Future Generation Computer Systems 114 (2021) 448–480

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Explicit aspects extraction in sentiment analysis using optimal rules combination

Mohammad Tubishat [a], Norisma Idris [a,*], Mohammad Abushariah [b]

[a] Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[b] Computer Information Systems Department, King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan

ARTICLE INFO

Article history:
Received 19 November 2019
Received in revised form 19 July 2020
Accepted 16 August 2020
Available online 22 August 2020

Keywords:
Sentiment analysis
Dependency-based rules
Pattern-based rules
Improved whale optimization algorithm
Explicit aspect extraction

ABSTRACT

Aspect extraction represents a core task of aspect-based sentiment analysis. This study presents a supervised aspect extraction algorithm for explicit aspect extraction from formal and informal texts. To accomplish the new algorithm, 126 aspect extraction rules are combined to cover both formal and informal texts, because customer reviews are a mix of formal and informal texts. These 126 rules include certain dependency-based rules and pattern-based rules from previous studies, in addition to newly developed rules intended to overcome prior rules' weaknesses. In addition, many aspect extraction rules have remained unexplored by previous studies. However, many of these 126 rules are irrelevant and should be removed. Thus, a prober selection of the included rules is required. Therefore, in this study we also improved the Whale Optimization Algorithm (WOA) to address rules selection problem with an improved algorithm called improved WOA (IWOA). Two major improvements were included into IWOA. The first improvement is the development of a new update equation based on Cauchy mutation to improve WOA population diversity. The second improvement is the development of a new local search algorithm (LSA) to solve WOA local optima. The IWOA algorithm is applied on the full set of rules to select best rules subset and remove low quality rules. Finally, a new pruning algorithm (PA) has been developed to remove incorrect aspects and retain correct aspects. The Results on seven benchmark datasets demonstrate that IWOA+PA outperforms all other state-of-the-art baseline works and most recent works.

1. Introduction

The emergence of different online shopping websites has changed our daily habits. Websites such as Amazon, Taobao, and countless others allow people to post and share their reviews and feedback about the purchased products [1]. Many customers refer to these reviews before making their buying decisions. About 70% of customers find these reviews as one of the major trusted sources in making their decisions. Manufacturers also use these reviews to analyze the defects of their products, as the pre-discovery of product defects can maximize the reputation of the manufacturers, which in turn increases their sales [2]. In addition, manufacturers use these reviews to explore the most important product aspects by customers. According to Akalamkam and Mitra [3], pre-purchase decisions by a huge numbers of customers are based on conducting pre-search through online reviews provided by previous customers. As about 90% of consumers depend on online reviews for making a purchase decision [4], this habit is now considered as a major step by customers before deciding to buy any products online. However, the process of reviewing and retrieving the reviews by customers is time consuming and cumbersome due to the huge volume of available information. Thus, automatic solutions to this process have been developed which are collectively categorized as Sentiment Analysis (SA).

SA is defined as the detection of authors' opinion expressions toward entities or their facets [5]. Millions of people daily express and share their opinions about different commercial products [6]. Therefore, SA has attracted popularity by research community for processing the huge amount of available data on social media such as blogs, collaborative media, social networks, and online communities [5,7].

SA is a branch of affective computing research [8]. The primary purpose of SA is to classify text, audio, and video [9] into either negative or positive, or even neutral opinions [10]. According to Cambria et al. [11], SA is a big suitcase which encapsulating several natural language processing (NLP) tasks. These NLP tasks include word polarity disambiguation [12]; concept extraction [13]; sarcasm detection [14]; aspect extraction [15]; and subjectivity

* Corresponding author.
E-mail addresses: mtubishat@siswa.um.edu.my (M. Tubishat), norisma@um.edu.my (N. Idris), m.abushariah@ju.edu.jo (M. Abushariah).

https://doi.org/10.1016/j.future.2020.08.019
0167-739X/

detection [16]. SA has been applied in such diverse applications as dialogue systems [17]; e-health [18]; community detection [19]; financial forecasting [20]; manufacturing applications [21]; and countless others.

SA can be performed at different levels including the document, sentence, or aspect (feature) levels [22]. At the document and sentence levels, the task is based on extracting all opinion words in a sentence or document, after which it outputs the opinion to the whole document or sentence. Whereas, aspect level sentiment analysis is concerned with the extraction of product aspects and the opinion words mentioned in the reviews toward each aspect [22]. Analysis at the document and sentence levels is ultimately based on the overall document or sentence without considering specific analysis to each aspect, while SA at aspect level considers each aspect and it associates each aspect with its related opinion words. One of the most important tasks in aspect level sentiment analysis is the aspect or feature extraction step [22], where the aspect can be either explicit or implicit. The aspect is considered as explicit if it is mentioned explicitly in the text. For example, in the review, *"the screen is very nice"*, the aspect *screen* is mentioned explicitly. On the other hand, if the aspect is mentioned implicitly in a text, it is known as implicit aspect. For example, in a review about a hand phone, *"the phone is expensive"*, the opinion word *expensive* is an indicator for *price* implicit aspect of the hand phone [23]. The implicit aspect extraction methods which were developed by the previous studies can be unsupervised, semi-supervised, or supervised [24].

Aspect extraction plays an important role in aspect-based sentiment analysis [25–27], and represents an important phase for product and feature ranking applications [1,28–30]. In addition, aspect extraction represents the most important phase in an aspect retrieval system [31]. Several studies have been conducted into explicit aspect extraction. The extraction techniques used in previous studies can be classified as either unsupervised such as [23,32–36], semi-supervised [28,37–43], or supervised [44–49].

Many previous studies used either a dependency-based approach, such as [36,39,50–52] or a pattern-based approach [34,53–55]. Extraction types based on pattern-based or dependency-based rules give promising results. However, according to previous studies [55–58], aspect extraction approaches based on dependency relations have the problem of generating error results, which occur as the dependency relations accuracy are based on the grammatical correctness of the reviews, while not all online reviews follow English grammar rules. Also, these reviews usually are mixed types of structured and unstructured text. In addition, there is no restriction from the website on the customer to follow language rules when they post his/her review. Customers may follow the language rules in writing and sometimes may violate some of it [55]. The pattern-based approach is good for such unstructured text, as the patterns mimic the ways users write their reviews [55].

Based on our preliminary findings, there are some problems in the current works. For example, methods which are based on topic modeling can only extract the general aspects and ignore the fine-grained aspects. In methods which are based on statistics for aspect extraction, these methods results are based on the size of the datasets used. Therefore, if the size of the dataset is small, then the results are unreliable. In addition, some methods utilized bootstrapping method for aspect extraction. However, this method suffers from the problem of error propagation and extracts many incorrect aspect words. Moreover, aspect extraction which is based on using Conditional Random Fields (CRF) cannot be applied to sentences with long patterns. Furthermore, works which are based on using dependency relations or syntactic pattern rules are characterized by some issues namely: (1) limited number of rules were used for extraction; (2) many aspect extraction rules were not explored; (3) the existence of irrelevant rules which result in extracting many incorrect aspect words; and (4) the unavailability of suitable aspects pruning algorithm. Therefore, further works are required to overcome these issues.

Thus, in this paper, the main aim is to propose an aspect extraction algorithm to extract aspects from both formal and informal texts. To achieve this aim, we have come up with a number of contributions. These contributions (in bold) and its importance can be summarized by following key points:

- ***Many new extraction rules are developed in this study***. *These rules are developed to overcome the weakness of existing rules. In addition, to overcome the unexplored aspect extraction rules by previous studies.*
- ***The combination of rules from different types***. *These rules include the new developed rules in this study and the rules used from previous studies. This rules combination is necessary to take advantage of both types of rules at the same time. In addition, they cover both formal and informal texts.*
- ***Improved Whale Optimization Algorithm (IWOA) for rules selection problem.*** *The development of IWOA algorithm is important to help in selecting the best subset of rules from the full set of rules. If we have n rules, then if we want to try all possible rules subsets it will be NP-hard problem, because it requires 2^n to find the optimal rules subset from all possible subsets. Therefore, IWOA will be used to solve this problem.*
- ***Two major improvements were included into WOA***. *These improvements are to solve WOA weakness and make it fit for rules selection problem.*
- ***The development of new update equation using Cauchy Mutation***. *This first improvement to WOA is necessary to improve the diversity of WOA (rules diversity) by diversifying the combination of each IWOA solution (selected and unselected rules).*
- ***The development of new local search algorithm (LSA).*** *LSA is combined with WOA to improve the current best solution (exploitation) [this will help select the rules which further improve the extraction performance], then it avoids WOA from being stuck at local.*
- ***The development of three phases aspects Pruning Algorithm (PA).*** *The PA algorithm is important to purify the candidate aspects. In addition, it can improve the extraction performance by solving the problem of pruning less frequent occurring aspects which were pruned by previous studies.*

Therefore, as detailed above, this study proposes a novel IWOA+PA algorithm for aspect extraction which is characterized by novelty in a number of points, such as the development of new aspect extraction rules, the extraction of aspect from reviews regardless if its structured or not, the application of IWOA for rules selection, the improvements to WOA, and the three phase pruning algorithm.

The proposed IWOA+PA algorithm utilized supervised approach for aspect extraction, since the selection of optimal rules is based on training dataset. The proposed methodology of IWOA+PA will be performed on a number of steps. First, IWOA selects the optimal subset of rules from the full set of rules based on a training dataset. Then, the best selected rules will be applied on the testing dataset. The candidate aspects which are the results of the best selected rules by IWOA will go to next step. These candidate aspects will be purified by using PA algorithm. The details of these steps will be discussed in the following sections.

The remainder of the paper is organized as follows. Section 2 discusses related works on extraction methods. Section 3 presents the workflow of the proposed aspect extraction algorithm. Section 4 presents the aspect extraction rules used in this study.

Section 5 explains the details of IWOA algorithm while Section 6 describes the details of the aspect pruning algorithm. Section 7 presents the experimental results and analysis. Finally, Section 8 concludes the finding of this study.

2. Related works on aspect extraction

This section presents the current and state-of-the-art works which were conducted for explicit aspect extraction. Several works have been conducted for explicit aspect extraction which can be grouped into three approaches: supervised, unsupervised, and semi-supervised. The supervised method requires training data and the semi-supervised requires little training data, while the unsupervised approach does not require any training data [22]. The details of each method with an example of works conducted using it are presented in the following subsections.

2.1.1. Unsupervised extraction methods

As mentioned earlier, no training data is required to extract the explicit aspects in unsupervised extraction methods. The following studies represent some examples of work conducted using unsupervised method based on the technique they used.

The following are examples of unsupervised works which utilized frequency-based method for aspect extraction are as follows: In a work by Hu and Liu [23], they utilized frequency-based method for aspect extraction. They considered noun and noun phrases as candidate aspects and the nearest adjectives as its opinion words. They extracted frequent aspects using Apriori algorithm by finding frequent nouns and noun phrases. The reported performance of the proposed method on customer review datasets [23] was 80% recall and 72% precision. Later, Popescu and Etzioni [32] improved and extended the previous work of Hu and Liu [23]. Their proposed method was based on frequency and Pointwise mutual information (PMI). They used web similarity measure to improve the precision of aspect extraction performance based on PMI. PMI was calculated for each candidate aspect and based on PMI value they proved or discarded the given aspect. The experiment was conducted on a customer review dataset [23] with 77% recall and 94% precision. The work conducted by [59] represents another improvement over [23] work. A new algorithm called High Adjective Count (HAC) was developed to extract features. HAC is mainly based on the concept that the noun which is frequently opinionated by reviewer can be an aspect. HAC starts by extracting all nouns and adjectives and set a counter for each noun which is initialized by zero. If it finds the closest adjective to the corresponding noun on its left or right side the counter will be incremented by 1. In the end, the nouns with a score higher than a pre-specified threshold will be considered as an approved aspects and other aspects will be discarded. The best achieved results was on DVD player dataset from customer review datasets [23] with 80% precision. However, a common limitation of these studies that they pruned low frequent aspects.

Some examples of unsupervised works which utilized dependency relation rules for aspect extraction are the following. In the work of Zhuang et al. [60], the proposed method is based on using word frequencies and four dependency relation rules for aspect extraction. The experiments were conducted on movie reviews and the best result was 52.9% using F-measure. Hai et al. [35] used one domain specific corpus and another independent domain corpus with three relation rules to extract explicit aspects from Chinese reviews. The technique started by extracting all candidate aspects using three syntactic dependency rules. In addition, two relevance measures were defined, namely extrinsic-domain relevance (EDR) and intrinsic-domain relevance (IDR). EDR finds how relevant the candidate aspect is to the independent corpus, while IDR finds how relevant the candidate aspect is to the domain specific corpus. If the EDR value of a given aspect is less than a specified EDR threshold, and IDR is greater than a specified IDR threshold, then the given aspect will be approved as a correct aspect. These experiments were conducted on two Chinese datasets about hotels and cellphones, and the results were 52.26% F-measure on hotel dataset and 63.6% F-measure on cellphone dataset. However, this work also used a statistic approach, and the results were not reliable when the dataset used was small [39].

Wu et al. [61] combined dependency rule-based approach with deep gated recurrent unit (GRU) network for aspects extraction. They reported F-measure results of aspects extraction on Restaurant and Laptop domains of SemEval 2014 datasets as 76.15% and 60.75% respectively. In addition, 63.36% on SemEval 2015 and 64.24% on SemEval 2016. However, they missed many correct aspects as they considered aspects as noun or noun phrase only, and they did not take care of semantic rules. In work by Luo et al. [62], they proposed a method called ExtRA (Extraction of Prominent Review Aspects). They used WordNet, Probase as a Knowledge source, and number of dependency relation rules. However, not all possible dependency extraction rules were explored and miss many correct aspects as they considered aspects as noun or noun phrase only. Furthermore, a common limitation of these studies that many incorrect aspects were resulted because they did not perform aspects pruning.

A study by Dragoni et al. [63] constructed a dependency graph and used three dependency rules to extract aspects based on that graph. They performed the experiments on SemEval 2015 and SemEval 2016 datasets about restaurants and laptop reviews. The achieved results using F-measure were 60% on restaurant dataset and 51% on laptop dataset from SemEval 2015. In addition, the results using F-measure were 67% on restaurant dataset and 57% on laptop dataset from SemEval 2016. Another example by Li et al. [64] proposed a bootstrapping method using a number of dependency relation rules for extracting aspects and its related opinion words. They used six dependency relations rules combined with sentiment information for extracting explicit aspects, where these relations were based on the grammatical relation between opinion words and aspects. The new aspects were used to generate new relations rule through the bootstrapping approach used. Moreover, two measures were defined including Prevalence and Reliability to evaluate the confidence of the extracted aspects and the relations rule used for extraction. The extracted aspects were grouped into a number of clusters with a weight allocated for each cluster. An aspect with low confidence was saved if it is grouped in the cluster which contains aspects with high confidence values. The experiment was conducted on customer review datasets [23] with a result was 89% using F-measure. Furthermore, Samha [65] proposed a method for aspect extraction which is mainly based on dependency relation rules for aspect extraction. She defined five new dependency relation rules for aspect extraction, and these five rules were combined with eleven dependency rules from previous studies [39,50,66,67]. She used some preprocessing such as removing symbols, lemmatization, POS tagging, and finding of dependency relations using Stanford dependency parser. The experiments were conducted on customer reviews datasets [23] and the best result achieved was 71.8% using F-measure. However, these methods which applied dependency rules for aspect extraction share the following limitations: (1) not all used rules were efficient for extraction; (2) not all possible dependency extraction rules were explored; and (3) the extraction of aspects based on dependency rules has a problem when dealing with informal written reviews. In addition, most of the methods either do not apply aspects pruning or use aspect pruning and need some improvements.

Some examples of unsupervised works which utilized syntactic patterns for aspect extraction are as follows. In [68], the proposed method is mainly based on the combination of Natural Language Processing (NLP) technique with statistical techniques for aspect extraction. Four pattern rules were used to extract aspects. The aspect approved is based on its frequency in both the reviews and the background corpus. The proposed technique was applied on Chinese dataset about mobile reviews with the best achievement of 74.04% using F-measure. Additionally, Moghaddam and Ester [33] proposed an extraction method based on combining frequency-based approach with several part-of-speech (POS) patterns to extract and filter incorrect aspects. They improved on Hu and Liu [23]'s work by combining frequency-based approach with several POS patterns to extract and filter incorrect aspect words. They also used known product aspects together with product reviews as input to the system for extracting explicit aspects. However, many aspects could not be extracted, as it was based on previous knowledge only. The best results were 80% precision and 87% recall on a customer review dataset [23]. In [69], aspects were extracted by using bootstrapping method and the process started by POS tagging all sentences in the reviews. From the POS tagged sentence, the parts that matched one of the four patterns were extracted as candidate aspects. In the following step, heuristics rules were applied for removing aspects with no opinion word in the sentence that contain that aspect. A new score was proposed called A-Score which is based on the mutual relation between words and the frequency of aspect. The algorithm selected the aspect with the highest A-score and added it to the seed sets. Two pruning methods were used including subset-support pruning by removing meaningless word from multi words aspects, and superset-support pruning, which is based on removing single word aspects that is a part of multiword aspects were applied. The proposed technique was applied on a customer review dataset [23] with the reported performance using F-measure was 72.9%. Furthermore, Htay and Lynn [34] defined eight pattern rules and considered all nouns and noun phrases as correct aspects. The experiments were conducted on customer reviews datasets [23] with an F-measure of 79%. Also, Maharani et al. [53]'s work based on using pattern-based method for aspect extraction. In their study, they defined new patterns for aspect extraction and used patterns from the previous studies including patterns from [34,70]. They tried to use different combination of these patterns rules and the best result was achieved when all patterns were used. The experiments were conducted on customer reviews datasets from [23,71] with the best achievement using F-measure was 67.2%. In a recent study by Asghar et al. [54], pattern-based method was also used to extract aspects and their related opinion words. They defined ten patterns extraction rules for aspect extraction. The experiments were conducted on customer reviews datasets [23] where the F-measure was 77.16%. However, these methods which applied pattern rules for aspect extraction share the following limitations: (1) not all used rules are efficient for extraction; (2) not all possible extraction pattern rules are explored; and (3) pattern-based rules are better than dependency-based rules for informal text, but for long sentence dependency-based rules are better. In addition, most of the methods either do not apply aspects pruning or use aspect pruning and need some improvements.

Some examples of unsupervised works which utilized topic modeling method for aspect extraction are as follows. In work by Brody and Elhadad [72], they considered each sentence in the reviews as one separate document to solve the problem of low frequent aspects. They applied local Latent Dirichlet Allocation (LDA) on these collections of documents for aspect extraction. Moreover, the extracted topics from each document were considered as the approved aspects. The experiments were conducted on restaurant reviews dataset with a precision of 86.8%, based on a review conducted by Rana et al. [73]. In another work by Moghaddam and Ester [74], they also used topic model for extracting aspects. They extended the standard LDA to Interdependent Latent Dirichlet Allocation (ILDA) model for extracting aspects and their sentiments. This was based on the assumption that there was an interdependency relation exists between aspect and opinion words. They conducted experiments on a number of electronic products, using Average Rand Index for aspects evaluation. The value achieved was 83% based on review conducted in [75]. In a study by Chen et al. [76], they also used topic model method for aspect extraction. They combined topic modeling with prior knowledge for extracting aspects. The prior knowledge was learned by applying LDA on multiple reviews domains, learnt the aspects in each domain, and learned the shared aspects among domains as prior knowledge. The experiments were conducted on electronics products from different domains using an evaluation metric precision@n where n was 5 and 10. The reported performance results were based on review conducted in [75], with average precision@5 of 90% and average precision@10 of 85%. However, common limitation of these works which utilized topic modeling, that topic modeling can only extract the general aspects and ignore the fine-grained aspects.

García-Pablos et al. [77] proposed W2VLDA technique for aspects extraction. They combined LDA topic model, maximum entropy classifier and hyper-parameters using continuous word embeddings. They reported F-measure results of aspects extraction on Restaurant domain of SemEval 2016 datasets as 72%. Ekinci and İlhan Omurca [78] proposed a model called Concept-LDA, this model combined LDA topic model with semantic knowledge based on the named entities and concepts which were extracted from Babelfy. The achieved results using F-measure was 75% on 10 public datasets. However, topic modeling can only extract the general aspects and ignore the fine-grained aspects. Chauhan et al. [79] combined rules-based method with bidirectional LSTM for aspects extraction. the rules were used to extract nouns and noun phrases as possible aspects. These candidate aspects were also used in training of attention-based bi-LSTM model to improve the extraction of noun phrases. The achieved results using F-measure were 79.01% on restaurant dataset and 73.30% on laptop dataset from SemEval 2016. However, they missed many correct aspects as they considered noun or noun phrase only. Also, they did not extract aspects in sentence without opinion words. Sokhin et al. [80] proposed neural network model that combined convolutional multi-attention mechanism for aspects extraction. They reported average F-measure results of aspects extraction on Restaurant domain of SemEval 2016 datasets as 80.33%. However, this work is unable to identify many correct aspects. Liang et al. [81] proposed Multi-level Sentence-Word Interaction Transfer (MSWIT) model for aspects extraction. In this model, they utilized multi-layer Bi-LSTM, multi-head attention layer, CRF, and multiple fully connected layers. They reported F-measure results of aspects extraction on SemEval 2014–2015 datasets as 61.73% and 46.17% respectively on laptop datasets. In addition, 46.81% and 41.19% on Restaurant datasets. Furthermore, a common limitation of these studies that many incorrect aspects were resulted because they did not perform aspects pruning.

2.1.2. Semi-supervised extraction methods

In the semi-supervised extraction methods as mentioned earlier, little training data is required to extract the mentioned explicit aspects. The following works represent some examples of works conducted using semi-supervised method.

Some examples of semi-supervised works which utilized dependency relation for aspect extraction are as follows: A study by Wu et al. [37] extracted aspects expression and their opinion expression using dependency parsing method at phrase level. They considered only noun phrases (NPs) and verb phrases (VPs) as possible candidate aspects expressions. In addition, they considered opinion words as the words surrounding candidate aspect expression based on existing opinion words dictionary. In addition, they used a tree kernel with Support vector machine (SVM) for extracting the relations between the candidate aspects and opinion words. The experiments were conducted on customer reviews datasets [23,82] datasets with a result of 57% using F-measure. In [39,83] based on propagation and dependency relation rules method for extracting aspects, a new algorithm called Double Propagation (DP) was proposed. The DP algorithm is based on the idea that a syntactic relation exists between aspects and opinion words. They defined eight dependency relations rules, where they considered only nouns and noun phrases as possible aspects and adjectives as possible opinion words. DP requires seed opinion words to start propagation. DP also uses these opinion words with dependency rules to extract new aspects and vice versa. In addition, they used several pruning techniques including clause pruning, frequency pruning, and global pruning. However, the DP algorithm has error propagation problem, and it is good for medium size datasets only [56]. For large datasets, it will extract many incorrect aspects through the propagation process. These incorrect aspects occur because many non-opinionated adjectives will be extracted as opinion words which in turn will be used and extract many incorrect aspects [56]. The experiments were conducted over customer reviews datasets [23] with a result 86% using F-measure.

In [56], DP was improved by adding two types of rules including: (1) no-pattern rule with list of words to use with; and (2) part-whole rule for identifying more aspects which require finding the class concepts of the corpus. In addition, to find aspect relevance they ranked these aspects using Hyperlink-induced topic search (HITS) algorithm and frequency-based approach. The best achieved performance was on a mattress dataset with 77% precision and 64% recall. Another study by Hai et al. [84] proposed a bootstrapping method based on utilizing three types of dependency relations which exist between the aspect and opinion word. The process started with an aspects seed list to start the bootstrapping process. Then, all noun/noun phrase were extracted as candidate aspects, while adjectives and verbs were extracted as possible opinion words. They also combined the bootstrapping approach with two new types of association models including Likelihood Ratio Tests (LRT) and Latent Semantic Analysis (LSA) which are used to find the degree of association between aspects and opinion words. The two proposed bootstrapping models were called LRTBOOT and LSABOOT. The experiments were conducted on two Chinese datasets with 61.9% F-measure on hotel reviews and 73.25% F-measure on cellphone reviews. However, this approach cannot extract non-frequent features and the bootstrapping based approach has the problem of error propagation [56]. In a study by Kumar and Raghuveer [50], they proposed an aspect extraction method which is mainly based on using dependency relation rules. They defined 11 rules using a combination of different types of dependency relations. The experiments were conducted on customer reviews datasets [23] with the best achievements of 82% recall and 73% precision.

A study by Yan et al. [43] proposed an extraction method which is based on dependency relation rules and NodeRank algorithm. They used four dependency relation rules to extract aspects and opinion words. In addition, the extracted aspects were expanded using a synonym lexicon. The technique started by preprocessing the reviews, then parsing the reviews and extracted possible aspects and opinion words based on the four dependency relations. They then created a network based on the extracted aspect-opinion pairs. Moreover, the NodeRank algorithm which is an extension of PageRank algorithm was used to rank all extracted pairs in the network. In the constructed network, any aspect opinion pair with a NodeRank greater than the given threshold was approved, and the given aspect added to the final list of aspects. The final list of aspects was expanded by finding the synonyms of each aspect from the synonym lexicon. The experiments were carried on Chinese reviews about three different electronic products with an average F-measure of 73.96%.

Further improvements were made in the DP algorithm. For example, in a study by Liu et al. [85], they improved the DP by employing aspect associations and semantic similarity. The best results were achieved when they used the AER (Aspect Extraction based on Recommendation) method. The experiments were conducted on customer reviews datasets [23] with a result of 87% using F-measure and on electronics products datasets from [86] with a result of 83% using F-measure. After that, Kang and Zhou [52] also improved the DP algorithm with new rules. Aspects were extracted by extending DP with some new rules including comparative rules, and indirect-dependency rules. In addition, part–whole relation rules were also defined, and all these rules were applied for aspect extraction. In addition, two pruning techniques were used for pruning incorrect aspects including self-filtering by removing the feature if it is not part of any multi-word candidate aspects and mutual exclusion [39]. Also, they utilized pruning based on the frequency and semantic similarity using WordNet. However, pruning based on WordNet is not reliable as many correct aspects are not included in WordNet. The experiments were conducted on customer reviews datasets [23] where the F-measure was 87%. In another study by Wang et al. [87], they applied DP rules for aspect extraction on eight different products from Amazon website. The data collected included reviews about the same products in English and Chinese languages. However, no performance results were reported. However, these methods which applied dependency rules for aspect extraction share the following limitations: (1) not all used rules were efficient for extraction; (2) not all possible dependency extraction rules were explored; and (3) the extraction of aspects based on dependency rules has a problem when dealing with informal written reviews. In addition, most of the methods either do not apply aspects pruning or use aspect pruning and need some improvements.

Thera are also some examples of semi-supervised works which utilized syntactic patterns for aspect extraction. For example, Samha et al. [42] proposed an aspect extraction technique which is based on syntactic patterns method. In their study, frequent syntactic patterns with a total of 11 frequent patterns were used for aspect extraction. In the beginning, opinion lexicon created by Hu and Liu [23] was used to match any opinion word in the review to one of the 11 patterns, then extract the corresponding aspect. The experiment was conducted on a customer review dataset [23] with 77% F-measure. Another example which is based on a syntactic patterns method for aspect extraction is the work of Rana and Cheah [55], where they considered the noun/noun phrase as the possible correct aspects. In addition, they utilized an opinion lexicon [23] in the extraction process. The pattern rules were used for aspect extraction. Also, two pruning methods were applied based on the frequency and Normalized Google Distance (NGD). However, they used many patterns which produce many incorrect aspects. Furthermore, long used patterns may introduce many incorrect aspects. Whereas the NGD pruning used is unreliable as Google consider any two words co-occurred on the same page as semantically related words, but the words may be close together or very far from each other on the web page. The experiments were conducted on customer reviews datasets [23] with an F-measure of 89%. Sequential syntactic pattern method was also used in [88] for aspect extraction. Based on the assumption that there is an association between the aspects and opinion words, they defined ten sequential pattern rules for aspect extraction. The proposed method worked by first searching for any noun or noun phrase in the sentence, then it will match that sentence with the ten rules. If the sentence matches with any of these rules, then the given noun/noun phrase will be extracted as a correct aspect; otherwise it will be discarded. The experiments were conducted on customer reviews datasets [23] with an F-measure is 89%. However, these methods which applied pattern rules for aspect extraction share the following limitations: (1) not all used rules are efficient for extraction; (2) not all possible extraction pattern rules are explored; and (3) pattern-based rules are better than dependency-based rules for informal text, but for long sentence dependency-based rules are better. In addition, most of the methods either do not apply aspects pruning or use aspect pruning and need some improvements.

More examples of semi-supervised works which utilized alignment model for aspect extraction include the work conducted by Liu et al. [57], where a partially supervised word alignment model (PSWAM) was proposed for extracting aspects. PSWAM is based on the idea that noun/ noun phrases as the candidate aspects and adjective as the opinion words. This PSWAM model was combined with some syntactic patterns to find possible relations between candidate aspects and opinion words, then determine the degree of association in each pair. In addition, a graph algorithm was applied to find the confidence of each candidate aspect. Moreover, the aspect with confidence value greater than the threshold was approved. The experiment was conducted on a customer review dataset [23] with 86% F-measure. Furthermore, in [58], they also used an alignment word model for extracting aspects and opinion words, where the alignment model was used for identifying relations between opinion words and aspects. In addition, they considered noun/noun phrase as candidate aspects and adjectives/verbs as possible opinion words candidates. Then, they estimated the confidence of each extracted aspect or opinion word using graph based ranking algorithm. Lastly, candidate aspects or opinion words with confidence greater than the threshold were approved as correct aspects and opinion words. The experiments were conducted on three datasets with 86.6% F-measure on a customer review dataset [23], COAE 2008 Chinese datasets with 77.5% F-measure, and Large dataset which contains English and Chinese reviews with 82% F-measure. However, in these works, the common limitation that infrequent aspects were pruned.

There are also some examples of recent semi-supervised works. Work by Feng et al. [89] represents one of the most recent semi-supervised work using the topic model with synonym recognition method for aspect extraction. In their work, they combined LDA with synonym recognition to extract explicit aspects and used TF-IDF to discard nouns which do not represent an aspect. In addition, they defined aspects similarity rules. The experiment was conducted on a customer review dataset [23] with 87% F-measure. Agerri and Rigau [90] modeled aspects extraction as sequence labeling, and they combined clustering features in the developed aspects extraction system. They reported F-measure results of aspects extraction as 84.11% on SemEval 2014 dataset, 70.90% on SemEval 2015 dataset and 73.51% on SemEval 2016 dataset. However, based on paper authors, the developed model cannot be generalized to deal with extraction of unknown aspects. In addition, many incorrect aspects where resulted, because they did not perform aspects pruning. Another recent work conducted by Chauhan and Meena [91] considered only nouns and noun phrases as possible aspects. They extracted these aspects with the help of an opinion lexicon which they created based on the contextual information of the datasets used. In addition, they discard incorrect aspects based on frequency using their prepared lexicon and similarity methods based on NGD. The proposed method was called Domain-Specific aspect term extraction method (DomSent). The experiment was conducted on a customer review dataset [23] with 86% F-measure. Park et al. [92] utilized LDA topic model to extract aspects, then they developed Topic Cleaning (TC) algorithm to find the clean topics from these topics generated by LDA through using of topics split and merge. For splitting and merging of topics, they proposed post-processing technique based on clustering and word embedding. However, methods which are based on topic modeling can only extract the general aspects and ignore the fine-grained aspects. In addition, many incorrect aspects where resulted, because they did not perform aspects pruning.

2.1.3. Supervised extraction methods

In the supervised extraction methods as mentioned earlier, training data is required to extract the mentioned explicit aspects. The following works represent some examples of works conducted using the supervised method.

First, some examples of supervised works which utilized deep learning method for aspect extraction are as follows. A study by Poria et al. [26] combined a deep neural network with the word embedding model and five linguistic rules for aspect extraction. They used seven layers deep neural network for extracting the aspects. The experiments were conducted on two datasets including customer review datasets [23] with 88% F-measure and SemEval 2014 dataset with 84.74% F-measure. Li and Lam [93] proposed aspect extraction method based on using Two LSTMs. In addition, they combined these two LSTM with neural memory operations and extended memories. They reported F-measure results of aspects extraction on Laptop domains of SemEval 2014 datasets as 77.58% and 73.44% on restaurant dataset from SemEval 2016. However, based on Ma et al. [94], they failed to catch the sentence overall meaning. In addition, many incorrect aspects where resulted, because they did not perform aspects pruning.

Also, in [95] they used deep learning method for aspect extraction. They combined Convolutional Neural Network (CNN) with two types of embedding including domain specific and general embedding. Furthermore, this combined embedding with CNN was applied on two datasets about laptop and restaurants with F1-score achieved as 81.59% for laptop dataset and 74.37% for restaurant dataset. However, the approach still has problems such as being unable to extract the aspects if there is a conjunction. Ma et al. [15] proposed an LSTM model which incorporates a hierarchical attention mechanism and common-sense knowledge for aspect-based sentiment analysis. In the most recent work by Da'u et al. [96] based on deep learning method for aspect extraction, they proposed multichannel convolutional neural network (MCNN) architecture for the aspect extraction which represents an extension of CNN architecture, where the new MCNN model combined two types of channels including POS embedding and word embedding channel. The experiment was conducted on a customer review dataset [23] with 89% F-measure. However, the limitation of this study that many incorrect aspects were resulted because they did not perform aspects pruning.

Shu et al. [97] proposed a controlled CNN (Ctrl) model for aspects extraction using two control layers. They considered the problem as a sequence labeling task. They reported F-measure results of aspects extraction on Laptop domain of SemEval 2014 datasets as 82.73% and 75.64% on restaurant dataset of SemEval 2016. Barnaghi et al. [98] combined CNN with two embedding layers: a domain specific layer and another general layer. They reported F-measure results of aspects extraction on Laptop domain of SemEval 2014 datasets as 78.26% and 73.81% on restaurant dataset of SemEval 2016. Kumar et al. [99] combined Bi-LSTM with CRF for aspects extraction by considering the problem as a sequence labeling task. In addition, they introduced a set of new sequential tags for aspects extraction. They reported F-measure results of aspects extraction on Laptop domain of SemEval 2014 datasets as 71.28% and 80.03% on restaurant dataset. Akhtar et al. [100] used Bi-directional Long Short-Term Memory (Bi-LSTM) followed by self-attention mechanism for aspects extraction. They reported F-measure results of aspects extraction on Laptop domain of SemEval 2014 datasets as 78.57% and 83.36% on restaurant dataset. However, these works are unable to identify many correct aspects. Furthermore, a common limitation of these studies that many incorrect aspects were resulted because they did not perform aspects pruning.

There are also some examples of supervised works which utilized Conditional Random Fields (CRF) for aspect extraction. CRF was first applied in [101] for aspect extraction. To model the sequential dependency relation which exist between words, they used linear-chain CRFs. In addition, they also used Skip-chain CRFs for dependency over long distance, while Tree CRFs were used to exploit the syntactic structure. The experiments were conducted over two datasets about movie and product reviews with the reported performance was 83.7% F-measure on movie reviews and 80.1% F-measure on products reviews. Furthermore, CRF was also adopted in [102] for aspect extraction. CRF was applied for aspect extraction using a combination of different features to learn CRF such as token, POS, word distance, short dependency path, and opinion distance. The model was applied on different datasets about cars, cameras, movies, and web-services. The best achieved performance was on movie reviews with 70.2% F-measure. Later, CRFs was also used in [47] for aspect extraction. In the first step, they preprocessed the data by removing special symbols, misspelling words correction, and noun stemming. In the second step, they prepared the datasets by labeling it with the corresponding tags. In the third step, they trained the CRFs model using the labeled data. Finally, they applied the trained CRFs model on the testing data to label the aspects. The experiments were conducted on customer reviews datasets [23] and other data collected from Amazon about digital cameras with 86.2% F-measure. Whereas in [103], CRFs was combined with three types of language features for aspect extraction. These features were word feature, POS feature, and sentence structure feature. The experiments were conducted on Hu and Liu [23] dataset and another part was collected from Amazon with 75.6% F-measure. Xiang et al. [104] proposed Multi-Feature Embedding (MFE) clustering using CRF model for aspects extraction. They reported F-measure results of aspects extraction on SemEval 2016 datasets as 84.33% on Restaurant dataset and 76.53% on laptop dataset. In addition, 70.31% on SemEval 2015 and 73.81% on SemEval 2016. However, these works which applied CRFs for aspect extraction share the following limitations: (1) CRF not suitable for long range patterns. (2) many incorrect aspects where resulted because they did not perform aspects pruning.

Another example of supervised work which utilized dependency relations method for aspect extraction is found in the works of Liu et al. [51]. In this work they extended the dependency-based extraction rules of double propagation algorithm [39] by adding more dependency rules. Furthermore, two algorithms were compared for selecting the best rules from the set of rules by using a greedy algorithm or Simulating Annealing (SA) algorithm. Based on reported results, the best achieved results were obtained when SA was used at 87.9% F-measure on a customer review dataset [23]. However, the limitation of this study that many incorrect aspects were resulted because they did not perform aspects pruning. In [105], they employed BI-LSTM-CRF (Bidirectional Long–Short Term Memory Conditional Random Field) model with three types of dependency relation rules for the Chinese language to extract aspects. At first, BI-LSTM-CRF was used for extracting aspects and their related opinion words. After that, they checked if the aspect was related to the extracted opinion word using the dependency relation rules. The experiments were conducted on Chinese product reviews with the achieved result was 81.86% using F-measure. In [106], they utilized a seven layers deep convolutional neural network (CNN) for aspect extraction. In addition, they combined CNN with four dependency rules to improve aspect extraction performance. The experiment was conducted on one dataset form [23] on a Nikon camera dataset with 88.6% precision and 90.5% recall.

An example of a recent work is the work conducted in [107], which is based on using NPL features with NGD and ConceptNet method for aspect extraction. They considered candidate aspects as nouns and noun phrase only, then they used NGD for computing word similarity and discard incorrect aspects. Furthermore, they also refined these aspects by using ConceptNet. The experiment was conducted on a customer review dataset [23] with 86% F-measure. However, it is unable to extract multi-aspects in the sentence. In addition, it missed many correct aspects as they considered aspects as noun or noun phrase only.

2.2. Dependency-based and pattern-based works

Many studies conducted in literature used dependency relations rules for extracting aspect and its related opinion words such as the works mentioned previously in Sections 2.1.1, 2.1.2, and 2.1.3. In these works, Stanford dependency relation which represents a binary grammatical relationship between words in the sentence was used for aspect extraction. In addition, patterns-based extraction was used in many studies as the works mentioned in Sections 2.1.1, 2.1.2, and 2.1.3. These works were based on the idea that a pattern contains sequence of tags. These tags described the aspect, opinion word, and relation between these tags. Therefore, the following Sections 2.2.1 and 2.2.2 provide a complete presentation of all works which have utilized dependency relations and pattern rules for aspect extraction. This represents a deep survey over these types of works which give a clear overview of these works.

2.2.1. Dependency-based work

The previous works which utilized dependency rules for aspect extraction are presented in Table 1, which gives more insight into these works as a summary of the proposed extraction technique, the language and datasets used, and the reported results.

2.2.2. Pattern-based works

Previous works which utilized pattern rules for aspect extraction are presented in Table 2, which gives more insight into these works as a summary of the proposed extraction technique, the language and the datasets used, and the reported results.

It has been found that most of the methods, whether supervised, semi-supervised, or unsupervised, have commonly used dependency-based, or pattern-based method for aspect extraction. However, there are some limitations in these two methods as discussed above. For example, in methods which applied dependency rules for aspect extraction, they share the following limitations: (1) not all used rules were efficient for extraction; (2) not all possible dependency extraction rules were explored; and (3) the extraction of aspects based on dependency rules has a problem when dealing with informal written reviews. Dependency rules give better results if the reviews follow English grammar rules. However, reviews available online are a mixed combination of formal and informal text.

Methods which apply pattern rules for aspect extraction share the following limitations: (1) not all used rules are efficient for

Table 1
Dependency-based works.

Work	Technique	Language	Datasets	Results
Zhuang et al. [60]	Word frequencies and four dependency relation rules	English	movie reviews	52.9% using F-measure
Wu et al. [37]	Dependency parsing method at phrase level + SVM	English	Customer reviews datasets [23,82] datasets	57% using F-measure
Qiu et al. [39,83]	Double Propagation (DP) using eight dependency rules + (clause, frequency, and global pruning)	English	Customer reviews datasets [23]	86% using F-measure
Zhang et al. [56]	DP improved with no-pattern rule and part-whole rule + HITS algorithm	English	Mattress dataset	77% precision and 64% recall
Hai et al. [84]	Bootstrapping (LRTBOOT and LSABOOT)	Chinese	Hotel and cellphone reviews	61.9% F-measure on hotel reviews and 73.25% on cellphone reviews
Kumar and Raghuveer [50]	Defined eleven dependency rules	English	Customer reviews datasets [23]	82% recall and 73% precision
Hai et al. [35]	Domain specific corpus and independent domain corpus + three dependency relation rules + two measures (EDR and IDR)	Chinese	Hotels	52.26% F-measure
			Cellphones	63.6% F-measure
Li et al. [64]	Bootstrapping using six dependency rules + two measures (Prevalence and Reliability)	English	Customer review datasets [23]	89% using F-measure
Yan et al. [43]	Four dependency relation rules + NodeRank algorithm + synonym lexicon	Chinese	Three different electronic products	73.96% using F-measure
Samha [65]	Five new dependency relation rules with eleven dependency rules used from previous studies	English	Customer review datasets [23]	71.8% using F-measure
Liu et al. [51]	Extended DP by more rules + (greedy algorithm or Simulating Annealing (SA))	English	Customer reviews datasets [23]	87.9% using F-measure
Liu et al. [85]	DP improved using aspect associations and semantic similarity	English	Customer reviews datasets [23]	87% using F-measure
			Electronics products datasets from [86]	83% using F-measure
Kang and Zhou [52]	DP algorithm improved with new rules (comparative rules, indirect-dependency rules, and part–whole relation rules) + Pruning based on self-filtering, mutual exclusion [39], and frequency and semantic similarity	English	Customer reviews datasets [23]	87% using F-measure
Xiong et al. [105]	BI-LSTM-CRF + three dependency relation rules	Chinese	Product reviews	81.86% using F-measure
Wang et al. [87]	DP	EnglishChinese	Eight different products	N/A
Dragoni et al. [63]	Dependency graph + three dependency relation rules	English	SemEval 2015	60% restaurant 51% laptop using F-measure
			SemEval 2016	67% restaurant 57% laptop using F-measure
Ray and Chakrabarti [106]	CNN + four dependency rules	English	One dataset form [23] on Nikon camera dataset	88.6% precision and 90.5% recall

extraction; (2) not all possible extraction pattern rules are explored; and (3) pattern-based rules are better than dependency-based rules for informal text, but for long sentence dependency-based rules are better. In addition, most of the methods either do not apply aspects pruning or use aspect pruning and need some improvements. Further work is required to solve these limitations.

This study proposes an aspect extraction algorithm that differs from these works in the following aspects: (1) the current study proposes to combine both dependency relation and syntactic pattern rules to overcome the problems of both methods and take their advantages; (2) it overcomes the limitations of previous studies by developing a number of new aspect extraction rules to overcome the problem of unexplored rules and to overcome the weakness of existing rules; (3) the proposed algorithm can be applied on reviews which contain both formal and informal texts; (4) the proposed IWOA algorithm can be provided with any number of aspects rules regardless of the rules quality, and is able select the optimal subset of rules from these rules while it discards the irrelevant and low quality rules; and (5) the proposed aspect pruning algorithm can solve the problem of discarding low frequent aspects which was pruned out by previous studies, because it has three phases based on frequency, product manual, and direct opinion association.

Table 2
Pattern-based works.

Work	Technique	Language	Datasets	Results
Li et al. [68]	Four pattern rules + frequency	Chinese	Mobile reviews	74.04% using F-measure
Moghaddam and Ester [33]	Pattern rules + frequency + known product aspects	English	Customer review dataset [23]	80% precision and 87% recall
Bagheri et al. [69]	bootstrapping + four patterns + two pruning (subset-support pruning + superset-support pruning)	English	Customer review dataset [23]	72.9% using F-measure
Htay and Lynn [34]	Defined eight pattern rules	English	Customer review dataset [23]	79% using F-measure
Samha et al. [42]	Defined eleven patterns + opinion lexicon [23]	English	Customer review dataset [23]	77% using F-measure.
Maharani et al. [53]	Defined new patterns and used patterns from [34,70]	English	Customer reviews datasets from [23,71]	67.2% using F-measure
Asghar et al. [54]	Defined ten patterns rules	English	Customer review dataset [23]	77.16% using F-measure
Rana and Cheah [55]	Pattern rules + two pruning methods (NGD + frequency)	English	Customer review dataset [23]	89% using F-measure.
Rana and Cheah [88]	Defined ten sequential pattern rules	English	Customer review dataset [23]	89% using F-measure.

3. The workflow of the proposed aspect extraction algorithm

Fig. 1 shows the main workflow of the proposed aspect extraction algorithm. The general process of aspect extraction consists of a number of phases. In the first phase, data preprocessing is carried out on the customer reviews datasets to prepare the dataset for aspect extraction. The reviews datasets are split into sentences. Then, for each sentence, the dependency relations in that sentence will be extracted using Stanford dependency parser. In addition, for each sentence, the sequence of POS tags will be determined by the Stanford POS tagger. The set of POS tags and dependency relations of each sentence will be used in the next phases. In the second phase, there are a combination of 126 rules, including dependency and pattern rules from the previous studies [34,39,42,50,53–55,60,65,103,108] together with the newly developed rules. In the third phase, these rules will be used by IWOA to select optimal combination rules subset from the full set of rules. The selected rules will be applied on the testing dataset for extracting aspects. Finally, the extracted aspects will be passed to the pruning phase where the correct aspects will be saved as final approved aspects, while the incorrect aspects will be discarded.

4. Aspect extraction rules

The main task of this study is aspect extraction, and to achieve this task, many rules have been used. These rules include rules from previous studies [34,39,42,50,53–55,65,70,83,108] in addition to new rules which were created in this study to improve aspect extraction accuracy. These new rules were created based on analysis and observation of aspects patterns in the datasets. The full set of rules used in this study contains 126 rules. While some of these rules can extract the same aspects, but one of these common rules is better than other in terms of precision as it contains more restrictions. Also, many of these rules may extract incorrect aspects. Therefore, a proper selection of included rules is required. This research uses three types of rules: dependency-based rules, sequential pattern-based rules, and new created rules. One reason for using three types of rules is that some rules are better for structured text, while other rules are better for unstructured text. Another reason is that in some sentences there is no pattern or dependency rule from previous studies that can match with these sentences to extract the aspect contained in these sentences. Moreover, there are cases in which two or more rules can extract the same aspects, but one rule from these rules is better in terms of precision.

The full set of rules which are used in this study are presented in Tables 3–5, where the *'nsubj', 'amod', 'prep', 'csubj', 'xsubj', 'dobj', 'iobj', 'conj', 'xcomp', 'compound', 'nmod', 'neg', 'det', 'nsubjpass', 'nummod', 'appos',* and *'cop'* are dependency relations from Stanford dependency parser which describe a relation between two words. Furthermore, OP means any opinion word from opinion lexicon [23], NN (any type of noun), VB (any type of verb), NNS (Plural noun), JJ (any type of adjective), RB (any type of Adverb), VBP (non3rd person singular present), VBZ (3rd person singular present), VBD (past tense verb), IN (Preposition or subordinating conjunction), DT (Determiner), TO (to), PRP (Personal pronoun), CC (Coordinating conjunction), and VBN (past participle verb). These represent the tags of Stanford part-of-speech (POS) tagger.

4.1. Dependency-based extraction rules

The dependency-based extraction rules used from previous studies are presented in Table 3. In Table 3, each rule is explained with an example from [23]'s datasets.

4.2. Sequential pattern-based extraction rules

The pattern-based extraction rules which are used from previous studies are presented in Table 4. Each rule is explained with an example from [23]'s datasets.

The rules presented in Tables 3 and 4 are selected based on a number of factors such as the achieved results in their studies, number of experiments were conducted in this work to check these included rules, and manual observation and comprehensive analysis of the datasets to check which rules can be included in the final set of 126 rules.

In addition, as an improvement to many of these used rules, an OP/JJ combination was used in these rules to cover the majority of opinion words. This combination will cover all possible types of opinion words. The reason for OP/JJ combination because the previous studies consider the opinion word as either adjective or the words which are contained in the opinion lexicon only, but not both. For example, in the sentence *"main dial is not* ***backlit****"*, backlit represents an adjective which represents the opinion words in the sentence, but the word ***backlit*** is not found in opinion lexicon. So, the previous studies, which were based on opinion lexicon only for defining an opinion word, cannot extract the aspect ***dial***. Another example, in the sentence *"i would* ***recommend*** *this* ***product*** *to anyone."* the aspect "***product***" cannot be extracted by previous studies, because they considered an adjective as the only correct opinion word, and the opinion word ***recommend*** is a verb and found only in the opinion lexicon. Therefore, this problem was handled by rule 71 using the OP/JJ combination.

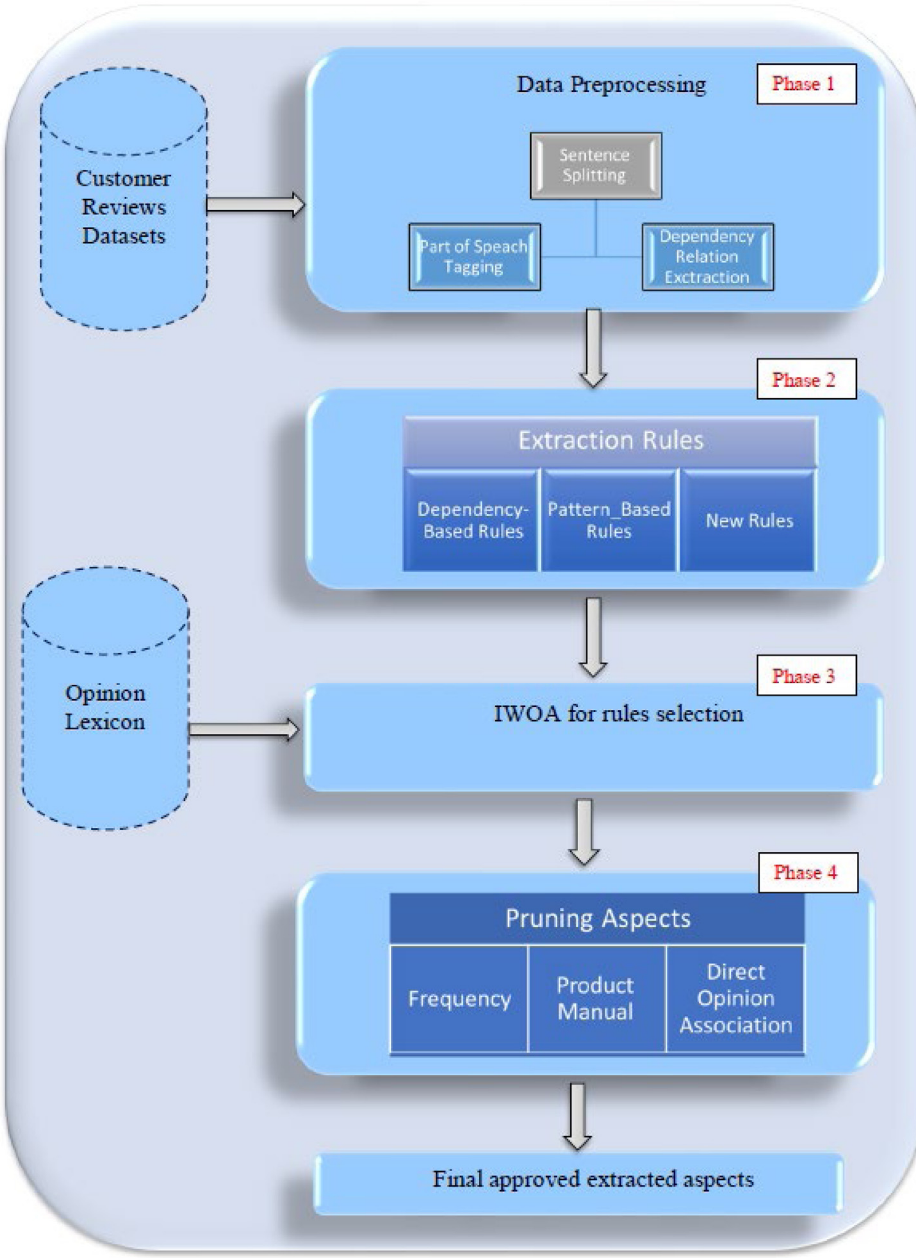


Fig. 1. The workflow of the proposed aspect extraction algorithm.

4.3. Formulation of new aspect extraction rules

To complement the aspect extraction rules used in previous studies, a comprehensive analysis and observation of customer reviews datasets [23] was conducted to come up with these new rules, built based on experiments and manual observation of Hu and Liu [23]'s datasets, and by finding frequent occurring rules which have not been explored by the previous studies. In addition, some of the new rules were developed to extract aspects which cannot be extracted by the previous rules. Furthermore, some new rules can extract the same aspects which can be extracted by the rules used in previous studies. However, the new rules are characterized by minimizing the number of incorrect extracted aspects which can be obtained by the previous rules. This is achieved by adding more restrictions on the newly developed rules to minimize the extraction of incorrect aspects. Table 5 presents these new rules. Each rule is explained with an example from [23]'s datasets.

To give more details about how these rules work, some rules are explained. For example in Table 3, **Rule #7** (amod(NN,OP)): if 'amod' relation found in the sentence with the first argument is NN and the second argument is OP, then the first argument will be extracted as an aspect. The following example "the poor manual". There is "amod(manual-3, poor-2)" relation between "***poor***" opinion word and "***manual***" aspect, then "***manual***" will be extracted as an aspect [60]. Another example in Table 4, **Rule #19 (NN** VBZ JJ/OP): For example, the sentence ***"audio*** *is excellent*." matches with this rule as the tagged sentence is "audio/NN is/VBZ excellent/JJ", where NN "***audio***" represents the extracted aspect and "***excellent***" represents an opinion word from OP [42]. An example from Table 5, **Rule #43** (**NN NN** * OP where * means any pattern found but no NN exist in the pattern in between): For example, the sentence *"their **customer service** is very*

Table 3
Dependency-based extraction rules.

Rule#	Rule	Example [23]	References
1	nsubj(JJ/OP,**NN**)	***video*** *was poor*	Zhuang et al. [60]
2	ReL1(H1,**NN1**) and ReL2(H1,**NN2**) such that ReL1 and ReL2 any dependency relation from [‘nsubj’, ‘amod’, ‘prep’, ‘csubj’, ‘xsubj’, ‘dobj’, ‘iobj’]	*proven canon built* ***quality*** *and* ***lens***	Qiu et al. [39]
3	nsubj(VB1,H1) and dobj(VB1,**NN**)	*honestly, i love this* ***player***.	Zhuang et al. [60]
4	nsubj(H1,**NN**) and xcomp(H1,JJ/OP)	*it 's* ***size*** *also makes it ideal for travel*	Samha [65]
5	amod(NN1,OP/JJ) and conj(**NN1**,**NN2**)	*it plays original* ***dvds*** *and* ***cds***	Kumar and Raghuveer [50]
6	nmod(OP/JJ,**NNS**)	*i find the lack of entertaining* ***games*** *on this phone quite disturbing.*	Samha [65]
7	amod(**NN**,OP)	*the poor* ***manual***.	Zhuang et al. [60]
8	ReL1(H1,**NN**) and ReL2(H1, OP/JJ) such that ReL1 and ReL2 any dependency relation from [‘nsubj’, ‘amod’, ‘prep’, ‘csubj’, ‘xsubj’, ‘dobj’, ‘iobj’]	*this* ***camera*** *has a major design flaw.*	Qiu et al. [39]
9	nsubj(**NN**,OP/JJ)	*my only gripe about the hardware is the* ***buttons***	Zhuang et al. [60]
10	dobj(OP/JJ,**NN**)	*i especially like the more commonly used* ***buttons***.	Qiu et al. [39]
11	nsubj(OP/JJ,NN1) and compound(**NN1**,**NN2**)	*i find* ***onscreen displays*** *annoying.*	Kumar and Raghuveer [50]
12	conj(**NN1**,**NN2**)	*proven canon built* ***quality*** *and* ***lens***	Qiu et al. [39]
13	amod(NN1,OP/JJ) and compound(**NN1**,**NN2**)	*overall, the g3 delivers what must be considered the best* ***image quality***	Kumar and Raghuveer [50]
14	neg(OP/JJ, H1) and nsubj(OP/JJ,**NN**)	*the* ***colors*** *on the screen are not as crisp as i 'd have liked them to be.*	Kumar and Raghuveer [50]
15	ReL(**NN**, OP/JJ) such that ReL any dependency relation from [‘nsubj’, ‘amod’, ‘prep’, ‘csubj’, ‘xsubj’, ‘dobj’, ‘iobj’]	*definetely a great* ***camera***	Qiu et al. [39]
16	ReL(OP/JJ, **NN**) such that ReL any dependency relation from [‘nsubj’, ‘amod’, ‘prep’, ‘csubj’, ‘xsubj’, ‘dobj’, ‘iobj’]	*the* ***manual*** *is relatively clear*	Qiu et al. [39]
17	nsubj(OP1/JJ1,**NN**) and cop (OP1/JJ1,H1)	*the* ***menus*** *are easy to navigate.*	Huang et al. [103]

Table 4
Sequential pattern-based extraction rules.

Rule#	Rule	Example [23]	References
18	**NNS** VBP OP /JJ	***controls*** *are poorly designed*	Samha et al. [42]
19	**NN** VBZ JJ/OP	***audio*** *is excellent.*	Samha et al. [42]
20	JJ/OP **NN NN**	*it 's a very nice* ***dvd player***.	Htay and Lynn [34]
21	RB JJ/OP **NN**	*it is a very amazing* ***product***.	Htay and Lynn [34]
22	OP **NN**	*poor* ***reliability***.	Htay and Lynn [34]
23	**NN** OP	*creative* ***software*** *stinks.*	Maharani et al. [53]
24	**NN** IN NN	***audio*** *on video also lacking.*	Maharani et al. [53]
25	**NN** IN DT NN	*the* ***construction*** *of the player is the cheesiest i have ever seen.*	Maharani et al. [53]
26	NN IN DT **NN**	*overall, a good buy for the* ***price***.	Maharani et al. [53]
27	OP to **VB**	*it has refused to* ***read*** *second discs.*	Asghar et al. [54]
28	JJ to **VB** such that JJ not in OP	*it is exceedingly simple to* ***navigate***	Asghar et al. [54]
29	JJ1 **JJ2 NN**such that JJ2 not in OP	*the g3 has much sharper* ***white offsets***.	Rana and Cheah [55]
30	**NN** VBZ/VBP DT OP/JJ NN	*the* ***manual*** *does a fine job.*	Rana and Cheah [108]
31	**NN** * PRP/DT OP* any pattern found but no NN	***apex*** *is the best cheap quality brand for dvd players.*	Rana and Cheah [108]
32	VB OP/JJ **NN**	*get great* ***reception***.	Rana and Cheah [108]
33	NN VB OP/**JJ**	***audio*** *is excellent*	Rana and Cheah [108]

poor.” matches with this rule. The tagged sentence is “their/PRP$ customer/NN service/NN is/VBZ very/RB poor/JJ”, where NN NN “***customer service***” represents the extracted aspect *and* “***poor***” represents the opinion word in OP.

These new rules were developed to extract aspects that have not been extracted by rules that were used in previous studies or to minimize the incorrect extracted aspects. For example, “i am really **impressed** by this **dvd player**” In this example, the extracted aspect is ***dvd player***, which can be extracted by the new developed rule number 62, and ***impressed*** is not an adjective and found in opinion lexicon only. In case of considering opinion lexicon or adjective only as an opinion word as in previous studies, the extraction algorithm will miss many correct aspects and cannot extract these aspects. Also, new rules were also developed to improve the low precision problem of previous studies' rules. There are new rules which can extract the same aspects which were extracted by previous rules, but with better precision based on the added restrictions to these new rules. In another example, in rule number 66 in the sentence *“the* ***vibration*** *is not* ***top***”, the previous studies extract ***vibration*** as the correct aspect but when they extracted it, they consider ***top*** as the opinion word. However, in this example before ***top*** opinion word, there is “***not***” word, this restriction was added to rule number 66 to extract accurate and reliable results. Also, in rule number 88 in the sentence *“my favorite being the* ***games*** *and the* ***pim****, and the* ***radio***.”, in this case by using regular expression in rule 88 it can extract all these aspects “***games***”, “***pim***”, and “***radio***”, which were not extracted by previous studies. More examples of using restrictions in the newly developed rules includes rule number 89 (DT NN VB * JJ/OP **NN** where * means any pattern found but no NN exist in the pattern in between). For example, in the sentence *“the g3 is loaded with many useful* ***features***.”, some of the previous studies extracted ***g3*** as the correct aspect, but the opinionated target aspect in this sentence is ***features***. Thus, rule 89 solve this problem and extract ***features*** as the correct aspect. These

Table 5
New extraction rules.

Rule#	Rule	Example [23]
34	(Any POS but not NN) **NN** VBZ RB OP/JJ	*the **manual** is relatively clear.*
35	amod(**NN**, Any POS) such that OP found on left or right with any pattern in between but no NN	*in addition it comes with a sleek and powerful **headset**.*
36	nmod(**NN**, Any POS) such that OP found on left or right with any pattern in between but no NN	*overall it is the best **camera** on the market.*
37	nummod(**NN**, Any POS) such that OP found on left or right with any pattern in between but no NN	*usb 2.0 **transfer** is insanely fast.*
38	appos(**NN**, Any POS) such that OP found on left or right with any pattern in between but no NN	*i love my new **nomad**, its great !.*
39	det(**NN**, Any POS) such that OP found on left or right with any pattern in between but no NN	*the **price** was right.*
40	nsubjpass(Any POS, **NN**) such that OP found on left or right with any pattern in between but no NN	*the **interfac**e used could be better designed.*
41	nmod(Any POS, **NN**) such that OP found on left or right with any pattern in between but no NN	*everything else about the **camera** is great.*
42	cop(**NN**, Any POS) such that OP found on left or right with any pattern but no NN	*this is a wonderful **camera**.*
43	**NN NN** * OP * any pattern found but no NN	*their **customer service** is very poor.*
44	**NN** RB (RB not in OP) * OP * any pattern found but no NN	***interface** practically seamless*
45	**NN** VBD (VBD not in OP) * OP * any pattern found but no NN	*the included **earbuds** were uncomfortable*
46	(Any POS but not NN) OP **NN**	*very comfortable **camera***
47	NN of **NN**	*it does play a wide range of **formats**.*
48	DT **(Any POS) NN** VBZ JJ/OP	*the **sound quality** is okay*
49	DT **NN** IN * JJ/OP * any pattern found but no NN	*you can see the **interface** as modern or classical.*
50	DT **NN NN** * JJ/OP * any pattern found but no NN	*the **software interface** supplied was very easy to use*
51	**NN** NN VBZ/VBP RB RB	*the **mms** technology is very well integrated with this phone, which you will enjoy.*
52	**NN NN** VB OP/JJ	***replacement battery** is expensive.*
53	**NN** (any POS) RB **JJ/OP**	*the **controls** are very intuitive*
54	OP (any POS) OP/JJ **NN**	*like the smart **volume***
55	DT **NN NN** VB RB	*the **lens cover** is surely loose*
56	DT **NN NN** VB OP/JJ	*the **storage capacity** is great for me –*
57	DT **NN** VB RB * OP [such that RB not in OP] * any pattern found but no NN	*the **headphones** are n't the best*
58	DT **NN** VB JJ/OP	*the **grain** was terrible*
59	OP * **NNS** * any pattern found but no NN	*lack of good **accessories**.*
60	DT **NN** * OP * any pattern found but no NN	*this **camera** is closest to perfect*
61	RB JJ/OP * **NN** * any pattern found but no NN	*very flexible and powerful **features***
62	RB OP/JJ * **NN NN** * any pattern found but no NN	*i am really impressed by this **dvd player**.*
63	OP/JJ IN DT **NN**	*either way, can't go wrong with this **price***
64	OP * **NN NN** * any pattern found but no NN	*i knew this before hand, and it is not that bad there is no **tiff format**.*
65	DT JJ/OP **NN**	*overall it is a great **unit**.*
66	DT **NN** VBZ/VBP RB JJ/OP	*the **vibration** is not top.*
67	**NN NN** VBZ JJ/OP	***color screen** is good*
68	DT RB JJ/OP **NN**	*it 's a very intuitive **program***
69	DT VBZ/VBP DT JJ/OP **NN**	*this is a good **deal**.*
70	RB VB JJ/OP **NN**	*does not provide enough **volume***
71	OP/JJ DT **NN**	*i would recommend this **product** to anyone.*
72	DT JJ/OP JJ/OP **NN**	*the nomad zen could use a little sturdier **construction***
73	**NN** RB OP/JJ	***interface** practically seamless*
74	VBZ/VBP RB OP/JJ **NN**	*the phone is very light **weight**.*
75	NN VB JJ/OP **NN**	*the zen has minimal **stoppage** between tracks*
76	OP/JJ TO VB (Any POS) **NN**	*very convenient to scroll in **menu***
77	JJ/OP JJ/OP **NN**	*polyphonic sweet **tunes***
78	JJ/OP **NN** RB	*bad **picture** yet.*
79	DT **NN** VBZ/VBP VBN	*the **display** is hinged*

(*continued on next page*)

represent some examples which show the importance of the newly developed rules.

As outlined above, the rules used in previous studies were included in the full set of 126 rules based on a number of factors, at first based on the achieved results in their study. The next step was making a number of experiments to check the performance of each included rule based on the customer reviews datasets used [23].

From the above discussion, the novelty of the proposed new extraction rules is clearly noticeable. In addition, the combination

Table 5 (*continued*).

Rule#	Rule	Example [23]
80	**NN** CC **NN** VBP JJ/OP	***option**s and **control**s are easy.*
81	PRP VBZ JJ/OP **NN**	*it takes great **pictures***
82	RB OP/JJ DT **NN**	*just received this camera two days ago and already love the **features** it has.*
83	DT **NN** **NN** VBZ/VBP JJ/OP	*the **menu options** are uncreative*
84	(Any POS but not NN) **NN** **NN** VBZ JJ/OP	*the little **jog dial** seems weak and quirky.*
85	DT **NN** VBZ/VBP JJ/OP	*kind of bulky and the **wheel** is awkward, but i can deal with that.*
86	**NN** **NN** VB (Any POS) JJ/OP	*the **picture quality** are so great.*
87	DT **NN** VBZ/VBP DT JJ/OP	*this **phone** is a winner.*
88	PRP JJ/OP * DT **NN** * any pattern	*my favorite being the **games** and the **pim**, and the **radio**.*
89	DT NN VB * JJ/OP **NN** * any pattern found but no NN	*the g3 is loaded with many useful **features**.*
90	DT NN * DT NNS * JJ/OP * any pattern	*the **player** itself has all sorts of problems.*
91	JJ/OP NN * NN NN * any pattern found but no NN	*my favorite features, although there are many, are the **speaker phone**, the radio and the infrared.*
92	DT **NN** RB VB	*the **calls** constantly drop in my area and i experince mega-static*
93	**NN** VB RB OP/JJ TO VB	*the **menus** are very easy to navigate.*
94	DT RB JJ/OP **NN** (Any POS but not NN)	*but at least youre starting with the most photorealistic **images** ive ever seen from a camera.*
95	JJ/OP IN DT (Any POS) **NN**	*i am bored with the silver **look**.*
96	nsubj(NN1,NN2) and amod(NN1,JJ/OP)	***t-mobile** was a pretty good server.*
97	PRP VB (Any POS) **NN**	*it takes wonderful **pictures** very easily in auto mode.*
98	JJ/OP NN (any pos) DT **NN**	*the only drawback is the **viewfinde**r is slightly blocked by the lens.*
99	DT **NN** * RB RB OP/JJ * any pattern found but no NN	*when talking the **voice** is not very clear*
100	DT **NN** **NN** * DT NN * JJ/OP * any pattern	*the **battery life** on this phone is surreal.*
101	PRP VB DT JJ/OP **NN**	*it is a perfect **phone**.*
102	OP (Not OP) **NN** **NN**	*excellent polyphonic **ringing tones.***
103	**NN** **NN** * JJ/OP NN * any pattern found but no NN	***sunset feature** takes incredible pics in the morning, and the evening !.*
104	OP/JJ (Any POS) NN CC NN	*the system is terrific in **size** and **design***
105	RB JJ/OP (Any POS) **NN**	*that is solved with the very comfortable handsfree **ear-piece** which is included.*
106	**NN** **NN** (such that no JJ/OP found in sentence)	*the zen does not have a **stop button** !*
107	DT JJ/OP **NN** **NN**	*the 2nd dvd player had a faulty **power supply** which caused to occasionally not turn on.*
108	DT (Any POS) **NN** * OP * any pattern found but no NN	*the included **earbuds** work quite well.*
109	compound(**NN1**,**NN2**) such that OP, no, not found on left or OP on right with any pattern in between but no NN	*well **flash photos** are never great.* *the zen does not have a **stop button** !*
110	**JJ** **NN** * OP (such that JJ not in OP and * any pattern found but no NN)	***manual functionality** is excellent.*
111	RB JJ/OP IN DT **NN**	*i am extremely pleased with this **camera**.*
112	OP * **JJ** **NN** (such that JJ not in OP and * any pattern found but no NN	*the radio feature has superb **sound quality**.*
113	DT **NN** * DT NN * OP/JJ * any pattern	*the **sound** of the player is pretty good.*
114	JJ/VBN/OP * DT **NN** * any pattern found but no NN	*i would not be inclined to purchase an **apex** product again.*
115	JJ/OP NN (Any POS) **NN**	*this phone has a very cool and useful feature the **speakerphone***
116	RB VB * **NN** [* mean any pattern that contain OP/JJ and no NN in between]	*i will never buy a creative **product** again.*
117	DT **NN** **NN** VBZ JJ/OP	*the **auto mode** is good.*
118	nsubj(OP/JJ, NN1) and conj(**NN1**, **NN2**)	*the **options** and **controls** are easy to use and logically laid out.*
119	xcomp(OP,**VB**)	*easy to **use***
120	amod(NN1,JJ/OP) and amod(**NN1**,**JJ**) [such that JJ not in OP]	*simply, the canon g3 is the best **digital camera***
121	nmod(JJ1/OP1,NN1) and nsubj(JJ1/OP1,**NN2**)	***door** broke after a month.*
122	nmod(OP/JJ,NN1) and compound(**NN1**,**NN2**)	*i 've been pleased with the **picture quality.***
123	dobj(H1,**NN1**) and amod(**NN1**,JJ/OP)	*you can also assign special **rings** to special people when they call.*
124	nsubj(JJ/OP,**NN1**) and det(**NN1**,H1)	*the **prints** are beautiful.*
125	nmod(JJ/OP,NN1) and conj(**NN1**,**NN2**)	*i am very pleased with its **quality** and **durability**.*
126	nsubj(JJ/OP,**NN1**) and amod(**NN1**,H1)	*main **dial** is not backlit*

of different rules types is able to cover any formal or informal review type. Furthermore, as shown from the presented examples above, the new rules can extract aspects which could not be extracted by previous rules. Also, the new rules can improve the precision based on the added restriction on these rules.

As discussed above and presented in Tables 3–5, many rules are available for aspect extraction. However, a subset of these rules is not good in aspect extraction. Thus, a proper selection of the included rules is required to select the efficient subset of rules from the full set of rules and to discard the irrelevant ones. To select optimal rules combination, IWOA will be used. The details of IWOA algorithm are presented in Section 5.

5. Improved whale optimization algorithm for rules selection

5.1. Standard whale optimization algorithm

One of the recently developed optimization algorithms which has proven its ability in solving a number of applications is the Whale Optimization Algorithm (WOA). WOA mimics the hunting mechanism used by humpback whales to catch their prey [109]. Based on a number of studies, WOA has been applied to different application areas and given promising results in as varied fields as energy applications [110–112], solving computer networks problems [113–115]; feature selection problem [116–124]; image processing applications [125–128]; to find the optimal weights of the neural network [129]; spammer identification problem [130]; clustering applications [131]; neural network training [132]; classification applications; [133,134]; for improving power system [135]; for solving constrained economic load dispatch problems [136]; requirements prioritization problem [137]; finding the minimal cost of network [138]; finding concrete columns' compressive strength [139]; design problems [140]; feature extraction in modulation signal [141]; parameter optimization [142]; antenna design [143]; speed prediction of wind [144]; link prediction [145]; and for finding the optimal planning for robot path [146].

Based on these previous works, WOA was successfully applied in several fields and proved its suitability and outperformance in solving these problems. Moreover, WOA has a number of features which include the following: WOA is efficient and simple [109], in terms of memory WOA more efficient than PSO, the solutions in WOA are updated either based on randomly selected solution or based on the current best solution, WOA has better exploration ability in comparison to PSO [122], WOA requires fewer parameters to adjust [147,148], and the WOA algorithm is simple to implement and has fast convergence [149]. These features of WOA and its superiority in comparison with other state-of-the-art optimization algorithms motivated this study to apply WOA to rules selection problem.

The standard WOA shows competitive results in comparison to other optimization algorithms as shown from previous works. However, WOA as other optimization algorithms may stick in local optima and it has a solutions diversity problem. Therefore, to improve and adapt WOA for rules selection problem, two major improvements were included into the original WOA. The first improvement is by using Cauchy mutation (CM) to improve WOA solutions diversity (the rules selection), and make a balance between WOA exploration and exploitation. The second improvement is to improve WOA exploitation by developing and employing a local search algorithm (LSA). Thus, WOA will use LSA to check if there is a better solution than the current best solution. LSA select and deselect extraction rules in the current best solution based on precision and recall values of these rules. The idea of WOA was developed by Mirjalili and Lewis [109], who formulated the humpback whale hunting behavior mathematically using a number of equations. The humpback uses a bubble-net hunting technique to encircle and catch their prey. This bubble-net hunting technique represents the intelligent part of WOA. In addition, Mirjalili and Lewis [109] formulates how whales in nature communicate and live. In nature, the whales hunt and catch groups of small fish which are swimming near the water surface. At first, whales begin diving about 12 meters deeper than the small fishes. After that, whales start creating and sending huge numbers of '9' or circles shaped bubbles. The whales will encircle the group of small fishes inside these bubbles. In the next phase, all whales in the group start swimming towards the surface and hunt these small fishes [109]. The WOA algorithm can be represented by three phases as follows:

(1) Encircling prey phase: At this phase, the search-agents (whales) will specify the possible locations of the prey. After that, the whales will encircle the prey. In the next step, WOA will find the fitness value of each search agent and will assign the search agent with the best fitness value to the best solution value $\vec{X}^*$. Next, the remaining whales will move and update their positions with reference to the initial best solution $\vec{X}^*$. The main equations for WOA search process are represented by Eqs. (1) and (2) [109]:

$$\vec{D} = \left|\vec{C} \cdot \vec{X}^* (t) - \vec{X}(t)\right| \tag{1}$$

$$\vec{X} (t + 1) = \vec{X}^* (t) - \vec{A} \cdot \vec{D} \tag{2}$$

where $\vec{A}$ and $\vec{C}$ variables are coefficient vectors, t represents the current WOA iteration, $\vec{X}^*$ represents the best solution so far, and $\vec{X}$ represents the search agent position. Eqs. (3) and (4) are used for calculating $\vec{A}$ and $\vec{C}$ values [109]:

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r} \tag{4}$$

where the value of $\vec{a}$ in Eq. (3) will decrease linearly from 2 to 0 over WOA iterations, while $\vec{r}$ is a vector with random values over [0,1].

(2) Bubble-net attacking phase (Exploitation phase): This phase is composed of two mechanisms, including shrinking encircling and updating position using upward spiral mechanism. Where the Shrinking encircling mechanism is achieved using Eq. (3). And the upward spiral mechanism is achieved using Eq. (5):

$$\vec{X} (t + 1) = \overrightarrow{D'} \cdot e^{bl} \cdot cos\,(2\pi l) + \overrightarrow{X^*}(t) \tag{5}$$

where $\overrightarrow{D'} = \left|\vec{X}^* (t) - \vec{X}(t)\right|$ is the distance between the whale position X and the prey. Furthermore, b is a constant value which defines the shape of upward spiral movement and l is a random number value over [−1,1].

In WOA, Mirjalili and Lewis [109] it is assumed that the whales have a 50% probability to switch between shrinking encircling and upward spiral mechanism when they are moving around the prey. Therefore, these two mechanisms are modeled using the following Eq. (6) [109]:

$$\vec{X} (t + 1) = \begin{cases} \vec{X}^* (t) - \vec{A} \cdot \vec{D} & if\ p < 0.5 \\ & \text{(Encircling)} \\ \overrightarrow{D'} \cdot e^{bl} \cdot cos\,(2\pi l) + \overrightarrow{X^*}\,(t) & if\ p \geq 0.5 \\ & \text{(upward spiral)} \end{cases} \tag{6}$$

where p value is a random number over [0,1], which will be used by WOA to select between two mechanisms (Encircling, upward spiral).

(3) Search for a prey phase (Exploration phase): In this phase, the value of $\vec{A}$ vector will specify whether to update each whale position with reference to the current best solution or with reference to a randomly selected solution.

Therefore, Eqs. (7) and (8) are used to update the search agent position based on $\vec{A}$ vector value [109],

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{X_{rand}} - \vec{X}\right| \tag{7}$$

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A} \cdot \vec{D} \tag{8}$$

where X_{rand} is the randomly selected solution. Fig. 2 shows the WOA algorithm [109].

As shown from Fig. 3, which shows the workflow of WOA and how WOA is used and utilized for rules selection. At first WOA as with other optimization algorithms it starts by creating number randomly generated solutions (search agents). Each solution has a dimension (***Dim***) which will be specified based on the problem to be solved. In this study, ***Dim*** of each solution is equal to 126, which is dependent on the number of rules. Therefore, WOA will create, as shown in Figs. 2 and 3 *n* solutions. In this case, each solution contains a subset of randomly selected rules from the full set of 126 rules. Each solution is represented as one dimensional vector of size (***Dim***) equal to 126 as [R1, R2, R3, ……., R126]. Each vector cell value can be "0" or "1". The vector cell with value "0" means that the corresponding rule is not selected, while the vector cell with value "1" means that the corresponding rule is selected. For example, if in a specific solution at index 30 of the solution vector the value is 0, then this means that Rule#30 is not selected. Another example, if in the solution vector at index 67 the value is 1, then this means that Rule#67 is selected. After each solution is initialized, then the given selected rules in that solution will be applied on the training data. Then, the f-measure (fitness value) for each solution will be determined. The best solution based on f-measure will be assigned to X^* *(prey)* as the current best solution. Then, WOA will execute the main loop on these solutions to update its solutions locations. This main loop will be iterate based on *maximum_iterations* value.

In each loop iteration the solutions positions will be updated using equations 2,5, or 8. At the start of the loop in each iteration, WOA Updates *a, A, C, l,* and *p* values. As mentioned before in [109] they assumed that the whales have 50% probability to switch between shrinking encircling and upward spiral mechanism. Therefore, this was accomplished in WOA algorithm using *p* value as shown in Figs. 2 and 3. Based on *p* value, if it is less than 0.5 and the value of *A is less than 1*, then WOA will execute Eq. (2) (**Exploitation (local search) using Encircling prey phase**). If the value of *p* value is less than 0.5 and *A is greater than or equal 1, then Eq.* (8) (**Exploration (global search) using** (**Search for a prey phase))** will be used to update the solution position. If the value of *p* value is *greater than or equal* 0.5 then WOA will execute Eq. (5) (**Exploitation (local search) using upward spiral mechanism)** to update the solution position. In order for WOA to balance between exploration and exploitation, $\vec{A}$ vector is used. The value of $\vec{A}$ vector will specify whether to update each whale position with reference to the current best solution or with reference to a randomly selected solution. After updating the solutions positions (vectors), its contents (cells) values will be changed, in which each solution the given cell content will be set or reset. At the end of each iteration, WOA will again apply the selected rules in each solution to the training dataset as shown in Fig. 3. In addition, for each solution it will find its fitness value (using f-measure). Now, the best solution based on f-measure will be assigned to X^*. As mentioned before, the main loop will be executed *maximum_iterations* times. Therefore, at the end of WOA main loop execution, the selected rules in the returned WOA best solution X^* will be applied on the testing dataset, as shown in Fig. 3.

5.2. The Cauchy mutation operator

The one-dimensional Cauchy density function is defined by using Eq. (9) [150]:

$$f(x, m, \mu) = \frac{1}{\pi}\frac{m}{m^2 + (x-\mu)^2}, \ -\infty < x < \infty \tag{9}$$

Based on Eq. (9), when the values of $m = 1$ *and* $\mu = 0$. Then it will be a standard Cauchy distribution and it is defined by using Eq. (10) [150]:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \ -\infty < x < \infty \tag{10}$$

Then, the value of the standard Cauchy distribution function can be determined using Eq. (11) [150]:

$$CM(0,1) = \tan\left[(\xi - 0.5)\,\pi\right] \quad \text{such that } \xi \epsilon U[0,1] \tag{11}$$

where the value of ξ is a random value over [0,1].

Several optimization algorithms in the literature were improved by using Cauchy mutation. For example, Wu and Law [151] Improved particle swarm optimization (PSO) exploration ability by using Cauchy mutation. Ali and Pant [152] improved Differential evolution (DE) and preventing it from falling into local optima by using Cauchy mutation. Wang et al. [153] improved Firefly algorithm (FFA) global search ability by using Cauchy mutation. Whereas [154] improved the Krill herd (KH) algorithm to prevent it from being trapped into local optima and improve its population diversity by using Cauchy mutation. Zou et al. [155] improved the balance between the speed and accuracy, and population diversity of PSO by using Cauchy mutation. Li et al. [156] improved gravitational search algorithm (GSA) exploration ability by using Cauchy mutation. Pappula and Ghosh [157] combined Cauchy mutation with cat swarm optimization (CSO) algorithm to prevent CSO from becoming stuck into local optima and to solve CSO premature convergence problem. Zhang et al. [150] improved the diversity of the solutions of Salp Swarm Algorithm (SSA) using Cauchy mutation. This was motivated by previous studies which used Cauchy mutation to improve their optimization algorithms and after using Cauchy mutation achieved better performance. Then, the idea of Cauchy mutation is combined with standard WOA algorithm to improve its population diversity and make a balance between its exploitation and exploration.

5.3. Local search algorithm

In the literature, there are many optimization algorithms which were improved by using local search techniques. For example, Oh et al. [158] improved GA searchability and prevented it from falling into local optima by using Local search operations. Asadzadeh [159] improved solutions diversity in GA and enhanced the best solution by using local search techniques. Toksari [160] improved Ant Colony Optimization (ACO) by using Iterated Local Search (ILS) algorithms. Ou et al. [161] improved the exploitation of Bird-mating optimization (BMO) by using hill climbing. Moradi and Gholampour [162] improved PSO exploitation and solved its premature convergence problem by using local search. Nekkaa and Boughaci [163] improved Harmony search (HS) exploitation by using a stochastic local search (SLS). Mavrovouniotis et al. [164] improved ACO by using local search operator. Mafarja and Mirjalili [165] improved WOA by using simulated annealing (SA). The purpose of using SA was to check if there is a better solution than the current one and it is executed at the end of every WOA loop iteration. In [166], they improved the position of each particle in PSO by using variable neighborhood search (VNS). Shehab et al. [167] improved Cuckoo Search Algorithm (CSA) exploration and convergence speed by

```
Initialize the search agents (whales) population X_i (i = 1, 2, ..., n)
Calculate the fitness of each search agent
X*=the best search agent
while (t < maximum_iterations)
 for each search agent
 Update a, A, C, l, and p values
   if1 (p<0.5)
       if2 (|A| < 1)
             Update the position of the current solution position using Eq. (2)
      else if2 (|A|≥1)
             Select a solution randomly search agent (X_rand) from list of solutions
             Update the position of the current search agent by the Eq. (8)
      end if2
   else if1 (p≥0.5)
        Update the current search position using Eq. (5)
   end if1
 end for
 Check if any search agent (whale) that goes outer the search space and amend it
 Calculate the objective value of each search agent
 Update X* if there is a better solution
 t=t+1
end while
return X*
```

Fig. 2. WOA algorithm [109].

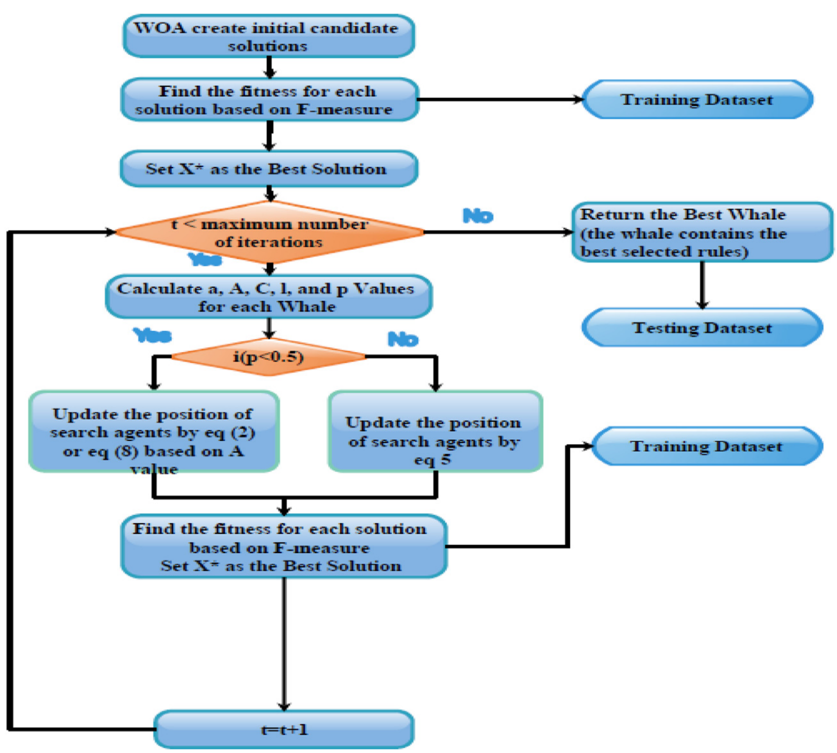


Fig. 3. Workflow of WOA algorithm for aspect extraction.

using hill climbing. Abdel-Basset et al. [168] improved WOA by using local search strategy. Furthermore, Riahi and Kazemi [169] improved the exploitation ability of ACO by using SA as a local search operator. Sakamoto et al. [170] improved PSO convergence speed using hill climbing. Kato et al. [171], improved each particle position in PSO by using Random-Restart Hill Climbing. Lin and Guan [172] improved PSO exploitation by using ILS. Abed-alguni and Alkhateeb [173] solved the premature convergence of CSA by using the β-hill-climbing algorithm [174]. Sulaiman et al. [175] improved artificial bee colony (ABC) exploitation by using evolutionary gradient search (EGS). Zhao et al. [176] improved the exploitation ability of Biogeography-based optimization (BBO) by using variable neighborhood search (VNS). Pei et al. [177] improved convergence speed of Bat algorithm (BA) by using VNS. Yan et al. [178] improved Coral reefs optimization (CRO) exploitation by using SA as a local operator at the end of each CRO iteration. Motivated by these studies which obtained better results by using local search algorithm for improving their optimization algorithms, and to make it suitable for rule selection problem, a new Local Search Algorithm (LSA) was developed to improve WOA exploitation and to prevent it from getting stuck in local optima. LSA works by setting and resetting of rules selection in the current best solution based on the precision and recall of these rules. The new LSA algorithm is shown in Fig. 4.

LSA will be used at the end of each WOA iteration to improve the best solution X^*. LSA algorithm iterates *MaxLIter* times. The input to LSA algorithm represents a list of tuples which contains 126 rules where each tuple contains rule number and rule precision in the first list, while the second list contains each rule with its recall value. At the first step of LSA, the selected rules in current *NSol* will be ranked in ascending order based on rules precision and the result will be saved into the ***P1*** list. In addition, the unselected rules in current *NSol* will be ranked in descending order based on rules precision and the result will be saved into ***P0*** list. After that, the best-selected rules in *NSol* will be in ***P1*** upper half, while the worst selected rules in *NSol* will be in ***P1*** lower half. Whereas, the best-unselected rules in *NSol* will be in ***P0*** lower half, while the worst unselected rules in *NSol* will be in ***P0*** upper half. Furthermore, based on recall values of each rule. The selected rules in current *NSol* will be ranked in ascending order based on rules recall values and the result will be saved into ***R1*** list, while the unselected rules in current *NSol* will be ranked in descending order based on rules recall values and the result will be saved into ***R0*** list. After that, the best-selected rules in *NSol* will be in ***R1*** upper half, while the worst selected rules in *NSol* will be in ***R1*** lower half. Whereas, the best-unselected rules in *NSol* will be in ***R0*** lower half, while the worst unselected rules in *NSol* will be in ***R0*** upper half. In the following steps, five rules will be randomly selected from ***P1*** lower half and reset to 0 (unselected rules) in the *NSol* as these rules represent selected rules with low precision. Also, five rules will be randomly selected from ***P0*** lower half and reset to 1 (selected rules) in the *NSol* as these rules represent unselected rules with high precision. Furthermore, five rules will be randomly selected from ***R1*** lower half and reset to 0 (unselected rules) in the *NSol* as these rules represent selected rules with low recall. Whereas, five rules will be randomly selected from ***R0*** lower half and reset to 1 (selected rules) in the *NSol* as these rules represent unselected rules with high recall. Next, the fitness value of *NSol* will be evaluated on the training dataset by using F-measure. If the *NSol* fitness value better than the current *Leader_pos* fitness value, LSA will update the current *Leader_pos* with *NSol* value.

5.4. Improved WOA based on using Cauchy mutation and LSA for rules selection

This section presents the details of the improvements introduced into standard WOA. As shown in Fig. 5, Cauchy mutation is used to improve the population diversity of WOA and to balance between WOA exploitation and exploration. In addition, LSA is used at the end of WOA iteration to improve its exploitation ability and prevent it from being stuck at local optima. As illustrated in Fig. 5, the first improvement to the standard WOA is based on introducing new mutation rate. For this mutation rate, if ***Mrate*** greater than or equal 0.5 then the new developed Cauchy mutation Eq. (13) will be used to update the position of the current solution; otherwise, Eq. (2) will be used to update the position of the current solution. CM in Eq. (13) represents the Cauchy mutation value which will be determined using Eq. (11).

$$\overrightarrow{D2} = \left|\vec{X}^*(t) - \vec{X}(t)\right| \tag{12}$$

$$\vec{X}(t+1) = \vec{X}(t) + CM * \overrightarrow{D2} \tag{13}$$

The second improvement to WOA, as presented in Fig. 5, includes the use of LSA algorithm at the end of each WOA iteration. LSA will be used to find a better solution than the current best solution. LSA works by adding and dropping rules from the current best based on precision and recall of these rules.

Rules selection by IWOA: IWOA will be applied on a full set of rules to select the best subset of rules combinations while discarding irrelevant rules. In the obtained results of IWOA, the selected rules in each solution are represented as a sequence of 1 and 0. The existence of "1" value in the solution means that the corresponding rule is selected, while the existence of "0" value means that the corresponding rule is unselected. The following steps outline the details of the IWOA algorithm for rules selection:

1. **IWOA Initialization:** In this step, IWOA randomly generates a number of search-agents based on the population size (number of whales). Each whale (solution) contains a subset of randomly selected rules from the full set of 126 rules. Next, IWOA will evaluate the objective values of each solution based on F-measure and the best one will be set as best solution X^*.
2. **Updates Solutions (Whales) Positions:** In this step, IWOA updates each solution position with reference to a randomly selected solution using Eq. (8) or with reference to X^* using either Eq. (2) or (13). The new improvement in this phase is represented by adding a random variable rate and the new Eq. (13). If the random variable rate is less than 0.5, then Eq. (2) will be used to update the current solution position; otherwise, Eq. (13) will be used to update the current solution position.
3. **Apply LSA:** At this phase, LSA will be applied at the end of the current IWOA iteration on current best solution X^*. If LSA algorithm finds a better solution than the current X^* solution, then it will replace the X^* with the new solution, and this step will repeat ***MaxLIter*** times.
4. **IWOA stop execution:** IWOA will iterate and repeats steps 2 and 3 t times (where the variable t used to specify the number of iterations IWOA must iterate over all solutions). In addition, IWOA will update the current best solution X^* if a better solution is found at the end of each iteration.
5. **Best Solution:** At the end of IWOA execution, the best solution represents the set of optimal rules combination which obtained by train IWOA on the training dataset.
6. **Dataset Testing:** The best selected rules subset based on the best solution X^* will be applied to the testing dataset.

6. Pruning algorithm (PA)

PA is required to improve the precision of the extracted aspects. After application of best rules subset selected via IWOA, a set of initial candidate aspects will be obtained. Therefore, PA will be applied on these aspects to remove incorrect aspects and retain correct aspects. Aspects pruning is required because sometimes the extraction rules may extract many incorrect aspects. To avoid these incorrect aspects and retain correct aspects, a three-phase PA is developed. This PA algorithm includes frequency-based pruning, manual-based pruning, and pruning based on direct opinion association. The PA algorithm is shown in Fig. 6 and the details of each pruning phase are as follows:

Pruning based on frequency: The frequency pruning was used in many studies such as [23,39,52,55]. In this study, we count the frequency of each extracted aspect. In the next step, we will prune the aspects with a frequency less than the specified thresholds. The thresholds are set to 2 for single words aspects and 1 for multi-word aspects. If the aspect is a single word with a frequency greater than 2 it will be approved as a correct aspect; otherwise, it will be pruned out and kept for the next two pruning phases. In addition, if the aspect is multi-word aspect with a frequency greater than 1, it will be approved as a correct aspect; otherwise, it will be pruned out and kept for the next

```
Input: all 126 rules are ranked based on their precision, all 126 rules are ranked based on their recall
R=1
MaxLIter=T   (T represents the number of times LSA will repeat)
NSol=Leader_pos (Leader_pos represents the current best solution at end of WOA iteration)
While (R < MaxLIter)
        P1= Sort All Selected Rules in NSol in ascending order based on precision
        P0= Sort All Non-Selected Rules in NSol in descending order based on precision
        P1_Worst_half = P1 half which contains the worst selected rules based on precision
        P1_Best_half = P1 half which contains the best selected rules based on precision
        P0_Worst_half = P0 half which contains the worst Non-selected rules based on precision
        P0_Best_half = P0 half which contains the best Non-selected rules based on precision
        Select 5 rules randomly from P1_Worst_half and reset its values to 0 in NSol
        Select 5 rules randomly from P0_Best_half and set its values to 1 in NSol
        R1= Sort All Selected Rules in NSol in ascending order based on recall
        R0= Sort All Non-Selected Rules in NSol in descending order based on recall
        R1_Worst_half = R1 half which contains the worst selected rules based on recall
        R1_Best_half = R1 half which contains the best selected rules based on recall
        R0_Worst_half = R0 half which contains the worst Non-selected rules based on recall
        R0_Best_half = R0 half which contains the best Non-selected rules based on recall
        Select 5 rules randomly from R1_Worst_half and reset its values to 0 in NSol
        Select 5 rules randomly from R0_Best_half and set its values to 1 in NSol

        NSolVal =Evaluate Fitness value of NSol based on F-measure
        IF (NSolVal > Leader_score):
            Leader_pos= NSol
            Leader_score= NSolVal
        EndIF
        R=R+1
End While
```

Fig. 4. LSA algorithm.

two pruning phases. However, not all pruned out aspects are incorrect aspects, but these may contain correct aspects which are infrequently mentioned in reviews and do not meet frequency pruning thresholds. Therefore, to overcome this problem, we have developed pruning based on product manual.

Pruning based on Product Manual: Normally, each electronic product is accompanied with a PDF product manual (also called user guide, instruction manual or owner's manual). The product manual is a small book that contains all details about the product such as safety instruction, product technical specification, installation instructions, setup instructions, operations instructions, maintenance instructions, service locations, and warranty information.[1] Therefore, for each aspect that was pruned out in the frequency pruning phase, if the aspect frequency in the manual is greater than a specified manual threshold, it will be approved as correct aspect; otherwise, it will be pruned out and kept for the last pruning phase. In this phase, the manual thresholds are set to 2 for single words aspects and 0 for multi-word aspects. If the aspect is a single word with its manual frequency is greater than 2, it will be approved as a correct aspect; otherwise, it will be pruned out and kept for the last pruning phase. In addition, if the aspect is a multi-word aspect with its manual frequency is greater than 0 it will be approved as a correct aspect; otherwise, it will be pruned out and kept for the last pruning phases. However, not all pruned aspects are incorrect aspects. They may contain correct aspects which are not mentioned in the manual or not meet manual pruning thresholds. Therefore, to overcome this problem we have developed a pruning method based on single aspect in the sub sentence, single aspect in the sentence, or the pruned aspect with a direct relation with opinion word.

Pruning based on a single aspect in sub sentence, single aspect in a sentence, or aspect has a direct relation with opinion word: In this phase, for each aspect that was pruned out in the product manual phase, we do the following steps. In the first step we check if there is only one aspect in the sentence [39] which contains the pruned aspect; then, we add that aspect to the list of approved aspects; otherwise, if two or more aspects exist in the sentence that contains the pruned aspect, we will go the second step.

In the second step, we check if the pruned aspect has direct opinion relation such as 'amod', 'nsubj' with an opinion word in the sentence, or if there is only one aspect exists in the sub sentence which contains the pruned aspect, then we add that aspect to the list of approved aspects. Lastly, if none of the two conditions are satisfied, then the PA algorithm will discard that aspect.

Fig. 6 shows the aspect pruning algorithm, where CA represents the initial candidate aspects which were obtained from the rules which were selected by IWOA. In addition, FA represents the final approved aspects, and NFA is the non-frequent aspects. Therefore, pruning algorithm works in phase order by applying frequency-based pruning at first, then pruning based on product manual and the last phase which is pruning based on one aspect in the sub sentence, one aspect in the sentence or the aspect has direct opinion relation with opinion word. The following are examples of the application of PA algorithm, where these examples were taken from [23] datasets. In the sentence "*it 's great to switch to **spot metering** and actually see it working on the lcd screen*" the aspect "***spot metering***" has a frequency of 1 in Canon dataset.

Therefore, based on frequency pruning it was pruned out, but based on the product manual of the Canon Camera, it was approved. In addition, in the sentence "*this camera also has a great feel and **weight** to it*" "***weight***" aspect does not satisfy frequency threshold in Canon dataset, then based on manual "***weight***" was approved as a correct aspect. Furthermore, other examples which are approved based on product manual from all datasets as the following which are indicated in bold: "*the included **memory card***

[1] https://en.wikipedia.org/w/index.php?title=User_guide&oldid=872157779.

```
Initialize the search agents (whales) population Xi (i = 1, 2, ..., n)
Calculate the fitness of each search agent
X*=the best search agent
while (t < maximum_iterations)
  for each search agent
  Update a, A, C, l, and p values
    if1 (p<0.5)
      if2 (|A| < 1)
          Mrate=rand(0,1)
          If3  (Mrate < 0.5)
              Update the position of the current solution position using Eq. (2)
          else
              Update the position of the current solution position using Eq. (13) using Cauchy
              mutation
          end if3
      else if2 (|A|≥1)
          Select a solution randomly search agent (Xrand) from list of solutions
          Update the position of the current search agent by the Eq. (8)
      end if2
    else if1 (p≥0.5)
      Update the current search position using Eq. (5)
    end if1
  end for
  Check if any search agent (whale) that goes outer the search space and amend it
  Apply LSA algorithm on X* to check if there is a better solution
  Update X* if there is a better solution
  t=t+1
end while
return X*
```

Fig. 5. Proposed IWOA algorithm based on Cauchy mutation and LSA algorithm.

is too small", "*the* ***macro*** *works great for medical photographs and the auto mode is terrific for point and shoot*",

"*the* ***lens cover*** *is surely loose*", "*the* ***service*** *from the supplier was exceptional*", "*the* ***volume key*** *can be hard to press*", "*the* ***pc sync*** *feature is superb that comes with nokia pc suite software*", "***audio*** *on video also lacking*".

The following are examples of pruning based on direct opinion association, sub-sentence, or sentence. For example, in the sentence "*and for those that are interested the* ***recharger*** *works anywhere in the world and is quite small*", in this sentence it contains two nouns include "***recharger***" and "***world***", but "***recharger***" has direct opinion association with opinion word "***work***" via "***nsubj***" relation, then "***recharger***" was approved as a correct aspect. Furthermore, in the sentence "*basic* ***usage*** *is easy, but the remote has a lot of buttons that i have n't used*" the sentence has more than one aspect, but "***usage***" aspect was approved because there is only one aspect in the sub-sentence. Also, "***usage***" aspect has a direct opinion association with "***easy***" opinion word via "***nsubj***" relation. In the sentence "*the little* ***jog dial*** *seems weak and quirky and i hope i do n't figure out a way to break it.*", then the "***jog dial***" aspect was approved because there is only one aspect exists in the sub sentence, where the end of the sub-sentence indicated by "***and***". This is another example which was approved based on that there is only one aspect exists in the sub-sentence, where "**,**" indicates the end of sub-sentence "*simple* ***click buttons***, *back buttons volume and display are very easy to read, access and use*". In addition, the aspect "***click buttons***" has "***amod***" relation with opinion word "***simple***". In the example "*the* ***grain*** *was terrible*", the aspect "***grain***" was approved because there is only one aspect exists in the sentence. Another example based on that one aspect exists in the sub-sentence is "*solid, high-quality* ***construction***". In the example "*i am very pleased with its quality and* ***durability***", in this sentence "***durability***" aspect was approved because there is only one aspect exists in the sub sentence which indicated by "***and***". In example, "*the* ***voice quality*** *is very good, and it gets great reception*" the aspect "***voice quality***" was approved because there is only one aspect exists in the sub-sentence indicated the end of sub-sentence by "**,**" and aspect "***voice quality***" has a direct opinion association with opinion word "***good***" via "***nsubj***" relation. In the sentence "the **vibrate setting** is loud" the aspect "***vibrate setting***" was approved as a correct aspect because there is only one aspect exists in the sentence. In the sentence "***touchups***, *redeye, and so on are very easy to alter, and correct*", the aspect "***touchups***" was approved because there is only one aspect exists in the sub-sentence which indicated by "**,**". In the sentence "*the design and* ***construction*** *are excellent*", the "***construction***" aspect was approved because there is only one aspect exists in the sub-sentence. In sentence "*awesome camera with huge* ***print quality***" it has two aspects "***camera***" and "***print quality***", in this case, based on direct opinion association, the aspect "***print quality***" was approved because it has direct opinion association with "***huge***" opinion word via "***amod***" relation. In the sentence, "*great feature list, poor* ***reliability***" the aspect "***reliability***" was approved because the there is only one aspect in the sub-sentence which indicated by "," and the "***reliability***" aspect has direct opinion relation with "***poor***" opinion word via "***amod***" relation. Therefore, based on the proposed PA algorithm, any unrelated extracted aspects or noise will be removed automatically. In addition, only related aspects will be approved.

As shown and discussed from the provided examples and details, aspects pruning is important, because sometimes the extraction rules will extract incorrect aspects that are not frequent in the datasets. To remove these incorrect aspects, a frequency pruning can be used, since the aspect word that occurs frequently is normally considered as the correct aspect [23]. However, sometimes pruned aspects are correct aspects, but do not occur frequently in the dataset. Therefore, manual pruning will be used

```
CA= Aspects received from application rules selected by IWOA
// Frequency Pruning Phase (Phase 1)
for each aspect in CA do:
        If frequency(aspect) in dataset > threshold1: (threshold1 =2 for single word aspects and 1 for
        multiword aspects)
            Add aspect to the list of final correct aspects FA // Final Aspects
            NFA=CA-FA              //NFA is Non-frequent aspects (pruned aspects)
      end if
end for
// Product Manual Pruning Phase (Phase 2)
for each aspect in NFA do:
        If frequency(aspect) in manual > threshold2: (threshold2 =2 for single word aspects and 0 for
        multiword aspects)
           Add aspect to the list of final correct aspects FA // Final Aspects
           NFA=CA-FA              //NFA is Non-frequent aspects (pruned aspects)
      end if
end for
// Direct opinion association, sub-sentence, or sentence Pruning Phase (Phase 3)
for each aspect in NFA do:
        If only one aspect found in current sentence
            Add aspect to the list of final correct aspects FA // Final Aspects
           NFA=CA-FA              //NFA is Non-frequent aspects (pruned aspects)
       else if more than one aspect found in sentence:
               if the pruned aspect has direct opinion association with an opinion word with relation  such as
               'amod','nsubj' or the subsentence has only one aspect: // sub-sentence can be indicated  by
               and,or,but,",",".":

                       Add aspect to the list of final correct aspects FA // Final Aspects
                      NFA=CA-FA              //NFA is Non-frequent aspects (pruned aspects)
               else
                     Discard aspect
      end if
 end for
 Return FA as the final aspect list
```

Fig. 6. Proposed aspects pruning algorithm.

to look for these candidate aspects. Furthermore, sometimes an aspect which is a pruned based on frequency and manual can be a correct aspect but does not frequently occur or is not mentioned in the manual. Thus, to solve this issue, a pruning based on direct opinion association is used to check these candidate aspects. Therefore, as shown in Fig. 7, the PA algorithm works in phase order. Such that for each candidate aspect, the first step is to check its frequency, then it will approve the given aspect if its frequency is greater than the frequency threshold; otherwise, the PA algorithm will check the given aspect based on product manual. In this phase, PA will approve the given aspect if its frequency is greater than the manual frequency threshold; otherwise PA algorithm will check the given aspect based on direct opinion association. In this phase, PA will approve the given aspect if it matches with one of the specified direct opinion association rules; otherwise PA algorithm will discard the given incorrect aspect. The above-mentioned examples give an examples of PA application.

The proposed aspect extraction (AE) algorithm is shown in Fig. 8, in which the AE algorithm is composed from number of phases. The first phase is the full set of extraction rules which contains 126 rules. However, using all these rules for extraction will improve the recall, but will decrease the extraction precision. Therefore, these rules will be an input to IWOA. IWOA will use the training dataset to find optimal subset of rules from the full set of rules. IWOA starts by creating number of randomly generated solutions. Each solution is a vector with ***Dim*** of 126. As detailed before for the WOA algorithm, IWOA will update the solutions inside the main loop until it reaches the *maximum_iterations* value. The difference between WOA and IWOA is that in IWOA we develop new update position equation based on CM to improve solutions diversity (the selected and unselected rules). In addition, at the end of each iteration in IWOA, IWOA will call LSA on current best solution to further improve it by setting or resetting rules in the solution. At the end of IWOA application, IWOA will return the best solution as a combination of 0 and 1. For a case example, assume we have the following settings which will be used by IWOA:

(1) The number of rules (problem dimension) ***Dim=126***
(2) The population size is 10
(3) The number of iterations *maximum_iterations*=30

Then, IWOA at initialization will generates randomly 10 different solutions, where each solution is one-dimensional vector with 126 cells. Each cell can be either 0 or 1. After finishing IWOA, it will return the best solution. For example, the following represents an example of the best solution returned by IWOA.

[0,0,0,1,0,0,1,1,1,0,0,0,0,0,0,1,0,0,1,0,0,1,1,0,0,1,1,0,0,1,0,0,0,1,0, 1,1,0,0,1,0,1,1,1,0,0,0,1,1,0,0,0,1,0,0,0,1,1,0,1,1,1,1,1,1,1,0,0,0,0,1,1, 0,0,1,0,1,0,0,0,1,0,0,0,1,0,0,0,1,1,0,1,0,0,1,1,1,1,1,0,0,0,0,1,1,0,0,1,0, 0, 1,0,1,1,1,1,1,0,0,0,1,1,1,1,0,0]

The cells with 0 value indicate that the corresponding rule was not selected, while the cells with 1 value indicate that the corresponding rule was selected. Therefore, based on the above solutions, we only consider the selected rules which are represented by the following rules numbers.

(4, 7, 8, 9, 16, 19, 22, 23, 26, 27, 30, 34, 36, 37, 40, 42, 43, 44, 48, 49, 53, 57, 58, 60, 61, 62, 63, 64, 65, 66, 71, 72, 75, 77, 81, 85, 89, 90, 92, 95, 96, 97, 98, 99, 104, 105, 108, 111, 113, 114, 115, 116, 117, 121, 122, 123, 124).

For example, in the best solution index 12 the value is 0, which means that rule#12 not included (not selected) in the best solution. Another example at index 27 the value is 1, which means

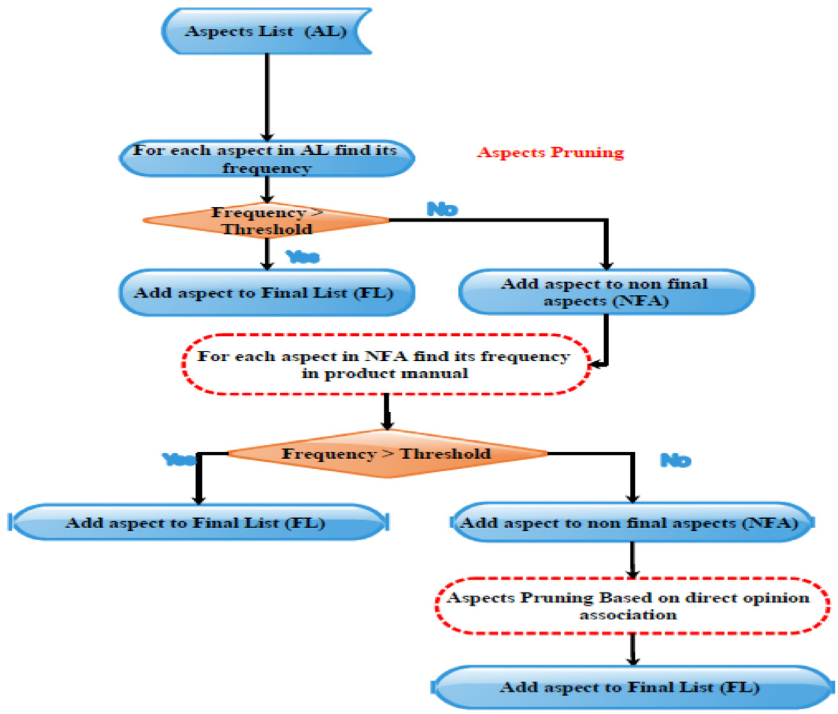


Fig. 7. Workflow of aspects pruning algorithm.

that rule#27 is included in the best subset of rules. Therefore, this best subset of rules will be used on the testing dataset. The candidate aspects which will result from application of these rules will go to the last phase which includes the application of PA algorithm. PA is used to purify these extracted aspects such as the examples provided before on the application of different phases of PA algorithm.

7. Experimental results and analysis

7.1. Evaluation datasets

To test and evaluate the performance and effectiveness of the proposed aspect extraction algorithm we used customer review datasets [23]. It is a benchmark dataset for aspect extraction which was used by the majority of researchers in aspect extraction [22]. These datasets contain customer reviews of five electronic products. The datasets are as follows: two digital cameras types (Canon D1, Nikon D2), mobile phone (Nokia D3), MP3 player (Creative D4), and DVD player (Apex D5). These datasets were annotated manually by Hu and Liu [23], in which each sentence was labeled with the contained aspect. The details and statistics of these datasets are presented in Table 6. The other benchmark datasets which are used from [86], where these datasets include reviews about computer and speaker. In addition, they annotated the datasets manually by labeling each sentence with the contained aspects. The details and statistics of these datasets are presented in Table 7.

7.2. Evaluation metrics and comparison baselines

The metrics which were used for algorithms evaluation include precision, recall, and F-measure. Precision is the ratio of the correct extracted aspects based on the gold standard to the total number of extracted aspects. The recall is the ratio of correct extracted aspects based on the gold standard to the total number of aspects in the gold standard. F-measure is the harmonic mean of recall and precision. The baseline methods that we compared our work with include the state-of-the-art aspect extraction works and the latest ones:

Table 6
Dataset1.

Data	Product	#Sentences
D1	Canon digital camera	597
D2	Nikon digital camera	346
D3	Nokia Cellphone	546
D4	Creative Mp3 player	1716
D5	Apex DVD player	740

Table 7
Dataset2.

Data	Product	#Sentences
D6	Computer	531
D7	Speaker	689

1. Double Propagation (DP) [39], DP represents one of the famous aspect extraction algorithm, wherein DP eight dependency relations rules were used for aspect extraction.
2. Htay [34], in which syntactic pattern rules were used for aspect extraction.
3. RubE [52], which represents one of the recent works. In RubE they extended DP rules with some new rules.
4. Two-fold rule-based model (TF-RBM) [55] which represents one of the recent work, where sequential patterns were used for aspect extraction.
5. Rule Selection using a Local Search algorithm (RSLS) [51], which is one of the recent aspect extraction algorithms where they extended DP rules with new rules. In addition, in RSLS they applied Simulated annealing for rules selection.
6. Convolutional neural networks with linguistic patterns CNN + LP [26].

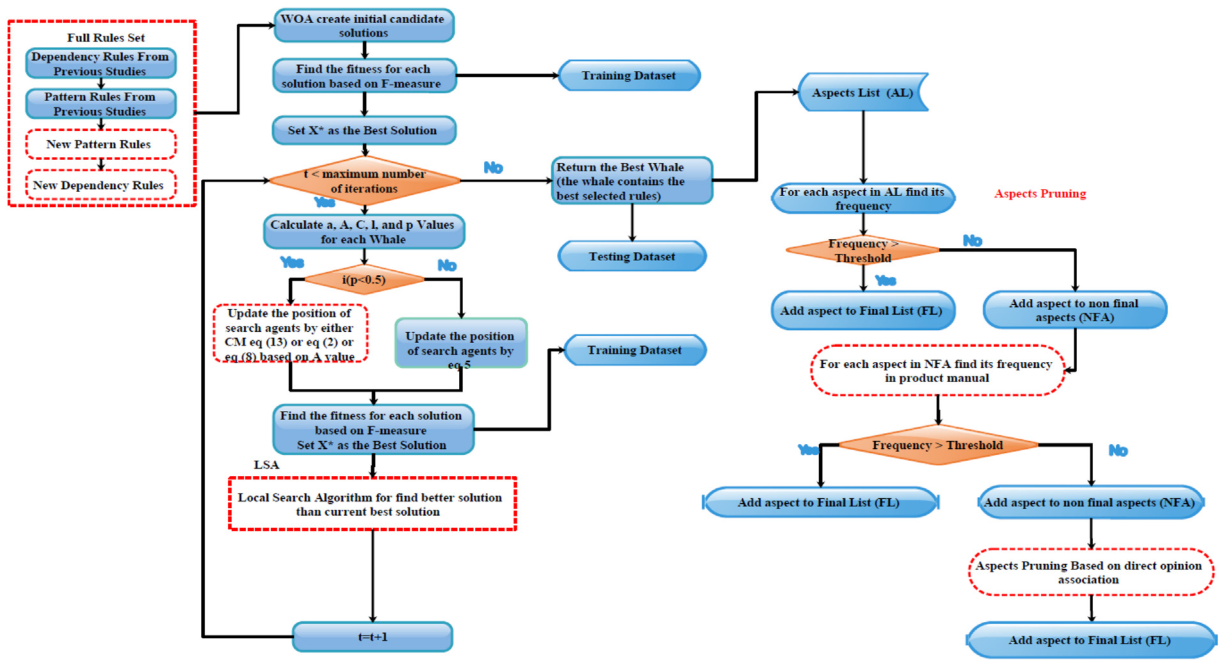


Fig. 8. Proposed Aspect Extraction (AE) Algorithm.

7. SPR (Sequential Pattern-Based Rules) [88], which represents a recent work for aspect extraction which used sequential patterns for aspect extraction.
8. Feng [89] represents one of the recent work for aspect extraction. In this technique, they adopted topic modeling with synonym recognition method for aspect extraction.
9. NGD+CNET [107], which also represents one of the recent work for aspect extraction. In this technique, they used NPL features with NGD and ConceptNet method for aspect extraction.
10. DomSent [91], which represents the latest and most recent work for aspect extraction. In this technique, they considered noun/noun phrase as possible aspects. In addition, they used the opinion lexicon which they created.
11. MCNN [96], which also represents the latest and most recent work for aspect extraction. This method based on using MCNN deep architecture for aspect extraction.
   All these baseline works used the same datasets [23] in their experiments. In addition, the results of each method were taken from their reported results in the original papers are as follows: DP [39], Htay [34], RSLS [51], CNN+LP [26], RubE [52,55], TF-RBM [55], SPR [88], Feng [89], NGD+CNET [107], DomSent [91], and MCNN [96].

7.3. Data cleaning and pre-processing

To obtain a clean text, we removed all human annotations and the words which includes special characters. For example, in the following sentence "camera[+2]##i love this camera." which is taken from the datasets, the cleaning includes removing the annotation, [+2], and ## to get the sentence which contains the aspect to be extracted "i love this camera".

After cleaning the dataset, some pre-processing steps are required. Each review is tokenized using sentence tokenizer. After getting each sentence, each sentence will be parsed using Stanford tagger to obtain the part-of-speech (POS) of each word in the sentence to use it with syntactic pattern rules. In addition, for each sentence, Stanford dependency parser will be used to extract the list of dependency relations within a sentence to use them with dependency relation rules.

7.4. Experimental results

To evaluate the performance of the proposed work, six experiments were conducted in this study. In the first experiment, the full set of 126 aspect extraction rules were applied to each dataset. In the second experiment, we applied IWOA on the 126 rules for selecting the best rules combination. In addition, in the second experiment, one of the datasets was used for IWOA training, while another different dataset was used for testing the selected rules by IWOA. In the third experiment, we compared the performance results of IWOA with the standard WOA algorithm results. In the fourth experiment, we compared the performance results of IWOA with seven well-known optimization algorithms results. In the fifth experiment, we applied PA on the extracted aspects resulted from applying IWOA. In the last experiment, we compared the proposed aspect extraction algorithm IWOA+PA with other state-of-the-art aspect extraction works.

In all experiments, we used Stanford Dependency Parser for dependency parsing and Stanford tagger for part-of-speech (POS) tagging. In addition, for all optimization algorithms, the population was set to 10 and the iterations to 30. The parameter settings for all optimization algorithm (FFA, Sine Cosine Algorithm (SCA) [179], SSA, PSO, IWOA, WOA, Moth-flame optimization (MFO) [180], Multi-verse optimizer (MVO) [181], Grey wolf optimizer (GWO) [182]) are shown in Table 8. In the conducted experiments, F-measure is the fitness value that we used for testing all optimization algorithms performance.

7.4.1. Experiment 1: Results obtained from applying full set of 126 rules on datasets

In this experiment, the full set of 126 rules were applied to each dataset. The results of this experiment are shown in Table 9. As shown in Table 9, the recall in each dataset is very high in comparison with the precision. Thus, it is confirmed from the obtained results the importance of rules selection for aspect extraction because using all rules will result in extracting many incorrect aspects. Therefore, we applied the IWOA algorithm on the 126 rules to select the optimal subset of rules combination and ignore irrelevant and low-quality rules.

Table 8
Parameters settings of the algorithms used.

Algorithm	Parameter
FFA	Alpha = 0.5, Beta_min = 0.20, and Gamma = 1 set as in [183,184]
SCA	a = 2 set as in [179]
SSA	c_2 and c_3 random numbers over [0,1] set as in [185]
PSO	Acceleration constants (C1 = 2, C2 = 2), and Inertia Weight (W1 = 0.2, W2 = 0.9) set as in [186,187]
IWOA	a = [2,0], b = 1 set as in [109], and LSA iteration = 10
WOA	a = [2,0], and b = 1 set as in [109]
MFO	b = 1 set as in [180]
MVO	Maximum wormhole existence probability = 1, and Minimum wormhole existence probability = 0.2 set as in [181]
GWO	a = [2,0] set as in [182]

Table 9
Results obtained from applying the full 126 rules set on each dataset.

Data	Precision	Recall	F-measure
D1	0.73	0.96	0.83
D2	0.74	0.97	0.84
D3	0.76	0.98	0.86
D4	0.74	0.97	0.84
D5	0.76	0.95	0.84
Avg	**0.75**	**0.97**	**0.84**

Table 10
Results obtained from applying IWOA for rules selection.

Testing data	Precision	Recall	F-measure	Training data
D1	0.86	0.94	0.90	D5
D2	0.84	0.94	0.89	D3
D3	0.89	0.94	0.91	D5
D4	0.85	0.96	0.90	D1
D5	0.86	0.92	0.89	D3
Avg	**0.86**	**0.94**	**0.90**	

7.4.2. Experiment 2: Results obtained from applying IWOA for rules selection

In this experiment, we applied IWOA on the full set of rules to select the optimal subset of rules combination for aspect extraction. The results of IWOA are shown in Table 10, where one dataset was used for training and another dataset was used for testing in this phase.

As shown in Table 10, there is a clear improvement on the precision with 11% and a slight decrease in the recall by about 3%. Also, there is another improvement in F-measure with 6%. To confirm the superiority of IWOA, the performance of IWOA was compared with the standard WOA performance in experiment 3.

7.4.3. Experiment 3: IWOA comparison with native WOA

In this experiment, a comparison between the performance results of the standard WOA and the proposed IWOA are shown in Table 11. Based on the bold results from all datasets in Table 11, there is a clear outperformance of IWOA over WOA on all datasets. These results proved the ability of IWOA to escape from local optima based on using the LSA algorithm and the ability of IWOA to generate diverse solutions based on using CM equation.

As shown from Fig. 9, IWOA outperforms WOA with 2% in precision, 4% in recall, and 3% in terms of F-measure respectively. Our experiments confirm the superiority of IWOA over WOA.

7.4.4. Experiment 4: IWOA comparison with other optimization algorithms

To provide additional support for the superiority of IWOA, it was compared with 7 well-known optimization algorithms including PSO [188], FFA [189], GWO [182], MFO [180], SCA [179], MVO [181], and SSA [185]. The results of experiment four are shown in Tables 12 and 13.

Table 11
IWOA results comparison with native WOA.

Testing data	WOA			IWOA			Training data
	P	R	F	P	R	F	
D1	0.85	0.91	0.88	**0.86**	**0.94**	**0.90**	D5
D2	0.83	0.90	0.86	**0.84**	**0.94**	**0.89**	D3
D3	0.82	0.92	0.87	**0.89**	**0.94**	**0.91**	D5
D4	0.84	0.92	0.88	**0.85**	**0.96**	**0.90**	D1
D5	0.85	0.87	0.86	**0.86**	**0.92**	**0.89**	D3
Avg	0.84	0.90	0.87	**0.86**	**0.94**	**0.90**	

As displayed in Tables 12 and 13, it is clearly noticed that IWOA algorithm outperforms all other algorithms over all datasets as indicated by bold fonts in both tables. Our experiments results have further confirmed the ability of IWOA to balance between exploitation and exploration, solutions diversity, and avoid falling into local optima. As the result of comparing IWOA with other state-of-the-art optimization algorithms, and as shown in Fig. 10, it confirmed the superiority of IWOA over all other algorithms. IWOA outperforms GWO with 1% in precision, 6% in recall, and 4% in terms of F-measure respectively. Also, IWOA outperforms MVO with 2% in precision, 4% in recall, and 3% in terms of F-measure respectively. Furthermore, IWOA outperforms MFO with 3% in precision, 3% in recall, and 3% in terms of F-measure respectively. In addition, IWOA outperforms PSO with 5% in precision, 2% in recall, and 4% in terms of F-measure respectively. Whereas, IWOA outperforms SSA with 3% in precision, 4% in recall, and 4% in terms of F-measure respectively. IWOA outperforms SCA with 3% in precision, 3% in recall, and 3% in terms of F-measure respectively Finally, IWOA outperforms FFA approach with 3% in precision, 3% in recall, and 3% in term of F-measure respectively. Therefore, the IWOA algorithm has the ability to outperform all of these optimization algorithms which is confirmed by the achieved results.

Our proposed approach is supervised in terms of rules selection. To select the optimal subset of rules, one dataset from any domain is required to train IWOA on it to select the optimal subset of rules and discard low quality rules. The dataset must have each sentence annotated with the corresponding aspects. Thus, any annotated dataset can be used for training IWOA, after which the selected subset of rules can be applied to any dataset regardless of its domain. In addition, this is clearly shown from the conducted experiments where one dataset was used for training IWOA to select the optimal rules, while another dataset from another domain was use for testing the selected rules.

The following rules represents an example of frequently selected rules by IWOA, where the number represents the rules numbers, as shown in Tables 3–5:

(4, 7, 8, 9, 16, 19, 22, 23, 26, 27, 30, 34, 36, 37, 40, 42, 43, 44, 48, 49, 53, 57, 58, 60, 61, 62, 63, 64, 65, 66, 71, 72, 75, 77, 81, 85, 89, 90, 92, 95, 96, 97, 98, 99, 104, 105, 108, 111, 113, 114, 115, 116, 117, 121, 122, 123, 124).

7.4.5. Experiment 5: Results obtained from applying the pruning algorithm

There are many incorrect aspects in the initial list of aspects which were obtained by the selected rules by IWOA. Therefore, pruning is required to remove incorrect aspects while retaining correct aspects. The results after applying pruning algorithm are shown in Table 14. Based on the obtained results from Table 14, there are additional improvements in precision and F-measure values with increase 6% and 2% respectively. Whereas, there is a slight drop in recall by 1%. Fig. 11 shows the impact of each phase on aspect extraction performance.

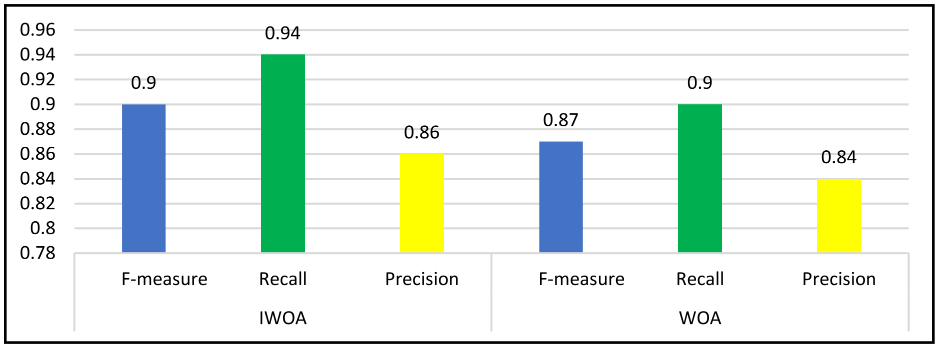


Fig. 9. IWOA comparison with WOA.

Fig. 10. IWOA comparison with other optimization algorithms.

Table 12
IWOA results comparison with other optimization algorithms.

Test data	FFA			SCA			SSA			PSO			IWOA			Training data
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	
D1	0.85	0.93	0.89	0.85	0.91	0.88	0.81	0.89	0.85	0.81	0.93	0.87	**0.86**	**0.94**	**0.90**	**D5**
D2	0.81	0.92	0.86	0.83	0.93	0.88	0.83	0.92	0.87	0.80	0.91	0.85	**0.84**	**0.94**	**0.89**	**D3**
D3	0.86	0.89	0.87	0.83	0.88	0.85	0.82	0.91	0.86	0.84	0.92	0.88	**0.89**	**0.94**	**0.91**	**D5**
D4	0.82	0.93	0.87	0.82	0.92	0.87	0.83	0.93	0.88	0.81	0.93	0.87	**0.85**	**0.96**	**0.90**	**D1**
D5	0.82	0.87	0.84	0.82	0.91	0.86	0.84	0.87	0.85	0.81	0.90	0.85	**0.86**	**0.92**	**0.89**	**D3**
Avg	0.83	0.91	0.87	0.83	0.91	0.87	0.83	0.90	0.86	0.81	0.92	0.86	**0.86**	**0.94**	**0.90**	

Table 13
IWOA results comparison with other optimization algorithms.

Test data	MFO			MVO			GWO			IWOA			Training data
	P	R	F	P	R	F	P	R	F	P	R	F	
D1	0.81	0.93	0.87	0.83	0.93	0.88	0.83	0.90	0.86	**0.86**	**0.94**	**0.90**	**D5**
D2	0.81	0.91	0.86	0.83	0.91	0.87	0.83	0.87	0.85	**0.84**	**0.94**	**0.89**	**D3**
D3	0.88	0.88	0.88	0.86	0.88	0.87	0.87	0.85	0.86	**0.89**	**0.94**	**0.91**	**D5**
D4	0.81	0.93	0.87	0.84	0.90	0.87	0.84	0.91	0.87	**0.85**	**0.96**	**0.90**	**D1**
D5	0.85	0.90	0.87	0.85	0.86	0.85	0.84	0.86	0.85	**0.86**	**0.92**	**0.89**	**D3**
Avg	0.83	0.91	0.87	0.84	0.90	0.87	0.85	0.88	0.86	**0.86**	**0.94**	**0.90**	

Table 14
Results obtained from applying pruning algorithm (PA).

Data	Precision	Recall	F-measure
D1	0.91	0.93	0.92
D2	0.91	0.92	0.91
D3	0.93	0.93	0.93
D4	0.93	0.95	0.94
D5	0.92	0.91	0.91
Avg	**0.92**	**0.93**	**0.92**

7.4.6. Experiment 6: Results comparison with baseline methods

The comparison results of IWOA+PA with other aspect extraction baseline works are presented in Tables 15–17. From these results, IWOA+PA is superior to all other baseline works in terms of precision, recall, and F-measure as indicated by bold font. These results confirmed the ability of our proposed approach IWOA+PA to perform more consistently because it can balance between the recall and precision. From these results, it is shown that IWOA+PA outperforms DP with 4% in precision, 10% in recall, and 6% in terms of F-measure respectively. Also, IWOA+PA outperforms RubE with 5% in precision, 5% in recall, and 5% in terms of F-measure respectively. Furthermore, IWOA+PA outperforms

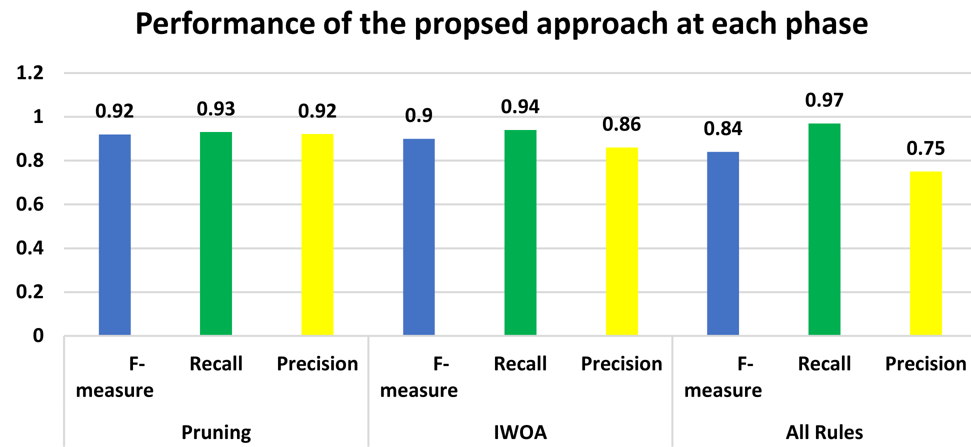


Fig. 11. Impact of each phase on performance.

TF-RBM with 5% in precision, 1% in recall, and 3% in terms of F-measure respectively. In addition, IWOA+PA outperforms RSLS with 7% in precision, 2% in recall, and 4% in terms of F-measure respectively. Whereas, IWOA+PA outperforms Htay work with 19% in precision, 7% in recall, and 13% in terms of F-measure respectively. Moreover, IWOA+PA outperforms CNN+LP approach with 2% in precision, 7% in recall, and 4% in terms of F-measure respectively. IWOA+PA outperforms SPR with 6% in precision, 2% in recall, and 3% in terms of F-measure respectively. As shown from Table 17, IWOA+PA outperforms Feng method with 5% in precision, 5% in recall, and 5% in terms of F-measure respectively. Also, IWOA+PA outperforms MCNN with 3% in terms of F-measure. However, in MCNN paper, they did not report the results values of precision and recall. Further, IWOA+PA outperforms NGD+CNET with 4% in precision, 9% in recall, and 6% in terms of F-measure respectively. Finally, IWOA+PA outperforms DomSent with 4% in precision, 8% in recall, and 6% in terms of F-measure respectively. Generally speaking, IWOA+PA algorithm dominates all other state-of-the-art and recent aspect extraction works based on the obtained results.

To confirms the superiority of the proposed IWOA+PA algorithm, we also compared it with baseline works using new datasets from other domains as shown in Tables 18–21.

As shown in Table 18, using all rules will improve the recall, but it decreases the precision. Therefore, proper selection of the included rules is required. As shown from the results, the use of IWOA algorithm improve the precision by 14%. To further improve the precision, as shown in Table 18, the use of PA algorithm further improves the precision by 9%.

To check the performance of IWOA, we also compared it with other optimization algorithms, as shown from Tables 19 and 20. These results confirmed the superiority of IWOA over all other optimization algorithms. IWOA outperforms all other baseline optimization algorithms in terms of precision, recall, and F-measure as shown in bold. Therefore, the main reasons behind IWOA results and outperformance are clearly based on a number of points including: (1) IWOA ability to balance between exploration and exploitation; (2) IWOA population diversity based on the new CM update equations (diversity of selected and unselected rules); and (3) The ability of IWOA to avoid local optima by using the new LSA algorithm (which can improve the current best solution by set or rest rules).

Now, as shown in Table 21, IWOA+PA also outperforms other baseline algorithms as shown in bold fonts by using the new datasets. This also confirms the superiority of IWOA+PA in comparison with other works. The results of AER are taken from [85]. In addition, DP, CRF, and RSLS results from [51]

As shown from the comprehensive conducted experiments using different types of datasets, IWOA+PA achieved remarkably better results than other works. These results are expected because the included rules can cover all text types. However, in these rules some perform better. This is because one rule can extract aspects with high recall, but it is also extracting many incorrect aspects with low precision. On the other hand, there are rules that can extract aspects with high recall with less incorrect extracted aspects and higher precision. Therefore, IWOA is used to select the optimal combination among these rules. IWOA can take any set of rules regardless of their quality and returns the best subset of selected rules. At each iteration of IWOA it will diversify these rules based on the achieved performance. Also, at the end of each IWOA iteration, LSA algorithm will improve the current best solution by set or reset rules. Hence, IWOA can iteratively make an improvement over the selected subset of rules. Finally, at the end of IWOA execution, IWOA will return the optimal subset of rules from the full set of rules. The previous discussion shows that rules selection is an important phase, because including all rules for extraction will improve recall, but will also result in more errors (incorrect aspects) and low precision.

7.5. Discussion of results

To give deeper insight over our achieved results, the following outlines how the results are improved. In addition, a comparison with related baseline works. In this study, a number of contributions helped to achieve these results. These can be detailed by the following subsections.

(1) Effect of rules combination: the use of different rules types has improved the recall of aspect extraction. This is clearly noticeable from the achieved results based on Tables 9 and 18 with an average recall is 97%. This result is expected as there are a number of new rules which extract aspect that was not extracted by previous studies. For example, Rule# 88 when was applied to a sentence such as *"my favorite being the* ***games*** *and the* ***pim****, and the* ***radio****.*", it extracted three aspects from this sentence, which cannot be extracted by previous studies when a sentence contains more than one aspect. In addition, the new improvement which was added to the rules using OP/JJ combinations, where it improved the extraction performance. This is because a majority of studies in the literature considered an adjective as an opinion word or simply used an opinion lexicon, but not both, resulting in missing many correct aspects. In another example, in a majority of studies in literature they considered aspect just as noun or noun phrase, but do not consider verb aspects. Number of rules were used in this study to take care of this issue and extract

Table 15
Results comparisons including Precision (P), Recall (R), and F-measure to other approaches.

Data	DP			RubE			TF-RBM			RSLS			IWOA+PA		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
D1	0.87	0.81	0.84	0.87	0.86	0.86	0.80	0.89	0.84	0.85	0.91	0.88	0.91	0.93	0.92
D2	0.90	0.81	0.85	0.90	0.86	0.88	0.87	0.93	0.90	0.89	0.94	0.91	0.91	0.92	0.91
D3	0.90	0.86	0.88	0.90	0.91	0.90	0.92	0.93	0.92	0.83	0.90	0.86	0.93	0.93	0.93
D4	0.81	0.84	0.82	0.87	0.90	0.88	0.86	0.93	0.90	0.82	0.91	0.86	0.93	0.95	0.94
D5	0.92	0.86	0.89	0.90	0.85	0.87	0.88	0.90	0.89	0.86	0.90	0.88	0.92	0.91	0.91
Avg	0.88	0.83	0.86	0.87	0.88	0.87	0.87	0.92	0.89	0.85	0.91	0.88	**0.92**	**0.93**	**0.92**

Table 16
Results comparisons including Precision (P), Recall (R), and F-measure to other approaches.

Data	Htay			CNN+LP			SPR			IWOA+PA		
	P	R	F	P	R	F	P	R	F	P	R	F
D1	0.74	0.92	0.82	0.93	0.85	0.89	0.81	0.92	0.87	0.91	0.93	0.92
D2	0.71	0.81	0.76	0.83	0.87	0.85	0.86	0.96	0.90	0.91	0.92	0.91
D3	0.74	0.82	0.78	0.93	0.88	0.90	0.89	0.95	0.92	0.93	0.93	0.93
D4	0.70	0.76	0.73	0.93	0.86	0.89	0.89	0.91	0.89	0.93	0.95	0.94
D5	0.78	0.97	0.87	0.90	0.84	0.87	0.86	0.81	0.84	0.92	0.91	0.91
Avg	0.73	0.86	0.79	0.90	0.86	0.88	0.86	0.91	0.89	**0.92**	**0.93**	**0.92**

verb aspects. Furthermore, previous studies which were based on considering aspects as noun or noun phrase only, have a problem when dealing with a sentence which have more than one noun and sometimes it will extract the incorrect noun as the target aspect. However, the opinion word in the sentence is related to the other noun. We resolved this issue by adding restriction to the developed rules as shown in the rules table. Therefore, the combination of rules with different types improved the recall, but still suffers from lower precision. This lower precision is expected, because we have several rules included with different quality. Thus, proper selection of included rules is an important step, as confirmed by IWOA improvement on precision results.

(2) Effect of rules selection: choosing the optimal rules combination is an important step to improve the extraction precision. This is clearly observed from IWOA achieved results over all datasets in comparison with standard WOA and other optimization algorithms, as shown in Tables 11–13, 19–20. The achieved results by IWOA improved the precision of extraction process while maintaining high recall results. These results were achieved because of the two major improvements included into IWOA. From these results, IWOA outperformed other optimization algorithms over all metrics including recall, precision, and F-measure. This outperformance is justified because IWOA has the ability to escape from being stuck in local optima. In addition, IWOA can improve the solution diversity by new CM equation. Also, the new LSA algorithm also improved the best solution of IWOA by setting or resetting included rules based on their recall and precision. Therefore, IWOA starts by random number of solutions, then it will iteratively improve these solutions. Lastly, IWOA will result in returning the best solution which includes the optimal combination of rules.

(3) Effect of Pruning: PA is an important step to further improve the extraction precision while maintain high recall results. This is clearly justified from the obtained results as shown in Tables 14 and 18. The proposed PA algorithm solved an important issue which was in previous studies which includes the pruning of infrequent aspects. For example, In the sentence "*it 's great to switch to* ***spot metering***" from [23] datasets, the aspect "***spot metering***" has a frequency of 1 in Canon dataset. In previous studies which is based on pruning out infrequent aspects, it will discard ***spot metering*** aspect and consider it as incorrect aspect. However, because PA works in three phases as outlined before it will approve this aspect as a correct aspect. This aspect will go into these three phases one by one by PA. At phase 2 in PA, it will check the manual and PA will approve this aspect as a correct aspect. Another example why the proposed algorithm outperformed other related works is the third phase of PA algorithm. For example, in the sentence "*and for those that are interested the* ***recharger*** *works anywhere in the world and is quite small*", this example has two nouns including "***recharger***" and "***world***". However, based on phase three of PA it will approve "***recharger***" as a correct aspect because it has direct opinion association with opinion word "***work***". These examples represent some forms of improvements over extraction performance by PA which resulted in achieving superior results in comparison to other baseline works.

To further discuss the results of IWOA+PA in comparison with baseline works the following discussion is outlined. An example of related works which used dependency relation rules for extraction are DP, RubE, and RSLS. As shown in the previous result tables, we observed that IWOA+PA results are remarkably superior to DP in terms of recall, precision, and F-measure. The outperformance of IWOA+PA over DP is justified because DP has a problem of error propagation which result in extracting many incorrect aspects and it is obvious from DP precision results in comparison to IWOA+PA. We can also see that DP miss many aspects as it does not cover all types of extraction rules, which is clearly observed from DP recall results. Furthermore, we observed that IWOA+PA achieved better results than RubE work. The main reasons for RubE poorer results than IWOA+PA are the same as DP reasons, because RubE represents an extension of DP. Also, the precision of both DP and RubE is lower than IWOA+PA. This is because no rules selection was applied in DP and RubE. Thus, the use of all rules regardless of their quality will extract many incorrect aspects. Also, RSLS represents another improvement of DP for extracting aspects. As shown from the results, RSLS results are obviously lower than IWOA+PA. There are number of reasons for RSLS poor results. One reason is that RSLS is based on DP which used propagation for extraction, and propagation will result in extracting many incorrect aspects, where this is obvious from DP precision results in comparison to IWOA+PA. In addition, no pruning was made in RSLS after extraction. These works DP, RubE, and RSLS considered opinion word as adjective only, which result in missing many correct aspects. Whereas, this problem is solved in IWOA+PA by using both adjective and opinion lexicon for opinion words.

More related works are TF-RBM, SPR, and Htay in which they employed patterns for aspect extraction. The results clearly indicate that IWOA+PA outperforms Htay baseline work. The key reasons for poor Htay results as they do not include all possible patterns rules for extraction. Thus, they missed many incorrect aspects which is clearly noticed from its poor recall in comparison to IWOA+PA. In addition, no pruning was applied, and this clearly confirmed from the its lower precision in comparison to IWOA+PA. Moreover, IWOA+PA outperforms SPR for the same reasons as Htay and this can be clearly noticed from lower results of SPR in comparison with IWOA+PA. In another example, IWOA+PA obtained superior results compared to TF-RBM based on the achieved results. One reasons why TF-RBM obtained lower

Table 17
Results comparisons including Precision (P), Recall (R), and F-measure to other approaches.

Data	Feng			MCNN			NGD+CNET			DomSent			IWOA+PA		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
D1	0.85	0.84	0.84	N/A	N/A	0.90	0.88	0.80	0.84	0.83	0.86	0.84	0.91	0.93	0.92
D2	0.89	0.87	0.88	N/A	N/A	0.87	0.81	0.85	0.83	0.89	0.87	0.88	0.91	0.92	0.91
D3	0.87	0.90	0.88	N/A	N/A	0.89	0.91	0.87	0.89	0.94	0.86	0.90	0.93	0.93	0.93
D4	0.88	0.87	0.87	N/A	N/A	0.89	0.90	0.87	0.88	0.83	0.92	0.87	0.93	0.95	0.94
D5	0.88	0.90	0.89	N/A	N/A	0.89	0.89	0.82	0.85	0.89	0.76	0.82	0.92	0.91	0.91
Avg	0.87	0.88	0.87	N/A	N/A	0.89	0.88	0.84	0.86	0.88	0.85	0.86	**0.92**	**0.93**	**0.92**

Table 18
Results comparisons over each phase of IWOA+PA including Precision (P), Recall (R), and F-measure over D6 and D7 datasets.

Data	All			IWOA			IWOA+PA		
	P	R	F	P	R	F	P	R	F
D6	0.72	0.96	0.84	0.88	0.94	0.91	0.96	0.88	0.92
D7	0.74	0.97	0.85	0.86	0.94	0.90	0.95	0.92	0.93
Avg	0.73	0.97	0.85	0.87	0.94	0.91	**0.96**	**0.90**	**0.93**

precision in comparison with IWOA+PA is the use of NGD for pruning. A disadvantage of NGD is that it considers any two words in the same page on the internet as related words, but this not always correct. These works TF-RBM, SPR, and Htay did not apply rules selection of the included rules. This is clearly noticed in the poor precision in comparison with IWOA+PA.

More works which IWOA+PA outperformed are represented by the following recent works. IWOA+PA gets better results than MCNN based on the achieved results. One key reason of IWOA+PA outperformance over MCNN was that no pruning was applied in MCNN, and this can be obviously noticed from MCNN results which is lower than IWOA+PA results. On the other hand, it is shown from the results that IWOA+PA clearly achieved superior results to DomSent. The main reasons why DomSent obtained poorer results than IWOA+PA include: (1) DomSent considered aspects as noun or noun phrase only, and opinion words based on their lexicon only. Therefore, this resulted in missing many correct aspects and this clearly noticed from lower DomSent recall in comparison with IWOA+PA; and (2) DomSent used NGD for pruning which has a problem as mentioned before and this clearly noticed from lower DomSent precision in comparison with IWOA+PA. Further, IWOA+PA achieved superior results to NGD+CNET. The poor results of NGD+CNET in comparison to IWOA+PA due to number of reasons. One reason that NGD+CNET is unable to extract multi-aspects in the sentence. Other reason that NGD+CNET considered aspects as noun or noun phrase only which results in miss many correct aspects. This is can be clearly confirmed from lower recall results of NGD+CNET in comparison to IWOA+PA. Also, NGD+CNET used NGD for pruning which has a problem as detailed before and this can be observed from poor precision results of NGD+CNET in comparison to IWOA+PA. Moreover, IWOA+PA outperformed Feng work, which can be noticed from the results. The major reasons for poorer Feng results include: (1) Feng used Topic model which has disadvantage of extracting general aspects and ignore infrequent aspects and this can be observed from its recall results in comparison to IWOA+PA; and (2) In Feng no pruning was used and this can be obviously noticed from Feng lower precision results in comparison to IWOA+PA results. Also, IWOA+PA obtained superior results to AER. The main reasons for AER poorer results than IWOA+PA are the same as DP reasons, because AER represents an improvement of DP. In addition, AER shares the same problems with DP and RubE such as no rules selection was conducted, and this clearly noticed from lower precision results of AER in comparison to IWOA+PA results. Lastly, IWOA+PA achieved better results than CRF. The main reasons for lower CRF performance including CRFs based approach is not suitable for long range patterns. Therefore, miss many aspects and this clearly noticed from lower recall results of CRF in comparison to IWOA+PA results. Another reason, in CRF no pruning was applied and this can be observed from poorer results of CRF precision in comparison to IWOA+PA results.

As outlined, IWOA+PA outperformed baseline works over all metrics (recall, precision, and F-measure). This outperformance of IWOA+PA is achieved based on a number of reasons including: (1) the combination different rules types which can be applied into formal, informal, or both types of reviews, which confirmed the ability of IWOA+PA to achieve higher recall than other works; (2) the restrictions which were added to the included rules to solve the issues in previous studies, which proved that these restrictions improved both recall and precision of IWOA+PA; (3) the new developed rules confirmed its ability to improve both precision and recall; (4) the rules selection by IWOA which shows an improvement over precision while maintaining a high value of recall; and (5) the use of the PA algorithm, which shows also an improvement on precision while maintaining a high value of recall. PA solved the issue of discarding infrequent aspects which were pruned by previous studies. Also, IWOA solved the issues of previous studies which include the use of any extraction rules regardless of their quality. Figs. 12 and 13 show a comparison between IWOA+PA and baseline works over Dataset1 and Dataset2 in terms of F-measure. The outperformance of IWOA+PA is clear over all other works. This outperformance is justified as detailed before.

Based on IWOA+PA results and comparison with state-of-the-art and most recent works, it is clearly noticeable that the ability of IWOA+PA to be outperform all other methods. These results confirmed IWOA+PA ability to perform consistently overall performance metrics. One of the main reasons for achieving better results by IWOA includes its ability to select the optimal combinations of rules. In addition, it is ability to balance between performance metrics by selecting the rules which can improve recall while it also improves precision. The improvement in recall resulted from the combination of different rules types. In addition, the improvement in precision resulted from the use of IWOA based on the introduced improvement. IWOA will select the best rules while it maximizes the performance. Also, another improvement on precision resulted from the use of PA algorithm, where PA purifies the candidate aspects and keeps the correct aspects only.

The advantages of the proposed IWOA+PA aspect extraction algorithm include the following: IWOA+PA can be applied on any type of reviews which include formal text, informal texts, or both types. In addition, IWOA can select the best subset of rules from the full set of rules. Lastly, PA can improve the extraction performance by pruning the incorrect aspects while retaining the correct aspects.

8. Conclusion

Previous studies which in common use either dependency relations or syntactic pattern rules for aspect extraction still face

Table 19
IWOA results comparison with other optimization algorithms over D6 and D7 datasets.

Test data	FFA			SCA			SSA			PSO			IWOA			Training data
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	
D6	0.81	0.90	0.85	0.80	0.91	0.85	0.79	0.93	0.86	0.81	0.92	0.86	**0.88**	**0.94**	**0.91**	**D7**
D7	0.82	0.91	0.86	0.81	0.88	0.84	0.80	0.91	0.85	0.84	0.90	0.87	**0.86**	**0.94**	**0.90**	**D6**
Avg	0.82	0.91	0.86	0.81	0.90	0.85	0.80	0.92	0.86	0.83	0.91	0.87	**0.87**	**0.94**	**0.91**	

Table 20
IWOA results comparison with other optimization algorithms over D6 and D7 datasets.

Test data	MFO			MVO			GWO			WOA			IWOA			Training data
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	
D6	0.80	0.91	0.85	0.78	0.92	0.85	0.81	0.91	0.86	0.81	0.92	0.86	0.88	0.94	0.91	D7
D7	0.82	0.91	0.86	0.82	0.92	0.86	0.81	0.90	0.85	0.82	0.92	0.87	0.86	0.94	0.90	D6
Avg	0.81	0.91	0.86	0.80	0.92	0.86	0.81	0.91	0.86	0.82	0.92	0.87	0.87	**0.94**	**0.91**	

Table 21
Results comparisons to other baseline approaches over D6 and D7 datasets.

Data	DP			RSLS			CRF			AER			IWOA+PA		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
D6	0.74	0.89	0.81	0.86	0.84	0.85	0.75	0.67	0.71	0.87	0.80	0.83	0.96	0.88	0.92
D7	0.71	0.91	0.80	0.83	0.83	0.83	0.85	0.74	0.79	0.81	0.84	0.82	0.95	0.92	0.93
Avg	0.73	0.90	0.81	0.85	0.84	0.84	0.80	0.71	0.75	0.84	0.82	0.83	**0.96**	**0.90**	**0.93**

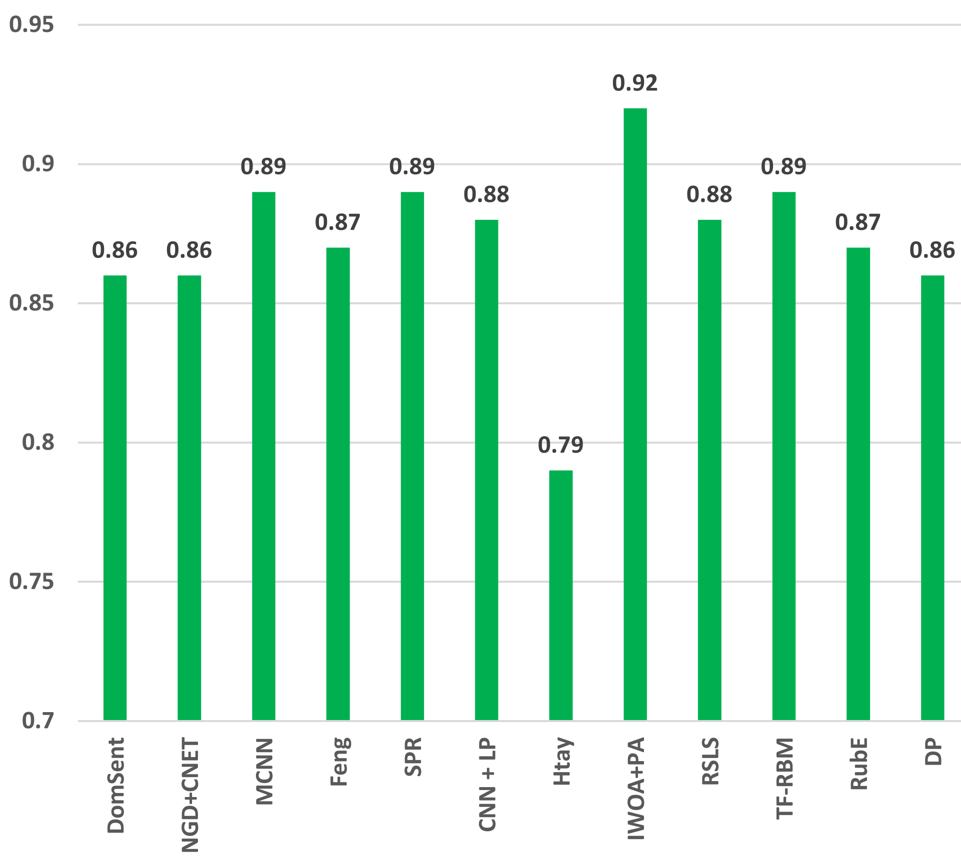


Fig. 12. IWOA+PA F-measure comparison with other baseline works using Dataset1.

some limitations. Therefore, to address these limitations we have proposed the IWOA+PA aspect extraction algorithm, which can identify explicit aspects from formal, informal texts, or both types of texts. Additionally, IWOA can select the optimal subset of rules from the set of original rules, while PA can remove incorrectly extracted aspects and retains the correct aspects. To fulfill the required tasks of IWOA+PA, we proposed to comprise 126 rules including dependency relation and sequential patterns rules from previous studies and the newly developed rules. The new rules were developed to overcome the shortcomings of previous studies. In addition, a combination of different rules types is proposed to cover all review types, including both formal and informal texts.

However, not all included rules in the full set of rules are of equal performance in aspect extraction. Hence, the subset which represents the best rules from the full set of 126 rules must

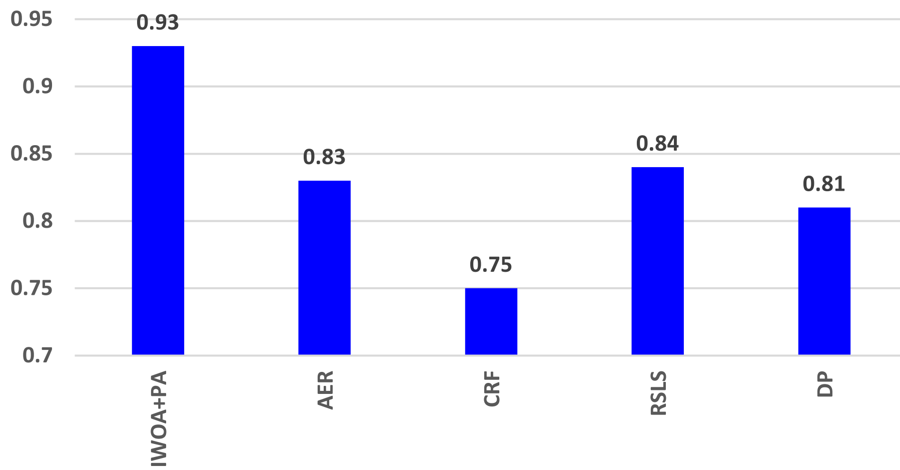


Fig. 13. IWOA+PA F-measure comparison with other baseline works using Dataset2.

be selected carefully. Therefore, we improved WOA algorithm to solve its problems and make it fit for rules selection problem and called it as IWOA. IWOA was used to select the optimal subset combination from these rules and discard low quality rules. In this study, IWOA represents an improved variant of WOA. Two main improvements were included into original WOA to solve its problems and make it suitable for rules selection problem. The two improvements were introduced into WOA includes the use of Cauchy mutation and LSA algorithm. The first improvement includes the development of update position equation based on using Cauchy mutation to improve the solutions diversity, while the second improvement includes the development of new LSA algorithm to prevent WOA from falling into local optima. To further improve extraction performance, PA was proposed and used after the application of IWOA. Several experiments were conducted in this study by comparing IWOA+PA with the state-of-the-art aspect extraction works and the most recent aspect extraction works. In addition, seven benchmark datasets from different domain types were used in the experiments. The results from the conducted experiments revealed that IWOA+PA outperformed all other state-of-the-art works and most recent works in terms of recall, precision, and F-measure.

The key advantages of the proposed IWOA+PA aspect extraction to sentiment analysis research community is that it can be applied to any type of reviews including formal or informal or both types. In addition, IWOA can be applied to any type of rule on any language, not only English. Therefore, just feed IWOA with the aspect extraction rules on a given language, then it will select the best subset of rules.

In future work, we plan to develop an algorithm for implicit aspect extraction and to apply IWOA+PA to other data domains. Another possible direction is to define extraction rules for other languages such as Arabic and Malay and apply IWOA+PA for extracting aspects in these languages.

CRediT authorship contribution statement

Mohammad Tubishat: Conceptualization, Data curation, Resources, Investigation, Methodology, Writing - original draft, Implementation, Software, Validation, Writing - review & editing. **Norisma Idris:** Funding acquisition, Project administration, Resources, Supervision, Writing - review & editing. **Mohammad Abushariah:** Validation, Supervision, Writing - review & editing.

Declaration of competing interest

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

Acknowledgment

This research work is supported by the University of Malaya, Malaysia Grant – Faculty Program (Project No: GPF007D-2018).

References

[1] Y. Liu, J.-W. Bi, Z.-P. Fan, Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory, Inf. Fusion 36 (2017) 149–161.

[2] D. Law, R. Gruss, A.S. Abrahams, Automated defect discovery for dishwasher appliances from online consumer reviews, Expert Syst. Appl. 67 (2017) 84–94.

[3] K. Akalamkam, J.K. Mitra, Consumer pre-purchase search in online shopping: Role of offline and online information sources, Bus. Perspect. Res. 6 (1) (2018) 42–60.

[4] K. Zhao, A.C. Stylianou, Y. Zheng, Sources and impacts of social influence from online anonymous user reviews, Inf. Manag. 55 (1) (2018) 16–30.

[5] E. Cambria, D. Das, S. Bandyopadhyay, A. Feraco, A Practical Guide to Sentiment Analysis, Springer, 2017.

[6] A. Hussain, E. Cambria, Semi-supervised learning for big social data analysis, Neurocomputing 275 (2018) 1662–1673.

[7] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107.

[8] S. Poria, E. Cambria, R. Bajpai, A. Hussain, A review of affective computing: From unimodal analysis to multimodal fusion, Inf. Fusion 37 (2017) 98–125.

[9] S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, L.-P. Morency, Context-dependent sentiment analysis in user-generated videos, in: Paper Presented at the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017.

[10] I. Chaturvedi, E. Cambria, R.E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis: Survey and challenges, Inf. Fusion 44 (2018) 65–77.

[11] E. Cambria, S. Poria, A. Gelbukh, M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intell. Syst. 32 (6) (2017) 74–80.

[12] Y. Xia, E. Cambria, A. Hussain, H. Zhao, Word polarity disambiguation using bayesian model and opinion-level features, Cogn. Comput. 7 (3) (2015) 369–380.

[13] E. Cambria, S. Poria, R. Bajpai, B. Schuller, Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives, in: Paper Presented at the Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016.

[14] S. Poria, E. Cambria, D. Hazarika, P. Vij, A deeper look into sarcastic tweets using deep convolutional neural networks, 2016, arXiv preprint arXiv:1610.08815.

[15] Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in: Paper Presented at the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[16] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, E. Cambria, Bayesian network based extreme learning machine for subjectivity detection, J. Frankl. Inst. 355 (4) (2018) 1780–1797.

[17] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, M. Huang, Augmenting end-to-end dialogue systems with commonsense knowledge, in: Paper Presented at the Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[18] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, J. Munro, Sentic computing for patient centered applications, in: Paper Presented at the IEEE 10th International Conference on Signal Processing Proceedings, 2010.

[19] S. Cavallari, V.W. Zheng, H. Cai, K.C.-C. Chang, E. Cambria, Learning community embedding with community detection and node embedding on graphs, in: Paper Presented at the Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017.

[20] F.Z. Xing, E. Cambria, R.E. Welsch, Natural language based financial forecasting: a survey, Artif. Intell. Rev. 50 (1) (2018) 49–73.

[21] X. Chi, T.P. Siew, E. Cambria, Adaptive two-stage feature selection for sentiment classification, in: Paper Presented at the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2017.

[22] T.A. Rana, Y.-N. Cheah, Aspect extraction in sentiment analysis: comparative analysis and survey, Artif. Intell. Rev. 46 (4) (2016) 459–483.

[23] M. Hu, B. Liu, Mining and summarizing customer reviews, in: Paper Presented at the Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004.

[24] M. Tubishat, N. Idris, M.A. Abushariah, Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open challenges, Inf. Process. Manage. 54 (4) (2018) 545–563.

[25] E. Marrese-Taylor, J.D. Velásquez, F. Bravo-Marquez, A novel deterministic approach for aspect-based opinion mining in tourism products reviews, Expert Syst. Appl. 41 (17) (2014) 7764–7775.

[26] S. Poria, E. Cambria, A. Gelbukh, Aspect extraction for opinion mining with a deep convolutional neural network, Knowl.-Based Syst. 108 (2016) 42–49.

[27] M.S. Akhtar, D. Gupta, A. Ekbal, P. Bhattacharyya, Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis, Knowl.-Based Syst. 125 (2017) 116–135.

[28] J. Yu, Z.-J. Zha, M. Wang, T.-S. Chua, Aspect ranking: identifying important product aspects from online consumer reviews, in: Paper Presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011.

[29] W. Xue, T. Li, N. Rishe, Aspect identification and ratings inference for hotel reviews, World Wide Web 20 (1) (2017) 23–37.

[30] Q. Zhou, C. Zhang, Detecting dietary preference of social media users in China via sentiment analysis, Proc. Assoc. Inf. Sci. Technol. 54 (1) (2017) 523–527.

[31] A. Caputo, P. Basile, M. de Gemmis, P. Lops, G. Semeraro, G. Rossiello, SABRE: A Sentiment Aspect-Based Retrieval Engine Information Filtering and Retrieval, Springer, 2017, pp. 63–78.

[32] A.-M. Popescu, O. Etzioni, Extracting product features and opinions from reviews, in: Natural Language Processing and Text Mining, Springer, 2007, pp. 9–28.

[33] S. Moghaddam, M. Ester, Opinion digger: an unsupervised opinion miner from unstructured product reviews, in: Paper Presented at the Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010.

[34] S.S. Htay, K.T. Lynn, Extracting product features and opinion words using pattern knowledge in customer reviews, Sci. World J. (2013).

[35] Z. Hai, K. Chang, J.-J. Kim, C.C. Yang, Identifying features in opinion mining via intrinsic and extrinsic domain relevance, IEEE Trans. Knowl. Data Eng. 26 (3) (2014) 623–634.

[36] S. Poria, E. Cambria, L.-W. Ku, C. Gui, A. Gelbukh, A rule-based approach to aspect extraction from product reviews, in: Paper Presented at the Proceedings of the Second Workshop on Natural Language Processing for Social Media (SocialNLP), 2014.

[37] Wu, Q. Zhang, X. Huang, L. Wu, Phrase dependency parsing for opinion mining, in: Paper Presented at the Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3, 2009.

[38] C.-P. Wei, Y.-M. Chen, C.-S. Yang, C.C. Yang, Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews, Inf. Syst. E-Bus. Manag. 8 (2) (2010) 149–167.

[39] G. Qiu, B. Liu, J. Bu, C. Chen, Opinion word expansion and target extraction through double propagation, Comput. linguist. 37 (1) (2011) 9–27.

[40] K. Liu, L. Xu, J. Zhao, Syntactic patterns versus word alignment: Extracting opinion targets from online reviews, in: Paper Presented at the Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013.

[41] L. Xu, K. Liu, S. Lai, Y. Chen, J. Zhao, Mining opinion words and opinion targets in a two-stage framework, in: Paper Presented at the Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2013.

[42] A.K. Samha, Y. Li, J. Zhang, Aspect-based opinion extraction from customer reviews, 2014, arXiv preprint arXiv:1404.1982.

[43] Z. Yan, M. Xing, D. Zhang, B. Ma, EXPRS: An extended pagerank method for product feature extraction from online consumer reviews, Inf. Manag. 52 (7) (2015) 850–858.

[44] W. Jin, H.H. Ho, R.K. Srihari, A novel lexicalized HMM-based learning framework for web opinion mining, in: Paper Presented at the Proceedings of the 26th Annual International Conference on Machine Learning, 2009.

[45] Y. Choi, C. Cardie, Hierarchical sequential learning for extracting opinions and their attributes, in: Paper Presented at the Proceedings of the ACL 2010 Conference Short Papers, 2010.

[46] P. Jiang, C. Zhang, H. Fu, Z. Niu, Q. Yang, An approach based on tree kernels for opinion mining of online product reviews, in: Paper Presented at the Data Mining (ICDM), 2010 IEEE 10th International Conference on, 2010.

[47] Chen, L. Qi, F. Wang, Comparison of feature-level learning methods for mining online consumer reviews, Expert Syst. Appl. 39 (10) (2012) 9588–9601.

[48] S. Li, R. Wang, G. Zhou, Opinion target extraction using a shallow semantic parsing framework, in: Paper Presented at the Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012.

[49] F.L. Cruz, J.A. Troyano, F. Enríquez, F.J. Ortega, C.G. Vallejo, 'Long autonomy or long delay?'the importance of domain in opinion mining, Expert Syst. Appl. 40 (8) (2013) 3174–3184.

[50] V.R. Kumar, K. Raghuveer, Dependency driven semantic approach to product features extraction and summarization using customer reviews, in: Advances in Computing and Information Technology, Springer, 2013, pp. 225–238.

[51] Liu, Z. Gao, B. Liu, Y. Zhang, Automated rule selection for opinion target extraction, Knowl.-Based Syst. 104 (2016) 74–88.

[52] Y. Kang, L. Zhou, Rube: Rule-based methods for extracting product features from online consumer reviews, Inf. Manag. 54 (2) (2017) 166–176.

[53] W. Maharani, D.H. Widyantoro, M.L. Khodra, Aspect extraction in customer reviews using syntactic pattern, Procedia Comput. Sci. 59 (2015) 244–253.

[54] M.Z. Asghar, A. Khan, S.R. Zahra, S. Ahmad, F.M. Kundi, Aspect-based opinion mining framework using heuristic patterns, Cluster Comput. (2017) 1–19.

[55] T.A. Rana, Y.-N. Cheah, A two-fold rule-based model for aspect extraction, Expert Syst. Appl. 89 (2017) 273–285.

[56] Zhang, B. Liu, S.H. Lim, E. O'Brien-Strain, Extracting and ranking product features in opinion documents, in: Paper Presented at the Proceedings of the 23rd International Conference on Computational Linguistics: Posters, 2010.

[57] K. Liu, H.L. Xu, Y. Liu, J. Zhao, Opinion target extraction using partially-supervised word alignment model, in: Paper Presented at the IJCAI, 2013.

[58] Liu, L. Xu, J. Zhao, Co-extracting opinion targets and opinion words from online reviews based on the word alignment model, IEEE Trans. Knowl. Data Eng. 27 (3) (2015) 636–650.

[59] M. Eirinaki, S. Pisal, J. Singh, Feature-based opinion mining and ranking, J. Comput. Syst. Sci. 78 (4) (2012) 1175–1184.

[60] L. Zhuang, F. Jing, X.-Y. Zhu, Movie review mining and summarization, in: Paper Presented at the Proceedings of the 15th ACM International Conference on Information and Knowledge Management, 2006.

[61] C. Wu, F. Wu, S. Wu, Z. Yuan, Y. Huang, A hybrid unsupervised method for aspect term and opinion target extraction, Knowl.-Based Syst. 148 (2018) 66–73.

[62] Z. Luo, S. Huang, K.Q. Zhu, Knowledge empowered prominent aspect extraction from product reviews, Inf. Process. Manage. 56 (3) (2019) 408–423.

[63] M. Dragoni, M. Federici, A. Rexha, An unsupervised aspect extraction strategy for monitoring real-time reviews stream, Inf. Process. Manage. 56 (3) (2019) 1103–1118.

[64] Li, Z. Qin, W. Xu, J. Guo, A holistic model of mining product aspects and associated sentiments from online reviews, Multimedia Tools Appl. 74 (23) (2015) 10177–10194.

[65] A.K. Samha, Aspect-based opinion mining using dependency relations, Int. J. Comput. Sci. Trends Technol.(IJCST) 4 (2016).

[66] B. Agarwal, S. Poria, N. Mittal, A. Gelbukh, A. Hussain, Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach, Cogn. Comput. 7 (4) (2015) 487–499.

[67] T. Chinsha, S. Joseph, A syntactic approach for aspect based opinion mining, in: Paper Presented at the 2015 IEEE International Conference on Semantic Computing (ICSC), 2015.

[68] Z. Li, M. Zhang, S. Ma, B. Zhou, Y. Sun, Automatic extraction for product feature words from comments on the web, in: Paper Presented at the Asia Information Retrieval Symposium, 2009.

[69] A. Bagheri, M. Saraee, F. de Jong, An unsupervised aspect detection model for sentiment analysis of reviews, in: Paper Presented at the International Conference on Application of Natural Language To Information Systems, 2013.
[70] P.D. Turney, Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: Paper Presented at the Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2002.
[71] X. Ding, B. Liu, P.S. Yu, A holistic lexicon-based approach to opinion mining, in: Paper Presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008.
[72] S. Brody, N. Elhadad, An unsupervised aspect-sentiment model for online reviews, in: Paper Presented at the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.
[73] T.A. Rana, Y.-N. Cheah, S. Letchmunan, Topic modeling in sentiment analysis: a systematic review, J. ICT Res. Appl. 10 (1) (2016) 76–93.
[74] S. Moghaddam, M. Ester, ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews, in: Paper Presented at the Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2011.
[75] Z. Madhoushi, A.R. Hamdan, S. Zainudin, Aspect-based sentiment analysis methods in recent years, Asia-Pacific Journal of Information Technology and Multimedia 8 (1) (2019).
[76] Chen, A. Mukherjee, B. Liu, Aspect extraction with automated prior knowledge learning, in: Paper Presented at the Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014.
[77] A. García-Pablos, M. Cuadros, G. Rigau, W2vlda: almost unsupervised system for aspect based sentiment analysis, Expert Syst. Appl. 91 (2018) 127–137.
[78] E. Ekinci, S. İlhan Omurca, Concept-LDA: Incorporating babelfy into LDA for aspect extraction, J. Inf. Sci. 46 (3) (2020) 406–418.
[79] G.S. Chauhan, Y.K. Meena, D. Gopalani, R. Nahta, A two-step hybrid unsupervised model with attention mechanism for aspect extraction, Expert Syst. Appl. (2020) 113673.
[80] T. Sokhin, M. Khodorchenko, N. Butakov, Unsupervised neural aspect search with related terms extraction, 2020, arXiv preprint arXiv:2005.02771.
[81] T. Liang, W. Wang, F. Lv, Weakly-supervised domain adaption for aspect extraction via multi-level interaction transfer, 2020, arXiv preprint arXiv:2006.09235.
[82] N. Jindal, B. Liu, Opinion spam and analysis, in: Paper Presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, 2008.
[83] G. Qiu, B. Liu, J. Bu, C. Chen, Expanding domain sentiment lexicon through double propagation, in: Paper Presented at the IJCAI, 2009.
[84] Z. Hai, K. Chang, G. Cong, One seed to find them all: mining opinion features via association, in: Paper presented at the Proceedings of the 21st ACM international conference on Information and knowledge management, 2012.
[85] Liu, B. Liu, Y. Zhang, D.S. Kim, Z. Gao, Improving opinion aspect extraction using semantic similarity and aspect associations, in: Paper Presented at the AAAI, 2016.
[86] Liu, Z. Gao, B. Liu, Y. Zhang, Automated rule selection for aspect extraction in opinion mining, in: Paper Presented at the IJCAI, 2015.
[87] Y. Wang, Z. Wang, D. Zhang, R. Zhang, Discovering cultural differences in online consumer product reviews, J. Electron. Commer. Res. 20 (3) (2019) 169–183.
[88] T.A. Rana, Y.-N. Cheah, Sequential patterns rule-based approach for opinion target extraction from customer reviews, J. Inf. Sci. 45 (5) (2019) 643–655.
[89] J. Feng, W. Yang, C. Gong, X. Li, R. Bo, Product feature extraction via topic model and synonym recognition approach, in: Paper presented at the CCF Conference on Big Data, 2019.
[90] R. Agerri, G. Rigau, Language independent sequence labelling for opinion target extraction, Artif. Intell. 268 (2019) 85–95.
[91] G.S. Chauhan, Y.K. Meena, DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis Smart Systems and IoT: Innovations in Computing, Springer, 2020, pp. 103–109.
[92] S.-M. Park, S.J. Lee, B.-W. On, Topic word embedding-based methods for automatically extracting main aspects from product reviews, Appl. Sci. 10 (11) (2020) 3831.
[93] X. Li, W. Lam, Deep multi-task learning for aspect term extraction with memory interaction, in: Paper Presented at the Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017.
[94] D. Ma, S. Li, F. Wu, X. Xie, H. Wang, Exploring sequence-to-sequence learning in aspect term extraction, in: Paper Presented at the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019.
[95] H. Xu, B. Liu, L. Shu, P.S. Yu, Double embeddings and CNN-based sequence labeling for aspect extraction, 2018, arXiv preprint arXiv:1805.04601.
[96] A. Da'u, N. Salim, I. Rabiu, A. Osman, Weighted aspect-based opinion mining using deep learning for recommender system, Expert Syst. Appl. 140 (2020) 112871.
[97] L. Shu, H. Xu, B. Liu, Controlled CNN-based sequence labeling for aspect extraction, 2019, arXiv preprint arXiv:1905.06407.
[98] P. Barnaghi, G. Kontonatsios, N. Bessis, Y. Korkontzelos, Aspect extraction from reviews using convolutional neural networks and embeddings, in: Paper Presented at the International Conference on Applications of Natural Language To Information Systems, 2019.
[99] A. Kumar, S. Verma, A. Sharan, ATE-SPD: simultaneous extraction of aspect-term and aspect sentiment polarity using bi-LSTM-CRF neural network, J. Exp. Theor. Artif. Intell. (2020) 1–22.
[100] M.S. Akhtar, T. Garg, A. Ekbal, Multi-task learning for aspect term extraction and aspect sentiment classification, Neurocomputing (2020).
[101] Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, H. Yu, Structure-aware review mining and summarization, in: Paper Presented at the Proceedings of the 23rd International Conference on Computational Linguistics, 2010.
[102] N. Jakob, I. Gurevych, Extracting opinion targets in a single-and cross-domain setting with conditional random fields, in: Paper Presented at the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010.
[103] S. Huang, X. Liu, X. Peng, Z. Niu, Fine-grained product features extraction and categorization in reviews opinion mining, in: Paper Presented at the Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on, 2012.
[104] Y. Xiang, H. He, J. Zheng, Aspect term extraction based on MFE-crf, Information 9 (8) (2018) 198.
[105] H. Xiong, H. Yan, Z. Zeng, B. Wang, Dependency parsing and bidirectional LSTM-CRF for aspect-level sentiment analysis of chinese, in: Paper Presented at the JIST (Workshops & Posters), 2018.
[106] P. Ray, A. Chakrabarti, A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis, Appl. Comput. Inform. (2019).
[107] A. Nawaz, S. Asghar, S.H.A. Naqvi, A segregational approach for determining aspect sentiments in social media analysis, J. Supercomput. 75 (5) (2019) 2584–2602.
[108] T.A. Rana, Y.-N. Cheah, Sequential patterns-based rules for aspect-based sentiment analysis, Adv. Sci. Lett. 24 (2) (2018) 1370–1374.
[109] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[110] H. Zhao, S. Guo, H. Zhao, Energy-related co2 emissions forecasting using an improved LSSVM model optimized by whale optimization algorithm, Energies 10 (7) (2017) 874.
[111] Y. Chen, R. Vepa, M.H. Shaheed, Enhanced and speedy energy extraction from a scaled-up pressure retarded osmosis process with a whale optimization based maximum power point tracking, Energy 153 (2018) 618–627.
[112] A. Saha, L.C. Saikia, Performance analysis of combination of ultra-capacitor and superconducting magnetic energy storage in a thermal-gas AGC system with utilization of whale optimization algorithm optimized cascade controller, J. Renew. Sustain. Energy 10 (1) (2018) 014103.
[113] M.M. Ahmed, E.H. Houssein, A.E. Hassanien, A. Taha, E. Hassanien, Maximizing lifetime of wireless sensor networks based on whale optimization algorithm, in: Paper Presented at the International Conference on Advanced Intelligent Systems and Informatics, 2017.
[114] A.R. Jadhav, T. Shankar, Whale optimization based energy-efficient cluster head selection algorithm for wireless sensor networks, 2017, arXiv preprint arXiv:1711.09389.
[115] D. Parambanchary, V.M. Rao, WOA-NN: a decision algorithm for vertical handover in heterogeneous networks, Wirel. Netw. (2018) 1–16.
[116] G.I. Sayed, A. Darwish, A.E. Hassanien, J.-S. Pan, Breast cancer diagnosis approach based on meta-heuristic optimization algorithm inspired by the bubble-net hunting strategy of whales, in: Paper Presented at the International Conference on Genetic and Evolutionary Computing, 2016.
[117] H. Zamani, M.-H. Nadimi-Shahraki, Feature selection based on whale optimization algorithm for diseases diagnosis, Int. J. Comput. Sci. Inf. Secur. 14 (9) (2016) 1243.
[118] M. Canayaz, M. Demir, Eature selection with the whale optimization algorithm and artificial neural network, in: Paper Presented at the 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017.
[119] M. Sharawi, H.M. Zawbaa, E. Emary, Feature selection approach based on whale optimization algorithm, in: Paper Presented at the 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI), 2017.
[120] H.F. Eid, Binary whale optimisation: an effective swarm algorithm for feature selection, Int. J. Metaheuristics 7 (1) (2018) 67–79.

[121] A.E. Hegazy, M. Makhlouf, G.S. El-Tawel, Dimensionality reduction using an improved whale optimization algorithm for data classification, Int. J. Mod. Educ. Comput. Sci. 10 (7) (2018) 37.

[122] M. Mafarja, S. Mirjalili, Whale optimization approaches for wrapper feature selection, Appl. Soft Comput. 62 (2018) 441–453.

[123] M. Tubishat, M.A. Abushariah, N. Idris, I. Aljarah, Improved whale optimization algorithm for feature selection in arabic sentiment analysis, Appl. Intell. (2018) 1–20.

[124] A.G. Hussien, A.E. Hassanien, E.H. Houssein, S. Bhattacharyya, M. Amin, S-shaped binary whale optimization algorithm for feature selection, in: Recent Trends in Signal and Image Processing, Springer, 2019, pp. 79–87.

[125] M.A. El Aziz, A.A. Ewees, A.E. Hassanien, Whale optimization algorithm and moth-flame optimization for multilevel thresholding image segmentation, Expert Syst. Appl. 83 (2017) 242–256.

[126] A. Mostafa, A.E. Hassanien, M. Houseni, H. Hefny, Liver segmentation in MRI images based on whale optimization algorithm, Multimedia Tools Appl. 76 (23) (2017) 24931–24954.

[127] M.A. El Aziz, A.A. Ewees, A.E. Hassanien, Multi-objective whale optimization algorithm for content-based image retrieval, Multimedia Tools Appl. 77 (19) (2018) 26135–26172.

[128] G. Hassan, A.E. Hassanien, Retinal fundus vasculature multilevel segmentation using whale optimization algorithm, Signal Image Video Process. 12 (2) (2018) 263–270.

[129] I. Aljarah, H. Faris, S. Mirjalili, Optimizing connection weights in neural networks using the whale optimization algorithm, Soft Comput. 22 (1) (2018) 1–15.

[130] A.-Z. Ala'M, H. Faris, M.A. Hassonah, Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts, Knowl.-Based Syst. 153 (2018) 91–104.

[131] J. Nasiri, F.M. Khiyabani, A whale optimization algorithm (WOA) approach for clustering, Cogent Math. Statist. 5 (1) (2018) 1483565.

[132] R. Bhesdadiya, P. Jangir, N. Jangir, I.N. Trivedi, D. Ladumor, Training multi-layer perceptron in neural network using whale optimization algorithm, Indian J. Sci. Technol. 9 (19) (2016) 28–36.

[133] A. Tharwat, Y.S. Moemen, A.E. Hassanien, Classification of toxicity effects of biotransformed hepatic drugs using whale optimized support vector machines, J. Biomed. Inform. 68 (2017) 132–149.

[134] N.P. Karlekar, N. Gomathi, OW-SVM: Ontology and whale optimization-based support vector machine for privacy-preserved medical data classification in cloud, Int. J. Commun. Syst. 31 (12) (2018) e3700.

[135] P.R. Sahu, P.K. Hota, S. Panda, Power system stability enhancement by fractional order multi input SSSC based controller employing whale optimization algorithm, J. Electr. Syst. Inf. Technol. 5 (3) (2018) 326–336.

[136] A. Kumar, V. Bhalla, P. Kumar, T. Bhardwaj, N. Jangir, Whale Optimization Algorithm for Constrained Economic Load Dispatch Problems—a Cost Optimization Ambient Communications and Computer Systems, Springer, 2018, pp. 353–366.

[137] A. Alzaqebah, R. Masadeh, A. Hudaib, Whale optimization algorithm for requirements prioritization, in: Paper Presented at the 2018 9th International Conference on Information and Communication Systems (ICICS), 2018.

[138] J. Ghahremani-Nahr, R. Kian, E. Sabet, A robust fuzzy mathematical programming model for the closed-loop supply chain network design and a whale optimization solution algorithm, Expert Syst. Appl. 116 (2019) 454–471.

[139] Y. Moodi, S.R. Mousavi, A. Ghavidel, M.R. Sohrabi, M. Rashki, Using response surface methodology and providing a modified model using whale algorithm for estimating the compressive strength of columns confined with FRP sheets, Constr. Build. Mater. 183 (2018) 163–170.

[140] A. Mukherjee, N. Chakraborty, B.K. Das, Whale optimization algorithm: An implementation to design low-pass FIR filter, in: Paper Presented at the 2017 Innovations in Power and Advanced Computing Technologies (I-PACT), 2017.

[141] Y. Miao, M. Zhao, V. Makis, J. Lin, Optimal swarm decomposition with whale optimization algorithm for weak feature extraction from multi-component modulation signal, Mech. Syst. Signal Process. 122 (2019) 673–691.

[142] L. Sai, F. Huajing, A WOA-based algorithm for parameter optimization of support vector regression and its application to condition prognostics, in: Paper Presented at the 2017 36th Chinese Control Conference (CCC), 2017.

[143] P. Yuan, C. Guo, J. Ding, Y. Qu, Synthesis of nonuniform sparse linear array antenna using whale optimization algorithm, in: Paper Presented at the 2017 Sixth Asia-Pacific Conference on Antennas and Propagation (APCAP), 2017.

[144] S. Osama, A. Darwish, E.H. Houssein, A.E. Hassanien, A.A. Fahmy, A. Mahrous, Long-term wind speed prediction based on optimized support vector regression, in: Paper Presented at the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017.

[145] R. Barham, I. Aljarah, Link prediction based on whale optimization algorithm, in: Paper Presented at the 2017 International Conference on New Trends in Computing Sciences (ICTCS), 2017.

[146] T.-K. Dao, T.-S. Pan, J.-S. Pan, A multi-objective optimal mobile robot path planning based on whale optimization algorithm, in: Paper Presented at the 2016 IEEE 13th International Conference on Signal Processing (ICSP), 2016.

[147] L.-L. Li, J. Sun, M.-L. Tseng, Z.-G. Li, Extreme learning machine optimized by whale optimization algorithm using insulated gate bipolar transistor module aging degree evaluation, Expert Syst. Appl. (2019).

[148] F. Luan, Z. Cai, S. Wu, T. Jiang, F. Li, J. Yang, Improved whale algorithm for solving the flexible job shop scheduling problem, Mathematics 7 (5) (2019) 384.

[149] M. Azizi, R.G. Ejlali, S.A.M. Ghasemi, S. Talatahari, Upgraded whale optimization algorithm for fuzzy logic based vibration control of nonlinear steel structure, Eng. Struct. 192 (2019) 53–70.

[150] Q. Zhang, H. Chen, A.A. Heidari, X. Zhao, Y. Xu, P. Wang, …, C. Li, Chaos-induced and mutation-driven schemes boosting salp chains-inspired optimizers, IEEE Access 7 (2019) 31243–31261.

[151] Wu, R. Law, Cauchy mutation based on objective variable of Gaussian particle swarm optimization for parameters selection of SVM, Expert Syst. Appl. 38 (6) (2011) 6405–6411.

[152] M. Ali, M. Pant, Improving the performance of differential evolution algorithm using Cauchy mutation, Soft Comput. 15 (5) (2011) 991–1007.

[153] Wang, W. Wang, H. Sun, S. Rahnamayan, Firefly algorithm with random attraction, Int. J. Bio-Inspired Comput. 8 (1) (2016) 33–41.

[154] Wang, S. Deb, A.H. Gandomi, A.H. Alavi, Opposition-based krill herd algorithm with Cauchy mutation and position clamping, Neurocomputing 177 (2016) 147–157.

[155] Y. Zou, P.X. Liu, C. Yang, C. Li, Q. Cheng, Collision detection for virtual environment using particle swarm optimization with adaptive cauchy mutation, Cluster Comput. 20 (2) (2017) 1765–1774.

[156] Li, N. Zhang, X. Lai, J. Zhou, Y. Xu, Design of a fractional-order PID controller for a pumped storage unit using a gravitational search algorithm based on the Cauchy and Gaussian mutation, Inform. Sci. 396 (2017) 162–181.

[157] L. Pappula, D. Ghosh, Synthesis of linear aperiodic array using Cauchy mutated cat swarm optimization, AEU-Int. J. Electron. Commun. 72 (2017) 52–64.

[158] I.-S. Oh, J.-S. Lee, B.-R. Moon, Hybrid genetic algorithms for feature selection, IEEE Trans. Pattern Anal. Mach. Intell. 26 (11) (2004) 1424–1437.

[159] L. Asadzadeh, A local search genetic algorithm for the job shop scheduling problem with intelligent agents, Comput. Ind. Eng. 85 (2015) 376–383.

[160] M.D. Toksari, A hybrid algorithm of ant colony optimization (ACO) and iterated local search (ILS) for estimating electricity domestic consumption: Case of Turkey, Int. J. Electr. Power Energy Syst. 78 (2016) 776–782.

[161] T.-C. Ou, W.-F. Su, X.-Z. Liu, S.-J. Huang, T.-Y. Tai, A modified bird-mating optimization with hill-climbing for connection decisions of transformers, Energies 9 (9) (2016) 671.

[162] P. Moradi, M. Gholampour, A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy, Appl. Soft Comput. 43 (2016) 117–130.

[163] M. Nekkaa, D. Boughaci, Hybrid harmony search combined with stochastic local search for feature selection, Neural Process. Lett. 44 (1) (2016) 199–220.

[164] M. Mavrovouniotis, F.M. Müller, S. Yang, Ant colony optimization with local search for dynamic traveling salesman problems, IEEE Trans. Cybern. 47 (7) (2017) 1743–1756.

[165] M.M. Mafarja, S. Mirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection, Neurocomputing 260 (2017) 302–312.

[166] Y. Marinakis, A. Migdalas, A. Sifaleras, A hybrid particle swarm optimization–variable neighborhood search algorithm for constrained shortest path problems, European J. Oper. Res. 261 (3) (2017) 819–834.

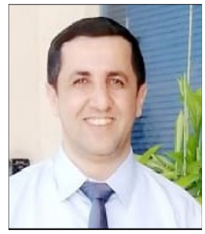
[167] M. Shehab, A.T. Khader, M.A. Al-Betar, L.M. Abualigah, Hybridizing cuckoo search algorithm with hill climbing for numerical optimization problems, in: Paper Presented at the 2017 8th International Conference on Information Technology (ICIT), 2017.

[168] M. Abdel-Basset, G. Manogaran, D. El-Shahat, S. Mirjalili, A hybrid whale optimization algorithm based on local search strategy for the permutation flow shop scheduling problem, Future Gener. Comput. Syst. 85 (2018) 129–145.

[169] V. Riahi, M. Kazemi, A new hybrid ant colony algorithm for scheduling of no-wait flowshop, Oper. Res. 18 (1) (2018) 55–74.

[170] S. Sakamoto, K. Ozera, M. Ikeda, L. Barolli, Implementation of intelligent hybrid systems for node placement problem in WMNs considering particle swarm optimization, hill climbing and simulated annealing, Mob. Netw. Appl. 23 (1) (2018) 27–33.

[171] E.R.R. Kato, G.D. de Aguiar Aranha, R.H. Tsunaki, A new approach to solve the flexible job shop problem based on a hybrid particle swarm optimization and random-restart hill climbing, Comput. Ind. Eng. 125 (2018) 178–189.
[172] G. Lin, J. Guan, A hybrid binary particle swarm optimization for the obnoxious p-median problem, Inform. Sci. 425 (2018) 1–17.
[173] B.H. Abed-alguni, F. Alkhateeb, Intelligent hybrid cuckoo search and β-hill climbing algorithm, J. King Saud Univ.-Comput. Inf. Sci. (2018).
[174] M.A. Al-Betar, $$ β $$-Hill climbing: an exploratory local search, Neural Comput. Appl. 28 (1) (2017) 153–168.
[175] N. Sulaiman, J. Mohamad-Saleh, A.G. Abro, A hybrid algorithm of ABC variant and enhanced EGS local search technique for enhanced optimization performance, Eng. Appl. Artif. Intell. 74 (2018) 10–22.
[176] F. Zhao, S. Qin, Y. Zhang, W. Ma, C. Zhang, H. Song, A hybrid biogeography-based optimization with variable neighborhood search mechanism for no-wait flow shop scheduling problem, Expert Syst. Appl. 126 (2019) 321–339.
[177] J. Pei, X. Liu, W. Fan, P.M. Pardalos, S. Lu, A hybrid BA-VNS algorithm for coordinated serial-batching scheduling with deteriorating jobs, financial budget, and resource constraint in multiple manufacturers, Omega 82 (2019) 55–69.
[178] C. Yan, J. Ma, H. Luo, A. Patel, Hybrid binary coral reefs optimization algorithm with simulated annealing for feature selection in high-dimensional biomedical datasets, Chemom. Intell. Lab. Syst. 184 (2019) 102–111.
[179] S. Mirjalili, SCA: a sine cosine algorithm for solving optimization problems, Knowl.-Based Syst. 96 (2016) 120–133.
[180] S. Mirjalili, Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm, Knowl.-Based Syst. 89 (2015) 228–249.
[181] S. Mirjalili, S.M. Mirjalili, A. Hatamlou, Multi-verse optimizer: a nature-inspired algorithm for global optimization, Neural Comput. Appl. 27 (2) (2016) 495–513.
[182] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[183] I. Aljarah, A.-Z. Ala'M, H. Faris, M.A. Hassonah, S. Mirjalili, H. Saadeh, Simultaneous feature selection and support vector machine optimization using the grasshopper optimization algorithm, Cogn. Comput. 10 (3) (2018) 478–495.
[184] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications, Future Gener. Comput. Syst. 97 (2019) 849–872.
[185] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp swarm algorithm: A bio-inspired optimizer for engineering design problems, Adv. Eng. Softw. 114 (2017) 163–191.
[186] R.A. Ibrahim, A.A. Ewees, D. Oliva, M.A. Elaziz, S. Lu, Improved salp swarm algorithm based on particle swarm optimization for feature selection, J. Ambient Intell. Humaniz. Comput. (2018) 1–15.
[187] N. Singh, F. Chiclana, J.-P. Magnot, A new fusion of salp swarm with sine cosine for optimization of non-linear functions, Eng. Comput. (2019) 1–28.
[188] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: Paper Presented at the Micro Machine and Human Science, 1995. MHS'95. Proceedings of the Sixth International Symposium on, 1995.
[189] X.-S. Yang, Firefly algorithms for multimodal optimization, in: Paper Presented at the International Symposium on Stochastic Algorithms, 2009.



Mohammad Tubishat received his Ph.D. degree in Computer Science (Artificial Intelligence — Natural Language Processing) from the University of Malaya in 2019. In addition, M.Sc. in Computer and Information Sciences from Yarmouk University in 2004. Furthermore, B.Sc. degree in Computer Science from Yarmouk University in 2002. His research interests include natural language processing, data mining, artificial intelligence, machine learning, optimization algorithms, data science, and sentiment analysis.

Norisma Idris joined Faculty of Computer Science and IT, University of Malaya in 2001. She is currently an Associate Professor and head of Artificial Intelligence Department. She received her Ph.D. in Computer Science from University of Malaya in 2011. Her research interest is on Natural Language Processing where the main focus is on developing efficient algorithms to process texts and to make their information accessible to computer applications, mainly for text normalization and sentiment analysis.

Mohammad Abushariah is working as an Associate Professor at the Department of Computer Information Systems, King Abdullah II School of Information Technology, The University of Jordan. He obtained his bachelor degree in information technology from the International Islamic University Malaysia, Malaysia in 2005. He then obtained his Master degree in software engineering and Ph.D. in computer science and information technology specialized in Natural Language Processing (NLP) and speech processing from University of Malaya, Malaysia in 2007 and 2012, respectively. He is a supervisor of various Ph.D. and Master research students specialized in Arabic NLP, speech processing, and software engineering. He has over 50 publications in ISI journals, IEEE international conferences, and technical reports. He was appointed as the Guest Editor of a special issue (Volume 19, No. 2) on Arabic Natural Language Processing and Speech Recognition: A study of algorithms, resources, tools, techniques, and commercial applications, which was published in 2016 by the International Journal of Speech Technology (IJST), Springer. In January 2017, he was appointed as an External Associate Editor for the Malaysian Journal of Computer Science (MJCS), which is indexed by ISI/WoS (JCR). In April 2015, he was appointed as an Editorial Team Member of the International Journal of Open Information Technologies (INJOIT). He also serves as a reviewer for various specialized international journals indexed in ISI/WoS (JCR) and conferences. In April 2014, he was appointed as a Consultant of NLP and speech recognition for Samsung R&D Institute in Jordan working with a distributed team with Samsung Headquarter in South Korea. His research interests include: Arabic NLP, Arabic speech processing, text and speech corpora, language resources production, speaker recognition using biometrics, speech emotion recognition, Arabic sign language, sentiment analysis, and software engineering.