A Combination Method for Improving Text Summarization

Zeinab Mashreghi
Department of Computer science, Kashan Branch, Islamic Azad University, Kashan, Iran

Mahdi Esmaeili
Department of Computer science, Kashan Branch, Islamic Azad University, Kashan, Iran

ABSTRACT
As the volume of online and electronic information increasingly has grown, quickly and accurately access to these important resources is a big challenge. Text analytics can help by transposing words and sentences in unstructured data into high-quality information. Text summarization is one of the applications of text mining, has been of interest to researchers. In addition to text summarization, using optimization algorithms can be influenced results. In this paper, has been presented a hybrid approach for English multi-document summarization. As name suggest, a text summarization system produces summary of original documents. Combination of text mining and optimization algorithms is main ways this research, to improve results and reduce redundancy in summary sentences and simultaneously summary sentences have the most relevant. Similarity measures are cosine and overlap. Using multi-objective particle swarm optimization algorithm improved results. The experimental results of the method on two data sets DUC2005 and DUC2007 show improvements in the three assessment criteria associated. Result of summarization about 3 percent in ROUGE-1, in ROUGE-2 at 2 percent and the benchmark ROUGE-SU for approximately 1.5% compared with the previous methods improves.

Keywords
Text Mining, Multi-Document Summarization, Multi-objective Particle Swarm Optimization

1. INTRODUCTION
With the increasing rate of text sources on the World Wide Web, the range of information available to users are added every day. Although, this rapid growth has many advantages, but it has problems too. The first challenge is how the user can achieve to its required information. The use of information retrieval system is one of the most common search methods and required information in today's world. The output of a routine recovery information system is a list of titles that each has been accompanied by a short explanation. Unfortunately, in most cases, these titles are very long, and checking all the retrieved documents is impossible. In general, the users check only the first few dozen documents and ignore the rest. Most users tend to have short query plan, so the users face with another challenge: how to select the useful information. Automatic text summarization system is a solution to problem. Three major advantages of summary automatic production by machine are:

- The size of the summary is controllable, it means that the machine can provide the summary due to the intended compression
- Its content is predictable
- It can be identified that each part of the summary is related to which part or parts of the original text.

In general, there are two main parts for summary: Extraction and Abstraction.

- Extraction is provided due to the statistical and intuitive criteria or the combination of these two. Since there are not syntactic and semantic changes in the production of these types of summaries, they can be considered as extraction of sentences.
- Abstraction is an interpretation of the original text. In producing the abstract, sentences concepts of the main text are overwritten in shorter form. For example, the sentence, "He ate apple, grape and cherries" can be written as, "He ate the fruits".

The main idea of text mining is to find small pieces of information from large volume of text data without having to read all of it. Summarization one application of text mining, reduces the amount of text in a document while maintaining its original meaning. Multi document summarization is very important in order to summarize multiple document with the same topic. There are two challenges in summarization of numerous documents: reaching to the lowest redundancy and maximum content coverage. Also, the maximum coherence is mentioned for some approaches. In the final stage of the summary, that is the time to select the best sentence for summarization, the global selection approach choose the best sentence that have the most important overlap and the lowest redundancy[1]. In this research has been presented an optimization method for minimum redundancy and maximum overlap in order to improve the text summary results. The rest of this article is organized as follows:

In the second section of literature, the work done in the summary text is examined. The third section is devoted to the presentation of the proposed solution. In the fourth section, the results are analyzed and evaluated. The fifth section is devoted to the conclusion of the materials presented and eventually the final section deals with the recommendations and future work in this area.

2. LITERATURE SURVEY
In recent years, the production of multi-document summarization has gained a lot of attention among researchers. The multi-document summarization system concentrates on the intensive production of documents with the features of the original documents. Often, text summarization technique with the use of extraction method (which the outstanding sentences in multiple documents are selected) are represented as a summary.

A multi-document summarization system has been developed based on the sentence extraction using the principle of vertex cover algorithm that selects automatically the related sentences which are covered by predominant concepts of incoming document [2]. This frame work represents the documents as a weighted undirected graph with sentences as the vertices and the similarity between the sentences as the edge weight between the corresponding vertices in the graph.

(Meena and Gopalani 2015) check the important techniques and also the method employed in automatic text summarization using genetic algorithm [3]. They also want to check growth and improvements in automatic text summarization techniques in implementing the technique of evolutionary algorithm.

An optimization-based approach represents document unsupervised to the automatic summarization. (Rasim et al. 2015) in the proposed method, text summarization is modeled as a Boolean programming problem [4]. They create a modified differential revolution in order to solve the optimization problem. This model generally attempts to optimize three properties, namely, (1) relevance: summary should contain informative textual units that are relevant to the user. (2) redundancy: summaries should not contain multiple textual units that convey the same information and (3) length: summary is bounded in length.

(Esther et al. 2014) make the summary produced by machine as a rough summary, and use the binomial distribution to identify the importance of each sentence in a rough summary [5]. The summary has been refined by removing the similar sentences, so that only informative sentences remain in a summary. The method of removing the presented redundancy to the summary obtained from an existing summary system with fuzzy-based summarization model is applied as a case study.

(Rasim et al. 2012) examine in another article, a multi-document summarization as an optimization problem that requires optimization more than one objective function simultaneously [6]. In this study, in making summary from multiple documents, there is an attempt to balance two objectives; content coverage and redundancy. Our goal is to examine three basic aspects of the problem. For example, designing an optimization model, optimization solution, and finding a solution for the best summary. Multi-document summarization model, as a QBP[1], where the objective is a harmonious fusion of content covering objectives and redundancy, will be solved. This problem has been solved by using binary evolutionary algorithm.

They consider the summary of a document as a multi-objective optimization problem, including four objective functions: the information coverage, importance, redundancy, and coherence of the text [7]. These functions identify the possible functions based on the core word, and measure the main subject.

The magazine has also been discussed in articles on topics summarize text:

Automated Summarization of the text is now become an important aspect as it makes the meaning of documents easy to understand and easy to read. Automated summarization is the process of decreasing a text document with a computer system to be able to develop a synopsis that retains the main points associated with document this is certainly initial. Once the irritating dilemma of information overload is continuing to grow, and as the total amount of data has increased, so has fascination with automated summarization. A typical example of the application of summarization technology such as for example Bing and Document summarization is another. There are number of clustering algorithms which have been used in the past as clustering plays significant role in summarizing of the documents. Yadav and Singh in 2016 discussed about the existing clustering algorithms. In This paper also has been proposed a hybridized algorithm based on the combination of fuzzy C-Means and Particle Swarm Optimization. In the last, they compared their proposed algorithm results with the existing clustering algorithms [8].

3. THE PROPOSED METHOD

In this part, first an explanation will be given about the particle swarm optimization algorithm along with its advantages and disadvantages. Then the variables and formulas used in the proposed method will be examined. At last, the proposed method steps will be examined.

3.1 Particle Swarm Optimization Algorithm

Particle swarm optimization algorithm[2] is a social search algorithm, which is modeled on the social behavior of the birds group. At first, this model was used in order to discover the patterns that govern the flights of the birds and their sudden change of direction and the optimal shape of the group. In PSO, the particles flow in the search space. Changing the location of particles in search space is influenced by experience and knowledge of themselves and their neighbors. So, the position of particle mass affects on how to find a particle. The modeling result of this social behavior, is a search process, in which the particles tend toward the successful areas. The particles learn from each other and go to the best of their neighbors, based on the knowledge gained. The basis of PSO is: each particle adjusts its location in the search space according to the best place it has ever been or the best place in the whole of its neighborhood. The algorithm flowchart of PSO is as follows:

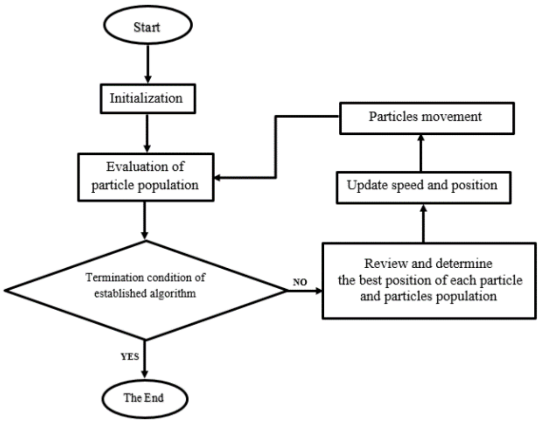

Figure 1: algorithm flowchart of PSO

This algorithm has some advantages compared to the other optimization algorithm. The followings are some samples:

- Memory usage
- Cooperation and information sharing between particles
- High-speed convergence
- Better flexibility against the optimum problem
- Easily implement and run

[1] QBP(Quadratic Boolean Programing)

[2] PSO(Particle Swarm Optimization)

Besides the benefits and features of the particle swarm optimization algorithm, this algorithm has a series of limitations and disadvantages that affects its performance.

- premature convergence
- Stuck in a local optimum
- Reduction of population diversity

3.2 The Formulation of the Issue

In this part, variables and formulas used in the proposed method are examined. Table 1 introduces the variables used in TS-MOPSO method. Variable N introduces the number of documents. So the documents are known as an array respectively. $D=\{D_1, D_2, \ldots ,D_N\}$. The array of a title of each document is known as Title, that is known as $Title=\{Title_1,Title_2,...,Title_N\}$. The number of sentences in each document is different. The number of sentences has been defined by the variable M. matrix S will be the sentences array of each document that is defined as $S=\{S_1,S_2,\ldots,S_M\}$. The number of terms are specified by K in each sentence. So the array of terms is considered as $T=\{T_1,T_2,\ldots.T_K\}$. There is a value for each word based on the repetition that is known as W_T_i. Document criteria are actually the sentences criteria contained in the document. Sentences quantitative criteria in each document are placed in the matrix called F, that is used as $F=\{F_1,F_2,\ldots.F_M\}$. Because there are three quantitative indexes for each sentence, each F_i is known as $F_i=\{Len_i,Pos_i,W\_Si\}$. There are two similarity criteria. The similarity of sentence i sentence is known as Sim_Title_i variable. Sentences overlapping are stored in Overlap matrix. That, the overlapping of i sentence with j is known by Overlap (i, j) variable.

Table 1: Variables and their definitions

Name	Definition
$\boldsymbol{N}$	number of documents
$\boldsymbol{D}$	documents
$\boldsymbol{D_i}$	i document
$\boldsymbol{M}$	number of sentences in each document
$\boldsymbol{S}$	sentences
$\boldsymbol{S_i}$	i sentence
$\boldsymbol{K}$	number of terms in each sentence
$\boldsymbol{T}$	Terms
$\boldsymbol{T_i}$	i term
$\boldsymbol{W\_T_i}$	weight of i term
$\boldsymbol{Title}$	titles
$\boldsymbol{Title_i}$	i title
$\boldsymbol{Len_i}$	length value quantitative parameter of i sentence
$\boldsymbol{Pos_i}$	position quantitative parameter of i sentence
$\boldsymbol{W\_S_i}$	weight quantitative parameter of i sentence
$\boldsymbol{Sim\_Title_i}$	i sentence similarity parameter with title
$\boldsymbol{Overlap\ (i,j)}$	i and j sentence overlapping

3.2.1 Extraction of Document Criteria

Sentences criteria of each document is divided into two quantity and similarity categories. Quantity criteria are independent and particular to the same sentence, but similarity criteria are calculated dependently. These criteria state the relationship between two sentences. Specifically, keeping these criteria determines the similarities of sentences. Based on the quantitative criteria, it can be determined which of the two similar sentences are more important.

3.2.1.1 Quantitative Criteria

By using this kind of criteria, scoring text sentences is done by using three features. These scores are very effective in selecting sentences.

- Sentence length criteria

According to the fact that, usually the larger sentences contains more important information, sentence length is considered as a quantity property. Equation (1) shows how to calculate this criteria.

$$\boldsymbol{Len_i = \frac{C_{T_i}}{C_{Long}}} \tag{1}$$

In equation (1), C_{Ti} is the number of terms in i sentence and C-Long is the number of words of longer sentence. Due to the denominator of C_{Long}, longer sentences gain weight in proportion to the longest sentence.

- Sentence position criteria

In most texts, especially the news texts, the first and the last sentence of each paragraph are more important than its middle sentence. Equation (2) shows how to calculate this criteria.

$$\boldsymbol{Pos_i = \mathrm{Max}\ \left(\frac{1}{Pos_{P_i}}, \frac{1}{C_{P_i}-Pos_{P_i}+1}\right)} \tag{2}$$

In equation (2) Pos_{Pi} is the position of i sentence in the paragraph. C_{pi} is the number of sentences in the paragraph. According to the two available fractions in Max function, if the sentence is closer to the beginning of paragraph, the first fraction will be high and will be the output of criteria, and if the sentence is closer to the end of the paragraph, the second fraction will be high and will be the output of criteria.

- Sentence weight criteria

In order to calculate this criterion, there is a necessity to calculate the weight of terms at first. Equation (3) shows how to calculate the weight of terms.

$$\boldsymbol{W\_T_i = C\_All\_T_i \times \log\frac{M}{C\_S\_T_i}} \tag{3}$$

In equation (3), in order to calculate the weight of term T_i, C-ALL-T_i is the number of occurrences of the T_i in the terms of document. C-S-T_i is the number of sentences containing T_i. Due to the weight of each term, the total weight of sentence is calculated by using equation (4).

$$\boldsymbol{W\_S_i = \frac{\sum_{i=1}^{K} W\_T_i}{Max(W\_S)}} \tag{4}$$

3.2.1.2 Similarity Criteria

Similarity is one of the most important criteria type in text processing for summarization. The reason is the recognition of similar sentences removal of an item from them in the final text summary.

- ➢ Cosine similarity criterion with document title

The most important sentence of a text is the title or the document that the text content is briefly explained. Cosine similarity criterion is the most popular and the most common symmetrical criterion to assess the similarity of a text. Equation (5) shows how to calculate sentence i cosine similarity with the document title.

$$Sim\_Title_{\,i} = \frac{\sum_{n=1}^{K}(W\_S_i\_T_n \times W\_Title\_T_n)}{\sqrt{\sum_{n=1}^{K} W\_S_i\_T_n \times \sum_{n=1}^{K} W\_Title\_T_n}} \qquad (5)$$

In the above equation, $W\_S_i\_T_n$ is the weight of nth term in ith sentence and $W\_Title\_T_n$ is the weight of nth term in the title.

- ➢ Sentence overlapping criterion

One of the most important similarity criterion in summary is sentences overlapping criterion. If the number of sentences in a document is exactly M, the criterion length of each sentence will be M-1. It means that there will be a need to a matrix (that its length is M-1) for each sentence. Equation (6) shows how to calculate the overlapping criterion of sentence i in proportion to the sentence j.

$$Overlap\,(i,j) = \frac{\sum_{n=1}^{K}(W\_S_i\_T_n \times W\_S_j\_T_n)}{\sum_{n=1}^{K}(W\_S_j\_T_n{}^2)} \qquad (6)$$

In the above relation, $W\_S_i\_T_n$ is the nth weight in ith sentence and $W\_S_j\_T_n$ is the weight of nth term in jth sentence.

3.2.2 Data Normalization

In this part, firstly the extracted criteria by previous stage are normalized in order to be in a range. The aim of normalizing is the integration of data and lowering the efficiency of it. Normalization formula is applied to the numerical data. Due to the different normalization method, like Min Max Normalization, z-score Normalization, Normalization by Decimal Scaling, Data types and etc…. In this paper has been chosen Z-score Normalization, because it controls better the effect of data outside the normal range [9]. In Z-score Normalization, also has been used of Mean Absolute Deviation instead of deviation from criteria in order to lower the effect of data outside the normal range. So first has been calculated Mean Absolute Deviation according to equation (7), and then has been normalize data according to equation (8).

$$MAD = \frac{\sum |X_i - \mu_X|}{N} \qquad (7)$$

$$norm_{data} = \frac{X_i - \mu_x}{MAD} \qquad (8)$$

In the above relation, N is the total number of data, μ_X is the mean of data and X_i is the amount of data.

3.3 The Process

In this research has been labeled the proposed method as TS-MOPSO. Figure 2 shows the performance structure of each stage that have been explained partially in the following:

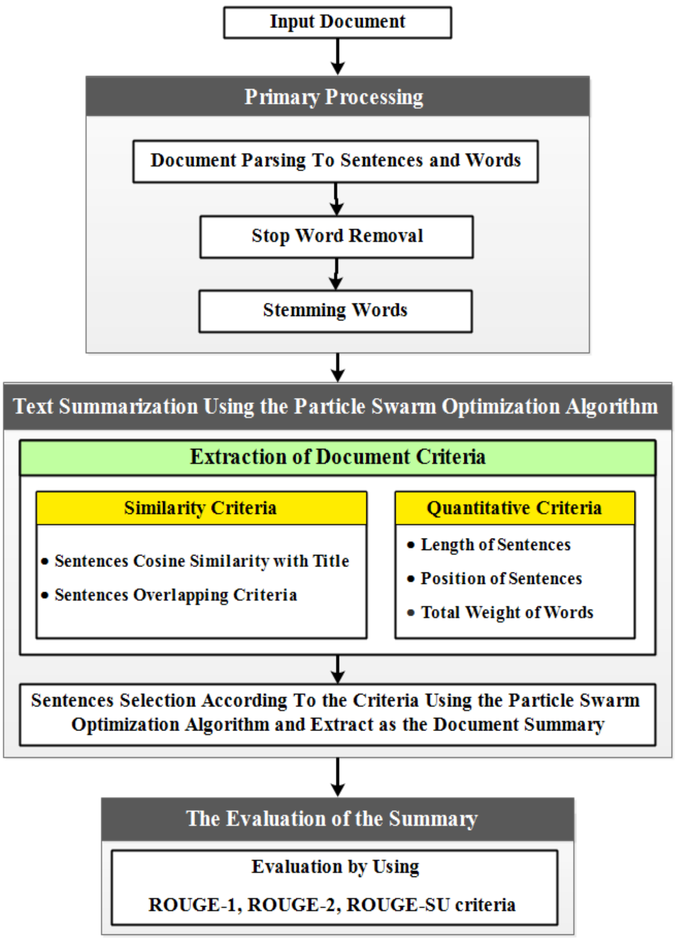

Figure 2: the structure of the proposed method

3.3.1 The Primary Process Level

In this level, at first the input documents are decomposed to sentences and then the sentences are decomposed to words. This is because has been used the words in order to extract the weight of sentences. In the following, the stop words are deleted and eventually the roots of words are extracted, that the comparison and finding similar words are done with higher accuracy.

The documents used in this article are DUC2005 and DUC2007 reference XML, in the first stage. These documents should be implemented to an acceptable structure. Each XML document has a title that is introduced as <HEADLINE>. The text of each document is placed in <TEXT> label. The text has a number of paragraph that is marked with <P> label. There are one or several sentences in each paragraph significantly. According to the conditions, there is a table for each document where two types of documents are placed. The first data is the document title and the second is the document primary text.

- ➢ Document parsing to terms and sentences

Document context contains of several sentences. In this part, S collection is formed of sentences. In the next step, each decomposed sentence and T collection (that is the terms in each sentence) are achieved.

- ➢ Word stop removal

Because the increasing speed of performance is important in finding important sentences and summarization, word stop should be ignored. The words are the repeated ones In English language that do not play an effective role in sentence similarity like; a, about, all, am, did, has, have…

The omission of these words has lots of effect in the performance speed of the proposed method. The numbers of used word stop in the proposed method are 119 words.

- Words stemming

In the following section the words are rooted for similarity recognition accuracy and accurate record of repeated words. Words etymology is used to convert them to the simplest position. This includes removing ed of the past tense, removing ing of the present tense and removing the plural s,es of the names. Similarity recognition accuracy will be much higher by doing this properly.

3.3.2 Text Summarization by Using Particle Swarm Optimization Algorithm

This part has two important processes. In first part the sentences are valued based on the quantitative and similarity criteria. Then top sentences are selected and extracted as summary by using Particle Swarm Optimization Algorithm.

3.3.3 Summary Evaluation

After extracting the final summary by ROUGE-1, ROUGE-2 and ROUGE-SU, the summary is evaluated and its semantic similarity with the original text with the original text is examined. ROUGE tool is the most popular tool to evaluate the automatic summarization that it has also used in other Natural Language Processing[3] and Information Retrieval[4] applications. ROUGE stands for "Recall-Oriented Understudy for Gisting Evaluation" that it means Reminded-based assessment for summary. This tool contains criteria in order to determine automatically the quality of summary by comparing them to the summaries produced by humans (ideal summaries).these criteria calculate the number of units that overlap between human and system summary like n-gram, words sequence and pairs of words. Among these criteria, ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S can be mentioned.

- ROUGE-N Evaluation criterion

ROUGE-N is a method (based on n-gram) between a systems summary and a sets of human summaries. ROUGE-N is calculated by equation (9).

$$\boldsymbol{ROUGE-N} = \frac{\sum_{S\epsilon\{Reference\ Summaries\}}\sum_{gram_n}\boldsymbol{Count_{match}(gram_n)}}{\sum_{S\epsilon\{Reference\ Summaries\}}\sum_{gram_n}\boldsymbol{Count(gram_n)}} \qquad (9)$$

In this equation, n is retrieved from n-gram length, and $count_{match}$ ($gram_n$) and ($gram_n$) are the maximum number of n-gram, that occurred both in summary produced by system and reference summary (summary produced by human). It's clear that ROUFGE-N criterion is a Recall-Related Measure, because the denominator of the equation is the total number of n-gram that exists in reference summaries.

- ROUGE-S Evaluation criterion

Skip-bigram is called to each pair of word (sequentially) in sentence. ROUGE-S is calculated by measuring the number of common skip-bigram between system and reference summaries. This measure is calculated by calculating the number of nonconformities in system and reference summaries. In this paper has been used this criterion more in the evaluation of machine translation.

- ROUGE-SU Evaluation criterion

The main problem of ROUGE-S method is that no score is considered for candidate statements (system summary sentences) which have no common skip-bigram with the reference summary. To achieve this goal, ROUGE-SU is calculated as tally unit according to the singularities.

3.4 Writing the Theme with the Help of Multi-Objective Particle Swarm Optimization Algorithm

The structure of multi-objective particle swarm optimization algorithm is divided into five parts to select important sentences based on the criteria produced in the previous section.

1. Production of the particles initial position based on the number of document sentences.

2. Finding the value of each member of the population using function of fit.

3. Particles movement toward the best particle.

4. Remove particles with low value.

5. Select the best particle.

- Production of the initial population

In the particle swarm algorithm, each member of population is known as particle. In the proposed structure, the array particle is bit. The initial population or particles are formed randomly. Each particle has a number of 0 and 1 that is known as the array of particle position or X. if the solution to the problem needs to find Nvar response, the next optimization of a particle will be 1*Nvar. This array is defined as follows:

X = [bit1, bit 2, … , bit Nvar]

If the value is 0, the sentence with this index is ignored and if 1 is placed in cell the sentence is selected according to table 2.

Table 2: the structure of a sets of indexes

Sentence ID	Sentence 1	Sentence 2	....	Sentence M
bit	0	1	...	1
Choice	Not selected	Is selected	...	Is selected

So the bit array is used for Multi-objective Objective Particle Swarm Algorithm population, that each particle is composed of a number of 0 and 1. For example if the document has 14 sentences, the particles produced (as initial population) will have 14 bits that is 0 or 1. If the bit is 1, the sentence relevant to this index is selected and if it is 0, the sentence with this index is ignored. The selection of the indexes is done by accident. The important point in the particles production in Alpha Summarization. Alpha Summarization determines the number of sentences needed in order to summarize. More

[3] NLP(Natural Language Processing)

[4] IR(Information Retrieval)

precisely, Alpha Summarization controls the number of particles 0 and 1 in order not to be more than limit. Equation (10) specifies the number of possible ones in each particle.

$$\boldsymbol{Ones\_Bit} = \propto \times \mathbf{M} \quad (10)$$

$\propto$ Is Alpha summarization in the above equation.

- The value of the particles

Valuation function has two objectives. The first objective is known as F1. The value of the selected sentences is achieved by using the length, position, weight and similarity with title criteria. The second objective is achieved by using sentences overlapping as F2. In first objective, there is a weight for each criterion, because it is composed of 4 criteria. Equation (11) shows how to calculate the first objective.

$$\boldsymbol{F1} = \sum_{i=1}^{M} \boldsymbol{bit_i} \times (\boldsymbol{w_1} \times \boldsymbol{Len_i} + \boldsymbol{w_2} \times \boldsymbol{Pos_i} + \boldsymbol{w_3} \times \boldsymbol{W_{S_i}} + \boldsymbol{w_4} \times \boldsymbol{Sim\_Title_i}) \quad (11)$$

w_1 is the weight of length criterion, w_2 is the weight of position criterion, w_3 is the weight of sentence criterion, and w4 is the weight of similarity with title criterion. Equation (12) shows how to calculate the second objective.

$$\boldsymbol{F2} = \sum_{i=1}^{M} \boldsymbol{bit_i} \times \sum_{j=1}^{M} \boldsymbol{bit_j} \times \boldsymbol{Overlap(i,j)} \quad (12)$$

The important fact about F2 is that if this amount is increased at any rate, sentence overlapping will be increased too. This means that the selective strings imply repetitious concepts and as a result the summary is not properly done. Because Multi-objective Particle Optimization Algorithm acts in a reduction form. (It means that the smaller amount of function of fit is more valuable).the final value of function of fit (that consists of F1 and F2) will be as follows:

$$\boldsymbol{F} = \frac{1}{\boldsymbol{F1}} + \boldsymbol{F2} \quad (13)$$

- Moving toward the best particle in the continuous space

Three factors are very important for particles motion:

1. Particles previous speed.

2. Its distance to the best experience of current particle.

3. Its distance to the best experience of all particles.

Figure 3 shows an example of how to move a particle in a continuous space.

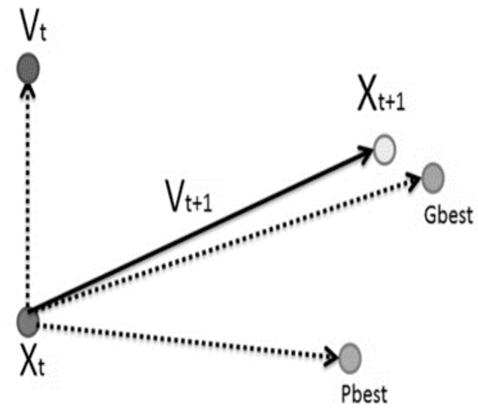

Figure 3: how to move the particle in a continuous space.

In figure3, X_t is the current position of particle. Pbest is the best experience and position of particle. And Gbest is the best experience in the whole particle. V_t is the current displacement velocity and V_{t+1} is the next phase displacement velocity. The important fact is that if the particles with a lower value is exactly transferred to the optimized particle, there will be no new answers to the problem in practice, and therefore performing displacement is in vain and with no new result. Displacement function is described the following equation:

$$\mathbf{V_{t+1}} = \mathbf{V_t} + \mathbf{rand} \times (\mathbf{Pbest} - \mathbf{X_t}) + \mathbf{rand} \times (\mathbf{Gbest} - \mathbf{X_t}) \quad (14)$$

$$\mathbf{X_{t+1}} = \mathbf{X_t} + \mathbf{V_{t+1}} \quad (15)$$

But this structure must be changed in the discrete space like the analysis in this article. To change this structure should act as follows.

- Moving toward the best particle and the best previous position in discrete space

For the initial speed of each particle, a matrix with length M (the number of strings in document) is created. 0, 1 is placed randomly in each cell of the matrix. The important fact in this part is the minimum and maximum particle velocity factor. According to equation (16) has been create two random numbers (between zeroes to M) as minimum and maximum velocity.

$$\propto \mathbf{V_{Low}} = \boldsymbol{Rand}\ (\mathbf{1}, \boldsymbol{M}) \qquad \exists\ \mathbf{0} < \propto \boldsymbol{V_{Low}} < \propto \boldsymbol{V_{High}} < \boldsymbol{M} \qquad \propto \mathbf{V_{High}} = \boldsymbol{Rand}\ (\mathbf{1}, \boldsymbol{M}) \quad (16)$$

For each particle, a random number is considered as the initial velocity factor $\propto \boldsymbol{V_t}$, that $\propto \boldsymbol{V_t}$ is between the minimum and maximum velocity. More precisely, velocity factor controls the number of ones of velocity vector in order not to be more than the limit. Equation (17) specifies the number of allowed and possible ones in each particle.

$$\boldsymbol{Ones\_Bit\_V} = \propto \boldsymbol{V_t} \quad (17)$$

According to this rule for particles velocity, in the next stage, at first t random number is created with the formula Rand(t=m/2) in order to find velocity vector. Then t and length-t are considered as the first and second point. Then, according to the existing rule in figure 4 and 5, the velocity bits of the best particle, current velocity bits of the combined particle and the next stage velocity are produced.

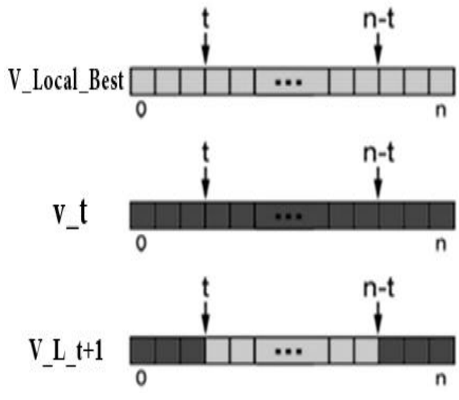

Figure 4: how to produce particle new velocity in the local space

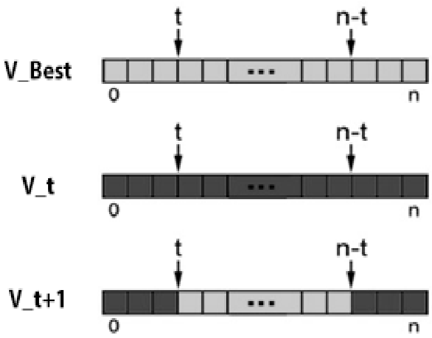

Figure 5: how to produce particle new velocity in the global space

For particle displacement based on the new velocity, has been acted as follows, that if the bit of new velocity was 1, the position bit of optimized particle is placed in the new position of corresponding bit. But if the bit of new velocity is zero, the bit of previous position is placed in the new position of corresponding bit. An example of displacement performance of the proposed method has been shown in figure 6 and 7.

Current position	1	1	0	0
The best current position of particle	0	0	0	1
V_L_t+1	1	0	0	0
The new local position	0	1	0	0

Figure 6: how to displace and determine the new position of particle in the local space

The new local position	0	1	0	0
The position of optimized particle	0	0	0	1
V t+1	1	0	1	1
The new position	0	1	0	1

Figure 7: how to produce and determine the new position of particle

An important factor in displacement is the summarization coefficient that must be applied in this part. That instead of applying the total velocity vector, the summarization coefficient can be applied on the current position. Or after the production of new particle, one can check the number of its ones and apply the limitation.

- Select the next generation of particle and remove particles with low value

The roulette wheel is used for selection. The roulette wheel is a selection where in an element that is more fit should be selected [10]. In fact, has been ascribed a cumulative probability for each element in proportion to the fitting number. And with this possibility, the chance to choose each element is determined. The selected particles are chosen as the next generation of Multi-objective Particle Swarm Optimization Algorithm. Figure 8 shows an example of roulette wheel selector structure based on the fitness.

	String s_i	Fitness $f(s_i)$	Relative Fitness
s_1	10110	2.23	0.14
s_2	11000	7.27	0.47
s_3	11110	1.05	0.07
s_4	01001	3.35	0.21
s_5	00110	1.69	0.11

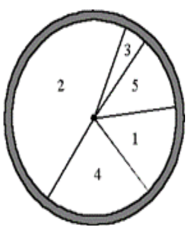

Figure 8: an example of roulette wheel selector structure based on the fitness

4. EVALUATION AND ANALYSIS OF RESULTS

In this part the performance results of the TS-MOPSO proposed method for text summarization based on Particle Swarm Optimization are examined. According to the importance and necessity of text summarization methods comparison, which is used to select the optimized method and for better evaluation. Because of the combination of text summarization and particle swarm optimization has been used, this method have been better solutions. The proposed method is compared with the reference article method by name TS-DE[5] .

4.1 Similarity Parameters

Due to the use of Particle Swarm Algorithm in this part, parameters of this algorithm are shown in table 3.

Table 3: specifications of Particle Swarm Optimization Algorithm

Termination condition	The initial weights of the valuation function	Alpha summarization	Number of generation	The number of initial population
Number of generation	0.25	0.6	200	200

4.2 TS-DE Algorithm

In this structure, differential evolution algorithm is used for text summarization [4].

4.3 Simulation Results

In this part, simulation results are investigated for two DUC2005 and DUC2007 Data sets. The results of the proposed method of TS-MOPSO are compared with TS-DE algorithm. In three tests, the results of the proposed combined method have been investigated. In the first evaluation, has been investigated the metrics of ROUGE-1, ROUGE-2 and ROUGE-SU in two TS-MOPSO and TS-DE algorithms in DUC2005 data collections. Because in the reference article TS-DE algorithm has been checked with similarity different criteria, different similarity criteria are considered in this article too.

Figure9 shows the comparing chart of the proposed method and TS-DE algorithm in DUC2005 data set. According to the results, it is clear that the proposed method has higher quality in ROUGE-1, ROUGE-2 and ROUGE-SU. The proposed algorithm has specifically higher quality compared to other methods due to the use of multi-objective function of fit in the proposed method and the use of different criteria similarity. Better sentences can be selected as summarization candidate due to the optimized performance of TS-MOPSO algorithm, and eventually the summary is the closest text to the original one in evaluation by these three indicators in DUC2005 data set. In the second experiment has been checked the amount of ROUGE-1, ROUGE-2 and ROUGE-SU metrics in DUC2007

[5] text summarization with differential evolutionary

data set in two TS-DE, TS-MOPSO algorithm. Figure 10 shows the comparing chart of the proposed method and TS-DE algorithm in DUC2007 data set. Due to the results, it is obvious that the proposed method has higher quality in the three mentioned evaluation metrics.

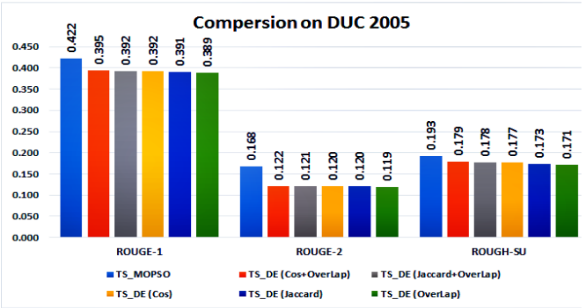

Figure 9: comparing the proposed method and TS-DE algorithm in DUC2005 data set

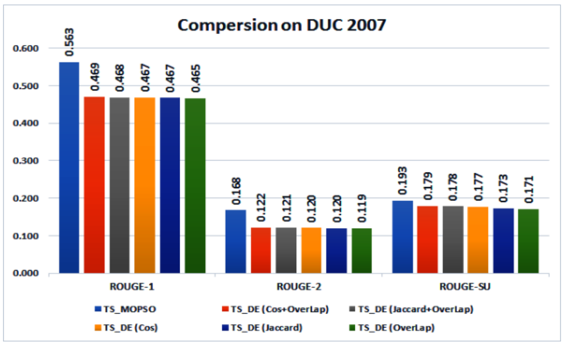

Figure 10: comparing the proposed method and TS-DE algorithm in DUC2007 data set

Using two types of standard weight and similarity criterion in the proposed method is one of the strength of this approach. The optimal choice of sentences by TS-MOPSO algorithm leads to the production of a strong summary. According to the performance assessment results, two better and more accurate syllabic words have been produced in the proposed method. The reason is the special and bold use of the weight of words along with overlapping words to each other.

One of the important parameter in the proposed method is the summarization factor. 0.6 has been considered in the proposed method. In the third experiment, Alpha or summarization factor has been changed in each metric from 0.1 to 0.9. Figure 11, 12 shows the parameter changes of summarization factor on three ROUGE-1, ROUGE-2 and ROUGE-SU metrics in DUC2005, 2007 datasets respectively.

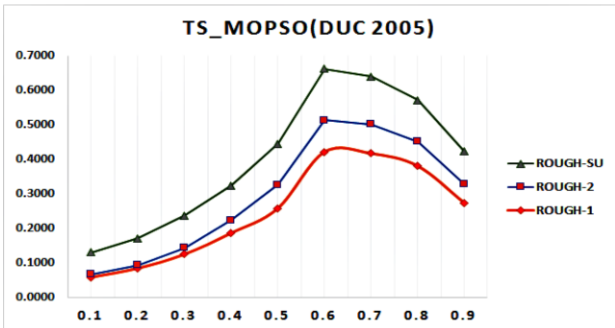

Figure 11: parameter changes of summarization factor in DUC2005 data set

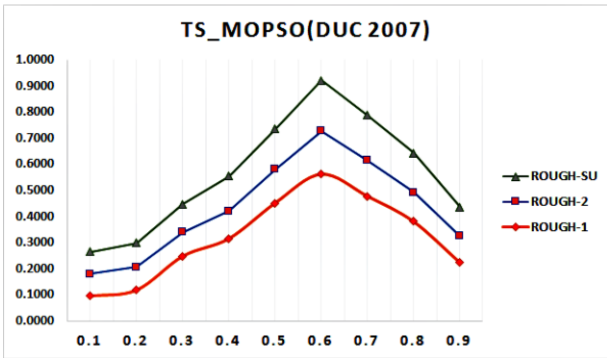

Figure 12: parameter changes of summarization factor in DUC2007 data set

Due to the results of simulation, the proposed method performance is better compared to the other methods. The results of different method comparison on DUC2005 summarization has been shown in table 4. Table 5 shows the results of different method comparison on DUC2007 summarization.

Table 4: the results of different method comparison on 2005 summarization

method	ROUGE-1	ROUGE-2	ROUGE-SU
TS_MOPSO	0.4215	0.1675	0.1927
TS_DE (Cos+OverLap)	0.3947	0.1218	0.1790
TS_DE (Jaccard+OverLap)	0.3921	0.1211	0.1778
TS_DE (Cos)	0.3916	0.1203	0.1766
TS_DE (Jaccard)	0.3908	0.1201	0.1725
TS_DE (OverLap)	0.3893	0.1187	0.1712

Table 5: the results of different method comparison on DUC2007 summarization

method	ROUGE-1	ROUGE-2	ROUGE-SU
TS_MOPSO	0.5627	0.1675	0.1927
TS_DE (Cos+OverLap)	0.4693	0.1218	0.1790
TS_DE (Jaccard+OverLap)	0.4681	0.1211	0.1778
TS_DE (Cos)	0.4673	0.1203	0.1766
TS_DE (Jaccard)	0.4669	0.1201	0.1725
TS_DE (OverLap)	0.4652	0.1187	0.1712

5. CONCLUSION AND FUTURE WORK

TS-MOPSO algorithm is an improved method of PSO based on multi-objective, using sentence criteria like: sentence length, sentence position and sentence similarity with title has a significant impact on an effective summarization. The optimized performance of TS-MOPSO algorithm leads to the better selection of sentences as summary candidate, and eventually in evaluating the summary produced by the metric ROUGE-SU will be the closest text to the original one. Because of the combination of text summarization and particle swarm optimization has been used, this method have been better solutions. According to the results obtained from the present contexts and for further evaluation of results and expansion and completion of this study, the following suggestions are presented:

- The use of newer optimization algorithms

The use of other new evolutionary algorithms like; mixed SFL algorithm (SFLA), Bat algorithm. In contrast to the genetic algorithm where in the attributes and capabilities is inherited by parents for children. In this algorithm, each individual acquires useful traits and characteristics by searching around (local search). It means that evolution goes forward individually in addition to the evolution of the population.

- The use of parallel algorithms to increase velocity

One of the main problems of the proposed method is the low speed of text processing and summary extraction. Significantly, the use of parallel algorithms increase the speed of information processing and eventually decrease time price.

6. REFERENCES

[1] Aggarwal, C.C. and Zhai, C., 2012. *Mining text data*. Springer Science & Business Media.

[2] John, A. and Wilscy, M., 2015. Vertex Cover Algorithm Based Multi-document Summarization Using Information Content of Sentences. *Procedia Computer Science*, *46*, pp.285-291.

[3] Meena, Y.K. and Gopalani, D., 2015. Evolutionary algorithms for extractive automatic text summarization. *Procedia Computer Science*, *48*, pp.244-249.

[4] Alguliyev, R.M., Aliguliyev, R.M. and Isazade, N.R., 2015. An unsupervised approach to generating generic summaries of documents. *Applied Soft Computing*, *34*, pp.236-250.

[5] Hannah, M.E., Mukherjee, S. and Balaramar, S., 2014. A Redundancy Elimination Approach towards Summary Refinement. *IERI Procedia*, *10*, pp.245-251.

[6] Alguliev, R.M., Aliguliyev, R.M. and Hajirahimova, M.S., 2012. GenDocSum+ MCLR: Generic document summarization based on maximum coverage and less redundancy. *Expert Systems with Applications*, *39*(16), pp.12460-12473.

[7] Huang, L., He, Y., Wei, F. and Li, W., 2010, April. Modeling document summarization as multi-objective optimization. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on* (pp. 382-386). IEEE.

[8] Yadav, R.K. and Singh, S., 2016. A New Approach to Automated Summarization based on Fuzzy Clustering and Particle Swarm Optimization. *International Journal of Computer Applications*, *148*(1).

[9] Han, J., Pei, J. and Kamber, M., 2011. *Data mining: concepts and techniques*. Elsevier.

[10] Lipowski, A. and Lipowska, D., 2012. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*, *391*(6), pp.2193-2196.