ISSN: 2321-8169
1895 - 1899

A Survey on Feature Selection Algorithms

Amit Kumar Saxena
Professor, Department of CSIT
Guru Ghasidas Vishwavidyalaya
Bilaspur, C.G., India
Amitsaxena65@rediffmail.com

Vimal Kumar Dubey
Research Scholar, Department of CSIT
Guru Ghasidas Vishwavidyalaya
Bilaspur, C.G., India
vimal.dubey15@gmail.com

Abstract- One major component of machine learning is feature analysis which comprises of mainly two processes: feature selection and feature extraction. Due to its applications in several areas including data mining, soft computing and big data analysis, feature selection has got a reasonable importance. This paper presents an introductory concept of feature selection with various inherent approaches. The paper surveys historic developments reported in feature selection with supervised and unsupervised methods. The recent developments with the state of the art in the on-going feature selection algorithms have also been summarized in the paper including their hybridizations.

Keywords: *Feature selection, data mining, machine learning and pattern recognition*.

*****

INTRODUCTION

Modern electronic gazettes have been used very often to record every piece of information to store in their datasets. E.g. an EEG [1] analyst can record large data for sleep analysis [2]. Thus the datasets are becoming larger and larger day by day. The data in so generated datasets can be in a proper format like a tabular data (or even a spreadsheet) having a fixed number of columns or unformatted like text of this paper or tweets. It results in very high dimension and high instances dataset [3, 4]. By dimension, we mean number of columns or properties or features specifically. By instances, we will mean the number of records. For this paper, we assume only formatted databases. Thus every instance will have same number of features with some values each (missing or noisy values in exceptional cases). Some features in a database may be irrelevant and few can even be noisy. However many features can be very essential for data mining [5] process. So removing irrelevant and noisy features and retaining critical and useful features in a dataset is an important task. Eliminating redundant, noisy and irrelevant features from datasets is defined as Dimensionality Reduction. Dimensionality reduction is used in face image dataset [6,7,8,9], micro array dataset and speech signals [6,7], digit images [6,7,8], letter images[8,10] for classification or clustering. Feature selection refers to selecting a subset of features from a complete set of features in a dataset. Feature selection
is used for the classification of various data sets like Abalone, Glass, Iris, Letter, Shuttle, Spam base, Tae, Vehicle, Waveform, Wine, Wisconsin and Yeast data [7]. Feature selection and dimensionality reduction possess mostly a common goal which is to reduce the number of harmful, irrelevant and noisy features in a dataset for smooth and fast data processing purposes.

FEATURE SELECTION

Feature extraction and feature selection are used as two main techniques for Dimensionality Reduction [11,12,13]. Getting (or creating) a new feature, from the existing features of datasets, is termed as feature extraction [12]. Feature selection [11] is the process of selecting a subset of features from the entire collection of available features of the dataset. Thus for feature selection, no preprocessing is required as in case of feature extraction. Usually the objective of feature selection is to select a subset of features for data mining or machine learning applications [14]. Feature selection [14,15,16] can be achieved by using supervised and unsupervised methods. The process of Feature selection is based mainly on three approaches viz. filter, wrapper [17] and embedded (Fig. 1).

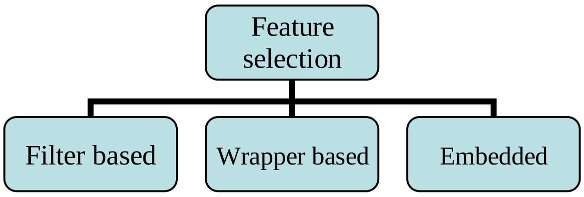

Fig. 1: Three approaches of Feature Selection

Filter based feature selection Algorithms: Removing features on some measures (criteria) come under the filter approach of feature selection. In the filter based feature selection approach, the goodness of a feature is evaluated using statistical or intrinsic properties of the dataset. Based on these properties, a feature is adjudged as the most suitable feature and is selected for machine learning or data mining [18] applications. Some of the common approaches of feature selection are Fast Correlation based filter, see Fig (2) (FCBF) [19] Correlation based feature selection (CFS) [20].

Wrapper based feature selection Algorithms*:* In the wrapper approach of feature selection, subset of features is generated and goodness of subset is evaluated using some classifier. An example of wrapper approach is GA-KDE-Bayes [21].

Embedded based feature selection Algorithms: In this approach, some classifier is used to rank the features in the dataset. Based on this rank, a feature is selected for the required application. SVM-RFE is one of the embedded feature selection approaches [22], refer Fig 3.

LITERATURE SURVEY

There has been an immense amount of work in the area of feature selection which credits its importance. In this section, we tour the growth of feature selection on various paradigms. This section is divided into two parts: Initial study on feature selection and then most recent developments in feature selection models.

Initial study: At their beginning, datasets with small number of features or variables were used for machine learning. As time progressed, number of features also kept on increasing in datasets. A feature selector model was proposed

input: S(F1, F2, ..., FN , C) // a training data set , a predefined threshold
output: S_{best} // an optimal subset
begin
for i = 1 to N do begin
calculate $SU_{i,c}$ for F_i;
if ($SU_{i,c} \geq \delta$)
append F_i to S^1_{list};
end;
order S^1_{list} in descending $SU_{i,c}$ value;
F_p = getFirstElement(S^1_{list});
do begin
F_q = getNextElement(S^1_{list}, Fp);
if (F_q <> NULL)
do begin
$F^1_q = F_q$;
if ($SU_{p,q} \geq SU_{q,c}$)
remove F_q from S^1_{list};
F_q = getNextElement(S^1_{list}, F^1_q);
else F_q = getNextElement(S^1_{list}, F_q);
end until (F_q == NULL);
Fp = getNextElement(S^1_{list}, F_p);
end until (F_p == NULL);
$S_{best} = S^1_{list}$;
end;

Fig (2) [19]

based on pre-weighted feature observations [23]. In the linear feature selection theory, it is shown that Eigen values, used therein do not affect the values of divergence in reduced dimension space [24]. F-entropies based probability of error model is used for feature selection in [25]. Karhunen-Loeve expansion based Linear Transformation which was applied on physiological variable for 100 patients for feature extraction resulted in selection of only two variables [26]. A method on the basis of approximation of class conditional densities by a mixture of parameterized densities of a special type is used for feature selection [27]. Medical axis transformation with artificial neural network (ANN [28]) is used for human chromosome classification [29]. Adaptive partitioning based discriminability measure was used as a criterion in a classifier-independent feature selection procedure [30]. Distinction Sensitive Learning Vector Quantization (DSLVQ) is used with weighted distance function for the selection of most informative input feature [31]. Fourier transforms (FT [32]) are also used by Wu etal to reduce the number of features [33].

Algorithm SVM-RFE:
Inputs:
X0 = [x1, x2,... xk,...xl]^T // Training examples
y = [y1, y2,... yk,...yl]^T // Class labels
Initialize:
s = [1, 2,... n] // Subset of surviving features
// Feature ranked list
r = [ ]
Repeat until s =[]
X = X0(:,s) // Restrict training examples to good feature indices
α = SVM-train(X, y) // Train the classifier
// Compute the weight vector of dimension length(s)
w = $\sum_k$ αk yk xk
// Compute the ranking criteria
ci = (wi)^2 , for all i //
// Find the feature with smallest ranking criterion
f = argmin(c)
// Update feature ranked list
r = [s(f), r]
// Eliminate the feature with smallest ranking criterion
s = s(1:f − 1, f + 1:length(s))
Output:
Feature ranked list r.

Fig. 3 [22]

Recent developments An unsupervised learning approach with maximum projection and minimum redundancy for feature selection is proposed by Wang [4]. Regularized self representation model is developed for dimensionality reduction in case of unlabeled data [5]. Matrix factorization [34] problem is modified for feature selection in case of unlabeled data [6]. Maximal relevance and maximal complementary (MRMC) based on neural networks is developed for feature selection [35]. Ant colony optimization (ACO [36]) with some modifications termed ABACO is used for feature selection [10]. Graph regularized Non Matrix Factorization (GNMF) is developed for feature selection [7]. An objective function is defined which finds a subspace such that all samples of the subspace are very far from all other samples. This objective function is optimized iteratively and resulted in unsupervised feature selection [37]. Mutual Information-based multi-label feature selection method using information interaction method is developed and is used in feature selection for multi label datasets. This algorithm measures dependency among features [38]. A new combined, document frequency and Term frequency feature selection method (DTFS) is developed for email classification [39]. Memetic feature selection algorithm for multi-label classification is developed to prevent premature convergence and gives better accuracy [40]. Fuzzy-rough [41] feature selection based on forward approximation is developed and used for feature selection on high dimensional dataset [42]. Fast Fourier Transform (FFT [43]) based feature selection algorithm is developed for mechanical system. It is highly robust with fast response and automation [44]. Modularity Qvalue-Based Feature Selection (CMQFS) model is developed using two methods. First one is mutual information-based criterion and second one relevant independency between features [45]. Kernelized fuzzy rough sets (KFRS) and the Memetic algorithm (MA) is hybridized for transient stability assessment of power systems. Memetic algorithm based on Binary Differential Evolution (BDE [46]) and Tabu Search

(TS [47]) is employed to obtain the optimal feature subsets with maximized classification rate [48]. A novel ensemble algorithm is developed for feature selection and named as Robust Twin Boosting Feature Selection (RTBFS) [49]. Conjoint analysis is applied for feature selection in the scenario of consumer preferences over the potential products [50]. Cuttlefish Algorithm (CFA) is implemented for feature selection in intrusion detection system. Decision tree is taken as a classifier [51]. In Autism dataset, eight feature selection methods are applied and finally a model is developed for fusion of these results. Most responsible gene is selected from autism dataset using clusterization with PCA [52] and Statistical Characteristics [53]. Gravitational search algorithm [54] is applied for feature search algorithm in machine learning [55]. Local information based feature selection algorithm is developed by Peng and Xu[58]. Modified Binary particle swarm optimization (MBPSO) is developed for feature subset selection. It optimizes support vector machine's (SVM [59]) kernel parameters as well as controls the premature convergence of the algorithm [60]. Fuzzy criteria are applied for the minimization of the classification error rate and the minimization of the feature cardinality [61]. Neighbourhood evidential decision error theory is used to find relevant features [62]. Adaptive feature selection using validity index is developed for Text streaming clustering [63]. Greedy Randomized Adaptive Search Procedure (GRASP) Metaheuristic, filter-Wrapper based algorithms are used for feature selection. It is a multi start constructive methods construct a solution in first stage and then runs for improving solutions [64]. Conservative subset selection method is introduced for missing values in the datasets. The proposed algorithm selects the minimal subset of features that renders other features without making any assumptions for the rest of missing data [65]. Genetic algorithm (GA[66]) is implemented as a filter model based feature subset selection method. From trained neural network (NN[32]), two types (input-hidden and hidden-output) of weights are extracted. General formula for each node is then generated and genetic algorithm is used to optimize these formulae as these formulae are based on inputs only [67]. Ant colony optimization [36] algorithm is implemented for feature selection in case of text categorization [68]. Hierarchical search framework and Tabu search method combines for getting optimal feature subset [69]. Bit based feature selection method is developed. It has two phases, in first phase bitmap indexing matrix is created from given dataset and in second phase a set of relevant features are selected for classification process and judged by domain expertise [70]. Hybrid method based on ant colony optimization and artificial neural networks (ANNs) is used for feature selection [71]. Group method of data handling (GMDH) algorithm is developed, where features are ranked according to their predictive quality using properties unique to learning algorithms [72]. Relative importance factor (RIF) is a measure developed to get information about less relevant features. Removing these less relevant features from dataset, results in higher accuracy and less computational time [73]. Rough sets theory given by Pawlak [74], has also been used for feature selection [75]. Feed forward neural network is used for feature selection. It is trained with an augmented cross-entropy error function. This function forces the neural network to keep low derivatives of the transfer functions of neurons when learning a classification task. It is applied over three real-world classification problems and gets higher accuracy [76]. A wrapper based evolutionary, population-based, randomized search algorithm FSS-EBNA (Feature Subset Selection by Estimation of Bayesian Network Algorithm), is developed. Naive-Bayes [77] and ID3 [78] learning algorithms are used as evaluator of feature subset. It uses EDA (Estimation of Distribution Algorithm) paradigm, avoids the use of crossover and mutation operators to evolve the populations, as in Genetic Algorithms. Evolution is guaranteed by the factorization of the probability distribution of the best solutions found in a generation of the search [79]. Genetic algorithm with Sammon's stress function as the fitness function is used to reduce the dimension of the datasets using unsupervised feature selection and still preserving the topology of the high dimensional dataset [80].

CONCLUSION

Feature selection is an important part of most of the data processing applications including data mining, machine learning and computational intelligence. The feature selection methods can be categorized under filter, wrapper and embedded approaches. This paper presented a review of some of the feature selection algorithms spanned over last five decades. The Unsupervised and supervised feature selection using various soft computing techniques viz. Neural networks, fuzzy sets, meta-heuristics, genetic algorithms, PSO, ACO, rough sets (or even their ensembles) are surveyed in the paper. It is concluded that feature selection is quite useful in the dimensionality reduction especially of big datasets as reported in the literature. It is also to note that no single feature selection algorithm can be universally effective for all datasets i.e. No free lunch [81]. Looking to its essence and importance, feature selection can play a major role in Big Data Analysis that has been identified as the future challenge and scope for the academicians, industrialists and researchers.

References

[1] B. Swartz, "The advantages of digital over analog recording techniques," Electroencephalography and Clinical Neurophysiology, Vol. 106 (2), pp. 113–117, 1998.

[2] R Agarwal, and J. Gotman, "Computer-Assisted Sleep Staging," IEEE Transactions on Biomedical Engineering, Vol. 48, No. 12, pp. 1412-1423, 2001.

[3] G. Dy, C. Brodley, A. Kak, L. Broderick, and A. Aisen, "Unsupervised feature selection applied to content-based retrieval of lung images," IEEE Trans. Pattern Anal. Mach. Intell., vol 25, pp. 373–378, 2003.

[4] J. Tang, and H. Liu, "Unsupervised feature selection for linked social media data," in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 904–912, 2012.

[5] J. Han, and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Second Edition, 2006.

[6] S. Wang, W. Pedrycz, Q. Zhu , and W Zhu, "Unsupervised feature selection via maximum projection and minimum redundancy," Knowledge-Based Systems vol. 75, pp. 19–29, 2015.

[7] P. Zhu, W. Zuo , L. Zhang, Q. Hu, and S. Shiu, "Unsupervised feature selection by regularized self-representation," Pattern Recognition vol. 48, pp. 438–446, 2015.

[8] S. Wang, W. Pedrycz, Q. Zhu, and W. Zhu, "Subspace learning for unsupervised feature selection via matrix factorization," Pattern Recognition vol. 48, pp. 10–19, 2015.

[9] J. Wang, J. Huang, Y. Sun, and X. Gao, "Feature selection and multi-kernel learning for adaptive graph regularized nonnegative matrix factorization," Expert Systems with Applications, vol. 42, pp. 1278–1286, 2015.

[10] S. Kashef, and H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection," Neurocomputing Vol. 147, pp. 271–279, 2015.

[11] H. Liu, and H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Springer, 1998.

[12] I. Guyon, and A. Elisseeff (Eds.), Feature Extraction: Foundations and Applications, Springer, vol. 207, 2006.

[13] A. Saxena, M. Kothari, and N. Pandey, Evolutionary Approach to Dimensionality Reduction, Encyclopedia of Data Warehousing and Mining, Edition 2 , Vol. II, pp. 810-816, 2009.

[14] I. Guyon, and A. Elisseeff, "An introduction to variable and feature selection," J. Mach.Learn. Res. 3 pp. 1157–1182, 2003.

[15] J.G. Dy, and C.E. Brodley, "Feature selection for unsupervised learning," J. Mach.Learn. Res. 5 pp. 845–889, 2004.

[16] H. Liu, and J. Ye, On similarity preserving feature selection," IEEE Trans. Knowledge Data Engg. Vol.25, pp. 619–632, 2013.

[17] R. Kohavi, and G. John, "Wrappers for feature selection," Artificial Intelligence, vol. 97(1-2) , pp. 273–324, 1997.

[18] V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzos, J.M. Benitez, and F.Herrera, "A review of microarray datasets and applied feature selection methods," Information Sciences, Vol. 282 pp.111-135, 2014.

[19] L. Yu, and H. Liu, "Feature Selection for high-dimensional data: A fast Correlation-based filter solution," in: Machine Learning-International Workshop then Conference-, vol. 20 pp. 856, 2003.

[20] M. Hall, Correlation-Based Feature Selection for Machine Learning. Phd Thesis, Citesse, 1999.

[21] M. Wanderley, V. Gardeux, R. Natowicz, A. Barga, and Ga-kde-bayes, "An evolutionary wrapper method based on non-parametric density estimation applied to bioinformatics problems," in, 21st European Symposium on Artificial Neural Networks-ESANN, pp.155-160, 2013.

[22] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Lear., Vol. 46(1-3), pp. 389-422, 2002.

[23] Y. Chien, and K. Fc, "Selection and Ordering of Feature Observations in a Pattern Recognition System," Information and Control 12, pp. 395-414, 1968.

[24] D. Brown, and M. O'MaUey, "The role of Eigen values in linear feature selection theory," Journal of Computational and Applied Mathematics, volume 3, No.4 , pp. 243–249, 1977.

[25] M. Ben-Bassat , "f-Entropies, Probability of Error, and Feature Selection," Information and Control, vol. 39, pp. 227-242, 1978.

[26] E. Artioli, G. Avanzolini, and G. Gnudi, "Extraction of discriminant features in post-cardiosurgical intensive care units," International Journal of Bio-Medical Computing, Vol. 39, pp. 349-358, 1995.

[27] P. Pudil, J. Novovicova, N. Choakjarernwanit, and J. Kittler, "Feature Selection based on the Approximation of class densities by finite mixtures of special type," Pattern Recognition, Vol. 28, No. 9 pp. 1389-1398, 1995.

[28] S. Haykin, Neural Networks, Second Edition, Pearson Prentice Hall, 2005.

[29] B. Lerner, H. Guterman, I. Dinstein and Y. Romem, "Medical Axis Transform-Based Features and Neural Network for Human Chromosome Classification," Pattern Recognition, Vol, 28, No. 11, pp. 1673-1683, 1995.

[30] A. Kohn, L. Nakano, and M. SILVA, "A Class Discriminability measure based on feature space partitioning," Pattern Recognition, Vol. 29, No. 5, pp. 873-887, 1996.

[31] M. Pregenzer, G. Pfurtscheller, and D. Flotzinger, "Automated feature selection with a distinction sensitive learning vector quantizer ," Neurocomputing Vol. 2, pp. 19-29, 1996.

[32] G. Groot, "Fast Fourier transformation filter program as part of a system for processing chromatographic data with an Apple II microcomputer, " Trends Anal. Chem., Vol 4, pp. 134- 137, 1985.

[33] W. Wu, B. Walczak, W. Penninckx, and D.L. Massart, "Feature reduction by Fourier transform in pattern recognition of NIR data," Analytica Chimica Acta., Vol. 331, pp. 75-83, 1996.

[34] L. Elden, Matrix Methods in Data Mining and Pattern Recognition, The Society for Industrial and Applied Mathematics, Philadelphia, 2007.

[35] S. Chernbumroong, S. Cang, and H. Yu, "Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition," Expert Systems with Applications., Vol. 42, pp.573–583, 2015.

[36] M. Dorigo and L. M. Gambardella, Ant Colony System: A cooperative learning approach to the traveling salesman problem, IEEE Transactions on evolutionary computation, Vol. 1(1), pp. 53-66, 1997.

[37] J. Yao, Q. Mao, S. Goodison, V. Mai, and Y. Sun, "Feature selection for unsupervised learning through local learning," Pattern Recognition Letters , Vol.53, pp. 100–107, 2015.

[38] J. Lee, and D. Kim, "Mutual Information-based multi-label feature selection using interaction information," Expert Systems with Applications., Vol. 42, pp. 2013–2025, 2015.

[39] Y. Wang, Y. Liu, L. Feng, and X. Zhu, "Novel feature selection method based on harmony search for email classification," Knowledge-Based Systems, vol. 73, pp. 311–323, 2015.

[40] J. Lee, and D.W. Kim, "Memetic feature selection algorithm for multi-label classification," Information Sciences., Vol. 293, pp. 80–96, 2015.

[41] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," Pattern Recognit., Vol. 40, , pp. 3509–3521, 2007.

[42] Y.Qian, Q. Wang, H. Cheng, J. Liang, and C. Dang, "Fuzzy-rough feature selection accelerator," Fuzzy Sets and Systems, vol. 258, pp. 61–78, 2015.

[43] P. Duhamel, and M. Vetterli, "Fast Fourier Transforms: A Tutorial Review and a State of the Art," Signal Processing, Vol. 19 , pp. 259-299, 1990.

[44] S. Gowid, R. Dixon, and S. Ghani, "A novel robust automated FFT-based segmentation and features selection algorithm for acoustic emission condition based monitoring systems," Applied Acoustics, vol. 88, pp. 66–74, 2015.

[45] G. Zhao, Y. Wu, F. Chen, J. Zhang, and J. Bai, "Effective feature selection using feature vector graph for classification," Neurocomputing, vol. 151 pp.376–389, 2015.

[46] R. Storn, and K. Price, “Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces,” Journal of Global Optimization, vol. 11, pp. 341–359, 1997.

[47] F. Glover, M. Laguna, Tabu Search, Kluwer Academic Publishers, 1997.

[48] X. Gu, Y. Li, and J. Jia, “Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm,” Electrical Power and Energy Systems, vol. 64, pp. 664–670, 2015.

[49] S. He, H. Chen, Z. Zhu, D. Ward, H. Cooper, M. Viantd, J. Heath, and X. Yao, “Robust twin boosting for feature selection from high-dimensional Omics data with label noise,” Information Sciences, Vol. 291, pp. 1–18, 2015.

[50] S. Maldonado, R. Montoya, R. Weber , “Advanced conjoint analysis using feature selection via support vector machines,” European Journal of Operational Research., Vol. 241, pp. 564–574, 2015.

[51] A. Eesa, Z. Orman, A. Mohsin, and A. Brifcani, “A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems,” Expert Systems with Applications, Vol. 42, pp.2670–2679, 2015.

[52] I. Jolliffe, Principal Component Analysis, second edition Springer-Verlag, 2002.

[53] T. Latkowski, and S. Osowski, “Data mining for feature selection in gene expression autism data,” Expert Systems with Applications, Vol. 42, pp. 864–872, 2015.

[54] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, “GSA: A Gravitational Search Algorithm,” Information Sciences, Vol. 179, PP. 2232–2248, 2009.

[55] X. Han, X. Chang, L. Quan, X. Xiong, J. Li, Z. Zhang, and Y. Liu, “Feature subset selection by gravitational search algorithm optimization,” Information Sciences, Vol. 281, pp. 128–146, 2014.

[56] X. Peng, and D. Xu, “A local information-based feature-selection algorithm for data regression,” Pattern Recognition, Vol. 46, pp. 2519–2530, 2013.

[57] N. Christianini, and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, 2000.

[58] S. Vieira, L. Mendonc, G. Farinha, and J. Sousa, “Modified binary PSO for feature selection using SVM applied to mortality prediction of septic patients,” Applied Soft Computing Vol. 13, pp.3494–3504, 2013.

[59] S. Vieira, J. Sousa, and U. Kaymak, “Fuzzy criteria for feature selection,” Fuzzy Sets and Systems, Vol. 189, pp. 1 –18, 2012.

[60] Z. Su, and P. Wang, “Minimizing neighborhood evidential decision error for feature evaluation and selection based on evidence theory,” Expert Systems with Applications, Vol. 39, pp. 527–540, 2012.

[61] L. Gong, J. Zeng, and S. Zhang, “Text stream clustering algorithm based on adaptive feature selection,” Expert Systems with Applications, Vol. 38, pp. 1393–1399, 2011.

[62] P. Bermejo, J. Gámez, and J. Puerta, “ A GRASP algorithm for fast hybrid (filter-wrapper) feature subset selection in high-dimensional datasets,” Pattern Recognition Letters, Vol. 32, pp.701–711, 2011.

[63] A. Aussem, and S. Morais, “A conservative feature subset selection algorithm with missing data,” Neurocomputing, Vol. 73, pp. 585–590, 2010.

[64] M. Mitchell, an Introduction to Genetic Algorithms, MIT Press, 1999.

[65] M. ElAlami, “A filter model for feature subset selection based on genetic algorithm," Knowledge-Based Systems, Vol. 22, pp. 356–362, 2009.

[66] M. Aghdam, N. Ghasem-Aghaee, and M. Basiri , “Text feature selection using ant colony optimization,” Expert Systems with Applications, Vol. 36, pp. 6843–6853, 2009.

[67] I. Oduntan, M. Toulouse, R. Baumgartner, C. Bowman, R. Somorjai, and T. Crainic, “A multilevel tabu search algorithm for the feature selection problem in biomedical data,” Computers and Mathematics with Applications, Vol. 55, pp. 1019–1033, 2008.

[68] W. Chena, S. Tseng, and T. Hong, “An efficient bit-based feature selection method,” Expert Systems with Applications, Vol. 34, pp.2858–2869, 2008.

[69] R. Sivagaminathan, and S. Ramakrishnan, “A hybrid approach for feature subset selection using neural networks and ant colony optimization,” Expert Systems with Applications, Vol. 33, pp. 49–60, 2007.

[70] R. Abdel-Aal, “GMDH-based feature ranking and selection for improved classification of medical data,” Journal of Biomedical Informatics, Vol. 38, pp.456–468, 2005.

[71] F. Zhu, S. Guan, “Feature selection for modular GA-based classification,” Applied Soft Computing, Vol. 4, pp. 381–393, 2004.

[72] Z. Pawlak,” Rough sets”, International Journal of Parallel Programming, Vol. 11 (5), pp. 341–356, 1982.

[73] R. Swiniarski, A. Skowron, “Rough set methods in feature selection and recognition,” Pattern Recognition Letters, Vol. 24, pp. 833–849, 2003.

[74] A. Verikas, M. Bacauskiene, “Feature selection with neural networks,” Pattern Recognition Letters , Vol. 23, pp. 1323–1335, 2002.

[75] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” Machine Learning, pp. 131–163, 1997.

[76] J. Quinlan, Induction of Decision Trees. Mach. Learn., Vol. 1, pp. 81-106, 1986.

[77] I. Inza, P. Larrañaga, R. Etxeberria, and B. Sierra, “Feature Subset Selection by Bayesian network-based optimization,” Artificial Intelligence , Vol. 123, pp. 157–184, pp. 2000.

[78] A. Saxena, N. R. Pal, and M. Vora, “Evolutionary Methods for Unsupervised Feature Selection Using Sammon’s Stress Function,” Springer, Fuzzy Inf. Eng., pp. 229-247, 2010 .

[79] H. David, and W Macready, “No Free Lunch Theorems for Optimization,” IEEE Transactions On Evolutionary Computation, Vol. 1, No. 1, pp. 67-82, 1997.