

Swarm Intelligence Based Feature Selection for High Dimensional Classification: A Literature Survey

Thinzar Saw^{a*}, Phyu Hnin Myint^b

^{a,b}*Univeristy of Computer Studies, Mandalay, Myanmar*

^a*Email: thinzarsaw@gmail.com*

^b*Email: phyuhninmyint@gmail.com*

Abstract

Feature selection is an important and challenging task in machine learning and data mining techniques to avoid the curse of dimensionality and maximize the classification accuracy. Moreover, feature selection helps to reduce computational complexity of learning algorithm, improve prediction performance, better data understanding and reduce data storage space. Swarm intelligence based feature selection approach enables to find an optimal feature subset from an extremely large dimensionality of features for building the most accurate classifier model. There is still a type of researches that is not done yet in data mining. In this paper, the utilization of swarm intelligence algorithms for feature selection process in high dimensional data focusing on medical data classification is form the subject matter. The results shows that swarm intelligence algorithms reviewed based on state-of-the-art literature have a promising capability that can be applied in feature selections techniques. The significance of this work is to present the comparison and various alternatives of swarm algorithms to be applied in feature selections for high dimensional classification.

Keywords: Feature Selection; Swarm Intelligence; High Dimensional Classification.

1. Introduction

The rapid development of modern technology had led to the generation of huge amounts of data that involve a large number of features and data obtained from biomedical field, for the purpose of early diagnosis of diseases. These data often have the characteristics of high dimensions, which pose a high challenge to existing data mining and machine learning algorithms. However, a huge number of non-relevance and repetitive attributes exist in original dataset. Also, the large dimensionality and little instance convey extraordinary troubles to the data preprocessing. Subsequently, the researchers have developed various methodologies [1–4] to arrangement with these issues.

* Corresponding author.

When the number of features in data increases, the computational cost also increases. To overcome this problem, it is necessary to search a way to reduce the dimensionality of data in consideration. In general, there are two processes often used: feature (subset) selection, and feature extraction.

Feature selection (FS) works by eliminating the features that are not relevant or are redundant. By removing these features, FS can reduce the number of features of the data, increase the learning process, clarify the learned model, and/or improve the performance [5,6]. On the other hand, feature extraction (or feature construction) [7–8], creates novel variables as combinations of others to moderate the dimensionality strictly connected to feature selection. The key distinction is that feature selection chooses a subset of first features while feature extraction makes new features from the original features. Feature selection technique is mainly addressed in this paper [10].

FS can be defined as the process of finding the best informative feature subsets in order to escape the scourge of dimensionality and maximize the classification accuracy. The idea for feature selection is to be able to determine the best attributes that are suitable to the classification task. A comprehensive search for the best attribute subset of a given dataset is essentially unbearable in most circumstances. The various kinds of search techniques such as complete search, greedy search, heuristic search, and random search [5,11,12] have been applied to feature selection process.

However, most existing feature selection strategies still undergo from declining in local optima and high computational cost [13,14]. Therefore, an effective search procedure is needed to better solve feature selection problems. In recent years, swarm intelligence (SI) has emerged as a fast and reasonably precise method in solving complex search problems, such as classifying high-dimensionality medical data. However, there aren't any comprehensive views on the strengths and weaknesses of these approaches at the side of their most appropriate application areas. Therefore, this paper aims to study SI based feature selection techniques for high dimensional medical data with thousands or tens of thousands of features, with the desire selection only a small feature subsets and attaining better classification accuracy than usage of all features. The process of studying the articles is by considering into how the algorithms adjusted to the problem in attribute selections specifically in improving the classification accuracy, reducing the numbers of features without conceding the quality of the results and evaluating the performance in term of error rate and processing time. The rest of this paper is structured as follows. Section II defines the basic concept of feature selection (FS) and swarm intelligence (SI). Section III reviews the typical SI algorithms for feature selection. Section IV discusses the related works on SI used in feature selection for high dimensional classification. Lastly, conclusion is described in Section V.

2. Background

2.1. Feature Selection (FS)

Feature selection is a challenging task due primarily to a huge search area, wherever the full range of probable solutions is 2^n for a dataset with n attributes [5,6]. The task is flatter more interesting as n is growing in several areas with the advances in the data collecting techniques and therefore the improved complexity of these issues. It's been applied to be a useful and cost-effective approach to arrange large-dimensional data for data

mining and machine learning. The general procedure for feature selection technique is illustrated in Figure 1 [5].

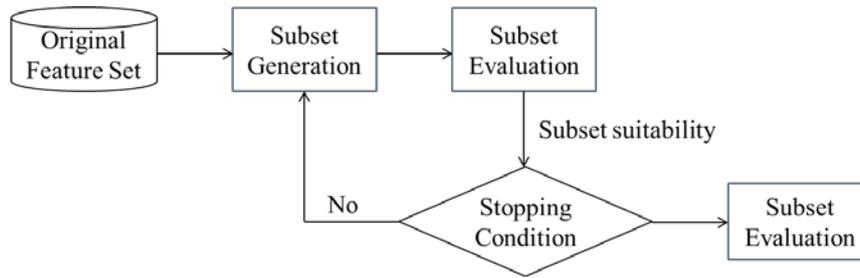


Figure 1: General Feature Selection Process [5]

Feature selection is the course of finding a subset of related attributes in an original dataset. It enhances the classification accuracy by eliminating unrelated and unnecessary feature. It can be distinguished three models according to their relationship with learning methods: filter, wrapper and embedded methods.

Filter methods evaluate how much the attributes are relative to the problem by checking properties of the data without developing learning algorithms. This method is extremely effective [15]. Generally, a feature relevancy score is calculated and then the small scoring features are removed. Subsequently, high ranked features are performed as an input to the classification process. Compared with wrapper model, it has the less computational cost. The core shortcoming of this filter method is that it neglects the result of the features subset within the induction rule, as expressed in [16].

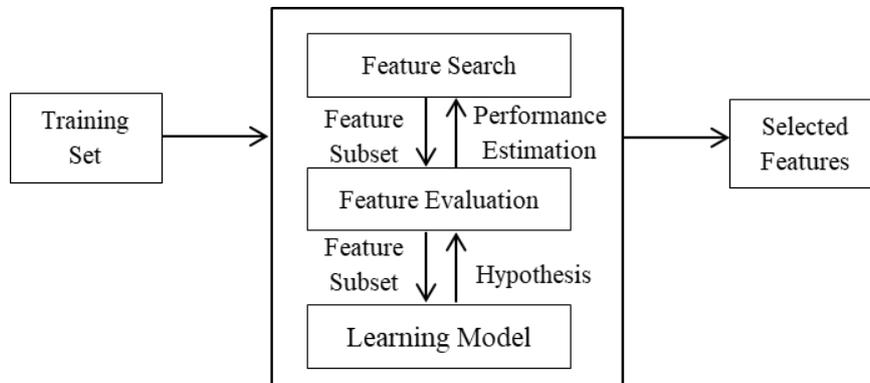


Figure 2: A general framework of wrapper models [15]

Wrapper methods select the features by taking into account each probable subset of them and the features subsets are ranked in keeping with their extrapolative influence, whereas considering the inbuilt classifier as a recorder. It considers the relevance and redundancy information of the features in additionally to their impacts at the predictive strength. It makes use of a specific learning algorithm to assess the value of the chosen attributes. Specified a predefined learning rule, a general work flow of the wrapper techniques is shown in Figure 2 [15].

Compared with filter methods, wrapper methods attain higher prognostic accuracy estimations, due to the fact that they do not forget the biases of the learning algorithms. However, wrapper methods are very costly over

computation [16].

Embedded model carry out feature selection as portion of the learning process and are typically precise to specified machine learning model [17]. This approach is a trade-off between the two methods by inserting the feature selection into the model building. Accordingly, this model yield improvement of both filter and wrapper models: they are so much less computationally rigorous than wrapper methods, due to the fact that they don't need to run the gaining knowledge of models frequently to evaluate the features, and they encompass the interaction with the training model [15].

2.2. Swarm Intelligence (SI)

Swarm intelligence (SI) system consists typically of a population of uncomplicated agents interacting nearby with each other and with their surroundings. The inspiration regularly derives from nature, mainly biological systems. The agents trail quite simple procedures, and although there is no centralized management structure uttering however individual agents ought to behave. Emergence of "intelligent" worldwide behavior might be additionally caused to the interactions between agents. SI algorithms such as Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC), Particle Swarm Optimization (PSO) and other systems prove its powerful ability in biological issues. These algorithms have currently been proven to supply suitable outcomes in a huge type of actual-world applications [18].

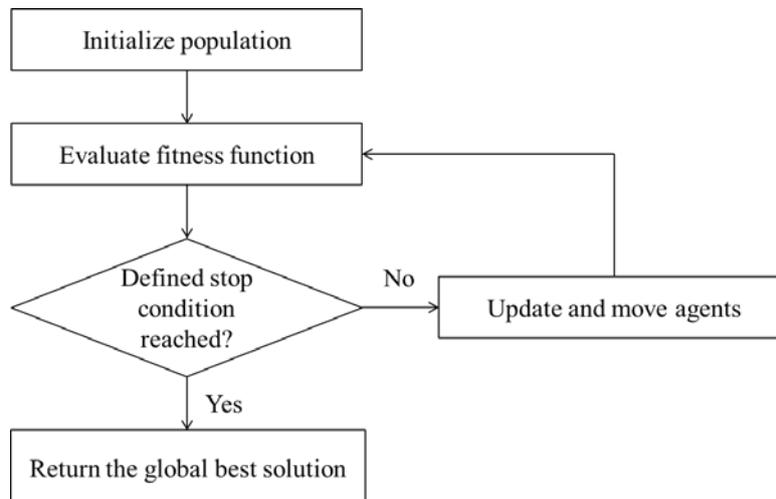


Figure 3: General Framework of SI algorithm [19]

Figure 3 presented the general framework of SI algorithms, in which the look for best solution is performed through swarm of agents. Every agent holds a candidate solution. The search starts with random initialization of the agents in line with the matter. This can be tracked with the aid of assessment of the quality of the candidate solutions proposed by the agents. The third stage is that the new subset generation that differs from one algorithm to any other. The generation of novel candidate subsets tracks the algorithms' sources of inspiration [19].

The evolution of swarm algorithms starting from earliest algorithm which is PSO followed by ACO, AFSA, ABC, Flower Algorithm (FA), Bats Algorithm (BA) and Salp Swarm Algorithm (SSA), the recent algorithms. The basic concept of each algorithms been inspired by the biological behavior of the swarms colony are also highlighted. The strengths and drawbacks for each algorithm have been pointed out based on previous literature. The literature review stated in this paper is primarily focused on fifteen tasks issued among 2010 and 2018. It is perceived that PSO, ACO and ABC are the widespread choice of meta-heuristic search algorithms applied to high dimensional classification [20].

3. Typical swarm intelligence algorithms for feature selection

3.1. Particle Swarm Optimization (PSO)

PSO is a standout amongst the most powerful and dominant swarm intelligence techniques [18] introduced by Kennedy and Eberhart in 1995 [21], which is inspired by social behaviors found in a flock of migrating birds or fish tutoring for tackling optimization problems. In PSO, a flock contains many entities referred to particles interconnecting among group to explore for the best solutions while moving in the large search space. Each particle contains a position, a candidate solution of the problem and a velocity, normally an n-dimension vector of numerical values. The velocity additionally has indistinguishable structure from position, which denotes the speed and direction that the particle move in the subsequent iteration. In each repetition, the velocity of a particle is primarily updated based on their own experience (local best, pbest) and the experience of other around them (global best, gbest). Equation (1) and (2) are utilized to update the velocity and position of every particle.

$$v_{id}(t+1) = \omega \times v_{id}(t) + r_1 c_1 (p_{id}(t) - x_{id}(t)) + r_2 c_2 (p_{gd}(t) - x_{id}(t)) \quad (1)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (2)$$

where $v_{id}(t)$ and $x_{id}(t)$ are velocity and position of particle i in d dimension at time t , respectively. p_{id} and p_{gd} are local best and global best positions in dimension d . c_1 and c_2 are two positive coefficients named learning factors, and r_1 and r_2 are two arbitrary function uniformly distributed in $[0, 1]$. The operator ω is the inertia weight employed as an improvement to control the impact of the previous history of velocities on the current velocity and also plays the role of balancing the global search and the local search. A predefined maximum velocity, v_{max} to the range $[-v_{max}, v_{max}]$ commonly limits the velocity values.

One interesting characteristic of PSO is that it does not use the gradient of the function, thus, objective functions need not to be differentiable. Moreover, the basic PSO is astonishingly simple. Adding variants to the original implementation can help it adapt to more complicated problems. PSO has been successfully applied in many areas other than bioinformatics including industrial applications and power systems.

3.2. Ant Colony Optimization (ACO)

ACO become applied through Dorigo and his colleagues [9] to catch the shortest path between the nest of ants and a supply of food. This is performed utilizing pheromone trails, which ants store at whatever point they

travel, as a shape of indirect communication. ACO is additionally created to solve discrete combinatorial optimization problems such as the travelling salesman problem (TSP). A consideration lot of bioinformatics complications like the sequence alignment, gene mapping and feature selection of the gene expression data are much like TSP. This causes ACO reasonable for large-scale bioinformatics optimization problems [19].

The primary thought is to build the problem to be comprehended as a search for an ideal path in a weighted graph, known as construction graph, and to apply synthetic ants to look for excellent paths. A key point in the improvement of any ACO algorithm is to decide the fitness feature based totally on which the components of a problem's construction graph will be merited with an abnormal state pheromone path, and to decide how ants will abuse these potential components when constructing new solutions.

Some weak point of current adaptation of ACO are as follows: 1) Space complexity of enfolding pheromone in core memory is extraordinary and the most outstanding current solution to resolve this trouble is utilizing candidate pheromone values which is excessively unyielding in numerous conditions. 2) Recent variants of ACO utilize the local search procedures in their standard algorithm; however they are not ready to utilize pheromone information to work all the more effectively. 3) The time complexity of choosing next move is great. Therefore, Ismkhan [23] expressed these weaknesses with more subtleties and furthermore proposes some new and successful techniques to dispense with these disadvantages. These strategies improve the overall performance of ACO and empower it to apply to expansive scale problems.

3.3. Artificial Bee Colony (ABC)

ABC is one of the most recent swarm search optimization algorithms. It was first proposed by Dervis Karaboga in 2005 [24]. It is an optimization algorithm inspired by the intelligent foraging conduct of a honey bee swarm for locating an ultimate solution. This algorithm is professed to be as basic and easy to implement as PSO and DE [25]. The ABC consists of three groups of bees which might be the onlooker bees, scouts and employed bees.

Every one of these bees has different function appointed to them so as to finish the algorithm's procedure. The employed bees recognize a food source and retain the locality of that food source in their recollections. The employed bees share the information with onlooker bees which remain within the hive. The onlooker bee accepts the information of the food source from the employed bee in the hive. From that point forward, one of the food sources is chosen to assemble the nectar. The scout bees find new food sources in the encompassing territory of hive and the new nectar. When a passerby onlooker bee and scout bee choice a food source they turn into employed. The following equation (3) is used to calculate the likelihood estimation of the food source.

$$P_i = \frac{fit_i}{\sum_{n=1}^N fit_n} \quad (3)$$

where N denotes the number of food sources and represents the fitness value of the solution. On the opposite hand, an employed bee is turned into a scout or onlooker bee while their food sources grow to be empty. A new candidate of food source is calculated by applying equation (4):

$$x_i^j = x_{min}^j + rand(0,1) * (x_{maz}^j - x_{min}^j) \quad (4)$$

where x_i^j represents the location of the food source and x_{maz}^j and x_{min}^j represents the lower bound and upper bound of the j dimension respectively. The feasible explanation to a problem below deliberation is signified via the food source in that algorithm in an iterative search procedure. In ABC, exploration is accomplished by using scout bees where the employed and onlooker bees are accountable for exploitation [26].

Comparing ABC algorithm with PSO and ACO methods, it carries the advantages of fast convergence, high flexibility, and strong robustness.

4. Feature Selection using SI in high dimensional classification

In this sector, the related works in feature selection using swarm intelligence on high dimensional classification is studied. As stated by means of the swarm search algorithms applied, the related works reviewed are grouped.

4.1. PSO used in Classification

Current research has revealed that PSO is a capable method to feature selection. However, it also certainly gets stuck in local optima, especially for gene selection problems with a large search space. Therefore, various approaches have been proposed to solve this problem.

In [27], hamming distance is applied compared to normal Euclidian distance measurement. Two different objects may look-alike in an enormous feature space; therefore, the work claimed that Euclidian distance might not be suitable for high dimensional data. The important features subset was selected by the velocity update of particle in binary PSO framework (HDBPSO). In this work, the velocity is updated by the adapted Hamming distance as a proximity measure. The results are evaluated based on validity indices and classification accuracies. Binary data were used in the proposed method for the comparison with GA and multi-objective GA. The results shows HDBPSO performs excellently as compared to others.

Yiyuan Chen and his colleagues [3] suggested a confidence based and cost effective feature selection (CCFS) technique using binary PSO. In CCFS, two novel points are performed: feature confidence and feature cost. Among these two ideas, feature confidence is employed to update the position of particle and enhance the performance of the FS. Feature cost is integrated into the scheme of the fitness function. The results through experiments on Lung Cancer dataset of the proposed method is compared with three different filter methods such as PCFS, InfoGain and CFS and wrapper methods such as BestFirst, GSFS and GSBS methods. The experimental results demonstrated that the proposed CCFS method points to improve the learning accuracy rate and decrease the number of selected features and the total costs of them.

Binh et. al [28] developed a new hybrid FS algorithm in a single evolutionary process to attain smaller feature subsets with well accuracy in a shorter time. In this approach, PSO based local search heuristic considering symmetric uncertainty measure is proposed to enhance the solution and a new hybrid fitness function is applied to calculate the goodness of the chosen features. The experimentation results on eight high dimensional datasets

demonstrated that the lesser feature subsets is obtained with considerably better classification accuracy than the original feature sets. The proposed hybrid FS algorithm nominated more related features in a shorter time with a reduced computational cost when compared to other PSO based FS methods. Additionally, the authors also suggested that the balance among the exploration and exploitation in PSO search towards improving the performance of PSO in FS on high dimensional data.

Threshold Controlled Binary Particle Swarm Optimization (TC-BPSO) along with Multi-Class Support Vector Machine (MC-SVM) is developed in [29] to make the most of the classification accuracy at the cost of least possible selected features. For the selection features, TC-BPSO is used while the MC-SVM is utilized to compute the classification accuracy. The goal of this FS algorithm is to eliminate the irrelevance and repetitive features in order to decrease the feature size. The results gained through experimentations on ten different datasets showed that the suggested method overtakes in all cases such as feature selection, accuracy and computational complexity. Feature selected through the proposed method are computationally faster than the Conventional BPSO along with the existing methods.

4.2. ACO used in Classification

Ant algorithms are easy to implement and cover wide range of applications [22], but their performance dramatically decreases in dealing with large-scale problems. The proposed algorithm in [30] is a new hybrid method, which combines ACO based local search with symmetric uncertainty measure to find the optimal subset of the features. The main goal of this hybrid FS algorithm is to evaluate the best features. Experimentation results on the large-scale gene expression datasets showed that the proposed method ACO-LS performs better when compared with existing methods; these results are measured in terms of prediction accuracy, smaller feature size and effectiveness. Additionally, the approach also considered the balance among local and genetic search towards improving the search quality and effectiveness of ACO-LS.

A hybrid Ant Bee Algorithm (ABA) which combines Ant Colony Optimization (ACO) with Artificial Bee Colony (ABC) algorithm in fuzzy expert system for encoding the solution variables using a modified form of representation was introduced in [31]. The optimal rule set of the combinatorial optimization is formed by the implementation of ACO in proposed work. The representation of the membership function as continuous number is done by ABC. This hybrid method is implemented on several gene expression data sets which include Receiver Operating Characteristic (ROC) analysis have been done to every datasets. To compare the performance of proposed approach with other algorithms, the value of area under ROC curve is used. The results obtained experiments is reported to have the best value when compared to BCGA, RCGA, PSO and GA.

Amirreza Rouhi and his colleagues [32] developed a hybrid method based on binary ant colony algorithm (BACO) to reduce the dimensionality of features by combining number of filter methods and then Advanced BACO meta-heuristic is applied to the reduced feature sets to select the best feature subset. Five well-known high-dimensional microarray datasets was implemented to measure the performance of the proposed method. In this paper, classification error rate and number of selected features as the measures of evaluating and comparing the performance of tested algorithms. The results obtained by Naïve Bayes classifier demonstrated the

effectiveness of the proposed method for high dimensional microarray data.

4.3. ABC used in Classification

ABC algorithm was selected for gene selection problems in various works. Its prevalence is also contributed by its rare parameters contrasted with different optimization algorithms. ABC algorithm was utilized to discover a subset of genes which is then used to train distinctive Artificial Neural Network (ANN); Multilayer Perceptron (MLP), Radial Basis Function Neural Network (RBF) and SVM [19].

In [33] M. S. Uzer and his colleagues proposed a hybrid approach that uses the artificial bee colony (ABC) algorithm for feature selection to select the related features in a shorter time and to reduce the dimension of the feature vector and then support vector machines (SVM) is used for classification to determine the accuracy rate. The main purpose of this paper is to test the effect of elimination of the unimportant and redundant features of the datasets to enhance the classification accuracy. *k*-fold cross-validation was used for the classifier reliability improvement. For the diagnosis of hepatitis, liver disorders, and diabetes datasets from the UCI database, the proposed system reached classification accuracies of 94.92%, 74.81%, and 79.29%, respectively.

Minimum Redundancy Maximum Relevance (mRMR) method joined with ABC algorithm (mRMR-ABC) is presented in [34]. mRMR is employed to evaluate the relevancy and redundancy of features used for continuous and discrete datasets. This technique can be determined the best features set and simplify the classifier to be trained correctly. Fitness value of the ABC algorithm applied in this effort can be defined by SVM classifier for classification accuracy. The experimental results showed that the suggested approach achieves enhanced performance and superior improvement using small number of prognostic genes when verified using microarray datasets and compared to formerly recommended approaches.

Beatriz and his colleagues [35] proposed a method for classifying DNA array. In this proposed approach, firstly, swarm intelligence algorithm is performed on feature selection process to find the subset of attribute (genes). After that, different ANNs are trained with selected feature subsets. Lastly, four different datasets is demonstrated to valid the accuracy and test the relevance of features to correctly classify the data. Through several experiments, ANN attained better results contrast with the MLP, SVM and the RBF neural networks touched with the distance classifier. Moreover, it is established that the suggested method is capable of choosing the right feature sets to identify, forecast and classify a certain disease with better accuracy. In [36], a multi-objective ranking binary artificial bee colony algorithm based on decomposition (MORBABC/D) method is planned to select the optimal subset of feature from the original high dimensional data. Fisher-Markov selector is applied to take a fixed number of microarray data. After that, a novel binary update approach is suggested to trade off the exploration and exploitation facility. Finally, the proposed method is performed for selection of feature, and exciting learning machine is worked as the classifier with ten-fold cross-validation techniques. The proposed approach is evaluated on eight microarray dataset in order to show the effectiveness and efficiency of the algorithm. The results through experiments presented that the proposed approach can effectively make simpler feature selection by selecting a smaller number of attributes desired and provide advanced classification accuracy matched with other state-of-the-art algorithms.

4.4. High Dimensional Classification in FS

In high dimensional classification tasks, instances belong to one of several classes such as cancer dataset or normal dataset, the main objective is to classify these instances and produce classified instances based on related measurements. Training classifiers on such high-dimensionality small-instance sets is a challenging problem that has received an increasing attention from the research community. To address this challenge, feature selection as a pre-processing step is performed and followed it by applying a classification algorithm that trains model complexity through regularization [37]. Feature selection technique improves the performance of large dimensionality dataset in terms of evaluation measures such as accuracy, stability, and specificity. In addition, the classification performance may rely on the performance of the feature selection process. Therefore, a comparative study is made to analyze the state-of-arts research experiments for this task. Table (1) illustrates the comparative study of swarm intelligence algorithms for feature selection on high dimensional datasets that have been used in the literatures. In brief, although PSO or other SI techniques have proven some achievement on solving high-dimensional classification for feature selection, there are still some potential restrictions, which include time-consuming and stagnation into local optima. Meanwhile, for the reason that many high-dimensional datasets, particularly gene expression data, normally have a few number of instances, n-fold cross validation is required. How to carry out n-fold cross validation in multiple runs is tough to design and computationally costly. In this way, investigating a powerful and efficient feature selection technique for high-dimensional data and the impact of feature selection bias are as yet open issues [38].

5. Conclusion and Recommendation

This survey shows a study on different existing feature selection techniques that have been applied on high dimensional classification. Moreover, the critical problem of swarm intelligence based feature selection approach for high dimensional classification found in literature is reviewed. Basic concept of feature selection and the comparative study of different meta-heuristic approaches used in feature selection process are discussed as well. From the stated feature selection techniques above, swarm search algorithms is more suitable for better performance specifically in high dimensional classification, which is considered as an important portion that may affect any learning model. The study shows that feature selection as a critical step in data mining and machine learning applications and it provides an effective way to analyze the high dimensional data by reducing not relevance and redundant data. However, the selection of the correct FS algorithm constitutes an enormous challenge for machine learning problems. To address this problem, many FS approaches have been suggested in recent year. In this paper, the usage of PSO, ABC, ACO meta-heuristic algorithms for high dimensional datasets is presented. Based on the related works in feature selection, the swarm intelligence algorithms gave to enhance the classification accuracy with the smaller number of selected attributes for handling various types of high dimensional data for classification especially biomedical data. According to the comparison table, meta-heuristic search algorithms like PSO, ACO, etc., is mostly used for feature selection method to search large space. This study only acquires from the previous literatures and also considerably clarifies the feature selection methods that most applied for high dimensional classification. However, this paper does not emphasize about the embedded based feature selection techniques which are used on high dimensional classification.

Table 1: A comparative study on feature section methods for high dimensional classification using swarm intelligence

Sr. No.	Year	Methods used	Dataset used	Performance Evaluation
1.	2013	ACO as feature selection method based on rough set theory (RST)	Breast cancer	Proposed model give the low number of informative genes
2.	2013	Hybrid system between swarm intelligence algorithm ACO as a feature selection method and SVM as a classifier	Colon dataset, CNS dataset, Lung cancer dataset as a binary class samples and Glioma dataset as a multiclass dataset	By trying different number of features in many trials. It improves the SVM classifier in different datasets.
3.	2014	Hybrid system between decision tree (DT) as a classifier and particle swarm optimization (PSO) as a gene selection method	Brain tumor, leukemia, SRBCT, LUNG cancer, prostate tumor	Performance of PSODT proves that it can earn better percentage in classifying gene related to different kinds of cancer.
4.	2014	PSO approach with both local search and reset global _{best} (PSO-LSRG) as a feature selection and KNN with LOOCV as a classifier	DLBCL, 9 Tumors, Prostate Tumors, 11 Tumors and Lung cancer [41]	The PSO based FS approach significantly decreased the size of the feature set, but the feature numbers are still large.
5.	2015	Genetic Bee Colony (GBC) algorithm: to find the optimal feasible solution in order to optimized the accuracy of SVM classifier	Colon, Leukemia1, Lung, SRBCT, Lymphoma, Leukemia2	In both binary and multi-class cancer classification, the proposed algorithm is a potential approach for explaining the dimensional reduction problem.
6.	2015	Hybrid wrapper-filter feature selection method based on a Binary Differential Evolution (BDE) algorithm with a rank-based filter FS method.	Six well-known datasets from Kent Ridge Biomedical Dataset Public Repository. [39]	The proposed hybrid method gives the facility to successfully reduce the original attribute subset size by more than 99% and displays the strength as well as the capability of the methods to increase the accuracy rate.
7.	2016	Improved shuffled frog leaping algorithm (ISFLA) for feature selection and 1-nearest neighbor (1-NN) as a evaluator	Nine well-known biomedical data from [41]	The improved algorithm effectively reduces the number of dataset features whilst simultaneously achieving better classification accuracy.
8.	2016	Chaotic binary particle swarm optimization and Local search (CBPSOL) for feature subset selection and K-nearest neighbors for classification	Eight datasets from UCI repository [40]	Proposed algorithm has a robust search ability in the problem space and can capably catch smaller feature subsets in high-dimensional synthetic datasets
9.	2017	Support Vector Machines recursive feature elimination (SVM-RFE) as a prefilter step and Binary Dragon Fly (BDF) as a feature selection method	Six DNA microarray datasets from Kent Ridge (KR) Bio-Medical Dataset Repository [39]	The proposed model is efficient and provides a higher classification accuracy rate using a reduced number of genes.
10.	2017	Hybrid FS approach combine with Advanced Binary Ant Colony Optimization (ABACOH) and improved binary gravitational search algorithm (IBGSA) for	Lung Cancer, Leukemia, Colon, SRBCT, Prostate	Proposed model yield 81.43%, 96.50%, 94.11%, 92.94% and 94.26%, for the five datasets. Higher than other techniques.

feature selection

11.	2018	Improved Gravitational Algorithm (IBSGA) as a feature selection process and k-Nearest Neighbor as a classifier	Binary Search	Lung Cancer, DLBCL, SRBCT, Prostate	Leukemia,	The proposed approach might be a suitable solution for feature selection for high-dimensional data, especially micro-array data and obtain better results.
-----	------	--	---------------	-------------------------------------	-----------	--

Acknowledgements

The authors are thankful to the anonymous reviewers for sharing their valuable comments that definitely improved the quality of the paper and gratitude to people who have contributed towards the development of this manuscript from the planning state to the finish state.

References

- [1] Mustafa Serter, et. al., "Feature selection method based on artificial bee colony algorithm and support vector machines for medical datasets classification." Hindawi Publishing Corporation, The Scientific World Journal, Volume 2013, Article ID 419187, 10 pages.
- [2] Fong, Simon, et al. "Feature selection in life science classification: metaheuristic swarm search." IT Professional 16.4 (2014): 24-29.
- [3] Chen, Yiyuan et al. "An effective feature selection scheme for healthcare data classification using binary particle swarm optimization." 2018 9th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2018.
- [4] Thara, L., et al., "Swarm Intelligence Based Feature Selection Algorithms and Classifiers for Gastric Cancer Prediction." International Conference on Intelligent Data Communication Technologies and Internet of Things. Springer, Cham, 2018.
- [5] M. Dash and H. Liu, "Feature selection for classification," *Intell. Data Anal.*, vol. 1, nos. 1–4, pp. 131–156, 1997.
- [6] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.
- [7] A. S. U. Kamath, K. De Jong, and A. Shehu, "Effective automated feature construction and selection for classification of biological sequences," *PLoS One*, vol. 9, no. 7, 2014, Art. ID e99982.

- [8] W. A. Albukhanajer, et al., "Evolutionary multi-objective optimization of trace transform for invariant feature extraction," in Proc. IEEE Congr. Evol. Comput. (CEC), Brisbane, QLD, Australia, 2012, pp. 1–8.
- [9] Elbeltagi, Emad, Tarek Hegazy, and Donald Grierson. "Comparison among five evolutionary-based optimization algorithms." *Advanced engineering informatics* 19.1 (2005): 43-53.
- [10] Xue, Bing, et al. "A survey on evolutionary computation approaches to feature selection." *IEEE Transactions on Evolutionary Computation* 20.4 (2016): 606-626.
- [11] H. Liu, H. Motoda, et al., "Feature selection: An ever evolving frontier in data mining," in Proc. JMLR Feature Sel. Data Min., vol. 10. Hyderabad, India, 2010, pp. 4–13.
- [12] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.
- [13] A. Unler and A. Murat, "A discrete particle swarm optimization method for feature selection in binary classification problems," *Eur. J. Oper. Res.*, vol. 206, no. 3, pp. 528–539, 2010.
- [14] Y. Liu et al., "An improved particle swarm optimization for feature selection," *J. Bionic Eng.*, vol. 8, no. 2, pp. 191–200, 2011.
- [15] Wang, Suhang, Jiliang Tang, and Huan Liu. "Feature selection." *Encyclopedia of Machine Learning and Data Mining* (2016): 1-9.
- [16] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [17] Mashhour, Emad Mohamed, et al. "Survey on different Methods for Classifying Gene Expression using Microarray Approach." *International Journal of Computer Applications* 150.1 (2016).
- [18] Kennedy, James. "Swarm intelligence." in: A.Y. Zomaya (Ed.), *Handbook of nature-inspired and innovative computing*. Springer, Boston, MA, 2006, pp. 187-219.
- [19] Brezočnik, Lucija, et al., "Swarm Intelligence Algorithms for Feature Selection: A Review." *Applied Sciences* 8.9 (2018): 1521.
- [20] B. Bin, et al., "Comparison on Swarm Algorithms for Feature Selections/Reductions." *International Journal of Scientific & Engineering* (2014).
- [21] J. Kennedy, R. Eberhart, *Particle Swarm Optimization*, in: *The Proceedings of the IEEE International Conference on Neural Networks*, Washington, DC, 1995, pp. 1942–1948

- [22] R. J. Mullen, D. N. Monekosso, S. A. Barman and P. wwwRemagnino, "Review: a review of ant algorithms," *Expert systems with applications*, vol. 36, no. 6, pp. 9608-9617, 2009.
- [23] Ismkhan, H., 2017. Effective heuristics for ant colony optimization to handle large-scale problems. *Swarm and Evolutionary Computation*, 32, pp.140-149.
- [24] Karaboga D. An idea based on honey bee swarm for numerical optimization. Technical Report TR06. Erciyes University. 2005.
- [25] Karaboga D. Artificial bee colony algorithm. *Scholarpedia*. 2010: 6915.
- [26] Zamri, et al. "Review on the Usage of Swarm Intelligence in Gene Expression Data." *International Conference for Innovation in Biomedical Engineering and Life Sciences*. Springer, Singapore, 2017.
- [27] Banka, H., Dara, S.: A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation. *Pattern Recogn. Lett.* 52, 94–100 (2015)
- [28] Tran, B., Zhang, M. and Xue, B., 2016, July. A PSO based hybrid feature selection algorithm for high-dimensional classification. In *Evolutionary Computation (CEC), 2016 IEEE Congress on* (pp. 3801-3808).
- [29] Gupta S., Baghel, A., and Iqbal, A., 2018. Threshold Controlled Binary Particle Swarm Optimization for High Dimensional Feature Selection, *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.10, No.8, pp.75-84.
- [30] Kannan, S. Senthamarai, and N. Ramaraj. "A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm." *Knowledge-Based Systems* 23.6 (2010): 580-585.
- [31] GaneshKumar, Pugalendhi, et al. "Hybrid ant bee algorithm for fuzzy expert system based sample classification." *IEEE/ACM transactions on computational biology and bioinformatics* 11.2 (2014): 347-360.
- [32] Rouhi, Amirreza, and Hossein Nezamabadi-pour. "A hybrid method for dimensionality reduction in microarray data based on advanced binary ant colony algorithm." *2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC)*. IEEE, 2016.
- [33] M. S. Uzer et al. (Selçuk University) "Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification" Hindawi Publishing Corporation, *The Scientific World Journal*, Volume 2013, Article ID 419187, 10 pages
- [34] Alshamlan, Hala, Ghada Badr, and Yousef Alohal. "mRMR-ABC: a hybrid gene selection algorithm

for cancer classification using microarray gene expression profiling." *Biomed research international* 2015 (2015).

[35] Garro, Beatriz A., Katya Rodríguez, and Roberto A. Vázquez. "Classification of DNA microarrays using artificial neural networks and ABC algorithm." *Applied Soft Computing* 38 (2016): 548-560.

[36] Li, Xiangtao, Mingjie Li, and Minghao Yin. "Multiobjective ranking binary artificial bee colony for gene selection problems using microarray datasets." *IEEE/CAA Journal of Automatica Sinica* (2016).

[37] Liang Lan, Slobodan Vucetic," Improving accuracy of microarray classification by a simple multi-task feature selection filter", *Int. J. Data Mining and Bioinformatics*, Vol. 5, No. 2, 2011

[38] Tran, Binh, et al. "Investigation on particle swarm optimisation for feature selection on high-dimensional data: Local search and selection bias." *Connection Science* 28.3 (2016): 270-294.

[39] Kent Ridge Bio-Medical Dataset Repository, URL: <http://datam.i2r.a-star.edu.sg/datasets/kcbd/index.html>.

[40] UCI Machine Learning Repository, URL: <http://archive.ics.uci.edu/ml/>

[41] <http://www.gems-system.org>.