ISSN 2455-4863 (Online) **www.ijisset.org** *Volume: 5 Issue: 8 | 2019*

ICA on FA for Feature Extraction in Supervised Classification

Md Shamim Reza[1], Sabba Ruhi[2]

[1]Assistant Professor, Department of Mathematics, Pabna University of Science and Technology, Pabna, Bangladesh
[2]Assistant Professor, Department of Statistics, Pabna University of Science and Technology, Pabna, Bangladesh

Abstract: -*Independent Component Analysis (ICA) and Factor Analysis (FA) are two statistical dimensionality reduction techniques that popularly uses in the field of data mining, intelligence science and machine learning community. Both of the techniques are popular in nature but their projection have different intensions. In this paper, we propose an integrated framework based on ICA and FA in classification analysis. To test the effectiveness of the combined method, the extensive experiments conducted on variety of situation using synthetic and real data analysis. The experimental results shows that ICA and FA does not works well independently, but their combined features space is useful to discover intrinsic information of high dimensional data and significantly improves the classification performance in supervised classification. Based on the experimental results, it is concluded that ICA on FA is a viable alternative to reduce multivariate high dimensional data as well as improve classification performances in supervised classification task.*

Keywords: *ICA, FA, Machine learning, supervised learning etc.*

1. INTRODUCTION

Information explosion has occurred in most of the sciences and researches due to advances in data collection and storage capacity in last few decades [1]-[2]. Modern day datasets are very rich in information with data collected from millions of devices and sensors. This makes the data high dimensional and it is quite common to see datasets with hundreds of features andis not unusual to see it go to tens of thousands. Dealing with thousands and millions of features is a must-have skill for any data scientist to figure out how to use it. Dimensionality reduction is a very useful way to retrieve the intrinsic structure of data as well as to find out the patterns in machine learning applications. It reduces the number of features in the dataset without having to lose much information and keep the model's performance. Feature selection and dimensionality reduction are two method that aim at solving these problems by reducing the number of features and thus the dimensionality of the data.

The most common and useful unsupervised feature transformation is PCA proposed by Pearson in the early 20th century [4]. PCA reduces the dimensionality through orthogonal transformation of data while attempting to preserve as much as variation as possible in the data. Variants of PCA have been developed, some of common are probabilistic PCA leads to an EM-algorithm [5], kernel PCA [6], Gaussian process based probabilistic PCA [7] etc. PCA and its extensions have been successfully applied in a large number of domains such as face recognition, object detection, and computer vision application. The drawbacks of PCA is that these extracted components are not always independent and invariant under transformation, which may contradict to many supervised classification assumptions [8]. Another linear transformation method that commonly used in classification system is LDA, proposed by Fisher. It uses class label to compute the between class and within class matrix, seeks the directions along which the classes are best-separated [9]. However, LDA is a very powerful and useful method for feature extraction, it requires enough training sample in each class and the application of LDA is limited when classes have significant difference between means [10].

Factor Analysis (FA) is also uses in dimension reduction as well as classification task that applied to a set of observed variables to find underlying factors (subsets of variables) from which the observed variables were generated.

Among the very recent development ICA is found to be very useful and effective technique helps to extract representative features in pattern classification. ICA is closely related to PCA and factor analysis. However, ICA technique that is much more powerful.Jutton and Herault [11] originally proposed it for solving blind source separation (BSS) problem. Although ICA was initially developed to solve the BSS problem, past studies have shown that ICA can serve as an effective feature extraction method of improving the classification performance in both supervised classification [12], [13] and unsupervised classification [14]. It has also been found that ICA may help to improve the performance of various classifiers, such as SVM, artificial neural networks, decisions trees, hidden Markov models, and the naive Bayes classifier. In earlier studies, Kwak et al. [15] also showed that ICA outperforms as feature extraction method for face recognition than PCA and LDA.

PCA,FA and ICA are unsupervised and popular in dimensionality reduction but they do not guarantee to generate useful information individually in supervised and unsupervised classification. In this paper, we use both of ICA and FA projection approach together through some statistical criterion. The remaining of

this paper is organized as follows: Section 2discuss approaches of dimensional reduction techniques, Section 3 describes proposed integrated framework. Section 4describes the data sets, experimental results and final Section draws the overall finding that obtain from experiments.

2. DESCRIPTION OF MODEL

2.1 Factor Analysis

Factor analysis can be considered as an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix.The essential purpose of factor analysis is to describe, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors.All variables within a particular group are highly correlated among themselves, but have relatively small correlations with variables in a different group.

Mathematically, suppose the observable random vector **X**, with k components, has mean μand correlation matrix Σ. The factor model postulates that **X** is linearly dependent upon a few unobservable variables$F_1, F_2, \cdots, F_d$ called common factors, and $k$additional sources of variation$\epsilon_1, \epsilon_2 \cdots \epsilon_k$, called errors or, sometimes, specific factors. In particular, the factor analysis model is

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \cdots + l_{1d}F_d + \epsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \cdots + l_{2d}F_d + \epsilon_2$$

$$X_k - \mu_k = l_{k1}F_1 + l_{k2}F_2 + \cdots + l_{kd}F_d + \epsilon_k$$

In matrix notation,

$$\mathbf{X} - \mu = \mathrm{LF} + \epsilon$$

The coefficient l_{ij} is called the loading of the ith variable on the jth factor, so the matrix **L** is the matrix of factor loadings. Note that the ith specific factor ϵ_i is associated onlywith the ith response X_i. The k deviations $X_1 - \mu_1, X_2 - \mu_2, \cdots, X_k - \mu_k$are expressedin terms of $k + d$ random variables $F_1, F_2, \cdots, F_d, \in_1, \in_2, \cdots, \in_k$ which are unobservable.

2.2 Independent Component Analysis

Independent Component Analysis (ICA), which is also a statistical and computational technique for revealing hidden factors that underlie sets of random variables. ICA defines a generative model for the observed multivariate data, which is typically assumed to be nongaussian. As before, in the model, the data variables are assumed linear combinations of some unknown latent variables, and the coefficients of the system are also unknown. The latent variables are also assumed nongaussian and mutually independent and they are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA.

The ICA model is

$$x = f(\theta, s) \qquad (1)$$

Where$x = \left(x_1, x_2, \cdots, x_p\right)^T$ is an observed vector and $f$is a general unknown function with parameter θ that operates on statistically independent latent variables listed in the vector $s = \left(s_1, s_2, \cdots, s_q\right)^T$. A special case of Eq.1 is obtained when the function is linear, and we can write

$$\mathrm{x} = \mathbf{A}\mathrm{s} \qquad (2)$$

Where A is an unknown p ×q mixing matrix. Eq.1 and Eq.2 consider x and $s$as random vectors. The linear model is identifiable under the following fundamental restrictions: at most one of the independent components $s_j$may be Gaussian, and the matrix **A** must be of full column rank. The identifiability of the model is proved in the case p = q, because we have no idea about the number of sources. The goal of ICA is to find a linear mapping **W** such that each component of an estimate y of the source vector can be obtained using the following expression:

$$\mathrm{y} = \mathbf{W}\mathrm{x} = \mathbf{WA}\mathrm{s}$$

The original sources s are exactly recovered when **W** is the pseudo-inverse of **A**up to some scale changes and permutations.

3. PROPOSED METHODOLOGY

The idea of the proposed method is very simple. In the approach, FA apply to the original data, we then retain those factors that can explain at least 90% of the total variation, which is the optimal number of factors or the optimal reduced dimensionality is considering relevant components the goal of our analysis, then standard ICA algorithm applying on extracted factors to optimize the components are independent. Thereafter, standard fourth central moment kurtosis is applied on the extracted independent components to find out meaningful features set in supervised classification, proposed framework showed pictorially details in

Chrat-1.

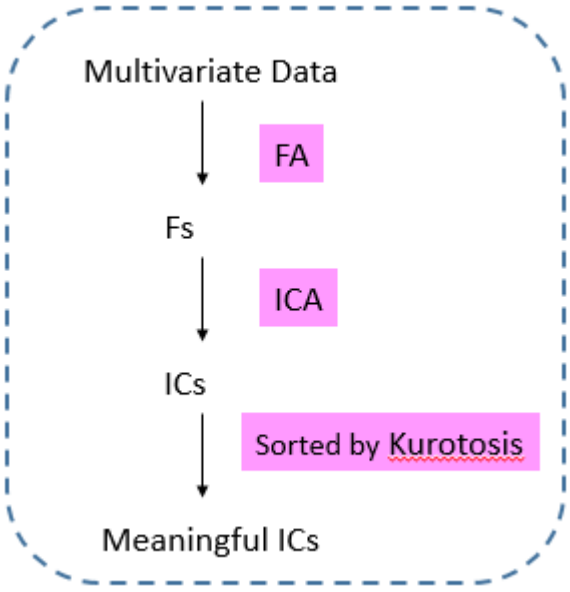

Chart -1: Proposed ICA on FA framework

4. EXPERIMENTAL RESULTS

In the section presents the evaluation of the performance of the proposed feature extraction approaches on one simulated and 3 benchmark UCI machine learning datasets. In the simulation study, 10000 data simulated each from four univariate probability distributions: Normal, student t-distribution Chi-square and uniform. Reshaping each of generated 10000 data into 8 features and 1250 observations individually. Then, the concatenation of final data frame stands 8 features 5000 observations with four distinct class. Proposed method then compared with others traditional linear dimensionality reduction methods, such as principal component analysis (PCA), linear discriminate analysis (LDA) and feature selection approach based on random forest recursive feature elimination (RF-RFE). Accuracy (%) using support vector machine classifier are describes in Table 1 and Chart-2.

Table 1: Classification accuracy (%) using SVM classifier

Dataset	***RF-RFE***	***PCA***	***LDA***	***FA***	***ICA on FA***
Synthetic	60.16	59.08	60.54	59.90	60.80
Sonar	77.95	86.71	89.42	88.40	91.45
Pima	77.08	76.95	77.21	77.93	78.43
Ionosphere	93.70	91.29	90.80	92.52	94.26

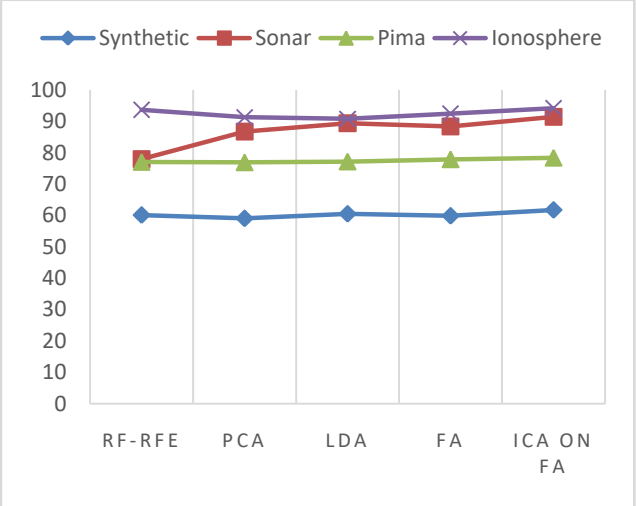

Chart -2: Results comparison on dataset

5. CONCLUSIONS

Feature extraction based on unsupervised method is a challenging task due to lack of class information in supervised classification. To overcome this limitation, we have proposed anovel method in this work. We have empirically compared our proposed methods with other existing methods on simulated and real data sets. Our analysis indicates that ICA and FA, could able to generate a representative feature for classification. The performances of the proposed approach clearly show the necessity of dimensionality reduction in the field of data mining and pattern classification.

REFERENCES

[1] P. Sharma. "*The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes)*". Analytics Vidhya, 2018, 13.

[2] W. J. Frawley, G. P. Shapiro and C. J Matheus. "*Knowledge discovery in databases: An overview*". AI magazine, 1992, 13(3): 57.

[3] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[4] K. Pearson. "LIII. *On lines and planes of closest fit to systems of points in space*". The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559–572.

[5] S. T. Roweis. "*EM algorithms for PCA and SPCA*". In: *Advances in neural information processing systems*, 1998: 626–632.

[6] B. Scholkopf, A. Smola and K-R Muller. "*Nonlinear component analysis as a kernel eigenvalue*

problem". Neural computation, 1998, 10(5): 1299–1319.

[7] N. Lawrence. "*Probabilistic non-linear principal component analysis with Gaussian process latent variable models*". Journal of machine learning research, 2005, 6(Nov): 1783–1816.

[8] A. R. Webb. *Statistical pattern recognition*. John Wiley & Sons, 2003.

[9] A. M Martinez and A. C Kak. "*Pca versus lda*". IEEE Transactions on Pattern Analysis &Machine Intelligence, 2001, (2): 228–233.

[10] M. Zhu and A. M Martinez. "*Subclass discriminant analysis*". IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(8): 1274–1286.

[11] C. Jutten and J. Herault. "*Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*". Signal processing, 1991, 24(1): 1–10.

[12] N. Kwak, Chong-Ho Choi and J. Y. Choi. "*Feature extraction using ICA*". In: International Conference on Artificial Neural Networks, 2001: 568–573.

[13] L. Fan, Kim-Leng Poh and Peng Zhou. "*Partition-conditional ICA for Bayesian classification of microarray data*". Expert Systems with Applications, 2010, 37(12): 8188–8192.

[14] A. Kapoor, T. Bowles and J. Chambers. "*A novel combined ICA and clustering technique for the classification of gene expression data*". In: Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on, 2005: v–621.

[15] N. Kwak. "*Feature extraction for classification problems and its application to face recognition*". Pattern Recognition, 2008, 41(5): 1701–1717.

AUTHORS' BIOGRAPHIES

Dr. Md. Shamim Reza is working as Assistant prof. of department of Mathematics in Pabna University of Science & Technology.He graduated PhD fromPeking University, China. His research focuses on machine learning and medical imaging.

Assistant Prof. Dr. Sabba Ruhi is working as Chairman, department of Statistics in Pabna University of Science & Technology. She graduated PhD from Rajshahi University. Her research focuses on statistical learning theory and Bio-statistics.