

A Framework for Categorize Feature Selection Algorithms for Classification and Clustering

Ali Asghar NADRI ^{1,*}, Farhad RAD ^{2,*}, Hamid PARVIN ²

¹ Department of Computer Engineering, College of Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran.

² Department of Computer Engineering, College of Engineering, Yasouj Branch, Islamic Azad University, Yasouj, Iran.

Abstract

At first, one of the dimension reduction techniques so called feature selection is explained. The Concepts, principles and existing feature selection methods for classification and clustering are also described. Then, a categorizing framework consisting of the procedures of finding selected subsets, including Search-based procedures and non-search based, evaluation criteria and data mining tasks will be completed and developed. During the grouping of Feature selection algorithms, categorizing framework represent guidelines to select appropriate algorithm(s) for each application. In categorizing, similar algorithms which follow the same process of selected subset finding and have the same evaluation criteria, are placed in the one block. Empty blocks indicates that no algorithm has been designed for them and this is a motive to design new algorithm.

Keywords: Feature selection, generator function, evaluation criteria, classification, clustering, categorizing framework.

1. INTRODUCTION

With ever-increasing advances in data processing technology, the cumulative volume of datasets and the number of features, which often swell waste of data, makes it hard to supply the needed resources, including storage and processing of data. Therewith, the explosive growth of data has result in increased noise. Curse of dimensionality is the most important consequent of increasing data dimensions. In addition to, despite the large number of features, learning models are prone to overfitting and performance degradation [2]. Several strategies have been propounded in the literature to deal with such consequences. Some strategies proposed for data reduction are *dimension reduction*, *data reduction*, and *data compression* [28]. In this article, the authors focus on a one dimension reduction techniques for reducing the number of features, removal of redundant features and noise data, called *feature selection*.

Feature selection is a dynamic field closely connected to data mining and other data processing techniques [48], and in various fields of machine learning and data mining is one of the subsidiaries of feature extraction. It is better in contexts where readability and interpretability are issues of concern, because the discounted values of the main features are preserved in the reduced space. It allows for the features with different data models to be combined. Therefore, the have to select a limited set from among them becomes apparent. Constraints and considerations such as reduction of computations and the runtime, avoiding the curse of dimensionality, memory limits, among others oblige one to select the minimum number of features to be applied in classification of future data.

Feature selection has *advantages* such as: 1) Adjustment of curse of dimensionality issue and avoidance of overfitting and improve performance of the model, 2) Dimension reduction of feature space, reduction of the needed computations and memory requirements, 3) Improve in visualization and data understanding as well as better data comprehensibility, 4) Reduction of training time and

final model utilizing time and increase of speedup of data mining algorithm [48], 5) Removal of redundant and irrelevant features or noisy data and increase of accuracy of the final model with a focus on potentially useful features for improve the quality of data, 6) better Comprehension of data behavior and detection of hidden patterns in data as well as receive a deeper insight of the data generation processes that lead to the acquisition of knowledge; and some of *disadvantages* of feature selection: 1) Attempting to find an effective and representative subset (small and limit) of the available features as an additional step in data mining, 2) Searching the existing virtual space, which adds another dimension to the problem, namely, finding a subset of optimal relevant features, and 3) additional time in the learning phase and 4) This step would be an exploratory process (delving into various aspects of the problem to isolate the most important dimensions and the most effective).

2. STEPS OF THE FEATURE SELECTION

Feature selection process is divided into four steps: generation procedure, evaluation function, stopping criteria and validation procedure.

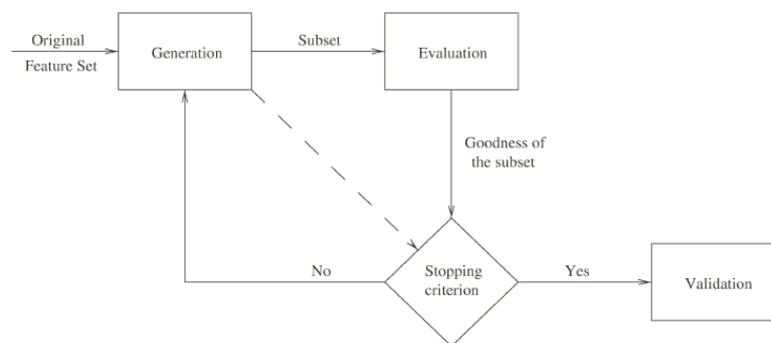


Fig. 1: Steps of feature selection

2.1. Generation Function (Selecting a Subset)

Two well-known strategies for finding subsets are search-based and non-search based strategies to be considered briefly below:

2.1.1. Search strategies

This is also referred to as subset generation in which a candidate subset is determined for evaluation in each state of the search space [14]. Two most important issues at this stage are as follows:

a) *Successor generation at every step*: This step has to do with deciding about the starting point of the search that determines its trajectory. That is to say, in every situation, forward, backward, compound, weighting and the random strategies can be used to decide on possible starting points of search as explicated below: 1) Forward strategy: in this strategy, the selected subset is first empty and later the features that are not selected yet are added to the selected subset in case they can decrease the classification error rate. The process continues in the same manner until the final subset is selected. 2) Backward strategy: unlike the forward strategy, in this strategy first all features are considered as the selected subset. Then the features that decrease the classifier error rate are removed from the selected subset. The process continues until the final subset is obtained. 3) Compound strategy: this strategy is to do with k sequential forward steps and L sequential backward steps. That is to say, forward or backward steps are performed in order to discover new interactions occurring between features. 4) Random strategy: it generates a random mode at any stage. Other

operators are limited with some criteria such as the number of features or error rate (in each stage).
5) Weighing strategy: It assigns weights to all features of the desired solution iteratively until all possibilities have been taken care of based on repetitive sampling of the existing set of samples.

b) Search organization: the Search organization supervises the feature selection process through heuristic, complete, and random strategies. The following are the main types of feature selection.

1) *Exponential search*: it is an optimized search-based strategy through which we can obtain the optimal subset.

2) heuristic search. This way allows for a feature to be selected from among all the next features repeatedly. Easy implementation is the main advantage of this way. Sequential forward search (SFS), sequential backward search (SBS) and, the bidirectional selection are examples of this method.

3) random search: This method starts with a randomly selected subset to achieve the optimized subset either through successor completely random subset generation as in Las Vegas or through a kind of sequential search that applies randomness (avoidance of local optima in the search space) in the sequential search approach [48].

4) *Genetic algorithm*: That is a flexible and powerful technique of random search for finding approximate solutions to optimization and search problems. It uses pattern matching to find the optimal solution and genetic evolution as to solve problems. The problems dealt with by this algorithm need to do with inputs turning into a solution through a process modeled on genetic evolution. Evaluation function built in this algorithm evaluates the solutions as candidates and in case the exit condition is met, the evaluation step comes to an end. The algorithm permits for the most appropriate solutions rather than the best ones to be selected.

2.1.2. Non-search Strategies

These strategies, also referred to as mathematical optimization, aim at reducing redundancy of the features and maximize the relationship between the feature and the target variable. A mathematical search utilizes the necessary and sufficient conditions that are proved to be true for the answer for the optimization. There are two types of optimization:

1) *Combinatorial Optimization method*: a branch of optimization which deals with the optimization problems that are generally hard to solve. These issues are usually solved in the best possible way by efficient examination of a space of all possible answers. Travelling Salesman Problem (TSP), Quadratic Assignment Problem (QAP), timeline and scheduling issues are instances of Combinatorial Optimization issues.

2) *Game theory method*: a way for finding the selected subset that solves the problem through different methods more efficiently. It delivers better results. This way is known as the science of interactive decision-making. In this case, the problem is first simulated in a solvable manner through the concepts of game-theory. Learning from the environment takes place while navigating the path, allowing for the best regions for navigation to be selected in the future repetitions [1]. Reinforcement learning and Monte Carlo methods are applied to navigate the environment in this way. Each subset of the all features is a state of state space and adding each feature to the set is also considered as an act that leads us towards a new state. In fact, each state is an unseen state relative to the before. At each stage, one feature that has obtained the highest scores in the previous steps and has led the issue towards a better state is selected.

2.2. Evaluation Functions

The characteristics of the data, the learning algorithm and the concept of target should be taken into account for evaluate the features of the candidate. Two categories of evaluation criteria

are applied in methods of feature selection: (i) *independent criteria*: They are independently of the algorithm used. The main types of these criteria are:

1) *distance measure*, also known as separability, provides a measure of divergence, and discrimination. For a two-class problem, a feature X is preferred to another feature Y if X induces a greater difference between the two-class conditional probabilities than Y ; if the difference is zero, then X and Y are indistinguishable. An example is the Euclidean distance measure [14].

2) *Dependence measures or correlation measures*: These qualify the ability to predict the value of one variable from the value of another. The coefficient is a classical dependence measure and can be used to find the correlation between a feature and a class [14]. This measure, which is strongly related to class, is preferably used for classification problems.

3) *Information gain or Uncertainty*: it determines the information gain of a specific feature, and is the most widely used measure for the definition of nonlinear dependencies among features.

4) *consistency-based measures*: these measures find out the minimally sized subset that satisfies the acceptable inconsistency rate, which is usually set by the user. The opposite is a situation where the two samples have the same feature values but belong to classes. Some criteria of the filter model used in feature selection for clustering are feature dependency, entropy-based distance (features with less entropy are more distinctive) and Laplace points [2].

(ii) *Dependent criteria*: In these measure, the candidate subset of features is evaluated by the mining algorithm. An example of dependent measure is classifier error rate-based, which is typically applied in wrapper algorithms where the algorithm evaluates the subsets by itself. Ordinarily attempts are made to reduce the difference between the target and the predicted value by the model towards zero. The difference between the two values will called the error rate. Thus, attempts are made to minimize this rate.

2.3. Stopping Criterion (Termination)

Without a suitable *stopping criterion* the feature selection process may run exhaustively or forever through the space of subsets [14]. Condition or criterion of stopping determines the end of the feature selection process that can be in accordance with the generative or evaluation function, such that: 1) the search is complete; 2) we have reached a certain bound (e.g. minimum number of specified features or maximum number of repetitions); 3) addition or removal of features doesn't lead to the generation of better subsets; and 4) an optimal subset based on the evaluation function is achieved.

2.4. Evaluation of Results (Validation)

Finally, to stop the selection process, stopping criteria must be determined. Feature selection process stops at validation procedure [37]. That is not the part of feature selection process. Rather, feature selection method must be validated by carrying out various tests and comparisons with previously established results or comparison with the results of competing methods using artificial datasets, real world datasets, or both [37]. The validity of the selected dataset would be check by various tests, real datasets or test sets. The classification process is accompanied with a testing phase but in clustering there is no need to test the model because the data label is determined at the same time they are being classified, and the model should learn from inside the data. It should be noted that the feature set used in the predicting and testing phase, is the feature set used in the training phase.

3. Methods of Feature Selection

Filter Method. As an ad hoc preprocessing feature selection method, it takes into account the general characteristics of the data such as information-gain or statistical dependencies regardless of the evaluation criteria to select the subset of features. An Independent measure is used to evaluate the subset of features. This method is computationally fast and is dependent of the used classifier. In

some cases, it ignores useless features, while these features may be useful in combination with other features. Feature selection for this type of data is done in two ways as follows [75]:

Ranking Methods

Most filter methods view the issue of feature selection as a ranking issue. Univariate and bivariate filter methods are ranking methods most often used for micro-array data analysis:

1) Univariate methods: They are subdivided into two groups, namely, *parametric* and *non-parametric* methods in accordance to their objectives. It is also noteworthy that in some cases due to the intrinsic importance of the features and regardless of their likely dependence to one another the univariate filter methods are used.

- *Parametric methods*: they allow for the data to be drawn from a given probability distribution, on these some more or less explicit assumption these method are based.
- *Non-parametric methods*: they allow for data to be drawn on the bases of some unknown distribution. To quantify the difference in expression between classes based on some estimate scoring function is used.

2) Bivariate methods: in accordance with their discrimination power between two or more conditions, ranking pairs of genes can be performed either using a “greedy strategy” or “all pair strategy.”

- *Greedy strategies*: they first rank all genes by individual ranking, using one of the criteria provide by univariate ranking methods; subsequently, the highest scoring gene is paired with the gene g_j that gives the highest gene pair score. After selecting first pair, g_s that is next highest ranked gene paired with the gene g_r , it maximizes the pair score, and so on.
- *All-pairs strategy*: unlike the greedy methods, all pair's strategies compute the pair scores for all pairs. As such, they examine all possible gene pairs.

Filter Methods-Space Searching Approach

That is an optimization strategy that will provide the most informative and least redundant subset of features among the whole set. This strategy follows three main steps described below:

- *Multivariate methods*. In this method, the features are evaluated batches and in comparison to other features, that's why this way is able to identify and control redundant features.

Wrapper Method. The model uses the prediction accuracy of the algorithm applied to determine the quality of the selected features and acts as a black box; that means that it has no parameter exchange with the outside world, and represents the simplicity of feature selection procedure in this approach. The Algorithm serves as evaluation criterion for features subset, too. The main disadvantage of this method is its high complexity that result in uncontrollability of problem solving. In other words, the mining algorithm controls the selection of features subset. Feature selection is done with immediate intervention of the classifier and based on its evaluation and the features subsets are introduced in accordance with their usefulness for the predictor. The search for finding a good feature subset is done by the algorithm itself as a part of the evaluation function. This method results in performance than the filter methods, but training sets with low volume usually tend to result in overfitting.

Wrapper methods for feature selection are either *Deterministic* or *randomness*. In the deterministic method (examples: SFS-SBE-Beam Search), the search starts from the existing feature space, in forward or backward manners. Randomness (examples SA-GA- Random Hill Climbing): Compared to the deterministic method, the next features subset is randomly searched [38].

Hybrid Method. This method is used when you want to have an additional evaluation of the features, or attend the samples behavior more effectively. Although, the calculations of this method are more than the calculations in the filter and wrapper method, this method delivers more accurate results for classification and goodness cluster for clustering. This method is generally used for large datasets and more complex goals. Hybrid algorithms are proposed to take advantage of the above-mentioned models and to avoid pre-description of a stopping criterion, in large datasets. The accuracy of this method is usually comparable to the accuracy of the wrapper method and its

performance is comparable to that of filter method. The quality of the clustering algorithm results in this method will determined the stopping condition.

Embedded Method. A characteristic of this method is the adjustment of the feature selection process. Feature selection is carried out implicitly within the classifier context and is considered as part of the training procedure. While classifying, the algorithm decides which features to be used or ignored. Thus, it is difficult to control the selection of appropriate number features, often result in redundancy of features. The predictive model is generated and trained along with the feature selection. This approach would be more effective if the user has an initial insight of the issue and its intended use.

4. A Framework for Categorize Feature Selection Algorithms

Feature selection allows domain experts to interpret a decision model by reducing the number of variables to analyze [58]. The goal of data refining and data mining can also be different needs including acquisition of useful knowledge, legal requirements [48], discovery of problem, more efficient decision-making, adopting appropriate policies and so on. Instead of selecting a public method for feature selection and acceptance of its implications for the final subset, different feature selection methods can be combined using hybrid approaches. Although the use of hybrid methods needs additional computational resources, their application is feasible. Generally, for each application, the most appropriate algorithm based on domain knowledge, data characteristics, or design the algorithm based on that is selected. Therefore, the user must not only have knowledge domain (which is usually not a problem of users) but also must know the technical details of existing algorithms [48].

A three-dimensional categorization including the strategies for finding the selected subset (through search and non-search procedures), evaluation criteria and data mining tasks (classification and clustering) are introduced and many well known and widely used algorithms will be included in the categorizing. The advantage of this approach is to help the user for use the required algorithms. However, selection of an appropriate choice from among the existing algorithms is a challenging problem. Only a limited number of algorithms are always used from among these algorithms.

		Search-based strategies							
		Complete		Heuristic				Genetic algorithm	
				Sequential		Random			
Evaluation criteria	filter	distance	B&B [53], BFF [78], BOBRO [4], OBLIVIN [41], Ichis- ska84a, Icho-skla84b[14]	Relief [34], Relief [36], RelifS [45], SFS[60], Segen's[66], SBS[29], BDA[77], sege84, Quei-Gels84, koll-saha96, Moor [14], LRS[38]	kmedoids, k-means[2]	MRmr[20]	CHCGA[13,24], EDAs[38]		
		information	MDLM [67], CARDIE [7], DFL[70]	DTM [7], Koller' [35], FG [46], FCBF [79], BSE [8], QMI-P[12]	Dash's [17], SBUD [16], EWKM,TF, IDF, TSC-AFS, HPStream[2]				
		dependency	Bobrowski's [67]	CFS [27], RRESET[50], POE+ACC [52], DVMM [70], MODTREE[39]	Mitra's[38], PreSET[30], DBSCAN, OPTICS, BIRCH, Chameleon, AGNES, DIANA, MCFS[2], MICAP[82], MCFS, LFSBSS[2]				
		consistency	FOCUS [3], ABB [47], MIFESI [56], Schlimmer's [63], SCH193[14]	Set Cover[15], WINNOW[42], SPEC, LS[2]	SPEC, LS[2]	LIV, QBB [46], LVF[43]			
	wrapper	predictive accuracy or cluster goodness	BS [21], AMB&B[25], FSLC[30], FSBC[31], CARDIE[7], OBLIVIN[42]	SBS-SLASH [8], WSFG[48], WSBG [48], BDS, PQSS [21], RC [22], SS [51], Queiros', BSE [8], K2-AS [68], RACE [51], SBS-W [19], Schemata Search[40],Queiros and Gelsema[44], SVM-RFE[80], 21SVM [5], Grafting Algorithm, OSFS[73]	AICC [32], FSSEM [23], ELSA [33], CLARANS, Diana, Agnes, BIRCH, ROCK,CAMELEON, DBSACN, OPTICS, DenClue, STING, WaveCluster, CLIQUE, EM, SOM, COBWEB, pCluster, COD, constrained clustering, pam, clara[28], FSSC-2, W-k-means-4, TF, TF-IDF, rX2, FTC, TSC- AFS[2]	SA, RGSS[21], LVW [44], RMHC-PF[69], GA[76], RVE[71]			
hybrid	(filter+wrapper)		BBHFS[18], Xing's [49], HGA[55], SU-GA-W[72], RCM[81], SMOTE[54]	Dash-Liu's, LUFs[2]					
embedded	(filter+wrapper)		ID3[62], C4/5[61], BlogReg, SBMLR[9], DEFS, ELM[59], WINNOW[42]				SVEGA[65]		
		classification	clustering	classification	clustering	classification	clustering	classification	clustering
Data mining tasks									
Ls and spec are for supervised and non-supervised learning.									

Fig. 2.1: Part 1 of Main Categorizing Framework

Non-search-based strategies	
Mathematical optimization methods	
Combinatorial optimization	Game theory-based (reinforcement learning & Monte Carlo)
GA [76], HTCO[57], RGOA, 2AOM, IMCS, BMG, A&F, B&M[11]	RLID[1], P-game tree, MCTS, UCT, flat UCB, BAST, SP-MCTS, TDMC(λ), α -AMAF[6], sv&relief [26], Batch RL[43], FSSH[70]
	UCT, flat UCB, BAST, SP-MCTS, α -AMAF are in category 'monte carlo tree algorithm' TDMC(λ) is a Reinforcement Learning Based Method.
Both approaches are semi-supervised.	

Fig. 2.2: Part 2 of Main Categorizing Framework

4.1. Three-Dimensional Categorizing

This categorizing is an extended framework with three dimensions including subset finding procedure, evaluation criteria and data mining tasks. Methods such as genetic algorithms (a search based method) are also included in the group. Optimization techniques (non-search methods) such as mathematical methods (*combinatorial optimization methods* and *game based theory*) are also included in this categorization as methods for finding features subset. The above mentioned categorizing is developed for In order to better understand the inner instrument of each algorithm

and the commonalities and differences among them [48]. Accordingly, it can be figured out that two important factors in the design of algorithms include: subset finding strategies and evaluation criteria, which are the two dimensions of this framework. Search strategies include: complete Search, heuristic search, random search and genetic algorithm. On the basis the evaluation criteria the algorithms are divided into four categories: filter, wrapper, hybrid and embedded algorithms. The third dimension includes data mining Tasks (classification and clustering tasks) that is based on the availability of the information class. Many block algorithms can be distinguished according to these three dimensions. There are also a number of empty blocks which are mainly related to feature selection for clustering [48]. Categorizing framework plays several important roles: (1) reveals the relationships between different algorithms. The algorithms of each block are very similar (feature subset generation strategies and similar evaluation criteria and for a specific type of data mining tasks). (2) According to this framework we can focus on a relatively limited number of algorithms, (regardless of their bodies) to select the appropriate algorithm for a method (classification or clustering). (3) Empty blocks suggest that no fruitful research has been conducted in those given cases and no algorithm is designed for them and at the same time determine the framework of the subsequent algorithms for that state. (4) By grouping of the existing algorithms, users can compare them according to this framework. (5) Attempt to design an integrated system for intelligent feature selection using categorizing framework and existing algorithms.

Thus, an intelligent feature selection system that can automatically suggest the most appropriate algorithm(s) for a particular application is needed. Therefore, we will be forced to use integrated systems and hybrid algorithms for solving new problems in the future data mining systems. In this regard, we will develop an integrated approach for intelligent feature selection of with their details.

5. Summary and Conclusion

In this study, a structure was defined in order to design an intelligent feature selection system which provides a suitable feature selection algorithm(s) for each application and in accordance with the effective factors in problem, or if needed, can design a suitable feature selection algorithm for the intended application. In this regard, a three-dimensional categorization consisting of strategies for finding selected subsets, evaluation criteria and data mining tasks was developed. Each categorizing block contains algorithms with similar characteristics for data mining tasks such as classification and clustering.

Acknowledgments

The authors are very grateful to the anonymous reviewers and editor. Their many helpful and constructive comments and suggestions helped us significantly improve this work.

References

- [1]. Hazrati Fard, S. M., Hamzeh, A., Hashemi, S. (2012). A Reinforcement Learning Based Method for Feature Selection on the Imbalanced Datasets. *Iranian Journal of Medical Informatics (2012) Vol 2, Issue 1, pp: 5-10.*
- [2]. Alelyani, S., Tang, J., Liu, H. (2013). Feature Selection for Clustering: A Review. *Data Clustering: Algorithms and Applications, CRC Press.*
- [3]. Almuallim, H., Dietterich, T. G. (1991). Learning with many irrelevant features. *In Proc. of the Ninth National Conference on Artificial Intelligence*, pp. 547-552.
- [4]. Bobrowski, L. (1988). Feature Selection Based on Some Homogeneity Coefficient. *In Proc. of 9th International Conference on Pattern Recognition*, pp. 544-546. Article (CrossRef Link).

- [5]. Bradley, P. S., Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. *ICML 1998: Machine Learning Proceedings of the Fifteenth International Conference, San Francisco, California*. PP: 82–90, Morgan Kaufmann.
- [6]. Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P.I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., Colton, S. (March 2012). A survey of Monte Carlo tree search methods. *Computational Intelligence and AI in Games, IEEE Transactions on*, 4(1): 1 - 43.
- [7]. Cardi, C. (1993). Using Decision Trees to Improve Case-Based Learning. in *Proc. of the 10th International Conference on Machine Learning, Amherst, MA, Morgan Kaufmann*, pp. 25-32.
- [8]. Caruana, R., Freitag, D. (1994). Greedy Attribute Selection. In *Proc. of the 11th International Conference on Machine Learning, New Brunswick, NJ, Morgan Kaufmann*, pp. 28-36.
- [9]. Cawley, G. C., Talbot, N. L. C., Girolami, M. (2006). Sparse multinomial logistic regression via bayesian L1 regularisation. In *NIPS*.
- [10]. Chandrashekar, G., Sahin, F. (2014). A survey on feature selection methods. Elsevier Science, ISSN: 0045-7906, *Computers and Electrical Engineering* 40 (2014) pp: 16–28.
- [11]. Chou, C. A. (2011). Mathematical optimization methods for clustering and classification with biological and medical applications. <http://hdl.rutgers.edu/1782.1/rucore10001600001.ETD.000063358> , A dissertation submitted to the Graduate School/New Brunswick Rutgers, The State University of New Jersey in partial fulfillment of the requirements for the degree of Doctor of Philosophy.
- [12]. Chow, T. W. S., Huang, D. (JANUARY 2005). Estimating Optimal Feature Subsets Using Efficient Estimation of High-Dimensional Mutual. *IEEE Transactions on Neural Networks*, vol. 16, NO. 1, PP: 213-224.
- [13]. Cordon, O., Damas, S., Santamaria, J. (2006). Feature-based image registration by means of the chc evolutionary algorithm. *Image Vis Comput* 2006; 24, PP: 525–33.
- [14]. Dash, M., Liu, H. (1997). Feature selection for classification. *Intelligent data analysis*, 1(1-4):131–156.
- [15]. Dash, M. (1997). Feature selection via set cover. In *Proc. of IEEE Knowledge and Data Engineering Exchange Workshop*, pp. 165-171.
- [16]. Dash, M., Liu, H., Yao, J. (1997). Dimensionality reduction of unsupervised data. In *Proc. of the Ninth IEEE International Conference on Tools with AI (ICTAI'97)*, pp. 532-539.
- [17]. Dash, M., Choi, K., Scheuermann, P., Liu, H. (2002). Feature selection for clustering – a filter solution. In *Proc. of the Second International Conference on Data Mining*, pp. 115-122.
- [18]. Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. In *Proc. of the Eighteenth International Conference on Machine Learning*, pp: 74-81.
- [19]. Devijver, P., Kittler, A. J. (1982). Pattern Recognition – A Statistical Approach. *Prentice Hall, London, GB*.
- [20]. Ding, C., Peng, H. (2003). Minimum Redundancy Feature Selection from Microarray Gene Expression Data. In: *Proc. Second IEEE Computational Systems Bioinformatics Conf.*, pp. 523-528.
- [21]. Doak, J. (1992). An Evaluation of Feature Selection Methods and Their Application to Computer Security. *Technical report, Univ. Of California at Davis, Dept. Computer Science*.
- [22]. Domingos, P. (1997). Why does bagging work? A bayesian account and its implications. In *Proc. of the Third International Conference on Knowledge Discovery and Data Mining*, pp: 155-158.
- [23]. Dy, J.G., Brodley, C. E. (2000). Feature subset selection and order identification for unsupervised learning. In *Proc. 17th International Conference on Machine Learning*, pp: 247-254. Morgan Kaufmann, 2000.
- [24]. Eshelman, L. (1991). The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. In: *Rawlins GJE, editor. In foundations of genetic algorithms. Morgan Kauffman; 1991*.

- [25]. Foroutan, I., Sklansky, J. (1987). Feature selection for automatic classification of non-gaussian data. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-17, no. 2, pp.187-198. Article (CrossRef Link).
- [26].Gore, S., Govindaraju, V. (2013). Feature Selection using Cooperative Game Theory and Relief Algorithm. *Proceedings of KICSS'2013*, pp. 114-125.
- [27]. Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. *In Proc. of the 17th International Conference on Machine Learning*, pp: 359-366.
- [28]. Han, J., Micheline, K. (February 2011). *Data Mining: Concepts and Technique* 2nded. *The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, March 2006*. ISBN 1-55860-901-6 (0975 – 8887), Volume 15, No. 7.
- [29]. Hastie, T., Tibshirani, R., Friedman, J. (2001). *The Elements of Statistical Learning*. Springer, 2001. Article (CrossRef Link).
- [30]. Ichino, M., Sklansky, J. (1984). Feature selection for linear classifier. *In Proc. of the Seventh International Confernce on Pattern Recognition*, pp: 124-127.
- [31]. Ichino, M., Sklansky, J. (1984). Optimum feature selection by zero-one programming. *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-14, no. 5, pp. 737-746. Article (CrossRef Link).
- [32]. Jain, A., Zongker, D. (1997) .Feature Selection: Evaluation, Application, and Small Sample Performance. *Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153-158. Article (CrossRef Link).
- [33]. Kim, Y., Street, W., Menczer, F. (2000). Feature selection in unsupervised learning via evolutionary search. *In Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. PP: 365-369. Article (CrossRef Link).
- [34]. Kira, K., Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. *In Proc. of the Tenth National Conference on Artificial Intelligence*, pp. 129-134.
- [35].Koller, D., Sahami, M. (1996). Toward optimal feature selection. *In Proc. of the Thirteenth International Conference on Machine Learning*, pp. 284-292.
- [36]. Kononenko, I. (1994). Estimating attributes: Analysis and extension of RELIEF. *In Proc. of the Sixth European Conference on Machine Learning*, pp: 171-182.
- [37]. Kumar, V., Minz, S. (June 2014). *Feature selection: A literature review*. Smart Computing Review, Vol. 4, No. 3, pp: 211–229.
- [38]. Ladha, L., Deepa, T. (May 2011). Feature selection methods and algorithms. *International Journal on Computer Science and Engineering (IJCSE)*, ISSN: 0975-3397, pp: 1787-1797, Vol. 3 No. 5.
- [39]. Lamirel, j. c., Cuxac, p., Hajlaoui, k., Sreevallabh Chivukula, A. (2014). A new feature selection and feature contrasting approach based on quality metric. *International Workshop on Quality Issues, Measures of Interestingness and Evaluation of Data mining HAL, Id: hal-00960127*, <https://hal.archives-ouvertes.fr/hal-00960127> .
- [40]. Langley, P. (1994). Selection of relevant features in machine learning. *In: Proceedings of the AAAI Fall Symposium on Relevance*, 127–131.
- [41]. Langley, P., Sage, S. (1994). Oblivious Decision Trees and Abstract Cases. *In Working Notes of the AAAI94 Workshop on Case Based Reasoning*, pp: 113–117.
- [42]. Littlestone, N. (1988). Learning Quickly when Irrelevant Attributes Abound: A New Linear Threshold Algorithm. *Machine Learning*, vol. 2, pp. 285-318. Article (CrossRef Link).
- [43]. Liu, H., Setiono, R. (1996). A probabilistic approach to feature selection - a filter solution. *In Proc. of the Thirteenth International Conference on Machine Learning*, pp. 319-327.
- [44]. Liu, H., Setiono, R. (1996). Feature selection and classification - a probabilistic wrapper approach. *In Proc. of the Ninth International Conference on Industrial and Engineering Applications of AI and ES*, pp. 419-424.
- [45]. Liu, H., Motoda, H., Yu, L. (2002). Feature selection with selective sampling. *In Proc. of the Nineteenth International Conference on Machine Learning*, pp. 395-402.

- [46]. Liu, H., Motoda, H. (1998). Feature Selection for Knowledge Discovery and Data Mining. *Kluwer Academic Publishers, London, GB*.
- [47]. Liu, H., Motoda, H., Dash, M. (1998). A monotonic measure for optimal feature selection. In: Proceedings of the European Conference on Machine Learning, pp. 101–106 (1998) 6.
- [48]. Liu, H., Yu, L. (April 2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. In: *IEEE Transactions on Knowledge and Data Engineering*. vol. 17, no. 4, pp: 491-502.
- [49]. Loscalzo, S., Yu, L., Ding, C. (2009). Consensus group based stable feature selection. In *Proc. of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*, pp: 567-576. Article (CrossRef Link).
- [50]. Miller, A. J. (2002). Subset Selection in Regression. *Chapman Hall/CRC, 2nd edition*. Article (CrossRef Link).
- [51]. Moore, A. W., Lee, M. S. (1994). Efficient Algorithms for Minimizing Cross Validation Error. in *Proc. of the 11th International Conference on Machine Learning*, Morgan Kaufmann, New Brunswick, NJ, pp. 190-198.
- [52]. Mucciardi, A. N., Gose, E. E. (1971). A comparison of seven techniques for choosing subsets of pattern recognition properties. *IEEE Transactions on Computers*, vol. 20, pp.1023-1031. Article (CrossRef Link).
- [53]. Narendra, P. M., Fukunaga, K. (1977). A branch and bound algorithm for feature subset selection. *IEEE Transaction on Computer*, vol. 26, no. 9, pp: 917-922. Article (CrossRef Link).
- [54]. Naseriparsa, M., bidgoli, A. M., Varae, T. (May 2013). Hybrid Feature Selection Method to Improve Performance of a Group of Classification Algorithms. *International Journal of Computer Applications (0975 – 8887) Volume 69– No.17, pp: 28-35*.
- [55]. Oh, I. S., Lee, J. S., Moon, B. R. (2004). Hybrid genetic algorithms for feature selection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(11):1424–1437.
- [56]. Oliveira, A. L., Vincentelli, A. S. (1992). Constructive induction using a non-greedy strategy for feature selection. In *Proc. of the Ninth International Conference on Machine Learning*, pp. 355-360.
- [57]. Oreski, S. (2014). Hybrid Techniques of Combinatorial Optimization with Application to Retail Credit Risk Assessment, *Artificial Intelligence and Applications*, Volume 1, Number 1, pp.21-43.
- [58]. Paul, J., Dupont, P. (2014). Kernel methods for mixed feature selection. 22th European Symposium on Artificial Neural Networks (ESANN'14), 2014, ISBN 978-287419095-7, pp: 301-306. Available from <http://www.i6doc.com/fr/livre/?GCOI=28001100432440> .
- [59]. P T, B., Subashini, P. (July 2014). Optimal Feature Subset Selection Using Differential Evolution and Extreme Learning Machine. *International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064, Volume 3 Issue 7, Paper ID: 0201524*.
- [60]. Pudil, P. J., Novovicova, J. (August 2002). Novel Methods for Subset Selection with Respect to Problem Knowledge. *Intelligent Systems and their Applications*, IEEE, ISSN: 1094-7167, Vol: 13, Issue: 2, PP: 66-74.
- [61]. Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. *Morgan Kaufmann*. ISBN: 1-55860-238-0.
- [62]. Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1):81–106.
- [63]. Rui, Y., Huang, T. S., Chang, S. F. (1999). Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation* 10, 39–62. Article (CrossRef Link).
- [64]. Saeys, Y., Inza, I., Larranaga, P. (2007). A Review of Feature Selection Techniques in Bioinformatics. *Oxford journals: Bioinformatics*, vol. 23, no. 19, pp: 2507-2517.
- [65]. Sasikala, S., Balamurugan, S-A. A., Geetha, S. (2015). A Novel Feature Selection Technique for Improved Survivability Diagnosis of Breast Cancer. *Published by Elsevier B.V*,

2nd International Symposium on Big Data and Cloud Computing (ISBCC'15), Procedia Computer Science 50 (2015) 16–23, DOI: 10.1016/j.procs.2015.04.005.

- [66]. Segen, J. (1984). Feature selection and constructive inference. *In Proc. of the Seventh International Conference on Pattern Recognition*, pp. 1344-1346.
- [67]. Sheinvald, J., Dom, B., Niblack, W. (1990). A modelling approach to feature selection. *In Proc. of the Tenth International Conference on Pattern Recognition*, pp. 535-539. Article (CrossRef Link).
- [68]. Singh, M., Provan, G. M. (1996). Efficient Learning of Selective Bayesian Network Classifiers. *In Proc. of the 13th International Conference on Machine Learning, Morgan Kaufmann*, pp. 453-461.
- [69]. Skalak, D. B. (1994). Prototype and feature selection by sampling and random mutation hill climbing algorithms. *In Proc. of the Eleventh International Conference on Machine Learning*, pp. 293-301.
- [70]. Slonim, N., Bejerano, G., Fine, S., Tishbym, N. (2002). Discriminative feature selection via multiclass variable memory markov model. *In Proc. of the Nineteenth International Conference on Machine Learning*, pp. 578-585.
- [71]. Stracuzzi, D. J., Utgoff, P. E. (2002). Randomized variable elimination. *In Proc. of the Nineteenth International Conference on Machine Learning*, pp: 594-601.
- [72]. Štrumbelj, E., Kononenko, J. (2010). An Efficient Explanation of Individual Classifications using Game Theory. *Journal of Machine Learning Research 11*, pp: 1–18 (2010) 13.
- [73]. Tang, J., Alelyani, S., Liu, H. (2013). Feature Selection for Classification: A Review. *Data Classification: Algorithms and Applications, CRC Press*.
- [74]. Tan, M., Tsang, I. W., Wang, L. (2014). *Towards ultrahigh dimensional feature selection for big data*. *The Journal of Machine Learning Research 15* (1), 1371-1429.
- [75]. Tyagi, V., Mishra, A. (April 2013). A Survey on Different Feature Selection Methods for Microarray Data Analysis. *International Journal of Computer Applications (0975 – 8887)*, Volume 67– No. 16.
- [76]. Vafaie, H., Imam, I. F. (1994). Feature selection methods: genetic algorithms vs. greedy-like search. *In Proc. of the International Conference on Fuzzy and Intelligent Control Systems*.
- [77]. Wang, H., Lo, S. H., Zheng, T., Hu, I. (2012). Interaction-Based Feature Selection and Classification for High-Dimensional Biological Data. *Journal of Data Science 10*, PP: 653–672.
- [78]. Xu, L., Yan, P., Chang, T. (1988). Best first strategy for feature selection. *In Proc. of the Ninth International Conference on Pattern Recognition*, pp: 706-708.
- [79]. Yu, L., Liu, H. (2003). Feature selection for high-dimensional data: a fast correlation-based filter solution. *In Proc. of the twentieth International Conference on Machine Learning*, pp. 856-863.
- [80]. Yu, Y. (2005). SVM-RFE Algorithm for Gene Feature Selection. *Technical report, University of Delaware*.
- [81]. Zeng, Z., Zhang, H., Zhang, R., Zhang, Y. (2014). A Hybrid Feature Selection Method Based on Rough Conditional Mutual Information and Naive Bayesian Classifier. *Hindawi Publishing Corporation Volume (2014), Article ID 382738, 11 page*.
- [82]. Zhao, X., Deng, W., Shi, Y. (2013). Feature Selection with Attributes Clustering by Maximal information Coefficient. *ITQM 2013, Procedia Computer Science 17 (2013) 70 – 79, doi: 10.1016/j.procs.2013.05.011*.