

A Comparative Study of Meta-Heuristic and Conventional Search in Optimization of Multi-Dimensional Feature Selection

Khin Sandar Kyaw, Department of International Business Management, Didyasarin International College, Hatyai University, Thailand*

Somchai Limsiroratana, Department of Computer Engineering, Prince of Songkla University, Thailand

Thampas Sattayaraksa, Department of International Business Management, Didyasarin International College, Hatyai University, Thailand

ABSTRACT

The algorithmic-based search approach is ineffective at addressing the problem of multi-dimensional feature selection for document categorization. This study proposes the use of meta heuristic-based search approach for optimal feature selection. Elephant optimization (EO) and ant colony optimization (ACO) algorithms coupled with Naïve Bayes (NB), support vector machine (SVM), and J48 classifiers were used to highlight the optimization capability of meta-heuristic search for multi-dimensional feature selection problem in document categorization. In addition, the performance results for feature selection using the two meta-heuristic-based approaches (EO and ACO) were compared with conventional best first search (BFS) and greedy stepwise (GS) algorithms on news document categorization. The comparative results showed that global optimal feature subsets were attained using adaptive parameters tuning in meta-heuristic-based feature selection optimization scheme. In addition, the selected number of feature subsets were minimized dramatically for document classification.

KEYWORDS

Cost Function, Nature-Inspired Algorithm, Pheromone, Support Vector Machine Algorithm, Swarm Intelligence

INTRODUCTION

Recently, document classification has become a main technology that deals with knowledge discovery process in various applications such as business intelligence model, medical intelligence model, social media intelligence model, and so on. The performance of document classification mainly depends on the quality of selected feature subset from the feature vector. Therefore, feature selection has become a major requirement to ensure relevant feature for the classification model (Kotsiantis, 2014). Selection of optimal feature subset from high dimensionality data for accurate classification model is becoming a tough computational research gap. Furthermore, text feature selection can be regarded as NP-hard problem (Abdollahi et al. 2019) because the number of feature combinations escalate exponentially for multi-dimensional data.

In the operation level, there are four main types of feature selection method (Remeseiro & Bolon-Canedo, 2019) such as filter (Cherrington et al. 2019), wrapper (El Aboudi & Benhlima, 2016), embedded (Hameed et al. 2018), and hybrid (Solorio-Fernández et al, 2019). Meanwhile, feature exploration level includes two optional feature searches approach such as an algorithmic based

DOI: 10.4018/IJAMC.292517

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

conventional search (Appel, 2014) and heuristic based intelligence search (Sharma & Kaur, 2020). Although several feature selections schemes exist, many employ brute force or exhaustive search (Xue et al., 2016) in which all features combination are considered exhaustively and insufficient for high-dimensional feature selection problem. To overcome this limitation, meta-heuristic based optimization algorithms (Beheshti & Shamsuddin, 2015) provides solution for non-linear, high-dimensional complex feature selection because they may provide global optimal feature subset using randomization and heuristic-based search capability. In the capability of meta-heuristic scheme, decentralization and randomization are performed by all groups of meta-heuristic algorithm for searching task. However, the objective function is the major driver of meta-heuristic search mechanism for specific application problem. In contrast, conventional algorithmic-based search such as best first search (BFS) (Clausen & Perregaard, 1999), greedy stepwise search (GSS), and ranker search (RS) (Drotár et al. 2019), uses exhaustive search in which only the best scored features are selected locally and therefore prone to bias for feature selection from classes with rich feature scores.

In addition, as data science evolves, meta-heuristic intelligence has gained grounds for studying the characters of complex data. Because of the adaptive search mechanism of meta-heuristic approach, it is relevant for exploring optimal feature by striking a balance between exploitation and exploration search scheme. Search results are then plugged with the objective function of the feature selection scheme and learning models to evaluate the quality of selected feature subsets which matched or at variance with the cost function of the specific problem. Meta-heuristic algorithms such as bat search (Yang, 2013), cuckoo search (Shehab et al. 2017), flower pollination search (Abdel-Basset & Shawky, 2019), firefly search (Alomoush et al., 2018) mimic nature-inspired search. In the nature-inspired search process, all agents search randomly in different directions, then shares information and experience regarding searching and compare their results for better outcome. Furthermore, the search process of individual organism in nature is performed based on collaboration (de-centralization) and are very helpful to reduce work overloaded and enhance output (solution).

This study aims to develop a meta-heuristic based search feature selection model that can provide global optimal features with reduced feature subset size for document categorization model. It employed a comparative approach for the optimization of multi-dimensional feature selection in document categorization using swarm and nature-inspired based meta-heuristic search, with filter approach and conventional search method. In the implementation of the proposed system, Ant Colony optimization (ACO) and Elephant optimization (EO)-based algorithms are embedded in the correlation-based feature subset filter approach to optimize the feature selection process. Meanwhile, merit function is used as the objective functions to guide the meta-heuristic based feature search process while searching for global optimal feature in the multi-dimensional feature space. Principal component analysis (PCA) is applied to compress the selected feature before subjecting it to the learning model. Finally, evaluation of the proposed methods was tested on news dataset using three classifiers.

Furthermore, this research investigates the efficiency of parameter tuning in meta-heuristic based searching scheme for multi-dimensional feature selection in document classification problem. The performance results in terms of accuracy (%), root relative squared error (RRSE) and computation time (CT) using the two meta-heuristic algorithms were compared with the conventional search methods. The remaining part of paper is presented in the following order, related works, background theory, system implementation, comparative experimental results and discussion, conclusions and future works.

RELATED WORKS

Multi-dimensional feature selection is a combinatorial optimization problem (Xu et al. 2009) which includes identifying the optimal feature subsets within a set. In this section, works related to feature

selection in document categorization and meta-heuristic based feature selection optimization in various application areas are described.

Several authors have described various feature selection methods. Harish and Revanasiddappa (2017), reviewed feature selection scheme for high dimensionality problem of text categorization. Popular feature selection schemes such as term frequency-inverse document frequency (TF-IDF) (Spärck, 2004), mutual information (MI) (Vergara & Estévez, 2014), ambiguity measure (AM) (Liu, 2013), information gain (IG) (Shang et al. 2013), chi-square (Şahin & Kılıç, 2019), term strength (TS) (Yang, 1995), symbolic feature selection (SFS) (Harish et al. 2010), and term frequency-relevance frequency (tf-rf) (Lantt et al., 2005), were presented. They observed eight feature selection schemes coupled with five classifiers for text categorization and then evaluated experimental results on three standard datasets which included 20 Newsgroup, 4 Universities and Reuters 21578. In addition, they used F-measure to evaluate classification results. The feature selection methods assigned high score values to distinctive features and the system allowed these distinctive features to input to the categorization model. According to the experimental results, symbolic feature selection (SFS) provided the best classification results by selecting distinctive features for text categorization model. In specific, SFS provided better F-measure on 20 Newsgroups dataset for four types of classifiers including NB, KNN, CBC and SVM. In addition, SFS achieved the best classification result with value of 93.4% for 4-Universities dataset. Similarly, SFS outperformed the best classification result with value of 94.0% for Reuters-21578 dataset.

In Ghosh et al. (2019), the authors proposed a hybrid wrapper-filter feature selection scheme using ant colony swarm optimization algorithm to generate optimal feature subset with far less computational cost. They used ten popular datasets from the UCI repository and NIPS2003 to evaluate the proposed system performance and separated into three groups of datasets by considering the number of features, for example, small (number of features < 10), medium (10 < number of features < 100), and large (number of features > 100). In addition, KNN and MLP classifiers were used to analyze the results for number of selected features (%), and accuracy (%) upon population size and number of iterations. The proposed system revealed that (WFACOFS) was superior to ACO, BCOFS, HMOGA, ME-PSO, and SA algorithms. Seven datasets except MK1, HR and MN provided better accuracy for MLP classifier results. On computation time analysis, the proposed WFACOFS required a less time in comparison to the wrapper method (TFSACO). Two new datasets, namely, facial emotion recognition (FER) and microarray datasets, were used for evaluation of robustness. In both datasets, the proposed system gained significant improvement in accuracy for the entire feature set. Summarily, the proposed WFACOFS reduced the feature dimension and increased the accuracy significantly.

A new elephant search optimization algorithm combined on deep neural network to effectively and efficiently analyze large microarray data by selecting the best gene expressions from large volume of microarray data was proposed (Panda, 2018). Firefly search (FFS) was used as comparison model to highlight the effectiveness of the Elephant search scheme in the feature selection process. In addition, the experimental results obtained from the proposed system were compared for suitability in future Bioinformatics research. The experiments were carried out on ten cancer microarray datasets from UCI machine learning repository. The authors concluded that the proposed Elephant search method selected the most relevance genes from many redundant genes present in the dataset and provided good accuracy results. Furthermore, the proposed system proved equally significant to the best available method in the literature when tested with ANPVA and Post hoc Tukey HSD Test to check the suitability of proposed system.

In Jiang et al. (2018), a new feature selection scheme based on hybridization of mutual information feature selection (MIFS) filter and modified binary cuckoo search algorithm (MBCS) wrapper method was proposed. The accuracy of K-nearest neighbor (KNN) classifier was used as the fitness function to evaluate the quality of selected feature. Six standard datasets from UCI machine learning repository were used to test the performance of the proposed system. The experimental results proved that the

proposed system achieved high classification performance with significantly low computational time and number of features.

The authors in Fong et al. (2014) proposed swarm search feature selection approach (SS-FS) to select optimal feature subset from an extremely complex large array of features. The experimental results proved that SS-FS gave a high accuracy classification result on two empirical biomedical datasets testing. Three measurements such as average error rate, total consuming time, and length of feature subset, were used to evaluate the proposed system. In addition, experimental results were compared with a baseline benchmark result of correlation-based feature selection. The experimental results showed that the proposed system outperformed the baseline correlation-based feature selection method and the non-FS methods. Furthermore, the proposed system achieved a very low error rate and reduced the feature size to half the original size. They proved that the proposed system can be applied as a powerful feature selection tool in biomedical informatics and analytics.

In Fong et al. (2013), wrapper based feature selection scheme with swarm search optimizer was proposed to overcome the NP-hard computational problem for selecting an optimal set of features from high dimensional dataset. In addition, correlation-based filter selection was used as a baseline benchmark for comparing the performance of the proposed system. Three popular classification algorithms and three contemporary meta-heuristic methods were used for this proposed system. Moreover, average error rate and maximum iterations were used as stopping criteria. Time consumption and dimension reduction were used for performance comparison. The experimental results proved that the three-swarm search-based feature selection models provided low error rate of around 4%. However, most of the swarm search methods needed more computation time to find an optimal feature subset for good classification accuracy. Nevertheless, the proposed swarm search-based feature selection system offered a good balance between low error rate and reasonable time efficiency.

A new improved cat swarm optimization algorithm (ICSO) was proposed for feature selection optimization problem in a big data text classification experiment (Lin et al. 2016). The experimental results showed that ICSO with term frequency-inverse document frequency (TF-IDF) was more accurate than TF-IDF alone. This experiment used a statistic best solution of 3600 as a termination condition and measured the average classification accuracy over ten cross validation process. Support vector machine was used to build the classification model. According to the experimental results, the proposed ICSO yielded greater classification accuracy than traditional CSO on testing of five UCI datasets. In addition, the proposed ICSO was used for real-world application of Chinese text classification in which textual data was collected from open data platform for food culture in Taiwan food categories. The experimental results proved that ICSO + TF-IDF + SVM provided higher accuracy than TF-IDF + SVM, and the proposed ICSO reduced the selected feature compared to experiment without ICSO.

A feature selection algorithm based on Ant Colony Optimization (ACO) and Artificial Neural Network (ANN) was proposed for text categorization (Joseph Manoj et al. 2019). Reuters-21578 dataset was used to test the proposed system performance. The proposed system results were compared with the results of genetic algorithm, information gain, ant colony optimization, and CHI square selection. The precision and recall of the proposed system attained a better result than other four feature selection algorithms. In addition, the predicted accuracy rate (PAR) of the proposed ACO-ANN system outperformed PAR result of baseline, ant colony optimization algorithm and genetic algorithm. Furthermore, the micro averaged F_1 of the proposed ACO-ANN system was better than the other four feature selection algorithms.

In AlFarraj et al. (2019), an optimized feature selection method based on fireflies gravitational ant colony optimization (FGACO) approach was proposed to solve big data predictive analytical problem. Experimental results proved that the proposed system provided better sensitivity, specificity, accuracy, error rate and the number of selected features based on time. Furthermore, the performance of the proposed system was evaluated on different benchmarks of datasets including Varibench protein dataset, PDB dataset, bank marketing dataset, and lung cancer dataset. The experimental

results of the proposed system were compared with the results of traditional GSA, FA, and ACO methods. The proposed system selected the number of features lesser than the other three traditional methods. Moreover, the proposed system provided a minimum root means square error (RMSE) and high accuracy in terms of sensitivity and specificity when compared to the other traditional methods. In summary, the proposed system gained an average efficiency of 98.4625% with minimum time consumption for feature selection.

In addition, a new text feature selection method integrated with ant colony optimization (ACO) to reduce the dimension of feature sets for sentiment analysis problem was proposed (Ahmad et al. 2019). The k-nearest neighbor (KNN) was used as a classifier to evaluate the quality of selected candidate subset for optimum features. The proposed system performance was compared with two hybrid algorithms, namely, the genetic algorithm with information gain (IG-GA) and information gain with rough set attribute reduction (IGRSAR). Five datasets were used to evaluate the proposed algorithm in terms of precision and recalls. The experimental results showed that the proposed system achieved a higher F-score than the other two algorithms. In addition, the experimental results indicated that the proposed ACO-KNN was effective in feature extraction and superior to the baseline algorithms. Based on the experimental results, the proposed system provided less mean-square error (MSE) and minimum feature subset length.

In Cheng et al (2013), the authors proposed a new swarm intelligence idea for big data analytic. They analyzed difficulty of big data problem such as high dimensionality problem, dynamical data problem, and multi-objective problem. They provided two typical examples of big data analytics application such as intelligent transportation system, and wireless sensor networks, and described the capability of swarm intelligence to fulfill the research gap of big data analytic problems in real-world applications.

In Manikandan & Kalpana (2019), a novel feature selection method based on fish swarm optimization (FSO) was proposed to overcome the combinatorial problems in big data analysis. Amazon product review datasets were used to evaluate the proposed system. The proposed system performance was compared with mutual information (MI) method-based feature selection system. The experimental results showed that the proposed FSO with CART method improved accuracy by 7.91% when compared with MI-CART. In addition, the experimental result of FSO with random forest increased by 7.31% when compared with MI-random forest. Moreover, the proposed system averagely reduced the number of features selected to 59.32% for various number of iterations.

A general framework of feature selection based on swarm intelligence search was proposed to solve a NP-hard problem for selecting optimal feature subset from high dimensional data (Fong et al. 2013). The experiments were carried out by testing three swarm intelligence algorithms with three different classification algorithms. Furthermore, the twelve pairs of proposed swarm search-based feature selection were tested across five well-known datasets from UCI repository. The experimental results proved that the proposed swarm search-based feature selection system was able to attain low error rates in classification.

In Liu & Wang (2019), the authors described a survey related with nature-inspired meta-heuristics for feature selection in classification. They showed various metaheuristic algorithms for feature selection in various areas. The survey included various feature selection approaches such as wrapper, filter, and hybrid, with different metaheuristic algorithms. In addition, they discussed the reasons for emerging metaheuristic algorithms in feature selection that included several significant challenges for solving feature selection problems such as accuracy, stability, scalability, and computational cost.

Cuckoo search based feature selection optimization for sentiment analysis was proposed (Kumar et al. 2018). They used supervised soft computing such as SVM, multilayer perceptron, decision tree, Naïve Bayesian, and k-nearest neighbor, and traditional TF-IDF feature extraction before plugging with novel feature optimization model. Benchmark Kaggle tweets dataset was used to test the proposed system performance. Empirical analysis validated that the proposed system outperforms elementary supervised algorithms based on conventional TF-IDF score. In addition, experimental

results showed that the proposed system successfully tackled trade-off between optimum feature selection and performance, by yielding a significant reduction in the feature set and an increase in performance at the same time.

In Zhang et al. (2018), the authors proposed a modified Firefly optimization algorithm (FA) for feature selection in classification and regression models. Efficiency of the proposed system was evaluated on 29 classification and 11 regression benchmark datasets. The experimental results showed statistically significant improvements over other FA variants and classical search methods.

In Kumar & Jaiswal (2019), the authors proposed a general framework for swarm intelligence-based optimal feature selection model for predictive sentiment model. They used two swarm intelligence algorithms, namely, binary grey wolf and binary moth flame, and two benchmark Twitter corpus datasets, to prove the performance of the proposed framework. Furthermore, five baseline classifiers were used for feature training. The experimental results validated that the proposed swarm intelligence based optimal feature selection outperforms baseline supervised learning algorithms. The binary grey wolf algorithm provided an average improvement of 9.4% in accuracy with an approximate 20.5% average reduction in features. In addition, the binary moth flame algorithm achieved an average accuracy improvement of 10.6% with an approximate 40% average reduction in features.

Furthermore, text document classification using swarm intelligence in which ant clustering algorithm (ACA) was used for grouping web document (Vizine et al. 2005). Two modifications such as a metric and a cooling schedule for a user-defined parameter, were used to evaluate the similarity degree of text data and to enhance the convergence properties of the ant algorithm. In addition, the sets of IEEE WCCI-1998 CD documents were used to evaluate the proposed system. Furthermore, experimental results of proposed system were compared with evolutionary computation (EC), artificial neural network (ANN) and fuzzy system (FS). According to the experimental results, nine different clusters were determined by the proposed algorithm, and only four objects were left isolated.

In Nguyen et al. (2020), the authors proposed sentiment classification for online social media using whale optimization algorithm and social impact theory based optimization (SITO). The proposed system results were compared with thirty-three supervised learning algorithms on testing of real IMDB, Polarity, and Amazon data sets. Precision, accuracy, F-measure, recall, and MCC were used to evaluate the proposed system performance. The proposed system provided a good result for sentiment analysis problem and outperformed the other thirty-three supervised machine learning algorithms in terms of accuracy, precision, and F-measure. In addition, the adapted intelligent optimization algorithms promised recall and MCC values.

In Akyol & Alatas (2020), the authors provided a comprehensive survey of swarm intelligence (SI) approaches for feature selection in data mining. It included feature selection approach based on three common swarm intelligence algorithms including PSO, ABC and ACO. The comprehensive survey involved the description of representation and corresponding search mechanism of swarm intelligence. It also included survey for many studies attempting to apply SI for feature selection and improve the selection performance. The authors concluded that the performance can be improved by modifying the search mechanism which is different for different algorithms due to variations in characteristics.

Authors in Kashef & Nezamabadi-Pour (2013) proposed a new feature selection algorithm based on advanced binary ant colony optimization (ABACO). Three datasets from UCI machine learning repository were used to evaluate the performance of the proposed system in terms of accuracy and average feature reduction. In the process of updating best route, only the best ant with the smallest mean square error (MSE) for the classifier is allowed and it continues for all iterations and end when the best feature subset with the least MSE of the classifier is obtained. The experimental results showed that the proposed system (ABACO) provided an effective search capability for finding minimal feature subset and achieved optimal feature subset classification accuracy with minimum number of selected features.

In Ke et al. (2010), a multi-objective ACO algorithm (MOACO) was proposed for rough feature selection. The proposed model adopted an elitist strategy to enhance the convergence performance and used crowding comparison operator to diversify the constructed solutions spread. Gene expression datasets were used for numerical experiments and a modified non-dominated sorting genetic algorithm was used as a comparison model. The experimental results showed that the proposed MOACO provided competitive solutions, efficient for rough feature selection, with few numbers of features for better classification results when compared to MNSGA-II.

Dadaneh et al. (2016), proposed unsupervised probabilistic feature selection based on ant colony optimization (UPFS). In calculating the proposed feature selection, inter-feature information was used to measure the similarity between the features, and the amount of redundancy between present feature and all former selected features. Performance evaluation of the proposed system was carried out on 10 datasets using 15 supervised and unsupervised feature selection methods by different classifiers. The experimental results showed that the proposed system provided better efficiency for selection of optimal features than the previous related methods.

In Peng et al. (2018), the authors proposed an improved feature selection algorithm based on Ant Colony optimization (FACO) to enhance the performance of classifier. The proposed FACO system included a fitness function, and the optimization of pheromone updating rule for eliminating redundant features and avoiding local optimum feature. The experiment was carried out on KDD CUP99 dataset. According to the experimental results, accuracy of the classifier was significantly enhanced by selecting the global optimal feature using FACO algorithm.

In Ozger et al. (2017), a comparative study on binary Artificial Bee Colony optimization methods for feature selection was performed. It employed variants of binary ABC algorithms for solving feature selection problem. The performance of the proposed system was evaluated on 10-UCI datasets using measurements of execution time, number of selected feature and error. The experimental results showed that the proposed model provided global search ability with low computation time and classification performance.

Schiezaro & Pedrini (2013), proposed data feature selection using Artificial Bee Colony algorithm to achieve good accuracy by selecting the most relevant subset of features from the dataset. The proposed system used 10-UCI machine learning repository datasets to compare the performance with other relevant approaches such as ACO, PSO, and genetic algorithms. According to the experimental results, selected features provided by the proposed system achieved better accuracy than the original feature set for all datasets. In addition, the proposed system obtained superior accuracy than other methods.

In Shunmugapriya & Kanmani (2017), a hybrid feature selection algorithm based on Ant and Bee Colony optimization (AC-ABC) was proposed to solve optimization problem in feature selection for classification. The proposed system removed stagnation behavior of the ants and avoided time consumption for global search by the employed bees. The experiments were carried out on thirteen UCI benchmark datasets to evaluate the proposed algorithm and J48 classifier was used to measure the performance accuracy for classification. ACO-NN Hybrid, Hybrid ACO, ACO based FS, ABC-FS, PSO-SVM, Catfish Binary PSO, IQR Bee, ABC-DE Hybrid, and ACO-PSO Hybrid were used as comparison models. The results showed that the proposed system promised better classification accuracy by selecting optimal features. In specifics, the proposed system provided reduced size of feature subset with increased classification accuracy, low computation cost, and quick convergence.

Artificial Bee Colony-based support vector machines was proposed for feature selection and parameter optimization for rule extraction (Kuo et al. 2018). C5 decision tree was applied to extract rules from SVM. Eight datasets were used to evaluate the effectiveness of the proposed system. The proposed ABC-SVM-DT system achieved better classification accuracy with reduced complexity in the final decision tree when compared to genetic algorithm and particle swarm optimization algorithm. The experiments were conducted using ten UCI benchmark datasets and the results were compared with five classification methods including DT, SVM-DT, PCA-SVM-DT, PSO-SVM-DT

and GA-SVM-DT. The experimental results showed that the proposed ABC-SVM-DT obtained better accuracy than the other compared methods.

Uzer et al. (2013), proposed a hybrid approach for feature selection based on Artificial Bee Colony algorithm and support vector machine in medical datasets classification. Liver diseases and diabetes diagnostics dataset from UCI machine learning repository were used to demonstrate the performance of the proposed system. Average classification accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were used as measurement schemes for performance of the proposed system. According to the experimental results, the proposed system achieved classification accuracies of 94.92%, 74.81%, and 79.29%, and the authors concluded that the performance accuracy of the proposed system attained higher accuracy than the other models for pattern recognition applications. Table 1 presents the summary of some mentioned algorithms with their different characteristics including feature selection scheme, control parameter and measurement, and searching policy.

BACKGROUND THEORY

The nature of text and the role of feature selection is described to highlight the research problem. These includes a description of the unstructured multi-dimensional properties of text, challenges associated with the search for global optimal feature in text feature selection process, difference between meta-heuristic based search and conventional search for feature selection process to emphasize the research gap within multi-dimensional feature selection in text document categorization, theories and calculations for swarm intelligence-based Ant Colony, and nature-inspired intelligence-based Elephant search policies. Moreover, the nature of Best First algorithmic-based search (BFS) and Greedy Stepwise algorithmic based search (GS) with a description of control parameters to elaborate the limited capability of conventional search at solving multi-dimensional problem is presented. Three classification learning models were used to evaluate the quality of selected feature subset based on meta-heuristic intelligence. In addition, the evaluation schemes for measuring the performance of the proposed model are explained.

Nature of Text and Role of Feature Selection

Text is characterized by complexity, multi-dimensionality, and unstructured nature. Individual word in each document is regarded as feature, hence computation gets complicated in the feature selection process as the number of features grows exponentially in text categorization. Feature selection is one of the most important stage in data pre-processing because it can reduce the feature dimensions. Furthermore, the vast number of features from high-dimensional text feature space can degrade the performance of categorization model to often almost fruitless. This phenomenon comes from the analysis of data in high-dimensional feature spaces. To overcome this problem, feature selection process can be applied for more efficient analysis by maximizing the relevant and minimizing the redundant features.

However, there is a lack of scheme for optimum feature selection among data mining researchers. Normally, the total number of features considered in feature selection process is based on 2^n combination, signifying that the complexity rate doubles for each feature. Though several ways of selecting local best feature subset with corresponding search strategies already exist, most employs conventional search with feature evaluation schemes which are inefficient for solving complex feature since it only works with static and linear search. For nonlinear relations between features and concept targets, selecting a feature with high local score for specific target class might yield inaccurate result. Therefore, meta-heuristic based optimization for non-linear search is a suitable alternative for complex multi-dimensional feature space search since it can heuristically guide search agent by synchronization with feature evaluation scheme (Hall, 1999) when searching for global feature subset matched with objective function.

Table 1. Comparison of different characteristics of mentioned algorithms

Algorithm	Characteristics	Description
Symbolic Feature Selection	Feature Selection Scheme	Unconventional filter approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Degree of similarity and dissimilarity ■ Features are of interval valued type, the degree of similarity between class representative vectors is estimated based degrees of overlapping of features. ■ The relative overlapping of interval type features is not equal and hence the degree of similarity between two symbolic vectors may not necessarily be symmetric.
	Searching Policy	<ul style="list-style-type: none"> ■ Calculate the sum of the total correlations of each column of matrix with all other columns. ■ Select the feature which have their respective total correlation greater than the average correlation and subsequently retain only those for representation. ■ These selected features are responsible for high cohesion of the classes of documents.
Hybrid Wrapper-Filter Feature Selection Scheme using Ant Colony Swarm Optimization Algorithm	Feature Selection Scheme	Unconventional hybrid feature selection
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Number of ants ■ Exploitation balance factor, exploration balance factor, number of iterations ■ Pheromone evaporation factor ■ Pheromone evaluation factor ■ Weight of accuracy ■ Weight of number of features ■ Similarity
	Searching Policy	<ul style="list-style-type: none"> ■ The value of similarity is 1 if the features are completely correlated and 0 if the features are completely independent. ■ The objective is to include non-correlated features to increase the recognition capability of the system. ■ Select the feature which has high pheromone (objective function).
Elephant Search Optimization	Feature Selection Scheme	Unconventional highly nonlinear, multimodal global optimization technique
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Visual range of each elephant using Euclidean distance ■ Number of iterations ■ Stop criteria
	Searching Policy	<ul style="list-style-type: none"> ■ The best possible solution is selected through an iterative process that is represented by the lifetimes of the searching elephants. ■ The local searches are led by some chief female elephants to a higher likelihood of obtaining best results. ■ The male elephants are rangers to lead the elephant clan so that the whole elephant can go out of local optimum.
Hybridization of Mutual Information Feature Selection (MIFS) Filter and Modified Binary Cuckoo Search Algorithm (MBCS)	Feature Selection Scheme	Unconventional hybrid feature selection
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Sigmoid function, mutation operator, accuracy (objective function) ■ Filter and wrapper feature evaluation
	Searching Policy	<ul style="list-style-type: none"> ■ A solution binary vector is employed which describe 1 for a feature will be selected to compose the new dataset and 0 otherwise. ■ Moreover, the mutation operator in the genetic algorithm is used to replace the nest elimination process. ■ Binary cuckoo search uses the probability of discovery to abandon the nest to maintain population diversity.

continued on following page

Table 1. Continued

Algorithm	Characteristics	Description
Swarm Search with Bat algorithm (BA), Particle Swarm Optimization (PSO), and Wolf search algorithm (WSA)	Feature Selection Scheme	Unconventional wrapper approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ▪ BA used the control parameters including A (loudness), R (pulse rate), populations, Q_{min}, Q_{max}, minimum and maximum boundary ▪ PSO utilized the several control variables including populations, c1 and c2 (constant values). ▪ WSA includes visual distance, escape distance, escape probability, population control parameters.
	Searching Policy	<ul style="list-style-type: none"> ■ Initialize search agents and randomly assign a different feature combination as a subset. ■ Generally, calculate the fitness of the objective function for each agent, rank the agents by their fitness values, and record the one that has the highest fitness as the most promising solution using local search and optimize the current best solution. ■ In PSO, the swarming particles have velocities. Recording the best local solution and global solution in each generation is required. ■ In BA, rank the current best position in each generation according to the pulse rate and loudness; then update the velocity and position for each agent. ■ In WSA, first randomly move to update the position; second, check if they are ready to swarm towards to the closest and best companions; third, activate the escape function at random-relocate agents that are qualified to jump dimensions.
Improved Cat Swarm Optimization Algorithm	Feature Selection Scheme	Unconventional clustering algorithm
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ N, number of initial feasible solutions ■ MR, mixture ratio ■ SMP, seeking memory pool ■ SPC, self-position consideration ■ SRD, range of the selected dimension ■ CDC, count of dimension to change ■ D, number of dimensions of dataset features ■ C, a constant value in tracing mode
	Searching Policy	<ul style="list-style-type: none"> ■ Set the termination condition as a static best solution of 3600 s and calculated average classification accuracy over ten cross-validated runs. ■ Initialize N random solutions with its own position and velocity. ■ Calculate the fitness of each cat to determine the best solution. ■ Randomly assign the cats to seeking mode or tracing mode. ■ If cats find a better new solution than the old best solution, update the best solution.
ACO-ANN based Feature Selection Algorithm	Feature Selection Scheme	Unconventional artificial neural network classification approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Number of populations ■ Iteration ■ Crossover probability ■ Mutation probability ■ Initial pheromone value
	Searching Policy	<ul style="list-style-type: none"> ■ Initializes the numbers of ants. ■ Probability and state distribution rules are used by the ant to choose the features. ■ A global updating rule is applied on the subset of features according to the objective of least classification error. ■ Local updating rule is applied on the ants which are left further to maintain or decrease the pheromone level of ants.

continued on following page

Table 1. Continued

Algorithm	Characteristics	Description
Optimized Feature Selection Algorithm based on Fireflies with Gravitational Ant Colony Algorithm (FGACO)	Feature Selection Scheme	Unconventional optimized feature selection
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Attractiveness ■ Intensity values ■ Minkowski distance ■ Gravitational search algorithm
	Searching Policy	<ul style="list-style-type: none"> ■ Calculate the transition probability values of the features. ■ Estimate the attractiveness value of the feature for computing distance between the features. ■ Find the intensity value of each feature by determining it with minimum function. ■ Calculate the force direction of the feature according to its attractiveness and intensity value. ■ Estimate the change of force direction. ■ According to the feature mass and force direction, examine the best fitness value. ■ Calculate the feature acceleration and velocity value. ■ Convert the calculated velocity value into the probability value. ■ Update the position of each feature value. ■ Compare the calculated velocity value with the feature mass value; if it is nearer to the mass value, it is considered the optimal feature. ■ Repeat this process continuously until the optimal features can be detected from the feature set.
Fish Swarm Optimization	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Positions, maximum move step length, step, vision radius, bulletin board ■ Population size, n ■ Congestion factor, δ ■ Feedback probability, P_b ■ Attenuation factor, θ ■ Maximum iteration, max_iterate
	Searching Policy	<ul style="list-style-type: none"> ■ Value of the feature subsets of each fish are arbitrarily prepared. ■ The fitness function assesses the fitness value of the attribute subcategory of every fish. ■ Search for an optimum result executes three lookout steps like follow, swarm, and prey for each fish. ■ If the terminal circumstance of algorithm is satisfied, stop the yield of the optimum attribute subset; else begin the algorithm to continue through added reiteration starting.
Meta-heuristic in the Wrapper, Filter and Embedded Feature Selection	Feature Selection Scheme	Unconventional feature selection approaches
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Several control parameters were used according to the different meta-heuristic algorithms. ■ However, the measurement for feature evaluation is defined according to the types of feature selection approach. ■ Filter feature selection methods evaluate feature subsets independent from learning algorithms but dependent on the dataset itself. ■ Wrapper feature selection methods use learning algorithms to evaluate the classification accuracy of the selected feature subsets. ■ Embedded feature selection algorithm can only be one of them. Some hybrid algorithms are the combination of two or three former ones.
	Searching Policy	<ul style="list-style-type: none"> ■ Meta-heuristic algorithms guide the filter, wrapper, and embedded feature selection scheme with the help of heuristic intelligence with their own nature of search approaches.
Binary Cuckoo Search for Optimized Feature Selection	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Loss parameter, p ■ Constant value, α ■ Number of nests, n ■ Dimension, d ■ Number of generations, T ■ c_1 and c_2 values
	Searching Policy	<ul style="list-style-type: none"> ■ Each cuckoo randomly selects a host (nest) to lay eggs. ■ The number of nests does not change over time and can be considered fixed. ■ The nests which are fitter (high quality) will have more chances of being present in the following generation. ■ If the host detects the egg laid by cuckoo, the egg can be discarded or the nest itself can be abandoned, in which case the host will build a new one. ■ Pa [0, 1] represents the probability of host detecting the egg (laid by cuckoo).

continued on following page

Table 1. Continued

Algorithm	Characteristics	Description
Feature Selection Using Firefly Optimization	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Population initialization, fitness value, attractiveness or offspring of the brighter neighboring firefly, termination ■ Chaotic map, c_k ■ Randomized vector, ϵ ■ Best memory, g_{best} ■ Second leader, s_{best}
	Searching Policy	<ul style="list-style-type: none"> ■ It employs diverse chaotic attractiveness movements. ■ Search agent enhanced local and global signals, scattering strategies of weak solutions, and the best and worst memories to increase search diversity and lead the search towards global optima.
Swarm intelligence based Optimal Feature Selection [Grey Wolves Optimization (GWO) and Binary Moth Flame Algorithm (BMF)]	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	In GWO, <ul style="list-style-type: none"> ■ Number of grey wolves in the pack, n ■ Number of iterations for optimization, Niter ■ Group of population including alpha, beta, delta and omega ■ Stopping criteria ■ Fitness values In BMF, <ul style="list-style-type: none"> ■ Number of search agents (moths), maximum iteration ■ Flames (fitness value) ■ Constant parameters such as b, r, t, σ
	Searching Policy	<ul style="list-style-type: none"> ■ In GWO, initialize n random binary solutions (wolves), and calculates of their fitness values. Assigns alpha, beta and delta wolves from amongst the pack, according to the fitness. The fitness is calculated by training the classifier according to the selected features as specified by the solution (wolf) and measuring the corresponding accuracy of the classifier. For each iteration, the wolf 'encircles' the prey and shifts closer to it. ■ In BMF, provides the continuous valued positions of solutions (flames and moths). Using a sigmoid function, the continuous valued method is transformed into discrete binary values. A solution binary vector l represents feature selected and 0 represents otherwise. A threshold is applied to convert the value output by sigmoid function into binary values.
Ant Cluster Algorithm	Feature Selection Scheme	Unconventional clustering algorithm
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ $k_c = 0.01$ (initial value) ■ Runtimes=20 ■ Cycle=30 ■ $\alpha=0.7$ ■ $k_d=0.06, k_{p_{min}}=0.001, Nants=1$ ■ Grid size: 13×13
	Searching Policy	<ul style="list-style-type: none"> ■ The stopping criterion of the algorithm becomes either a maximum number of cycles (1cycle = 10,000 steps of each ant) or a minimum value for k_c. ■ In both cases, the chosen value has to be such that ants are no longer picking up objects from the grid, thus resulting in a final, stable clustering solution.
Whale Optimization Algorithm (WOA) and Social Impact Theory based Optimization (SITO)	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	In WOA, <ul style="list-style-type: none"> • Whale population, iteration number, fitness function • a, A, C, l In SITO, <ul style="list-style-type: none"> • Population with N individuals $[0, 1]$ with d dimension, termination condition • Number of persuasives • Number of supporters
	Searching Policy	<ul style="list-style-type: none"> • Exploitation Phase performs Surrounding Hunting Air Bubble Attack Method. • Exploration phase includes hunting search. • The population with N individuals which randomly take values in the range of $[0, 1]$, are generated and each has d dimensions. • In each iteration, an individual changes its attitudes if the persuasive effect is greater than the supportive effect. In the state of change of attitude, the power of each will be recalculated. • The additional probability parameter representing the probability of spontaneous attitude change is activated with κ. Thus, if the persuasive effect is greater, the probability of change takes place by $(1 - \kappa)$, otherwise, the change takes place with probability.

continued on following page

Table 1. Continued

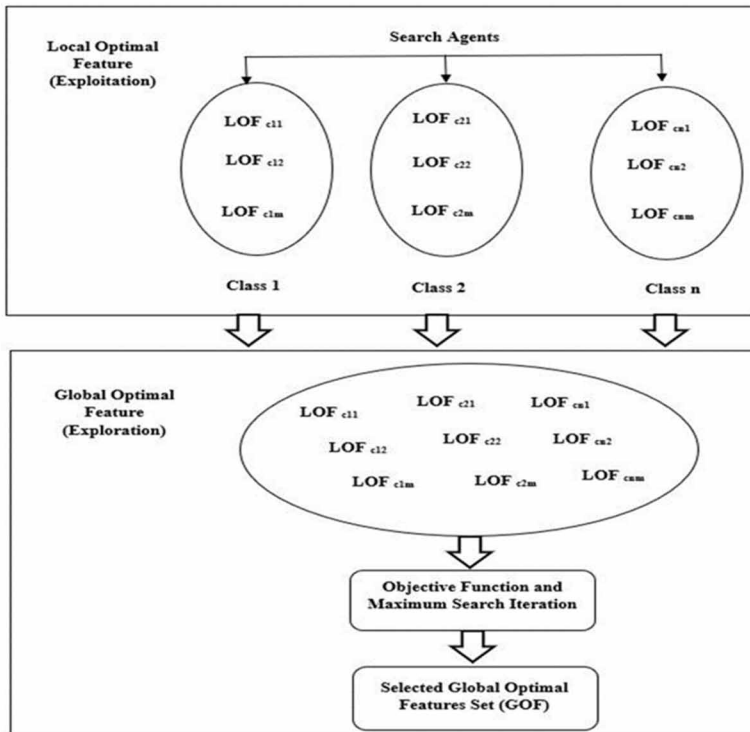
Algorithm	Characteristics	Description
Advanced Binary Ant Colony Optimization (ABACO)	Feature Selection Scheme	Unconventional feature selection approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Number of ants, m ■ Maximum number of iterations, I_{max} ■ Tunable parameters α, β and ρ ■ Initial pheromone level, τ_0 and the heuristic information, η of all n features by assigning equal values to τ and η.
	Searching Policy	<ul style="list-style-type: none"> ■ Initialize and generate m ants. ■ For each ant, construct a subset using transition rule. ■ Evaluate all constructed subsets and select the local and global best subsets. ■ If the termination is met, record the best subset. Otherwise, update τ and generate new m ants and repeat to construct a subset using transition rule.
Multi-objective ACO Algorithm (MOACO)	Feature Selection Scheme	Unconventional multi-objective feature selection approach
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Objective function includes the smallest feature subset of the original set and yield the highest classification quality. ■ Number of ants ■ Number of iterations ■ Other constant variables including α, β, τ and η.
	Searching Policy	<ul style="list-style-type: none"> ■ At each cycle, every ant constructs a solution and then pheromone trails are updated. ■ The algorithm stops iterating when the objective function (two objectives) is met.
Hybrid Feature Selection Algorithm based on Ant and Bee Colony Optimization (AC-ABC)	Feature Selection Scheme	Unconventional hybrid feature selection
	Control Parameter and Measurement	<ul style="list-style-type: none"> ■ Population of ants, K ■ EBees and OBees (SN) are the number of features (M) ■ Number of runs ■ Pheromone value, τ ■ Objective function, f_i ■ Fitness value, fit_i ■ Onlooker, v_i ■ Probability, p_j ■ Termination criteria
	Searching Policy	<ul style="list-style-type: none"> ■ Ant colony utilizes the exploitation procedure of employed and onlooker bees to determine the best ant and the best feature subset of the generation. ■ Employed bee instead of generating feature subsets, adapts the feature subsets produced by the ant. ■ Scout bees do not generate food sources, uses the newly generated feature subsets by the ants. ■ The feature subsets generated are evaluated the yielding optimal feature subset and maximum prediction accuracy gain.

- Reduced computation and memory storage cost;
- Improved classification results in term of accuracy, and declined error rate;
- Achieve more understandable causal relationship between features and classes with only significant features.

Nature of Meta-Heuristic and Conventional Search-Based Feature Selection

Meta-heuristic intelligence is an optimization scheme for achieving global optimal solution. There are many well-known groups of meta-heuristic intelligence such as swarm intelligence (Ma et al. 2017), nature-inspired intelligence (Vassiliadis & Dounias, 2009), bio-inspired intelligence (Yang & Karamanoglu, 2013) etc. Although there is an overlap between evolutionary computation (Vikhar, 2017) and swarm intelligence, metaheuristics show more characteristics of swarm intelligence, such as cuckoo search (CS), firefly algorithm (FA), and artificial bee colony (ABC) optimization. Recently, lots of newly proposed meta-heuristics have been introduced to solve feature selection problems for classification, such as Butterfly Optimization Algorithm (BOA), and Whale Optimization Algorithm (WOA). Therefore, it is necessary to investigate these newly proposed metaheuristics for feature selection in classification.

Figure 1. Functional diagram of global optimal feature selection



However, conventional approach can only retrieve local optimal feature from specific class by using linear calculation. The definition of linear search is appropriate for simple features which are not fussed on the hypothesis plan. Since most text feature is not simple, drawing exact boundary line that can separate the individual class correctly is challenging. In other words, conventional search only works with feature evaluation scheme that consider high score feature using measurement of information gain, gain ratio, mutual information, etc.

Global Optimal Feature Selection

The achievement of a global optimal feature is obtained by performing exploitation and exploration process together. Global optimal feature selection uses global selection scheme to select best feature spread across all labels of categories. Therefore, meta-heuristic intelligence-based feature selection scheme is an emerging solution for global optimal feature selection in high-dimensional feature selection process. In the process of global optimal feature selection scheme, several functional components are involved in searching for global optimal features, as shown in Figure 1.

In the exploitation stage, search agents are assigned for individual class to search for local optimal feature (LOF) and share observation among themselves. However, the exploration process is concerned with the global search coverage for the whole search space. The optimal feature selection process is proceeded by using objective function and maximum searching iteration. The local optimal features are evaluated in feature evaluation step by matching with objective function. The search process ends when selected local features matched with objective function and achieved maximum iteration for searching. Through the process of global optimal feature selection, the following profits can be attained:

- Reduced dimensions of unstructured text data;

Meta-heuristic algorithms can be used for optimizing the feature subset selection process within a filter approach by determining the global optimal feature subset that can achieve objective function of filter feature selection. Filter approaches utilize the merit function in calculating correlation between feature and its corresponding class. To achieve the most correlated feature subset, meta-heuristic algorithms play the role of feature searching in feature selection process. In the operation of meta-heuristic based search, several common and specific variables are included to compute for best feature subset that passes through the feature evaluation process. For example, objective function, iteration number, and population are common variables for meta-heuristic algorithms, while pheromone, and evaporation are specific variables for ant colony algorithm.

A good search strategy should provide good global search capability, rapid convergence to near optimal solution, good local search ability, and high computational efficiency. Search strategies can be categorized into three groups: exponential, sequential, and randomized. Exponential search, also called complete search, is the most exhaustive global search strategy. It starts from original feature set and guarantees optimal result. However, this strategy is generally impractical and computationally intensive especially for high dimensional data sets, and prohibitive and intractable for all but a small initial number of features. An example of this strategy is exhaustive search (Neumann & Matas, 2011) which evaluates all possible subsets to find the optimal subset. Sequential search, also called greedy hill-climbing search, adds, or subtracts one feature at a time. The most common sequential strategies are sequential forward selection (SFS) and sequential backward selection (SBS). It is relatively simple to implement, its complexity is polynomial with respect to the number of features, and it is robust to multi collinearity problems. However, these methods perform poorly on non-monotonic indices and may cause nesting effect (Hao et al. 2003) because once a feature is added (or deleted), it is not allowed to be deleted (or added) latter. Moreover, they are sensitive to feature interaction, hence they can easily be trapped into local minima (Maldonado & Weber, 2009). Sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS) (Nakariyakul & Casasent, 2008) were developed to overcome these problems by providing mechanisms to re-select the deleted features and delete the already added features. Some other examples of sequential search strategies are best first search, beam search—an optimized solution of best first search, and plus 1 take-away algorithm (PTA) (Kudo & Sklansky, 2000).

Randomized search strategy starts by randomly selecting the features and then proceeds with two different search strategies. The first uses the classical sequential or bidirectional search, e.g., simulated annealing (Ezugwu et al. 2017) and random hill-climbing (Goswami et al. 2019). The second uses strategies that have no regular movements, e.g., genetic algorithm (GA) (Babatunde et al. 2014), and Tabu search (Zhang & Sun, 2002) The second strategies can escape local optima in the search space, but they have a greater chance of producing incorrect results due to non-availability of mechanism to capture relationship between features.

Swarm Intelligence-Based Ant Colony Search

Swarm intelligence is a popular approach of computational intelligence, and matches with five principles including proximity principle, quality principle, principle of diverse response, principle of diverse stability, and principle of adaptability. Ant Colony Optimization algorithm (Dorigo & Stützle, 2000) is a popular swarm intelligence that emerged in the early 1990s. It is inspired by the observation on real ants in their search for the shortest paths to food sources. It is equipped with collective unsophisticated agents interacting with their environment. It includes three main categories of foraging task: employed bees which are responsible for exploiting the food and recruiting the others by dancing; onlooker bees choose the food by watching the movement of employed bees' dancing; and scout bees perform the exploration process. Employed bee becomes a scout bee in case of food source exhaustion. The population of food source is generated randomly using Equation (1):

$$x_{ij} = x_j^{min} + rand(0,1)(x_j^{max} - x_j^{min}) \quad (1)$$

where $i = 1 \dots SN$, $j = 1 \dots D$, SN is the number of food sources, D is the number of design parameters, x_j^{min} and x_j^{max} are lower and upper boundary of j^{th} dimension, correspondingly.

Employed bees performing local search for neighborhood sources (exploitation), can be generated using Equation (2):

$$v_{ij} = x_{ij} + \varphi_{ij} (x_{ij} - x_{kj}) \quad (2)$$

where i is the current solution, k is a neighbor solution chosen randomly, and φ_{ij} is a real random number of uniform distributions in the range $[-1, 1]$. Greedy selection is applied between current and its mutant solutions to select better population. In addition, combination of local search and Greedy selection are applied for every food source in the population. The onlooker bees search for better solution among the neighborhood of food sources and is selected stochastically depending on their fitness values using Equation (3). One of the selection schemes such as roulette wheel, tournament, stochastic university sampling, ranking based or others, is employed to select a better solution after calculating the value of individual food source probability:

$$p_i = \frac{fitness_i}{\sum_{i=1}^{SN} fitness_i} \quad (3)$$

If the solution cannot be improved by the local search for both phases of employed bees and onlooker bees, the counter is increased by one. The counter keeps the exploited and retained number of solutions in the population, which is used for determining exploitation sufficiency and exhaustion. The exhausted solution is replaced a new randomly solution generated by Equation (1) if the counter exceeds the limit.

Algorithm 1. Ant Colony Optimization Algorithm (ACO)

Input: Training data (D), Objective function $Merit(D)$, $D = (d_1, d_2, \dots, d_n)$.

Output: Optimal feature subsets with n features, d_{best} .

1. Generate the number of populations of ants, NP . // $NP=20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200$.
2. Define the evaporation value, e , with 0.9.
3. Define the pheromone value, $\alpha=2.0$.
4. Define the value of heuristic, $\beta=0.7$.
5. Define a maximum number of iterations, NI . // $NI=20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200$.
6. Initialize $CI = 0$ // current iteration or time step
7. **while** ($CI < NI$) // continue searching for the optimal feature subset if current iteration is less than the maximum number of iterations.
8. Get a population of ants for food source (feature) randomly, named x_{ij} , using Equation (1).
9. Search the feature using the exploration process which

coverages for the whole set of population.

10. Do the exploitation search to evaluate the quality of selected feature (F_i) according to the objective function in Equation (3).
11. Perform the local trial update by selecting the feature or solution among NP, named v_{ij} , with j position, randomly.
12. **if** ($F_i > F_j$) **then**
13. Replace j by the new solution
14. **end if.**
15. A probability of the worse pheromones are abandoned, and new ones are built.
16. Keep the best solution (feature) for ants.
17. Rank the solutions (features) for every iteration search of ants and find the current best.
18. Pass the current best solutions to next cycle. // update the best pheromone value for next iteration search.
19. **end while.**
20. **return** the optimal feature subset or solution (d_{best}).

Nature-Inspired Based Elephant Search

Elephant Optimization Algorithm (EO) (Deb et al. 2016) is one of the modern nature-inspired metaheuristic algorithm. In EO, gender ratio is used to handle the inclination on global exploration or local intensification. EO attempts to bridge the advantages of swimming ability of PSO and evolutionary ability of GA, by tapping on the dual benefits of swarm movements and evolution. The solutions of EO are enhanced across the spatial domain by allowing the position of best fitness in swarm and a better generation by the principle of retaining only the fittest in evolution. The best heuristic information is carried over to future generation, through evolutions, by updating current elephants intermittently with the best-found positions. In the calculation of EO, elephant population is divided into k clans. Each clan is led by the matriarch and Each member j of clan i moves considering the matriarch c_i using Equation (4):

$$x_{new,ci,j} = x_{ci,j} + \alpha (x_{best,ci} - x_{ci,j}) \times \gamma \quad (4)$$

where $x_{new,ci,j}$ means the new position of the elephant j in clan i , while $x_{ci,j}$ refers its old (current) position, $x_{best,ci}$ is the best solution of clan c_i , $\alpha \in [0, 1]$ represents a scale factor which determines the influence of the matriarch. The optimal elephant in a clan is updated using the following Equation (5):

$$x_{new,ci} = \beta \times x_{center,ci} \quad (5)$$

where $\beta \in [0, 1]$ that controls influence of the $x_{center,ci}$ which is defined as:

$$x_{center,ci,d} = \frac{1}{n_{ci}} \times \sum_{l=1}^{n_{ci}} x_{ci,l,d} \quad (6)$$

where $1 \leq d \leq D$ is the d^{th} dimension where D is total dimension of the space and n_{ci} is the number of elephants in clan i . In each generation for each clan m_{ci} the number of the elephants move to live

far from the clan. Elephants with the defeat or worst fitness values are chosen to be moved and their new position is calculated using Equation (7):

$$x_{defeat,ci} = x_{min} + (x_{max} - x_{min} + 1) \times rand \quad (7)$$

where x_{min} and x_{max} represent lower and upper bound of the search space, respectively. Parameter $rand \in [0, 1]$ represents a random number form uniform distribution. The pseudo code of elephant optimization algorithm is described in Algorithm 2.

Algorithm 2. Elephant Optimization Algorithm (EO)

Input: Training data (D), Objective function $Merit(D)$, $D = (d_1, d_2, \dots, d_n)$.

Output: Optimal feature subset with n features, d_{best} .

1. Generate a population of elephant, NP , and the maximum number of iteration or generation, NI . // $NP, NI = 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200$.

2. Define the mutation probability with 0.01.

3. Define the mutation type with bit flip.

4. Initialize the generation counter $t=1$.

5. **repeat**

6. Rank all the elephants as stated by their fitness value.

7. **for all** clans C_i in the population of elephant (NP) **do**

8. **for all** elephants j in the clan C_i **do**

9. Update $x_{ci, j}$ and generate $x_{new,ci, j}$ by Equation (4).

10. **if** $x_{ci, j} = x_{best,ci}$ **then**

11. Update $x_{ci, j}$ and generate $x_{new,ci, j}$ by

Equation (5).

12. **end if**

13. **end for**

14. **end for**

15. **for all** clans C_i in the population of elephant (NP) **do**

16. Substitute the defeat elephant in clan C_i by

Equation (6).

17. **end for**

18. Evaluate the population of elephant by the newly updated positions.

19. **until** $t < NI$

20. **return** the optimal solution (feature subset, d_{best}).

Algorithmic Based Best First Search (BFS) and Greedy Stepwise (GS)

The algorithmic-based BFS and GS methods are types of exhaustive search and is composed of three main variables which are shown in Table 2. Search index and direction, search policy, and evaluation schemes are worked together to search for local best features from the search space. They generate the best candidate feature among all possible combination of features from local search and are suitable for handling linear model. However, the computation cost is high for multi-dimensional feature selection problem because the algorithmic-based search is always consider based on the scheme of

2ⁿ combination. Therefore, the higher the number of features the higher the exponential computation cost. Table 2 shows three pairs of control variable combination for algorithmic-based search.

Table 2. Combination of control variable for algorithmic-based search

Search Index and Direction	Search Policy	Evaluation Scheme
Empty and Forward	Exponential Search	Consecutive Non-improving Nodes
Full and Backward	Sequential Search	Consecutive Non-improving Nodes
Range of index and Bi-directional	Random Search	Consecutive Non-improving Nodes

Classification Learning Models

Document categorization is a part of knowledge discovery applicable in medical (Raymer et al. 2003), social media (Ruz et al. 2020), online news classification (Li et al. 2016), and so on. The role of classification model is important in text documents categorization. Classification learning model performs the training and testing phase between the target label and other attributes on all the samples as a supervised learning model. Three classification models are built in this experiment to evaluate the performance of two meta-heuristic based optimization for feature selection process.

In Naïve Baye classification (John & Langley, 2013), Bayes theorem is used to build the learning network by selecting the strong independent assumptions among the features. It calculates the conditional probabilities of the independent features and applies the conditional prior probabilities generated from the training data to predict testing data. Furthermore, it can handle missing values and is effective for text categorization in which $d \in D$, where D denotes the training document set and d represent a bag of words. In addition, individual word $w \in d$ where w is feature words, and each document d is concerned with a class label $c \in C$, where C denotes the class label set.

Support vector machine (SVM) (Keerthi et al. 2001) is a linear classifier and supervised learning scheme. The operation of support vector machine uses the concept of hyper-plane in which kernel function is used for conversion of input vector into hyper-plane. The calculation of SVM considers the maximum distance among different data clusters.

Decision tree J48 model (Hormann, 1964) includes the internal node, branches, and leaf nodes. The topmost root node presents a generalized value whereas the lowest leaf note refers to more specific value. In addition, testing on individual attribute is performed by internal node while the outcome of the test is represented by branches and the class labels are held by leaf nodes. Furthermore, it can be used not only for classification but also for clustering. In classification process, the attribute values of the tuple for which the associated class label is unknown, are tested against the decision tree by tracing from the root to a leaf node and converting to classification rules. Different schemes such as entropy, mutual information, and information gain, are used to calculate quantified information for optimal values induction.

Performance Evaluation Schemes

Three different schemes including accuracy, relative root squared error (RRSE), and computation time (CT) are used to evaluate the proposed system performance. Accuracy is the quantity of confusion matrix in percentage while root relative squared error (RRSE) is the rate of error difference between predicted value and target value. In addition, computation time is the total consumed time for classification model in seconds. The mathematical formulas for accuracy and RRSE are described in Equation (8), and Equation (9) respectively:

$$Accuracy(\%) = \frac{(TP + TN)}{(TP + FP + TN + FN)} * 100\% \quad (8)$$

where TP is the number of true positive, TN is the number of true negative, FP is the number of false positive, and FN is the number of false negative:

$$RRSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{\sum_{i=1}^N (O_i - \bar{O}_i)^2}} \quad (9)$$

where O_i is the predicted values, P_i is the target values, and \bar{O}_i is the mean of target values:

TLM or TCM = Time to build model on full dataset

where TLM or TCM is the total computation time for classification model (seconds).

SYSTEM IMPLEMENTATION

The shortcomings observed from the literatures, the optimal feature subset would be sought on the fly without the need of setting an arbitrary subset length and putting the model under tougher tests of more features. Therefore, system implementation for the proposed model includes several processes like feature extraction, common mechanism of meta-heuristic search, feature selection based on two meta-heuristic algorithms. Dataset and experimental setting for our proposed model are described in the last part of this section.

Term Frequency Inverse Document Frequency Based Feature Extraction (TF-IDF)

In this proposed model, TF-IDF is used to extract important feature in which the weight of the document terms is calculated by the multiplication of term frequency (TF) and inverse document frequency (IDF). Detailed mathematical formula of TF-IDF (Medler, 1986) is shown in Equation (10):

$$TF - IDF_{d,t} = TF_{d,t} * IDF_t = \frac{n_{d,t}}{|d|} * \log \frac{|D|}{|\{d : t \in d\}|} \quad (10)$$

where $n_{d,t}$ is number of occurrences of t in d , $|D|$ is the number of categories and $|\{d : t \in d\}|$ if the number of documents with term t occurrences.

Common Mechanisms of Meta-Heuristic Search

Meta-heuristic search includes four fundamental steps: initialization of population, definition of termination, evaluation of fitness function, updating search agents and retrieving global best solution. Initialization of population size is the first stage in meta-heuristic search in which the number of candidate solution is defined (Yang et al. 2014).

Termination criteria is defined as a point of stopping the search execution of an algorithm. In meta-heuristic algorithms, the termination condition is either one, or is composed of two or more.

The basic common termination criteria are defined by the number of iterations or generations, and objective function.

In this proposed model, selected feature subsets from the meta-heuristic based search are plugged with correlation-based feature subset filter (CFS) (Hall, 1999) to evaluate the quality of feature subset. The objective function for the proposed system is shown in Equation (11), and the output of optimal feature subset is used as a recommended set of features for the classifiers:

$$\text{Objective Function} = \text{Merit}(D) \tag{11}$$

where merit function for correlation-based feature subset can be formulated in Equation (12), and D is the set of total number of features in the hypothesis:

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)\bar{r}_{ff}}} \tag{12}$$

where, M_s refers to the heuristic “merit” function for feature subset S with k features, \bar{r}_{ff} means the average intercorrelation between features, and \bar{r}_{cf} brings up the mean of feature to class correlation in which f belongs to S . Moreover, the numerator measures the predictability of the class set of features; and the denominator calculates the level of redundancy among features of different class. CFS is composed with two acceptance rules including irrelevant feature (which were ignored due to low correlation with the interested), and redundant features (which were screened out due to high correlation of feature with each other). Furthermore, the condition for feature acceptance was dependent on the prediction of class in the scope of instance space, but not considered on other features. The selected feature subset was evaluated by equipped with the classifiers to achieve maximum accuracy with minimum relative square error, and computational cost.

For updating and retrieving global optimal solution, the previous selected features were updated with the best optimal feature for each iteration. The policy of updating solution included two generals: random search mechanism for initialization of the random feature subset; and comparison mechanism to compare the selected randomization of neighborhood candidate subset with an initialized feature

Table 3. Parameter settings for ACO and EO based feature search

Parameters	ACO	EO
Chaotic coefficient	4.0	4.0
Chaotic type	Logistic mapping	Logistic mapping
Iteration	20,30,40,50,60,70,80,90,100,150,200	20,30,40,50,60,70,80,90,100,150,200
Objective type	Merits	Merits
Population size	20,30,40,50,60,70,80,90,100,150,200	20,30,40,50,60,70,80,90, 100,150,200
Start set	1	1
Objective function	Merit	Merit
Search direction	Forward	Forward
Others	Evaporation (0.9)	Mutation Probability (0.01)
	Heuristic (0.7)	Mutation Type (Bit-flip)
	Pheromone (2)	

subset. New global candidate solution was updated, if the value of new subset of neighbor is greater than the old one, and the process was repeated until the end criterion such as the maximum number of iterations with the objective function is achieved.

ACO-Based and EO-Based Optimization of Feature Selection Models

In both ACO and EO-based search process, the merit function is used to guide the search process in filter feature selection approach. Table 3 describes the parameter setting for ACO and EO-based feature search. In the search process of ACO, the number of population and iteration are defined randomly in the initialization process. In this research, 20 to 200 population and iteration values were used for this experiment. For instance, the number of outputs for selected feature subset (380 features = 20 features or population * 19 iteration search) for 20 values in population and iteration in filter feature selection process. The random 20 features were selected from the feature vector and search process was executed until defined maximum iteration (200). In addition, the update process of global optimal feature (shortest path in nature of ant search) is performed by measuring pheromone levels.

The heuristic value guides ACO search for the most promising feature (solution). Objective function (merit) is used heuristic value (0.7) to reach the promising feature value among several local best feature for individual class. In addition, evaporation of pheromone is used to overcome the local optima by eliminating old high pheromone value in local search to favor the exploration search in new areas. In the evaporation value (0.9) is used in this case study and it is about half of the pheromone value for avoiding the convergence to a local optimal feature. In other words, previous path selected by the first ant will continue to be extensive to the following ants if pheromone evaporation is not considered. Although most of the parameters were same as ACO algorithm, searching rules were different for EO. In addition, bit flip mutation was intended for updating individual feature value by comparing two features while the mutation probability (0.01) is used for each bit inversion rate.

Dataset and Experimental Settings

In this section, the dataset used to evaluate the proposed model, and the experimental settings for the list of feature optimizations selection methods are described. In addition, the parameter settings for correlation-based feature subset selection and principal component analysis-based feature reduction are presented. Three learning algorithms for document categorization are also presented.

Table 4 shows dataset information and extracted features using TF-IDF based feature transformation scheme. The proposed model uses news dataset with five labels from BBC website (Greene & Cunningham 2006) with a total number of instances of 2,222. Although dataset was not so large, it showed complexity and high-dimensional data properties because the text dataset was unstructured. Thus, it is suited for problem of multi-dimensional feature since the number of extracted features is more than the total number of instances.

Table 4. Dataset and extracted feature

Target Class	Number of Instances and Data Type
Business	510 [Text, Nominal]
Technology	400 [Text, Nominal]
Entertainment	386 [Text, Nominal]
Politics	417 [Text, Nominal]
Sport	509 [Text, Nominal]
Total number of instances	2,222
Total number of extracted features	2,591

Table 5 describes six pairs of experiments for the proposed model. Ant Colony Optimization algorithm and Elephant Optimization algorithm-based search with correlation-based feature subset filter approach, and three classifiers are run as the experiments.

Table 5. List of optimizations of feature selection methods for experiments

Experiment Abbreviation	Classifier	Metaheuristic Intelligence	Approach
CFS-ACO-NB	Naïve Baye	Ant Colony Optimization	Swarm
CFS-ACO-SVM	Support Vector Machine	Ant Colony Optimization	Swarm
CFS-ACO-J48	Decision Tree	Ant Colony Optimization	Swarm
CFS-EO-NB	Naïve Baye	Elephant Optimization	Nature-inspired
CFS-EO-SVM	Support Vector Machine	Elephant Optimization	Nature-inspired
CFS-EO-J48	Decision Tree	Elephant Optimization	Nature-inspired

Several common filter methods, for example, mutual information, information gain, Relief, Chi², are used in previous era of feature selection. However, correlation-based feature subset selection filter is used in this experiment. Table 6 shows the parameter settings for correlation-based feature selection, and principal component analysis feature reduction scheme. Normally, the nature of filter model does not depend on individual machine learning algorithm, but its evaluation consideration relies on the characteristics of data. Correlation-based feature subset filter includes two main processing units: feature evaluation and feature searching. In the process of feature evaluation, the correlation matrix score is implemented firstly, and merit function uses the correlation matrix score in considering features that are highly correlated with predictive label, but uncorrelated with other labels. In feature searching process, feature ranking is based on certain criteria- either by taking account of feature space or independently of each other. In other word, features from the searching process are sent to the evaluation process in which merit function is used as objective function and individual feature that has a high correlation with the class is predicted and added as selected feature subsets.

Table 6. Correlation-based feature subset filter (CFS) and principal component analysis (PCA)

CFS	Setting
Evaluation	Correlation Matrix Scores
Number of threads	1
PCA	Setting
Maximum attribute names	5
Maximum attributes value	-1
Variance coverage	0.95

Principal component analysis (PCA) is one of the most popular schemes for dimensionality reduction. It is accomplished by selecting enough eigenvectors to account for some percentage of the variance in the original data. In the process of PCA, variance coverage is set up at (0.95) to retain enough principal component attributes to account for the proportion of variance. In the calculation of variance coverage, mean values are calculated to normalize the features, and the standard deviation

between individual variables is figured out to implement the covariance matrix which is used to measure the variance of attributes. Engine value and vector calculation are performed for data transformation and are non-negative descending ordered eigenvalues. The components are chosen to form a feature vector or a column vector. Finally, principal components are formed by multiplying the transformed feature vector with transformed scaled features. In this proposed mode, five maximum number of attributes to include in transformed attribute are also defined. Furthermore, maximum number of principal component attribute is retained till (-1).

Table 7 shows parameter setting for three classifiers in document categorization. In Naïve Baye (NB) classifier, numeric estimator precision values are selected by analyzing the training data. Batch size is the preferred number of instances to process if batch prediction is being performed and is set up with (100) number of batches. In addition, kernel estimator is used rather than normal distribution. In support vector machine, the calibration models use logistic to obtain proper probability estimates. The number of folds of cross-validation is set up with (-1) and thus training data is used to generate calibration models. The epsilon (ϵ) is used for round-off error and the tolerance value is set up at default (0.001). In the parameter settings of J48, binary splits are used on nominal attributes when building the trees. The value of confidence factor (0.25) is used for pruning and taking account of subtree raising operation when pruning. Number of folds determines the amount of data used for reduced-error pruning. Laplace is used to smooth the counts at leaves nodes.

Table 7. Parameter settings for three classifiers in document categorization

Classifier	Parameter	Value
NB	Batch size	100
	Kernel estimator	Yes
SVM	Batch size	100
	Calibrator	Logistic
	Epsilon (ϵ)	0.000000000001
	Kernel	Poly Kernel
	Number of folders	-1
	Tolerance	0.001
J48	Batch size	100
	Confidence factor	0.25
	Number of folds	3

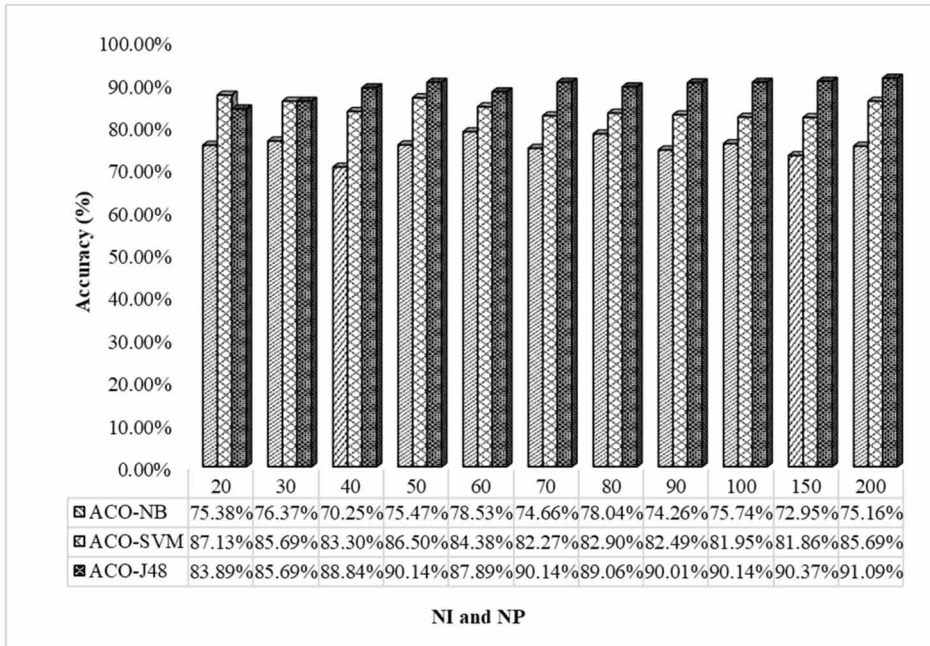
RESULTS AND DISCUSSION

Performance Optimization for Complex Text Feature Selection Using ACO

Figure 2 illustrates the accuracy performance for optimization of feature selection based on ACO search. The selected feature subsets for individual number of population (NP) and number of iteration (NI) in ACO-based search were imported to the classifiers and the performance of the selected feature subsets were evaluated. According to the results, the best accuracy was achieved in J48 classifier evaluation process, while the least accuracy was provided in Naïve Baye.

Though the proposed ACO-based optimization of feature selection method is not synchronized with learning model, accuracy values above 91% were achieved. The highest accuracy of J48 result

Figure 2. Performance optimization results using ACO according to time series (accuracy, %)



(91.01%) was obtained at the highest NI and NP values of ACO-based search. In contrast, the highest accuracy of SVM result (87.13%) was provided at the lowest NI and NP. Meanwhile, the peak accuracy of NB result (78.53%) occurred at medium value of NI and NP (60). In summary, the accuracy peak value varied according to the classifier models, and J48 classifier was the best appropriate classifier when working with ACO-based searching scheme for filter.

On RRSE measurement Figure 3, ACO-based J48 classifier result showed the least error value (42.05%) at NP and NI= 200. The second least error value (49.64%) occurred with ACO-based SVM classifier results, and value was achieved at the lowest NI and NP value. whereas the lowest error result (68.73%) was obtained at NI and NP= 60 for ACO-based Naïve Baye classifier.

According to the computation cost results using ACO in Table 8, the number of relevant features can be reduced drastically when compared to the original number of extracted feature (2,519). The smallest number of feature (382) was obtained at NI and NP= 20, and the highest one (1,015) occurred at highest NI and NP values. The lowest computation cost (CT) (0.18) is provided at the lowest NI and NP values (20) for Naïve Baye classifier, while the second lowest (0.89) was achieved at NI and NP= 30 for ACO-based J48 classifier. However, the highest computation cost occurred with ACO-based SVM classifier with its lowest computation cost (6.92) at NI and NP= 50, which was higher than the highest CT (0.59) at NI and NP= 150 in NB classifier and peak CT value (1.9) at NI and NP= 200 in J48 classifier.

Performance Optimization for Complex Text Feature Selection Using EO

According to the experimental results in Figure 4, the highest classifier results (90.23%) was achieved in EO-based J48 classifier. In contrast, the lowest classifier performance output (77.32%) was obtained in EO-based NB classifier. The classification accuracy results between best and worse is provided in EO-based SVM classifier and the highest value (88.03%) occurred at NI and NP= 30. The lowest accuracy (79.12%) for EO-based J48 and the lowest accuracy for EO-based SVM (80.96%) were

Figure 3. RRSE Results using ACO according to time series

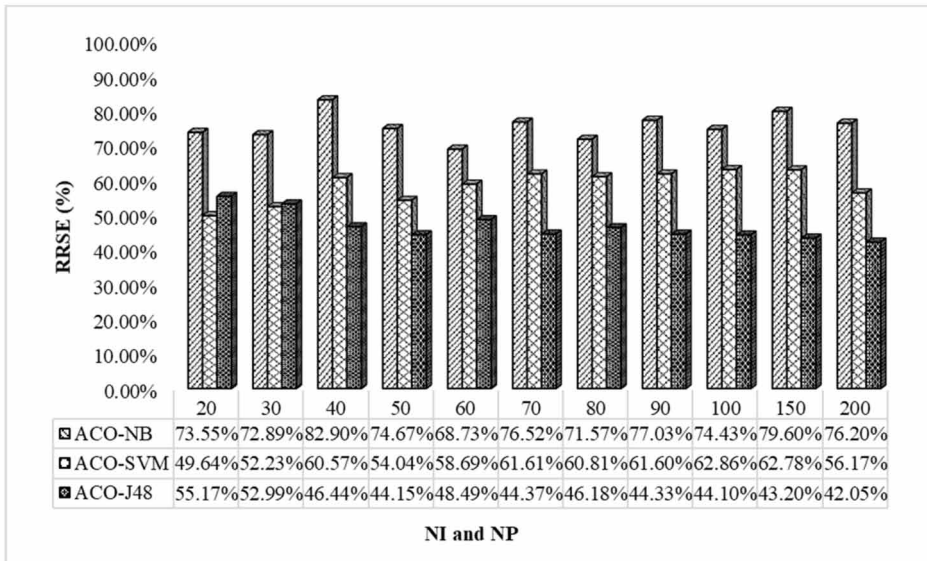
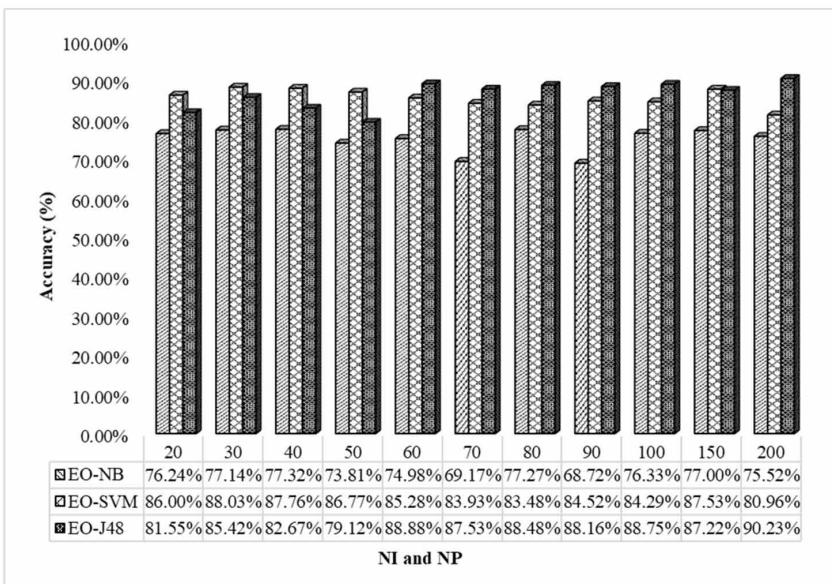


Table 8. Computation cost results using ACO according to time series

NI, NP	20	30	40	50	60	70	80	90	100	150	200
NF	382	482	774	850	674	929	889	904	842	989	1015
CT-NB	0.18	0.31	0.48	0.46	0.38	0.48	0.51	0.51	0.48	0.59	0.58
CT-SVM	9.34	7.4	7.84	6.92	7.38	8.41	8.46	7.96	8.26	9.1	9.78
CT-J48	0.9	0.89	1.52	0.98	1.25	1.63	1.64	1.77	1.58	1.61	1.9

Figure 4. Performance optimization results using EO according to time series (accuracy, %)



higher than the highest accuracy (77.32%) in EO-based NB. The J48 classifier was the best appropriate classifier with EO-based searching scheme for filter.

Figure 5 illustrates the root relative square error (RRSE) classification results for EO-based feature selection. Among three classifiers, J48 classifier results had the least error values, and its minimum error (44.47%) was achieved at NI and NP= 200. However, highest error values are reported in EO-based NB classifier results with peak value (85.89%) at NI and NP= 90. In addition, the lowest error value (71.24%) for EO-based NB classifier was higher than the peak error value (61.37%) of J48 classifier at NI and NP=50. Meanwhile, the error values line for SVM occurred between bottom and peak lines with peak error value of (63.96%) at NI and NP= 200 and its lowest error value of (47.60%) at NI and NP= 30. Maximum error value (63.96%) for SVM was lower than the minimum error (71.24%) for NB classifier.

Figure 5. RRSE results using EO according to time series

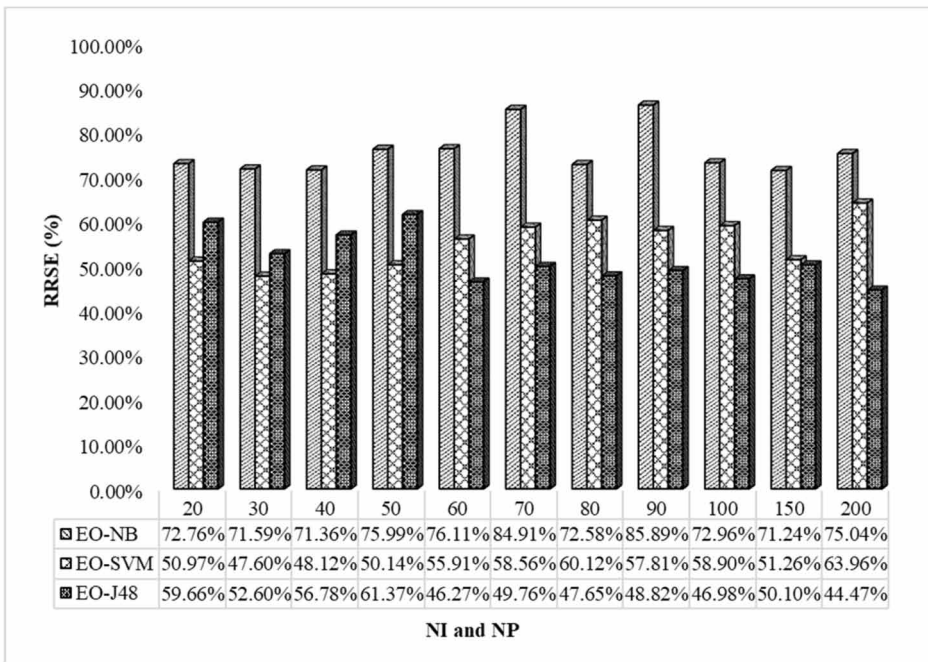


Table 9 shows the computation cost results using EO where the number of relevant features can be reduced to the smallest feature size (263) at NI and NP= 20. Although the highest feature size (976) occurred at highest NI and NP value, the feature size was less than the original feature size (2,591). The lowest computation cost (CT) (0.14) was obtained at the lowest NI and NP values (20) in Naïve Baye classifier, while the second lowest (0.63) was achieved at NI and NP= 50 in EO-based J48 classifier. However, the highest computation cost occurred in EO-based SVM classifier with lowest computation cost (6.41) at NI and NP= 90, which was higher than highest CT (0.51) at NI and NP= 200 for NB classifier and peak CT value (1.64) at NI and NP= 100 for J48 classifier.

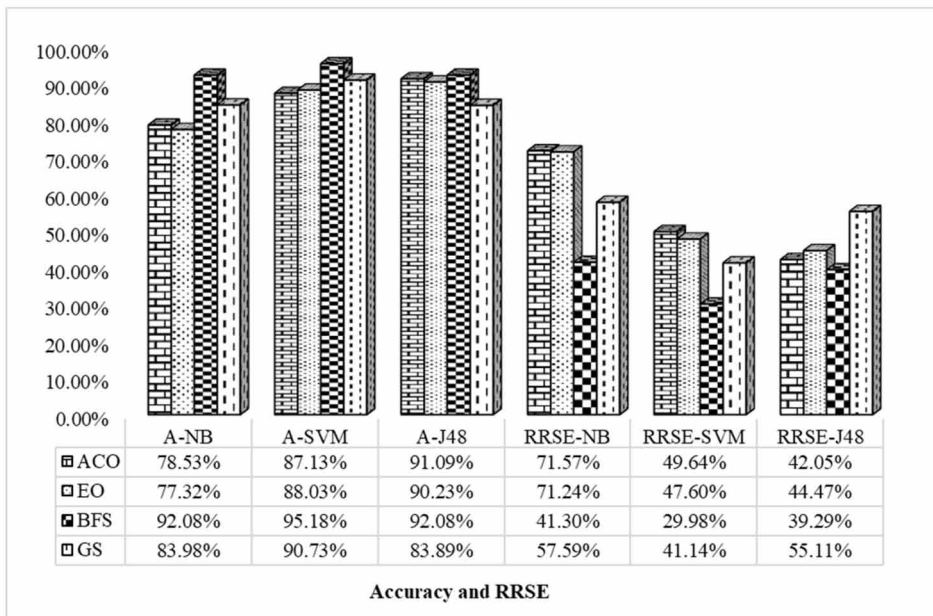
Performance Comparison: Meta-Heuristic Intelligence and Conventional Search

Figure 6 presents the comparative study for accuracy and error rate between meta-heuristic and conventional search. The J48 classifier results indicated that the accuracy results using both meta-

Table 9. Computation cost results using EO according to time series

NI, NP	20	30	40	50	60	70	80	90	100	150	200
NF	263	390	483	281	602	741	766	773	805	549	976
CT-NB	0.14	0.17	0.23	0.17	0.36	0.44	0.44	0.44	0.42	0.47	0.51
CT-SVM	8.36	8.15	7.22	7.7	8.46	12.65	8.55	6.41	7.72	8.43	8.98
CT-J48	0.72	0.67	0.74	0.63	1.1	1.09	1.04	1.06	1.64	0.83	1.57

Figure 6. Accuracy and RRSE results using meta-heuristic and conventional search



heuristic searches were above 90%, (91.09%) in ACO-based search and (90.23%) in EO-based search. Though both results based on meta-heuristic searches were not better than best first search accuracy result (92.08%), the number of selected features was reduced drastically in meta-heuristic based search and better result could be achieved if the searching parameters were adjusted and not fixed. Therefore, it is more appropriate for complex feature selection problem. However, greedy stepwise-based feature selection for document classification result (83.89%) was not better than accuracy results of both meta-heuristic search approaches. In the case of SVM and NB classifiers accuracy results, meta-heuristic based search approaches were not better for NI and NP ranges between 20 and 200 when compared to conventional search approaches. However, the results illustrated in Figure 2 and Figure 4 showed that meta-heuristic based search can vary in accuracy results according to the parameter tunings like NI and NP. This property proof that meta-heuristic based search approaches are more appropriate for complex feature selection but will require auto parameter tuning according to the simulation results from this comparative studies.

Table 10 summarizes the computation costs results for meta-heuristic intelligence and conventional-based search. It includes the number of selected feature (NF) and computation time (CT) measurements. The NF results in BFS and GS approach were like meta-heuristic-based search, but the algorithmic search method, was not appropriate for complex feature because it is based on 2ⁿ

Table 10. Computation cost results: Meta-heuristic intelligence and conventional search

Experiments	ACO	EO	BFS	GS
NF	382	263	102	60
CT-NB	0.18	0.14	0.09	0.03
CT-SVM	6.92	6.41	10.89	6.48
CT-J48	0.89	0.63	0.34	0.13

combinatorial approach. In the case of SVM, the computation cost for both meta-heuristic search (**6.92**) in ACO and (**6.41**) in EO, were better than conventional-based search results in BFS (10.89) and GS (6.48). However, NB classifier was more cost effective computationally for meta-heuristic based search and J48 classifier was the least cost effective. The performance measurement in terms of accuracy and RRSE were the reverse of the cost effectiveness. In the case of BFS and GS, computation costs for NB and J48 were more cost effective than SVM, but an opposite in meta-heuristic based search. A summary of computation cost analysis results showed that SVM classifier was more appropriate for meta-heuristic search in terms of computation cost complexity.

CONCLUSION AND FUTURE WORKS

In this comparative study, two meta-heuristic based filter feature selection approaches were compared with the traditional method. Result analysis for the quality of selected feature subsets in each time series were evaluated using three different classifiers. The numbers of selected features were effectively with six times reduction in the minimum number of iteration and population (20) whereas the maximum number of iteration and population (200) were halved when the proposed system was applied to a real-world text document categorization. However, results of NB and SVM were not better than traditional search-based classification in this comparative study.

For further studies, the wrapper approach should be applied for feature selection in specific application area and the objective function should be synchronized with the classifier performance using the measurements of accuracy, precision, recall, and error rate. In other word, wrapper feature selection approach can be synchronized with individual specific classifier for auto-parameter adjustment in which meta-heuristic based algorithm will be used to search for features that match with specific classifier requirement. In addition, hybrid feature selection scheme with meta-heuristic search or hybrid meta-heuristic search should be further explored and deployed depending on the problem data characteristic. Furthermore, distributed processing unit should investigate feature search in big data problems to achieve low computation cost with good performance.

ACKNOWLEDGMENT

This work was supported by the Department of International Business Management, Didyasarin International College, Hatyai University. The authors wish to thank Dr Ozioma F. Nwabor for his assistance throughout the manuscript preparation phase.

REFERENCES

- Abdel-Basset, M., & Shawky, L. A. (2019). Flower pollination algorithm: A comprehensive review. *Artificial Intelligence Review*, 52(4), 2533–2557. doi:10.1007/s10462-018-9624-4
- Abdollahi, M., Gao, X., Mei, Y., Ghosh, S., & Li, J. (2019). An ontology-based two-stage approach to medical text classification with feature selection by particle swarm optimization. *Proceedings of IEEE Congress on Evolutionary Computation, CEC*, 119–126. doi:10.1109/CEC.2019.8790259
- Ahmad, S. R., Bakar, A. A., & Yaakub, M. R. (2019). Ant colony optimization for text feature selection in sentiment analysis. *Intelligent Data Analysis*, 23(1), 133–158. doi:10.3233/IDA-173740
- Akyol, S., & Alatas, B. (2020). Sentiment classification within online social media using whale optimization algorithm and social impact theory-based optimization. *Physic Statistical Mechanics and Its Applications*, 540, 123094. doi:10.1016/j.physa.2019.123094
- AlFarraj, O., AlZubi, A., & Tolba, A. (2019). Optimized feature selection algorithm based on fireflies with gravitational ant colony algorithm for big data predictive analytics. *Neural Computing & Applications*, 31(5), 1391–1403. doi:10.1007/s00521-018-3612-0
- Alomoush, W., Omar, K., Alrosan, A., Alomari, Y. M., Albashish, D., & Almomani, A. (2018). Firefly photinus search algorithm. *Journal of King Saud University - Computer and Information Sciences*, 32(5), 599–607. doi:10.1016/j.jksuci.2018.06.010
- Appel, E. (2014). Search techniques. *Cybervetting*, 189–208. doi:10.1201/b17651-21
- Babatunde, O., Armstrong, L., Leng, J., & Diepeveen, D. (2014). A genetic algorithm-based feature selection. *International Journal of Electronics Communication and Computer Engineering*, 5(4), 899–905. https://ro.ecu.edu.au/ecuworkspost2013/653
- Beheshti, Z., & Shamsuddin, S. M. (2013). A review of population-based meta-heuristic algorithm. *International Journal of Advanced Soft Computing Application*, 5(1), 1-35.
- Boudia, M. A., Hamou, R. M., & Amine, A. (2018). Comparative study between two swarm intelligence automatic text summaries: Social spiders vs social bees. *International Journal of Applied Metaheuristic Computing*, 9(1), 15–39. doi:10.4018/IJAMC.2018010102
- Brezočnik, L., Fister, I., & Podgorelec, V. (2018). Swarm intelligence algorithms for feature selection: A review. *Applied Sciences (Basel, Switzerland)*, 8(9), 1521. doi:10.3390/app8091521
- Cheng, S., Shi, Y., Qin, Q., & Bai, R. (2013). Swarm intelligence in big data analytics. *International Conference on Intelligent Data Engineering and Automated Learning*, 417–426. doi:10.1007/978-3-642-41278-3_51
- Cherrington, M., Thabtah, F., Lu, J., & Xu, Q. (2019). Feature selection: filter methods performance challenges. *International Conference on Computer and Information Sciences, ICCIS*, 1-4. doi:10.1109/ICCISci.2019.8716478
- Clausen, J., & Perregaard, M. (1999). *Correlation-based feature selection for machine learning*. doi:10.1023/A:1018952429396
- Dadaneh, B. Z., Markid, H. Y., & Zakerolhosseini, A. (2016). Unsupervised probabilistic feature selection using ant colony optimization. *Expert Systems with Applications*, 53, 27–42. doi:10.1016/j.eswa.2016.01.021
- Deb, S., Fong, S., & Tian, Z. (2016). Elephant search algorithm for optimization problems. *The 10th International Conference on Digital Information Management*, 249–255. doi:10.1109/ICDIM.2015.7381893
- Dorigo, M., & Stützle, T. (2000). The ant colony optimization metaheuristic: algorithms, applications, and advances. *Handbook of Metaheuristics*, 250–285. doi:10.1057/palgrave.jors.2602644
- Drotár, P., Gazda, M., & Vokorokos, L. (2019). Ensemble feature selection using election methods and ranker clustering. *Information Sciences*, 480, 365–380. doi:10.1016/j.ins.2018.12.033
- El Aboudi, N., & Benhlima, L. (2016). Review on wrapper feature selection approaches. *Proceedings of International Conference on Engineering and MIS, ICEMIS*, 1–5. doi:10.1109/ICEMIS.2016.7745366

- Ezugwu, A. E. S., Adewumi, A. O., & Fr̄ncu, M. E. (2017). Simulated annealing based symbiotic organisms search optimization algorithm for traveling salesman problem. *Expert Systems with Applications*, 77, 189–210. doi:10.1016/j.eswa.2017.01.053
- Fong, S., Deb, S., Yang, X., & Li, J. (2014). Feature selection in life science classification: Metaheuristic swarm search. *IT Professional*, 16(4), 24–29. doi:10.1109/MITP.2014.50
- Fong, S., Yang, X. S., & Deb, S. (2013). Swarm search for feature selection in classification. *Proceedings of 16th IEEE International Conference on Computational Science and Engineering, CSE*, 902–909. doi:10.1109/CSE.2013.135
- Fong, S., Zhuang, Y., Tang, R., Yang, X. S., & Deb, S. (2013). Selecting optimal feature set in high-dimensional data by swarm search. *Journal of Applied Mathematics*, 2013, 1–18. Advance online publication. doi:10.1155/2013/590614
- Ghosh, M., Guha, R., Sarkar, R., & Abraham, A. (2019). A wrapper-filter feature selection technique based on ant colony optimization. *Neural Computing & Applications*, 1–19. doi:10.1007/s00521-019-04171-
- Goswami, S., Chakraborty, S., Guha, P., Tarafdar, A., & Kedia, A. (2019). Filter-based feature selection methods using hill climbing approach. *Natural Computing for Unsupervised Learning*, 213–234. 10.1007/978-3-319-98566-4_10
- Greene, D., & Cunningham, P. (2006). Practical solutions to the problem of diagonal dominance in kernel document clustering. *Proceedings of the 23rd International Conference on Machine Learning*, 377–384. doi:10.1145/1143844.1143892
- Hall, M. A. (1999). *Correlation-based feature selection for machine learning*. Academic Press.
- Hameed, S. S., Petinrin, O. O., Hashi, A. O., & Saeed, F. (2018). Filter-wrapper combination and embedded feature selection for gene expression data. *International Journal of Advances in Soft Computing and Its Applications*, 10(1), 90–105. <https://core.ac.uk/reader/328816260>
- Hao, H., Liu, C. L., & Sako, H. (2003). Comparison of genetic algorithm and sequential search methods for classifier subset selection. *Proceedings of the Seventh International Conference on Document Analysis and Recognition, ICDAR*, 765–769. doi:10.1109/ICDAR.2003.1227765
- Harish, B. S., Guru, D. S., Manjunath, S., & Kiranagi, B. B. (2010). A symbolic approach for text classification based on dissimilarity measure. *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, 104–108. doi:10.1145/1963564.1963581
- Hormann, A. M. (1964). Programs for machine learning Part II. *Information and Control*, 7(1), 55–77. doi:10.1016/S0019-9958(64)90259-1
- Jiang, Y., Liu, X., Yan, G., & Xiao, J. (2018). Modified binary cuckoo search for feature selection: A hybrid filter-wrapper approach. *Proceedings of 13th International Conference on Computational Intelligence and Security, CIS*, 488–491. doi:10.1109/CIS.2017.00113
- John, G. H., & Langley, P. (2013). Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, 338–345. doi/proceedings/ doi:10.5555/2074158
- Joseph Manoj, R., Anto Praveena, M. D., & Vijayakumar, K. (2019). An ACO–ANN based feature selection algorithm for big data. *Cluster Computing*, 22(2), 3953–3960. doi:10.1007/s10586-018-2550-z
- Kashef, S., & Nezamabadi-Pour, H. (2013). A new feature selection algorithm based on binary ant colony optimization. *Proceeding of the Fifth Conference on Information and Knowledge Technology*, 50–54. doi:10.1109/IKT.2013.6620037
- Ke, L., Feng, Z., & Shang, K. (2010). A multiobjective ACO algorithm for rough feature selection. *Second Pacific-Asia Conference on Circuits, Communications and System*, 1, 207–210. doi:10.1109/PACCS.2010.5627071
- Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation*, 13(3), 637–649. doi:10.1162/089976601300014493
- Kotsiantis, S. B. (2014). Feature selection for machine learning classification problems: A recent overview. *Artificial Intelligence Review*, 42(1), 157–176. doi:10.1007/s10462-011-9230-1

Kudo, M., & Sklansky, J. (2000). Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1), 25–41. doi:10.1016/S0031-3203(99)00041-2

Kumar, A., & Jaiswal, A. (2019). Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter. *Multimedia Tools and Applications*, 78(20), 29529–29553. doi:10.1007/s11042-019-7278-0

Kumar, A., Jaiswal, A., Garg, S., Verma, S., & Kumar, S. (2019). Sentiment analysis using cuckoo search for optimized feature selection on Kaggle tweets. *International Journal of Information Retrieval Research*, 9(1), 1–15. doi:10.4018/IJIRR.2019010101

Kuo, R. J., Huang, S. B. L., Zulvia, F. E., & Liao, T. W. (2018). Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. *Knowledge and Information Systems*, 55(1), 253–274. doi:10.1007/s10115-017-1083-8

Lantt, M., Sungt, S., Lowt, H., & Tant, C. (2005). A comparative study on term weighting schemes for text categorization. *Proceedings of IEEE International Joint Conference on Neural Networks*, 1, 546–551. doi:10.1109/IJCNN.2005.1555890

Li, X., Xie, H., Rao, Y., Chen, Y., Liu, X., Huang, H., & Wang, F. L. (2016). Weighted multi-label classification model for sentiment analysis of online news. *International Conference on Big Data and Smart Computing, BigComp*, 215–222. doi:10.1109/BIGCOMP.2016.7425916

Lin, K. C., Zhang, K. Y., Huang, Y. H., Hung, J. C., & Yen, N. (2016). Feature selection based on an improved cat swarm optimization algorithm for big data classification. *The Journal of Supercomputing*, 72(8), 3210–3221. doi:10.1007/s11227-016-1631-0

Liu, W., & Wang, J. (2019). A brief survey on nature-inspired metaheuristics for feature selection in classification in this decade. *Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control, ICNSC*, 424–429. doi:10.1109/ICNSC.2019.8743245

Liu, X., Zhang, J., & Guo, C. (2013). Full-text citation analysis: A new method to enhance scholarly networks. *Journal of the American Society for Information Science and Technology*, 64(9), 1852–1863. doi:10.1002/asi.22883

Ma, H., Ye, S., Simon, D., & Fei, M. (2017). Conceptual and numerical comparisons of swarm intelligence optimization algorithms. *Soft Computing*, 21(11), 3081–3100. doi:10.1007/s00500-015-1993-x

Maldonado, S., & Weber, R. (2009). A wrapper method for feature selection using support vector machines. *Information Sciences*, 179(13), 2208–2217. doi:10.1016/j.ins.2009.02.014

Manikandan, R. P. S., & Kalpana, A. M. (2019). Feature selection using fish swarm optimization in big data. *Cluster Computing*, 22(S5), 10825–10837. doi:10.1007/s10586-017-1182-z

Medler, J. T. (1986). Term-weighting approaches in automatic text retrieval. *Insect Systematics & Evolution*, 17(3), 323–337. doi:10.1163/187631286X00251

Nakariyakul, S., & Casasent, D. P. (2008). Improved forward floating selection algorithm for feature subset selection. *Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR*, 2, 793–798. doi:10.1109/ICWAPR.2008.4635885

Neumann, L., & Matas, J. (2011). Text localization in real-world images using efficiently pruned exhaustive search. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 687–691. doi:10.1109/ICDAR.2011.144

Nguyen, B. H., Xue, B., & Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. *Swarm and Evolutionary Computation*, 54, 100663. Advance online publication. doi:10.1016/j.swevo.2020.100663

Ozger, Z. B., Bolat, B., & Diri, B. (2017). A comparative study on binary artificial bee colony optimization methods for feature selection. *Focus on Swarm Intelligence Research and Applications*, (2), 109–128. doi:10.1109/INISTA.2016.7571836

Panda, M. (2018). Elephant search optimization combined with deep neural network for microarray data analysis. *Journal of King Saud University - Computer and Information Sciences*. 10.1016/j.jksuci.2017.12.002

- Peng, H., Ying, C., Tan, S., Hu, B., & Sun, Z. (2018). An improved feature selection algorithm based on ant colony optimization. *IEEE Access: Practical Innovations, Open Solutions*, 6, 69203–69209. doi:10.1109/ACCESS.2018.2879583
- Raymer, M. L., Doom, T. E., Kuhn, L. A., & Punch, W. F. (2003). Knowledge discovery in medical and biological datasets using a hybrid bayes classifier and evolutionary algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, Cybernetics*, 33(5), 802–813. doi:10.1109/TSMCB.2003.816922 PMID:18238233
- Remeseiro, B., & Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Computers in Biology and Medicine*, 112, 103375. doi:10.1016/j.compbiomed.2019.103375 PMID:31382212
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104. doi:10.1016/j.future.2020.01.005
- S., B., & B., M. (2017). A comprehensive survey on various feature selection methods to categorize text documents. *International Journal of Computer Applications*, 164(8), 1–7. 10.5120/ijca2017913711
- Şahin, D. Ö., & Kılıç, E. (2019). Two new feature selection metrics for text classification. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 60(2), 162–171. 10.1080/00051144.2019.1602293
- Schiezaro, M., & Pedrini, H. (2013). Data feature selection based on artificial bee colony algorithm. *EURASIP Journal on Image and Video Processing*, 1(1), 47. doi:10.1186/1687-5281-2013-47
- Shang, C., Li, M., Feng, S., Jiang, Q., & Fan, J. (2013). Feature selection via maximizing global information gain for text classification. *Knowledge-Based Systems*, 54, 298–309. doi:10.1016/j.knsys.2013.09.019
- Sharma, M., & Kaur, P. (2020). A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. *Archives of Computational Methods in Engineering*, 1–25. doi:10.1007/s11831-020-09412-6
- Shehab, M., Khader, A. T., & Al-Betar, M. A. (2017). A survey on applications and variants of the cuckoo search algorithm. *Applied Soft Computing*, 61, 1041–1059. doi:10.1016/j.asoc.2017.02.034
- Shunmugapriya, P., & Kanmani, S. (2017). A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm and Evolutionary Computation*, 36, 27–36. doi:10.1016/j.swevo.2017.04.002
- Solorio-Fernández, S., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. (2019). A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 53(2), 907–948. doi:10.1007/s10462-019-09682-y
- Spärck Jones, K. (2004). A statistical interpretation of term specificity and its application in retrieval. *The Journal of Documentation*, 60(5), 493–502. doi:10.1108/00220410410560573
- Uzer, M. S., Yilmaz, N., & Inan, O. (2013). Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification. *TheScientificWorldJournal*, 2013, 1–10. Advance online publication. doi:10.1155/2013/419187 PMID:23983632
- Vassiliadis, V., & Dounias, G. (2009). Nature-inspired intelligence: A review of selected methods and applications. *International Journal of Artificial Intelligence Tools*, 18(4), 487–516. doi:10.1142/S021821300900024X
- Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing & Applications*, 24(1), 175–186. doi:10.1007/s00521-013-1368-0
- Vikhar, P. A. (2017). Evolutionary algorithms: A critical review and its future prospects. *Proceedings of International Conference on Global Trends in Signal Processing, Information Computing and Communication, ICGTSPICC*, 261–265. doi:10.1109/ICGTSPICC.2016.7955308
- Vizine, A. L., De Castro, L. N., & Gudwin, R. R. (2005). Text document classification using swarm intelligence. *International Conference on Integration of Knowledge Intensive Multi-Agent Systems, KIMAS'05*, 134–139. doi:10.1109/KIMAS.2005.1427067
- Xu, Z., Lyu, M. R., King, I., & Cse, K. (2009). Non-monotonic feature selection. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1145–1152. doi:10.1145/1553374.1553520

- Xue, B., Zhang, M., Browne, W. N., & Yao, X. (2016). A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation*, 20(4), 606–626. doi:10.1109/TEVC.2015.2504420
- Yang, X. S., Chien, S. F., & Ting, T. O. (2014). Computational intelligence and metaheuristic algorithms with applications. *TheScientificWorldJournal*, 1-4, 1–4. Advance online publication. doi:10.1155/2014/425853 PMID:25610904
- Yang, X. S., & He, X. (2013). Bat algorithm: Literature review and applications. *International Journal of Bio-Inspired Computation*, 5(3), 141–149. doi:10.1504/IJBIC.2013.055093
- Yang, X. S., & Karamanoglu, M. (2013). Swarm intelligence and bio-inspired computation: An overview. *Swarm Intelligence and Bio-Inspired Computation*, 3–23. 10.1016/B978-0-12-405163-8.00001-6
- Yang, Y. (1995). Noise reduction in a statistical approach to text categorization. *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 256–263. doi:10.1145/215206.215367
- Yücelbas, Ş. (2020). Simple logistic hybrid system based on greedy stepwise algorithm for feature analysis to diagnose Parkinson's disease according to gender. *Arabian Journal for Science and Engineering*, 45(3), 2001–2016. doi:10.1007/s13369-020-04357-1
- Zhang, H., & Sun, G. (2002). Feature selection using tabu search method. *Pattern Recognition*, 35(3), 701–711. doi:10.1016/S0031-3203(01)00046-2
- Zhang, L., Mistry, K., Lim, C. P., & Neoh, S. C. (2018). Feature selection using firefly optimization for classification and regression models. *Decision Support Systems*, 106, 64–85. doi:10.1016/j.dss.2017.12.001

Khin Sandar Kyaw (PhD) is a Computer Engineer from the Prince of Songkla University, Thailand. She is currently a full-time lecturer at the International Business Management Department, Didyasarin International College, Hatyai University, Thailand. Her research interests are Data Mining, Nature-inspired and Swarm-based Optimization Algorithms, Natural Language Processing, Business Intelligence, and Web Development.

Somchai Limsiroratana attained the B.Eng degree in Electrical Engineering from Prince of Songkla University in 1991, the M.Arg and Dr.Arg degrees from division of environmental science and technology, Kyoto University in 2000 and 2005 respectively for the detection of fruits on natural background research. He has been working at Department of Computer Engineering, Prince of Songkla University since 1991. His research interests are agricultural image processing, medical image processing, optimization, and AI.

Tharnpas Sattayaraksa (PhD) is currently a full-time lecturer at the International Business Management Department, Didyasarin International College, Hatyai University, Thailand. His research interest is in business administration, finance, and industrial engineering.