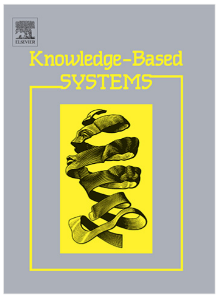
Accepted Manuscript



A survey of multiple types of text summarization based on swarm intelligence optimization algorithms

Mohamed Atef Mosa, Arshad Syed Anwar, Alaa Hamouda

PII: S0950-7051(18)30459-3
DOI: https://doi.org/10.1016/j.knosys.2018.09.008
Reference: KNOSYS 4492

To appear in: *Knowledge-Based Systems*

Received date : 1 March 2018
Revised date : 4 September 2018
Accepted date : 7 September 2018

Please cite this article as: M.A. Mosa, et al., A survey of multiple types of text summarization based on swarm intelligence optimization algorithms, *Knowledge-Based Systems* (2018), https://doi.org/10.1016/j.knosys.2018.09.008

A survey of multiple types of text summarization based on swarm intelligence optimization techniques

Mohamed Atef Mosa[1], Arshad Syed Anwar[2], Alaa Hamouda[3]
[1] National Authority for Remote Sensing and Space Sciences, Cairo, Egypt
[2] Faculty of Science and Engineering, University of Manchester, United Kingdom
[3] Faculty of Computer Engineering, Al-Azhar University, Cairo, Egypt
[1] mohammedatefmosa@gmail.com, m.atef@narss.sci.eg,
[2] arshad.syedanwar@postgrad.manchester.ac.uk, [3] alaa.hamouda@azhar.edu.eg

Abstract Due to the tremendous increment of data on the web, extracting the most important data as a conceptual brief would be valuable for certain users. Therefore, there is a massive enthusiasm concerning the generation of automatic text summary frameworks to constitute abstracts automatically from the text, web, and social network messages associated with their satellite content. This survey highlights, for the first time, how the swarm intelligence (SI) optimization techniques are performed to solve the text summarization task efficiently. Additionally, a convincing justification of why SI, especially Ant Colony Optimization (ACO), has been presented. Unfortunately, three types of text summarization tasks using SI indicate bit utilizing in the literature when contrasted with the other summarization techniques as machine learning and genetic algorithms, in spite of the fact that there are seriously promising outcomes of the SI methods. On the other hand, it has been noticed that the summarization task with multiple types has not been formalized as a multi-objective optimization (MOO) task before, despite that there are many objectives which can be considered. Moreover, the SI was not employed before to support the real-time summary approaches. Thus, a new model has been proposed to be adequate for achieving many objectives and to satisfy the real-time needs. Eventually, this study will enthuse researchers to further consider the various types of SI when solving the summarization tasks, particularly, in the short text summarization (STS) field.

Keywords
Natural Language Processing. Text Mining. Text Summarization. Swarm Intelligence. Ant Colony Optimization.

1. Introduction

The need for text summarization has largely emerged with the growth of publishing and information on the internet as social networks and information communication technologies are expanding quickly. An immense amount of data is being produced day by day as a result of the interactions and interchange of knowledge among users on the internet platforms. Which makes it an extremely difficult process to nominate the most relevant, important information. To surmount the problems of information explosion, automatic text summary has become a necessity. The process of summarization lessens the exertion and time required to identify the most salient and relevant sentences. Generally, a summary can be characterized as a text that is created from one or more texts that convey the most important information in the original text while being sufficiently short. The field of automatic summarization is over 50 years of age [1].

Since the summarization topic has been introduced long time ago, it has been solved by several diverse techniques and algorithms. Nonetheless, scarcity has been subsisted for solving text summarization through swarm intelligence (SI) algorithms. This research aims to motivate and increase the future work of using swarm techniques in an innovative manner to solve such problems which have proven their effectiveness in several areas. In addition, we have highlighted the state-of-the-art papers that have been used in summarizing the content of social media effectively [2] [3].

On the other hand, as far as we know, the swarm algorithms were just recently utilized in this field and there are not any reviews or surveys about automatic summarization using swarm intelligence techniques presented previously. Only some surveys have reviewed some of the applications that have presented summarization tasks based on other conventional techniques except SI [4] [5]. It intends to show a general figure of automatic text summarization by investigating several existing research studies that have been based on swarm techniques. Additionally, we addressed the utilized assessment strategies, the outcomes, and the related corpora. Besides, it talks about the most important issues in the field of the swarm and at last finish up with suggestions for a future work.

2. Various Types of Summarization Problems

On the basis of the number of texts, single, multi-document and short text summarizations are the three important categories of summarization. The task of producing an abstract from multiple documents is more complicated than the task of summarizing the single document. The main problem that appears in summarizing multiple documents is redundant, particularly in the short text problem. Some researchers manipulated with the redundancy problem by initially picking the sentences that are at the beginning of the paragraph and then measuring the similarity with the later sentences to select the best one [5]. Therefore, the Maximal Marginal Relevance approach (MMR) is suggested by [6] for reducing the redundancy. To produce the optimal results in multi-document and short text summarization, several researchers have investigated diverse approaches and algorithms to produce the optimal summary [2] [3] [7] [8].

Besides, the summarization task can also be categorized as an abstractive or extractive task. An extractive summary is a process of selecting a few appropriate sentences from the original text depending on the length of the summary. Some saliency features and scores are assigned to sentences in the original text and then the sentences that have the highest score are chosen to constitute the final summary. Whereas an abstractive summary is a task of producing an abstract summary which includes new other words that do not exist within the original text after the paraphrasing process. Therefore, it is much more complex than extractive summarization. Additionally, summarization can also be classified into two types: generic-summary or query-summary [9].

On the other hand, a summarization task can be unsupervised or supervised [10] [11]. A supervised system needs the training data to be able to select the most important content. In the training phase, the algorithm needs a large amount of annotated data for learning. These systems are labelled to positive and negative samples according to an occurrence of the sentence in the summary or not [12] [13]. Some popular classification methods are employed for accomplishing sentence classification tasks, such as decision tree (DT) [14], support vector machine (SVM) [9], and neural networks [10]. In contrast, unsupervised systems do not demand any training data. The output of summary is just generated by accessing only the original text. Thus, they are more suitable for any newly observed data lacking any sophisticated adjustments. Some systems apply swarm and heuristic rules to extract highly relevant summary [3] [10] [15] [16] [17] [18].

Concerning languages, there are three kinds of summaries: monolingual, multilingual and cross-lingual summaries [19]. If the language of the original source and the summary document is the same, it's a monolingual summarization system. When the source of the document has more than one language like English, Arabic, and Japanese and the final summary has the same languages, then it is termed as a multilingual summarization system. But when the source of the document is Arabic language and the extracted summary is in English or any other language other than Arabic, then it is named as a cross-lingual summarization system.

3. Traditional Techniques for Multiple Types of Summary

Regarding the field of text summarization, several studies have been developed and presented in the last past recent years. The authors in [20] build an extractive generic summarization system using the Rhetorical Structure Theory (RST) that generates a varied lengthy summary based on the request of the users. Initially, the RS-tree is built to be used to produce the primary summary. Next, based on five features, the score of each sentence in the primary summary is calculated. On the other hand, [21] suggested a new technique to summarize a single document based on the dependency between words obtained through a dependency parser and dependency between sentences obtained through RST. Both of these dependencies are figured by constructing a nested tree. To trim the nested tree lacking the significant content in the document, the summarization task has been formulated as an integer linear programming problem. Besides, [22] showed the extractive graph-based unsupervised mechanism for single document summarization using three important properties of summarization: 1) importance, 2) non-redundancy 3), local coherence. The summary depends on the rank of sentences in accordance with their importance, non-redundancy and local coherency through the optimization process. The authors in [23] proposed an optimization method to extract the summary by an efficient alternating direction method of multipliers (ADMM) algorithm. The more dissimilar sentences are introduced in the optimization framework to achieve diversity in the summary sentences. The authors in [24] built two methods to extract meaningful paragraphs from original documents to generate summaries. The first one used RST, while the second one is based on the VSM. Finally, the experimental results revealed that the RST technique is a better one. Moreover, the authors in [25] presented a hybrid summarization model based on RST mixed predefined rhetorical relations derived from various studies, such as [20].

On the other hand, multi-document summarization (MDS) is the process of separating vital data from the set of documents to produce a summary for specific users. In other words, it can be said that the MDS is a span of a single document summarization process. There are several systems that have been developed at the time of recent years for various aspects of MDS. The authors in [26] Presented (MDS-Sparse), composed of a two-level representation model for MDS based on three important properties: coverage, diversity, and sparsity. This model uses a simulated annealing algorithm to achieve summarization task. In addition, the authors in [27] proposed two key (Sen-Rich and Doc-Rich) methods for extracting multi-document summaries. Both consist of constructing clusters, extracting local key phrases, scoring documents and sentences, and extracting summary sentences. An algorithm is presented for summarizing manifold documents depending on the combination of multi generated summaries from different systems [28]. In this algorithm, four unsupervised systems are used to generate the basic summaries. Then to extract the candidate summaries, the basic summaries are combined on the sentence level. Finally, a supervised model is employed to pick the desirable summary from different perspectives. [29] proposed multi-document and a single summarization model. In the proposed framework, the summary is picked in light of the ranks of the terms. First, to extract summary sentences, the system used the discriminant analysis method and the clustering algorithm to rank the terms. Several papers are developed to enhance the clustering process as shown in [30]. The minimal-redundancy maximal-relevance (MRMR) [31] based the algorithm on selecting the features. Later, each sentence is ranked depending on three different strategies. Finally, the summary is generated.

Moreover, in recent years, social media (SM) has been a widespread platform in our daily life. With the colossal and expanding amount of online user-contributed comments on SM, users unnecessarily go over the whole comments to extricate the valuable. Several works developed recently can be classified into three categories: (1) topic and event detection, (2) summarization technique, and (3) filtering and rating. On the other hand, a novel unsupervised integrated score framework is generated as a generic extractive multi-document summary by ranking sentences depending on dynamic programming (DP) strategy [32].

Regarding the field of short text summarization, numerous researchers concentrated on making this problem as a classification and recommendation tasks. Some researchers focused on micro-blogging [33] [34] [35] [36]. A diversity of mechanisms was developed to satisfy the different needs of summarization. In [35], the gathering of short messages on a specific point is demonstrated in view of choosing the essential essence of what users are stating as a short summary. Additionally, a descriptive method is proposed to enable the adequate browsing of the massive collection of tweets by discovering the peaks of highly-discussed activity [33]. On the other hand, another system [36] is presented to show the latest breaking news automatically depending on the geographical location of tweets. Moreover, the authors in [34] suggested a system for classifying the tweets into four pre-defined categories. Next, a different strategy was utilized to generate the final summary.

Concerning the micro-blogs selection on Tweeter [14] [37] [38] [39]. El-Fishawy N. et al [14] suggested a multi-post method based on machine learning to summarize Arabic tweets specifically messages in Egyptian dialect, by gathering a group of tweets related to a specific topic. The system begins by applying natural language processing steps. Then, several features are extracted relating to the importance and similarities of the messages, post lengths, the number of followers, and the number of re-tweets. Later, features selection phase is applied to be an input for the regression model. To determine the actual weight for each feature, the model tree classifier is utilized to fulfil this issue. Finally, messages are then ordered and to eliminate redundancies to extract the summary. The proposal strategy [39] is presented to get the highest-quality tweets from the centre of the cluster. These tweets reflect mainly the key aspects of the event. [37] presented a model of tweets selection. In [38], the event timeline is divided using a modified Model, and then the system picks the closest messages to all others. Another study [40] is developed to drive bag-of-words features to classify the short messages into a number of topics. So, the users can effortlessly capture the desired information. In [41], the K-Means clustering algorithm is utilized to develop a word expansion methodology. In [42], a series of messages that describe the chain of interesting occurrences in an event is the main target to be extracted. Therefore, a modified hidden Markova model is employed to segment the event timeline, and then, the *n* messages closest to all other ones will be selected in each one. A new model is developed to strengthen the segmentation task [43].

Furthermore, the main motivation of the topic discovery on social media stream (SMS) is to support the users to understand the stream of social media. Grouping the topics for more directed examination is based on extracting a set of topics. Moreover, in [44] [45], multiple text sources are congregated with each other to strengthen the influence of topic summarization on Twitter. On the other hand, numerous studies attempted to relieve the most representative information by selecting the comments that represent the multiple opinions of the group or contain important information. The mechanisms [46] [37] [38] are superfluously developed to determine the significant comments and additionally to keep them and eliminate the more similar ones. The work of [46] aims to pick the best top-k informative short messages from a list of messages for a particular YouTube video.

4. Swarm intelligence (SI) overview

A swarm is a huge number of homogenous, simple agents interplaying locally between themselves, and their environment, with no centric control to permit a global interesting behaviour to emerge. SI algorithms have recently sprung up as a family of nature-inspired, population methods that are capable of producing low cost, fast, and robust solutions to several complex problems [47]. SI can, therefore, be defined as a relatively novel branch of Artificial Intelligence (AI) that is employed to model the collective behaving of societal swarms in nature, for example, honey bees, ant colonies, and bird flocks. Despite these specialists (bugs or swarm individuals) are moderately innocent with restricted capacities on their own, they are interacting together with certain behavioural patterns to cooperatively accomplish tasks necessary for their survival. The social connections among swarm individuals can be either direct or indirect [48].

In the previous decades, scholars and common researchers have examined the behaviour of social insects because of the astounding efficiency of these natural swarm systems. In the late-80s, computer scientists proposed the scientific perception of these natural swarm systems in the field of AI. In 1989, the articulation "Swarm Intelligence" was first presented by Beni G. and Wang J. [49] in the worldwide optimization structure as a set of algorithms for controlling robotic swarm. In 1991, Ant Colony Optimization (ACO) [50] [51] was presented and partners as a novel nature-enlivened meta-heuristic for the arrangement of hard combinatorial optimization problems (COP). In 1995, particle swarm optimization was advanced by Kennedy J. and Eberhart R. C. [52], Eberhart, R. and Kennedy, J. [53] and was first designed for simulating bird flocking social behaviour. By the late 90s, these two most widespread SI algorithms started to go beyond a purely scientific interest and to enter the domains of real-world applications. It is perhaps worth mentioning here that a number of years later, precisely in 2005, Artificial Bee Colony Algorithm was proposed by Karaboga, D. [54]. As a novel member of the family of SI algorithms. Since the computational modelling of swarms was proposed, there has been a stationary increase in the number of research papers employing the successful application of SI algorithms in several optimization tasks and research issues.

SI principles have been successfully applied in a variety of problem areas including function optimization problems, scheduling, structural optimization, finding optimal routes, and image/data analysis [55].

Computational modelling of swarms has been further applied to an extensive variety of different areas, including machine learning, dynamical frameworks, and operations research, bioinformatics and medical informatics, and they have been even applied in finance and business. This section supplies a convinced justification of why the swarm intelligence is an effective technique to solve multiple types of text summarization tasks.

4.1. Why SI especially ant colony optimization (ACO)

All optimization algorithms have common goals to achieve a minimum or maximum of the objective function. Besides, optimization methods such a function can be divided into two basic classes: numerical methods and nature inspired methods. Özgür Yeniay [56] showed the comparison between two popular heuristic methods with two commonly used gradient-based methods, namely Generalized Reduced Gradient (GRG) and Sequential Quadratic Programming (SQP), to obtain optimal conditions. The comparison results indicated that the heuristic methods outperform the traditional methods on the majority of the problems. Moreover, it is more reliable, robust and more efficient than the other methods. The nature algorithms showed noticeable achievement when solving Economic Load Dispatch (ELD) problem compared with the classical optimization algorithms [57].

A critical analysis of these SI-based algorithms was carried out by analysing their ways to mimic evolutionary operators, the ways of achieving exploration and exploitation in algorithms by using mutation, crossover, and selection by [58]. They finally stated that SI-based algorithms can have some advantages over traditional algorithms.

On the other hand, the main motive for choosing ACO among other meta-heuristic and SI algorithms like a genetic algorithm, simulated annealing, firefly algorithm, tabu search, ABC, CSO, etc., is threefold. Firstly, ACO functional of solving an optimization algorithm proved to be highly strong and could be easily understood. More accurately, it is a meta–heuristic, an algorithmic framework that can be adapted to various problems, typically are represented as a weighted graph called construction graph. Secondly, the convergence of ACO has been analytically proved by Gutjahr [59] whereas most other meta-heuristic algorithms their convergence is not yet proven using mathematical models, and results are solely depending on experimentation. Thirdly, it is evident from the literature that ACO outperforms local search methods like simulated annealing, tabu search, PSO, ABC, and Firefly etc. though different in terms of their vocabulary is almost analogous in the work implementation of ACO [60]. Thus, all these catalysts were the mainstay behind the use of ACO other than optimization algorithms. Eventually, we can notice that in table 1, the average success rate of traditional and nature-inspired optimization techniques in terms of time complexity, accuracy/efficiency, convergence, and stability are compared between. In contrast, the table 1 shows the comparison between ACO and other Met-heuristic methods in terms of time Convergence, accuracy/efficiency, and reliability is performed. Subsequently, we can notice that the ACO algorithm is the best one in most put forward aspects. So, we recommend the ACO to be the most considerable algorithm for solving such as these issues in the future.

Table 1: comparison between numerical and nature inspired/meta-heuristic methods

Parameter	Time complexity	Accuracy/efficiency	Convergence	Stability
Numerical	Low [49]	Low [48] [50]	Unsecured [75] [78]	Low [48] [50]
nature / meta-heuristic	High [49]	High [48] [50]	Just from experimentations [44]	High [48] [50]
ACO	High [49]	High [48] [50]	analytically proved [51]	High [48] [50]

5. Swarm intelligence techniques for multiple types of summary generation

This section presents numerous different types of summarization techniques based on the swarm intelligence algorithms, mainly to show the recent single-document, multi-document, and short text summarization approaches respectively.

5.1. SI-based single-document summarization

Regarding ACO algorithms, recently a few studies employed ACO mechanism in their works. HASSAN, O. F. [61] proposed an automatic single document text summarization using ACO to produce good summaries. After text pre-processing is applied, numerous types of text summarizations features have been proposed to extract Salient sentences from the text. These features are the number of title words, sentence length, and sentence position as in the text, numerical data, and thematic words. Later, the ACO-Based feature subset selection method starts by producing a number of ants, each selects one feature randomly. From these initial positions, each ant traverses edges to complete its path according to the probabilistically satisfied. The gathered subsets of the feature are evaluated by the best summation of feature subset. The algorithm is terminated when an optimal subset is found, or the algorithm gets the maximum number of times to generate the final summary. The deposit amount of pheromone is depending on the number of selected features. Finally, the authors used 100 documents for training from the DUC2002 data set. Each starts by pre-processing stage, then extracting the text features by scoring all document sentences as follows [61]

$$\boldsymbol{Score}(_{s_i}) = \sum_{j=0}^{5} \boldsymbol{s}(\boldsymbol{f}_j) * \boldsymbol{voop}(\boldsymbol{i}) \tag{1}$$

Where $Score(_{s_i})$ is the score of the sentence S_i, $s(f_j)$ is the score of the feature and $voop(i)$ is the value of the bit in ACO. Then the top 20% sentences are selected as a final summary. The ROUGE-1 is employed as a fitness function to measure if the summary is good enough. In addition, two of the Human-made summaries (H2-H1) is assigned as a reference summary. H2-H1 is compared with each other to facilely assess how close the performances of the proposed algorithms are against human performance. Moreover, some algorithms are chosen for comparison as a binary differential evolution based text summarization model H2-H1, genetic algorithm, and PSO. The proposed algorithm ultimately exceeded all other algorithms.

The particle swarm optimization (PSO) is applied to choose appropriate features successfully. It is used to pick a subset of features for the training of neural network and classification task. According to the successes of PSO, the study [15] adapted PSO in a text summarization area that depends on the extraction of the most meaningful sentences from the original text. They introduced the PSO method to examine the influence of feature structure on the feature selection procedure in text summarization. Most used features are combined with a linear combination to exhibit the significance of the sentence. The simply employed features were “keyword” and “first sentence similarity”. In contrast, “sentence centrality”, “title feature”, and “word sentence score” were the complex features. Later, cluster document sentences into a number of clusters equal the summary length. The PSO has been employed for each cluster to find out the most effective features after computing each feature score. To represent the feature’s existence, each bit takes the value of one or zero. The particle position was represented as shown in Fig. 1. Every bit takes the value of one or zero, which represents the case of one feature. If the bit contains the value 0, it means the analogous feature is unpicked, otherwise, the analogous feature is picked. Subsequently, the authors employed ROUGE-1 as a fitness function. One-hundred articles from DUC-2002 were used as a dataset for training the system. The complex features received higher weights than simple features in the result, which shows that feature structure is essential in the feature selection process.

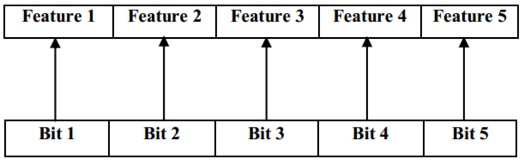


Fig. 1. The structure of a particle.

Furthermore, in [16] the dataset was divided into two sections (training and testing) to calculate the features weights. The authors used 129 documents as a dataset, 99 of them were used in the training phase, and the rest were used for testing. Consequently, the final selected sentences for the summary must be ranked in a descending manner. To evaluate the outcomes, the Microsoft word-summarizer and the first human summary that authors installed have been compared. The result showed that the PSO model exceeds the Microsoft word-summarizer and accomplished performance closer to the human model. Moreover, Binwahlan M. et al [17] continued optimizing the summarization issue based on the PSO model mixed with the maximal margin importance (MMI) technique [6]. Each sentences cluster is represented as one binary tree or more. The sentences in the binary tree are ordered according to the scores as shown in Eq. 2, and 3.

$$\boldsymbol{Score_{BT}(S_i) = impr(S_i) + (1 - (impr(S_i) * frindsNO(S_i)))} \quad (2)$$

$$\boldsymbol{impr(S_i) = avg(WSS(S_i) + SC(S_i) + SS_{NG(S_i)} + sim_{fsd(S_i)} + kwrd(S_i))} \quad (3)$$

Where WSS: the score of the word, SC: sentence centrality, SS_{NG}: the average of similarity features, the similarity of the sentence S_i with the first calculated sentence in the document using cosine similarity metric, and $kwrd$ is the keyword feature. $Score_{BT}(S_i)$ is the score of the sentence S_i in the binary tree, $impr(S_i)$ is the signification average of the sentence S_i and $frindsNO(S_i)$ is the number of the closest sentences. The main idea behind that work is to prevent the likely redundancy by selecting a sentence that has low relevance to previously selected sentences in the summary, at the same time, is largely relevant to the topic. Finally, the results are evaluated using ROUGE-1, ROUGE-2, and ROUGE-L respectively.

On the other hand, Binwahlan M. et al [18] proposed an automatic text summarization system from an abundance of text sources. The system has been implemented using swarm intelligence incorporated with fuzzy logic. Initially, the weights of sentences have been captured from the swarm module based on previously mentioned features. Later, these weights were used as inputs for the fuzzy inference system. The fuzzy logic including fuzzification, interface, and defuzzification has the advantage of training and developing in such problems, so the rules are completely understandable and simple to modify, add, and remove rules. The main objective of employing PSO for fabricating the features weights in accordance with the importance, which played a meaningful role in the discrimination between features significance. The score of each sentence S_i in the document is calculated as shown [18].

$$\boldsymbol{swarm\_import(S_i) = \sum_{j=1}^{5} \omega_j * score\, f_j(S_i)} \quad (4)$$

In the fuzzy logic, the authors used the trapezoidal membership function to determine the degree to which the input values belong to the appropriate fuzzy sets. Three fuzzy sets are used: high, medium and low. After getting the scores of all phrases produced by the fuzzy inference system, the phrases are re-ranked according to those scores. The top n sentences are picked to be the summary. A proposed system is generated in two forms. Where the diversity does not dominate the model behaviour and it works such as the fuzzy swarm-based method. In the second form, the behaviour of the system is variety dominated. In addition, the removing of the redundant sentences using the diversity selection gives the high performance as well. Combining all these mechanisms gives better performance than each one individually. Finally, the sentence score accounted for in accordance with their scores to constitute the summary. The experiments have been implemented based on the DUC 2002 document sets to assess the proposed method. The comparisons were performed between the proposed method and another three methods, the swarm model and the two benchmarks (MS-Word and H2-

H1) using the average precision, recall, and F-measure with ROUGE-1, ROUGE-2, and ROUGE-L. It was shown that the incorporation of fuzzy logic with swarm intelligence could achieve an appreciable performance during the construction of the final summary. Also, the results showed the higher performance of the algorithm outperformed the swarm model when the fuzzy model was mixed.

What is more, in [62], the authors proposed a novel approach for text summarization using multi-agent particle swarm optimization. First, the pre-processing has been performed on the input text, including separation of sentences, stop words removing, and steaming. Then, some features has been calculated as an adapt TF-IDF method [62].

$$G(T_{ij}) = \log \frac{N}{n_j} + 1 \tag{5}$$

Where N is the total number of sentences of the text, and n_j is the number of sentences that have the word j. After the weight of sentences has been given, the similarity matrix had been obtained using cosine metric. Subsequently, a MAPSO algorithm is applied using this information to extract key sentences and a summary of an original text. The initial parameters of the MAPSO algorithm had been set and initialized, where the number of maximum iterations was 100, the number of particle agent was 20. Then, a random number of sentences are assigned to each particle agent. The cost of each particle agent is calculated according to the readability of summarized sentences and the dependency of sentences with each other. The sentence dependency factor is the final summary discusses the same information. Moreover, the readability factor indicates that the gathered summary has a high degree of similarity. Finally, calculate the cost of each agent and select the best sentences. The proposed method was examined with a set of DUC 2002 standard documents and the results were evaluated by ROUGE method and were also compared with other methods. The comparison showed that the proposed system outperformed the other systems.

Regarding the Turkish language, a novel hybrid of artificial bee colony algorithm ABC combined with the semantic features and structured is proposed in [63] to summarize the Turkish language. The system used five features, three are semantic features whose importance is extracted from Turkish Wikipedia links. The proposed features are joined based on their weights using two novel approaches. The first utilizes the process of analytical hierarchical, which is based on several human judgments and depends on pairwise comparisons of the features. ABC algorithm is employed as a second approach to specifying the real weights of the features. In order to obtain the optimal weights, the corpus was 100 documents. It is split into a training set consisting of 88 documents and a test set consisting of the rest using a 5-fold cross-validation. A new Turkish corpus that includes 110 documents and 3 expert-generated the desired summary to confirm and evaluate the significance of the proposed hybrid system. Moreover, ROUGE-1 and F-measure matrices are used for evaluation. Finally, the experimental results showed that exploiting the features with each other gave a better performance than exploiting each feature individually.

On the other hand, other SI algorithms were employed in the summarization field, the bacterial foraging and cuckoo algorithms. The authors in [64] [65] recently presented new approaches for solving the summarization task. [65] Used cuckoo search optimization algorithm (CSOA) to enhance the performance of extractive-based summarization task. Firstly, after the pre-processing stage, the weight of sentences must be done based on the TFIDF method. Secondly, the cosine similarity is used to calculate similarities between sentences and keywords. Thirdly, important sentences are extracted from the main text by CSOA algorithm. After initialization of CSOA's parameters, random assignment of sentences to birds is presented. Later, the cost function is employed to assess the summarization value as following [65].

$$CF_s = \frac{\log(C*9+1)}{\log(M*9+1)} \tag{6}$$

Where CF_s is the coherence factor of sentences, C is the average distance of the available sentences, and M is the maximum weight of the sentences. Eventually, the final candidate summarized text is given based on three categories: 1) human judgments, 2) matching sentences together, 3) matching words together. The experiments have been implemented using the DUC 2002 document sets to assess the proposed system. The comparisons were performed between the proposed method and another three methods, PSO, MS Word, and BFOA using the average F-measure. Analysing of obtained results indicated reliability and better performance of the proposed approach.

Moreover, Nikoo, M. D., et al [64] presented a new method of automatic text summarization based on the bacterial foraging optimization. The main idea behind this method is weighting words using TF-IDF method, then calculating the sentences value, extracting the best sentences from the text. A bit string is used to represent the corresponding words. Each bit that corresponds a word in the text, can take only 1 or 0. Secondly, the bacterial foraging optimization employed to converge the solutions is obtained from each bacterium. Finally, the weight of sentences is determined by summing the total weights of their words. Rouge-1 is employed as an evaluation function to evaluate each candidate summary text.

5.2. SI-based multi-document summarization

Regarding the multi-document summarization (MDS), Donis-Díaz, C. A., et al [66] has identified a new model for linguistic summarization of numerical data (LDS) based on ACO. The ACO-LDS algorithm used the Max-Min ant system. The algorithm starts initializing the pheromones to a high value and figuring out the local heuristic desirability information for nodes in the graph. The heuristic information depends on the frequency of using the term F_u. Where F_u is the number of times, $u_{v_{ij}}$, the node v_{ij} has been used in the summary as follows [66]:

$$\boldsymbol{F_u = 1 - \left(u_{v_{ij}}/p\right)^e} \tag{7}$$

Term $\boldsymbol{p}$ represents the number of propositions added up till now to the partial summary and $\boldsymbol{e}$ is a parameter in [0, 1] to graduate the “power” of an influence of F_u.

Later, after each iteration of the algorithm, the local search is applied to all constructed solutions to enhance the produced summaries. According to the value of fitness function, the best summary is updated. Finally, the exit criteria terminate the algorithm when the stagnation condition appears, or the algorithm reaches the maximum number of iterations. Finally, the ACO-LDS algorithm overcomes measures of the goodness of the generated summary but fails to enhance the diversity of the final summary extracted by Hybrid GA-LDS. Good results that have been obtained in ACO-LDS for fineness are impressed by the constructive procedure used in ACO.

On the other hand, Yadav, R K, and Sugandha S. [67] also presented a hybridized algorithm using the combination of PSO and fuzzy C-Means. The approach used PSO with partitioning clustering algorithms based on Fuzzy-C-Means and K-means to handle such high-dimensional clustering over summarization of multi-documents. A number of used dataset documents is 1504, and the number of clusters equals two to five. The accuracy of the documents clustering has been evaluated according to three validity measures: Entropy, F-Measure, and similarity distance. In the last, they compared the proposed algorithms KPSO (K-Means Particle Swarm Optimization) and FCPSO (Fuzzy C-Means Particle Swarm Optimization) results with the existing clustering algorithms K-Means and Fuzzy CMeans. The quality of FCPSO cluster has achieved performance better than that KPSO as it deals perfectly with the overlapping nature of the documents which makes it more suitable to use.

Besides, an additional work related to MDS is applied. Some researchers Bazghandi M. et al [68], Aliguliyev R. M. [69] suggested generating a crossbreed system by mixing the mutation mechanism in the genetic algorithm with PSO. A set of document summarization system has been prepared as an input of the system [68], including the relevant/irrelevant information to the main topic of the text. The sentences were clustered based on PSO in a way that the most similar sentences had been gathered with each other within the same cluster (intra-cluster), the sentences within a cluster should be dissimilar from the sentences in other clusters. To overcome on violate the limitations or even appear the vector in decimal forms, the genetic mutation strategy has been mixed with the PSO algorithm. In terms of the experiment, a set of sports news from ISNA (Iranian Students News Agency) was chosen as the experimental corpus. Eight subsets have been administered for assessments and tests. The initial population of the particles was 40. Maximum iteration number was 250. Besides, three examiners have been asked to select the most important 10 sentence. The proposed system was compared with the clustering based on PSO and K-means algorithms. In [69] another generic multi-document summarization approach using sentences clustering is introduced. Depend on PSO and genetic mutation techniques, five clustering methods are presented. But in this method, another dissimilarity measure is employed, if the two terms appear in the same text which means they are semantically relatives. Besides, a new sentence extractive mechanism is developed using the recursive formula. This mechanism calculates the mean weight of a sentence with respect to the cluster, which they are assigned to. Finally, Experimental results on the DUC-2007 and DUC-2005 corpus have shown the good effects of the proposed methods. The summarization methods have been evaluated using ROUGE-2 and ROUGE-SU4 metric.

A multi-document summarization technique based on the benefits of clustering, semantic role labelling, and PSO to rank arguments in each cluster using optimized features [70]. First, divide the document into the separated sentences. Next, a semantic role parser is employed to obtain PAS from sentence collection set. The semantic similarity is created from pairwise similarities of PASs based on Jiang's similarity measure. Subsequently, the hierarchical clustering method depends on the similarity matrix of PASs is constructed to group semantically PASs that are more similar. Based on features selection, the optimal weights are determined to employ PSO to choose the top-ranked PASs from each cluster. Finally, Simple realization engine is utilized to produce sentences from the chosen PASs. Eventually, Experiment of this approach is performed with DUC-2002 and the performance is evaluated against the benchmark summarization systems using F-measure. Experimental results affirm that the proposed approach yields better results than other comparison summarization systems.

Furthermore, a successful model of honey bees searching follows for nectar in a field of summarization [71]. In this framework, the Jensen-Shannon (JS) divergence has been considered as a classical similarity metric for comparing the summary content and the source documents. The authors present an optimization framework to extract the summary from multi-documents. The framework has been developed based on two approaches; one is based on a genetic algorithm, the second is based on an ABC algorithm. In their experiments, they used two known datasets from the document understanding conference (DUC), namely (DUC-2002 and DUC-2003). The algorithms have been compared with the JS divergence metric for both the unigrams (JS-1) and the bigrams (JS-2) different. ROUGE-1 and ROUGE-2 methods have been used as an evaluation method based on human evaluation. Finally, the proposed algorithm has been compared with four known algorithms as TF-IDF weighting, LexRank, and KL-div Greedy [72]. The evaluation has shown a competitive performance of the proposed framework. Moreover, the analysis of results had shown that exciting complementary property of the swarm and genetic summarizers and of a strong baseline.

In addition, another novel multi-document summarization technique is presented based on ABC optimization [73]. It is used by companies for extracting significant facts about the specific product from their competitors' news to strategic business decision making. Initially, the system creates a bag-of-words vector space, runs the latent semantic analysis, uses just top 5%, represents the documents, runs the self-organizing map, and employs the K-Means algorithm to cluster the provided data. Later, ABC algorithm is applied to generate the final summary. The corpus consists of news items about multiple products of the company. So, the news items about specific categories of products will be clustered according to each product. Finally, the presented results were based on the collecting news items for an appointed consumer electronics company from authentic news sites. In terms of ROUGE and F-measure evaluation scores, the performance of the generated summary using the ABC algorithm is shown to be better in several aspects than the MEAD algorithm.

On the other hand, a novel Cat Swarm Optimization (CSO) based multi-document summarizer is proposed to address this problem [74]. The system begins by applying the pre-processing phase, by Sentence segmentation, tokenization, remove stop words and stemming. Subsequently, the authors have tried to build summaries from document sets with multiple objectives as non-redundancy, content coverage, cohesion, and readability, which are explained as follows [74].

$$\boldsymbol{f}(\boldsymbol{S}) = \boldsymbol{f}_{\boldsymbol{cov}}(\boldsymbol{S}) + \boldsymbol{f}_{\boldsymbol{coh}}(\boldsymbol{S}) + \boldsymbol{f}_{\boldsymbol{read}}(\boldsymbol{S}) \quad (8)$$

$$f_{cov}(S) = sim(S_i, O) \quad i = 1, 2, \ldots, n \tag{9}$$

$$f_{coh}(S) = 1 - sim(S_i, S_j) \quad where\ i \neq j \quad i, j = 1, 2, \ldots, n \tag{10}$$

$$f_{read}(S) = sim(S_i, O) \quad i = 1, 2, \ldots, n \tag{11}$$

Where O = the center of the main content collection of sentences. The content coverage of each sentence in the summary $f_{cov}(S)$ is represented as the similarity between S_i and O. That means, higher similarity values correspond to high content coverage. Likewise, the higher value of $f_{coh}(S)$ specifies the high connection between sentences and vice versa. $f_{read}(S)$ Measures the similarity between $\boldsymbol{S_i}$ and $\boldsymbol{S_i}$ to specify higher readability of the summary.

Later, the algorithm calculates the Informative score, and similarity for every sentence, and merge the least similar sentences in one document to use sentence weight and inter-sentence similarity as cat information. By applying for the cat's positions according to the fitness function is specified in equation (21). Save *gbest* as the best cat position (best solution). Finally, *gbest* is the vector of candidate summary sentences when the exit criteria are satisfied. In terms of the experiment, the performance is evaluated in terms of F-score, ROUGE score, positive predictive value (PPV), sensitivity S_{svt}, and summary accuracy S_{acc} on a benchmark DUC 2006 and DUC 2007. Ultimately, in most of the cases, CSO based model is showing better performance higher than the HS and CSO algorithm in summary generation.

5.3. SI based short text summarization

On the other work, regarding social network comments, [2] proposed recently and for the first time, ACO based user-contributed comments system. The proposed system has been introduced to summarize an Arabic short text on Facebook posts, specifically posts in Egyptian dialect, by selecting all comments associated with the specific post. The algorithm begins as shown in fig. 2 by applying the natural language processing (NLP) phase to generate the list of vectors, by removing any foreign language, redundant character, and symbols. Later, convert the colloquial terms to modern standard Arabic (MSA), remove stop words and root extraction processes are developed. Subsequently, a Semi-graph representation is constructed by excluding extremely long comments and the comments that have the closest similarity with each other according to a specific threshold. Accordingly, nodes represent the comments and directed edges designate the possible candidate comments. Two commonly adoptive metrics are employed to measure the similarity between comments: 1) Jaccard similarity, 2) Cosine similarity. The Cosine similarity is defined as [2].

$$\boldsymbol{Cosine}_{(C_i, C_j)} = \frac{\sum_{i=1}^{n} A_i . B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{12}$$

The cosine similarity approach is applied in text matching when the attribute vectors A_i and B_i are usually the term frequency vectors of the comments C_i and C_j. The second method is the Jaccard similarity that is presented as.

$$\boldsymbol{Jaccard}_{(C_i, C_j)} = C_i . C_j / C_i \cup C_j \tag{13}$$

The Jaccard method measures the similarity between finite sample sets. Subsequently, the number of ants is set on randomly the number of nodes. Each ant applies an arbitrary probability based on three features associated with two limitations, to make a clever decision which node should be visited later. Some of the effective features are selected to make the desirable heuristic information, as TF-IDF [14], PageRank, and a number of (likes, mentions, and replies) to constitute effective heuristic information as shown [2].

$$\boldsymbol{\eta_{ij}} = \frac{\lambda\,.PR(C_i)+\theta\,.TF-IDF(C_i)+\gamma\,.ISU(C_i)}{\omega\,.\sum_{w_i \,\epsilon\, C_i} w_i+\upsilon\,.Max[simi(C_i\,, C_{ps})]} \tag{14}$$

Where the PageRank algorithm $PR(C_i)$ is a walk style across a graph randomly based on the similarity between the comments based on the precedence [46]. TF-IDF is defined as follows, $tf(w_i)$ is the number the term w_i appears in the comments. $idf(w_i)$ is the logarithm of the number of the comments. $ISU(C_i)$ is the number of likes, replies, and mentions that the comment has received, divided by the number of comments that the word appeared in and the maximum similarity with the added comments in the partial solution.

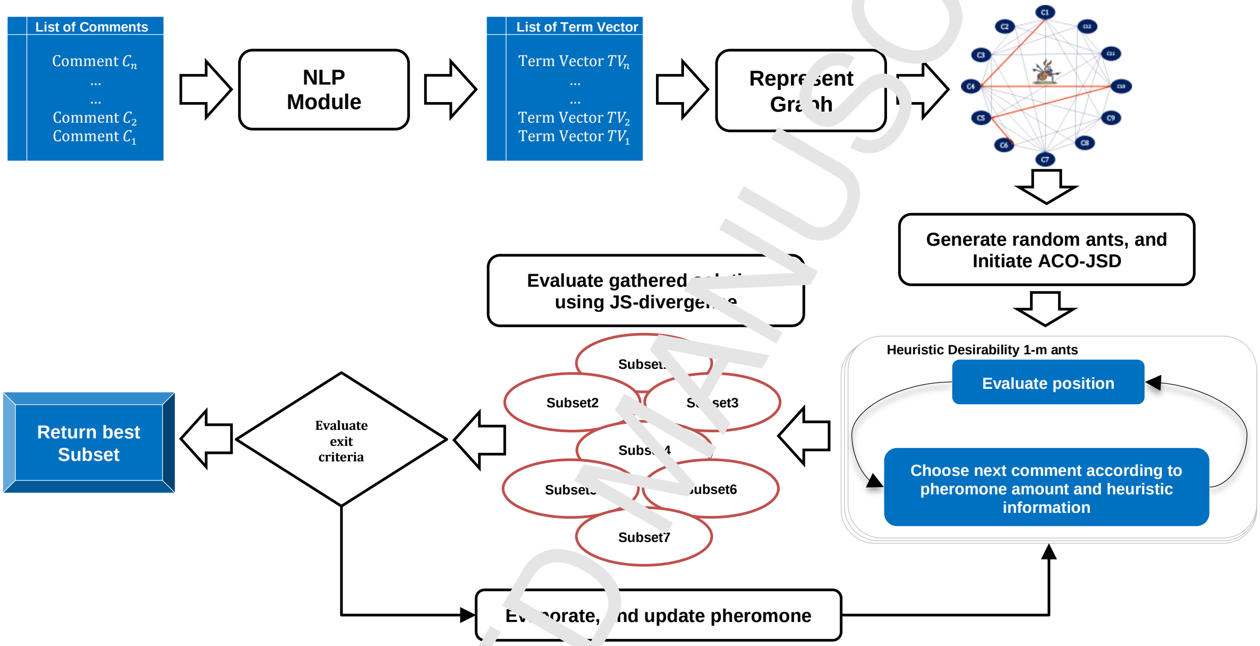


Fig 2. ACO based user-contributed summarization framework.

Afterward, the algorithm picks the feasible comment which has the highest probability and the amount of pheromone. After each iteration, the algorithm selects the visited paths for depositing pheromone and the additional amount of pheromone for the best paths. Moreover, the evaporation stage for all paths to help the ants forgetting the bad paths. Finally, when there is no room for improvement after a minimum iteration of 20 iterations, the algorithm utilizes the JS-Divergence method as a fitness function to measure the retention ratio of information (RRI) as for the list of comments. RRI is defined as follow [2].

$$\boldsymbol{RRI} = \frac{\textit{information ratio in the summary}}{\textit{information ratio in the list of comments}} \tag{15}$$

On the other hand, the influence of varying the parameters has been experimented using different settings of 10 independent executions of the algorithms. The number of ants, initial pheromone trail, evaporation rate, heuristic information, and the weights of variant kinds of heuristic information (λ, θ, γ, ω, υ) and the weights of pheromone are varied among candidate values except for one parameter that keeps unchanged to determine the desired various weights. Ultimately, in terms of the corpus, two data sets are used as benchmarks. One is gathered including the twenty seven posts with sixty comments, a corpus is collected from the Egyptian and Arabic news pages on Facebook [75]. The second data set is collected by the authors Including 100 posts are downloaded with associated comments from 20 celebrated Arabic Egyptian pages. Four volunteers have been invited to evaluate the quality of the generated summary for the algorithm and other traditional algorithms like

PageRank, LexRank, mead, TF-IDF, and Mead in agreement with some of the standards. In addition, these traditional methods have been mixed with k-means and hierarchical cluster algorithm to generate the diversity. 20 comment streams are used for human and automatic evaluation. The F-measures mechanism is used for evaluating the automatic summary as for the human summary. Calculating the ratio of information RRI, compression rate CR, and similarity of the summary have been used for automatic evaluation. The result showed that the system's performance was superior to all of the other systems.

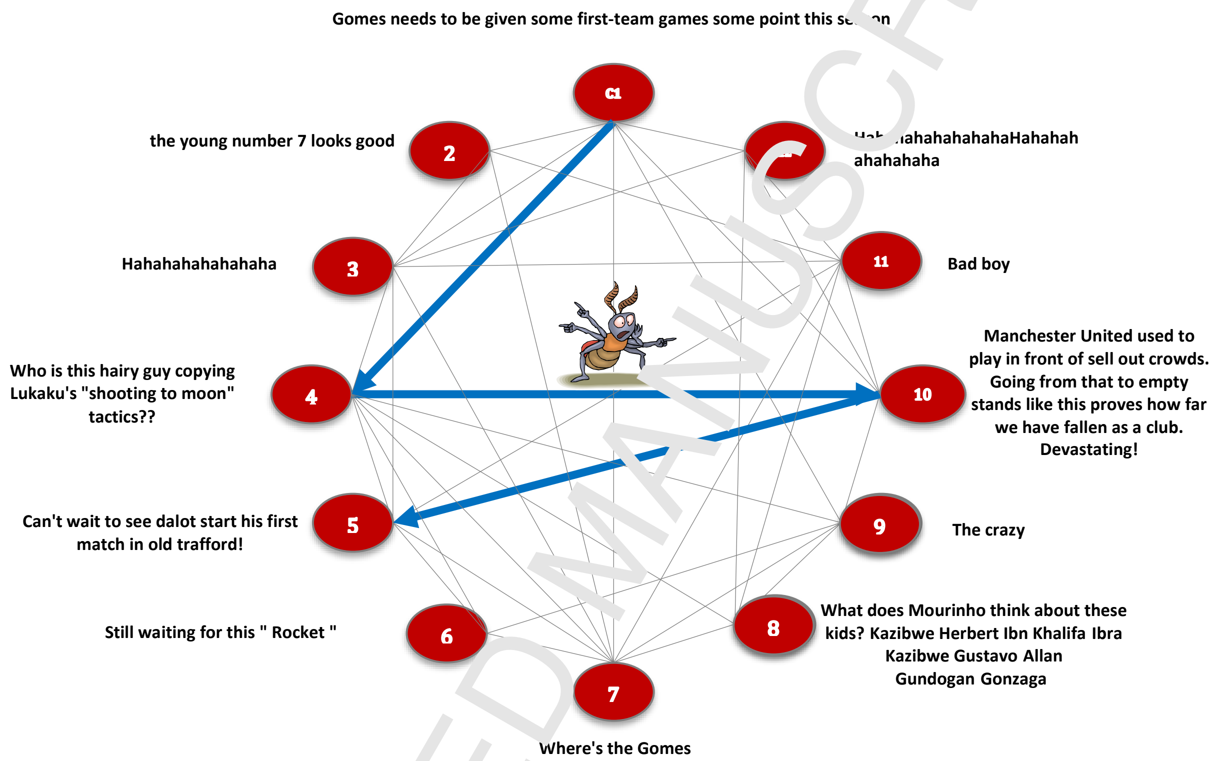


Fig. 3. ACO problem representation for user-contributed summarization

Furthermore, Mosa M. A. et al [3] continued optimizing the summarization task by adding a graph coloring mechanism [2]. The paper opens up an inventor field of short text summarization using a hybrid ant colony optimization, mixed with a graph coloring and local search mechanisms. Initially, the graph coloring algorithm has been utilised before to contract the number of comments by dividing the list of comments into smaller sets to protect the system from falling in the cycling and enhance the fulfilment of the approach. The graph coloring concept used to creates several colors, whereas each color has a varied collection of comments and topics that have been addressed by users. At the same time, each color should retain a higher ratio of retaining information based on Tri-vergence of probability distributions (TPD) [76]. The same comment can be affiliated with many colors. Later, the final summary is extracted from the best color. Jaccard method used to measure the similarity among the comments with each other. After that, the Simi-graph is constructed within each color by isolating the extremely lengthy comments as shown in fig. 3. Subsequently, activating ACO mixed with the local search to constitute the final summary. Each ant will construct its path within each color. The best one is selected in all colors. Meticulously, each ant applies an ingenious probability based on two celebrated features. The mutual information MI and the PageRank algorithm. MI is used to calculate the significance of the term into the color according to the existence of it in other colors and the original list of comments according to Eq. 16 [2].

$$MI = \frac{N_{11}}{N} log_2 \frac{N_{11}N}{N_{1.}N_{.1}} + \frac{N_{10}}{N} log_2 \frac{N_{10}N}{N_{1.}N_{.0}} + \frac{N_{01}}{N} log_2 \frac{N_{01}N}{N_{0.}N_{.1}} + \frac{N_{00}}{N} log_2 \frac{N_{00}N}{N_{0.}N_{.0}} \quad (16)$$

Where N shows the number of comments, N_{11} shows the number of comments that contain term t and are in color C, N_{10} shows the number of comments in which contain term t and are not in color C, N_{01} shows the number of comments that don't contain term t and are in color C, and N_{00} shows the number of comments which don't contain term t and aren't in color C. The PageRank algorithm is used to give a more considering for the earlier comments that have the similarity with the later comments. A heuristic desirability is employed cleverly to pick the comments that have higher mutual information and PageRank weighed. At the same time, these comments shouldn't have high similarity with the gathered comments into the solution and not lengthy.

Moreover, after completion of ACO algorithm, the author's suggested searching for the comments that have higher priority and have not been assigned to the summary. Their hypothesis is that the comment may be more important, attractive and has a great priority by the fans themselves, and considers how many likes, replies, and mentions it has. The authors considered not all granted like, mention and reply are of the same significance. The granted like by celebrated user who has numerous followers is more important than others. Apparently, counting how many followers the users have is an innovative way to determine the actual importance of these users. So, they employed a local search algorithm to update the solution by replacing the comment that has the lowest priority by the unselected one as shown in fig. 4 that has the highest priority as shown below [3].

$$\boldsymbol{priority(C_i)} = \sum_{i\,\epsilon\,\#LRM} \left[\frac{C^i_{\#Followers}}{\#MaxFollowers}\right]^{1}_{0.05} \tag{17}$$

Where $C^i_{\#Followers}$ the number of followers is that user has, $\#Maxfollowers$ is the maximum number of followers among all fans. Finally, when there is no additional improvement in the algorithm after ten times, the Tri-vergence of probability distributions is employed as a beneficial fitness function to evaluate the gathered solutions automatically. This method allows calculating the similarity among triplets of objects P, Q, and R. Where Q is the source of the original document, R as a summary to evaluate and P is all the summaries of the set excepting R.

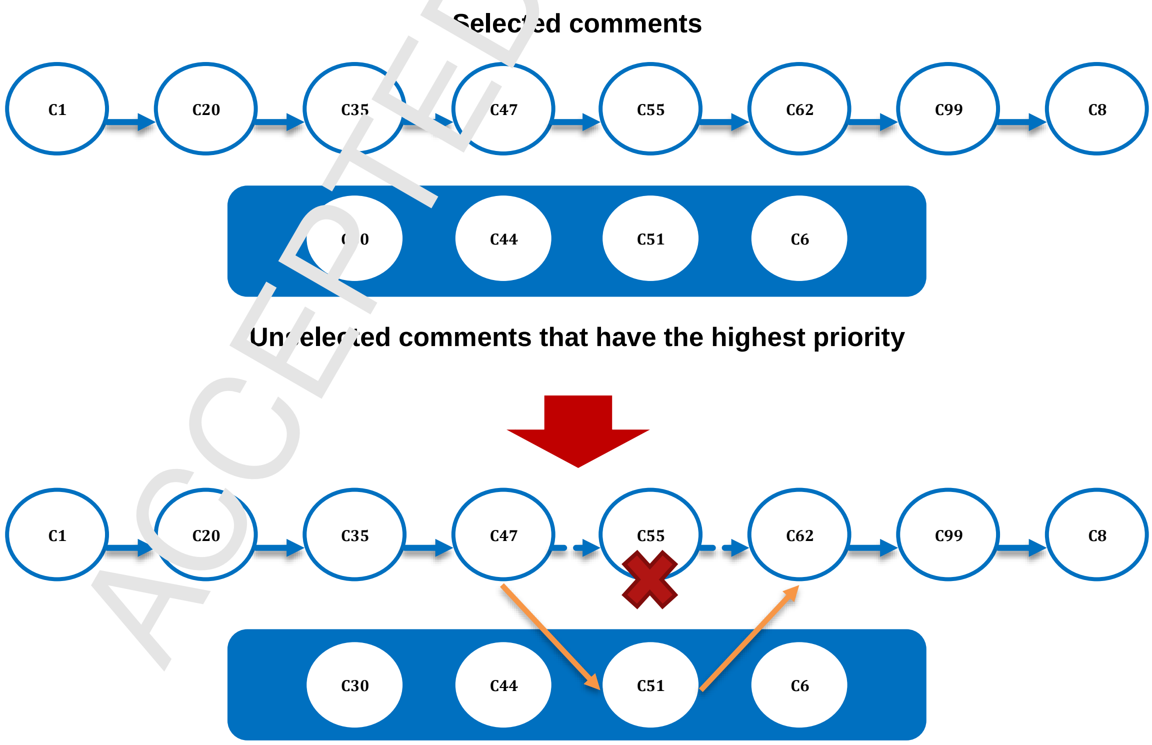


Fig. 4. Neighbourhood structure of LS

Regarding the corpus, the two data sets are used as benchmarks [75] [2]. The algorithm trained on 60 posts and tested on 40 posts associated with comments to adjustment the ACO parameters set. The algorithm has been compared with other traditional algorithms like MMR [6], PageRank, LexRank, TF-IDF, and Mead. These methods have been mixed with k-means and hierarchical cluster algorithm and with each other to make different solutions. Thirty comments are used for human and automatic evaluation based on ROUGE method. The automatic evaluation has been suggested to evaluate the different generated solutions as shown [3]:

$$\boldsymbol{Automatic - Evaluation} = \frac{\boldsymbol{Ratio\ information}}{\boldsymbol{Compression\ Rate\ Similarity}} \tag{18}$$

Eventually, the experiments showed that the algorithm has achieved higher performance than other alternative hybrid systems when ACO is applied. The performance increases when the local search is combined with the ACO algorithm until reached the peak when three algorithms graph coloring, ACO, and local search have been combined with each other.

6. Summary evaluation

Evaluation item is an exceedingly important part of automatic text summarization; still, it represents a challenging problem in this field [77]. There are several elements behind the challenges of summary evaluation and the obstacles of comparing various summarization systems. One motive is that to evaluate summary contents, a comparison is made with reference summaries [77]. In other words, there is a need for a benchmark corpus that contains both the reference summaries and the original documents to be summarized likewise. The problem with making such a corpus is no ideal 'unique' summary [78] whereas several summaries can be adequate for the same document, and even using unlike phrases, can create different versions of the same summary.

Besides, there are many elements beyond the challenges of summary evaluation and the obstacles of comparing various summarization systems. Furthermore, the generation of reference summaries based on human experts and time-consuming task is a costly task. It is not facilitated for humans to know what kind of data should show up in the summary. On the other words, information changes depending on the target of the summary and to select this information automatically is a very complicated issue. Generally, two ways for specifying the performance of a summarization task [78]: (1) extrinsic evaluation: the quality of summarization relies on how it affects other issues as information retrieval, Text classification, and question answering. (2) Intrinsic evaluation: the determined summary quality depends on the coverage ratio between the human-summary and machine-summary. Retention of the information is the important aspects on the basis of which a summary is good. But how to know which parts of the document are relevant and which are not, it is still a problem with this task.

6.1. Evaluation corpora

Several workshops and conferences addressing the area of automatic summarization have been organized since 1998, such as SUMMAC, DUC, and TAC. These conferences avail the summarization field in several ways, such as through providing researchers with required corpora and enabling them to evaluate and compare their methods. For instance, Document Understanding Conferences (DUCs) where a significant series of conferences that addressed the subjects of automatic summarization and were turned by the US Department of Commerce's National Institute of Standards and Technology (NIST). Gambhir, M., & Gupta, V. [4] compared and evaluated their results using DUC's benchmark. In 2008, DUC turned into a summarization path at Text Analysis Conference (TAC). TAC also entertains a sequencing of evaluation workshops that provide many tracks addressing different areas of NLP. The summarization issue was included from 2008 to 2014. Additionally, in some of the surveyed studies in summarization using SI, authors used their own corpus (see Table 3), and for this reason, it is not possible to compare the results of these studies. Although DUC 2001, 2002, 2003, 2005, 2007, and ISNA, in addition to the used languages are different including English, Arabic, Turkish, and Persian.

Table 3. Types of corpus produced by the surveyed literature and their summarization approach.

Reference	corpus
Binwahlan M. et al [18]	DUC-2002
Binwahlan M. et al [15]	DUC-2002
Peyrard, M., and Eckle-Kohler, J [71]	DUC-2002 and DUC-2003
Mosa, M. A. et al, 2017a [2]	Authors' corpus
Mosa, M. A. et al, 2017b [3]	Authors' corpus
HASSAN, O. F. 2015 [61]	DUC-2002
(Güran, A, 2013)	Authors' corpus
Chakraborti, S., and Dey, S. 2015 [73]	Authors' corpus
Binwahlan M. et al [18]	DUC-2002
Binwahlan M. et al [15]	DUC-2002
Asgari, H, 2014 [62]	DUC-2002
Bazghandi, M., et al, 2012 [68]	ISNA (Iranian Students News Agency)
Aliguliyev, R. M., 2010 [69]	DUC-2005 and DUC-2007
Khan, A., et al, 2015 [70]	DUC-2002
Nikoo, M. D., et al, 2012 [64]	DUC-2002
Mirshojaei, S. H., and Masoomi, B. 2015 [65]	DUC-2002
Rautray, R., and Balabantaray, R. C. 2017 [74]	DUC-2006 and DUC-2007

6.2. Evaluation methods

To evaluate the quality and the goodness of an extracted summary as well as to compare the results of various summarization systems, appropriate assessment methods are required. There are many numbers of methods used for summary evaluation like Pyramid method, ROUGE [79], etc.

This section shows the various evaluation methods which have been used in this survey study. Table 4 shows the different types of automatic summary evaluation and their used methods. For instance, ROUGE is a popular and widely used set of automatic evaluation metric. It consists of a package that automatically evaluates summaries and has been used by most of the surveyed systems such as [18] [15] [3] [64] [65] [74] [73] [68] [61]. In the ROUGE method, the number of common terms between a specific summary and a chosen of reference summaries is counted. Thus, it supports to evaluate the summary issue automatically. Besides, ROUGE includes five measures like ROUGE-L, ROUGE-N, ROUGE-W, ROUGE-SU, and ROUGE-S as follows.

- ROUGE-L calculates the ratio between the size of two summaries' longest common subsequence and size of reference summary.
- ROUGE-N quantifies the common N-gram units among a specific summary and a collection of reference summaries, where N determines the length of N-Grams.
- ROUGE-S evaluates the proportion of common skip bigrams among a specific summary and a collection of summaries.
- ROUGE-W It's the optimization over the simple longest common subsequence approach.
- ROUGE-SU is the average score between ROUGE-S and ROUGE-1 and it extends ROUGE-S with a counting term as unigram.

For example, ROUGE-N is a recall measure that evaluates the summary by computing the n-gram recall between the set of references summaries and the summary itself as follows [79]:

$$\boldsymbol{ROUGE-N} = \frac{\sum_{s\in MS}\sum_{n-gram\in S}\boldsymbol{Match}\ (n-gram)}{\sum_{s\in MS}\sum_{n-gram\in S}\boldsymbol{Count}\ (n-gram)} \quad (19)$$

Moreover, although some ROUGE measures correlate well with human evaluations based on DUC's human evaluation methods of English corpora at 2001, 2002 and 2003 conferences [79], the same assumption may not be correct with other methods of human evaluation or with other corpora in other languages, such as Arabic.

On the other hand, the summary task can be evaluated using other popular metrics which are precision, recall, and f-measure. They are also required to predict coverage between generated machine-summaries and human-summary automatically. It is common information retrieval metric [80] have been used by some studies such as [2] [68] [16] [73] to evaluate their generated summaries against other reference summaries automatically [80].

$$\textbf{Recall} = \frac{\textit{relevant unigrams} \cap \textit{retrieved unigrams}}{\textit{relevant unigrams}} \tag{20}$$

$$\textbf{Precision} = \frac{\textit{relevant unigrams} \cap \textit{retrieved unigrams}}{\textit{retrieved unigrams}} \tag{21}$$

Precision determines the fraction of the correct selected sentences between the system and the humans. Recall determines what ratio of the selected sentences of humans is even recognized by the machine. Precision and recall are antagonistic methods to one another. F-measure balances, precision, and recall by using a parameter β (used β = 1). Thus, F-measure is defined as follows:

$$\boldsymbol{F-measure} = \frac{(\beta^2+1)\,Precision \cdot Recall}{\beta^2\, Precision + Recall} \tag{22}$$

Although this metric has some of the drawbacks that were discussed in [80]. As there is no idealistic summary in using reference summaries for comparison, which strafes the summary for including good phrases that were not selected by the reference summary. Moreover, they strafe summaries that use sentences not selected by the reference summary as well, even if they are similar to chosen sentences. This problem is phrases common in multi-document summarization often refers to summaries that will select one phrase from a set of similar ones in the document set.

Other automatic methods were also used at short text summarization are Tri-vergence of probability distributions (TPD) [76] and Jensen–Shannon divergence [81] (JSD). TPD and JSD mechanisms of automatic evaluation have been employed to assess the quality of the short text summarization in [2] [3] [82] to ensure the summary has had the gist of the original document. They are the most attractive mechanisms used to evaluate the summary contents automatically. Obviously, these methods measure the ratio of information in the summary as for the original text. TPD theoretically allows calculating the similarity between triplets of objects. It is a statistical measure that compares three different probability distributions Q, P, and R simultaneously to ensure it captures the important retention information in the original text. The composite tri-vergence of Kullback–Leibler is defined as shown [76]:

$$\boldsymbol{T_c(P||Q||R) = \sum_{\sigma\in P} p_\sigma \log \frac{p_\sigma}{\left(\sum_{\omega\in Q} q_\omega \log\frac{q_\omega}{r_\omega}/N\right)}} \tag{23}$$

Where Q is the original document, R is an extracted summary to evaluate, and P is all other summaries of the set excepting R. ω is the terms belonging to Q; r_ω and q_ω are the probabilities of ω to occur in 'R' and 'Q' respectively, σ is the words belonging to P; p_σ is the probabilities of σ to occur in P. N represents the normalization parameter. On the other hand, The JS-divergence between two probability distributions P and Q is given by

$$\boldsymbol{J(P||Q) = \frac{1}{2}[D(P||A) + D(Q||A)]} \tag{24}$$

Where the $D(P||A)$ is defined as follow:

$$\boldsymbol{D(P||A) = \sum_w pP(w) \log_2 \frac{pP(w)}{pQ(w)}} \tag{25}$$

$A = (P/Q)/2$. In text summarization problem, the two distributions are the words in the summary and the original text. Clearly, according to Cabrera-Diego, L. et al [76], the analysis results showed that the tri-vergence can have a better performance, in comparison to the JS-divergence, when more than 16 summaries are analysed.

Table 4. Different types of evaluation used in the surveyed literature.

Reference	Evaluation method
Binwahlan M. et al [18]	ROUGE-1, ROUGE-2, and ROUGE-L
Binwahlan M. et al [15]	ROUGE-1
Peyrard, M., and Eckle-Kohler, J [71]	JS-1, JS-2, ROUGE-1, and ROUGE-2
Mosa, M. A. et al, 2017a [2]	F-measure, JSD, similarity and lengthy
Mosa, M. A. et al, 2017b [3]	ROUGE-1, TPD, similarity, and lengthy
HASSAN, O. F. 2015 [61]	ROUGE-1, and ROUGE-2
Bazghandi, M., et al, 2012 [68]	ROUGE-1, and F-measure
Chakraborti, S., and Dey, S. 2015 [73]	ROUGE, and F-measur
Binwahlan M. et al [18]	ROUGE-1, ROUGE-2, ROUGE-L, and F-measure
Binwahlan M. et al [15]	ROUGE-1, ROUGE-2, and ROUGE-L
Asgari, H, 2014 [62]	ROUGE-1
Aliguliyev, R. M., 2010 [69]	ROUGE-2, and ROUGE-SU4
Khan, A., et al, 2015 [70]	F-measure
Nikoo, M. D., et al, 2012 [64]	ROUGE-1
Mirshojaei, S. H., and Masoomi, B. 2015 [65]	ROUGE
Rautray, R., and Balabantaray, R. C. 2017 [74]	ROUGE-1, ROUGE-2, F-measure, PPV, S_{svt}, S_{acc}

7. Discussion

Few studies addressed swarm intelligence based text summarization by comparison with other traditional algorithms, such as machine learning and genetic algorithms. However, this situation is beginning to change recently when the short text summarization task has been solved using ACO [2] and graph coloring combines with ACO [3].

Mani I. [83] divided the summarization methods according to linguistic space levels into 1) shallow methods, in which the representation level does not emerge as the syntactic level & extractive summaries are commonly produced; 2) deeper methods, in which the representation level is leastwise at the semantic level and abstractive summaries are the output; and 3) hybrid approaches, which combine between the previous two methods. Although the syntactic-based approach is still popular in different natural language processing (NLP) tasks. NLP studies need to jump the curve [84] by adopting extra semantic approaches instead of depending on only syntactic ones. To the best of our knowledge, there exists no summary system based on SI that can generate abstractive summaries. It is expected that generating abstract summary will be one of the major challenges in the automatic summarization employing PSO that need to great efforts.

Most of the summarization studies followed the numerical approach, by applying a machine learning algorithm [e.g. Support Vector Machin, GAs, and mathematical regression, Naïve Bayesian, GP, and ANNs [43]]. This situation can be explained by perhaps this is because the authors do not imagine completely how to model the problem of summary in the optimization problem. Until 2016, no study has ever used SI in solving the short text summarization problem. However, this situation is beginning to change especially after the [2] [3] have been published.

On the other hand, we found that the SI has presented extremely successful approaches for solving the problem of the summary with the famous, widespread, different languages such as English, Arabic [2] [3], Persian [68], and Turkish [63] to extract the summary text. Table 2 shows the majority of types of summarization approaches associated with their types.

Table 2. Types of summaries produced by the surveyed literature and their summarization approaches.

Reference	Types of summaries	Summarization approach
Binwahlan M. et al [18]	Multi-document, and mono-lingual summaries	PSO and fuzzy
Binwahlan M. et al [15]	Single-document, and mono-lingual summaries	PSO
Peyrard, M., and Eckle-Kohler, J [71]	Multi-document, and mono-lingual summaries	ACO and genetic
Mosa, M. A. et al, 2017a [2]	Short text, and mono-lingual summaries	ACO
Mosa, M. A. et al, 2017b [3]	Short text, and mono-lingual summaries	ACO and graph coloring
HASSAN, O. F. 2015 [61]	Single-document, and mono-lingual summaries	ACO
(Güran, A, 2013)	Single-document, and mono-lingual summaries	ABC
Chakraborti, S., and Dey, S. 2015 [73]	Multi-document, and mono-lingual summaries	ABC
Binwahlan M. et al, 2009b [16]	Single-document, and mono-lingual summaries	PSO
Binwahlan M. et al, 2009c [17]	Single-document, and mono-lingual summaries	PSO
Asgari, H, 2014 [62]	Single-document, and mono-lingual summaries	PSO
Bazghandi, M., et al, 2012 [68]	Multi-document, and mono-lingual summaries	PSO and genetic mutation
Aliguliyev, R. M., 2010 [69]	Multi-document, and mono-lingual summaries	PSO and genetic mutation
Khan, A., et al, 2015 [70]	Multi-document, and mono-lingual summaries	PSO
Nikoo, M. D., et al, 2012 [64]	Single-document, and mono-lingual summaries	Bacterial Foraging
Mirshojaei, S. H., and Masoomi, B. 2015 [65]	Single-document, and mono-lingual summaries	Cuckoo Search
Rautray, R., and Balabantaray, R. C. 2017 [74]	Multi-document, and mono-lingual summaries	CSO

On the other hand, to the best of our knowledge, no summarization system has been formalized into a multi-objective optimization (MOO) task based on SI techniques. There is, however, an ongoing study that has begun in formalizing a multi-document summary problem into MOO problem [74]. In addition, although the authors in [74] named their approach MOO task, it is not clear from the paper if any new major contribution has been added to the summary text. Several previous studies can be adeptly formalized into MOO such as [2] [3] [15] [18] [74]. To be the approach more reliable and cleverness, different maximization or minimization objectives should be employed to promote the quality of the summary. The authors should be combines like the weighted sum, the normal boundary intersection (NBI) [85] methods, with SI to find a near-uniform spread of Pareto-optimal solutions in the frontier. The main purpose of using such an NBI method to determine the desired weights of the multi-objective. Consider the generic MO problem as the following [85].

$$\boldsymbol{MINIMUM\ or\ MAXIMUM}(\boldsymbol{f_1}(\boldsymbol{x}), \boldsymbol{f_2}(\boldsymbol{x}), \ldots, \boldsymbol{f_k}(\boldsymbol{x})) \tag{26}$$

$$\boldsymbol{s.t.x} \in \boldsymbol{X}, \tag{27}$$

Where the integer $k \geq 2$ is the number of objectives and the set X is the feasible set of decision vectors. The feasible set is typically defined by some constraint functions.

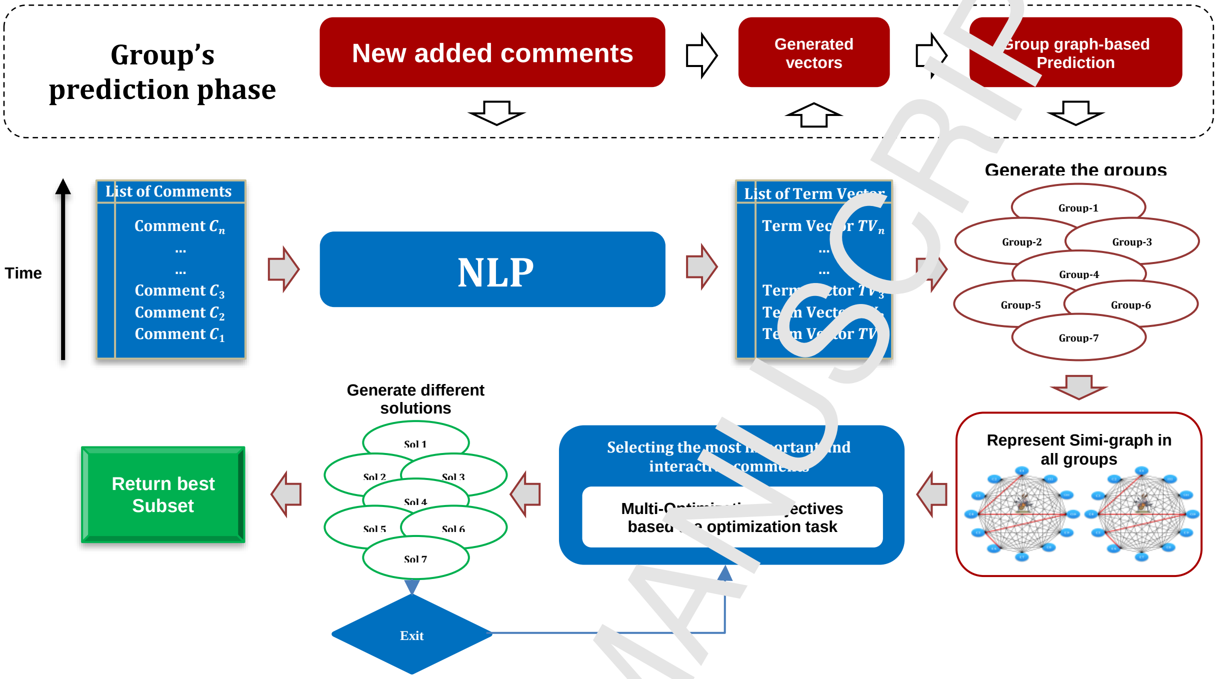


Fig. 5. A proposed real-time framework for STS.

Furthermore, to the best of our knowledge, no system based on SI techniques proposed a fully incremental algorithm aiming to provide the immediate and instant summary of real-time text streams, especially, social short messages. The main key is that designing more flexible groups that can be updated effortlessly using certain criteria that can meet the real-time needs.

In this article, a novel approach will be proposed in fig. 5. To model the summary task as an extended work of summarizing STS [2] [3]. We could re-formulate the short text summary problem into an MOO task and satisfied the real-time needs. Actually, regarding the short test summary, several features have been employed in [2] [3] can be considered objectives herein below: 1) Maximize the ratio of information (ROI) as for original text based on TPD [76]. 2) Maximize the popularity of comments (POC) using PageRank method [46]. 3) Maximize the more attractive comments by fans according to Eq. 17. 4) Minimization of the redundancy/similarity (SIM) between the extracted comments. 5) Minimization of the length (CR) of the summary. Moreover, the complex constraints need to be taken into account in the scheduling process such as the comment not admissible to be picked more one time in the summary, the value of the weight of each objective <=1, and the summation of weights are <=1.

In this section, the particulars of the approach are expressively showed. The overall process unravels to illustrative components as shown in Fig 5. Firstly, Natural Language Processing (NLP) module is used to remains the important terms in conjoined form and transform the comments to a set of n-terms separately. Secondly, we recommend usage ACO algorithm to solve such issue for the reasons that have been discussed in section 4.1. Moreover, to protect the ants in the ACO algorithm of likely cycling, it is better to shrink the solution area. One of the best techniques of constricting the solution space is the clustering method by grouping the similar comments together into the same group. On the other hand, to further accelerate the clustering task in the real-time phase, a fully incremental clustering algorithm is designed based on Naïve Bayes (NB) for instance to fulfill this issue. We already have several clusters/groups. NB is a straightforward and powerful algorithm for assigning new comments to the closest cluster (has a great similarity with its comments). It works on conditional probability. Conditional probability is the likelihood that a certain comment belongs to a certain cluster. Using the conditional probability, we can calculate the probability of an event using its prior knowledge (existing clusters). Below is the formula for calculating the conditional probability.

$$C_{NB} = \underset{c \in C}{argmax}\, P\,(C_j)\ \prod_{w \in W} P\,(w \mid C) \tag{28}$$

$$P\,(C_j) = \frac{N_c}{N} \tag{29}$$

$$P(w \mid c) = \frac{count\,(w\,, c)+1}{count\,(c)+|V|} \tag{30}$$

Where N_c is the number of comments in cluster c. N is the number of all comments. $count\ (w,\ c)$ is the number of word w in the cluster c. $count\ (c)$ is the number of all the words in cluster c. and $|V|$ is the number of unique words.

The main objective of the clustering task is to accelerate the updating process to match the real-time phase and to protect ants from a likely cycling. The newcomer comments will be assigned to the most appropriate cluster (most similar comments) with no need to cluster the task from scratch. In this work, the ACO algorithm is then employed to obtain the individual minima. Later, NBI is utilized to resolve the variable scalar values. By changing the scalar values, the Pareto frontier is constructed based on the output of the algorithms. Thirdly, a cyclic semi-graph is constructed within all groups taking into account isolating the extremely lengthy comments. Eventually, all the objectives we have, it's necessary to the trade-off between them so that do not arise some objectives and sacrifice the other. Therefore, the ACO is mixed with assistance mechanism like normal boundary instruction (NBI) algorithm to scale the different weights of multiple objectives. NBI and ACO are applied to all groups to discover which subset is interactive, representative, little, variant, and the highest ratio of information and lack redundancy. The main target beyond employing the NBI is to scalarization the different weights of different used objectives competently, where not all objective has the same importance.

Finally, it is predictable for the researchers when implementing the proposed approach, that it will achieve a noticeable performance. Due to the fact that the proposed approach is an extended work of STS approaches that have been presented by Mosa, M. A. et al, 2017a, Mosa, M. A. et al, 2017b [2] [3], the produced results should be compared with the previous two approaches. On the other hand, it has been noticed from this survey, the ACO is one of the best algorithms of swarm intelligence with regard to the convergence, accuracy, and stability, whereas the time of complexity is high. Therefore, we recommend that the ACO should be further considered in the near future works along with the consideration of some smart issues to enhance the ACO complexity. Some researchers have proved and recommended some aspects to reduce the convergence time of ACO as Huang, H. [42]. They stated that the convergence time could be reduced in some special cases when the number of ants is larger. Moreover, it can be drawn that a more pheromone rate, and pheromone rate difference per iteration can reduce the runtime of the brief ACO algorithms.

8. Conclusion and further study

The survey performed in this paper would serve as a perfect starting point for the researchers interested in the field of summarization to contribute to additional studies and mechanisms in this field. Unfortunately, this paper has shown that usage of SI methods in the summary task is quite limited, especially the developed techniques based on ACO compared to other literature summarization techniques. Moreover, an important justification and recommendation of why swarm intelligence especially ACO has been introduced. No summary problem has been formalized before into MOO task using SI. Therefore, a proposed framework is presented to cover this insufficient work. In addition, we recommended ACO mixed with NBI to solve this task. Extraction of the essence of the short messages is formulated as a multi-objective optimization (MOO) task. Later, ACO is utilized to achieve many objectives for generating a concise summary. The MOO summarization task is grasped by the NBI mechanism to resolve the variable scalar values. Since some users may demand the brief at any moment, several groups are established by possessing more similar content. Moreover, to satisfy the real-time needs, an inventive incremental algorithm is proposed to update the existing groups without having to create those groups from scratch based on Naïve Bayes mechanism. Ultimately, Evaluation methods, used corpus, and the types of summary have been presented. Indeed, there is a great opportunity for further research in summarization based on SI. Some good future work may be provided by the researchers with the help of some of the previously presented work to improve the summary generation techniques so that this research field progresses continuously.

References

1. Luhn HP (1958) the automatic creation of literature abstracts. I BM J Res Dev 2 (2):159–165
2. Mosa, M. A., Hamouda, A., & Marei, M. (2017a). Ant colony heuristic for user-contributed comments summarization. Knowledge-Based Systems, 118, 105-114.
3. Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. Artificial Intelligence Review, 47(1), 1-66.
4. Al-Saleh, A. B., & Menai, M. E. B. (2016). Automatic Arabic text summarization: a survey. Artificial Intelligence Review, 45(2), 203-234.
5. Sarkar K (2010) Syntactic trimming of extracted sentences for improving extractive multi-document summarization. J Comput 2:177–184.
6. Mosa, M. A., Hamouda, A., & Marei, M. (2017b). Graph coloring and ACO based summarization for social networks. Expert Systems with Applications, 74, 115-126.
7. Liu CY, Chen MS, Tseng CY. "Incrests: Towards real-time incremental short text summarization on comment streams from social network services". IEEE Transactions on Knowledge and Data Engineering. 27(11), 2015, 2986-3000.
8. AL-DHELAAN, M. "StarSum: A Simple Star Graph for Multi-document Summarization". In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015. p. 715-718.
9. Ouyang Y, Li W, Li S, Lu Q (2011) Applying regression models to query-focused multi-document summarization. Info Process Manag 47:227–237.
10. Fattah MA, Ren F (2009) GA, MR, FFNN, PNN and GMM based models for automatic text summarization. Compute Speech Lang 23:126–144. doi:10.1016/j.csl.2008.04.002.
11. Riedhammer K, Fav re B, Hakkani-Tur D (2010) Long story short- global unsupervised models for keyphrase based meeting summarization. Speech Commun 52:801–815.
12. Song W, Choi LC, Park SC, Ding XF (2011) Fuzzy evolutionary optimization modelling and its applications to unsupervised categorization and extractive summarization. Expert Syst Appl 38:9112–9121.
13. Chali Y, Hasan SA (2012) Query focused multi-document summarization: automatic data annotations and supervised learning approaches. Nat Lang Eng 18:109–145
14. El-Fishawy, N., Hamouda, A., Attiya, G. M., & Atef, M. (2014). Arabic summarization in twitter social network. Ain Shams Engineering Journal, 5(2), 411-420.
15. Binwahlan, M., S, N. and Suanmali, L. (2009a). Swarm based features selection for text summarization. International Journal of Computer Science and Network Security IJCSNS. 9(1), 175–179.
16. Binwahlan, M., S, N. and Suanmali, L. (2009b). Swarm based text summarization. In Computer Science and Information TechnologySpring Conference, 2009. IACSITSC'09. International Association of. IEEE, 145–150.
17. Binwahlan, M., S, N. and Suanmali, L. (2009c). Swarm Diversity Based Text Summarization. In Neural Information Processing. Springer, 216–225.
18. Binwahlan, M. S., Salim, N., & Suanmali, L. (2009d). Fuzzy Swarm Based Text Summarization 1.
19. Franco-Salvador, M., Rosso, P., & Montes-y-Gómez, M. (2016). A systematic study of knowledge graph analysis for cross-language plagiarism detection. Information Processing & Management, 52(4), 550-570.
20. AlSanie, W., Touir, A., & Mathkour, H. (2005). Towards an infrastructure for Arabic text summarization using rhetorical structure theory. Master's thesis, King Saud University, Riyadh.
21. Kikuchi Y, Hirao T, Takamura H, Okumura M, Nagata M (2014) Single document summarization based on nested tree structure. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, vol. 2, pp. 315–320
22. Parveen D, Strube M (2015) Integrating importance, non-redundancy and coherence in graph-based extractive summarization. In: Proceedings of the 24th international conference on artificial intelligence. AAAI Press. Pp. 1298–1304.
23. Yao JG, Wan X, Xiao J (2015a) Compressive document summarization via sparse optimization. In: Proceedings of the 24th international conference on artificial intelligence. AAAI Press. Pp. 1376–1382.
24. Ibrahim A, Elghazaly T (2013b) Rhetorical representation and vector representation in summarizing Arabic text. Natural language processing and information systems, lecture notes in computer science, vol 7934.Springer, Berlin, pp 421 – 424.
25. Ibrahim A, Elghazaly T (2013a) Improve the automatic summarization of Arabic text depending on rhetorical structure theory. I n: 2013 12th Mexican international conference on artificial intelligence (MICAI), pp 223 – 227.
26. Li C, Liu Y, Zhao L (2015a) Using external resources and joint learning for bigram weighting in ilp-based multi-document summarization. In: Proceedings of NAACL-HLT, pp 778–787.

27. El-Ghannam F, El-Shishtawy T (2013) Multi-topic multi-document summarizer. Int J Comput Sci Inf Technol 5(6): 77 – 90
28. Hong K, Marcus M, Nenkova A (2015) System combination for multi-document summarization. In: Proceedings of the 2015 conference on empirical methods in natural language processing. pp 107–117
29. Oufaida H, Nouali O, Blache P (2014) Minimum redundancy and maximum relevance for single and multi-document arabic text summarization. J King Saud Univ Comput Inf Sci 26(4):450–461 special Issue on Arabic NLP.
30. Ma, J., Jiang, X., & Gong, M. (2018). Two-phase clustering algorithm with density exploring distance measure. CAAI Transactions on Intelligence Technology, 3(1), 59-64.
31. Peng H, Long F, Ding C (2005) Feature selection b ased on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Tr ans Pattern Anal Mach Intell 27(8):1226–1238.
32. Wang, B., Zhang, J., Liu, Y., & Zou, Y. (2017). Density peaks clustering based integrate framework for multi-document summarization. CAAI Transactions on Intelligence Technology, 2(1), 26-30.
33. Marcus A, Bernstein MS, Badar O, Karger DR, Madden S, Miller RC. "TwitInfo: aggregating and visualizing microblogs for event exploration". Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI'11). 2011, 227–236.
34. Weng JY, Yang CL, Chen BN, Wang YK, Lin SD. "Imass: An intelligent microblog analysis and summarization system". Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACLHLT'11). 2011, 133–138.
35. Sharifi B., Hutton MA, Kalita JK. "Experiments in Microblog Summarization". Proc. of the 2nd IEEE International Conference on Social Computing (SocialCom'10). 2010, 49–56.
36. Sankaranarayanan J., Samet H., Teitler B. E., Lieberman M. D., and Sperling J. TwitterStand: News in Tweets. Proc. of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM-GIS'09), pages 42–51, 2009.
37. Takamura H, Yokono H, Okumura M. "Summarizing a document stream". In European Conference on Information Retrieval. Springer Berlin Heidelberg. 2011, 177-188.
38. Chakrabarti D, Punera K. "Event Summarization Using Tweets". Proc of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11). 2011, 66–73.
39. Becker H, Mor N, and Luis G. "Selecting Quality Twitter Content for Events". Proc. of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11). 2011, 442–445.
40. Rosa, K. D., Shah, R., Lin, B., Gershman, A., & Frederking, R. (2011). Topical clustering of tweets. Proceedings of the ACM SIGIR: SWSM.
41. Perez-Tellez F., Pinto D., Cardiff J., and Rosso P. On the Difficulty of Clustering Company Tweets. Proc. of the 2nd International Workshop on Search and Mining User Generated Contents (SMUC'10), pages 95–102, 2010.
42. Chakrabarti, D., & Punera, K. (2011). Event Summarization Using Tweets. ICWSM, 11, 66-73.
43. An, H., Wang, D., Pan, Z., Chen, M., & Wang, X. (2018). Text segmentation of health examination item based on character statistics and information measurement. CAAI Transactions on Intelligence Technology, 3(1), 28-32.
44. Liu F., Liu Y., and Weng F. Why is "SXSW" trending? Exploring Multiple Text Sources for Twitter Topic Summarization. Proc. Of the ACL/HLT Workshop on Language in Social Media (ACL-LSM'11), pages 66–75, 2011.
45. Zahra, K., Ostermann, F. O., & Purves, R. S. (2017). Geographic variability of Twitter usage characteristics during disaster events. Geo-spatial information science, 20(3), 231-240. 100
46. Khabiri, E., Caverlee, J., & Hsu, C. F. (2011, July). Summarizing User-Contributed Comments. In ICWSM.
47. Panigrahi, B. K., Shi, Y., & Lim, M. H. (Eds.). (2011). Handbook of swarm intelligence: concepts, principles and applications (Vol. 8). Springer Science & Business Media.
48. Belal M., Gaber J., El-Sayed H., and Almojel A., Swarm Intelligence, In Handbook of Bioinspired Algorithms and Applications. Series: CRC Computer & Information Science. Vol. 7. Chapman & Hall Eds, 2006. ISBN 1-58488-477-5.
49. Beni, G., & Wang, J. (1993). Swarm intelligence in cellular robotic systems. In Robots and Biological Systems: Towards a New Bionics? (pp. 703-712). Springer Berlin Heidelberg.
50. Dorigo, M., Maniezzo, V., & Colorni, A. (1991). The ant system: An autocatalytic optimizing process.
51. Colorni, A., Dorigo, M., & Maniezzo, V. (1992, September). An Investigation of some Properties of an``Ant Algorithm''. In PPSN (Vol. 92, pp. 509-520).
52. Kennedy, J., & Eberhart, R. (1995). Synthetic structure of industrial plastics (Book style with paper title and editor). In Proceeding of the 1995IEEE International Conference on Neural Networks (pp. 1942-1948).
53. Eberhart, R., & Kennedy, J. (1995, October). A new optimizer using particle swarm theory. In Micro Machine and Human Science, 1995. MHS'95. Proceedings of the Sixth International Symposium on (pp. 39-43). IEEE.
54. Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization (Vol. 200). Technical report-tr06, Erciyes University, engineering faculty, computer engineering department.
55. Van den Bergh, F., & Engelbrecht, A. P. (2002, October). A new locally convergent particle swarm optimiser. In Systems, Man and Cybernetics, 2002 IEEE International Conference on (Vol. 3, pp. 6-pp). IEEE.
56. Özgür Yeniay (2014). COMPARATIVE STUDY OF ALGORITHMS FOR RESPONSE SURFACE OPTIMIZATION. Mathematical and Computational Applications, Vol. 19, No. 1, pp. 93-104.

57. Santra, D., Mondal, A., & Mukherjee, A. (2016). Study of Economic Load Dispatch by Various Hybrid Optimization Techniques. In Hybrid Soft Computing Approaches (pp. 37-74). Springer India.
58. Yang, Xin-She. (2014): Swarm intelligence based algorithms: a critical analysis. Evolutionary intelligence 7.1 pp.17-28.
59. Gutjahr, W. J. (2000). A graph-based ant system and its convergence. Future generation computer systems, 16(8), 873-888.
60. Prakasam, A., & Savarimuthu, N. (2016). Metaheuristic algorithms and probabilistic behaviour: a comprehensive analysis of Ant Colony Optimization and its variants. Artificial Intelligence Review, 45(1), 97-130.
61. HASSAN, O. F. (2015). TEXT SUMMARIZATION USING ANT COLONY OPTIMIZATION ALGORITHM (Doctoral dissertation, Sudan University of Science and Technology).
62. Asgari, H., Masoumi, B., & Sheijani, O. S. (2014, February). Automatic text summarization based on multi-agent particle swarm optimization. In Intelligent Systems (ICIS), 2014 Iranian Conference on (pp. 1-5). IEEE.
63. Güran, A, NİLGÜN G B, and Mustafa Z G. "Efficient feature integration with Wikipedia-based semantic feature extraction for Turkish text summarization." Turkish Journal of Electrical Engineering & Computer Sciences 21.5 (2013): 1411-1425.
64. Nikoo, M. D., Faraahi, A., Hashemi, S. M., & Erfani, S. H. (2012). A Method for Text Summarization by Bacterial Foraging Optimisation Algorithm. IJCSI International Journal of Computer Science Issues, 9(4), 36-40
65. Mirshojaei, S. H., & Masoomi, B. (2015). Text Summarization Using Cuckoo Search Optimisation Algorithm. Journal of Computer & Robotics, 8(2), 19-24.
66. Donis-Díaz, C A., Rafael B, and Janusz K. "Using ant colony optimisation and genetic algorithms for the linguistic summarization of creep data." Intelligent Systems' 2014. Springer International Publishing, 2015. 81-92.
67. Yadav, R. K., & Singh, S. (2016). A New Approach to Automated Summarization based on Fuzzy Clustering and Particle Swarm Optimization. International Journal of Computer Applications, 148(1).
68. Bazghandi, M., Tabrizi, G. T., Jahan, M. V., & Mashahd, I. (2012). Extractive Summarization Of Farsi Documents Based On PSO Clustering. jiA, 1, 1.
69. Aliguliyev, R. M. (2010). Clustering Techniques and Discrete Particle Swarm Optimization Algorithm for Multi-Document Summarization. Computational Intelligence, 26(4), 420-448.
70. Khan, A., Salim, N., & Kumar, Y. J. (2015). A framework for multi-document abstractive summarization based on semantic role labelling. Applied Soft Computing, 30, 737-747.
71. Peyrard, M., & Eckle-Kohler, J. A General Optimization Framework for Multi-Document Summarization Using Genetic Algorithms and Swarm Intelligence.
72. Haghighi, A., & Vanderwende, L. (2009, May). Exploring content models for multi-document summarization. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 362-370). Association for Computational Linguistics.
73. Chakraborti, S., & Dey, S. (2015, October). Product news summarization for competitor intelligence using topic identification and artificial bee colony optimization. In Proceedings of the 2015 Conference on research in adaptive and convergent systems (pp. 1-6). ACM.
74. Rautray, R., & Balabantaray, R. C. (2017). Cat swarm optimization based evolutionary framework for multi document summarization. Physica A: Statistical Mechanics and its Applications, 477, 174-186.
75. Alaa El-Dine A.H., F. El-zahraa El-taher (2012). Automatic Summarization of Arabic post, the first International Conference for Faculty of Computers and Information.
76. Cabrera-Diego, L. A., Torres-Moreno, J. M., & Durette, B. (2016, June). Evaluating Multiple Summaries without Human Models: A First Experiment with a Trivergent Model. In International Conference on Applications of Natural Language to Information Systems (pp. 91-101). Springer International Publishing.
77. Lloret, E., & Palomar, M. (2012). Text summarisation in progress: a literature review. Artificial Intelligence Review, 37(1), 1-41.
78. El-Haj, M., Kruschwitz, U., & Fox, C. (2011, December). Exploring clustering for multi-document Arabic summarisation. In Asia Information Retrieval Symposium (pp. 550-561). Springer Berlin Heidelberg.
79. Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. In Text summarization branches out: Proceedings of the ACL-04 workshop (Vol. 8).
80. Nenkova A (2006) Summarization evaluation f or text and speech: issues and approaches. I n: INTERSPEECH.
81. Louis, A., & Nenkova, A. (2009, August). Automatically evaluating content selection in summarization without human models. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1 (pp. 306-314). Association for Computational Linguistics.
82. Mosa, M. A., Hamouda, A., & Marei, M. (2017c). How can Ants Extract the Essence Contents Satellite of Social Networks. LAP Lambert Academic Publishing, ISBN: 978-3-330-32645-3.
83. Mani, I. (2001). Automatic summarization (Vol. 3). John Benjamins Publishing.
84. Cambria E, White B (2014) Jumping nlp curves: a review of natural language processing research [review article]. Comput Intell Mag I EEE 9(2):48–57
85. Das, I., & Dennis, J. E. (1998). Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. SIAM Journal on Optimization, 8(3), 631-657.