Volume 3, No. 4, July- August 2012
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info

Enhanced Feature Selection Algorithm using Modified Fisher Criterion and Principal Feature Analysis

L. Arockiam
Associate Professor, Department of Computer Science
St.Joseph's College (Autonomous)
Tiruchirappalli, Tamil Nadu, India
larockiam@yahoo.co.in

V. Arul Kumar*
Research Scholar, Department of Computer Science
St.Joseph's College (Autonomous)
Tiruchirappalli, Tamil Nadu, India
v.arulkumar@ymail.com

Abstract: Dimensionality reduction is one of the key issues in various fields such as Data Mining, Machine Learning, Pattern Recognition, Image Retrieval, Text mining etc. This technique is used to reduce the dimensionality of features and improve the performance of learning algorithms. In general, dimensionality reduction is classified into two categories: feature selection and subspace learning. Recently, many researchers combined these two methods to improve the performance of the learning algorithm. In this paper, an enhanced feature selection algorithm is proposed namely, Modified Fisher Criterion Principal Feature Analysis (MFCPFA). The MFCPFA algorithm is developed by merging Modified Fisher Criterion (feature selection and subspace learning) and Principal Feature Analysis. To prove the effectiveness of proposed algorithm, the algorithm is tested with various datasets available in the UCI repository. The results show that the newly developed algorithm improves the classification and reduces the unwanted features.

Keywords: Feature selection, Fisher Criterion, Fisher Score, Linear Discriminant Analysis, and Principal Feature Analysis.

I. INTRODUCTION

Data mining is the process of extracting useful information from large amounts of data stores in various databases, data warehouse and other information repositories. The data collected from the real world applications contain lot of erroneous data. Data preprocessing is an important technique in data mining to rectify the erroneous data present in the dataset. Many data mining application contain high dimensional data. The High dimensionally decreases the performance of the learning algorithm and increase the time and space required for processing the data [1]. The above issue is resolved using Dimensionality Reduction technique (DR). The DR is divided into two methods, visually, feature selection and feature extraction. In the past research, the DR issue is carried out by either feature selection technique or subspace learning technique [2].

The feature selection technique is used to select the most discriminant feature from the original input feature set. The subspace learning technique is used to transform the original input features to a lower dimensional space. The popular subspace learning includes Linear Discriminant Analysis (LDA) [3, 4], Principal Component Analysis (PCA) [5, 6], Direct Linear Discriminant Analysis (DLDA) [7], Singular Value Decomposition (SVD) [8], Support Vector Machine (SVM) [9], Locality Preserving Projection (LPP) [10], Neighborhood Preserving Embedding (NPE) [11] and Null Space Linear Discriminant Analysis (NLDA) [12].

Fisher criterion [13] plays a vital role in dimensionality reduction. This criterion is used to select the features by minimizing the within-class distance and maximizing the between-class distance. Two methods are developed based on the Fisher criterion namely, Fisher score and Linear Discriminant Analysis (LDA). The Fisher score is a feature selection method and LDA is a subspace learning method. In this paper, an enhanced feature selection algorithm MFCPFA is proposed by combining two supervised methods and one unsupervised method. The two supervised methods Fisher score and LDA are developed based on the Fisher criterion. The Fisher score is used to select the subset of original features. LDA is used to reduce the feature space by maximizing the class separability. In recent days, various LDA algorithms are proposed. These algorithms concentrate on classification accuracy and computational time. This method is used the select the feature subset by reducing the feature space and it removes the noisy and irrelevant features in the original dataset. Principal Feature Analysis (PFA) is an unsupervised method which is used to remove the redundant features in the original dataset. Hence, in our proposed algorithm the supervised and unsupervised methods are combined to obtain the best feature subset by removing the irrelevant, redundant and noisy features present in the original dataset.

The paper is organized as follows: Section 2 supplies the overview of Feature selection. Fisher score and Linear Discriminant Analysis (LDA) are discussed in Section 3 and 4 respectively. Section 5 deals with Principal Feature Analysis. Section 6 describes the proposed algorithm and Section 7 presents our experimental results. Finally, Section 8 gives the conclusion of this study.

II. FEATURE SELECTION

Feature selection is the process of finding the pertinent subset of features from the original dataset using some selection criterion [14]. This technique mainly used to reduce the dimensionality of data by removing the irrelevant, redundant and noisy features. Reduction of the data dimensionality and removal of futile features have the advantages of reduced computational time and improved classification accuracy in various data mining applications. The feature selection algorithms are categorized into supervised [15], unsupervised [16] and semi-supervised [17] algorithms according to the utilization of label information. Based on the search strategy, feature selection algorithms fall into three approaches, namely filter approach, wrapper approach and embedded approach [18]. In filter approach, the features are selected using predefined criteria. The wrapper approach uses the learning algorithm to select the best feature subset. In hybrid approach, the subset of features are selected using both predefined criteria and learning algorithm. The feature selection technique is widely used in various research fields such as Data Mining [19], Machine Learning [20], Intrusion detection [21] and Text Mining [22].

III. FISHER SCORE

The main objective of the Fisher score is to select the features which have the similar values in the same class and the dissimilar values in different classes. The Fisher score [23] is calculated using the formula

$$\mathrm{FS} = \mathrm{tr}\left\{(\tilde{S}_b)(\tilde{S}_c + \gamma X)^{-1}\right\} \tag{1}$$

where

$\tilde{S}_b$ is the scatter matrix between-class

$\tilde{S}_c$ is the total scatter matrix

γX is the positive regularized parameter

tr is the trace value of the given scatter matrix

$$\tilde{S}_b = \sum_{n=1}^{m} k_n (\tilde{\mu}_n - \tilde{\mu})(\tilde{\mu}_n - \tilde{\mu})^T \tag{2}$$

$$\tilde{S}_c = \sum_{j=1}^{s} (x_j - \tilde{\mu})(x_j - \tilde{\mu})^T \tag{3}$$

where

$\tilde{\mu}$ is the total mean of the reduced data

k_n is the size of the n^{th} class reduced data

$\tilde{\mu}_n$ is the mean of the n^{th} feature

x_j is the number of samples of the j^{th} class in the reduced data

IV. LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA [24] is a well-known statistical approach for dimensionality reduction. It is used to determine the low-dimensional features from a high-dimensional space. In the past research, several dimensionality reduction techniques have been proposed. Among them LDA is regarded as the most popular dimensionality reduction method. According to Fisher's criterion, LDA has to find a linear projection matrix $W \in \mathbb{R}^{d\times m}$ by minimizing the ratio of the *within-class scatter matrix* and maximizing the ratio of the *between-class scatter matrix*, and is given below.

$$arg \max_{W} tr\left((W^T S_w W)^{-1}(W^T S_b W)\right) \tag{4}$$

Where the *between-class matrix* S_b, *within-class matrix* S_w, are calculated using the formula

$$S_b = \sum_{n=1}^{m} k_n (\mu_n - \mu)(\mu_n - \mu)^T \tag{5}$$

$$S_w = \sum_{n=1}^{m} \sum_{i=X_n} (p_i - \mu_n)(p_i - \mu_n)^T \tag{6}$$

Where

X_n is the index set of n^{th} class

μ_n is the mean vector of n^{th} class

The LDA have three scatter matrices within-class matrix, between-class matrix and total scatter matrix. The total scatter matrix calculated by the formula

$$S_t = S_w + S_b \tag{7}$$

Hence, the EQ. (4) is rewritten as

$$arg \max_{W} tr\left((W^T S_t W)^{-1}(W^T S_b W)\right) \tag{8}$$

V. PRINCIPAL FEATURE ANALYSIS (PFA)

The PFA [25] technique is derived from the Principal Component Analysis (PCA). It is an unsupervised dimensionality reduction technique.

Let *P* be a zero mean *m*-dimensional random feature vector and consider *X* be the covariance matrix. Let *D* be a matrix whose columns are the orthonormal eigenvectors of the matrix *X*.

$$X = D\Lambda D^T \tag{9}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_m \end{bmatrix}$$

Where $\lambda_1, \lambda_2, \ldots, \lambda_m$, be the eigenvalues of *X*. Let D_q be the first *q* column of *D*. Let $V_1, V_2, V_3 \ldots V_n \in R_m$ be the rows of D_m. Each vector V_i represents the projection of the i^{th} feature of the vector *P*. To find the best subset, row vector V_i used to cluster the features which are having high correlated measure. Finally, from each cluster one feature is obtained to form a feature subset. The algorithm can be summarized in the following five steps:

Step 1: Calculate the sample covariance matrix or true covariance matrix. In few cases correlation matrix is preferred to use instead of covariance matrix. The correlation matrix are computed by $N_{xy} = \frac{M[j_x j_y]}{M[j_x^2]M[j_y^2]}$

Step 2: Calculate the Principal components and eigenvalues of the Covariance/

Correlation matrix by $X = D\Lambda D^T$

Step 3: Construct the Matrix D_q is constructed from D by choosing the subspace dimension q. This can be chosen by deciding how much of the variability of the data is desired to be retained.

Step 4: Cluster the vector $|V_1|, |V_2|, \dots, |V_n| \in \boldsymbol{R^q}$ into p clusters using the *K*-means algorithm. Euclidean distance is used as a distance measure in *K*-Means algorithm.

Step 5: Find the corresponding vector V_i from each cluster which is closest to the mean of the cluster. Choose the corresponding features, S_i, as a principal feature. This step will yield the choice of p features.

VI. PROPOSED WORK

Fisher criterion plays an important role in dimensionality reduction technique. This criterion is used to select the features by minimizing the within-class distance and maximizing the between-class distance. Based on the Fisher criterion, two methods are developed based on the Fisher criterion namely, Fisher score is a feature selection method and Linear Discriminant Analysis (LDA) is a subspace learning method. In the past decades, many researchers used the Fisher score and LDA separately to find the best feature subset. In recent days, the researchers combined the two methods to reduce the dimensions and to obtain the finest feature subset. In our proposed algorithm the Modified Fisher Criterion (MFC) is the combination of Modified Fisher Score (MFS) and MLDA.

Modified Fisher Score (MFS)

A. Modified Fisher Score (MFS):

In our earlier study, the Fisher score is modified and it is named as Modified Fisher Score (MFS). The MFS calculates the Euclidean Norm for the given input matrix instead of calculating the trace value and this value is used as Fisher score to select the best features. The Fisher score MFS is calculated by the given formula and it is the replacement of the EQ. (1)

$$M_{FS} = \|(Q_m)(Q_n + \delta Pr)^{-1}\| \qquad (10)$$

$$Q_m = \sum_{i=1}^{n} s_i (\mu_i - \mu)(\mu_i - \mu)^T \qquad (11)$$

$$Q_n = \sum_{j=1}^{m} (z_j - \mu)(z_j - \mu)^T \qquad (12)$$

$$\mu = \sum_{i=1}^{n} s_i \mu_i \qquad (13)$$

where

μ_i is the mean of i^{th} class
s_i is the size of the i^{th} class
μ is the total mean
Q_m is between-class scatter matrix
Q_n is total scatter matrix
δPr is a positive parameter
z_j is the number of samples of the j^{th} class

B. Modified Linear Discriminant Analysis (MLDA):

In the MLDA, Euclidean Norm is computed for the within-class scatter matrix and between-class scatter matrix instead of trace value to reduce the subspace. While calculating the trace value it considers only the diagonal value of the given input matrix it became a drawback for finding the best feature subset. Hence, in MLDA method the whole matrix is considered to calculate the Euclidean Norm value. Hence, the EQ. (8) is rewritten as

$$M_{LDA} = \|(W^T S_w W)^{-1}(W^T S_b W)\| \qquad (4)$$

Where S_b is the between-class matrix , S_w is the within-class matrix.

C. Modified Fisher Criterion (MFC):

The MFC is the fusion of MFS and MLDA. By combining these two methods it is able to do feature selection and subspace learning simultaneously. It is computed using the formula

$$M_{FC} = \|(W^T Q_n W)^{-1}(W^T Q_m W)\| \qquad (10)$$

Where Q_n and Q_m are the total scatter matrix and between-class scatter matrix respectively.

D. MFCPFA Algorithm:

The objective of the dimensionality reduction is to find the best feature subset by reducing the dimension space. The above issue is a challenging task in data mining. To overcome the above issue a new algorithm is proposed by combining the MFC and PFA. The MFC is a supervised method which removes the noisy and irrelevant features but it is unable to remove the redundant features present in the data set. The PFA is an unsupervised method which removes the redundant features by exploring the correlation analysis between the features. In MFC, the *x* features are selected by accumulating them in the decreasing order until it surpasses the users' threshold value of 0.95. Then PFA method is used to cluster the selected x features into y clusters and in each cluster a feature is selected to from finest feature subset. The algorithm is summarized is presented below.

Step 1: Calculate the MFC measure for each feature and arrange it in decreasing order.

Step 2: Select the best *x* features till the MFC measure surpasses the users' threshold value.

Step 3: The selected *x* features are grouped into *y* clusters using PFA method.

Step 4: Finally, select a feature from each cluster and form a best feature subset.

VII. EXPERIMENT

An experiment is conducted on a publicly available dataset to prove the efficiency of the proposed MFCPFA algorithm. The efficiency of the algorithm is examined in terms of Classification accuracy. To evaluate the classification accuracy, K–Nearest Neighbor classifier is used. The proposed algorithm is compared with the existing algorithms: FS+PFA, LDDR and also with our own algorithm MFSPFA proposed in the earlier research work.

In the FS+PFA and LDDR algorithms, trace values are calculated for the given input matrix and the values are used as a Fisher criterion for selecting the relevant features. In our earlier algorithm MFSPFA, only the best feature subset is obtained but the subspace learning is not considered. In the proposed MFCPFA algorithm, the best feature subset is obtained by reducing the subspace.

Table 1. Classification accuracy of MFCPFA algorithm

Dataset	Features	Accuracy Percentage (%)
		MFCPFA
Ionosphere	32	96.11
Isolet	617	93.21
Multi- features	649	92.53
Leukemia	12558	94.85

Table 2. Classification accuracy of MFCPFA and existing algorithms

Dataset	Features	Accuracy Percentage (%)			
		FS+PFA	LDDR	MFSPFA	MFCPFA
Ionosphere	32	92.00	93.11	95.28	**96.11**
Isolet	617	88.13	90.22	92.86	**93.21**
Multi-features	649	83.58	85.13	90.18	**92.53**
Leukemia	12558	89.29	91.86	93.47	**94.85**

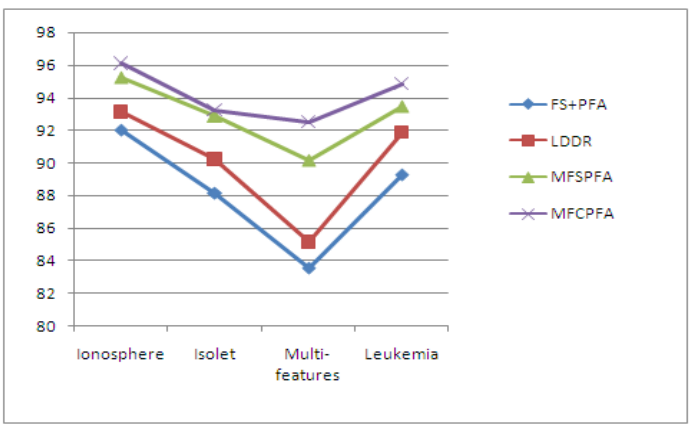

Figure 1. Comparison of MFCPFA and Existing Algorithms

The experiment shows the efficiency of the proposed algorithm and is analyzed through classification accuracy. For this experiment, four publicly available dataset are taken, they are, *Ionsphere*, *Isolet*, *Multi-features*, and *Leukemia*. In the above dataset the best feature subset are obtained using the MFCPFA algorithm. The *K*-Nearest Neighbor classifier is used to obtain the classification accuracy and their results are depicted in Table I. Based on the experiment result the proposed algorithm is compared with the existing feature selection algorithms and their accuracy results are shown in Table II and it is graphically shown in Figure 1. The results of the proposed algorithm have better classification accuracy compared to the existing feature selection algorithms.

VIII.CONCLUSION

In this paper, the proposed MFCPFA algorithm solves the feature selection and subspace learning issues. MFS and MLDA obtained the best feature subset by selecting the features and reduced the dimensions respectively. Then PFA technique is used to remove the redundant features from the selected features. To check the efficiency of the proposed algorithm, an experiment is conducted on a publicly available datasets. The experiment results on four datasets (*Ionsphere*, *Isolet*, *Multi-features*, and *Leukemia*) exhibit that the proposed algorithm is promising in improving the classification accuracy.

IX. REFERENCES

[1] Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, World of Computer Science and Information Technology Journal, Volume 2, Number 2,2012, pp.51-56, ISSN: 2221-0741.

[2] Nebu Varghese, Vinay Verghese, Gayathri. P,N. Jaisankar,A Survey of Dimensionality Reduction and Classification Methods, International Journal of Computer Science & Engineering Survey Volume 3, Number3, June 2012, pp. 45-54

[3] A. Phinyomark, H. Hu, P. Phukpattaranont, C. Limsakul,Application of Linear Discriminant Analysis in Dimensionality Reduction for Hand Motion Classification, Measurement Science Review, Volume 12, Number 3, 2012, pp.82-89.

[4] Zizhu Fan, Yong Xu, Zhang.D, Local Linear Discriminant Analysis Framework Using Sample Neighbors, IEEE Transactions on Neural Networks, Volume 22, Issue 7, July 2011, pp. 1119-1132.

[5] Shilpa lakhina, Sini Joseph, Bhupendra verma, Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD, International Journal of Engineering Science and Technology Volume 2, issue 6, 2010, pp. 1790-1799.

[6] Christos Boutsidis, Michael W. Mahoney, Petros Drineas, Unsupervised feature selection for principal components analysis, 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008, pp. 61-69, ISBN: 978-1-60558-193-4.

[7] Song. F, Zhang. D, Wang. J, Liu. H, Tao. Q, 2007. A parameterized direct LDA and its application to face recognition, Neurocomputing Journal, Volume 71, 2007, pp. 191–196.

[8] M. Ishteva, L. De Lathauwer, P.-A. Absil, S. Van Huffel, Dimensionality reduction for higher-order tensors: algorithms

and applications, International Journal of Pure and Applied Mathematics, Volume 42, Issue 3, 2008, pp. 337-343.

[9] Ireneusz Czarnowski, Cluster-based instance selection for machine classification, Journal of Knowledge and Information Systems, Volume 30, Number 1,2012,pp. 113-133.

[10] Zhao Zhang and Ning Ye, Locality preserving multimodal discriminative learning for supervised feature selection, Journal of Knowledge and Information Systems Volume 27, Number 3, 2011, pp. 473-490.

[11] Xianfa Cai, Jia Wei, Guihua Wen, Jie Li, A Neighborhood Preserving Based Semi-supervised Dimensionality Reduction Method for Cancer Classification, Journal of Advances in Intelligent and Soft computing, Volume 122, 2012, pp. 29-37.

[12] Masashi Sugiyama, Tsuyoshi Idé, Shinichi Nakajima, Jun Sese, Semi-supervised local Fisher discriminant analysis for dimensionality reduction, Journal of Machine Learning, Volume 78, Numbers 1-2, 2010, pp.35-61.

[13] Charles Bouveyron and Camille Brunet, Simultaneous model-based clustering and visualization in the Fisher discriminative subspace, Journal of Statistics and Computing, Volume 22, Number 1, 2012, pp.301-324.

[14] Qinghua Hu, Weiwei Pan, Lei Zhang, Zhang. D,Yanping Song, Maozu Guo, Daren Yu, Feature Selection for Monotonic Classification, IEEE Transactions on Fuzzy Systems, Volume 20 , Issue 1, 2012, pp. 69-81.

[15] Rassoul Noorossana, Karim Atashgar, Abbas Saghaei, An integrated supervised learning solution for monitoring process mean vector, International Journal of Advanced Manufacturing Technology Volume 56, Numbers 5-8, 2011, pp. 755-765.

[16] Pirooz Shamsinejadbabki, Mohammad Saraee, A new unsupervised feature selection method for text clustering based on genetic algorithms, Journal of Intelligent Information Systems Volume 38, Number 3, 2012, pp. 669-684.

[17] Zhao Zhang, Ning Ye, Learning a tensor subspace for semi-supervised dimensionality reduction,Journal of Soft Computing - A Fusion of Foundations, Methodologies and Applications Volume 15, Number 2 ,2011, pp. 383-395.

[18] Ruichu Cai, Zhifeng Hao, Xiaowei Yang, Han Huang, A new hybrid method for gene selection, Journal of Pattern Analysis & Applications, Volume 14, Number 1, 2011,pp. 1-8.

[19] Alex A. Freitas, A Review of evolutionary Algorithms for Data Mining, Soft Computing for Knowledge Discovery and Data Mining, Part II, 2008, pp. 79-111. Alex A. Freitas, A Review of evolutionary Algorithms for Data Mining, Soft Computing for Knowledge Discovery and Data Mining, Part II, 2008, pp. 79-111.

[20] Ruixi Yuan, Zhu Li, Xiaohong Guan, Li Xu, An SVM-based machine learning method for accurate internet traffic classification, Information Systems Frontiers, Volume 12, Number 2, 2010, pp. 149-156.

[21] You Chen, Yang Li, Xue-Qi Cheng, Li Guo, Survey and Taxonomy of Feature Selection Algorithms in Intrusion Detection System, Springer LNCS, Volume 4318, 2006, pp. 153-167.

[22] F. Artigas-Fuentes, R. Gil-Garc´ıa, J.M. Bad´ıa-Contelles, A High-dimensional Access Method for Approximated Similarity Search in Text Mining, International Conference on Pattern Recognition,2010, pp. 3155-3158.

[23] Jian-Bo Yang, Kai-Quan Shen,Chong-Jin Ong,Xiao-Ping Li, Feature Selection for MLP Neural Network: The Use of Random Permutation of Probabilistic Outputs, IEEE Transactions on Neural Networks, Volume 20, issue 12, December 2009,pp. 1911-1922.

[24] Shuiwang Ji, Jieping Ye, Generalized Linear Discriminant Analysis: A Unified Framework and Efficient Model Selection, IEEE Transactions on Neural Networks, Volume 19, Issue, 10, 2008, pp. 1768-1782.

[25] Yijuan Lu, Ira Cohen, Xiang Sean Zhou, Qi Tian, Feature selection using principal feature analysis, Proceedings of the 15^{th} International Conference on Multimedia,Germany, 2007, pp. 301-304, ISBN: 978-1-59593-702-5.