Overview of Particle Swarm Optimisation for Feature Selection in Classification

Binh Tran, Bing Xue, and Mengjie Zhang

Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand
{tran.binh,bing.xue,mengjie.zhang}@ecs.vuw.ac.nz

Abstract. Feature selection is a process of selecting a subset of relevant features from a large number of original features to achieve similar or better classification performance and improve the computation efficiency. As an important data preprocessing technique, research into feature selection has been carried out over the past four decades. Determining an optimal feature subset is a complicated problem. Due to the limitations of conventional methods, evolutionary computation (EC) has been proposed to solve feature selection problems. Particle swarm optimisation (PSO) is an EC technique which recently has caught much interest from researchers in the field. This paper presents a review of PSO for feature selection in classification. After describing the background of feature selection and PSO, recent work involving PSO for feature selection is reviewed. Current issues and challenges are also presented for future research.

Keywords: Particle swarm optimisation, feature selection, evolutionary computation, classification.

1 Introduction

In many fields such as data mining and machine learning, data sets may contain a large number of features. However, the redundant or irrelevant features may reduce the classification performance. In order to solve this problem, feature selection is proposed to pick a subset of features that are relevant to the target concept [11]. By removing the irrelevant and redundant features, feature selection could significantly shorten the running time, improve the classification accuracy, and/or simplify the structure of the learned classifiers or models [11]. However, feature selection is a difficult problem, especially when the number of features is large [49,28]. Therefore, the optimal solution cannot be guaranteed to be acquired except when an exhaustive search is performed. However, an exhaustive search often takes a long time [50]. In real-world applications, obtaining good solutions in a reasonable amount of time is more interested than being obsessed with optimal solutions.

EC techniques are population-based techniques with a set of genetically motivated operations. These operations are used by a population of candidate solutions to obtain the optimal or near-optimal solution of the problem. Recently, different EC algorithms have been applied to feature selection problems such as particle swarm optimisation (PSO) [3,23,27], genetic algorithms (GAs) [30,41], genetic programming (GP) [29,32], and ant colony optimisation (ACO) [16,35]. As a relatively new EC technique, PSO is

G. Dick et al. (Eds.): SEAL 2014, LNCS 8886, pp. 605–617, 2014.
© Springer International Publishing Switzerland 2014

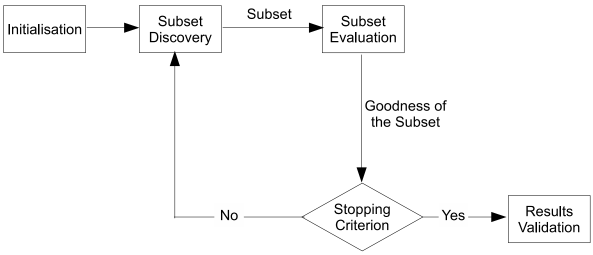

Fig. 1. General Feature Selection Process [11]

inspired by social behaviour such as birds flocking and fish schooling. Compared with other EC algorithms such as GAs, PSO is easier to implement and can converge more quickly [19]. It has been shown to be an effective method for feature selection problems [3,15,23,27]. This goal of this paper is to review recent work about PSO and its binary version called binary PSO (BPSO) [18] for feature selection in classification and to find the need for future research. The remainder of this paper is organised as follows. Feature selection is introduced in section 2 and Section 3 describes the standard PSO and BPSO algorithms. Section 4 reviews recent studies about PSO for feature selection in classification. The final section discusses current issues and challenges.

2 Feature Selection

Feature selection attempts to select the minimally sized subset of features that are necessary and sufficient to describe the target concept [11]. Its purposes include reducing the amount of data needed for learning, shortening the running time, improving the system accuracy, and increasing the comprehensibility of the learned model [24]. Fig. 1 shows the process of a typical feature selection method [11], which consists of five basic steps:

1. Initialisation: A feature selection algorithm starts with an initialisation procedure based on all the original features.
2. Subset discovery: A discovery procedure to generate candidate subsets. It is a search procedure [22], which can start with no features, all features, or a random subset of features. Many search techniques including conventional search methods and EC techniques are applied in this generation step to search for the best subset of features.
3. Subset evaluation: An evaluation function to measure the goodness of the generated feature subsets.
4. Stopping criterion: The algorithm will stop according to a given criterion, which can be based on the generation procedure or the evaluation function. The former can be a predefined number of features selected or a predetermined maximum number of iterations reached. The latter includes whether an optimal feature subset according to a certain evaluation function is obtained or whether addition or deletion of any feature does not produce a better subset.
5. Results validation: The validity of the selected subset is tested by carrying out tests on unseen data.

Table 1. Taxonomy of the reviewed papers

PSO for feature selection			
		Wrapper approach	**Filter approach**
Continuous PSO	Sing.Obj.	[3,15,23,27,45]	[14]
	Multi.Obj.	[48]	
Binary PSO	Sing.Obj.	[1,9,10,20,21,25,26,39,47,49,50,51]	[4,6,5,7,8,33,36,38,40]
	Multi.Obj.	[44]	[42,43]

Based on whether the subset evaluation process includes a learning algorithm or not, Langley [22] grouped different feature selection methods into two broad categories: filter approaches and wrapper approaches. Filter approaches are utilized to select features based on the evaluation criterion without using a learning algorithm. They are argued to be computationally less expensive and more general [7,50]. On the other hand, wrapper approaches implement a learning algorithm to construct a classifier in the evaluation procedure. They add or delete features to produce various feature subsets, and then measure the subsets depending on the performance of the developed classifier. Compared with filter approaches, wrapper approaches usually produce better results, especially when the classifier is designed to solve a particular problem. However, they are computationally expensive when the number of features is large [32]. In order to take advantage of both wrapper and filter approaches, recent studies proposed a hybrid approach in which filter methods were first used to select informative features before transferring to wrapper methods.

Naturally, an optimal feature subset is the smallest one that can obtain the highest classification quality, which makes feature selection a multi-objective problem [37]. Single-objective approaches can only produce one subset of features. Feature selection as a multi-objective problem producing several trade-off subsets can meet different user requirements in real-world applications. This paper briefly reviewed different PSO based feature selection algorithms, which can be seen in Table 1.

3 Particle Swarm Optimisation

3.1 Continuous Particle Swarm Optimisation

Particle swarm optimisation (PSO) is an evolutionary computation technique proposed by Kennedy and Eberhart in 1995 [12,17]. In PSO, each potential solution is called a bird or particle with no weight and no volume. The ith particle flies in a D-dimensional search space to find the optimal solution. There is a vector $x_i = (x_{i1}, x_{i2}, ..., x_{iD})$ presenting the position of particle i, where $x_{id \in [l_d, u_d]}$, $d \in [1, D]$, l_d and u_d are the lower and upper bounds of the dth dimension. The velocity of the ith particle is represented as $v_i = (v_{i1}, v_{i2}, ..., v_{iD})$. The best previous position of any particle is recorded as the personal best called $pbest$. The best solution visited by the whole swarm so far is the global best called $gbest$. The swarm is initialised with a population of random solutions. According to the $pbest$ and the $gbest$, the algorithm searches for the best solution by updating particles' positions and velocities using the following formulae:

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_1 * (p_{id} - x_{id}^{t}) + c_2 * r_2 * (p_{gd} - x_{id}^{t}) \tag{1}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{2}$$

where t means that the algorithm is going on the tth iteration. c_1 and c_2 are acceleration constants. r_1 and r_2 are random values uniformly distributed in [0, 1]. p_{id} presents the $pbest$ while p_{gd} presents the $gbest$. w is the inertia weight first introduced by Shi and Eberhart [34]. w can make a balance between the global search and the local search to improve the performance of PSO. The velocity v_{id}^{t} is limited by a maximum velocity, $v_{id}^{t+1} \in [-v_{max}, v_{max}]$ and v_{max} is predefined based on the problem to be solved. Eberhart and Shi [13] suggested v_{max} to be set at about $10-20\%$ of the dynamic range of the variable in each dimension.

3.2 Binary Particle Swarm Optimisation

PSO was originally proposed for continuous optimisation. Kennedy and Eberhart [18] developed a binary PSO (BPSO), which can be used for discrete problems. In BPSO, the position of each particle is encoded by a binary string. x_{id}, p_{id} and p_{gd} are restricted to 0 or 1. The velocity in BPSO represents the probability of an element in the position taking value 1. Equation (1) is still applied to update the velocity. A sigmoid function $s(v_{id})$ is introduced to transform v_{id} into the range of 0 and 1. BPSO updates the position of the particle according to the following formulae:

$$x_{id} = \begin{cases} 1, \text{ if } rand() < \dfrac{1}{1+e^{-v_{id}}} \\ 0, \ otherwise \end{cases} \tag{3}$$

where $rand()$ is a random number selected from a uniform distribution in [0,1]. v_{id} is transformed to [0,1] by a sigmoid limiting function.

4 PSO for Feature Selection

PSO shares many similarities with EC algorithms like GAs, but compared with GAs, PSO has its own advantages such as converging quickly and computationally inexpensive [52]. Both continuous PSO and binary PSO have been used for feature selection. Generally, when a continuous PSO algorithm is applied to feature selection problems, a particle in the swarm is formed by a vector of n real numbers, where n is the total number available features. In order to determine whether a feature will be selected or not, a threshold is needed to compare with the value in the vector. In BPSO, the representation of a particle is a n-bit binary string. The feature mask is Boolean that "1" represents that the feature is selected and "0" otherwise. A short review of recent work on PSO for feature selection will be presented in this section.

4.1 PSO Based Wrapper Feature Selection

Azevedo et al. [3] proposed a wrapper feature selection algorithm using PSO and support vector machines (SVM) for personal identification in a keystroke dynamic system. Experimental results showed that the proposed approach produced better performance than a GA with SVM regarding the classification error, processing time and the feature reduction rate. However, the false acceptance rate of the program was still high.

As a relatively new EC technique, PSO cannot avoid to have some disadvantages. Its high possibility to get stuck in local optima can be a typical example. Different strategies have been proposed to solve this problem. Yang et al. [50] proposed two BPSO based algorithms using two chaotic maps, a logistic map and a tent map to determine the inertia weights dynamically. The K-nearest neighbor (KNN) method with leave-one-out cross-validation (LOOCV) was applied in the wrapper models to evaluate the classification accuracies. Experiments showed that the proposed methods, especially BPSO with tent map, produced slightly higher classification accuracy than other methods, including sequential forward search (SFS), plus and take away (PTA), sequential forward floating search (SFFS), sequential genetic algorithm (SGA) and different hybrid genetic algorithms (HGAs).

Yang et al. [49] constructed a strategy for $gbest$ using Boolean operator to improve BPSO. When $gbest$ fitness was identical after three iterations, a Boolean operator 'and(.)' would 'and' each bit of the $pbest$ of all particles to create a new binary string. This new binary string would replace the old $gbest$. KNN with LOOCV was also applied to evaluate the classification accuracies in the experiments. Results illustrated that the proposed method usually achieved higher classification accuracy with fewer features than GA and BPSO. However, proposed BPSOs were not compared with other variations of PSO, which might produce better results. Chuang et al. [9] developed another resetting strategy for $gbest$ to improve the performance of BPSO for feature selection. $gbest$ would be reset to zero if the $gbest$ fitness maintained the same value after several iterations. The fitness of each particle was evaluated by KNN with LOOCV. Experiments were conducted on gene expression data sets. Results showed that this method effectively reduced the number of needed features and got the higher classification accuracy than the method created by Yang et al [49] in most cases.

Chuang et al [10] also applied another resetting strategy called catfish effect to improve BPSO. Similar to [9], if the fitness of $gbest$ stayed the same for a predefined number of iterations, ten percent of the population with the worst fitness would be forced to extreme positions which were either all 0s or all 1s randomly. The reported results of this method were better than BPSO and those such deterministic algorithms as SFS, PTA, SFFS and current stochastic algorithms for feature selection including simple GA and hybrid GAs on all data sets. However, only 5 runs were conducted for stochastic algorithms. Another improved BPSO algorithm was proposed in [26] for gene expression data. Speed concept was introduced to update particles' positions instead of velocity to increase the probability of not choosing a feature. In this way, PSO was able to find much smaller feature subsets than [9] and other compared methods. However, this method also reduced the accuracy in the cases, which might require a higher number of features in order to create a good prediction. Therefore, there were only six out of the ten data sets which had a higher average accuracy than [9]. Furthermore, except for some data sets which have 100% accuracy in all the runs, with a 10-run experiment, the standard deviations of the results on the remaining data sets were quite high (ranging from 0.3 to 0.9). There was also no statistical significance test done for the results.

In order to simultaneously maximise the classification accuracy and minimize the number of features selected, an aggregate fitness function was proposed for BPSO in [47]. This BPSO evolved in two stages. Classification error rate was the only measure used in the fitness function of the first stage. In the second stage, the subset size was added into the fitness function with an adaptive weight. This method evolved smaller feature subsets and higher classification performance than the standard BPSO and the PSO using only stage one. Alba et al. [1] combined geometric BPSO with SVM for feature selection where current position, $pbest$ and $gbest$ were used as three parents in a three-parent mask-based crossover operator to determine the new position for each particle. SVM with 10-fold cross-validation was applied in the fitness evaluation process. An aggregate fitness function was used to simultaneously maximise the classification accuracy and minimize the number of features with different weights. Compared to the second algorithm proposed in this paper which combined GA with SVM, this method performed slightly better with smaller feature subsets. Experiments also showed that the initialisation of the PSO produced a great influence in the performance, since it introduced an early subset of acceptable solutions in the evolution process.

Based on this motivation, Xue et al [45] proposed three mechanisms to initialize particles in PSO for feature selection in classification. While small initialisation method generated particles with a small number of roughly 10% of features selected, the large initialisation method generated particles with a random large number of more than 50% of features selected. The mixed method used the small initialisation for most of the particles (about two-thirds) and the large initialisation for the remainders. The mixed initialisation gave the best results of the three methods. Although achieving as a good classification performance as that of standard PSO, the feature subsets evolved by the mixed method are smaller than those of PSO in eight out of 14 data sets, thereby, reducing the computation time. Additionally, the paper also introduced three new updating mechanisms for $pbest$, $gbest$. Combining the mixed initialisation method with the new updating method achieved much smaller subsets and better or at least similar accuracy as the standard PSO and the two-stage algorithm [46,47].

Unler et al. [39] proposed a wrapper feature selection method for binary classification problems based on a modified BPSO and a logistic regression model. In this study, BPSO was modified by extending social learning to update the velocity of the particles. An adaptive feature subset selection strategy was developed, where the features were selected not only according to their independent likelihood calculated by BPSO, but also according to their contribution to the subset of features already selected. Meanwhile, this strategy maintained a list of features, which had already been considered for feature addition. Only a limited number of features were considered to be selected in the feature subset, thus the computational effort for the classification learning was reduced. Experimental results indicated that the proposed method outperformed tabu search and scatter search algorithms.

Based on a statistical clustering method and BPSO, Lane et al. [20] developed a new wrapper feature selection algorithm, where features are grouped into different clusters by the statistical clustering method based on their similarity, i.e. relatively homogeneous features in the same group. The proposed algorithm aimed to select one representative features from each cluster. The results show that by selecting only a very small number of features, the algorithm can achieve similar or even better classification performance than using all features. Later, multiple or zero features are allowed to be selected from each cluster in [21] to further improve the classification performance.

A wrapper multi-objective PSO was proposed by Xue et al [48] using the non-dominated sorting concept (NSPSOFS) and the crowding, mutation and dominance concept (CMDPSOFS) to evolve non-dominated solutions for feature selection in classification. The results showed that both algorithms achieved more and better solutions than existing deterministic feature selection algorithms, i.e. linear forward selection (LFS), greedy stepwise backward selection (GSBS), and stochastic algorithms as the standard PSO, and the two-stage PSO [47]. By using the strategies of maintaining the diversity of the swarm, CMDPSOFS outperformed NSPSOFS and other three well-known evolutionary multi-objective algorithms, namely non-dominated sorting-based multi-objective GA II, strength Pareto evolutionary algorithm 2 and Pareto archived evolutionary strategy on 12 benchmark data sets. The performance of continuous PSO and binary PSO for multi-objective feature selection is compared in [44]. The results show that continuous PSO generally achieved better performance than binary PSO. More PSO based filter feature selection algorithms can be seen from [2,25,51].

4.2 PSO Based Filter Feature Selection

Wang et al [40] proposed a filter feature selection approach based on an improved BPSO and rough sets theory. In this BPSO method, velocity was used to determine the number of bits should be changed in the binary position of the particle. The fitness function combined the dependency degree of classes on features calculated according to the rough sets theory and the proportion of the selected features. Experimental results showed that the improved BPSO was computationally less expensive than a GA using rough sets in terms of both memory and running time. However, the classification performance of the feature subsets was only tested on LEM2 algorithm, which had some bias for rough set based algorithms. A fuzzy sets based fitness function was introduced by Chakraborty [7] to build a BPSO based filter feature selection algorithm. Feature evaluation index [31] was used in the fitness function. It aimed to find a feature subset which had minimum intraclass ambiguity and maximum interclass ambiguity. Experiments with the Iris and Sonar data sets illustrated that the proposed BPSO performed better than GA did. However, both Iris and Sonar include a relatively small number of features.

Cervante et al. [6] developed two filter based approaches using BPSO. In the first algorithm, mutual information was used to measure the relevance between features and the class labels and the redundancy between a pair of features. Meanwhile, entropy was employed in the second method to measure the relevance of a group of features to the class labels and the redundancy within a group of features. Both algorithms used an aggregate fitness function combining the relevance level and the redundancy level with different weights. The results showed that the first algorithm evolved smaller subsets while the second one produced better classification accuracy. However, the classification performance using the feature subsets evolved by both algorithms was just as good as using all the features in three out of four data sets. This confirms one of the drawbacks of filter based approaches. Xue et al [43] further explored the effectiveness of mutual information and entropy in multi-objective feature selection. These two measures were combined with non-dominated sorting concept and the crowding, mutation and dominance concept as in [48] to form four different multi-objective BPSO algorithms: NSfsMI, NSfsE, CMDfsMI, CMDfsE. The results showed that the proposed multi-objective approaches achieved better solutions than the single-objective BPSO using the same measures [6]. The algorithms using entropy achieved better classification performance than those using mutual information. Although algorithms using mutual information selected a smaller number of features than those using entropy in single-objective algorithms, this observation did not appear in multi-objective algorithms. CMDfsMI and CMDfsE outperformed all the other methods in terms of both the number of features and the classification performance.

Some studies not only used PSO for feature selection but also employed PSO to optimize parameters for the classification algorithm used to evaluate the feature subsets. Lin et al. [23] proposed a wrapper feature selection approach (PSO+SVM), which simultaneously determined the parameters and picked a subset of features using continuous PSO. Radial basis kernel function (RBF) with two parameters was used in SVM. In PSO+SVM, each particle with $(n + 2)$ variables represents n features and 2 parameters. Experiments with 10-fold crossover validation demonstrated that the classification accuracy of PSO+SVM outperformed that of a grid search, a newton SVM and a Lagrangian SVM. The PSO+SVM approach could simultaneously determine the parameter values and find a subset of features without the lowering the classification accuracy.

Similarly, Huang et al. [15] also developed a wrapper method (PSO-SVM) for feature selection and parameters determination in one process. The difference between PSO-SVM and PSO+SVM [23] was that PSO-SVM used binary PSO and continuous PSO to simultaneously optimize the feature subset and SVM kernel parameters, respectively. Experiments with 10-fold crossover validation showed that PSO-SVM could determine the parameters, search the discriminating feature subset simultaneously and also achieve high classification accuracy.

Mohemmed et al. [27] proposed a hybrid method (PSOAdaBoost), which incorporated PSO with an AdaBoost framework for face detection. The PSOAdaBoost algorithm picked the best feature subset and determined the decision thresholds of AdaBoost simultaneously, so it could speed up the process of the training and increase the accuracy of weak classifiers. The method encoded the particle with the feature parameters and two centroids which were used to label an instance into the positive or negative class. Experimental results showed that PSOAdaBoost could be trained in a much shorter time and improve the performance of feature selection. This method used different learning algorithms and test problems with PSO+SVM [23] and PSO-SVM [15], but all of them could optimize the feature subset and parameters in one process.

A filter-based PSO was introduced by Guan et al [14] for feature selection in microarray data sets. Two informativeness metrics constructed based on ANOVA statistics were used to evaluate feature subsets. The experiment results on two binary-class data sets were compared with six methods including the Two-Phase EA/KNN, the SVM, the GA-SVM, the EA, the Redundancy based and the PSO-SVM. Although the proposed method always evolved the smallest subsets, it can only achieved the best accuracy on one dataset. More than ten independent runs should be conducted and statistical significant test should be done in order to have a stronger conclusion. A two-stage based filter feature selection algorithm using BPSO was proposed in [33], where k-means technique was used in the first stage to cluster features into k clusters. Then, signal-to-noise ratio score was used as a filter approach to select the best feature from each cluster. The k selected features were transfered into the second stage for PSO to search for the optimal subset. SVM, KNN and Probabilistic Neural Network were used to evaluate the goodness of the selected features. The results showed that the proposed method achieved higher accuracy with much smaller subsets than using all features. However, the experiment was applied on only four binary-class gene data sets. The efficiency of the proposed method were not discussed and compared with any other EC techniques. The different setting values of k used in k-means technique for different data sets might require some expert knowledge about microarray data. The work in [33] is similar to that in [20,21], but use different clustering methods to group features. It will be interesting to investigate their advantages and disadvantages by comparing with each other. More PSO based filter feature selection algorithms can be seen from [4,5,8,36,38,42].

5 Conclusions: Current Issues and Challenges

Many different EC algorithms have been applied to feature selection problems. In recent work, both filter and wrapper approaches were developed and feature selection was also regarded as multi-objective problems. This paper mainly reviewed recent PSO based feature selection approaches. In conclusion, some discussions about current issues and challenges as well as some possible research directions of PSO for feature selection are given as follows:

- Feature selection for *large-scale* classification with thousands or tens of thousands of features is still a challenging task, since most of the existing methods have difficulties to scale up to such high dimensions. These problems typical require a novel search mechanism and evaluation methods;
- High computational cost is one of the main problems in feature selection. Efficient evaluation or fitness *measures* can significantly speed up the feature selection process. This can be achieved by developing new filter measures, such as information theory measures, consistency measures, statistical measures, and fuzzy or rough sets based measures;
- In PSO or other EC based feature selection methods, the traditional *representation* is one of the issues limiting their performance on large-scale complex problems. Developing a novel representation is also an open issue but with very few existing works. The novel representation can be continuous or discrete (or binary) and fixed-length or variable length;
- Feature selection is an NP hard problem which requires a powerful global search technique. To improve the performance of an EC based approach, new *search mechanisms* are needed to develop, which may involve local search, hybrid different EC search mechanisms and so on;

- The number of features and the classification performance are often conflicting objectives, but only a small number of multi-objective feature works have been conducted. Meanwhile, developing new evaluation metrics and further selection method to choose a single solutions from a set of trade-off solutions are also an interesting topics, and
- Feature selection is not only important in classification, but also for other problems, such as clustering and regression problems. Feature selection in such domains will also be interesting.

Acknowledgment. This work is supported in part by the National Science Foundation of China (NSFC No. 61170180), the Marsden Funds of New Zealand (VUW1209 and VUW0806) and the University Research Funds of Victoria University of Wellington (203936/3337).

References

1. Alba, E., Garcia-Nieto, J., Jourdan, L., Talbi, E.G.: Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms. In: IEEE Congress on Evolutionary Computation (CEC 2007), pp. 284–290 (2007)
2. Aneesh, M., Masand, A.A., Manikantan, K.: Optimal feature selection based on image preprocessing using accelerated binary particle swarm optimization for enhanced face recognition. Procedia Engineering 30(5), 750–758 (2012)
3. Azevedo, G., Cavalcanti, G., Filho, E.: An approach to feature selection for keystroke dynamics systems based on PSO and feature weighting. In: IEEE Congress on Evolutionary Computation (CEC 2007), pp. 3577–3584 (2007)
4. Cervante, L., Xue, B., Shang, L., Zhang, M.: Binary particle swarm optimisation and rough set theory for dimension reduction in classification. In: IEEE Congress on Evolutionary Computation, pp. 2428–2435 (2013)
5. Cervante, L., Xue, B., Shang, L., Zhang, M.: A multi-objective feature selection approach based on binary pso and rough set theory. In: Middendorf, M., Blum, C. (eds.) EvoCOP 2013. LNCS, vol. 7832, pp. 25–36. Springer, Heidelberg (2013)
6. Cervante, L., Xue, B., Zhang, M., Shang, L.: Binary particle swarm optimisation for feature selection: A filter based approach. In: IEEE Congress on Evolutionary Computation (CEC 2012), pp. 881–888 (2012)
7. Chakraborty, B.: Feature subset selection by particle swarm optimization with fuzzy fitness function. In: 3rd International Conference on Intelligent System and Knowledge Engineering (ISKE 2008), pp. 1038–1042 (2008)
8. Chakraborty, B., Chakraborty, G.: Fuzzy consistency measure with particle swarm optimization for feature selection. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2013), pp. 4311–4315 (2013)
9. 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)
10. Chuang, L.Y., Tsai, S.W., Yang, C.H.: Improved binary particle swarm optimization using catfish effect for feature selection. Expert Syst. Appl. 38(10), 12699–12707 (2011)

11. Dash, M., Liu, H.: Feature selection for classification. Intelligent Data Analysis 1(4), 131–156 (1997)
12. Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: Proceedings of the Sixth International Symposium on Micro Machine and Human Science, MHS 1995, pp. 39–43 (October 1995)
13. Eberhart, R., Shi, Y.: Particle swarm optimization: developments, applications and resources. In: IEEE Congress on Evolutionary Computation, vol. 1, pp. 81–86 (2001)
14. Guan, J., Han, F., Yang, S.: A new gene selection method for microarray data based on PSO and informativeness metric. In: Huang, D.-S., Jo, K.-H., Zhou, Y.-Q., Han, K. (eds.) ICIC 2013. LNCS, vol. 7996, pp. 145–154. Springer, Heidelberg (2013)
15. Huang, C.L., Dun, J.F.: A distributed PSO-SVM hybrid system with feature selection and parameter optimization. Appl. Soft Comput. 8, 1381–1391 (2008)
16. Ke, L., Feng, Z., Xu, Z., Shang, K., Wang, Y.: A multiobjective ACO algorithm for rough feature selection. In: Second Pacific-Asia Conference on Circuits,Communications and System (PACCS 2010), vol. 1, pp. 207–210 (2010)
17. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
18. Kennedy, J., Eberhart, R.: A discrete binary version of the particle swarm algorithm. In: IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, pp. 4104–4108 (1997)
19. Kennedy, J., Spears, W.: Matching algorithms to problems: an experimental test of the particle swarm and some genetic algorithms on the multimodal problem generator. In: IEEE World Congress on Computational Intelligence, pp. 78–83 (1998)
20. Lane, M., Xue, B., Liu, I., Zhang, M.: Particle swarm optimisation and statistical clustering for feature selection. In: Cranefield, S., Nayak, A. (eds.) AI 2013. LNCS, vol. 8272, pp. 214–220. Springer, Heidelberg (2013)
21. Lane, M., Xue, B., Liu, I., Zhang, M.: Gaussian based particle swarm optimisation and statistical clustering for feature selection. In: Blum, C., Ochoa, G. (eds.) EvoCOP 2014. LNCS, vol. 8600, pp. 133–144. Springer, Heidelberg (2014)
22. Langley, P.: Selection of relevant features in machine learning. In: AAAI Technique Report FS-94-02, pp. 127–131 (October 1994)
23. Lin, S.W., Ying, K.C., Chen, S.C., Lee, Z.J.: Particle swarm optimization for parameter determination and feature selection of support vector machines. Expert Systems with Applications 35(4), 1817–1824 (2008)
24. Liu, H., Dougherty, E., Dy, J., Torkkola, K., Tuv, E., Peng, H., Ding, C., Long, F., Berens, M., Parsons, L., Zhao, Z., Yu, L., Forman, G.: Evolving feature selection. IEEE Intelligent Systems 20(6), 64–76 (2005)
25. Mohamad, M., Omatu, S., Deris, S., Yoshioka, M., Abdullah, A., Ibrahim, Z.: An enhancement of binary particle swarm optimization for gene selection in classifying cancer classes. Algorithms for Molecular Biology 8(1), 15 (2013)
26. Mohamad, M., Omatu, S., Deris, S., Yoshioka, M.: A modified binary particle swarm optimization for selecting the small subset of informative genes from gene expression data. Information Technology in Biomedicine 15(6), 813–822 (2011)
27. Mohemmed, A., Zhang, M., Johnston, M.: Particle Swarm Optimization based Adaboost for face detection. In: IEEE Congress on Evolutionary Computation, pp. 2494–2501 (2009)
28. Narendra, P., Fukunaga, K.: A Branch and Bound Algorithm for Feature Subset Selection. IEEE Transactions on Computers C-26(9), 917–922 (1977)
29. Neshatian, K., Zhang, M., Andreae, P.: A filter approach to multiple feature construction for symbolic learning classifiers using genetic programming. IEEE Transactions on Evolutionary Computation 16(5), 645–661 (2012)

30. Oliveira, L., Sabourin, R., Bortolozzi, F., Suen, C.: Feature selection using multi-objective genetic algorithms for handwritten digit recognition. In: 16th International Conference on Pattern Recognition, vol. 1, pp. 568–571 (2002)
31. Pal, S., Chakraborty, B.: Fuzzy Set Theoretic Measure for Automatic Feature Evaluation. IEEE Transactions on Systems, Man and Cybernetics 16(5), 754–760 (1986)
32. Purohit, A., Chaudhari, N., Tiwari, A.: Construction of classifier with feature selection based on genetic programming. In: IEEE Congress on Evolutionary Computation, pp. 1–5 (2010)
33. Sahu, B., Mishra, D.: A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data. Procedia Engineering 38(0), 27–31 (2012)
34. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: IEEE Congress on Evolutionary Computation, pp. 69–73 (1998)
35. Sivagaminathan, R.K., Ramakrishnan, S.: A hybrid approach for feature subset selection using neural networks and ant colony optimization. Expert Systems with Applications 33(1), 49–60 (2007)
36. Stevanovic, A., Xue, B., Zhang, M.: Feature selection based on pso and decision-theoretic rough set model. In: IEEE Congress on Evolutionary Computation, pp. 2840–2847 (2013)
37. Subbotin, S., Oleynik, A.: The multi objective evolutionary feature selection. In: International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science, pp. 115–116 (2008)
38. Unler, A., Alper Murat, R.B.C.: mr2pso: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. Information Sciences 20, 4625–4641 (2011)
39. 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)
40. Wang, X., Yang, J., Teng, X., Xia, W., Jensen, R.: Feature selection based on rough sets and particle swarm optimization. Pattern Recognition Letters 28(4), 459–471 (2007)
41. Waqas, K., Baig, R., Ali, S.: Feature subset selection using multi-objective genetic algorithms. In: 13th International Conference on INMIC, pp. 1–6 (2009)
42. Xue, B., Cervante, L., Shang, L., Browne, W.N., Zhang, M.: Binary PSO and rough set theory for feature selection: A multi-objective filter based approach. International Journal of Computational Intelligence and Applications 13(02), 1450009 (2014)
43. Xue, B., Cervante, L., Shang, L., Browne, W., Zhang, M.: A multi-objective particle swarm optimisation for filter-based feature selection in classification problems. Connect. Sci. 24(2-3), 91–116 (2012)
44. Xue, B., Zhang, M., Browne, W.N.: Multi-objective particle swarm optimisation for feature selection. In: Genetic and Evolutionary Computation Conference (GECCO 2012), Philadelphia, PA, USA, pp. 81–88. ACM (2012)
45. Xue, B., Zhang, M., Browne, W.N.: Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. Applied Soft Computing 18(0), 261–276 (2014)
46. Xue, B., Zhang, M., Browne, W.N.: Novel initialisation and updating mechanisms in PSO for feature selection in classification. In: Esparcia-Alcázar, A.I. (ed.) EvoApplications 2013. LNCS, vol. 7835, pp. 428–438. Springer, Heidelberg (2013)
47. Xue, B., Zhang, M., Browne, W.: New fitness functions in binary particle swarm optimisation for feature selection. In: IEEE Congress on Evolutionary Computation, pp. 1–8 (2012)
48. Xue, B., Zhang, M., Browne, W.: Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE Transactions on Cybernetics 43(6), 1656–1671 (2013)

49. Yang, C.S., Chuang, L.Y., Ke, C.H., Yang, C.H.: Boolean binary particle swarm optimization for feature selection. In: IEEE Congress on Evolutionary Computation (CEC 2008), pp. 2093–2098 (2008)
50. Yang, C.S., Chuang, L.Y., Li, J.C., Yang, C.H.: Chaotic maps in binary particle swarm optimization for feature selection. In: IEEE Conference on Soft Computing in Industrial Applications, pp. 107–112 (2008)
51. Yong, Z., Dunwei, G., Ying, H., Wanqiu, Z.: Feature selection algorithm based on bare bones particle swarm optimization. Neurocomputing 148(0), 150–157 (2015)
52. Yu, X.M., Xiong, X.Y., Wu, Y.W.: A PSO-based approach to optimal capacitor placement with harmonic distortion consideration. Electric Power Systems Research 71(1), 27–33 (2004)