IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 20, Issue 3, Ver. II (May. - June. 2018), PP 26-31
www.iosrjournals.org

Estimation of Classification Errors in Discriminate Variant Reduction Using Feature Projection

D.Ashok kumar[1], P.Velayudham[2]
[1]*(Department of Computer Science, Government Art College/ Bharathidasan University, India)*
[2]*(Department of Computer Science, Government Arts College, Bharathidasan University, India)*
Corresponding Author: D.Ashok Kumar

Abstract: *Dimensionality Reduction is the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. Dimensionality reduction techniques offer solutions that both significantly improve the computation time, and yield reasonably accurate clustering results in high dimensional data. The Reduction techniques are struggled to project the features for variant of discriminant information. The LDA is the best supervised reduction method for linear discriminate information based on the mean values. LDA required more features to project the classification errors. The proposed semi supervised reduction techniques the classification stage minimum distance classifier to identify the projected space. The root means square techniques automatically reconstruct the discriminate classification errors. This is the direction for supervised techniques into semi-supervised techniques. The feature projections have three stages to propose and their performances were compared with each other. First one used the principal component analysis (PCA) for dimension reduction; second one used the linear discriminant analysis (LDA) for dimension reduction. The third one used PCA for the first step of dimensionality reduction, and then used LDA for the next step of dimensionality reduction. The proposed techniques outperform the computation time and accurate clustering results.*

Date of Submission: 18-05-2018 Date of acceptance: 02-06-2018

I. Introduction

Data pre-processing is an important step of data mining that enables the abduction of irrelevant values from a dataset. The main aspect of data pre-processing [6] is dimensionality Reduction (DR). It means the transformation of high-dimensional data into a meaningful representation of reduced dimensionality. The aim of Dimensionality Reduction [11] is to providing low-dimensional representations of high-dimensional data sets to reduce redundant features for noise filtering, compression, clustering, and data mining. For examples include eigenfaces for face recognition, orthogonal decomposition in transform coding, and sparse PCA for cluster analysis. Two types of dimensional reduction methods the linear and Non-Linear, It also subdivides into convex and non-convex techniques. Convex techniques optimize an objective function that does not contain any local optima, where as non-convex techniques optimize objective functions that do contain local optima. The main drawback of PCA [5,7] is that the size of the covariance matrix is proportional to the dimensionality of the data points. As a result, the computation of the eigenvectors might be infeasible for very high-dimensional data (under the assumption that $n > D$).

I.1. Feature Reduction

Feature reduction [2] is a preprocessing techniques used to remove the noise of higher dimensional data. Feature reduction refers to mapping of original higher dimensional data into reduced lower dimensional space. The criterion for feature reduction [6] is differed from different problem settings. For supervised techniques [3, 4] to maximize the class discriminates and the unsupervised techniques to minimize the information loss. The necessary for feature reduction is curse of dimensionality, query accuracy degrades Vs dimension increases, intrinsic dimension is very small, visualization problem, and data compression provides the efficient storage and retrieval.

Feature projection [9, 10] could map high dimensional original features to an appropriate low dimensional space. It could be optimized by a learning criterion so that the features belonging to the same class are clustered and the generalization ability is improved. In addition, dimensionality reduction of feature projection reduces the processing time of pattern recognition, and is beneficial to meet the demand of the real-time control. The principal component analysis (PCA) is a conventional method, which can project high dimension data into a low dimension space and make the data not relevant in the low dimension space. PCA [1] is used to carry out the first step of dimensionality reduction considering that PCA merely generates a well-

described coordinate space of the features without considering class separation. The linear discriminant analysis (LDA) [3, 4] could reduce the dimensionality of features and meanwhile take the class reparability into account and within short processing time. Therefore LDA is used for the second step of dimensionality reduction considering that the combination of PCA and LDA projection may obtain better-integrated merits with excluding the classification errors. The proposed technique gives the attention for the classification stage of reduction, the minimum distance classifier to recognize the projected features and the root means square [12] take the average of mean for recurrently. This process is automatically reconstructing the classification errors into datasets. It improves the classification accuracy for clustering high dimensional data.
The outline of remaining chapter II discuss the Feature extraction, discuss the proposed idea in section III, present some experimental results in section IV and conclude with section V.

II. Feature Extraction

Feature Extraction [8, 9] is frequently used data pre-processing techniques for data mining applications, because it increases the predictive accuracy of data mining tasks. High dimensional data have large number of redundant features to suspect the processing speed for knowledge discovery. The redundant feature transformed into a reduced set of features into a feature vectors with minimal redundancy. This process is called feature extraction. The extracted features are expected to contain the relevant information from the input data. Feature extraction [6] involved to reduce the required resources called required memory and processor speed. The proposed linear dimensionality reduction technique to extract the image redundant features using canny edge detection. The matlab function "impyramid" reduce the redundant feature into a feature eigenvector. The dimension of extracted image is low to compare the original image.

III. Feature Projection and Classification

Feature vector becomes a high-dimensional vector. In order to recognize feature projection is necessary to map a high-dimension feature vector to an appropriate lower dimension space. Feature projection could make the feature vectors, which belong to the same class clustered, and improve the generalization ability of the classifier. In addition, dimensionality reduction of feature projection could reduce the computational cost. This is the time to find the best feature projection techniques to classify the higher dimensional data without the classification degrade. The figure 1 shows the block diagram of feature projection.

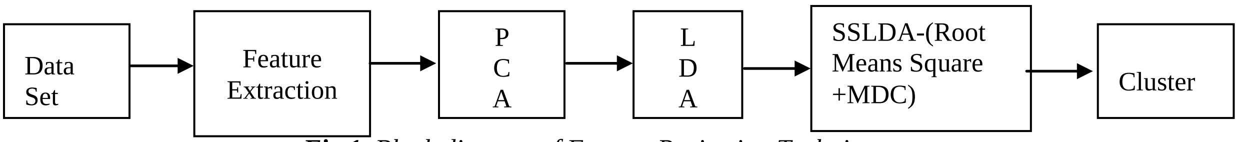


***Fig 1**. Block diagram of Feature Projection Techniques.*

The extracted data come to lower projection with the help of Principal Component Analysis, next for the higher projection apply for discriminant variant analysis using LDA for feature projection based on maximize the between class scatter, it needs more features to project the higher dimension and also occur the classification errors. Thus the proposed SSLDA the classification stage use a technique called Minimum Distance Classifier to recognized projected space to estimate the classification error with the help of root means square recursive process automatically reconstruct the classification errors.

$$Y_1=a_{11}X_1+a_{21}X_2+\ldots+a_{n1}X_n-a_1^rX$$

$$Y_2=a_2X_1+a_{22}X_2+\ldots+a_{n2}X_n-a_2^rX$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$Y_m=a_{1m}X_1+a_{2m}X_2+\ldots+a_{nm}X_n-a_{mr}X \quad (1)$$

As described in the previous work [2], PCA can project transform into solving the following optimization problem.

$$\begin{cases} \max a_1^r c a_1 \\ \text{s.t } a_1^r a_1 = 1 \end{cases} \quad (2)$$

To solve this problem the Lagrange multiplier is used, and the formula $Ca_i = \lambda a_i$ is gotten. And then the principal components could be gotten. In general, The i^{th} principal component of X can be obtained by solving the corresponding eigenvector of the i-th largest eigenvalue, and they are usually required to be independent of each other in order to make the information they contains not overlapping. Finally, a set of orthogonal basis (a_1, a_2,…, a_m), denoted by A, is obtained, and it can project n-dimension random variables $X=(X_1,X_2,\ldots,X_n)^T$ for m(m<n) dimension random variables $Y=(Y_1,Y_2,\ldots,Y_m)^T$ achieving dimensionality reduction meanwhile retaining the maximum information.

$$Y = A^T X \qquad (3)$$

As described in the previous work [3], LDA projects high-dimension vectors onto an optimal discriminate space to extract class information and reduce the vector dimension, and makes sure that the projected vectors have the largest between-class distance and the smallest within-class distance. The between-class scatter matrix and the within-class scatter matrix are defined.

$$S_b = \sum_{i=1}^{c} n_i (u_i - u)(u_i - u)^T \qquad (4)$$

$$Sw = \sum_{i=1}^{c} \sum_{X_k \in c\ class\ i} (ui - xk)(ui - xk)^T \qquad (5)$$

Where $u_i = 1/n \sum x$, ($x \in$ class i) is the mean of the i-th sample, $u = 1/m \sum x_i$ is the mean of the total samples, m is the number of the total samples, n_i is the number of the sample i, c is the number of the classes, and $n_1+n_2+\ldots+n_c=m$. LDA [3,4] needs the lower between-class coupling degree and higher within-class polymerization degree. Thus, the Fisher criterion is introduced.

$$J(W) = |W^T S_a W| \, / \, |W^T S_a W/ \qquad (6)$$

The optimal projection matrix W can be gotten by maximizing J(W). It is easy to prove, to maximize J(W), SBW= λ SWW must be met. A sample X can be projected onto the LDA [4] optimal discriminate space to get a new sample Y.

$$Y = W^T X \qquad (7)$$

The proposed semi-supervised projection scheme, the matrices A and W can be obtained sequentially in the training stage, and then a sample X can be projected onto the PCA. The LDA combination projected space to obtain a new sample Y with root means square based minimum distance classifier (MDC) is one of the simple and efficient classification methods. MDC firstly calculates the center of each class, and then finds the class, which has the minimum distance from the given input sample to the class center, and the input sample is classified into the minimum distance class. In this study, MDC is an appropriate classifier because the projected feature vectors are easy to identify in a low-dimensional space, thus a complex classifier is not necessary. The root means square techniques are automatically reconstruct the classification errors. Thus it is project very large dimensional data set into low dimensional data for less computing time.

IV.Experiment Results and Discussions

Comparison of three feature projection schemes PCA, LDA, and Semi-Supervised LDA were compared with each other. The Semi-supervised techniques outperform the PCA and LDA.

Fig 2. The classification accuracy vs dimensionality in the projected spaces.

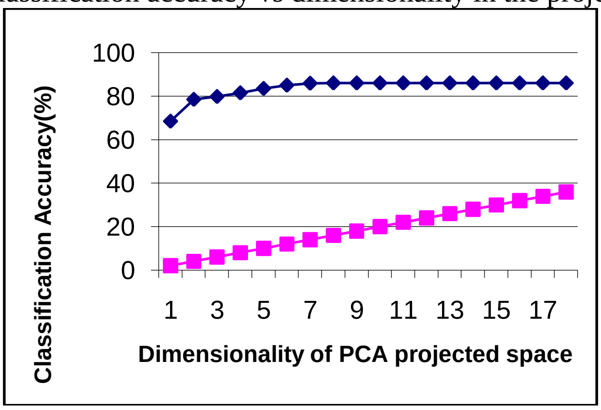


Fig.2.a). The Dimension of PCA,

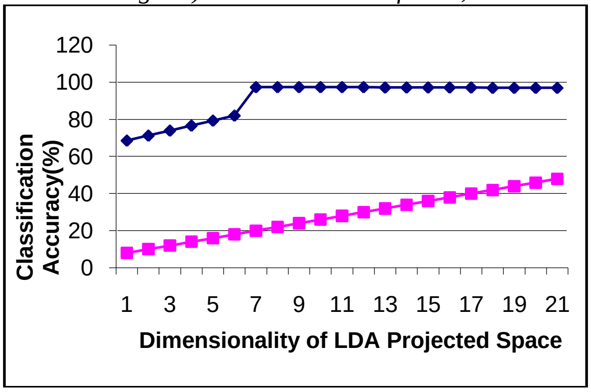


Fig.- 2. b). The Dimension of PCA+LDA

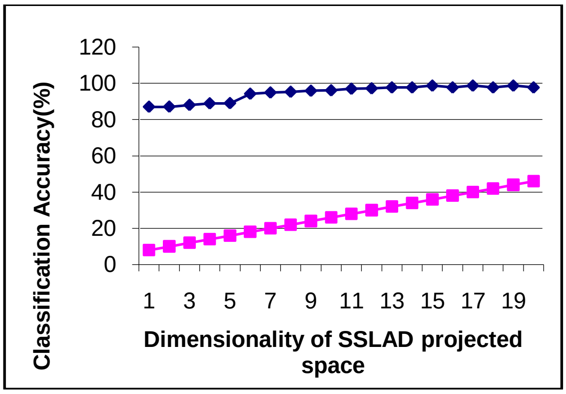


Fig.2. c). The Dimension of Semi Supervised LDA.

A). Performance Evaluation Parameters.

The proposed model is validated using four parameters namely the Accuracy of the classifier, Area under ROC Curve, Sensitivity and Specificity.

TP (True Positive): The number of examples correctly classified to that class.
TN (True Negative): The number of examples correctly rejected from that class.
FP (False Positive): The number of examples incorrectly rejected from that class.
FN (False Negative): The number of examples incorrectly classified to that class.

$$\text{Accuracy} = \frac{TP + TN}{\text{Tot. No of Instances}}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

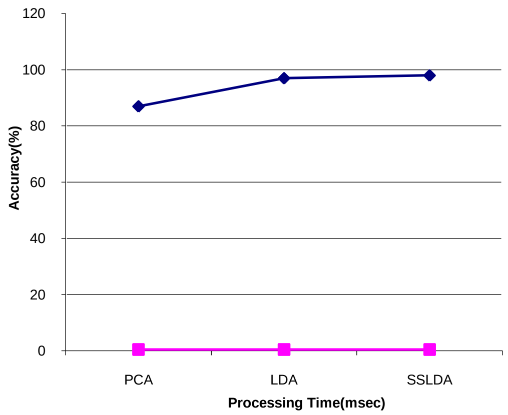
Table 1. Processing Time and Classification result of Projected PCA, LDA, and SS LDA.

Classification Performance	***Feature Projection Schemes***		
	PCA	***LDA***	***SSLDA***
Classification Accuracy(%)	85.6	97.4	97.7
Processing Time(msec)	0.5	0.51	0.49
Error Rate(%)	±2.2	±0.8	±0.6

Experimental results show that PCA had a poor classification performance compared with the other two schemes, and the classification accuracy of LDA and semi- Supervised LDA was very close to each other. PCA is a good tool for dimensionality reduction, yet lacks of the ability of the class separation. LDA conss the class grouping when it learns from the training samples to obtain a linear optimal projected matrix. Thus, SSLDA had a higher accuracy than PCA and supervised LDA for complex computation to the projection process; the processing time of these projection does in have big difference.

ROC describes the tradeoff between Sensitivity and Specificity, as well as the performance of the classifier, can be visualized and studied using the Receiver Operating Characteristic (ROC) curve.

Fig. 3 ROC Curve for Dimensionality Reduction Techniques Vs Classification Accuracy



B). Real Time Image with Reduction.

Dimension= 146 X 26

Fig. 4 Original Images – Before Reduction

In order to verify the performance of the proposed system to reduce the dimension of real-time image, the structure consisting of PCA, LDA feature projection, and Semi supervised LDA classification was employed.

Dim1= 73X63 Dim2= 37 X Dim3= 19X16 Dim3= 10X8

D4 D3 D2 D1

Fig. 5 PCA Reduced images – After Reduction
(Processing Time=0.5 msec)

The PCA reduce the 146X26 dimensional image into the above lower dimension images. The processing time is 0.5 msec, the LDA and the SSLDA the same image to reduce the fewer dimensions with same processing time.

Dim1=110X95 Dim2=55X48 Dim3=28X24 Dim4=14X12

D4 D3 D2 D1

Fig. 6 LDA Reduced images – After Reduction
(Processing Time=0.51msec)

Dim1=73X63 Dim2=37X32 Dim3=19X16 Dim4=14X2

D4 D3 D2 D1

Fig. 7 SSLDA Reduced images – After Reduction
(Processing Time=0.5 msec)

V. Conclusion

The classification accuracy rate is estimated for different reduction techniques using feature projection. The experimental results show that PCA compared with the other linear techniques, it had less classification performance while increase the dimensionality of datasets. The classification accuracy of LDA and semi-supervised LDA was very close to each other. PCA is a good tool for dimensionality reduction, yet lacks of the ability for the complex data class separation. LDA considers the expected class separabillity, when it learns from the training samples to obtain a linear optimal projected matrix. Since it takes more time to classify the complex variant datasets Thus, the proposed Semi-supervised LDA is outperforming to others for complex data dimensionality reduction and classification accuracy is little bit high, it does not bring the complex computation to the projection process, thus the processing time expected improvements. The classification obtained the classification accuracy of 97.7% and just needed the processing time of 0.49(msec). In feature this techniques apply for big data.

References

[1]. MM Steffi, JJR Jose -Comparative Analysis of Facial Recognition involving Feature Extraction Techniques ,2018.

[2]. K Wang, D Zhang, Y Li, R Zhang, "Cost-Effective Active Learning for Deep Image Classification," IEEETransactions on Circuits...Volume: 27 Issue: 12 Dec-2017.

[3]. L Shao, Z Cai, L Liu, K Lu "Performance evaluation of deep feature learning for RGB-D image/video classification," Volumes 385–386, April 2017- Elsevier, Pages 266-283 .

[4]. L. Liu, L. Shao, X. Zhen, and X. Li, "Learning discriminative key poses for action recognition." IEEE Transactions on Cybernetics, 2013.

[5]. Q Ding, J Han, X Zhao, Y Chen - IEEE Transactions on …, "Missing-data classification with the extended full-dimensional Gaussian mixture model: Applications to EMG-based motion recognition," - ieeexplore.ieee.org. Volume: 62 Issue: 8. 2015.

[6]. M Luo, F Nie, X Chang, Y Yang… - ieeexplore.ieee.org Adaptive unsupervised feature selection with structure regularizations", - IEEE transactions - Volume: 29 Issue: 4 . 2018

[7]. Srinivasalu Asadi, Dr. Ch. D. V. Subba Rao, V Saikrishna, "A Comparative Study of Face recognition with Principal Component Analysis and Cross-Correlation Technique", International Journal of Computer Applications (0975 – 8887), November 2010, Volume 10- No.8, pp. 17 – 25.

[8]. G.A. Iffat, S. S. Leslie, "Feature subset selection in large dimensionality domains", Pattern Recognition, vol. 43, no. 1, pp. 5-13, 2010.

[9]. AB Verónica Bolón, M Amparo, CN Sánchez - , "Artificial Intelligence: Foundations, Theory, and Algorithms Feature Selection for High-Dimensional Data,– Springer- Pages 95-124. 2017

[10]. G Haixiang, L Yijing, J Shang, G Mingyun. Learning from class-imbalanced data: Review of methods and applications, Elsevier - Volume 73, 1 May 2017, Pages 220-239 .

[11]. M. Loog, "Semi-supervised linear discriminant analysis through moment-constraint parameter estimation," Pattern Recognition Letters, vol. 37, pp. 24–31, Mar. 2014.

[12]. J. H. Krijthe and M. Loog, "Implicitly Constrained Semi- Supervised Least Squares Classification," Tech. Rep., 2013.

D.Ashok kumar "Estimation of Classification Errors in Discriminate Variant Reduction Using Feature Projection." IOSR Journal of Computer Engineering (IOSR-JCE) 20.3 (2018): 26-31.