A PERFORMANCE ANALYSIS OF EVOLUTIONARY BASED RANK FEATURE SELECTION AND PRINCIPLE COMPONENT ANALYSIS

[1]R. Abitha, [2]Dr. S. Mary Vennila
[1]Research Scholar, [2]Associate Professor
PG & Research Department of Computer Science
Presidency College, University of Madras, Chennai, India

Abstract

In Data Mining and Machine learning the first step of analysis undergo preprocessing of the features in a given space. Managing large number of features always increase problems for a model in terms of both accuracy and time complexity. Many research have been done on various feature selection , feature construction and feature reduction methods to reduce the number of properties to improve accuracy of the model built. In this paper a performance analysis have been done on Symmetrical Uncertainty (SU) feature selection embedded with evolutionary algorithm and Principle Component Analysis (PCA) for two datasets of Autism Spectrum Disorder from kaggle repository and data collected from special schools and tested the accuracy with specific classifier. The implementation results on selected UCI datasets with binary, continuous variables it explained that principle component analysis works best with expected accuracy. For limited data set especially with categorical variable features alike Autism spectrum disorder data filter selection with evolutionary algorithm formerly proposed by the author [14] is useful through decreasing the volume of initial features and improved accurate result by implementing better detection performance in the classification methods relating with other feature selectors.

Keywords: *Machine Learning, Feature Selection, Feature Reduction, Principle Component Analysis, Symmetrical Uncertainty, Evolutionary Algorithm.*

1. Introduction

Data analysts, researchers need superior and pertinent data from the enormous amount of collected data. In data mining there are major category for preprocessing. 1.Feature selection 2. Feature Reduction 3. Feature construction. The main issues here are “Curse of Dimensionality” refereed by Bellmen shows that as the dimensionality of the given data set increases, the amount of data essential for the analysis increase exponentially[11]. Next issue is as the dimensionality rises computational cost also rise[12]. To overcome the above said issues the feature extraction step is necessary in data mining. During this preprocessing from the properties provided in the data base only influence attributes are identified and used for the model. For this different methods are used by the analysts.

1.1 Feature Reduction:

Principle Component Analysis (PCA) is a well-known data preprocessing method to detent linear dependencies among the features of a data set. It works on the principle as that the features present in the data set are correlated with any extent. It reduces the attribute space by the smallest possible amount of information on the original data at the same time it maintains the variance among them. In this classic technique the same feature will not be used to train ,fit in a model for accuracy testing instead the original **n** features are substituted by **m** orthogonal new set of features by linear combinations that is called principle components. Here eigenvectors are playing major role in transforming the matrix in to orthogonal components. These components are orthogonal to each other and linearly uncorrelated to each other. The set of new features arranged in the decreasing order of their importance. The first component based on the cumulative proportion of its variance contributes to predict the result. Though PCA builds new set of features different from the original one it preserves as much as information of the original matrix. Following are the steps in PCA.

Step1 : Normalize the data and transform the data in a normal distribution.

Step 2: Calculate the covariance matrix of the transformed matrix.

Step3: Calculate Eigen value and Eigenvectors

Step4: Principle Components arranged with decreasing order of variance.

1.2 Algorithm for Calculating Principle Components[13]

1: Given a data matrix ($X = [x_1, x_2, \ldots, x_N]$), where N represents the total number of samples and xi represents the i^{th} sample.

2: Compute the mean of all samples as follows: $u = \frac{1}{N} \sum_{i=1}^{N} xi$

3: Subtract the mean from all samples as follows:

$$D = \{d1, d2, \ldots dn\} = \sum_{i=1}^{N} xi - u$$

4: Compute the covariance matrix as follows:

$\sum = \frac{1}{n-1} D * DT$:

5.Compute the eigenvectors V and eigenvalues λ of the covariance matrix (Σ).

6: Sort eigenvectors according to their corresponding eigenvalues.

7: Select the eigenvectors that have the largest eigenvalues W = {v1, …, vk}. The selected eigenvectors (W) represent the projection space of PCA.

8: All samples are projected on the lower dimensional space of PCA (W) as $Y = W^T D$.

1.3 Feature Selection

Following mentioned are three main categories for feature selection method.

Filter Method: This method does not require any classification machine learning algorithm. Depending on the various evaluation statistical measures features are selected.

These methods exhibit low computational complexity, scalable and more faster

Wrapper Method: In this method it needs a model for the learning phase as to choose useful sub set of attributes contribute to the accurate result. This wrapper method though results more accuracy suffered from high computational cost.

Embedded Method: Here both filter and wrapper methods can be combined or any feature selection method is combined with a classifier.

Feature selection technique supports in decreasing the dimensionality of features by eliminating noisy, redundant or irrelevant data through which improvisation in classification accuracy with the least processing of data has completed[14].Basic idea of knowledge discovery process is deriving an interesting new unknown pattern from the data base. As the real time high dimensional datasets include complicated information with errors, and in such a condition, classification method plays an important role. Feature selection is a simple method that tries to find out a subset of original features that have the same information regarding the whole datasets, without the loss of generality[18]. The main reason to use feature selection is to reduce computational cost, improved accuracy, and problem understanding.[18]

2. Related Works

The following describes the associated works done on the hybrid feature selection techniques for the different kinds of problems. Mustafa K. Masood Chaoyang Jiang Yeng Chai Soh in [1] has presented the proposed framework to work the occupancy estimation, a novel technique called Hybrid Feature-Scaled Extreme Learning Machine (HFS-ELM). The proposed method gives three and four tolerance accuracies of 413 around 90% and 95%, respectively.

Kangfeng Zheng, Xiujuan Wang in [2], firstly the joint maximal information entropy (JMIE) is determined to measure a feature subset. Next, a binary particle swarm optimization (BPSO) algorithm is proposed to explore the optimal feature subset. Finally, classification is performed on UCI corpora to check the performance of our proposed method compared to the traditional mutual information (MI) method, CHI method, as well as a binary version of particle swarm optimization-support vector machines (BPSO-SVMs) feature selection. Investigations show that FS- JMIE obtains an equal or better performance than MI, CHI, and BPSO-SVM. Further, FS-JMIE manifests moderately better robustness to the number of classes. Moreover, the method shows better time-efficiency and higher consistency than BPSO-SVM

Songyot Nakariyakul in [3], discusses a new hybrid method called the interaction information-guided incremental selection (IGIS) algorithm that operates interaction information to guide the search. In this paper, the empirical results for eleven high dimensional datasets describe that the proposed algorithm consistently exceeds prior feature selection techniques, while enquires a moderate amount of search time.

Asil Oztekin, Lina Al-Ebbini, Zulal Sevkli, Dursun Delen in [4], shows the hybrid GA-based feature selection technique along with the development and design of many highly accurate classification algorithms to classify the essential features in the feature-rich and large UNOS transplant dataset for lung transplantation.

Indu Jain Vinod Kumar Jain Renu Jain in [5], exhibits the two-phase hybrid model for cancer classification integrating Correlation-based Feature Selection (CFS) with improved-Binary Particle Swarm Optimization (iBPSO). The proposed iBPSO also checks the problem of immediate concurrence to the local optimum of traditional BPSO. The proposed design has estimated on 11 benchmark microarray datasets of different cancer types. Experimental results have compared with seven other well-known methods, and our model exhibited better findings regarding the number of selected genes and classification accuracy in most cases. In particular, it obtained up to 100% classification accuracy for seven out of eleven datasets with a minimal sized prognostic gene subset (up to 1:5%) for all eleven datasets.

Abitha R et al[14] proposed optimized feature selection method for ASD using cultural algorithm and presented expected result with high accuracy. Bai-Ning Jiang et al. [16] proposed a hybrid feature selection algorithm. In the first step Symmetric Uncertainty (SU) of the individual features is calculated and those features that have less SU than the threshold value are discarded. In the second step Genetic Algorithm based searching is employed on the left over features.[15][18]Naive Bayes classifier and SU are used to evaluate goodness of the feature subsets by 10 fold-cross validation. J. Zhou et al. [17][18] proposed an Ant Colony Optimization and Mutual Information based feature selection for equipment fault diagnosis. Regression estimation model and mean squared error are used for the feature subset evaluation. Both mutual information and classification accuracy are used for subset optimization. [18]

Vol. 12, No. 5, (2019), pp. 259 - 266

3. Hybrid Feature Selection Methodology

3.1 Symmetrical Uncertainty (SU)

SU is a filter based feature sub set selection. Symmetrical uncertainty can be used to calculate the fitness of features between each feature and the target class for feature selection [18]. Since it is independent of the model, selection by rule basically done based on the evaluation of correlation between the feature and target class using any one of the metrics information gain or mutual information or gain ratio and chi square method. In the first round the feature which has high value of SU gets high importance[18] and given higher rank. The symmetrical uncertainty (SU) [6] between target concept and features are employed to capture the best features for classification. This method is symmetric since $SU_{(A,B)}=SU_{(B,A)}$ and because of its symmetric nature number of comparisons will be reduced. The elements with higher SU values have the higher weightage which means highest rank. SU measures the relationship among A, B variables based on the information theory [7]. Which is a measure of uncertainty of a random variable.[21].It is calculated as follows

$$SU\ (A,B) = 2\ \frac{IG(A/\ B)}{(H(A) + H\ (B))}$$

$$\mathrm{IG(A/B)}\ = \mathrm{H(A)}\ -\ \mathrm{H(A/B)}$$

Where IG(A/B) is the information gain of feature A, that is an independent attribute and B is the class attribute[5]. H(A) is the entropy of feature A and H(B) is the entropy of feature B. Information gain has a desired property of symmetric.[5] H(A) is a priori probability and H(A/B) is a Posteriori probability[21]. The amount of Information given by a feature B about another feature A is effectively the same as that of the information given of feature A and the feature B. Estimating[5] the SU indicates the normalized range value [0 to1] and correction factor value is 2. If SU value is 1, then the information of one feature is predictable. If SU value is 0, then A, B are not associated. The use of symmetrical uncertainty has been proved to be useful in dimensionality reduction in previous studies [20][21]

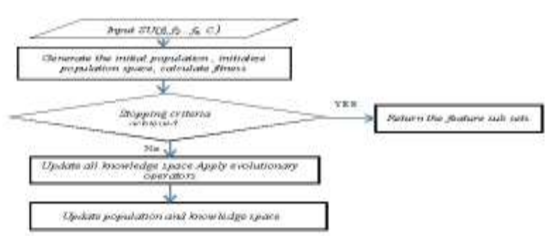
3.2 Cultural Algorithm

Cultural algorithm is an evolutionary algorithm mainly useful for finding optimized solution for a problem. Basically it works on the concept of "survival of fittest". The major Components in cultural algorithm are population space, belief space and a communication protocol in between these two spaces to exchange knowledge during evolution.

The belief space comprised of five knowledge sources namely normative, situational, domain, topographical and history knowledge source. Each knowledge sources are contributing their knowledge and memory to the evolution process and brands the evolutionary algorithm as if memory less which is basically memory less and retain the elite population across the generation and useful in finding optimal solution.

3.3 Procedure for proposed Hybrid Feature Selector Method (SU-CA)

The existing Cultural Algorithm (CA) and Symmetrical Uncertainty (SU) algorithm are combined to develop the Hybrid Feature Selector Algorithm. The framework of hybrid feature selector is as follows.



ISSN: 2233-7857 IJFGCN

Figure 1 Framework for SU-CA

4. Results and Discussion

The proposed SU-CA algorithm is used as rank based feature selection method to select sub set of the data set Autism Spectrum Disorder (ASD1) taken from Kaggle with 18 features and 1054 instances. The results of the proposed algorithm shows the data set with 10 predominant features useful in predicting the target variable with improved accuracy. The Principle Component Analysis is used as feature reduction method on the same (ASD1) data set The results are analyzed in the following sub sections. The same work was carried out with 78 features with 274 instances (ASD2) of collected data set.

4.1 Analysis of the proposed SU-CA and PCA for the data set ASD2

For ASD2 data set 78 orthogonal components have been generated in the decreasing order of their variance. The first component have 57% of proportion of variance which shows its importance in predicting the class. The collective value of first two components are 77% but shows accuracy of the model 68.9% From this study, the model will not give improved accurate result for ASD2 data set with the first two principle components

Table1 Summary of PCA

	Comp. 1	Comp.2	Comp. 3
Standard deviation	5.6985053	4.0170534	2.11630686
Proportion of Variance	0.5796357	0.2080371	0.07994485
Cumulative Proportion	0.5796357	0.8676728	0.94761761

Table2 Sensitivity, specificity

Number of cases in the data set	274
Samples predicted correctly	189
Accuracy	68.9%

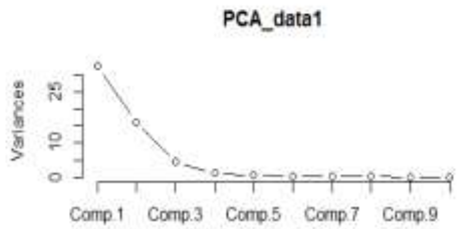


Figure 2 component importance-ASD2

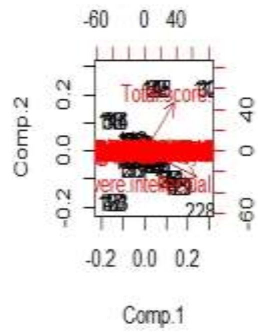


Figure3 Bar Plot-ASD2

4.2 Analysis of the proposed SU-CA and PCA for the data set ASD1

For the ASD1 data set 18 orthogonal components have been generated in the decreasing order of their variance. The first component have 24% of proportion of variance which shows its importance in predicting the class. From this study, the model will not give improved accurate result for ASD2 data set with the first single principle component. This shows though PCA has proven result in dimensionality reduction but for this two data sets one taken from Kaggle and the other collected one has categorical values its performance based on the accuracy is lesser than symmetrical uncertainty based feature selection. For the ASD1 data set also the result showed from the graphs the principle components are negatively correlated with each other.

Table3 Summary of PCA

	Comp .1	Comp. 2	Comp. 3
Standard deviation	1.9889	1.18484	1.08010
Proportion of Variance	0.2472	0.08774	0.07291
Cu tior			

Table4 Sensitivity, specificity

Number of cases in the data set	1054
Samples predicted correctly	710
Accuracy	67.3%

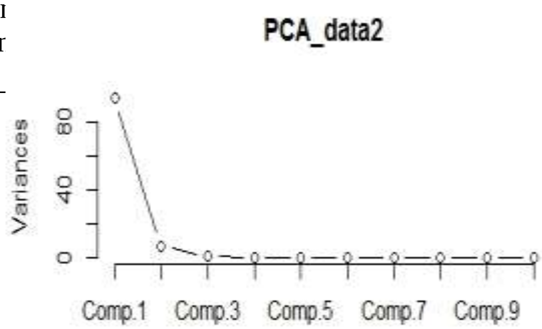


Figure4-Importance of PCA-ASD1

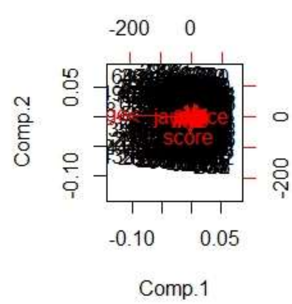


Figure5-Bar Plot for ASD1

4.3 Performances Analysis of SU-CA for ASD1 and ASD2

	No. of features obtained	Accuracy	TPR	TNR	Precision	F-Measure
SU-CA-ASD1	10	93.00	0.9301	0.8688	0.9303	0.9298
SU-CA-ASD2	25	89.51	0.8914	0.8184	0.8952	0.8932

5. Conclusion

In this work, hybrid feature selector combines SU and CA is applied on the data sets to achieve the goal of removing redundant and irrelevant features and to produce feature sub set of original features. The data reduction also done on the same data sets using Principle Component Analysis (PCA). The results are compared to choose the best method to generate feature sub sets. Based on the measures sensitivity, specificity and accuracy, precision, F-Measure the two methods are evaluated and it is proved the embedded SU with CA shows better performance than the PCA. But PCA is a highly accepted feature reduction method especially for binary and discrete variables. Since the data set used here in this work has categorical variables proposed method is more favorable than PCA. The reason may be the type of variables as categorical. The data classification also has enhanced with non-redundant and relevant features. For feature reduction techniques in nominal categorical data Multiple Correspondence Analysis (MCA) can be used which will be the future work on the same data sets.

Reference

[1] Mustafa K.Masood, Chaoyang Jiang, Yeng Chai Soh, "A Novel Feature Selection framework with Hybrid Feature-Scaled Extreme Learning Machine (HFS-ELM) for Indoor Occupancy Estimation," Energy and Buildings, pp. 1139-1151, 2018.

[2] Kangfeng Zheng, Xiujuan Wang, “Feature selection method with joint maximal information entropy between features and class," Pattern Recognition, Elsevier, pp.20-29, 2018.

[3] Nakariyakul, Songyot. "High-dimensional hybrid feature selection using interaction information-guided search." Knowledge-Based Systems (2018).

[4] Oztekin, Asil, et al. "A decision analytic approach to predicting the quality of life for lung transplant recipients: A hybrid genetic algorithms-based methodology." European Journal of Operational Research 266.2 (2018): pp. 639-651.

[5] Jain, Indu, Vinod Kumar Jain, and Renu Jain. "Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification." Applied Soft Computing 62 (2018), pp.203-215.

[6] Onan, Aytuğ, and Serdar Korukoğlu. "A feature selection model based on genetic rank aggregation for text sentiment classification." Journal of Information Science 43.1 (2017): 25-38.

[7] Wang, Yintong, et al. "An efficient semi-supervised representatives feature selection algorithm based on information theory." Pattern Recognition 61 (2017): 511-523.

[8] Liu, Jinghua, et al. "Feature selection based on quality of information." Neurocomputing 225 (2017): 11-22.

[9] Abualigah, Laith Mohammad, and Ahamad Tajudin Khader. "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering." The Journal of Supercomputing 73.11 (2017): 4773-4795.

[10] Gu, Shenkai, Ran Cheng, and Yaochu Jin. "Feature selection for high-dimensional classification using a competitive swarm optimizer." Soft Computing 22.3 (2018): 811-822

[11] R. E. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, USA, 1957.

[12] A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data Zena M. Hira and Duncan F. Gillies Department of Computing, Imperial College London,

[13] Alaa Tharwat Principle component analysis – a tutorial Int. J. Applied Pattern Recognition, Vol. 3, No. 3, 2016

[14] Abitha R ,Mary Vennila S. “A Swarm Based Symmetrical Uncertainty Feature Selection Method For Autism Spectrum Disorders.” 3rd International Conference on Inventive Systems and Control. IEEE,2019.

[15] Syed Imran Ali , Waseem Shahzad .A Feature Subset Selection Method based on Symmetric Uncertainty and Ant Colony Optimization, International Journal of Computer Applications (0975 – 8887) Volume 60– No.11, December 2012

[16] Bai-Ning Jiang Xiang-Qian Ding Lin-Tao Ma “A Hybrid Feature Selection Algorithm: Combination of Symmetrical Uncertainty and Genetic Algorithms” The Second International Symposium on Optimization and Systems Biology, Lijiang, China, pp. 152–157, 2008.

[17] J. Zhou, R. Ng, and X. Li, "Ant colony optimization and mutual information hybrid algorithms for feature subset selection in equipment fault diagnosis", in Proceeding of 10th International Conference on Control, Automation, Robotics and Vision, pp.898-903, 2008.

[18] A Feature Subset Selection Technique for High Dimensional Data Symmetric Uncertainty Bharat Singh, Nidhi Kushwaha, Om Prakash Vyas, journal of Data Analysis and Information Processing, 2014, 2, 95-105

[19] Liu, H. and Yu, L. (2005) Toward Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Transactions on Knowledge and Data Engineering, **17**, 491-502. http://dx.doi.org/10.1109/TKDE.2005.66

[20] Symmetrical Uncertainty-Based Feature Subset Generation and Ensemble Learning for Electricity Customer Classification, Minghao Piao , Yongjun Piao and Jong Yun Lee .MDPI 5 April 2019

[21] Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. 2004, 5, 1205–1224.

[22] R.Abitha, Dr. S.Mary Vennila. A Proficient Optimized Feature Selection Method Germane for Autism Spectrum Disorder Classification. International Journal of Scientific &Technology Research.(Under print).