iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57433

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

An Extensive Analysis of Feature Selection in Machine Learning

Mrs.D.Malarvizhi[1], Dr.A.Prakash[2]
[1]*Assistant Professor, Department of Computer Science Dr. N.G.P. Arts and Science College Coimbatore*
[2]*Professor, Department of Computer Science, Hindusthan College of Arts & Science Coimbatore*

Abstract: Researchers and engineers working in the disciplines of data mining and machine learning have difficulties while analysing high-dimensional data. A dimension reduction method called feature selection is used to pick characteristics that are pertinent to machine learning tasks. Improving the efficiency of machine learning algorithms, hastening the learning process, and creating basic models all depend critically on reducing the size of the dataset by removing superfluous and useless information. Numerous feature selection techniques have been put forth in the literature to find the pertinent feature or feature subsets in order to accomplish the goals of clustering and classification. The purpose of this study is to review the state of the art for these methods.

I. INTRODUCTION

With the rapid development of modern technology, tremendous new computer and internet applications have generated large amounts of data at an unprecedented speed, such as video, photo, text, voice, and data obtained from social relations and the rise of the Internet of things and cloud computing. These data often have the characteristics of high dimensions, which poses a high challenge for data analysis and decision-making. Feature selection has been proven in both theory and practice effective in processing high-dimensional data and in enhancing learning efficiency [1–3].

The amount of high-dimensional data that exists and is publically available on the internet has greatly increased in the past few years. Therefore, machine learning methods have difficulty in dealing with the large number of input features, which is posing an interesting challenge for researchers. In order to use machine learning methods effectively, pre-processing of the data is essential. Feature selection is one of the most frequent and important techniques in data pre-processing, and has become an indispensable component of the machine learning process [4].

Feature selection is referred to the process of obtaining a subset from an original feature set according to certain feature selection criterion, which selects the relevant features of the dataset. It plays a role in compressing the data processing scale, where the redundant and irrelevant features are removed.

Feature selection technique can pre-process learning algorithms, and good feature selection results can improve learning accuracy, reduce learning time, and simplify learning results [5–7].

In the process of feature selection, irrelevant and redundant features or noise in the data may be hinder in many situations, because they are not relevant and important with respect to the class concept such as microarray data analysis [8]. When the number of samples is much less than the features, then machine learning gets particularly difficult, because the search space will be sparsely populated. Therefore, the model will not able to differentiate accurately between noise and relevant data [9]. There are two major approaches to feature selection. The first is Individual Evaluation, and the second is Subset Evaluation. Ranking of the features is known as Individual Evaluation [10].

Feature selection, which has been a research topic in methodology and practice for decades, is used in many fields, such as image recognition [11–15], image retrieval [16–18], text mining [19–21], intrusion detection [22–24], bioinformatic data analysis [25–32], fault diagnosis [33–35], and so on.

According to the theoretical principle, feature selection methods can be based on statistics [36–40], information theory [41–46], manifold [47–49], and rough set [50–54]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

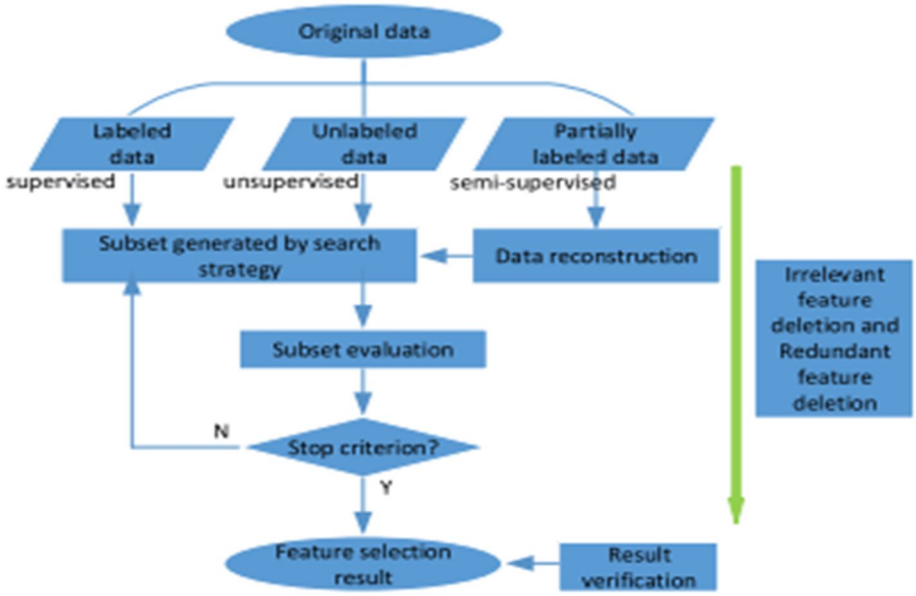


Fig 1. A Framework for feature selection

II. FEATURE SELECTION

Feature selection is the process of selecting relevant features, or a candidate subset of features. The evaluation criteria are used for getting an optimal feature subset. In high-dimensional data (number of samples << number of features), finding the optimal feature subset is a difficult task [55].

Let the original set of features and be an evaluation criterion to be maximized (optimized) and defined as L: A'⊆ A->R. The candidate subset of features can be considered under the following considerations [56]:

1) Let IAI = m & |A'|=n, then,L(A') is ,maximized, where m>n and A'⊂ A.
2) Set a threshold θ such that L(A') > θ; to find a subset of the feature with the smallest number (m>n).
3) Finding the optimization function L(A') with optimal feature subsets |A'|.

Table 1. General algorithm for feature selection

INPUTS:
X: Set of features of a data set having features
SG : Successor Generator Operator
E : Evaluation measure (dependent or independent
θ : Stopping Criteria OUTPUT: : Optimal feature set or weighted feature
OUTPUT:

$X_{opt.}$: Optimal feature set or weighted features
Initialize:
X' := Start_ point(X);
$X_{opt.}$ ={Best of X' using E };
Repeat:
X':=Search_Strategy (X',SG(E),X);
$X_{opt.}$ ={Best of X' according to E };
If E(X') >= E(X_{opt}) or E(X') == E(X_{opt}) & |X'| < | X_{opt}|)
Then X_{opt} = X;
Until Stop criteria is not found;

©IJRASET: All Rights are Reserved

III. TYPES OF FEATURE SELECTION

To achieve superior results in machine learning, a well-prepared and high-quality input dataset is required. In order to train our model and facilitate its improved learning, we gather a vast amount of data. The dataset typically includes some usable data mixed together with irrelevant and noisy data. Furthermore, the massive volume of data slows down the model's training process, and the model may not forecast or perform effectively if it contains noise and irrelevant data. Therefore, it is imperative to exclude these noises and less significant data from the dataset. To do this, feature selection techniques are applied.

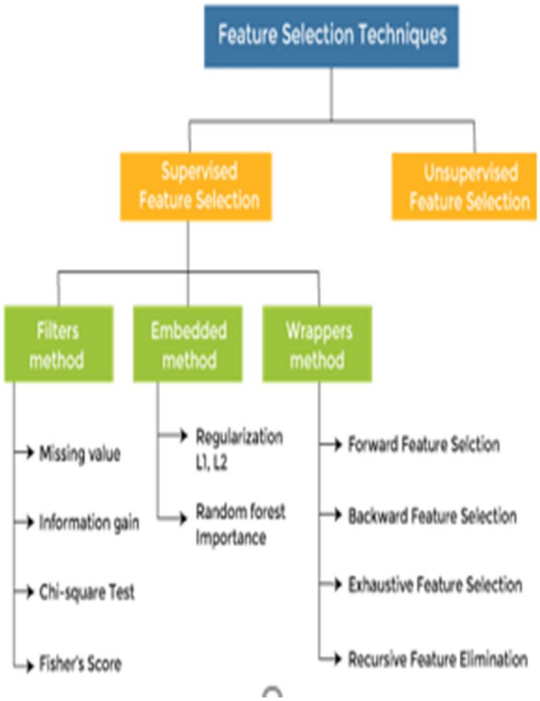


Fig 2. Types of feature selection

The goal of feature selection techniques in machine learning is to find the best set of features that allows one to build optimized models of studied phenomena.

The techniques for feature selection in machine learning can be broadly classified into the following categories:

1) *Supervised Techniques:* These techniques can be used for labeled data and to identify the relevant features for increasing the efficiency of supervised models like classification and regression. For Example- linear regression, decision tree, SVM, etc.
2) *Unsupervised Techniques:* These techniques can be used for unlabeled data. For Example- K-Means Clustering, Principal Component Analysis, Hierarchical Clustering, etc.

A. *Filter Methods*

Filter methods pick up the intrinsic properties of the features measured via univariate statistics instead of cross-validation performance. These methods are faster and less computationally expensive than wrapper methods. When dealing with high-dimensional data, it is computationally cheaper to use filter methods.

1) *Information Gain:* Information gain determines the reduction in entropy while transforming the dataset. It can be used as a feature selection technique by calculating the information gain of each variable with respect to the target variable.
2) *Chi-square Test:* Chi-square test is a technique to determine the relationship between the categorical variables. The chi-square value is calculated between each feature and the target variable, and the desired number of features with the best chi-square value is selected.
3) *Fisher's Score:* Fisher's score is one of the popular supervised technique of features selection. It returns the rank of the variable on the fisher's criteria in descending order. Then we can select the variables with a large fisher's score.
4) *Missing Value Ratio:* The value of the missing value ratio can be used for evaluating the feature set against the threshold value. The formula for obtaining the missing value ratio is the number of missing values in each column divided by the total number of observations. The variable is having more than the threshold value can be dropped.

B. Wrapper Methods

1) Wrappers require some method to search the space of all possible subsets of features, assessing their quality by learning and evaluating a classifier with that feature subset. The feature selection process is based on a specific machine learning algorithm we are trying to fit on a given dataset. It follows a greedy search approach by evaluating all the possible combinations of features against the evaluation criterion. The wrapper methods usually result in better predictive accuracy than filter methods.

2) Forward Feature Selection: This is an iterative method wherein we start with the performing features against the target features. Next, we select another variable that gives the best performance in combination with the first selected variable. This process continues until the present criterion is achieved.

3) Backward Feature Elimination: This method works exactly opposite to the Forward Feature Selection method. Here, we start with all the features available and build a model. Next, we the variable from the model, which gives the best evaluation measure value. This process is continued until the preset criterion is achieved.

4) Exhaustive Feature Selection: This is the most robust feature selection method covered so far. This is a brute-force evaluation of each feature subset. This means it tries every possible combination of the variables and returns the best-performing subset.

5) Recursive Feature Elimination: Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), the goal of recursive feature elimination (RFE) is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features, and each feature's importance is obtained either through a coef_ attribute or a feature_importances_attribute.Then, the least important features are pruned from the current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to select is eventually reached.

C. Embedded Methods

These methods encompass the benefits of both the wrapper and filter methods by including interactions of features but also maintaining reasonable computational costs. Embedded methods are iterative in the sense that takes care of each iteration of the model training process and carefully extract those features which contribute the most to the training for a particular iteration.

Table 1: List of some common selection algorithms

Method	Algorithms	Type/Search Strategy
Filter	mRMR ReliefF Chi-square Test	Univariate
	CFS (Correlation-Based Feature Selection) FCBF (Fast Correlation-Based Feature Selection) Markov Blanket-Based Filter Selection	Multivariate
Wrapper	SFS (Sequential Forward Selection) SBS (Sequential Backward Selection) Bidirectional Search	Deterministic Search
	Randomized Hill Climbing Genetic Search Simulated Annealing	Randomized Search
	Exhaustive Search Beam Search Branch and Bound	Exponential Search
Embedded	Decision Trees Random Forest Weighted Naïve Bayes	Tree-Based Methods
	Ridge Regression LASSO Elastic Net	Regularization Methods

IV. CONCLUSION

The literature on feature selection and feature selection stability is reviewed in the current study. High-dimensional dataset problems have spurred interest in dimension, or data, reduction techniques such as feature selection. Because of this, a wide range of feature selection strategies have been developed over time. Selecting the right method is essential to the feature selection process because these techniques employ various strategies to select pertinent features. Numerous studies have demonstrated how removing superfluous and unnecessary features enhances both the quality of the data analysis and the efficiency of machine learning algorithms. However, choosing the best feature set is not always possible, especially when features are closely related, and feature selection further complicates the learning process.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The models constructed using the feature subsets that selection algorithms choose, as well as their stability, define the quality of those algorithms. The robustness, or insensitivity, of the selection algorithm to small alterations in the training set is referred to as stability. Repetitive outcomes are produced by stable feature selection techniques. Because unstable algorithms mislead users in choosing the resulting subset of attributes and erode their trust in the algorithm and analysis process, the stability of the selection algorithm is a critical concern.

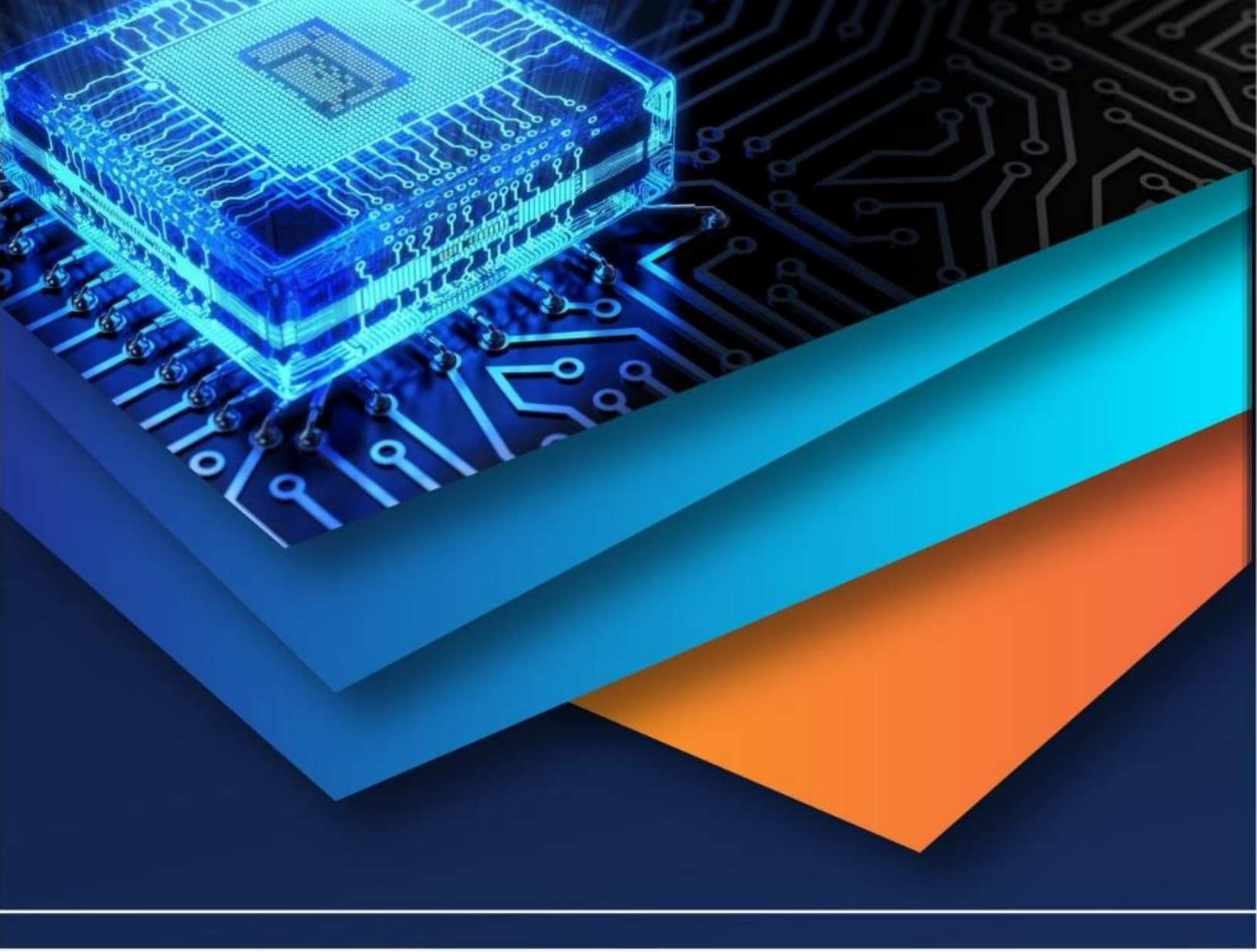
REFERENCES

[1] A.L. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artif. Intell. 97 (1997) 245–271.
[2] H. Liu, H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Springer Science & Business Media, 2012.
[3] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[4] A. Kalousis, J. Prados, M. Hilario, "Stability of Feature Selection Algorithms: a study on high dimensional spaces," Knowledge and information System, vol. 12, no. 1, pp. 95-116, 2007. Article (CrossRef Link)
[5] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, Advancing Feature Selection Research, ASU Feature Selection Repository (2010) 1–28.
[6] P. Langley, Selection of relevant features in machine learning, in: Proceedings of the AAAI Fall Symposium on Relevance, 1994, pp. 245–271.
[7] P. Langley, Elements of Machine Learning, Morgan Kaufmann, 1996.
[8] M. Dash, H. Liu, "Feature Selection for Classification," Intelligent Data Analysis, Elsevier, pp. 131-156, 1997.
[9] F. Provost, "Distributed data mining: scaling up and beyond," Advances in distributed data mining, Morgan Kaufmann, San Francisco, 2000.
[10] I. Guyon, A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003.
[11] A. Khotanzad, Y.H. Hong, Rotation invariant image recognition using features selected via a systematic method, Pattern Recognit. 23 (1990) 1089–1101.
[12] N. Vasconcelos, Feature selection by maximum marginal diversity: optimality and implications for visual recognition, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, pp. 762–769.
[13] N. Vasconcelos, M. Vasconcelos, Scalable discriminant feature selection for image retrieval and recognition, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2004.
[14] J.Y. Choi, Y.M. Ro, K.N. Plataniotis, Boosting color feature selection for color face recognition, IEEE Trans. Image Process. 20 (2011) 1425–1434.
[15] A. Goltsev, V. Gritsenko, Investigation of efficient features for image recognition by neural networks,, Neural Netw. 28 (2012) 15–23.
[16] D.L. Swets, J.J. Weng, Efficient content-based image retrieval using automatic feature selection, in: Proceedings of International Symposium on Computer Vision, 1995.
[17] D.L. Swets, J.J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 831–836.
[18] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, A simultaneous feature adaptation and feature selection method for content-based image retrieval systems, Knowl.-Based Syst. 39 (2013) 85–94.
[19] D.D. Lewis, Y. Yang, T.G. Rose, F. Li, Rcv1: a new benchmark collection for text categorization research, J. Mach. Learn. Res. 5 (2004) 361–397.
[20] L.P. Jing, H.K. Huang, H.B. Shi, Improved feature selection approach TFIDF in text mining, in: Proceedings of International Conference on Machine Learning and Cybernetics, 2002, pp. 944–946.
[21] S. Van Landeghem, T. Abeel, Y. Saeys, Y. Van de Peer, Discriminative and informative features for biomolecular text mining with ensemble feature selection, Bioinformatics 26 (2010) 554–560.
[22] G. Stein, B. Chen, A.S. Wu, K.A. Hua, Decision tree classifier for network intrusion detection with GA-based feature selection, in: Proceedings of the 43rd ACM Southeast Conference, 2005, pp. 136–141.
[23] F. Amiri, M.R. Yousefi, C. Lucas, A. Shakery, N. Yazdani, Mutual information-based feature selection for intrusion detection systems, J. Netw. Comput. Appl. 34 (2011) 1184–1199.
[24] A. Alazab, M. Hobbs, J. Abawajy, M. Alazab, Using feature selection for intrusion detection system, in: Proceedings of International Symposium on Communications and Information Technologies (ISCIT), 2012, pp. 296–301.
[25] H. Liu, J. Li, L. Wong, A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, Genome Inform. 13 (2002) 51–60.
[26] H. Liu, H. Han, J. Li, L. Wong, Using amino acid patterns to accurately predict translation initiation sites, In Silico Biol. 4 (2004) 255–269.
[27] Q. Song, J. Ni, G. Wang, A fast clustering-based feature subset selection algorithm for high-dimensional data, IEEE Trans. Knowl. Data Eng. 25 (2013) 1–14.
[28] G. Li, X. Hu, X. Shen, X. Chen, Z. Li, A novel unsupervised feature selection method for bioinformatics data sets through feature clustering, in: Proceedings of IEEE International Conference on Granular Computing, 2008, pp. 41–47.
[29] Y.F. Gao, B.Q. Li, Y.D. Cai, K.Y. Feng, Z.D. Li, Y. Jiang, Prediction of active sites of enzymes by maximum relevance minimum redundancy (mRMR) feature selection, Mol. Biosyst. 9 (2013) 61–69.
[30] D.S. Huang, C.H. Zheng, Independent component analysis-based penalized discriminant method for tumor classification using gene expression data, Bioinformatics 22 (2006) 1855–1862.
[31] C.H. Zheng, D.S. Huang, L. Zhang, X.Z. Kong, Tumor clustering using nonnegative matrix factorization with gene selection, IEEE Trans. Inf. Technol. Biomed. 13 (2009) 599–607.
[32] H.J. Yu, D.S. Huang, Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids, IEEE/ACM Trans. Comput. Biol. Bioinform. 10 (2013) 457–467.
[33] L. Wang, J. Yu, Fault feature selection based on modified binary PSO with mutation and its application in chemical process fault diagnosis, Adv. Nat. Comput. 3612 (2005) 832–840.
[34] T.W. Rauber, F. de Assis Boldt, F.M. Varejão, Heterogeneous feature models and feature selection applied to bearing fault diagnosis, IEEE Trans. Ind. Electron. 62 (2015) 637–646.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[35] K. Zhang, Y. Li, P. Scarf, A. Ball, Feature selection for high-dimensional machinery fault diagnosis data using multiple models and Radial Basis Function networks, Neurocomputing 74 (2011) 2941–2952.
[36] M. Vasconcelos, N. Vasconcelos, Natural image statistics and low-complexity feature selection, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 228–244.
[37] T. Khoshgoftaar, D. Dittman, R. Wald, A. Fazelpour, First order statistics based feature selection: a diverse and powerful family of feature seleciton techniques, in: Proceedings of 11th International Conference on Machine Learning and Applications (ICMLA), 2012, pp. 151–157.
[38] J. Gibert, E. Valveny, H. Bunke, Feature selection on node statistics based embedding of graphs, Pattern Recognit. Lett. 33 (2012) 1980–1990.
[39] M.C. Lane, B. Xue, I. Liu, M. Zhang, Gaussian based particle swarm optimisation and statistical clustering for feature selection, in: Proceedings of European Conference on Evolutionary Computation in Combinatorial Optimization, 2014, pp. 133–144.
[40] H. Li, C.J. Li, X.J. Wu, J. Sun, Statistics-based wrapper for feature selection: an implementation on financial distress identification with support vector machine, Appl. Soft Comput. 19 (2014) 57–67.
[41] L. Shen, L. Bai, Information theory for Gabor feature selection for face recognition, EURASIP J. Appl. Signal Process. (2006) 1–11.
[42] B. Morgan, Model selection and inference: a practical information – theoretic approach, Biometrics 57 (2001) 320.
[43] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238.
[44] F. Fleuret, Fast binary feature selection with conditional mutual information, J. Mach. Learn. Res. 5 (2004) 1531–1555.
[45] H.H. Yang, J.E. Moody, Data visualization and feature selection: new algorithms for nongaussian data, Adv. Neural Inf. Process. Syst. 12 (1999) 687–693.
[46] B. Bonev, Feature Selection Based on Information Theory, Universidad de Alicante, 2010.
[47] Z. Xu, I. King, M.R.T. Lyu, R. Jin, Discriminative semi-supervised feature selection via manifold regularization, IEEE Trans. Neural Netw. 21 (2010) 1033–1047.
[48] B. Jie, D. Zhang, B. Cheng, D. Shen, Manifold regularized multi-task feature selection for multi-modality classification in Alzheimer's disease, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2013, pp. 275–283.
[49] B. Li, C.H. Zheng, D.S. Huang, Locally linear discriminant embedding: an efficient method for face recognition, Pattern Recognit. 41 (2008) 3813–3821.
[50] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognit. Lett. 24 (2003) 833–849.
[51] Y. Chen, D. Miao, R. Wang, A rough set approach to feature selection based on ant colony optimization, Pattern Recognit. Lett. 31 (2010) 226–233.
[52] W. Shu, H. Shen, Incremental feature selection based on rough set in dynamic incomplete data, Pattern Recognit. 47 (2014) 3890–3906.
[53] J. Derrac, C. Cornelis, S. García, F. Herrera, Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection, Inf. Sci. 186 (2012) 73–92.
[54] J. Wang, K. Guo, S. Wang, Rough set and Tabu search based feature selection for credit scoring, Procedia Comput. Sci. 1 (2010) 2425–2432. [54] J.R. Quinlan, C4. 5: Programs for Machine Learning, Elsevier, 2014
[55] R. Kohavi, G.H. John, "Wrappers for feature subset selection," Artificial Intelligence, vol. 97, no. 1-2, pp. 273-324, 1997. Article (CrossRef Link)
[56] M. Kudo, J. Sklansky, "A Comparative Evaluation of medium and large–scale Feature Selectors for Pattern Classifiers," in Proc. of the 1st International Workshop on Statistical Techniques in Pattern Recognition, pp. 91-96, 1997.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)