# Feature Selection Using Particle Swarm Optimization : A Review

[1]Vijay Hasan Puri, [2]Jyoti Vashishtha

[1]M.Tech Scholar, [2]Associate Professor
[1]Department of Computer Science & Technology
[1]Guru Jambheshwar University of Science & Technology, Hisar, India.

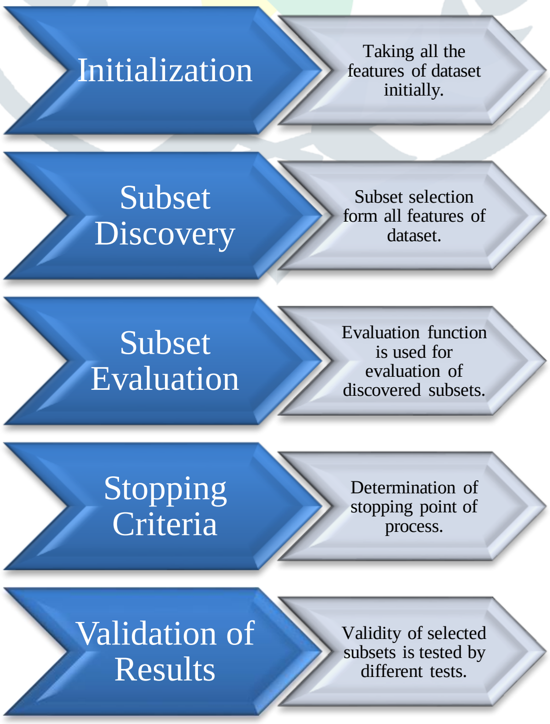
***Abstract :*** Classification problems occurred when unusual, irrelevant and redundant features are selected. This may degrade the classification performance and classification accuracy. Feature selection approach intended to select the less number of features which are relevant for improvement in classification performance than using all the features of corresponding dataset. Particle Swarm Optimization (PSO) approach is used for feature selection to achieve better results than traditional feature selection methods. Due to PSO's simplicity, many researchers attracted towards PSO. This paper proposes a concise audit on PSO approaches for highlight determination. Advancement, enhancements and alterations on PSO algorithms for feature selection are examined in this paper.

***Keywords* -** Feature Selection, Multi-Objective Feature Selection, Particle Swarm Optimization, Multi-Objective PSO.

## I. INTRODUCTION

Classification is very effective and important task of data mining in which model is generated from the training set to classify unknown data [1]. In classification, models are used to determine the class labels of objects whose class labels aren't known. Models are constructed from training set in which class labels are known. Classifier are constructed to determine the class label of test data where model is constructed from training data [1]. Basically classification is the process to classify the unknown data item from data set into one predefined class sets. There are some well-known classifier which are artificial neural network (ANN), K-nearest neighbor classifier (KNN), decision tree classifier (DT) and support vector machine (SVM) [2],[3],[4].

Actually, dataset often consists of number of features but all the features are not necessary for classification. Few features may be used for better classification. Dataset often flooded with irrelevant and redundant features which can be termed as unnecessary features which don't provide effective information [5],[6]. The way towards picking a subset of applicable or relevant features or non-redundant features for target finding operation is known as feature selection. Features which are irrelevant and redundant increases the computational complexities and affect the process of classification. Feature selection is concerned with main motive to remove the unusable, irrelevant and redundant feature for classification. Feature selection process have five important steps [1],[2],[5]:




1. Initialization : Initialization is the very first part which is concerned with taking all the features of dataset initially.
2. Subset Discovery : This step is concerned with selection of subsets from the all features of dataset. If dataset have N features than there will be $2^N$ feature subsets. So exhaustive search is practically impossible for some search strategies are used for evaluation. Complete, sequential, random and heuristic are commonly used search strategies.

3. Subset Evaluation : Discovered subsets are evaluated by evaluation function which is based on distance, information gain, dependency, consistency and classifier accuracy.
4. Stopping Criteria : This determines when the feature selection process will stop by checking stopping criteria which can be the search complete or reaching the bound limit of maximum iterations or selection of sufficiently good subset.
5. Validation of Result : Different test are carried out for validation of selected subsets.

Feature selection methods are of three types:

- Wrapper Method : Learning or classification algorithm is used in this approach which ensures that feature selection should consider classifier's characteristics. To evaluate the feature subset, this method uses the classifier error rate or accuracy of evaluation function obtained by learning algorithms [8],[9]. Wrapper methods provide higher performance of classifier than other simple filters. In this approach, a subset of features is selected and data is processed along with this selected feature subset to learning algorithm. Quality evaluation is done by the performance of learning algorithm. Many different algorithms can be used to select feature subset and wrapper methods have to train the classifier[10].
- Filter Method : This approach is independent for other learning algorithms. This approach doesn't use any classifier and provides a general view of feature selection. To evaluate a feature or subset of feature, a predetermined evaluation function is used. This measures the selective ability of the feature subset [10]. Different evaluation functions are used by different algorithms. There may be low performance if evaluation criteria doesn't matches to classifier.
- Embedded Method : This method obtained by combining the advantages of wrapper and filter method both. The name "embedded" is given because it uses two different evaluation methods. Independent measure and data mining algorithm both are used in this approach in which independent measure chooses the best subset and data mining chooses the finest subset from the best subsets [10],[11].

## II. Particle Swarm Optimization

In 1995, Eberhart and Kennedy developed an algorithm which is based on the natural behavior of bird flocking, known as Particle Swarm Optimization (PSO). PSO is totally based on intelligence and movement of particles of swarm and concept of natural social behavior to solve problem [12]. PSO concept can be understood by following example of birds searching for food. When a group of birds moves for searching food within a particular area then they don't know the position of food and moves random. When some of them found the food nearest to them then all other birds start following that bird for food. Individually each bird have limited potential but there cooperation in work made their behavior more intelligent [12]. PSO algorithms assigns fitness value to each particle and each particle randomly generate solutions and searches the optimal solution by iteratively changing the velocity and position based on previous velocity, own and group's flying experience. Optimal solution is determined at the end of iterative method [13].

PSO algorithm was developed by simulating the social behavior of flocking or birds and fitness concept is used in PSO. PSO uses a population of randomly generated particles which are associated with their respective velocity and position. Particle's velocity and position are changed on the basis of some equations which have parts as momentum, cognitive and social. Momentum ensures that velocity of particle cannot change quickly and uses information from previous velocity and position. Cognitive is concerned with the learning from flying experience of that particle. Social part is concerned with learning from flying experience of group [13],[14].

### 2.1 Continuous PSO:

A population of particles which are randomly generated and each particle is associated with its velocity and position, each particle corresponding to randomly generated solution. Optimal solution determined by PSO on changing the velocity and position iteratively which is based on moving experience of particle and group of particles towards gbest and pbest location. gbest is the best fitness value for whole population of the particles group and pbest is the particle's personal best fitness value that is achieved by the corresponding particle. Particle's velocity and position changes on the basis of following equation (i) & (ii) respectively :

Velocity : $$v_{id} = w * v_{id} + c_1 * rand() * (p_{id} - x_{id}) + c_2 * rand() * (p_{gd} - x_{id}) \quad ...(i)$$

Position : $$x_{id} = x_{id} + v_{id} \quad ...(ii)$$

Here $v_{id}$ and $x_{id}$ are ith particle's velocity and ith particle's position respectively in dimension d at time t, $p_{id}$ is pbest's location and $p_{gd}$ is gbest's location. On the basis of particle's distance of current position and previous velocity from gbest and pbest location, corresponding particle updates their velocity and position. Here, $c_1$ is Cognitive part which describes learning form particle from its flying or moving experience and $c_2$ is Social part which shows learning from flying or moving experience of particles group. The random function (*rand()*) generating random values in the range [0,1]. Here, $w$ is the weight of inertia or inertia weight which ensures the balance between local and global exploitation and exploration [12],[13],[14].

### 2.2 Binary PSO:

PSO was introduced and aimed to effectively handle real value problems which was further extended and improved to discrete or binary space to solve discrete/binary problems. Velocity was squashed using logistic function in discrete problems. Particles are moving in a restricted space from 0 to 1 on each dimension. Particle changes velocity according to equation …(i) except that $x_{id}$, $p_{id}$ and $p_{gd}$ are in range of 0 to 1 and $v_{id}$ must be constrained in the range of 0 to 1. Sigmoid function can word better to do this and particle's position is changed according to the following rule [15] :

If rand() < $S(V_{id})$
$x_{id}=1$;
else
$x_{id}=0$;

Where, $S(v) = \frac{1}{(1+e^{-v})}$

*S(v)* is sigmoid limiting transformation and rand() is randomly selected number from range [0,1].

## III. Multi Objective Optimization

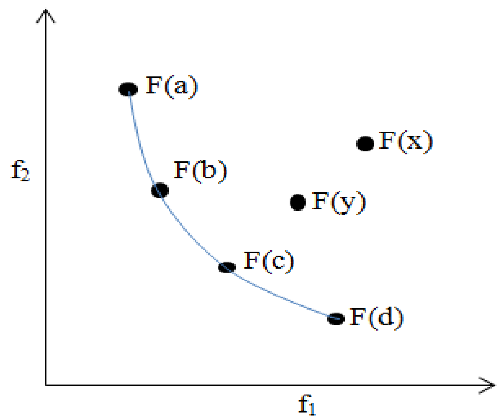
Multiple objectives are involved in Multi-objective optimization problem. In this optimization problem, two or more than two objective functions simultaneously optimized for effective optimization. In mathematical terms, this problem can be formulated as minimization issue or problem. Multi objective function can be composed as [16]:

$$\text{Minimizing, } F(x) = [f_1(x), f_2(x), \ldots, f_n(x)]$$

Subject to

$$Qi(x) \leq 0, i = 1, 2, \ldots k \quad \ldots\text{(ii)}$$

$$Ri(x) = 0, \quad i = 1, 2, \ldots l \quad \ldots\text{(iii)}$$




$f_1$ and $f_2$ function are minimized

Where *x* is considered as decision variable and *fi*(*x*) is a function of *x*, number of objective functions which are considered to be minimized are represented by *n* , and the problem have two constraint functions as $Qi(x)$ and $Ri(x)$. It is being connected in different fields of designing, engineering, science, financial matters and co-ordinations where ideal choices are required within the sight of exchange off between at least two conflicting objectives. A solution x dominates solution y if it fulfil following condition:

$$fi(x) <= fi(y) \;\&\&\; fi(x) < fi(y) \quad \ldots\text{(iv)}$$

If x and y solution of n objective problem satisfy the equation (iv), then these are called dominated solution. The solution become non-dominated or pareto optimal when the optimizing solution is not dominated by any other solutions . Non dominated solutions in search space represents a trade-off surface which termed as Pareto front. Features selection problem defined as a multi-objective problem dealing with two competing objectives, one is to minimize number of features and other is to minimize classification error [16],[17]. Multi-objective approach is the most studied field in the recent years has been applied in most of the data mining problem. Multi-objective approach has been used in the field of features selection using Particle swarm optimization to obtain pareto optimal solutions [18].

Multi-objective optimization algorithm determines an optimal solution which is called non-dominated and pareto optimal in which solution is non-dominated by other solutions. Multi-objective algorithms are intended to select the set or subset of solutions which are non-dominated. In terms of feature selection, there are two clashing objectives for feature selection which are intended to decrease classification error rate or increase classification accuracy and decrease the number of features selected [16],[17],[18].

## IV. Surveys on Feature Selection Methods

Generally, Feature selection approach is categorized in three parts as wrapper approach, filter approach and embedded approach. The most commonly or frequently used wrapper methods are SFS(Sequential Forward Selection) and SBS(Sequential Backward Selection), both performed the selection of features in a sequential manner [8][19]. Both algorithms are completely opposite of their working as SFS considers an empty set of features in starting and SBS initializes with all features of dataset. Selected or Candidate features are added in SFS feature set in sequence and sequentially removed in SBS algorithm. Both algorithms keep adding and removing the features as their working mechanism until they don't find any improvement in classification performance but both of these algorithms are suffering from nesting effect which may be because of similar objects and similar features.

To overcome this nesting effect, a new method is introduced which is known as Plus-i-take-away-r method. This algorithm was developed by integration of SFS and SBS which performs along a systematic addition and removal of features. In this method, i steps of SFS algorithm are followed by *r* steps of SBS algorithm for selecting the features. But suitable values selection of *i* and *r* is a cumbersome problem because both are user defined parameters. There may be two more possible solutions for nesting effect which are [19],[20] : SFFS ( Sequential Forward Floating Selection) and SBFS (Sequential Backward Floating Selection), which are taken under floating search methods category and both are same as Plus-i-take-away-r approach but SFFS and SBFS has dynamic control apart from Plus-i-take-away-r approach. SFFS and SBFS have the capability of adding and removing the features to the feature set at different stages of feature selection procedure until the algorithm obtained the desired number of features [21],[22].

A filter based algorithm was introduced for feature selection which is known as Relief. Relief generally ranks the features based on feature relevance property using a threshold value to select the feature subset. Relief only considers relevant features but It don't considers redundant features. Another filter based algorithm FOCUS was introduced using exhaustive approach for optimal feature subset with all features [23]. However, due to exhaustive approach FOCUS affected from high computational cost. We can obtain either optimal feature subset or computationally relevant and effective feature subset by using traditional feature selection approaches. There are very less chances to provide satisfactory results for large datasets by using traditional feature selection methods.

So, some more effective methods are needed for feature selection and moving towards nature based algorithm like PSO, Ant Colony etc. for feature selection was an effective step because of there global search ability. Some work discussing feature selection using PSO have been reviewed in this subsection.

Liam *et al.*[24 ] introduced two algorithms known as BPSO-P and BPSO-G which are used to find more optimal solution than traditional approaches. Mutual information and entropy (group of features) are used for evaluation of relevance between class labels and features. BPSO-P uses mutual information as a measure while BPSO-G uses entropy of group of features. Experiments performed on both algorithms giving results with suitable weight of relevance. Redundancy is also optimised in both algorithms. Both algorithms still have not been compared with others.

Chuang *et al.*[25] proposed an algorithm by resetting the value of gbest, whenever gbest became unchanged for several iterations. KNN (K -Nearest Neighbour) method with the collaboration of LOOCV approach is aimd to obtain the particle's fitness value. Experiments performed by Chuang *et al.*[17] are clearly ensuring that classification accuracy is more than other algorithms by this approach.

Yang *et al.*[26] suggested one more parameter to reset the value of gbest after several iterations when gbest remains unchanged. Replacement of old gbest fitness value with new gbest fitness value with Boolean function, when gbest remains same during three successive iterations.

Chuang *et al.*[27] proposed another method to prevent BPSO from getting stuck in local optima which is known as CBPSO ( Chaotic map embedded with BPSO) which is featured to adjust the weight of inertia. Various equations are used by chaotic map including logistic and tent map for updating weight of inertia effecting the search ability of proposed approach. Experiments performed are showing that CBPSO with tent maps performs better by achieving high classification accuracy than CBPSO with logistic map. Chuang *et al.*[28] proposed Catfish effect which enhanced the performance of BPSO. According to Catfish effect, new particles are placed on the place of old particles which are having worst fitness value for consecutive iterations. Catfish-BPSO approach performed better for feature selection and outperforms BPSO and deterministic algorithms.

All the work performed above are based on single objective to increase classification accuracy but there is also some work for feature selection using multi objective techniques. The work for feature selection is investigated frequently using evolutionary algorithms but first time Xue *et al.*[16] used PSO approach for feature selection. Instead of selecting binary PSO, Continuous PSO was selected because of particle's position was changing according to velocity of particle in BPSO. PSO considers both - velocity and position of the particle. Xue *et al.*[16] introduced two new feature selection methods using multi objective PSO, which are known as NSPSOFS(Non-Dominated Sorting PSO for Feature Selection) and CMDPSOFS(Crowding Mutation and Dominance based PSO for Feature Selection). NSPSOFS based on the non-dominated sorting concept while CMDPSOFS using the concept of crowding, mutation and dominance to find optimal feature set which is non-dominated solution. Different problems of local optima are sorted with CMDPSOFS [29]. NSPSOFS is less effective when avoiding stagnation in local optima while CMDPSOFS performed better. CMDPSOFS is more effective than NSPSOFS. Experimental results showing that both of the algorithms performing better than other algorithms for feature selection [30].

## V. Conclusions and Future Work

Particle Swarm Optimization (PSO) for feature selection intended to make increment in the classification accuracy and helped for selection of optimized feature subset. Many algorithms are introduced to select feature subset (Feature selection) for classification and made many modifications to overcome the limits of traditional feature selection methods. Algorithms such as NSPSOFS and CMDPSOFS increased the performance using PSO algorithm for feature selection. But still some more work can be done to increase classification accuracy in future using PSO. Researcher can use binary PSO which may help to select more optimized feature subset. Hybrid mutation is successfully applied for cost based feature selection but have not applied for feature selection without cost which may be another way to improvement.

## VI. References


[1] J. Han, M. Kamberand J. Pei, "Data Mining: Concepts and Techniques, Third Edition, MorganKaufmann", 2011.

[2] S. W. Lin, K. C. Ying, S. C. Chen and Z. J. Lee, "Particle swarm optimization for parameter determination and feature selection of support vector machines",*Expert Systems with Applications*, Vol. 35,No. 4, pp. 1817-1824, 2008.

[3] J. Vashishtha, D. Kumar, S. Ratnoo and Kapila, "Mining comprehensible and interesting rules: a genetic algorithm approach", *International Journal of Computer Applications*, Vol. 31, No. 1, pp.39–47, 2011.

[4] Neha and J. Vashishtha, "Particle Swarm Optimization based Feature Selection", *International Journal of Computer Applications*, Vol. 146, No.6, pp. 0975 – 8887, 2016.

[5] H. Liu, J. Sun, L. Liu, and H. Zhang, "Feature Selection with Dynamic Mutual Information", *Pattern Recognition*, Volume 42, Issue 7, pp. 1330-1339, 2009.

[6] A. Pathak and J. Vashishtha, "Classification Rule and Exception Mining Using Nature Inspired Algorithms", *International Journal of Computer Science and Information Technologies*, Vol. 6, No. 3, pp. 3023-3030, 2015.

[7] B. Yeoman, B. Xue and M. Zhang, "Particle Swarm Optimization for Feature Selection: A Size-Controlled Approach", In *Proceedings of the 13th Australasian Data Mining Conferences on Research and Practice in Information Technology, Sydney, Australia*, Vol. 168, 2015.

[8] S. Visalakshiand V. Radha, "Wrapper based Feature Selection and Classification for Real Time Dataset",*International Journal of Emerging Technologies in Computational and Applied Sciences*, Vol. 3, pp. 306–311, 2015.

[9] M. C. Lane, B. Xue, I. Liu and M. Zhang, "Particle Swarm Optimization and Statistical Clustering for Feature Selection", *Advances in Artificial Intelligence, Lecture Notes in Computer Science Heidelberg, Springer*,Vol. 8272, pp. 214–220, 2013.

[10] G. Kim, and S. Kim, "Feature Selection Using Genetic Algorithms for Handwritten character Recognition", *7th International Workshop on Frontiers in Handwriting Recognition*, pp. 103–112, Amsterdam, 2000.

[11] D. Cvetkovic, I. C. Parmee, "Preferences and their application in evolutionary multiobjective optimization", *IEEE Transactions on Evolutionary Computation*, Vol. 6, pp. 42-57, 2002.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization", In*Proceeding of International Conference on Neural Networks. IEEE*, Vol. 4, pp. 1942–1948, 1995.

[13] Y. Shi, "Particle Swarm Optimization", In *Proceeding of International Conference on Neural Networks Society IV, IEEE*, No. 1, pp. 8-13, 2004.

[14] Y. Shi and R. Eberhart, "A Modified Particle Swarm Optimizer", In *Proceeding of International Conference on Evolutionary Computation, Anchorage, AK, IEEE*, pp. 69–73, 1998.

[15] A. Unler and A. Murat. 2010. "A discrete particle swarm optimization method for feature selection in binary classification problems," In *European Journal of Operational Research*, Vol. 206, No. 3, pp. 528–539.

[16] B. Xue, M. Zhangand W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach", *Cybernetics, IEEE*, Vol. 43,No. 6, pp. 1656-167, 2013.

[17] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, V. G. da Fonseca, "Performance assessment of multiobjective optimizers: an analysis and review",*IEEE Transactions on Evolutionary Computation*, Vol. 7, pp. 117-132, 2003.

[18] Y. Zhang, D. W. Gong and J. Cheng, "Multi-objective Particle Swarm Optimization Approach for Cost-based Feature Selection in Classification", *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1-13, 2015.

[19] M. C. Lane, B. Xue,I. Liu and M. Zhang,"Gaussian Based Particle Swarm Optimization and Statistical Clustering for Feature Selection",*Evolutionary Computation in Combinatorial Optimization, Lecture Notes in Computer Science, Heidelberg, Springer*, Vol.8600, pp. 133–14, 2014.

[20] B. Xue, M. Zhang and W. N. Browne, "Particle swarm optimization for feature selection in classification: Novel initialization and updating mechanisms", *Applied Soft Computing, Elsevier*, Vol. 18, pp. 261–276, 2014.

[21] M. Dash and H. Liu, "Feature selection for classification", *Intelligent data analysis, Elsevier*, Vol. 1, No. 3, pp. 131–156, 1997.

[22] G. Chandrashekar and F. Sahin, "A survey on feature selection methods", *Computers & Electrical Engineering, Elsevier*, Vol. 40, No. 1, pp. 16–28, 2014.

[23] X. Wang, J. Yang, X.Teng, W. Xia and R. Jensen,"Feature selection based on rough sets and particle swarm optimization",*Pattern Recognition Letters, Elsevier*, Vol. 28, No. 4, pp. 459–471, 2007.

[24] L. Cervante, B. Xue, M. Zhang and L. Shang, "Binary particle swarm optimization for feature selection: A filter based approach", In *Proceedings of 2012 Congress on Evolutionary Computation, Brisbane, QLD, IEEE*, pp. 1–8, 2012.

[25] L. Y. Chuang, H. W. Chang, C. J. Tu and C. H. Yang, "Improved Binary PSO for Feature Selection Using Gene Expression Data",*Computational Biology and Chemistry, Elsevier*, Vol. 32, No. 1, pp. 29–38, 2008.

[26] C. S. Yang, L. Y. Chuang, C. H. Keand C. H. Yang,"Boolean binary particle swarm optimization for feature selection", In *Proceedings of Congress on Evolutionary Computation, Hong Kong, IEEE*, pp. 2093–2098, 2008.

[27] C. S. Yang, L. Y. Chuang, J. C. Li and C. H. Yang, "Chaotic maps in binary particle swarm optimization for feature selection" In *Proceedings of Conference on Soft Computing in Industrial Applications, Muroran, IEEE*,pp. 107–112, 2008.

[28] L. Y. Huang, S. W. Tsai and C. H. Yang, "Improved binary particle swarm optimization using catfish effect for feature selection", *Expert Systems with Applications, Elsevier*, Vol. 38, No. 10, pp. 12699–12707, 2011.

[29] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II",*Evolutionary Computation, IEEE*, Vol. 6, No. 2, pp. 182–197, 2002.

[30] M. R. Sierra and C. A. C. Coello, " Improving PSO Based Multi-objective Optimization Using Crowding, Mutation and ∈-Dominance. Evolutionary Multi- Criterion Optimization", *Lecture notes in computer science, Guanajuato, Mexico, Springer*, pp. 505–519, 2005.