Novel Initialisation and Updating Mechanisms in PSO for Feature Selection in Classification

Bing Xue, Mengjie Zhang, and Will N. Browne

School of Engineering and Computer Science
Victoria University of Wellington, Wellington, New Zealand
Email: (Bing.Xue, Mengjie.Zhang, Will.Browne)@ecs.vuw.ac.nz

Abstract. In classification, feature selection is an important, but difficult problem. Particle swarm optimisation (PSO) is an efficient evolutionary computation technique. However, the traditional personal best and global best updating mechanism in PSO limits its performance for feature selection and the potential of PSO for feature selection has not been fully investigated. This paper proposes a new initialisation strategy and a new personal best and global best updating mechanism in PSO to develop a novel feature selection algorithm with the goals of minimising the number of features, maximising the classification performance and simultaneously reducing the computational time. The proposed algorithm is compared with two traditional feature selection methods, a PSO based method with the goal of only maximising the classification performance, and a PSO based two-stage algorithm considering both the number of features and the classification performance. Experiments on eight benchmark datasets show that the proposed algorithm can automatically evolve a feature subset with a smaller number of features and higher classification performance than using all features. The proposed algorithm achieves significantly better classification performance than the two traditional methods. The proposed algorithm also outperforms the two PSO based feature selection algorithms in terms of the classification performance, the number of features and the computational cost.

Keywords: Particle Swarm Optimisation, Feature Selection, Classification.

1 Introduction

Classification problems usually have a large number of features, including relevant, irrelevant and redundant features. However, irrelevant and redundant features may reduce the classification performance due to the large search space, known as “the curse of dimensionality” [3, 6]. Feature selection is to select a subset of relevant features for classification, which could shorten the training time, simplify the learned classifiers, and/or improve the classification accuracy [6].

Feature selection is a difficult problem due mainly to the large search space, which increases exponentially with respect to the number of available features [6]. Therefore, an exhaustive search is practically impossible in most situations.

Different heuristic search techniques have been applied to feature selection, such as greedy search [3]. However, most of the existing algorithms still suffer from the problems of stagnation in local optima or being computationally expensive [3, 6]. In order to better address feature selection problems, an efficient global search technique is needed.

Evolutionary computation (EC) techniques are well-known for their global search ability. They have been applied to feature selection problems, such as genetic algorithms (GAs) [1], genetic programming (GP) [13], and particle swarm optimisation (PSO) [15]. PSO [14] is a relatively recent EC technique, which is computationally less expensive than some other EC algorithms. In PSO [14], a population of candidate solutions are encoded as particles in the search space. PSO starts with the random initialisation of a population of particles. Based on the best experience of one particle (*pbest*) and its neighbouring particles (*gbest*), PSO searches for the optimal solution by updating the velocity and the position of each particle according to the following equations:

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{1}$$

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_{1i} * (p_{id} - x_{id}^{t}) + c_2 * r_{2i} * (p_{gd} - x_{id}^{t}) \tag{2}$$

where x and v represent the position and the velocity. t denotes the tth iteration in the evolutionary process. $d \in D$ denotes the dth dimension in the D-dimensional search space. w is inertia weight. c_1 and c_2 are acceleration constants. r_{1i} and r_{2i} are random values uniformly distributed in [0, 1]. p_{id} and p_{gd} represent the elements of *pbest* and *gbest* in the dth dimension.

Many studies have shown that PSO is an efficient search technique for feature selection [2, 8, 15]. However, there are some limitations about current PSO for feature selection. Firstly, PSO has not been tuned to the feature selection task. Many initialisation strategies have been proposed in PSO to improve its performance [16]. However, no existing initialisation strategies are specifically proposed for feature selection. Secondly, the traditional *pbest* and *gbest* updating mechanism may cause missing good feature subsets with high classification performance and a small number of features (Discussions in Section 2.2). Therefore, the potential of PSO for feature selection has not been fully investigated.

1.1 Goals

The overall goal of this paper is to propose a new PSO based feature selection approach to selecting a smaller number of features and achieving similar or even better classification performance than using all features and traditional/existing feature selection methods. To achieve this goal, we propose a new initialisation strategy and a new mechanism for updating *pbest* and *gbest* in PSO to reduce the number of features without reducing (or even increasing) the classification performance. Specifically, we will:

– propose a new initialisation strategy in PSO to reduce the number of features without decreasing the classification performance of the evolved subset,

– develop a new updating mechanism to lead PSO to search for the feature subsets with high classification performance and small numbers of features,
– develop a new PSO based wrapper feature selection algorithm using the proposed initialisation strategy and updating mechanism, and
– investigate whether the proposed feature selection algorithm can outperform two traditional feature selection methods, a PSO based algorithm with the goal of only maximising the classification performance, and a PSO based two-stage algorithm considering both of the two main objectives.

2 Proposed Approach

Feature selection has the two main objectives of maximising the classification performance and minimising the number of features. However, most existing methods only aim to maximise the classification performance [2]. Some works combine these two objectives into a single fitness function [8, 18], but they need a predefined parameter to balance these two components, which is usually problem-dependent and hard to determine *a priori*. To solve this problem, we only include the classification error rate in the fitness function (Equation 3) because it is more important than the number of features. Meanwhile, we propose an initialisation strategy and a new *pbest* and *gbest* updating mechanism to reduce the number of features without decreasing or even increasing the classification performance, which also reduces the computational cost.

$$Fitness_1 = ErrorRate \tag{3}$$

2.1 New initialisation Strategy

The new initialisation strategy is motivated by the two traditional methods, forward selection [17] and backward selection [12]. Forward selection starts with an empty set of features and it usually selects a smaller number of features, but it may miss the optimal feature subset with a large number of features. Backward selection starts with the full set of features and it usually selects a large number of features, but the computational time is longer than forward selection.

Hence, we propose a new initialisation strategy to take the advantages of forward and backward selection and avoid their disadvantages. In this new strategy, particles are initialised using a small number of features. Therefore, the algorithm will start with searching the solution space with small feature subsets. This will also reduce the computational cost because the evaluation of a small feature subset in wrapper approaches takes less time than a large feature subset. However, if all the particles are initialised with small subsets, PSO may miss the medium or large feature subsets that can achieve the best classification performance. Therefore, in the proposed initialisation strategy, most particles are initialised using a small number of features (simulating forward selection) and other particles are initialised using large feature subsets (simulating backward selection). Meanwhile, through social interaction (updating *pbest* and *gbest*), PSO is expected to be able to reach and search the solution space with medium feature subsets if these feature subsets can achieve better classification performance.

2.2 New *pbest* and *gbest* Updating Mechanism

In PSO, particles share information through *pbest* and *gbest*, which can influence the behaviour of the swarm during the evolutionary process. Traditionally, the *pbest* and *gbest* are updated solely based on the fitness value of the particles (i.e., classification performance in feature selection problems). *pbest* of a particle is updated only when the fitness of the new position of the particle is better than the current *pbest*. In feature selection, the traditional updating mechanism has a potential limitation. If the classification performance of the particle's new position is the same as the current *pbest*, but the number of features is smaller, the particle's new position corresponds to a better feature subset. However, according to the traditional updating mechanism, the *pbest* will not be updated because their classification performance is the same.

To overcome this limitation, we propose a new *pbest* and *gbest* updating mechanism. In the new mechanism, the classification performance of the feature subset is used as the fitness function, which means the classification performance is still the first priority, but the number of features is also considered. *pbest* and *gbest* are updated in two situations. The first situation is that if the classification performance of the particle's new position is better than *pbest*, *pbest* will be updated and replaced by the new position. In this case, the number of features selected will be ignored. The second situation is that if the classification performance of the new position is the same as *pbest* and the number of features is smaller, the current *pbest* will be replaced by the particle's new position. After updating *pbest*, *gbest* of each particle is updated in the same way by comparing *gbest* with the *pbest* of the particle and its neighbours.

By adding the second situation, the proposed updating mechanism is expected to avoid the limitation of traditional updating mechanism. Where available, it will always select a better feature subset to be the *pbest* or *gbest*, which either has better classification performance or the same classification performance with a smaller number of features. This can help the algorithm filter out redundant features and make the feature subset with good classification performance and a small number of features to be the *leader* (*pbest* or *gbest*) of each particle and the whole swarm.

Note that in GP, each individual can be represented as a tree. The size of the trees can be considered in the selection process, known as parsimony pressure [11]. The parsimony pressure seems similar to the proposed *pbest* and *gbest* updating mechanism. However, they are different ideas in two aspects. Firstly, the parsimony pressure in GP changes the size of the trees while the proposed *pbest* and *gbest* updating mechanism does not change the size of the particles that is always the total number of features in the dataset. Secondly, the parsimony pressure is to control the size of the trees in GP, which can be used in any problem domain, but the number of features considered in the proposed *pbest* and *gbest* updating mechanism is particularly for feature selection problems to optimise one of the two main objectives, i.e., minimising the number of features.

Based on the new initialisation strategy and updating mechanism, a new feature selection algorithm is proposed named IniPG. The pseudo-code of IniPG

Algorithm 1: The pseudo-code of the proposed algorithm (IniPG)

```
begin
    initialise most of the particle using small feature subsets and the others
    particles using relatively large feature subsets;
    initialise the velocity of each particle;
    while Maximum Iterations or the stopping criterion is not met do
        evaluate the fitness of each particle on the Training set;
        for i=1 to Population Size do
            if fitness of particle i (x_i) is better than that of pbest then
                pbest = x_i ;                    // Update the pbest of particle i
            else if fitness of x_i is the same as pbest and |x_i| < |pbest| then
                pbest = x_i ;                    // Update the pbest of particle i
            if fitness of pbest of any neighbour is better than that of gbest then
                gbest = pbest ;                  // Update the gbest of particle i
            else if fitness of pbest of any neighbour is the same as gbest and
            |pbest| < |gbest| then
                gbest = pbest ;                  // Update the gbest of particle i
        for i=1 to Population Size do
            update the velocity and the position of particle i
    calculate the classification accuracy of the selected features on the Test set;
    return the position of gbest (the selected feature subset);
    return the training and test classification accuracies;
```

can be seen in Algorithm 1. PSO has two versions, which are continuous PSO [14] and binary PSO [9], but binary PSO has potential limitations [10]. Therefore, we will use continuous PSO to propose a novel feature selection algorithm. The representation of a particle is a "n" bits string, where "n" is the total number of features. The position value in each dimension (x_{id}) is in [0,1]. A threshold θ is needed to compare with the value of x_{id}. If $x_{id} > \theta$, the dth feature is selected. Otherwise, the dth feature is not selected.

3 Design of Experiments

3.1 Benchmark Techniques

To examine the performance of the proposed algorithm (IniPG), two traditional wrapper feature selection methods and two PSO based algorithms (ErRt and 2Stage) as benchmark techniques in the experiments.

The two traditional methods are linear forward selection (LFS) [5] and greedy stepwise backward selection (GSBS), which were derived from SFS and SBS, respectively. More details about LFS can be seen in the literature [5] and GSBS starts with all available features and stops when the deletion of any remaining feature results in a decrease in classification performance. ErRt only uses the classification error rate as the fitness function. 2Stage [18] employs a two-stage fitness function to optimise the classification in the first stage and take the

Table 1. Datasets

Dataset	#Features	#Classes	#Instances	Dataset	#Features	#Classes	#Instances
Wine	13	3	178	Zoo	17	7	101
Wisconsin Breast Cancer				Vehicle	18	4	846
(Diagnostic) (WBCD)	30	2	569				
Ionosphere	34	2	351	Lung	56	3	32
Hillvalley	100	2	606	Madelon	500	2	4400

number of features into account in the second stage [18]. Binary PSO was used in [18], but continuous PSO is employed in this paper to keep consistent with ErRt and IniPG for fair comparisons.

3.2 Datasets and Parameter Settings

Eight datasets (Table 1) are chosen from the UCI machine learning repository [4], which have different numbers of features, classes and instances. For each dataset, the instances are randomly divided into two sets: 70% as the training set and 30% as the test set.

K-nearest neighbour (KNN) was used in the experiment and K=5 (5NN). Weka [7] is used to run the experiments of using LFS and GSBS. All the settings in LFS and GSBS are kept to the defaults except that backward search is chosen in GSBS. The parameters of PSO in ErRt, 2Stage and IniPG are set as follows: $w = 0.7298$, $c_1 = c_2 = 1.49618$, $v_{max} = 6.0$, population size is 30, and the maximum iteration is 100. The fully connected topology is used. These values are chosen based on the common settings in [14]. According to our previous experiments, the threshold θ is set as 0.6 in the three PSO based algorithms. In IniPG, a major part of the swarm (2/3) is initialised using around 10% of the total number of features. The other minor part of the swarm (1/3) is initialised using more than half of the total number of features, where a random number (e.g. m, where m is between half and the total number of features) is firstly generated and m features is randomly selected to initialise this particle.

For each dataset, each each experimental test has been conducted for 40 independent runs. A statistical significance test, T-test, is performed between their classification performances and the significance level was selected as 0.05.

4 Experimantal Results and Discussions

Table 2 shows the experimental results of the proposed algorithm and the benchmark techniques. "All" means that all features are used for classification. "NO." represents the average number of features selected. "Ave", "Best" and "StdDev" indicate the average, the best and the standard deviation of the 40 test accuracies in ErRt, 2Stage or IniPG. "T-test" shows the result of the T-test, where "+" ("-") means that the classification performance of a benchmark technique is significantly better (worse) than that of IniPG. "=" indicates they are similar.

4.1 Results of Benckmark Techniques

Results of LFS and GSBS: according to Table 2, LFS selected a smaller number of features and achieved a similar or higher classification accuracy than using

Table 2. Experimantal Results

Dataset	Method	NO.	Ave(Best)	StdDev	T-test	Dataset	Method	NO.	Ave(Best)	StdDev	T-test
Wine	All	13	76.54		-	Zoo	All	17	80.95		-
	LFS	7	74.07		-		LFS	7	74.07		-
	GSBS	8	85.19		-		GSBS	8	85.19		-
	ErRt	8	95.96 (100)	1.83E-2	=		ErRt	9.18	95.5 (97.14)	90.3E-4	=
	2Stage	8	95.96 (100)	1.83E-2	=		2Stage	9.18	95.5 (97.14)	90.3E-4	=
	IniPG	6.78	95.12 (98.77)	1.87E-2			IniPG	6.58	95.52 (97.14)	71.3E-4	
WBCD	All	30	92.98		-	Vehicle	All	18	83.86		-
	LFS	10	88.89		-		LFS	9	83.07		-
	GSBS	25	83.63		-		GSBS	16	75.79		-
	ErRt	13.42	93.39 (94.74)	55.8E-4	-		ErRt	9.52	85 (87.01)	79E-4	=
	2Stage	5	93.54 (94.74)	75.1E-4	-		2Stage	8.65	84.95 (87.01)	77.9E-4	=
	IniPG	3.45	94.09 (94.74)	82.5E-4			IniPG	10.28	85.31 (87.01)	95.5E-4	
Ionosphere	All	34	83.81		-	Lung	All	56	70		-
	LFS	4	86.67		-		LFS	6	90		+
	GSBS	30	78.1		-		GSBS	33	90		+
	ErRt	12.58	88.4 (93.33)	2.14E-2	+		ErRt	27.35	72 (80)	6E-2	-
	2Stage	12.05	88.14 (91.43)	1.89E-2	+		2Stage	27.38	72.25 (90)	6.89E-2	-
	IniPG	3.2	87.14 (91.43)	1.88E-2			IniPG	6.22	78.75 (90)	6.4E-2	
Hillvalley	All	100	56.59		-	Madelon	All	500	70.9		-
	LFS	8	57.69		=		LFS	7	64.62		-
	GSBS	90	49.45		-		GSBS	489	51.28		-
	ErRt	47.32	57.54 (61.81)	1.52E-2	=		ErRt	258.1	76.55 (79.49)	1.22E-2	-
	2Stage	47.05	57.57 (61.81)	1.55E-2	=		2Stage	256.48	76.52 (79.36)	1.26E-2	-
	IniPG	12.72	57.95 (60.71)	1.48E-2			Initia	216.4	78.49 (84.23)	3.23E-2	

all features in most cases. GSBS could reduce the number of features, but only achieved better classification performance on a few datasets. In most cases, LFS outperformed GSBS in terms of both the number of features and the classification performance. The results indicate that LFS as a forward selection algorithm is more likely to obtain good feature subsets with a small number of features GSBS (backward selection) because of different starting points. Feature subsets selected by GSBS may still have redundancy. This also suggests that utilising the advantages of both forward selection and backward selection can improve the performance of a feature selection algorithm, which motivates the proposal of the new initialisation strategy in this work.

Results of ErFs: according to Table 2, in almost all datasets, ErRt achieved similar or better classification performance than using all features, and the evolved feature subsets only included around half of the available features. This suggests that PSO as an evolutionary search technique can be successfully used for feature selection problems.

Results of 2Stage: according to Table 2, 2Stage evolved feature subsets with around half (or less) of the available features and achieved better classification performance than using all features in almost all cases. 2Stage outperformed ErRt in almost all cases. However, 2Stage attempted to find a trade-off between the classification performance and the number of features, which means the reduction of the number of features might decrease the classification performance.

4.2 Results of IniPG

According to Table 2, in 11 of the 12 datasets, IniPG evolved feature subsets that selected less than half (or even close to than 10% in four datasets) of the available features, but achieved significantly better classification performance than using all features. Only in the Movement dataset is the average classification performance obtained by IniPG (94.62%) is less, by 0.2%, than that of using all features (94.81%), but the best accuracy (95.19%) is higher.

Comparisons Between IniPG and Two Traditional Methods (LFS and GSBS): in almost all datasets, IniPG achieved significantly better or similar classification performance to LFS, although the number of features is slightly larger in some cases. Comparing IniPG with GSBS, the number of features in IniPG is smaller than GSBS in all datasets and the classification performance of IniPG is significantly better than GSBS in 11 of the 12 datasets. This suggest that IniPG as a PSO based algorithm can search the solution space more effectively than both LFS and GSB. The initialisation strategy movitated by both forward selection and backward selection can help IniPG take the advantages of both forward selection and backward selection to obtain feature subsets with a smaller number of features and better classification performance than both LFS and GSB.

Comparisons Between IniPG and ErRt: according to Table 2, IniPG selected feature subsets including smaller numbers of features and achieved significantly better or similar classification performance than ErRt in almost all datasets (except for the Ionosphere dataset, where the number of features in IniPG is around one fourth of that in ErRt). This suggests that although ErRt and IniPG shared the same fitness function (Equation 3), the proposed initialisation strategy and *pbest* and *gbest* updating mechanism can help IniPG to effectively eliminate the redundant and irrelevant features to obtain a smaller feature subset with significantly better classification performance than ErRt.

Comparisons Between IniPG and 2Stage: according to Table 2, in almost all datasets, the classification performance of IniPG is significantly better or similar to that of 2Stage and the number of features is smaller. The reason might be that the fitness function in the second stage in 2Stage aims to find a balance between the classification performance and the number of features. Therefore, the reduction of the number of features will also decrease the classification performance. In IniPG, the fitness function only includes the classification performance during the whole evolutionary process. This ensures that the reduction of the number of features in IniPG will not reduce the classification performance. Meanwhile, the proposed initialisation strategy and *pbest* and *gbest* updating mechanism can help IniPG further remove the irrelevant or redundant features to reduce the number of features, which in turn could increase the classification performance. In addition, compared with 2Stage, another advantage of IniPG is that it does not need a predefined parameter to balance the relative importance of the classification performance and the number of features.

Note that simply increasing the number of iterations cannot help ErRt and 2Stage achieve the same performance obtained by IniPG. The main reason is that ErRt does not consider the number of features in the fitness function and 2Stage takes a trade-off between the classification performance and the number of features. IniPG simulates both forward and backward selection to duplicate their advantages, which helps IniPG pay more attention to small feature subsets, but does not miss the large feature subsets with high classification performance. Meanwhile, because of the new updating mechanism, for two feature subsets with the same classification performance, IniPG will select the smaller one as the new *pbest* or *gbest*. ErRt and 2Stage using traditional updating mechanism will not do this during the evolutionary training process. Therefore, ErRt and 2Stage can not achieve as good performance as IniPG in almost all situations.

4.3 Analysis on Computational Time

All the five methods used in the experiments are wrapper based feature selection approaches. Therefore, most of their computational time is spent on the fitness evaluation, which regards the training and testing classification processes.

LFS usually used less time than the other four methods because the forward selection strategy starts with a small number of features and the evaluation of a small feature subset takes less time than a large feature subset. GSBS cost less time than other three PSO based algorithms (ErRt, 2Stage and IniPG) on the datasets with a small number of features, but more time on the datasets with a large number of features, such as the Madelon and Isolet5 datasets. The reason is that GSBS starts with the full set of features, which needs much longer time for each evaluation. The number of evaluations in GSBS substantially increases in such large datasets while the number of evaluations in PSO based algorithms is still the same. Generally, 2Stage cost less time than ErRt because the size of the feature subsets evolved by 2Stage is smaller than ErRt during the evolutionary training process. For the same reason, the computational time of IniPG is less than both of ErRt and 2Stage.

5 Conclusions

This paper proposes a new PSO algorithm for feature selection problems (IniPG). In IniPG, *a new initialisation strategy* was proposed based on the ideas of two traditional feature selection methods (forward selection and backward selection) to utilise the advantages of these two methods. Meanwhile, *a new pbest* and *gbest updating mechanism* was proposed to overcome the limitation of the traditional updating mechanism in order to ensure the feature subset with the highest classification performance and their smallest number of features become the new *pbest* or *gbest*. IniPG was examined and compared with two traditional feature selection algorithms (LFS and GSBS), a PSO based algorithm with only the classification error rate as the fitness function (ErRt) and a PSO based two-stage algorithm (2Stage). Experimental results show that in almost all datasets, IniPG achieved significantly better classification performance than LFS and GSBS, although the number of features is larger than LFS in some cases. In almost all cases, IniPG outperformed ErRt and 2Stage in terms of the number of features and the classification performance, and used less computational time.

In the future, we will further tune the PSO algorithm for feature selection problems. We will also investigate multi-objective PSO for feature selection in classification problems. We will also investigate whether using a given learning algorithm in a wrapper feature selection approach can select a good or near-optimal feature subset for other learning algorithms in the future.

References

1. Banerjee, M., Mitra, S., Banka, H.: Evolutionary rough feature selection in gene expression data. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(4), 622–632 (2007)
2. Chuang, L.Y., Chang, H.W., Tu, C.J., Yang, C.H.: Improved binary PSO for feature selection using gene expression data. Computational Biology and Chemistry 32(29), 29– 38 (2008)
3. Dash, M., Liu, H.: Feature selection for classification. Intelligent Data Analysis 1(4), 131–156 (1997)
4. Frank, A., Asuncion, A.: UCI machine learning repository (2010)
5. Gutlein, M., Frank, E., Hall, M., Karwath, A.: Large-scale attribute selection using wrappers. In: IEEE Symposium on Computational Intelligence and Data Mining (CIDM '09). pp. 332–339 (2009)
6. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. The Journal of Machine Learning Research 3, 1157–1182 (2003)
7. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 27, 83–85 (2005)
8. Huang, C.L., Dun, J.F.: A distributed PSO-SVM hybrid system with feature selection and parameter optimization. Application on Soft Computing 8, 1381–1391 (2008)
9. Kennedy, J., Eberhart, R.: A discrete binary version of the particle swarm algorithm. In: IEEE International Conference on Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation. vol. 5, pp. 4104–4108 (1997)
10. Khanesar, M., Teshnehlab, M., Shoorehdeli, M.: A novel binary particle swarm optimization. In: Mediterranean Conference on Control Automation. pp. 1–6 (2007)
11. Luke, S., Panait, L.: Lexicographic parsimony pressure. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'02). pp. 829–836 (2002)
12. Marill, T., Green, D.: On the effectiveness of receptors in recognition systems. IEEE Transactions on Information Theory 9(1), 11–17 (1963)
13. Neshatian, K., Zhang, M.: Using genetic programming for context-sensitive feature scoring in classification problems. Connection Science 23(3), 183–207 (2011)
14. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: IEEE International Conference on Evolutionary Computation (CEC'98). pp. 69–73 (1998)
15. Unler, A., Murat, A.: A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research 206(3), 528–539 (2010)
16. Wang, H., Li, H., Liu, Y., Li, C., Zeng, S.: Opposition-based particle swarm algorithm with cauchy mutation. In: IEEE Congress on Evolutionary Computation (CEC'07). pp. 4750–4756 (2007)
17. Whitney, A.: A direct method of nonparametric measurement selection. IEEE Transactions on Computers C-20(9), 1100–1103 (1971)
18. Xue, B., Zhang, M., Browne, W.N.: New fitness functions in binary particle swarm optimisation for feature selection. In: IEEE Congress on Evolutionary Computation (CEC'2012). pp. 2145–2152 (2012)