in quantum rotation gate, which this method applies the constant values of the table regardless of the search status. This type of updating quantum individuals may have a negative impact on the search speed and efficiency of the algorithm, since the rotation gate operator does not pay attention to the current state of the search. To overcome this issue, in MTSQIGA, an adaptive strategy has been proposed to determine the direction and value of the rotating angel based on the search status. As a result, it has improved the performance of the canonical rotation gate for automatic text summarization problem.

However, the overall approach of our proposed quantum rotation gate in MTSQIGA is similar to the canonical one, it determines the value and direction of rotation angel $\Delta\theta$ in an adaptive scheme based on two factors. The first factor in determining $\Delta\theta$ is the difference between fitness value of each quantum offspring and the fitness value of the best quantum individual of the current generation. The second factor is the length of the generated summary in comparison with the user-specified size. In this approach, the changes of the $\Delta\theta$ value corresponding to each individual is much more when the difference between the fitness value of the current individual and the best quantum individual is high, and

$$\overrightarrow{X}_Q \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 & \alpha_9 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 \end{bmatrix}$$

Mutant Q-bit

$$\overrightarrow{X'}_Q \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \beta_5 & \alpha_6 & \alpha_7 & \alpha_8 & \alpha_9 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 & \alpha_5 & \beta_6 & \beta_7 & \beta_8 & \beta_9 \end{bmatrix}$$

Fig. 6. An example of applied Q-bit flip quantum mutation.

the length of the generated summary is less than the user-specified size. Thus, the current quantum individual alters faster toward the best quantum individual of the population. On the other hand, when the length of the summary is larger than the user-specified size, the slope of altering the selected individual toward the best individual of the population decreases in an adaptive scheme based on both the quality of the current individual and the length of the summary. Therefore, in this approach, two different scenarios can occur for calculating $\Delta\theta$, where in each of them, the direction and value of rotation angel is calculated as Eq. 19.

$$\Delta\theta_i = \gamma \left[0.005\pi + \left(0.05\pi - 0.005\pi\right) \times \left(\frac{\left|f_{X_Q} - f_{X_{Q^*}}\right|}{f_{X_{Q^*}}}\right) + 0.25\lambda\pi \times \frac{L_{user} - L_{summary}}{L_{user}}\right] \tag{19}$$

where f_{X_Q} and $f_{X_{Q^*}}$ represent the fitness of the current quantum individual and the best quantum individual in the current generation, respectively. L_{user} is the user-specified size and $L_{summary}$ represents the length of the summary which is generated by the current quantum individual. The control parameter γ can be set to 1 or -1 based on the length of the summary as shown in condition 20. This parameter will be set to 1 when the length of the generated summary by the current individual is less than or equal to the user-specified size, otherwise it is set to -1.

$$\gamma = \begin{cases} 1 & \text{if} \quad L_{summary} \leqslant L_{user} \\ -1 & \text{otherwise} \end{cases} \tag{20}$$

Furthermore, λ parameter in Eq. 19 is used for handling the two possible scenarios in which it can take one of two values 1 or -1 based on the current scenario (condition 21). The first scenario occurs when the i^{th} bit of the binary solution X_B generated by the current quantum individual X_Q is 0 while the i^{th} bit of the binary solution corresponding to the best quantum individual X_{Q^*} is 1. In this case, λ will be set to 1. The second scenario is the opposite of the first scenario, that is, the value of λ will be set to -1 when the i^{th} bit in X_B corresponding to X_Q is 1 and the i^{th} bit of binary solution X_{B^*} corresponding to X_{Q^*} is 0.

$$\lambda = \begin{cases} 1 & \text{if} x_{B_i} = 0 \quad \& \quad x_{B_i^*} = 1 \\ -1 & \text{elseif} \quad x_{B_i} = 1 \quad \& \quad x_{B_i^*} = 0 \end{cases} \tag{21}$$

Therefore, in our proposed method, the value and direction of the rotating angle in the proposed rotation gate are a function of the quality of the quantum individual and the length of the generated summary by that individual. The method of calculating rotation gate resulting from rotation angel ($\Delta\theta$) is given in Eq. 22 which is used to update the values of α and β corresponding to each Q-bit.

$$R\left(\Delta\theta_i\right) = \begin{bmatrix} \cos(\Delta\theta_i) & -\sin(\Delta\theta_i) \\ \sin(\Delta\theta_i) & \cos(\Delta\theta_i) \end{bmatrix} \tag{22}$$

$\Delta\theta$ makes the i^{th} Q-bit converge toward the state 0 or 1 based on the best quantum individual in the population of the current generation according to the matrix illustrated in Eq. 22, where $i = 1, 2, \ldots c, n$ and n represents the size of each quantum individual. After calculating the rotational gate, this operator updates each Q-bit according to following equation.

$$\left(\alpha_i^{'}, \beta_i^{'}\right) = R\left(\Delta\theta_i\right) \cdot \left(\alpha_i, \beta_i\right) \tag{23}$$

Furthermore, in order to control the length of the generated summary using the quantum interference operator, an approach is considered in this operator which handles the length of the summary based on the probability amplitudes α and β of each Q-bit after applying the rotation gate operator on the quantum individual. In this mechanism, probability of having state 1 or 0 in each Q-bit of a quantum individual is calculated based on the amount of increase or decrease in the value of α and β after applying the rotation gate and then the length of the summary is calculated based on this estimation which the system has made.

If the estimated length of the summary exceeds the user-specified size, the rotation gate operator will be applied on those Q-bits of the quantum individual which were not subjected to rotation gate and their corresponding sentences have obtained lower score from the measures based on cosine similarity as well. In this way, the proposed system guarantees generation of a summary with a maximum specified length. Fig. 7 illustrates how the quantum rotation gate affects on each Q-bit.

3.2.11. Replacement

This step concerns forming next generation of quantum population through replacing the current generation with the individuals from offspring and parents. In order to achieve this goal, quantum individuals from the parents and offspring populations will be selected to form next generation using a specific replacement strategy. The replacement operator applied in MTSQIGA is an elitism replacement which forms half of the next generation ($\frac{k}{2}$) from the best quantum individuals of the parents group and the other half ($\frac{k}{2}$) from the best individuals produced by the recombination operators. Moreover, Through this replacement strategy, the presence of the best quantum individual of the current generation in the next generation is guaranteed. Once the next generation of quantum population is formed, all of the above evolutionary steps are performed in an iterative way until one of the stopping criterion is reached.

3.2.12. Termination criterion

In our proposed approach, the combination of two termination criterion is used. The optimization process in MTSQIGA runs until one of these given termination conditions is satisfied in the quantum genetic algorithm. One of the stopping conditions in MTSQIGA is based on assessing the fitness of several quantum individuals of the previous and current generations. According to this termination criterion, we sort the quantum population of both the previous and current generation in descending order at first. Then, we select the first half of the current population and compute the average fitness of it. We do this for the quantum population of the previous generation as well. If the absolute difference between the quality of these two groups is less than a threshold value which is considered 0.05 in 20 consecutive generations, the QIGA will stop the evolution process.

In some cases, the first termination criterion may prevent QIGA from stopping evolution process since the absolute difference between the average of two consecutive generations is a little greater than 0.05:

$$0.05 < \left| fitness_{curr^*} - fitness_{prev^*} \right| \leqslant 0.05 + \epsilon$$

In this situation, the second termination criterion will play its role which is considered to be production of quantum population up to 500 generations. If each of these conditions is met by the generated solutions of the quantum population, MTSQIGA will return the best solution of the current generation as the most important sentences of the input document collection.

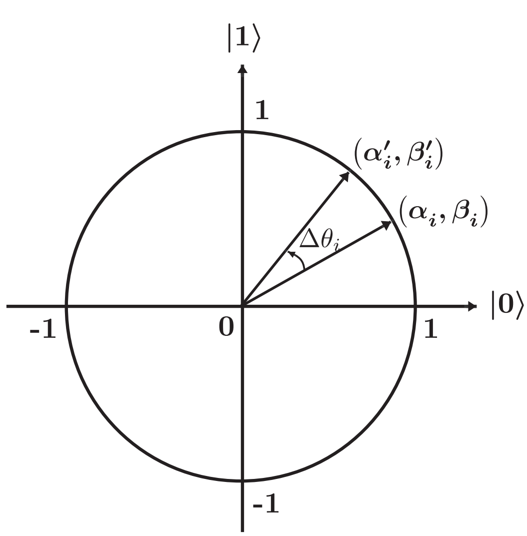

Fig. 7. An example of quantum rotation gate (Dahi et al., 2016).

3.2.13. Generation of the summary

Given that in multi-document summarization, sentences which are selected to appear in the final summary may belong to different documents of the input collection, then the semantic coherence between sentences of the summary can be reduced. In order to maintain the coherence of the content in the final summary, it is important to order appropriately the sentences of the summary in multi-document summarization. Thus, in our proposed system, two different strategies have been used to order the selected sentences. In the first strategy, which is a simple heuristic approach, sentences are ordered in the summary according to the date of their corresponding documents, that is, sentences which are selected from documents with older release date are placed before the sentences belongs to the input documents with a newer release date (Dang, 2006). In this way, the selected sentences take an initial order, but this method can not ensure the coherence of the sentences of the summary in many cases. Thus, in the proposed system, we have applied a more accurate approach to sort the selected sentences in addition to the first strategy. In our second strategy, one of the documents of the input document collection is considered as the reference document. Then, the system calculates the cosine similarity between the selected sentences and the sentences of the reference document. In the next step, for each of the selected sentences of the final summary, the most similar sentence from the reference document is extracted. Finally, the sentences of the summary are arranged based on the place of their corresponding sentence in the reference document. For example, a sentence of the summary that is corresponding to the fourth sentence of the reference document will place before the other sentence of the summary that is corresponding to the twentieth sentence of the reference document. Applying these two strategies will guarantee the cohesion and coherence of the sentences in the final summary.

4. Experiment and evaluation

In this section, we present the conducted experiments and evaluation results of our proposed multi-document summarization system on DUC 2005 and DUC 2007 datasets. First, We briefly introduce these two datasets in the following subsection. Then, the evaluation results of our proposed system on these datasets and comparison with several state-of-the-art multi-document summarization methods which have achieved the best result on the DUC 2005 and 2007 datasets in recent years are presented.

Given that the documents of the employed datasets are presented in XML format, each section of the text has a specified tag. Therefore, first of all, each XML document is parsed and different sections corresponding to a specific tag is extracted. After extracting the news text from all XML documents corresponding to the same topic, the text of these documents are aggregated into one large input text which the sentences and words of it are divided into individual sentences and words respectively using NLTK library (Bird et al., 2009). In the next step, the stopwords of each sentence are removed using a list which contains 214 most common English words. The word stemming step is performed using the Porter stemming algorithm which is implemented in NLTK. Furthermore, the spaCy[1] library is used for part-of-speech tagging. The value of the heuristic slope t is set to -0.6, based on our experiments on validation set, for the evaluation of our proposed system.

4.1. Dataset

The Document Understanding Conference (DUC) is a series of annual conferences aimed at evaluating automatic text summarization systems. This series of conferences is run by the National Institute of Standards and Technology (NIST) to make progress in automatic text summarization and enable researchers to participate in a prestigious scientific event.

The DUC 2005 and DUC 2007 datasets are both released by the DUC conference in 2005 and 2007, respectively. They are two of the most widely used open benchmark datasets for multi-document summarization which are specifically presented for evaluation of multi-document summarization systems (Dang, 2006).

DUC 2005 consists of 50 topics, each of which contains 25–50 documents, while DUC 2007 includes 45 topics with 25 documents in each topic. Each document relates to a news report from various news agencies such as The Associated Press, New York Times, and The Wall Street Journal. In these datasets, the reference summaries generated by human assessors for the evaluation of the multi-document summarization systems comprise approximately 250 words. Therefore, to be able to evaluate our proposed system using the available reference summaries, the summary which is generated by the system must consist maximum of 250 words. In DUC 2005, the number of reference summaries varies for several topics in which 30 topics include 4 reference summaries and remaining 20 topics contain 9 or 10 reference summaries. Since ROUGE evaluation method depends on the number of reference summaries, NIST removed some of the reference summaries for 20 topics with 10 reference summaries in order to simplify the evaluation, so that all the topics of DUC 2005 have exactly 9 reference summaries each. However, all topics of DUC 2007 contains 4 human summaries created by NIST assessors.

It is important to note that the main task of both datasets is query-focused text summarization, that is, each topic is associated with a query which represents the information need (Dang, 2006). However, our summarization system is a generic summarization approach. Therefore, in all of our experiments, we generate the final summary without considering the query provided by the dataset. Table 1 gives a brief description of the DUC 2005 and DUC 2007 datasets.

Table 1
A brief description of the DUC 2005 and DUC 2007 datasets.

Feature	DUC 2005	DUC 2007
Number of topics	50	45
Number of documents per topic	25–50	25
Total number of documents	1593	1125
Data source	TREC	AQUAINT
Reference summary length	250 words	250 words

4.2. Evaluation metric

In order to evaluate the performance of our proposed system, we used ROUGE (Recall-Oriented Understudy for Gisting Evaluation) toolkit version 2.0 (Ganesan, 2018), which is the most widely used tool for evaluating automatic summarization tasks. ROUGE has been adopted by DUC as the official evaluation method to measure the quality of system-generated summaries. ROUGE toolkit measures the quality of a system-generated summary through counting the number of overlapping units, such as N-grams, between the generated summary and a set of reference (human-created) summaries. In all experiments, we use three standard evaluation metrics of ROUGE toolkit, namely ROUGE-1, ROUGE-2, and ROUGE-SU4 for evaluating our summarization system and comparing it with some of the state-of-the-art methods, since they have been considered as the official metrics in the DUC 2005 dataset. ROUGE-1 and ROUGE-2 compare the unigram and bigram overlap between the system-generated summary and the reference summaries created by human, respectively. These two metrics are a special case of ROUGE-N which can be calculated using ROUGE-N formula as follows (Lin, 2004):

$$ROUGE-N = \frac{\sum_{S\in Referencesummary} \sum_{N-gram\in S} Count_{match}\left(N-gram\right)}{\sum_{S\in Referencesummary} \sum_{N-gram\in S} Count\left(N-gram\right)} \quad (24)$$

where N stands for the length of the N-gram, $Count_{match}(N-gram)$ represents the maximum number of matching N-grams co-occurring in a system-generated summary and a set of reference summaries, and $Count(N-gram)$ is the number of N-grams occurring in the reference summaries. Given that the denominator of the Eq. 24 is the total number of N-grams occurring in the reference summaries, ROUGE-N is a recall-oriented measure.

ROUGE-SU (skip-bigram co-occurrence averaged with unigram co-occurrence) is the latest metric introduced in ROUGE toolkit which is an extension of ROUGE-S (skip-bigram co-occurrence). Skip-bigram in ROUGE-S and ROUGE-SU is defined as any pair of words in their sentence order, allowing for arbitrary gaps (Lin, 2004). ROUGE-S measures the quality of a system-generated summary based on the number of overlapping skip-bigrams between the candidate summary and a set of reference summaries. The potential problem of ROUGE-S metric is that it does not give any score to a candidate sentence which does not have any skip-bigrams co-occurring with its references. In this case, according to the ROUGE-S metric, there is no difference between candidate sentences that don't have any common skip-bigrams with reference summaries and the candidate sentences which have no co-occurring unigrams in addition to skip-bigrams with reference summaries. To solve this problem, ROUGE-SU extends ROUGE-S with considering the number of common unigrams as a counting unit (Lin, 2004).

4.3. Evaluation and comparison

In this research, we apply three different objective functions as described in section 3.2.6 in order to evaluate the impact of each objective function on the performance of our proposed system on DUC 2005 and select the best one. Then, we employ our summarizer with the best objective function on DUC 2007 to ensure its performance. The first objective function is a combination of word frequency and sentence position metrics. The second objective function consists of three sentence scoring measures: coverage (cosine similarity between all sentences of the input text and the candidate summary), the cosine similarity between the candidate summary and titles of the input documents, and sentence length. Finally, the third objective function is the combination of 6 sentence scoring measures which is described in Eq. 18.

As QIGA is a stochastic algorithm, we produce 10 summaries for each topic to make sure of the proposed system. Thus the score of each topic reported here is averaged over the corresponding 10 generated summaries. ROUGE toolkit reports results in terms of precision, recall, and F-score based on the selected evaluation metrics (ROUGE-1, ROUGE-2, and ROUGE-SU4). Thus, we use F-score which is the harmonic average of the precision and recall to compare the performance of our proposed system with several state-of-the-art methods which have

[1] https://spacy.io

achieved the best results on DUC 2005 and DUC 2007 datasets.

One of the impressive control parameters in the process of evolutionary algorithms is the population size. In our method, we consider three different population sizes to investigate the effect of this control parameter on the performance of the proposed system. Fig. 8–10 show the results of our summarization system on DUC 2005 based on the intended population sizes in conjunction with first, second, and third (main) objective functions, respectively.

As shown in Figs. 8,9, when the population size is initialized using Eq. 6, the summarization system provides better results compared to the other two situations with the population size of $\frac{n}{4}$ and $\frac{n}{10}$. According to the results presented in Fig. 9, only in one situation in which the proposed system uses the second objective function, the performance of the proposed QIGA with population size of $\frac{n}{4}$ is better than the transformed value. We have also presented the results of our proposed system in term of ROUGE-SU4 evaluation metric based on different population sizes in conjunction with the presented objective functions for QIGA in Fig. 10.

Also, it can be seen in Fig. 10 that the performance of the presented quantum-inspired genetic algorithm is better than other scenarios when using third objective function along with transformed value in $[50, 100]$ interval using Eq. 6 for population size. Thus, we have initialized the population size using Eq. 6 and used the third objective function as the main fitness function in our proposed QIGA. The evaluation results of our proposed summarization system formed using the third objective function alongside the dynamic population size on DUC 2005 are shown in Table 2.

Based on the results obtained from the experiments on DUC 2005, the third objective function alongside dynamic population size indicates better performance. In order to guarantee the quality of our solution, we have evaluated it on DUC 2007. Table 3 and Fig. 11 demonstrates the results achieved during our experiments on DUC 2007 where the third objective function alongside different population sizes are employed.

In the remaining of this section, we compare the performance of the proposed summarization system on DUC 2005 and DUC 2007 with several prominent multi-document extractive summarization approaches proposed in recent years which have achieved the best results on these datasets. First, we employ MTSQIGA on DUC 2005 dataset. Following it, we have compared our solution with other methods on DUC 2007 dataset to ensure its performance. The results of comparing our solution with other methods on DUC 2005 are demonstrated in Table 4. To do so, we briefly describe the methods used in the comparison at first:

- DUC 2005 best participant system (S15) (Ye, Qiu, Chua, & Kan, 2005): Each year, several automatic text summarization systems were participated in the Document Understanding Conference (DUC) series to be evaluated in large-scale experiments. At the end each conference, the participating systems were ranked and the results of each system were reported by the conference program committee based on the evaluation process. One of the approaches used to compare with the proposed system is the top method with the highest ROUGE score participated in DUC 2005. This approach extracts salient sentences by applying a modified version of Maximal Marginal Relevance (MMR) method in which a semantic similarity measure named concept link is used to rank sentences based on four factors from WordNet: synonyms, hyponyms, derivational morphological variations, and Inflectional morphological variations.
- CBS (Radev, Jing, Styś, & Tam, 2004): Centroid-based summarization method used in MEAD summarizer to extract salient sentences. This method summarizes clusters of documents using information from the centroids of the clusters. A centroid is a set of most important words of a cluster used for extracting salient sentences that are most relevant to the general topic of the entire cluster. In CBS, the score of each sentence of a cluster is computed using a weighted linear combination of three features: centroid value, positional value, and First-sentence overlap.
- PLSA (Hennig, 2009): This method utilizes probabilistic latent semantic analysis (PLSA) technique to produce a query-focused summary. PLSA is a latent variable model for co-occurrence data which relates a new unobserved class variable with each observed ones. To achieve this goal, a PLSA model is trained on the term-sentence matrix including all sentences of the corpus and queries. Then, the system associates sentences of the input documents and queries with a representation in the latent topic space which is provided by the trained PLSA model. In the next step, five sentence-level features are used to compute the similarity of sentences and query over latent topic space. Finally, it ranks sentences based on the score obtained from the linear combination of individual feature scores in order to produce the summary.
- SRSum (Ren et al., 2018): SRSum is a hybrid neural summarization model which uses sentence relation features for sentence scoring phase. This model consists of five deep neural network sub-models to automatically learn sentence relation features from training data, namely PriorSum, SFSum, CSRSum, TSRSum, and QSRSum. Each sub-model is comprised of a kind of CNN, RNN or combination of

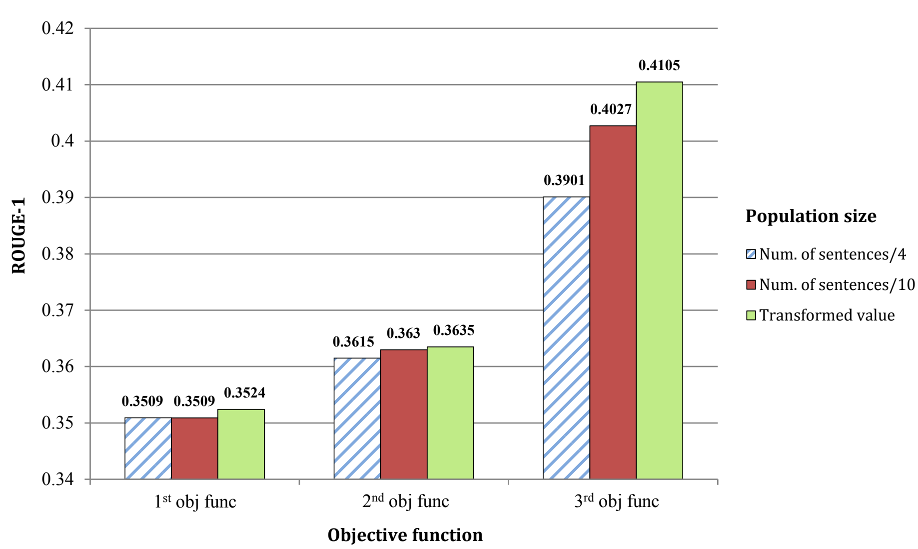

Fig. 8. The effect of various population size on the performance of our summarization system along with different objective functions on DUC 2005 based on ROUGE-1 evaluation metric. The 1st objective function consists of word frequency and sentence position. The 2nd one is a combination of coverage, cosine similarity with title, and sentence length. The 3rd objective function is constituted of six sentence scoring measures based on Eq. 18.

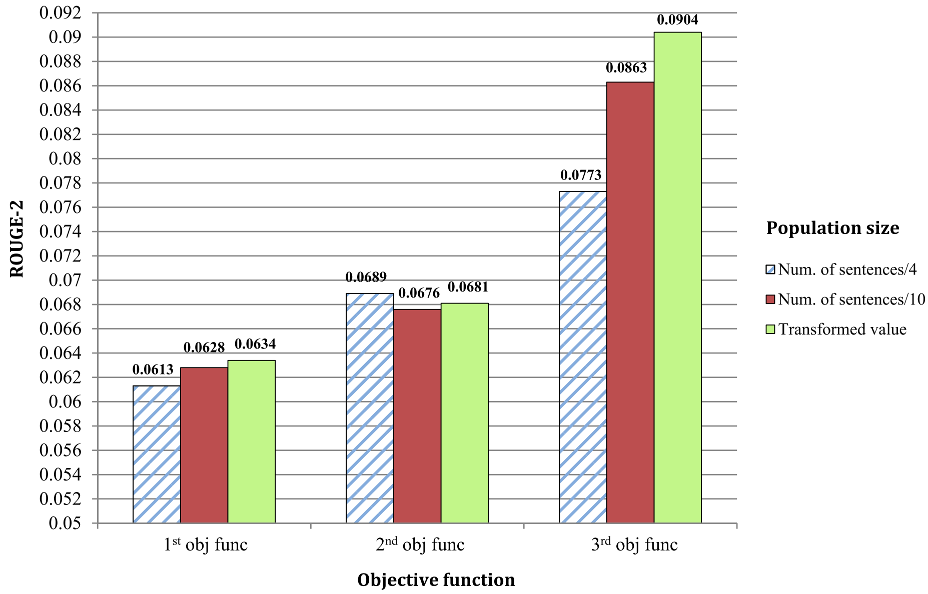

Fig. 9. The effect of various population size on the performance of our summarization system along with different objective functions on DUC 2005 based on ROUGE-2 evaluation metric. The 1st objective function consists of word frequency and sentence position. The 2nd one a combination of coverage, cosine similarity with title, and sentence length. The 3rd objective function is constituted of six sentence scoring measures based on Eq. 18.

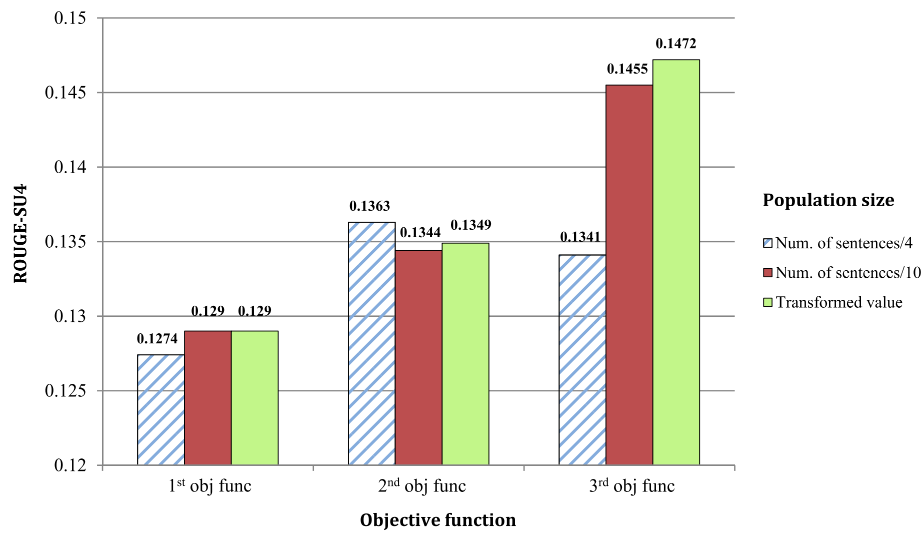

Fig. 10. The effect of population size on the performance of our summarization system along with different objective functions on DUC 2005 based on ROUGE-SU4 evaluation metric. The 1st objective function consists of word frequency and sentence position. The 2nd one is a combination of coverage, cosine similarity with title, and sentence length. The 3rd objective function is constituted of six sentence scoring measures based on Eq. 18.

Table 2
ROUGE scores of our proposed method with main objective function and transformed value for population size on DUC 2005.

Metric	ROUGE-1	ROUGE-2	ROUGE-SU4
Recall	0.4141	0.0905	0.1488
Precision	0.4071	0.0893	0.1457
F-score	0.4106	0.0898	0.1472

Table 3
ROUGE scores of our proposed method with main objective function and transformed value for population size on DUC 2007.

Metric	ROUGE-1	ROUGE-2	ROUGE-SU4
Recall	0.4779	0.1289	0.1892
Precision	0.4757	0.1286	0.1879
F-score	0.4767	0.1287	0.1885

them to measure one of the sentence relation features. Then, SRSum applies a MLP with three hidden layer as a decoder to transfer the outputs of these sub-models into a unified value as the final score of each main body sentence. Finally, the greedy algorithm is used to select summary sentences based on their score.

- CDDS (Alguliev et al., 2013): Constraint-driven document summarization emphasizes on generating summaries with maximum content coverage and high diversity by utilizing user-provided constraints and tune the constraint parameters. In CDDN, content coverage and diversity are measured by the similarity of the summary to the document and sum of a pairwise similarity within the

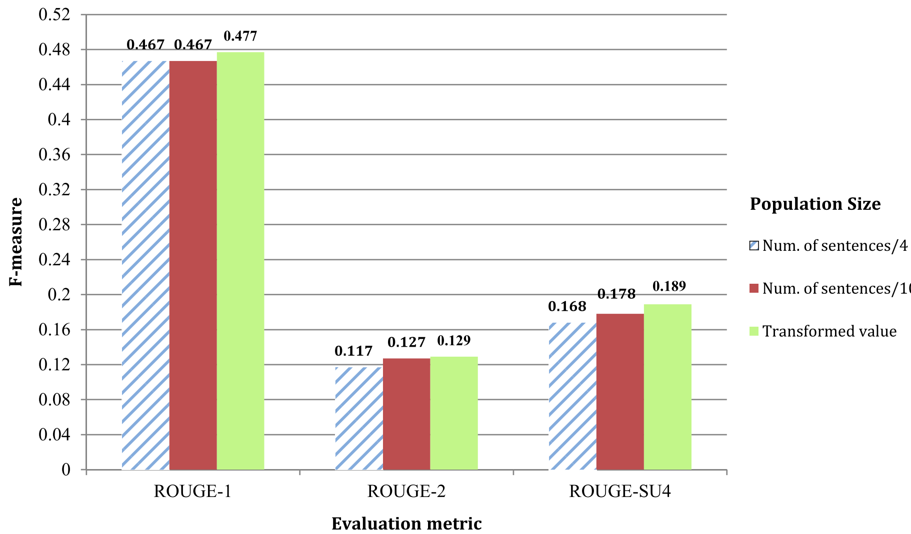

Fig. 11. The impact of population size on the performance of the proposed summarizer on DUC 2007. The main objective function shown in Eq. 18 has been used in this experiment.

Table 4
Comparison of our proposed method with main objective function and dynamic value as population size with other solutions on DUC 2005.

Method	ROUGE-1	ROUGE-2	ROUGE-SU4
MTSQIGA	**0.4106** (1)	**0.0898** (1)	**0.1472** (1)
MA-MultiSumm	0.4001 (2)	0.0868 (2)	0.1434 (2)
SRSum	0.3983 (3)	0.0857 (3)	–
CDDS	0.3956 (4)	0.0852 (4)	0.1434 (2)
DESAMC + DocSum	0.3937 (5)	0.0822 (6)	0.1418 (3)
MA&MR	0.3932 (6)	0.0834 (5)	0.1411 (4)
PLSA	0.3913 (7)	0.0811 (7)	0.1389 (7)
LFIPP	0.3905 (8)	0.0804 (8)	0.1403 (5)
MCMR	0.3891 (9)	0.079 (9)	0.1392 (6)
S15	0.3765 (10)	0.0738 (10)	0.1324 (8)
TMR	0.3775 (11)	0.0715 (11)	0.1304 (9)
CBS	0.3535 (12)	0.0638 (12)	0.1198 (10)
SNMF + SLSS	0.3501 (13)	0.0604 (13)	0.1172 (11)

candidate sentences of summary with each other, respectively. This approach is formulated as a quadratic integer programming problem in which binary PSO is applied to solve it.

- MR&MR (Alguliev et al., 2015): This summarization approach applies a combination of symmetric and asymmetric similarity measures which consists of weighted harmonic mean of cosine similarity as symmetric measure and overlap similarity as asymmetric measure. Then, it addresses text summarization task as a boolean programming problem where the differential evolution (DE) algorithm is used to optimize the problem.
- LFIPP (Alguliev et al., 2011): LFIPP models multi-document text summarization as a discrete optimization problem where an adaptive differential evolution algorithm is used to solve it. The model formulates sentence-extraction-based summarization as a linear fraction of sentence similarities in which it takes into account both the sentence-to-document similarity and the sentence-to-sentence similarity in order to select salient sentences of the input document collection and reduce redundancy in the generated summary, respectively. To solve the optimization problem, an adaptive DE is proposed in this study that applies a new trial vector generation strategy and an adaptive crossover rate. In this adaptive DE algorithm, the crossover rate is steadily adjusted according to the current generation and the maximum number of generations.
- TMR (Tang, Yao, & Chen, 2009): TMR (Topic Modeling with Regularization) is aimed at solving the problem of multi-topic based query-oriented summarization by employing a regularization framework to constrain and guide the topic model. The approach contains two sub-processes which are trained simultaneously. After the topic distribution of documents and the query is estimated in the first stage, the topic distribution is adjusted towards the query-specific topic distribution in the next phase. Actually, the purpose of these two sub-processes is minimizing the regularized data likelihood.
- SNMF + SLSS (Wang et al., 2008): This multi-document summarization approach consists of two main phases. In the first phase, the sentence similarity matrix is constructed using the sentence semantic similarity which is calculated based on both the semantic role analysis and word relation discovery. Having created pairwise sentence similarity matrix, the sentences of the document collection are clustered into groups by applying symmetric matrix factorization in the second phase. Finally, the most semantically important sentence of each cluster is selected based on both the internal and external information related to the cluster.
- MA-MultiSumm (Mendoza et al., 2014): MA-MultiSumm, which is a multi-document summarizer, employs a memetic algorithm based on CHC (Cross-generational elitist selection, Heterogeneous recombination, Cataclysmic mutation) evolutionary algorithm and a greedy local search to generate a generic extractive summary from source document collection. The role of local search operator in this approach is to optimize the offspring generated using half uniform crossover (HUX). This will help the memetic algorithm improve fitness value of its individuals for the next generation. In this method, the objective function has been defined as a linear combination of coverage and redundancy factors.
- DESAMC + DocSum (Alguliev et al., 2012): In this method, generic multi-document summarization is modeled as a modified p-median problem. The problem is formulated with taking into consideration four necessary features of an appropriate summary including relevance, information coverage, diversity in the summary, and length limit. A modified differential evolution algorithm, called DESAMC, is employed to solve the optimization problem. This DE algorithm uses a self-adaptive crossover and mutation strategies in which it can dynamically tune its crossover rate and mutation factor.
- MCMR (Alguliev et al., 2011): MCMR, that is an unsupervised method, provides a linear model for generic text summarization. In particular, it represents multi-document text summarization as an integer linear programming (ILP) problem where it attempts to optimize both coverage and redundancy simultaneously. This

method applies two optimization algorithms, namely branch-and-bound (B&B) and particle swarm optimization (PSO) algorithms to globally solve the problem.

- CAM-EK (Ya, Liu, & Guo, 2020): This query-focused summarization method presents a fine-grained and interactive word-by-word attention model which utilizes Compare-Aggregate framework to capture intent of query and fill the expression gap between query and document. This framework consists of 4 layers including word representation layer, attention layer, comparison layer, and aggregation layer, respectively. It also applies conceptual knowledge to improve the attention model.
- HyperSum (Wang, Li, Li, Li, & Wei, 2013): This solution employs hypergraph theory along with a semi-supervised sentence ranking algorithm for query-focused summarization. The nodes of the hypergraph in this model are sentences and the hyperedges can capture both pairwise relation as well as group relation between a topic and its subtopics, each of which is described by a group of sentences.
- TL-TranSum (Van Lierde & Chow, 2019): TL-TranSum is a hypergraph-based query-focused summarization method that takes into account a topic model based on semantic clustering of terms. Sentences forms the nodes of the graph, while the hyperedges of this graph are constituted by the themes grouping sentences covering the same topics. Finally, the most relevant sentences to the query are selected based on generating a traversal of the nodes.
- PPRSum (Liu, Wang, Zhang, & Xu, 2008): Liu et al. proposed a unified framework for query-focused extractive summarization based on Personalized PageRank method. This approach takes into account multiple features to select salient sentences of the corpus. At first, it trains a model based on sentence global information using Naïve Bayes. Then it generates a relevance model based on the query of each corpus. Finally, this approach computes Personalized Prior Probability based on both provided models and applies PPR to select most relevant sentences.

Further we compared our approach with other methods on DUC 2007 to ensure its performance. The results of this comparison are revealed in Table 5. As it can be seen from the table, our proposed solution shows better result in comparison with other methods in terms of ROUGE-1 and ROUGE-SU4 metrics.

According to the results shown in Table 4, our proposed summarizer (MTSQIGA) outperforms all other approaches which have been presented in recent years for multi-document text summarization on DUC 2005 dataset, that is, it demonstrates better performance in comparison with other state-of-the-art methods on the basis of all three evaluation metrics including ROUGE-1, ROUGE-2, and ROUGE-SU4. The results also represents better performance of our solution on DUC 2007 in terms of ROUGE-1 and ROUGE-SU4 evaluation metrics. This performance is due to several features of our system based on the analysis. First, MTSQIGA takes both sentence-to-document and sentence-to-topic similarities into account which leads to maximize the coverage of important and relevant information presented in the document collection. It also removes candidate sentences containing redundant information by computing cosine similarity in the summary. In addition to cosine similarity, it considers common proper nouns between sentences of each document and all topics of the source cluster which is another reason for selecting more relevant and informative sentences.

Table 5
Comparison of our proposed method with main objective function and dynamic value as population size with other solutions on DUC 2007.

Method	ROUGE-1	ROUGE-2	ROUGE-SU4
MTSQIGA	**0.4767** (1)	**0.1287** (2)	**0.1885** (1)
CDDS	0.4719 (2)	0.1237 (4)	0.1836 (2)
MA&MR	0.4697 (3)	0.1222 (6)	0.1793 (4)
MCMR	0.4685 (4)	0.1221 (7)	0.1753 (6)
PPRSum	0.4673 (5)	0.1195 (8)	0.1710 (7)
LFIPP	0.4668 (6)	0.1192 (9)	0.1687 (8)
CAM-EK	0.4572 (7)	0.1227 (5)	–
PLSA	0.4540 (8)	0.1168 (10)	0.1768 (5)
SRSum	0.4501 (9)	0.1280 (3)	–
TL-TranSum	0.4238 (10)	**0.1299** (1)	0.1799 (3)
HyperSum	0.4205 (11)	0.1117 (11)	0.1659 (9)

Second, the weighted combination of sentence scoring measures used in the objective function of the proposed QIGA has significant impact in achieving quality extract summaries in comparison with other approaches. However, we only applies statistical and cosine similarity based measures in the presented combination, the weights of these sentence scoring methods are set in such efficient way that leads to generate better quality extract summaries in comparison with other semantic and machine learning based approaches such as PLSA, SNMF + SLSS, and SRSum, in some of which there are different levels of semantic similarity analysis.

Finally, the proposed QIGA with modified quantum measurement and adaptive quantum rotation gate, which is a powerful and scalable meta-heuristic evolutionary algorithm, plays the most important role in achieving the best results.Our proposed summarizer outperforms other approaches in maximum 500 evaluation of the objective function where other systems such as MA-MultiSumm and DESAMC + DocSum find the optimal solution in 1500 and 50000 evaluation of the objective function, respectively. Thus, with regard to the superposition of states in quantum-inspired genetic algorithm, our summarization system can easily be applied to a document collection with a large number of source documents. In fact, the application of adaptive QIGA has made it feasible to easily employ the proposed system in the real world scenarios where the number of source document in a specific topic is far high. MTSQIGA also shows that adaptive quantum-inspired genetic algorithm can provide better results in comparison with canonical evolutionary algorithms in ATS.

On the basis of the results shown in Table 4 and Table 5, our proposed method has achieved the state-of-the-art results in overall. The algorithm with the second place on DUC 2005 is MA-MultiSumm, which utilizes a combination of an evolutionary algorithm (CHC) and greedy local search. This method generates summaries after 1500 evaluations of the objective function, which is approximately three times higher than MTSQIGA. In other words, our proposed method with much lower costs generates the best results. Our method is even more successful than existing deep learning based summarization algorithms (i.e., SRSum is a deep learning based summarization approach which takes third place in Table 4). Our analysis demonstrates that its performance is resulted by the use of a multilayer perceptron (MLP) with five deep neural network sub-modules where each of them is responsible for taking into account a type of similarity between sentences based on word embeddings. The results presented in Table 5 also demonstrates that MTSQIGA has achieved significant improvement in terms of ROUGE-1 and ROUGE-SU4. Only in terms of ROUGE-2 metric, our proposed summarization system concedes to TL-TranSum as this query-oriented method considers semantic features of the words to model the topic in the document collection. As it can be seen in both tables, the performance of several methods are reported with '-' in ROUGE-SU4 column since no result is reported for them on this evaluation metric.

5. Conclusion

In this paper, we have proposed an automatic multi-document text summarization system, called MTSQIGA, which generates extract summaries using an unsupervised technique. The main contribution of this paper is to apply a tuned quantum-inspired genetic algorithm on automatic text summarization field. In our approach, extractive summarization is modeled as an optimization problem where a QIGA with modified quantum measurement and adaptive quantum rotation gate is employed to find the globally optimal solution. The algorithm is proposed with emphasis on designing an effective objective function and modifying recombination operators to enhance the quality of the generated summary and control its length, simultaneously. It takes into account sentence-to-document and sentence-to-title similarities while reducing redundancy using intra-sentence similarity of candidate sentences. According to the sentence scoring measures used in the objective function, MTSQIGA is a language-independent approach and therefore it can be used for summarization in any languages. As QIGA is a powerful and scalable meta-heuristic search technique, this approach can easily be applied to document clusters with a far high number of documents. Also, thanks to the characteristic of Q-bits, the quantum population size is always in $[50, 100]$ interval based on a mapping function. This leads to increase the performance and speed of our proposed system. In our summarizer, quantum measurement and quantum rotation gate operators have the most important role in controlling the summary length. They also have great impact on convergence of the algorithm to the best global solution.

We have evaluated our summarization system by means of DUC 2005 and DUC 2007 standard benchmark datasets which demonstrates state-of-the-art results. We have compared our approach with some of the existing state-of-the-art summarization systems in terms of ROUGE-1, ROUGE-2, and ROUGE-SU4. The results of comparison demonstrate that our system outperforms all other approaches in terms of almost all ROUGE evaluation metrics. This paper also exposes that a well-tuned QIGA can provide promising results in comparison with many other canonical evolutionary algorithm in text summarization field.

Considering that QIGA performance highly depends on the choice of its hyperparameters value, one of the most challenging parts in our proposed system is finding the best setting for hyperparameters of the QIGA. Also figuring out the best combination of weights for sentence scoring measures in the objective function is another challenging task that we can improve the performance of MTSQIGA by employing machine learning methods to find the best tuning for it in the future. Furthermore, as another direction for future work, we intend to apply our system on other datasets, specially domain-specific ones, with different languages since our summarization approach is independent of the language. We would also like to extend our proposed objective function with semantic based sentence scoring methods such as sentiment analysis and co-reference chains as they can improve the performance of our summarization system. Combining additional statistical measures with the current ones is another potential future direction.

CRediT authorship contribution statement

Mohammad Mojrian: Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing - original draft. **Seyed Abolghasem Mirroshandel:** Conceptualization, Methodology, Supervision, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Al-Saleh, A., Menai, M. E. B. (2018). Ant colony system for multi-document summarization. In: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA. pp. 734–744.

Alami, N., Adlouni, Y. E., En-nahnahi, N., Meknassi, M. (2018). Using statistical and semantic analysis for arabic text summarization. In: Proceedings of the International Conference on Information Technology and Communication Systems, Springer International Publishing, Cham. pp. 35–50.

Alguliev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., & Mehdiyev, C. A. (2011). MCMR: Maximum coverage and minimum redundant text summarization model. *Expert Systems with Applications, 38*, 14514–14522.

Alguliev, R. M., Aliguliyev, R. M., & Isazade, N. R. (2012). DESAMC+DocSum: Differential evolution with self-adaptive mutation and crossover parameters for multi-document summarization. *Knowledge-Based Systems, 36*, 21–38.

Alguliev, R. M., Aliguliyev, R. M., & Isazade, N. R. (2013). CDDS: Constraint-driven document summarization models. *Expert Systems with Applications, 40*, 458–465.

Alguliev, R. M., Aliguliyev, R. M., & Isazade, N. R. (2015). An unsupervised approach to generating generic summaries of documents. *Applied Soft Computing, 34*, 236–250.

Alguliev, R. M., Aliguliyev, R. M., & Mehdiyev, C. A. (2011). Sentence selection for generic document summarization using an adaptive differential evolution algorithm. *Swarm and Evolutionary Computation, 1*, 213–222.

Aliguliyev, R. M. (2009). A new sentence similarity measure and sentence based extractive technique for automatic text summarization. *Expert Systems with Applications, 36*, 7764–7772.

Amancio, D. R., Nunes, M. G., Oliveira, O. N., da F. Costa, L. (2012). Extractive summarization using complex networks and syntactic dependency. Physica A: Statistical Mechanics and its Applications 391, 1855–1864.

Antiqueira, L., Oliveira, O. N., da Fontoura Costa, L., das Graças Volpe Nunes, M. (2009). A complex network approach to text summarization. Information Sciences 179, 584–599.

Aries, A., Zegour, D. E., Hidouci, W. (2019). Automatic text summarization: What has been done and what has to be done. arXiv preprint arXiv:1803.01937.

Baralis, E., Cagliero, L., Mahoto, N., & Fiori, A. (2013). GraphSum: Discovering correlations among multiple terms for graph-based summarization. *Information Sciences, 249*, 96–109.

Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media. first edition. pp. 1–128.

Bonissone, P. P., Subbu, R., Eklund, N., & Kiehl, T. R. (2006). Evolutionary algorithms + domain knowledge = real-world evolutionary computation. *IEEE Transactions on Evolutionary Computation, 10*, 256–280.

Canhasi, E., & Kononenko, I. (2016). Automatic extractive multi-document summarization based on archetypal analysis. In *Non-negative Matrix Factorization Techniques: Advances in Theory and Applications* (pp. 75–88). Berlin Heidelberg: Springer.

Christensen, J., Mausam, Soderland, S., Etzioni, O. (2013). Towards coherent multi-document summarization. In: Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies, Atlanta, Georgia. pp. 1163–1173.

Dahi, Z. A. E. M., Mezioud, C., & Draa, A. (2016). A quantum-inspired genetic algorithm for solving the antenna positioning problem. *Swarm and Evolutionary Computation, 31*, 24–63.

Dang, H. T. (2006). DUC 2005: Evaluation of question-focused summarization systems. In: Proceedings of the Workshop on Task-Focused Summarization and Question Answering, Association for Computational Linguistics, Sydney, Australia. pp. 48–55.

Fattah, M. A., & Ren, F. (2009). GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *Computer Speech & Language, 23*, 126–144.

Ferreira, R., de Souza Cabral, L., Freitas, F., Lins, R. D., de França Silva, G., Simske, S. J., & Favaro, L. (2014). A multi-document summarization system based on statistics and linguistic treatment. *Expert Systems with Applications, 41*, 5780–5787.

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review, 47*, 1–66.

Ganesan, K. (2018). ROUGE 2.0: updated and improved measures for evaluation of summarization tasks. arXiv preprint arXiv:1803.01937.

García-Hernández, R. A., & Ledeneva, Y. (2013). Single extractive text summarization based on a genetic algorithm. In *Pattern Recognition* (pp. 374–383). Berlin, Heidelberg: Springer.

Han, K. H., & Kim, J. H. (2000). Genetic quantum algorithm and its application to combinatorial optimization problem. In *Proceedings of the 2000 Congress on Evolutionary Computation* (pp. 1354–1360). La Jolla, CA, USA: IEEE.

Han, K. H., Park, K. H., Lee, C. H., & Kim, J. H. (2001). Parallel quantum-inspired genetic algorithm for combinatorial optimization problem, in. In *Proceedings of the 2001 Congress on Evolutionary Computation, IEEE* (pp. 1422–1429).

Hennig, L. (2009). Topic-based multi-document summarization with probabilistic latent semantic analysis. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing* (pp. 144–149). Borovets, Bulgaria: Association for Computational Linguistics.

Hey, T. (1999). Quantum computing: An introduction. *Computing & Control Engineering Journal, 10*, 105–112.

Isonuma, M., Fujino, T., Mori, J., Matsuo, Y., & Sakata, I. (2017). Extractive summarization using multi-task learning with document classification. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2101–2110). Copenhagen, Denmark: Association for Computational Linguistics.

Joshi, A., Fidalgo, E., Alegre, E., & Fernández-Robles, L. (2019). Summcoder: An unsupervised framework for extractive text summarization based on deep auto-encoders. *Expert Systems with Applications, 129*, 200–215.

Junan, Y., Bin, L., & Zhenquan, Z. (2003). Research of quantum genetic algorithm and its application in blind source separation. *Journal of Electronics (China), 20*, 62–68.

Li, B. B., & Wang, L. (2007). A hybrid quantum-inspired genetic algorithm for multiobjective flow shop scheduling. *IEEE Transactions on Systems, Man, and Cybernetics Part B (Cybernetics), 37*, 576–591.

Liao, G. C. (2012). Solve environmental economic dispatch of smart microgrid containing distributed generation system – Using chaotic quantum genetic algorithm. *International Journal of Electrical Power & Energy Systems, 43*, 779–787.

Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In: Proceedings of the ACL 2004 Workshop on Text Summarization Branches Out, Association for Computational Linguistics, Barcelona, Spain. pp. 74–81.

Lin, C. Y., Hovy, E. H. (1997). Identifying topics by position. In: Proceedings of the 5th conference on applied natural language processing (ANLP), San Francisco, CA, USA. pp. 283–290.

Lita, L. V., Ittycheriah, A., Roukos, S., Kambhatla, N. (2003). tRuEcasIng. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics, Association for Computational Linguistics, Sapporo, Japan. pp. 152–159.

Litvak, M., Last, M., Friedman, M. (2010). A new approach to improving multilingual summarization using a genetic algorithm. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Uppsala, Sweden. pp. 927–936.

Liu, Y., Wang, X., Zhang, J., & Xu, H. (2008). Personalized pagerank based multi-document summarization. *IEEE International Workshop on Semantic Computing and Systems*, 169–173.

Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development, 2*, 159–165.

Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2019). Graph-based text summarization using modified TextRank. In *Soft Computing in Data Analytics* (pp. 137–146). Springer Singapore.

Malossini, A., Blanzieri, E., & Calarco, T. (2008). Quantum genetic optimization. *IEEE Transactions on Evolutionary Computation, 12*, 231–241.

Mani, I. (1999). *Advances in Automatic Text Summarization*. Cambridge, MA, USA: MIT Press.

Mendoza, D., Cobos, C., León, E., Lozano, M., Rodríguez, F., & Herrera-Viedma, E. (2014). A new memetic algorithm for multi-document summarization based on CHC algorithm and greedy search. In *Human-Inspired Computing and Its Applications* (pp. 125–138). Cham: Springer International Publishing.

Mendoza, M., Bonilla, S., Noguera, C., Cobos, C., & León, E. (2014). Extractive single-document summarization based on genetic operators and guided local search. *Expert Systems with Applications, 41*, 4158–4169.

Mogren, O., Kågebäck, M., Dubhashi, D. (2015). Extractive summarization by aggregating multiple similarities. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, INCOMA Ltd. Shoumen, Hissar, Bulgaria. pp. 451–457.

Mutlu, B., Sezer, E. A., & Akcayol, M. A. (2020). Candidate sentence selection for extractive text summarization. *Information Processing & Management, 57*, Article 102359.

Narayanan, A., & Moore, M. (1996). Quantum-inspired genetic algorithms. *Proceedings of IEEE International Conference on Evolutionary Computation*, 61–66.

Peyrard, M., Eckle-Kohler, J. (2016). A general optimization framework for multi-document summarization using genetic algorithms and swarm intelligence. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, The COLING 2016 Organizing Committee, Osaka, Japan. pp. 247–257.

Peyrard, M., Gurevych, I. (2018). Objective function learning to match human judgements for optimization-based summarization. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), Association for Computational Linguistics, New Orleans, Louisiana. pp. 654–660.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*, 130–137.

Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. *Information Processing & Management, 40*, 919–938.

Ren, P., Chen, Z., Ren, Z., Wei, F., Nie, L., Ma, J., & de Rijke, M. (2018). Sentence relations for extractive summarization with deep neural networks. *ACM Transactions on Information Systems, 36*, 1–32.

Ribaldo, R., Akabane, A. T., Rino, L. H. M., & Pardo, T. A. S. (2012). Graph-based methods for multi-document summarization: Exploring relationship maps, complex networks and discourse information. In *Computational Processing of the Portuguese Language* (pp. 260–271). Berlin Heidelberg, Berlin, Heidelberg: Springer.

Rylander, B., Soule, T., Foster, J., Alves-Foss, J. (2001). Quantum evolutionary programming. In: Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation (GECCO'01), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. pp. 1005–1011.

Saini, N., Saha, S., Jangra, A., & Bhattacharyya, P. (2019). Extractive single document summarization using multi-objective optimization: Exploring self-organized differential evolution, grey wolf optimizer and water cycle algorithm. *Knowledge-Based Systems, 164*, 45–67.

Sanchez-Gomez, J. M., Vega-Rodríguez, M. A., & Pérez, C. J. (2018). Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach. *Knowledge-Based Systems, 159*, 1–8.

Sivanandam, S., & Deepa, S. N. (2007). *Introduction to Genetic Algorithms*. Berlin Heidelberg: Springer.

Suanmali, L., Salim, N., Binwahlan, M. S. (2011). Fuzzy genetic semantic based text summarization. In: Proceedings of the 2011 IEEE 9th International Conference on Dependable, Autonomic and Secure Computing, Sydney, Australia. pp. 1184–1191.

Talbi, H., Draa, A., Batouche, M. (2004). A new quantum-inspired genetic algorithm for solving the travelling salesman problem. In: 2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT '04., pp. 1192–1197.

Tang, J., Yao, L., Chen, D. (2009). Multi-topic based query-oriented summarization. In: Proceedings of the 9th SIAM International Conference on Data Mining, Nevada, USA. pp. 1148–1159.

Tohalino, J. V., & Amancio, D. R. (2018). Extractive multi-document summarization using multilayer networks. *Physica A: Statistical Mechanics and its Applications, 503*, 526–539.

Van Lierde, H., & Chow, T. W. (2019). Query-oriented text summarization based on hypergraph transversals. *Information Processing & Management, 56*, 1317–1338.

Verma, P., & Om, H. (2019). MCRMR: Maximum coverage and relevancy with minimal redundancy based multi-document summarization. *Expert Systems with Applications, 120*, 43–56.

Verma, P., & Om, H. (2019). A variable dimension optimization approach for text summarization. In *Harmony Search and Nature Inspired Optimization Algorithms* (pp. 687–696). Springer Singapore.

Wang, D., Li, T., Zhu, S., Ding, C. (2008). Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore. pp. 307–314.

Wang, W., Li, S., Li, J., Li, W., & Wei, F. (2013). Exploring hypergraph-based semi-supervised ranking for query-oriented summarization. *Information Sciences, 237*, 271–286.

Xiao, J., Yan, Y., Zhang, J., & Tang, Y. (2010). A quantum-inspired genetic algorithm for k-means clustering. *Expert Systems with Applications, 37*, 4966–4973.

Xu, J., Gan, Z., Cheng, Y., & Liu, J. (2020). Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5021–5031). Association for Computational Linguistics.

Xu, Y., Mei, X., Dai, Z., & Su, Q. (2014). Application of the improved quantum genetic algorithm. In *Computational Intelligence, Networked Systems and Their Applications* (pp. 122–128). Berlin, Heidelberg: Springer.

Ya, J., Liu, T., & Guo, L. (2020). A compare-aggregate model with external knowledge for query-focused summarization. In *Web Information Systems Engineering – WISE 2020* (pp. 68–83). Cham: Springer International Publishing.

Yadav, C. S., Sharan, A., Kumar, R., Biswas, P. (2016). A new approach for single text document summarization. In: Proceedings of the Second International Conference on Computer and Communication Technologies, Springer India, New Delhi. pp. 401–411.

Yasunaga, M., Zhang, R., Meelu, K., Pareek, A., Srinivasan, K., Radev, D. (2017). Graph-based neural multi-document summarization. In: Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017), Association for Computational Linguistics, Vancouver, Canada. pp. 452–462.

Ye, S., Qiu, L., Chua, T. S., Kan, M. Y. (2005). NUS at DUC 2005: Understanding documents via concept links. In: Proceedings of the Document Understanding Conference, Vancouver, Canada.

Zhang, G., Jin, W., Hu, L. (2003). A novel parallel quantum genetic algorithm. In: Proceedings of the 4th International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE. pp. 693–697.

Zhang, J., Li, H., Tang, Z., Lu, Q., Zheng, X., & Zhou, J. (2014). An improved quantum-inspired genetic algorithm for image multilevel thresholding segmentation. *Mathematical Problems in Engineering, 2014*, 1–12.

an Zhang, M., Deng, Y., Xia Chang, D. (2014). A novel quantum genetic clustering algorithm for data segmentation. In: Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation, ACM, New York, USA. pp. 1485–1486.

Zhang, X., Lapata, M., Wei, F., & Zhou, M. (2018). Neural latent extractive document summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 779–784). Brussels, Belgium: Association for Computational Linguistics.