INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

# RELIEF: Feature Selection Approach

**S. Francisca Rosario**
Head of the Department, Department of Computer Applications
Arignar Anna (Arts & Science) College, Krishnagiri, Tamil Nadu, India
**Dr. K. Thangadurai**
Head of the Department, PG & Research Department of Computer Science
Government Arts College (Autonomous), Karur, Tamil Nadu, India

***Abstract:***
*Feature subset selection is a technique for reducing the attribute space of a feature set. In other words, it is identifying a subset of features by removing irrelevant or redundant features. A good feature set contains a highly relevant feature which helps to improve the efficiency of the classification algorithms and to classify accurately. Relief is a feature selection algorithm for random selection of instances for feature weight calculation. The Relief algorithm adopts the random selection of instances for weight estimation. It uses the Monte Carlo Approaches for randomization selection of instances in the Relief. The efficiency and effectiveness of proposed algorithm is evaluated with Cotton Disease data sets provided by cotton research station and WEKA tool. Naïve Bayes and J48 are used as the classifiers. The classification results, in terms of classification accuracy and size of feature subspace to show the performance of the Relief algorithm.*

***Keywords:*** *Feature Selection, Relief, attribute, classification, randomization, Naïve Bayes, Weka, instances*

## 1. Introduction

Feature Selection is the process for selecting the subset of relevant features by removing redundant, irrelevant and noisy data from the original dataset. In real-world data, the representation of data often uses too many features, but only a few of them, may be related to the target concept. There may be redundancy in the original data sets. But they are not necessary to incorporate into the task of modeling. Feature subset selection is otherwise defined as the process of identifying and removing the irrelevant and redundant attributes as possible. This operation reduces the dimensionality of the data sets, which in turn to allow the learning algorithms to work faster and more effectively. The prime objective of the feature selection approach is machine learning as well as data mining with minimum feature to get maximum accuracy. Based on the nature of Feature selection approach, it has been categorized as filter approach, wrapper approach and hybrid approach. In filter approach, the features were selected based on the criteria which are independent of the particular learning algorithm applied to the data. In wrapper approach, the feature selection is based on a wrapper, which is a subset of attributes and are evaluated with a learning algorithm [BOYA11]

## 2. Literature Review

Feature selection is the process of choosing subset of features enough to describe the target concept. In earlier days, large numbers of algorithms or methods were proposed for feature selection. They are evaluated based on the concept of target value and are categorized into two as search strategy and algorithms evaluation strategy. Feature selection process will also reduce the dimensionality of the data sets. Thus the reduced feature set helps to allow the learning algorithms to operate faster and more effectively. The feature set having relevant feature results better the model in the classification. The feature set approach is said to be a good when the approach or algorithm eliminates the redundant and irrelevant feature effectively. At the same time, the feature selection technique should bring the minimum subset of features which is able to model the target most appropriately. Finding the minimum feature with more relevancies and less redundancy in the feature space would reduce system complexity and it will reduce the system processing time. This in turn saves the computation resources as well as processing time. In general, feature selection or feature reduction approaches are widely used in image processing, data mining and machine learning as well as artificial intelligence. Feature Selection plays a critical role in many domains for creating the model. There is a serious challenge with limited training samples for selecting useful features by existing feature selection algorithms.

[KIRA92] Kira and Rendell proposed the Relief Algorithm. Statistical method is used in Relief instead of Heuristic search. Relief requires linear time in the number of given features and number of training instances regardless of the target concept to be learned. It selects the statistically relevant features. Relief-F [KONO94, [KONO96] is the extension of Relief algorithm. This Relief F has enabled to work with noisy and incomplete datasets and to deal with multi-class problems. RelieveD [JOHN94] is a deterministic version of Relief. It uses all instances and all near-hits and near-misses of each instance. This results in the equivalent of running Relief for an infinite amount of time. The EUBAFES (Euclidean Based Feature Selection) [SCHE97] algorithm weights and selects features similarly to the Relief algorithm. It is also a distance-based approach that reinforces the similarities between instances that belong to the same class while deteriorating similarities between instances in different classes. A gradient descent approach is employed to optimize feature weights with respect to this goal. In real world data, the representation of data often uses too many features. But only a few features may be used to relate the target concept. In the bulk set of data, there is possibility of repetition of features. This is otherwise called as redundancy. In case of bulky or normal data sets, there will be irrelevant features sets. These irrelevant features do not have any influence for modeling. The novel feature selection algorithm should remove the redundant and irrelevant features. Feature selection is the process of identifying and removing the irrelevant and redundant features information as possible. This reduces the dimensionality of the data and may allow learning algorithms to operate faster and more effectively. This will in turn increase the accuracy level on future classification. Principle component analysis and compression (information theory) plays major role for feature selection by way of eliminating the features with less information for prediction [SAE 07]. These approaches have adopted the feature selection technique for different areas to improve the model [LIU 09]. J. Hua et.al [HUA 09] reports that comparison of some famous feature selection method in the area of bioinformatics. The methods are information Gain, Gini Index, t-test, Sequential Forward Selection (SFS).. According to him, the Feature selection in biological area is inevitable but quite challenging. Among the existing feature selection algorithms, the Relief and its variants are considered as successful one due to its simplicity and effectiveness. Relief algorithm was first proposed in [KIR 92]. After that lot of variants came to usage for feature selection. Every variant has its own merits and demerits depending on the nature of data sets. The key idea of Relief is to iteratively estimate feature weights according to their ability to discriminate between neighboring models. Relief was extended to handle noisy and missing data and solve multi class issues which the original Relief algorithm cannot deal with [ROB 03]. This Relief algorithm was named as Relief F. Subsequently, in order to explore the framework of expectation maximization, Iterative-Relief is put forward in [SUN 07]. Adaptive Relief is termed as A-Relief. This A-Relief algorithm offers an effective feature subset for the further identification [WEN 12]. Likewise, many variant of Relief algorithms are available in the feature selection domains.

**3. Preliminary Work on Feature Selection Relief Algorithm**

The Relief algorithm was proposed by Kira and Rendell in year 1992. According to Kira and Rendell [ Kira and Rendell, 1992b, Kira and Rendell, 1992a], The Relief algorithm weights each feature according to its relevance to the class. Initially, all weights are set to zero and then updated iteratively. In each iteration, this non-deterministic algorithm chooses a random instance $i$ in the dataset and estimates how well each feature value of this instance distinguishes between instances close to $i$. In this process two groups of instances are selected: some closest instances belonging to the same class and some belonging to a different class. With these instances, Relief will iteratively update the weight of each feature and it differentiates data points from different classes while, simultaneously, recognizing data points from the same class. At the end, a certain number of features with the highest weights is selected. In an alternative version, a threshold may be used in such a way that only the features with weights above this value are selected.

**4. Relief Algorithm**

The Relief algorithm was first described by Kira and Rendell [KIRA92] as a simple, fast, and effective approach to attribute weighing. The output of the Relief algorithm is a weight between −1 and 1 for each attribute, with more positive weights indicating more predictive attributes. The pseudo code for Relief is shown below. The weight of an attribute is updated iteratively as follows. A sample is selected from the data, and the nearest neighboring sample that belongs to the same class (nearest hit) and the nearest neighboring sample that belongs to the opposite class (nearest miss) are identified. A change in attribute value accompanied by a change in class leads up to weighting of the attribute based on the intuition that the attribute change could be responsible for the class change. On the other hand, a change in attribute value accompanied by no change in class leads to down weighting of the attribute based on the observation that the attribute change had no effect on the class. This procedure of updating the weight of the attribute is performed for a random set of samples in the data or for every sample in the data. The weight updates are then averaged so that the final weight is in the range [−1, 1]. The attribute weight estimated by Relief has a probabilistic interpretation. It is proportional to the difference between two conditional probabilities, namely, the probability of the attribute's value being different conditioned on the given nearest miss and nearest hit respectively [ROBN03].

```
set W[a] = 0 for each attribute a
for i = 1 to n do
        select sample si from data at random
        find nearest hit sh and nearest miss sm
        for each attribute a do
                ΔWi[a] = diff(a, si, sm) - diff(a,sj, sh)
```

***W[a] = W[a] + ΔWi[a]***
***end for***
***end for***
***for each attribute a do***
***W[a] = W[a] / n***
***end for***
***where diff(a, si, sj) = 0, if si[a] = sj[a]***
***= 1, if si[a] ≠ sj[a]***

**Pseudo code of Relief algorithms**

## 5. Experimental Methodology

The classification experiment is carried out in pest and weather data set in cotton to find the accuracy level of the Relief algorithm. The pest and weather data is taken from crop pest decision support system at Coimbatore cotton research station. Naive Bayes Classifier is used to evaluate the classification accuracy. Pest is the class attribute. The classification of the pests directly depends on 13 attributes such as crop, location, observation, week, pest value, maximum temperature, minimum temperature, relative humidity1, relative humidity2, rain fall, wind speed, sunshine hours and evaporation.

| Crop | Location | Pest | Observation | Week | Pest value | Max temp | Min temp | RH1 | RH2 | Rain fall | Wind speed | Sunshine hrs | Evaporation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cotton | Akola | Miridbug | 1995 | 1 | 3 | 29.1 | 16.5 | 85.7 | 63.9 | 32 | 3.8 | 6.5 | 5.6 |
| Cotton | Akola | Miridbug | 1995 | 2 | 0 | 27.2 | 20 | 95.1 | 75.7 | 46.2 | 5.7 | 4.3 | 2.8 |
| Cotton | Akola | Miridbug | 1995 | 3 | 11 | 27.3 | 17.7 | 93 | 59.6 | 2.5 | 4.4 | 5.7 | 2.6 |
| Cotton | Akola | Miridbug | 1995 | 4 | 1 | 29.2 | 16.1 | 93.4 | 52.6 | 0 | 3.8 | 8.1 | 3.7 |
| Cotton | Coimbatore | Mealybug | 1995 | 5 | 11 | 29.4 | 15.9 | 90.4 | 51.3 | 0 | 4.4 | 8.6 | 4 |
| Cotton | Coimbatore | Mealybug | 1995 | 6 | 9 | 31.7 | 18.5 | 94.1 | 52.9 | 0 | 4.7 | 8.7 | 3.8 |
| Cotton | Coimbatore | Mealybug | 1995 | 7 | 20 | 32.1 | 18.1 | 92.3 | 50.6 | 0 | 6.3 | 8.7 | 3.7 |
| Cotton | Coimbatore | Mealybug | 1995 | 8 | 11 | 33.4 | 21.4 | 92 | 52.1 | 0 | 6.7 | 8.5 | 5.3 |
| Cotton | Coimbatore | Mealybug | 2005 | 9 | 13 | 34.6 | 21.1 | 89.3 | 50.9 | 0 | 6.9 | 8.8 | 5.7 |

*Table 1: Pest and weather data set for cotton plant*

The evaluation measures such as Precision, Recall and F Measure are taken for experiments. In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labelled as belonging to the positive class) divided by the total number of elements labelled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labelled as belonging to the class). Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labelled as belonging to the positive class but should have been). A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score

## 6. Weka Experimental Editor

To perform the Relief experiment we used Weka (Waikato Environment for Knowledge Analysis)- a powerful open-source Java based machine learning tool that can be run on any computer that has a Java run time environment installed. Weka brings together many machine learning algorithms and tools under a common frame work with an intuitive graphical user interface. Weka has two primary modes: a data exploration mode and an experiment mode. The explorer provides easy access to all of Weka's data preprocessing, learning, attribute selection and data visualization modules in an environment that encourages initial exploration of the data. The experimenter allows large scale experiments to be run with results stored in a database for later retrieval and analysis. Figure 1 shows the configuration panel of the experimenter.

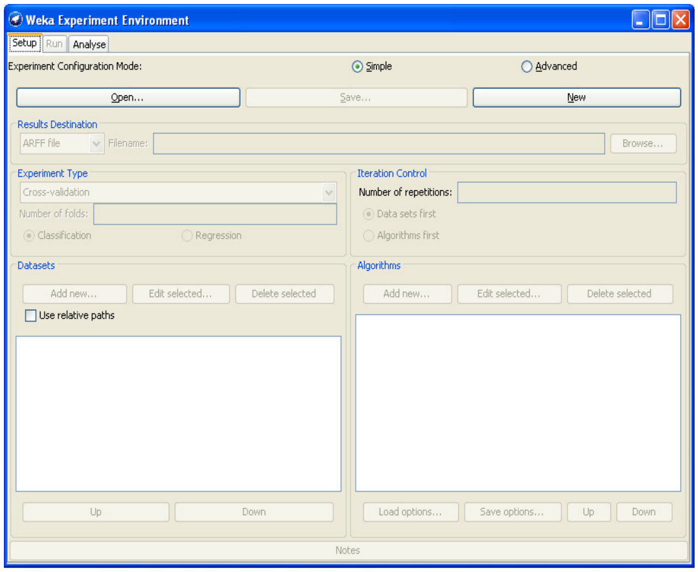


*Figure 1: Weka experimenter*

## 7. Experimental Results and Analysis

### *7.1. Classification and Attribute Selection Results*

The result shows that the class attribute pest classifies the data set in to two classes such as miridbug and mealbug. Six instances are classified as Miridbug class and three instances are classified as Mealybug. The detailed accuracy by class is shown in the weka screen. The true positive rate, false positive rate, precision, recall and fmeasure are computed and the results are produced through weka explorer. The classification result is shown by the confusion matrix. Precision is defined as the fraction of retrieved instances that are relevant and Recall is defined as the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. High precision denotes that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results. In measuring the accuracy of classification, a perfect precision score of 1.0 means that every result retrieved by a search was relevant whereas a perfect recall score of 1.0 means that all relevant documents were retrieved by the search. The weka screen represents the classification of pests before ranking and selecting the subset of attributes.. The average precision value for both the classes miridbug and mealybug is 0.444 and the average recall for both the classes is computed as 0.667.

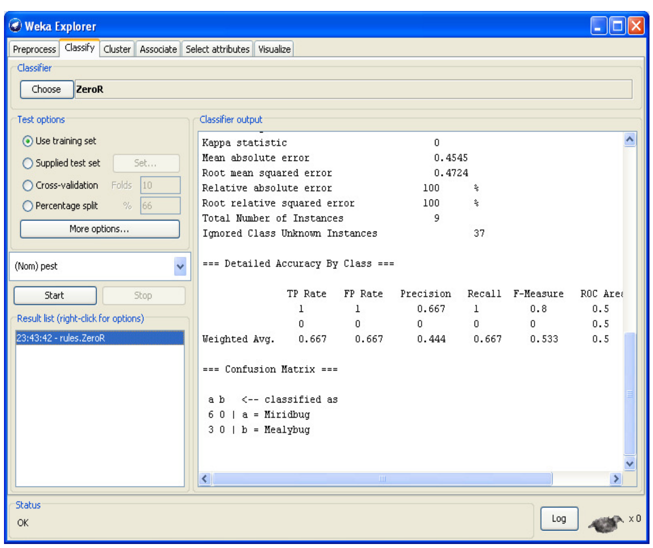


*Figure 2: Weka explorer screen which represents the classification of pests before ranking the attributes*

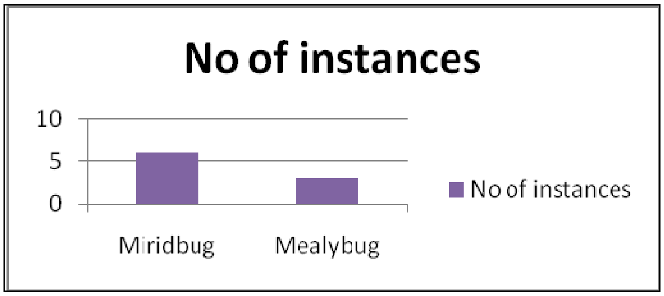


*Figure 3: Chart represents the classification of pests*

*7.2. Feature Selection Result*

The attributes are ranked using the search method. Training data in two classes such as Mealbug and Miridbug are evaluated. Relief is taken as the attribute evaluator for attribute ranking. The ranks of the selected attributes is provided by the weka tool.

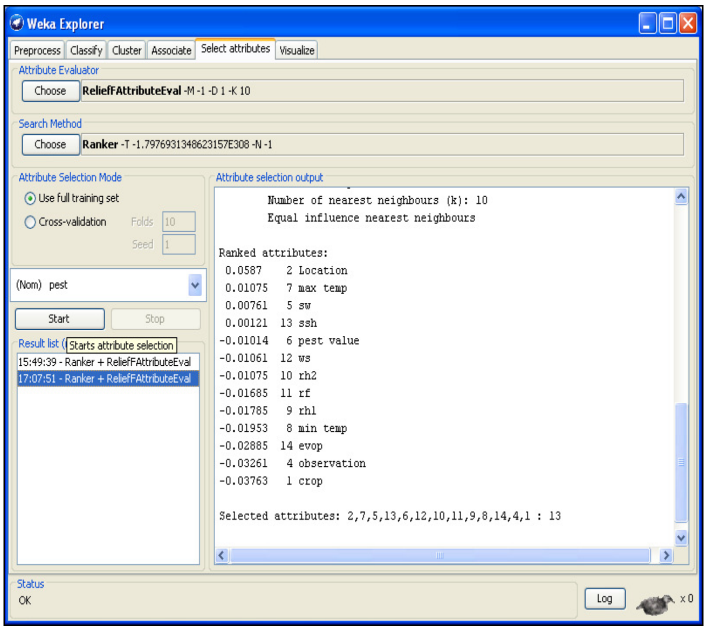


*Figure 4: Weka explorer screen which represents the ranking of attributes*

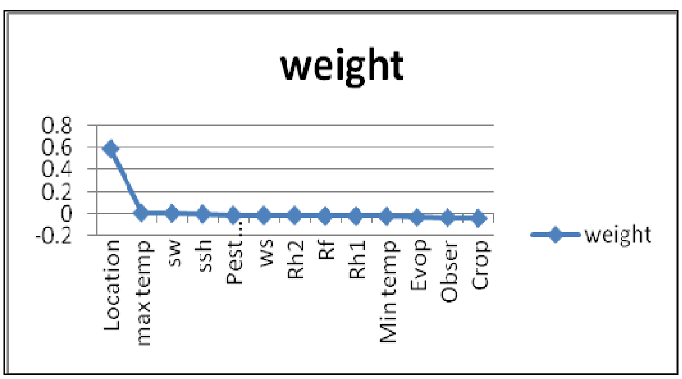


*Figure 5: Graph represents the ranking of attributes*

The experiment uses weka tool to simulate and program Relief algorithm and the data from cotton research station is imported to calculate feature weights. In the results shown by the weka screen the data weight value of attributes location, maximum temperature, pest value, sunshine hours, wind speed, minimum temperature, rain fall, RH1, RH2 and standard week are all above the threshold value -0.025, bigger than the weight value of observation, evaporation and crop whose value are below the threshold value.Therefore the three features namely evaporation, year and crop are rejected from the feature subset, with the rest of features to represent the classification feature subset. The accuracy of the classification after ranking the attributes and selecting the attributes with the highest weight is computed in terms of the accuracy measures precision and recall. The average precision value for the newly classified attributes is 0.917 and average recall value of the attributes is 0.889. In analyzing the accuracy measures for classification before attribute reduction, and after attribute reduction through ranking, the precision value of the new classification is 0.917 which is greater than 0.444, the classification result before attribute selection. Similarly the recall measure after ranking and reducing the subset in classification technique is 0.889 which is greater than the previously computed value 0.667. The experimental results are shown through the weka explorer screen.

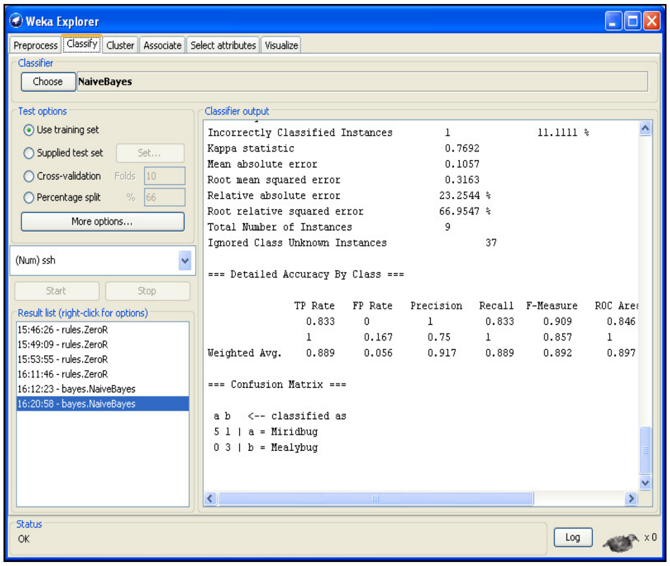


*Figure 6: Weka explorer screen which shows the classification of pests after attribute selection*

The graph shows the classification accuracy in terms of precision and recall before and after the feature selection.

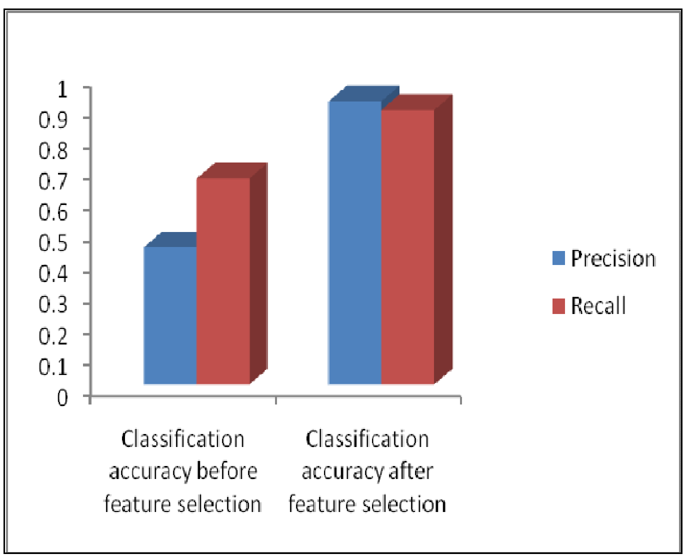


*Figure 7: Graph which represents the classification accuracy*

## 7. Conclusion

From the experiments, the algorithm studied with pest and weather data set obtained from cotton research station. The Relief algorithm is found to be out performed in the feature selection and classification accuracy level. So the experiments results reveal that Relief is considered to be the best for feature subset evaluation.

## 8. References


i. [GRE09] Greene CS, Penrod NM, Kiralis J, Moore JH: Spatially uniform ReliefF (SURF) for computationally-efficient filtering of gene-gene interactions. BioData ining 2009, 2:5

ii. [KON94] Kononenko I: Estimating attributes: analysis and extensions of RELIEF. In Proceedings of the European conference on machine learning on Machine Learning; Catania, Italy. Catania: Springer-Verlag New York, Inc; 1994:171-182

iii. [BAKE87] Baker, James E. (1987). "Reducing Bias and Inefficiency in the Selection Algorithm". Proceedings of the Second International Conference on Genetic Algorithms and their Application(Hillsdale, New Jersey: L. Erlbaum Associates): 14–21.

iv. [ROBN03] Robnik-Šikonja M, Kononenko I**:** Theoretical and empirical analysis of ReliefF and RReliefF.Machine Learning 2003, **53:**23-69.

v. [BOYA11] Boyang Li,Qiangwei Wang,Jinglu Hua, Feature Subset Selection: A Correlation -Based SVM Filter Approach, IEEJ Transactions On Electrical And Electronic Engineering, Volume 6, 2011,pp. 173-179.

vi. LIXI 03] Li-Xin Zhang, Jia-Xin Wang,Yan-Nan Zhao,Ze-Hong Yang," A Novel Hybrid Feature Selection Algorithm: Using ReliefFEstimation for GA –Wrapper search", Proceedings of the second International Conference on Machine Learning and Cybermetics,Xian,2-5 November 2003

vii. [DOAK92] Doak, J. (1992). An evaluation of feature selection methods and their application to computer security. Technical report, University of California, department of Computer Science, Davis, CA.

viii. [KONO94] Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European Conference on Machine Learning (ECML'94), pages 171–182.

ix. [KONO96] Kononenko, I., Robnik-Sikonia, M., and Pompe, U. (1996). ReliefF for estimation and discretization of attributes in classification., pages 31–40. Artificial Intelligence: Methodology, Systems, Applications. IOS Press.

x. [JOHN94] John, G., Kohavi, R., and Pfleger, K. (1994). Irrelevant features and the subset selection problem. In Proceedings of the Eleventh International Conference on Machine Learning (ICML'94), pages 121–129.

xi. G. Carneiro and N. Vasconcelos, ―Minimum Bayes Error Features for Visual Recognition by Sequential Feature Selection and Extraction,‖ Proc. Computer and Robot Vision Conf. (CRV '05), pp. 253-260, 2005

xii. M.A. Hall and G. Holmes, ―Benchmarking Attribute Selection Techniques for Discrete Class Data Mining,‖ IEEE Trans.Knowledge and Data Eng., vol. 15, no. 6, pp. 1437-1447, Nov./Dec 2003.

xiii. E. Choi and C. Lee, ―Feature Extraction Based on the Bhattacharyya Distance,‖ Pattern Recognition, vol. 36, no. 8, pp.1703-1709, 2003.

xiv. M. Dash and H. Liu, ―Feature Selection for Classification, Intelligent Data Analysis, vol. 1, nos. 1-4, pp. 131-156, 1997.

xv. K. Fukunaga and D.M. Hummels, ―Bayes Error Estimation Using Parzen and k-NN Procedures,‖ IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-9, no. 5, pp. 634-643, Sept. 1987.

xvi. R. Gilad-Bachrach, A. Navot, and N. Tishby, ―Margin Based Feature Selection—Theory and Algorithms,‖ Proc. 21st Int'l Conf.Machine Learning (ICML '04), 2004.

xvii. I. Guyon and A. Elisseeff, ―An Introduction to Variable and Feature Selection,‖ J. Machine Learning Research, vol. 3, pp. 1157- 1182, 2003.

xviii. Y. J. Li, D. F. Hsu, S. M. Chung, "Combination of multiple feature selection methods for text categorization by using combinatorial fusion analysis and rank-score characteristic," International jouranl on artificial intelligence tools, [Online].Available: DOI: 10.1142/S0218213013500012

xix. N. Garcia-Pedrajas, A. de Haro-Garcia, J. Perez-Rodriguez, "A scalable approach to simultaneous evolutionary nstance and feature selection," Information sciences, Elsevier, vol.228, no.1, pp. 150-174, 2013.

xx. W. J. Hu, K. S. Choi, Y. G. Gu, S. T. Wang, "Minimum maximum local structure information for feature selection," pattern Recognition Letters, Elsevier, vol.34, no.50, pp. 527-535, 2013.

xxi. I. Santos, X. Ugarte-Pedrero, F. Brezo, P. G. Bringas, J. M.Gomez-Hidalgo, "Noa: An information retrieval based

xxii. S. Baccianella, A. Esuli, F. Sebastiani, "Using micro documents for feature selection: The case of ordinal text classification," Expert Systems with Applications, Elsevier, vol.40, no.11, pp. 4687-4696, 2013.

xxiii. S. Shilaskar, A. Ghatol, "Feature selection for medical diagnosis : Evaluation for cardiovascular diseases," Expert Systems with Applications, Elsevier, vol.40, no.10, pp. 4146-4153, 2013.

xxiv. A. K. Farahat, A. Ghodsi, M. S. Kamel, "Efficient greedy feature selection for unsupervised learning," Knowledge and information systems, Springer, vol.35, no.2, pp. 285-310, 2013.

xxv. A. Kaban, "Non-parametric detection of meaningless distances in high dimensional data," Statistics and Computing, Springer, vol.22, no.2, pp. 375-385, 2012.

xxvi. H. X. Guo, X. W. Liao, K. J. Zhu, C. Ding, Y. H. Gao, "Optimizing reservoir features in oil exploration management based on fusion of soft computing," Applied soft computing, Elsevier, vol.11, no.1, pp. 1144-1155, 2011.

xxvii. J. Yan, N. Liu, S. C. Yan, Q. Yang, W. G. Fan, W.Wei,etc, "Trace-Oriented Feature Analysis for Large-Scale Text Data Dimension Reduction," Knowledge and data engineering, IEEE Transactions on, vol.23, no.7, pp. 1103-1117, 2010.

xxviii. A. Ozcift, A. Gulten, "Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases," Digital signal processing, Elsevier, vol.23no.1 pp. 230-237, 2013.

xxix. X. Y. Liu, H. Q. Wang, D. M. Xu, "The application of adaptive partitioned random search in feature selection problem,"in Advanced Data Mining and Applications 2005, 2005 1st International Conference on vol.3584, 2005,pp. 307-314.

xxx. L. Xu, J. C. Sun, Y. Tian, "Feature selection using rough entropy-based uncertainty measures in incomplete decision systems," Knowledge-based systems, Elsevier, vol.36, no.1, pp. 206-216, 2012.