IJCTA, 9(10), 2016, pp. 4801-4808

Dimensionality Reduction Methods Classical and Recent Trends : A Survey

***Shikha Agarwal *Prabhat Ranjan **R. Rajesh**

Abstract : High dimensionality is the problem for many research areas. There are huge number of dimensionality reduction methods are available. Broadly they are grouped into two categories feature selection and feature extraction. Feature selection methods select a subset of features based on some criteria while feature extraction methods transform the data in to the lower dimensional space. This paper presents a survey of classical and modern dimensionality reduction methods. Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization, Artificial Neural Network and Artificial Immune System are few modern nature inspired methods which have been applied for feature selection problem. To find the best feature selection methods, an experiment has been conducted using classical feature extraction, classical feature selection and nature inspired genetic algorithm and particle swarm optimization. Experimental results reveal that modern nature inspired particle swarm optimization is outperforming the other methods.

Keyword : High dimensionality, Feature selection, Nature inspired methods, Particle swarm optimization

1. INTRODUCTION

High dimensional data is defined as very high number of features as compared to the number of samples. This high dimensionality brings three unique features [1] namely; noise accumulation [2], spurious correlations [3] and incidental endogeneity [4], [5]. Due to these characteristics high dimensional data usually produces the surprising results like concentration of measure [6] and empty space phenomenon [7].

Analyzing high dimensional data requires simultaneous analysis of many features. Suppose for a feature selection problem, total number of features is p, one has to evaluate (2P-1) subsets of features to find minimum optimum features. For a small value of $p = 20$, this formula will produce 1048575 (around one million) subsets which are need to be evaluated. Therefore if number of variables are more than hundred than it will impose statistical, mathematical and computational challenges. Similarly for a classification problem, there are always a threshold number of features beyond which the performance of classifier will degrade rather than improve.

Dimensionality reduction methods are categorized into two categories, feature extraction and feature selection. Feature extraction transforms the higher dimension data into lower dimension while feature selection methods pick up some important features from pool of features to reduce the dimensionality.

In this paper section II deals with classical feature selection and feature extraction methods. Section III covers the nature inspired particle swarm optimization and other methods. Section IV presents experiment using different feature selection methods. Sections V present a discussion on experimental results. Section VI concludes the paper.

* Department of Computer Science, Central University of South Bihar, Patna, 800014, India shikhaagarwal@cub.ac.in, prabhatranjan@cub.ac.in

** Department of Computer Science, Central University of Kerala, Kerala, India kollamrajeshr@ieee.org

2. P DIMENSIONALITY REDUCTION

Feature extraction and feature selection are two broad categories of dimensionality reduction. Feature extraction reduces the dimensionality by projection of higher dimension vector into lower dimension vector. For a given feature set X= $\{x_i \mid i = 1, 2, 3, \ldots, m\}$, extracted features are defined as a transformed feature vector Y such that

$$\mathrm{Y} = f(\mathrm{X}) \tag{1}$$

where x_i is the i^{th} feature in feature set. f is a transformation function, Y is defined as Y= $\{\mathrm{Y}_i \mid i = 1, 2, 3, \ldots, m'\}$ and $m' < m$.

Some of the feature extraction methods, but not limited are; Principle Component Analysis (PCA) [8], linear discriminant analysis (LDA) [9], Factor Analysis (FA) [10], Locality Preserving Projection (LPP) [11], Non negative matrix factorization [12][13], Self Organizing map [14][15], Curvilinear Component analysis [16], Kernel PCA [17], nonlinear PCA [18], Laplacian eigen maps [19], manifold alignment [20], Diffusion maps [21], Hessian Locally linear embedding [22], Isomap [23], Locally linear Embedding [24], and Sammon mapping [25] etc.

Feature selection method selects the subset of features from complete set of the features without any transformation. For a given set of features vector X = $\{x_i \mid i = 1, 2, 3, ..., m\}$, selected feature subset is defined as a vector X = $\{x_i \mid i = 1, 2, 3, ..., m\}$, where x_i is the i^{th} feature in feature set and $m' < m$.

The main categories of feature selection methods are, filter, wrapper, embedded and ensemble feature selection methods.

A. Filter Feature Selection

In this method features are selected as pre-process without any information from fitness evaluator. Filter feature selection commonly apply some scoring function to rank each feature. Then a fixed number of top ranked features are selected.

1. Chi-square test [26] Chi square test checks the independence of two events. In feature selection two events are occurrence of a feature and class and score is calculated using (2).

$$\lambda^2 = \sum_{i=1}^{m}\sum_{j=1}^{c}\frac{(\mathrm{O}_{ij} - \mathrm{E}_{ij})^2}{\mathrm{E}_{ij}} \tag{2}$$

   where O_{ij} and E_{ij} are the observed and expected frequencies of ith feature in j^{th} class. m and c are total number of features and class respectively.

2. Information Gain [27] Information Gain method selects the subset of feature using information theory based on probability which measures the difference between probability of class in a data set and probability of class level for a given variable in data set. The amount of information gained using a feature with respect to class defined by (3).

$$\mathrm{IG} = \mathrm{H}(\mathrm{P}_c) - \mathrm{H}(\mathrm{P}_c \mid x) \tag{3}$$

   Where, $\mathrm{H}(\mathrm{P}_c)$ is the independent entropy of class c and $\mathrm{H}(\mathrm{P}_c \mid x)$ is the entropy of class c with respect to variable x.

3. Correlation Coefficient [28] Correlation based feature selection methods find highly correlated subset of features in which each feature is highly correlated with class variable. Pearson correlation coefficient is used as the fitness evaluator for the feature subset given by (4).

$$\text{fitness} = \frac{m' r_{cx}}{\sqrt{m' + (m' + 1) r_{xx}}} \tag{4}$$

   Where, m' is the number of feature in a subset, r_{xx} is the number of feature in a subset, r_{cx} is the measure of correlation between features in a subset.

4. Relief [29] Relief methods select the attributes based on the ability of attribute to distinguish two samples that are near to each other. Algorithm first randomly selects a sample, then search for two nearest neighbor one from same class called nearest hit (H) and other one from different class called nearest miss (M). Then according to the (5) attribute weight is updated.

$$\omega(x)_{new} = \omega(x)_{old} - \text{diff}\,(X_i, H, x) + \text{diff}\,(X_i, M, x) \tag{5}$$

$\omega(x)$ new and (x)old are the new and old weight of attribute x. diff(X_i, H, x) and diff(X_i, M, x) is the difference of value at attribute x for instances H and M respectively. Higher the average weight assigned to any attribute higher is the importance of the feature for classification.

B. Wrapper Feature Selection

Wrapper methods evaluate a subsets of features according to their performance for a given classifier. The three main categories of wrapper methods are, forward greedy wrapping, backward greedy wrapping, forward backward wrapping.

1. **Forward wrapping :** This method continuously adds feature one by one until no further improvement in the classification can be achieved. Hill climbing [30] method is an example of forward selection method.
2. **Backward wrapping :** This method start with full set of features and keep on removing one by one until no further improvement in classification performance can be achieved. Stepwise regression [31] elimination is the popular example of backward wrapper methods.
   INTERACT method is based on the interaction of features using backward elimination with C-consistency measurement. C-consistency of the irrelevant features will be minimum and for relevant feature it will be high [32]. Recursive feature elimination (SVM-RFE) is a wrapper method which performs backward elimination. SVM-REF finds the features which lead to the largest margin of class separation, and uses the weight vector was a ranking criterion [33].
3. **Forward backward wrapping :** In this method, features can be added as well as removed from the data until no further positive changes can be achieved. Hui Li et. al. [34] proposed a wrapper approach based on SVM for financial Distress identification (FDI) dataset.

C. Embedded feature selection methods

Embedded techniques have embedded feature selection mechanism for classification. Decision tree [35] classifiers of data mining have embedded feature selection methods which automatically determine optimal feature subset while building the tree. Other example of embedded methods is random forests [36] and methods based on regularization techniques like LASSO.

D. Other methods of Feature Selection

IG-GA is a two stage hybrid filter wrapper approach proposed by Karzynski et al [37]. Bio-geography based optimization (BBO) method is developed by Dan Simon [38] which works on the migration behavior of species and concept of mutation. In BBO-SVM wrapper method evaluates the fitness of selected subset of features using SVM as classifier [39]. IG-BBO-KNN-NN is proposed by Kumar et al [40]. Loris Nanni [41] have combined multiple feature reduction approaches for improving classification performance. A hybrid statistical pattern recognition algorithm is proposed [42][43] in which features with high class-correlation are first selected and redundant features with high inter correlation are eliminated. Yu et al [44] proposed predominant correlation method. Yu et al [45] also proposed fast correlation based filter method based on linear correlation and information theory.

3. NATURE INSPIRED METHOD

Nature inspired methods are computational model of nature that mimics the solution finding capability of nature. The main goal of most of the nature inspired algorithm is to find global optimum solution of a problem. Two key factors common in all nature inspired algorithm are exploration and exploitation. Exploration leads to the random search of solution for finding global optima and exploitation finds the local optimum in explored solution space. A balance is very essential between these two key factors to achieve optimum solution.

A. Particle Swarm Optimization [46]

PSO is a population based stochastic search method. It is inspired by the flock of birds, fish and other animals searching for food. The ith particle is m dimensional vector denoted by $X_i = (x_1, x_2, x_3, \ldots, x_m)$. Here m is the number of feature or dimension of the search space. The optimal particle of the swarm represents the global solution gbest and best position of any particle based on its fitness is represented as local best or pbest. Position of each particle is a candidate solution. Based on the fitness pbest and gbest are gets calculated which are used to update velocity and position (6) and (7) given by Kennedy and Eberhart [1] to update the velocity and position.

$$v(t+1) = v(t) + c_1 r_1(\text{pbest}(t) - X(t)) + c_2 r_2(\text{gbest}(t) - X(t)) \quad (6)$$

$$X(t+1) = x(t) + v(t+1) \quad (7)$$

Where, $v(t+1)$ and $v(t)$ are particle velocity at time $(t+1)$ and t respectively. pbest(t) and gbest(t) are local and global best position of particle respectively at time t. c_1 and c_2 acceleration (learning) factors, and rand, r_1 and r_2 are random numbers. X(t) is the position of particle at time t.

In 1998 Shi and Eberhart [47] introduced the coefficient in equation of velocity (omega w) (8) which are needed to control the velocity and balance between exploration and exploitation

$$v(t+1) = wv(t) + c_1 r_1(\text{pbest}(t) - X(t)) + c_2 r_2(\text{gbest}(t) - X(t)) \quad (8)$$

Where, w is the inertia weight.

Rajesh and Shikha have proposed some modifications in improved binary PSO [48] for dimensionality reduction problem [49][50].

B. Other Nature Inspired methods for Feature selection

Palomino and Liang has proposed FM-PGA, A map reduced-based hybrid of FM-test and parallel GA [51]. Nalepa and Kawulok have developed a mimetic algorithm for fast and efficient selection of valuable training set for SVM [52]. Kashef et al have developed advance binary Ant Colony Optimization [53]. Chen et. al proposed any colony optimization for feature selection (ACO-FS) [54]. An ACO based under sampling method was proposed by Yu et al [55]. In artificial immune system weighted feature selection (AIS-WFS) weights are assigned to features to improve the accuracy of classification [56]. In Bare Bones PSO (BBPSO) algorithm a reinforced memory strategy is designed to update the local leaders of the particles for avoiding the degradation of outstanding feature of the particles [57]. An improved artificial immune recognition system developed by Wang et al using the opposite sign tests for feature selection [58]. K. Anitha has proposed gene selection method based on rough set theory using attribute reduction by quick-reduct based genetic algorithm [59]. Kalaiselvi et al proposed PSO-LSVM based method for classification and predication [60].

4. EXPERIMENT AND RESULTS

Different dimensionality reduction methods have been applied on bench marking Car, Diabetes, and Arrhythmia and Breast cancer datasets. Details of datasets are given in Table I. Nine different dimensionality reductions are included in experiment which is filter methods and wrapper methods. Included filter methods are Chi square, Information gain, Correlation Coefficient and Relieff and wrapper methods are ITRACT, Incremental Wrapper. Nature inspired PSO, Evolutionary method and genetic algorithm are applied. The experiment is performed using weka 3.7.

Table 1. Description of Datasets

Dataset	*No. of Samples*	*No. of Features*	*No. of Classes*
Car	1728	7	4
Diabetes	768	9	2
Arrhythmia	452	452	16
Breast Cancer	97	24481	2

After dimensionality reduction performance of selected features is evaluated using K-Nearest Neighbor classifier. The results are shown in Table II. From Table II it is clear that new nature inspired methods are outperforming the classical methods. Among modern nature inspired methods particle swarm optimization is out performing.

5. CONCLUSION

In this study, a small survey of classical and nature inspired methods for dimensionality reduction has been presented. The experimental result after dimensionality reduction using K-nearest neighbor (K-NN) as evaluator shows that although all nature inspired methods are outperforming the classical methods. And among all nature inspired methods performance of PSO is noticeable. This survey could help the researchers to select feature selection method while dealing with high dimensional datasets like breast cancer data set.

Table 2. Experimental Results showing classification accuracy, precision, recall, fall positive rate (FPR), ROC analysis and selected features (SF) on bench marking data sets.

Methods	*Datasets*	*Accuracy*	*Precision*	*Recall*	*FPR*	*ROC*	*SF*
Chi-square	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	68.099	0.682	0.681	0.379	0.655	4
	Arrhythmia	52.6549	0.509	0.527	0.297	0.424	250
	Breast Cancer	71.6783	0.689	0.717	0.527	0.693	5
Information Gain	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	68.099	0.682	0.681	0.379	0.655	4
	Arrhythmia	52.6549	0.509	0.527	0.297	0.424	250
	Breast Cancer	71.6783	0.689	0.717	0.527	0.693	5
Correlation Coefficient	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	70.1823	0.702	0.702	0.357	0.681	3
	Arrhythmia	52.6549	0.509	0.527	0.297	0.424	250
	Breast Cancer	71.6783	0.689	0.717	0.527	0.693	5
Relieff	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	70.1823	0.696	0.702	0.378	0.650	5
	Arrhythmia	52.6549	0.509	0.527	0.297	0.424	250
	Breast Cancer	71.6783	0.689	0.717	0.527	0.693	5
INTRACT	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	68.099	0.682	0.681	0.379	0.655	4
	Arrhythmia	55.7522	0.503	0.558	0.296	0.634	24
	Breast Cancer	70.979	0.678	0.710	0.544	0.620	8
Incremental Wrapper	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	68.099	0.682	0.681	0.379	0.655	4
	Arrhythmia	55.97	0.377	0.560	0.388	0.586	1
	Breast Cancer	69.9301	0.691	0.699	0.446	0.599	1
PSO	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	66.7969	0.669	0.668	0.396	0.643	3
	Arrhythmia	55.0885	0.536	0.551	0.231	0.659	23
	Breast Cancer	70.2797	0.676	0.703	0.519	0.616	7

Methods	*Datasets*	*Accuracy*	*Precision*	*Recall*	*FPR*	*ROC*	*SF*
Evolutionary Computing	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	66.7969	0.669	0.668	0.396	0.643	3
	Arrhythmia	40.44	0.399	0.4000	0.272	0.560	6
	Breast Cancer	70.2797	0.676	0.703	0.519	0.616	7
Genetic Algorithm	**Car**	94.213	0.943	0.942	0.042	0.983	5
	Diabetes	66.7969	0.669	0.668	0.396	0.643	3
	Arrhythmia	49.115	0.478	0.491	0.287	0.600	36
	Breast Cancer	70.2797	0.676	0.703	0.519	0.616	7

6. ACKNOWLEDGMENT

Shikha Agarwal, awarded with Senior Research Fellowship (Grant No. (09/1144(0001)2015EMR-I) wish to acknowledge the financial support of Council of Scientific & Industrial Research (CSIR), India.

7. REFERENCES

1. J. Fan, F. Han and H. Liu, "Challenges of Big Data Analysis," Natl Sci Rev, vol. 1, no. 2, pp. 293-314, 2014.
2. J. Fan and Y. Fan, "High Dimensional classification using feature annealed independence rules," Ann. Stat., vol. 36, no. 6, pp. 2605-2637, 2008.
3. J. Fan, S. Guo and N. Hao, "Variance estimation using refitted crossvalidation in ultrahigh dimensional regression," J R Stat Soc Series B Stat Methodology, vol. 74, no. 1, pp. 37-65, 2012.
4. Y. Liao and W. Jiang, "Posterior consistency of nonparametric conditional monent restricted modeld," Ann. Stat., vol. 39, no. 6, pp. 3003-3031, 2011.
5. J. Fan and Y. Liao, "Endogeneity in ultrahigh dimension," Ann. Stat., vol. 42, pp. 872-917, 2014.
6. D. Donoho, "High-Dimensional Data Analysis: The Curse and Blessings of dimensionality," AMS Math Challenges Lecture, pp. 1-32, 2000.
7. D. W. Scott and J. R. Thompson, "Probability density estimation in higher dimensions," in Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface, J.E. Gentle Eds. Amsterdam, New York, 1983.
8. K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space," Philosophical Magazine, vol.2, no. 11, pp. 559-572, 1901.
9. R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, vol. 7, no. 2, pp. 179-188, 1936, doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227.
10. R. B. Cattell Factor analysis, New York, Harper & Brothers, 1952.
11. X. He, P. Niyogi, "Locality preserving projections," Proceedings of the Conference on Advances in Neural Information Processing Systems, pp. 585-591, 2003.
12. I. Dhillon, S. Sra "Generalized Nonnegative Matrix Approximations with Bregman Divergences," Neural Information Processing Systems (NIPS), pp. 283-290, Dec, 2005.
13. R. Tandon, S. Sra, "Sparse nonnegative matrix approximation: new formulations and algorithms," Technical Report, Max Planck Institute for Biological Cybernetics, 2010.
14. T. Kohonen, T. Honkela, "Kohonen Network, " Scholarpedia, vol. 2 , no. 1, 1568, 2007. DOI: 10.4249/scholarpedia.1568
15. T. Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps," Biological Cybernetics, vol. 43, no. 1, pp. 59-69, 1982. doi:10.1007/bf00337288.
16. P. Demartines and J. Hérault, "Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets," IEEE Transactions on Neural Networks, vol. 8, no. 1, pp. 148-154, 1997.
17. B. Schölkopf, A. Smola, K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural

Computation," MIT Press Cambridge, MA, USA, vol. 10, no. 5, pp. 1299-1319, 1998, doi: 10.1162/089976698300017467.

18. M. Scholz, F. Kaplan, C. L. Guy, J. Kopka, J. Selbig, "Non-linear PCA: a missing data approach," In Bioinformatics, vol. 21, no. 20, pp. 3887-3895, Oxford University Press, 2005.
19. M. Belkin and P. Niyogi, "Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering," Advances in Neural Information Processing Systems, vol. 14, pp. 586-691, MIT Press, 2001.
20. C. Wang, S. Mahadevan, "Manifold Alignment using Procrustes Analysis" The 25th International Conference on Machine Learning. pp. 1120-1127, 2008.
21. S. Lafon, "Diffusion Maps and Geometric Harmonics," PhD Thesis, Yale University, May, 2004.
22. D. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," Proc Natl Acad Sci, vol. 13, no. 100, pp. 5591-5596, 2003.
23. J. B. Tenenbaum, V. de Silva, J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," Science, vol. 290, pp. 2319-2323, 2000.
24. S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol 290, pp. 2323-2326, 2000.
25. J. W. Sammon, "A nonlinear mapping for data structure analysis," IEEE Transactions on Computers, vol. 18, pp. 401-402, 1969, DOI: 10.1109/t-c.1969.222678.
26. F. Yates, "Contingency table involving small numbers and the 2 test," Supplement to the Journal of the Royal Statistical Society, vol. 1, no. 2, pp. 217-235, 1934.
27. T. M. Mitchell, "Machine Learning" The Mc-Graw-Hill Companies, 1997, Inc. ISBN : 0070428077.
28. A. L. Edwards, "The Correlation Coefficient, " Ch. 4 in An Introduction to Linear Regression and Correlation. San Francisco, CA: W. H. Freeman, pp. 33-46, 1976.
29. L. Bai, Z. Wang, Y. H. Shao and N. Y. Deng, "A novel feature selection method for twin support vector machine," Knowledge Based Systems, Elsevier, vol. 59, pp. 18, 2014.
30. S. J. Russell, P. Norvig, "Artificial Intelligence: A Modern Approach," (2nd ed.), Upper Saddle River, New Jersey, Prentice Hall, pp. 111-114, ISBN 0-13-790395-2.
31. M. A. Efroymson, "Multiple regression analysis," In Ralston, A. and Wilf, HS, editors, Mathematical Methods for Digital Computers. Wiley, 1960.
32. Z. Zhao and H. Liu, Searching for Interacting Features, Proceedings of International Joint Conference on Artificial Intelligence, pp. 1156-1161, 2007.
33. I. Guyon, "Gene Selection for Cancer Classification using Support Vector Machines," Machine Learning, vol. 46, pp. 389-422, 2002.
34. H. Li, C.-J. Li, X. J. Wu and J. Sun, "Statistics-based wrapper for feature selection: An implementation onfinancial distress identification with support vector machine," Applied Soft Computing, Elsevier, vol. 19, pp. 57-67, 2014.
35. L. Rokach, O. Maimon, "Data mining with decision trees: theory and applications," World Scientific Pub Co Inc., ISBN-13: 978-9812771711.
36. J. R. Quinlan, "Induction of Decision Trees," Machine Learning, , Kluwer Academic Publishers, vol. 1, pp. 81-106, 1986.
37. T. K. Ho, "Random Decision Forests," Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, pp. 278-282, 1995.
38. M. Karzynski, A. Mateos, and J. Dopazo, "Using a Genetic Algorithm and a Perceptron for Feature Selection and Supervised Class Learning in DNA Microarray Data," Artificial intelligence review, vol. 20, no. 1, pp. 39-43, 2005.
39. D. Simon, "Biogeography Based Optimization," IEEE transactions on evolutionary computation, vol. 12, no. 6, pp. 702-713, 2008.
40. A. P. Kumar, P. Valsala, "Feature Selection for high Dimensional DNA Microarray data using hybrid approaches," Bioinformation, vol 19, no. 6, 2013
41. L. Nanni, S. Brahnam, A. Lumin, Combining multiple approaches for gene microarray classification," Bioinformatics, vol. 28, no. 8, pp. 1151-1157, 2012.

42. A. Sundaram, N. L. Venkata, R. S. Parthasarathy, "Hybrid SPR algorithm to select predictive genes for effectual cancer classication," Turk J Elec Eng & Comp Sci, vol. 21, pp. 2357-2366, 2013.

43. Y. Saeys, T. Abeel, Y. V. Peer, "Robust feature selection using ensemble feature selectiontechniques," W. Daelemans et al. (Eds.): ECML PKDD, Part II, LNCS (LNAI) 5212, pp. 313-325, 2008.

44. L. Yu, H. Liu, "Efficiently Handling Feature Redundancy in High Dimensional Data," ACM, August 27, 2003.

45. L. Yu, H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," Proceedings of the Twentieth International Conference on Machine Learning, 2003.

46. J. Kennedy and R. C. Eberhart, Particle swarm optimization In: IEEE International Conference on Neural Networks, Perth, Australia, pp. 1942-1948, 1995.

47. Y. Shi and R. C. Eberhart, "A modified particle swam optimizer," IEEE Word Congress on Computational Intelligence, pp. 69-73, 1998.

48. 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, vol. 32, pp. 29-38, 2008.

49. R. Rajesh and S. Agarwal, "Some Modification in Particle Swarm Optimization," The 18th Online World Conference on Soft Computing in Industrial Applications, 2014. In press

50. S. Agarwal and R. Rajesh, "Enhanced Velocity BPSO and Convergence Analysis on Dimensionality Reduction," in proceedings of Recent Advances In Mathmatics, Statistics and Computer Science, 2015, in press.

51. R. A. Palomino and L. R. Liang, "Cloud Parallel Genetic Algorithm for Gene Microarray Data Analysis," 23rd IEEE International Conference on Tools with Artificial Intelligence, pp. 1082-3409, 2011.

52. J. Nalepa and M. Kawulok, "A memetic algorithm to select training data for support vector machine," GECCO'14, Proceedings of conference on Genetic and Evolutionary Computation, pp. 573-580. New York, NY, USA: ACM, 2014.

53. S. Kashef and H. Nezamabadipour, "An Advanced ACO algorithm for feature subset selection," Neurocomputing, vol. 147, pp.271-179, 2014.

54. B. Chen, L. Chen and Y. Chen, "Efficient ant colony optimization for images feature selection," Signal Processing, vol. 93, pp. 1566-1576, 2013.

55. H. Yu, J. Ni and J. Zhao, "ACOSampling: An ant colony optimization-based undersampling method for classification imbalanced DNA microarry data," Neuro computing, vol. 101, pp. 309-318, 2013.

56. V.Karunakaran and D. M. Suganthi, "Multi objective artificial immune system weighted feature selection for classification," Int. J. Research in Computer Application and robotics, vol. 2, no. 5, pp. 96-101, 2014.

57. Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature Selection algorithm based on bare bones particle swarm optimization," Neurocomputing, vol. 148, pp. 150-157, 2015.

58. K. J. Wang, K. H. Chen and M. A. Angelia, "An improved artificial immune recogniotion system with the opposite sign test for feature selection," Knowledge Based Systems, vol.71, pp.126-145. 2014.

59. K. Anitha, "Gene Selection Based on Rough Set," International Journal of Computing Algorithm, vol. 1, no. 2, pp. 38-41, 2012. DOI: 10.20894/IJCOA.101.001.002.004, ISSN: 2278-2397.

60. C. Kalaiselvi, G. M. Nasira, "Classification and Prediction of Heart Disease from Diabetes Patients using Hybrid Particle Swarm Optimization and Library Support Vector Machine Algorithm," International Journal of Computing Algorithm, vol. 4, no. 2, pp. 54-58, 2015. doi: 10.20894/IJCOA.101.004.002.001, ISSN: 2278-2397.