ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Comparative study of PCA and ICA in the field of Data Reduction

Vidya Mohanty[*1], P.Annan Naidu[2], Dr.Maya Nayak[3]
[*1] M.Tech Scholar, [2] Assitant Professor, CSE,Centurion University, India
[3] Professor, Orissa Engineering College, India
vidya.mohanty@gmail.com

Abstract

This paper presents the results of a comparative study of Pca and Ica in the field of data reduction. In particular, we compare the two feature extraction techniques- independent component analysis (ICA) and Principal component analysis (PCA) to project microarray data into statistically independent components and genes are clustered according to their mean distances from the calculated centroid. We test the statistical significance of enrichment of gene annotations within clusters. Result shows PCA outperforms ICA in constructing functionally coherent clusters on microarray Breast Cancer Wisconsin, Primary Tumours,, Parkinson's tele monitoring and ecoli data and hepatitis data set.

Keywords: Gene expression clustering, hierarchical cluster, K-means cluster, fuzzy c means, PCA, ICA, cancer data set

Introduction

Microarray analysis techniques are used in interpreting the data generated from experiments on DNA, RNA, and protein microarrays, which allow researchers to investigate the expression state of a large number of genes - in many cases, an organism's entire genome - in a single experiment. Such experiments can generate very large volumes of data, allowing researchers to assess the overall state of a cell or organism. These large data amount can be difficult to analyze, especially in the absence of good gene annotation. Before any kind of microarray data can be analysed for differential expression several steps must be taken. Raw data must be quality assessed to ensure its integrity. Unprocessed raw data will always be subject to some form of technical variation and thus must be preprocessed to remove as many unwanted sources of variation as is possible, to ensure that results are of the highest attainable level of accuracy. Ideally, the data being assayed should be preprocessed using several different methods, the results of which should be compared to identify which method is of the highest level of suitability. The most appropriate method should then be used to preprocess the raw data before differential expression analysis.In an expression matrix, each gene corresponds to one row and each condition/sample to one column. Common tasks in clustering analysis of expression data include i) grouping genes by their expressions over conditions/samples, ii) grouping conditions/samples based on the expression of genes, and iii) finding subgroups of genes and conditions/samples such that the identified genes share similar expression patterns over a specified subset conditions/samples.

Unsupervised (hypothesis-free) approaches are important for discovering novel biological mechanisms, for revealing genetic regulatory networks and for analyzing large datasets for which little prior knowledge is available. Here we apply ICA and PCA as a versatile unsupervised approach for feature extraction for microarray analysis, and evaluate its performance with different clustering techniques k means, Hierarchical and FCM. In clustering, the data consist only of the gene expression values. The analytical goal is to find clusters of samples or clusters of genes such that observations within a cluster are more similar to each other than they are to observations in different clusters. Cluster analysis can be viewed as a data reduction method in that the observations in a cluster can be represented by an 'average' of the observations in that cluster.

Unsupervised analysis methods for microarray data can be divided into three categories: clustering approaches, model based approaches and projection methods. Clustering approaches group genes and experiments with similar behavior [6-10], making the data simpler to analyze [11]. Clustering

ISSN: 2277-9655
Impact Factor: 1.852

methods group genes that behave similarly under similar experimental conditions, assuming that they are functionally related. Most clustering methods do not attempt to model the underlying biology. A disadvantage of such methods is that they partition genes and experiments into mutually exclusive clusters, whereas in reality a gene or an experiment may be part of several biological processes. Model-based approaches first generate a model that explains the interactions among biological entities participating in genetic regulatory networks, and then train the parameters of the model on expression datasets [12-16]. Depending on the complexity of the model, one challenge of model-based approaches is the lack of sufficient data to train the parameters, and another challenge is the prohibitive computational requirement of training algorithms. Projection methods linearly decompose the dataset into components that have a desired property. Both PCA and ICA are used to reduce measured data into a smaller set of components. PCA - utilizes the first and second moments of the measured data, hence relying heavily on Gaussian features .ICA - exploits inherently non-Gaussian features of the data and employs higher moments. PCA is probably the optimal dimension-reduction technique according to the sum of squared errors [17]. Applied to expression data, PCA finds principal components, the eigen arrays, which can be used to reduce the dimension of expression data for visualization, filtering of noise and for simplifying the subsequent computational analyses [18,19]. In contrast to PCA, ICA decomposes an input dataset into components so that each component is statistically as independent from the others as possible. A common application of ICA is in blind source separation (BSS) problems [20]: suppose that there are M independent acoustic sources - such as speech, music, and others - that generate signals simultaneously, and N microphones around the sources. Each microphone records a mixture of the M independent signals. Given N mixed vectors as the signals received from the microphones, where N ≥ M, ICA retrieves M independent components that are close approximations of the original signals up to scaling. ICA has been used successfully in BSS of neurobiological signals such as electroencephalographic (EEG) and magneto encephalographic (MEG) signals [21-23], functional magnetic resonance imaging (fMRI) data [24] and for financial time series analysis [25,26]. ICA can also be used to reduce the effects of noise or artifacts of the signal [27] because usually noise is generated from independent sources. Most applications of ICA assume that the source signals are mixed linearly into the input signals, and algorithms for linear ICA have been developed extensively [28-32]. In several applications nonlinear mixtures may provide a more realistic model and several methods have been developed recently for performing nonlinear ICA [33-35]. Liebermeister [36] first proposed using linear ICA for microarray analysis to extract expression modes, where each mode represents a linear influence of a hidden cellular variable. However, there has been no systematic analysis of the applicability of ICA as an analysis tool in diverse datasets, or comparison of its performance with other analysis methods.Here we apply PCA and ICA to microarray data analysis and project the genes into clusters.

Multivariate Statistical Technique

Multivariate Statistical Technique

Multivariate data analyses can extract information from large data sets containing observations related to a wide range of variables. Principal component analysis (PCA) (Joliffe, 1986) and PLS are two multivariate projection methods that can handle problems associated with most microarray data such as missing values, the presence of more variables than observations, and noise.

PCA is the oldest and best known of the multivariate projection techniques. The central idea of PCA is to reduce the dimensionality of a data set, X, while retaining as much as possible of the variation present in the data.The reduction is accomplished by introducing a new setof variables, the principal components, which are linear combinations of the original variables and uncorrelated to each other. The principal components can be determined using the NIPALS algorithm (Wold, 1966) or by singular value decomposition (SVD) .As in PCA, principal components are constructed to reduce the dimensions of X. . Consider a data matrix Xnxp with *p* component in each random vector *X*. A linear function β^T have maximum variance. So that,

$$\beta_1^T x = \beta_{11}x_1 + \beta_{12}x_2 + \cdots + \beta_{1p}x_p$$

$\beta_2^T x$, $\beta_2^T x$, $\beta_3^T x$ …,$\beta_n^T x$ are uncorrelated and $\beta_k^T x$ is the k th PC. We are interested to find variance and structure of correlation and covariance of p variables. Generally PCA concentrate on variance rather than correlation and covariance. We will assume that the mean vector is 0 and $\sum$ (singular and positive semi-definite) is covariance matrix. Using this PCs we want to find out grouping or clusters in multivariate data set for subsequent procedure.

ISSN: 2277-9655
Impact Factor: 1.852

Independent Component Analysis (ICA)

When data cannot be ensembled (hence, most likely non-Gaussian via Central Limit Theorem); when raw data appear to be very noisy; when a sensor records several source signals simultaneously .Mathematically, PCA is adequate if the data are Gaussian, linear, and stationary. If not, then higher order statistics begin to be essential. Recently, ICA has been used by biomedical scientists as an unsupervised approach to explore gene expression features and discover novel underlying biological information from large microarray data sets. Gene expression data provided by microarray technology is considered a linear combination of independent components having specific biological interpretations. Let the $n \times k$ matrix $\boldsymbol{X}$ denote microarray gene expression data with k genes under n samples or conditions. x_{ij} in $\boldsymbol{X}$ is the expression level of the j^{th} gene in the i^{th} sample. Generally speaking, the number of genes k is much larger than that of the samples n, $k>>n$.

Suppose that the data have been preprocessed and normalized (i.e., each sample has zero mean and standard deviation); then the ICA model for gene expression data is the same as

$$\boldsymbol{X} = \boldsymbol{AS}$$

The $k \times n$ matrix $\boldsymbol{X}$ is used to denote k genes under n samples. In this case, the transform, $\boldsymbol{X}^T$ is used in the ICA model: $\boldsymbol{X}^T = \boldsymbol{AS}^T$. So, $\boldsymbol{X}^T$ here denotes the same $n \times k$ matrix as used in the ICA model.In ICA modeling of microarray data, the columns of $\boldsymbol{A} = [a_1, a_2, \ldots, a_n]$ are the $n \times n$ latent vectors of the gene microarray data, $\boldsymbol{S}$ denotes the $n \times k$ gene signature matrix or expression mode, in which the rows of $\boldsymbol{S}$ are statistically independent to each other, and the gene profiles in $\boldsymbol{X}$ are considered to be a linear mixture of statistically independent components $\boldsymbol{S}$combined by an unknown mixing matrix $\boldsymbol{A}$. characteristically latent variables have been obtained, the corresponding elementary modes can be identified, which yields useful information for classification. Also, the distribution of gene expression levels generally features a small number of significantly overexpressed or under-expressed genes which form biologically coherent groups and may be interpreted in terms of regulatory pathways.

K-means clustering

K-means clustering partitions the input data set into K subsets. A description of K-means clustering is as follows:

1. Initial K cluster centers (or average expression vectors) are located randomly or by using prior knowledge, and all the clustering objects are assigned to the closest center.
2. The cluster center is then updated for each cluster and this is used to compute the distances between clusters.
3. Cluster objects are moved iteratively between clusters and intra- and inter-cluster distances are measured with each move. Objects can remain in the new cluster only if they are closer to it than to their former cluster.
4. The centers for each cluster are recalculated after each move.
5. The iteration proceeds until moving any more objects would increase intra-cluster distances and decrease inter-cluster dissimilarity.

Hierchical clustering

Hierarchical Clustering is the most popular method for gene expression data analysis. In hierarchical clustering, genes with similar expression patterns are grouped together and are connected by a series of branches (clustering tree or *dendrogram*). Experiments with similar expression profiles can also be grouped together using the same method.

Fuzzy C Means

In fuzzy clustering data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. One of the most widely used fuzzy clustering algorithms is the Fuzzy C-Means (FCM) Algorithm (Bezdek 1981). The FCM algorithm attempts to partition a finite collection of n elements into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of c cluster centres partition matrix , where each element wij tells the degree to which element xi belongs to cluster cj . Like the k-means algorithm, the FCM aims to minimize an objective function. The standard function is:

$$w_k(x) = \frac{1}{\sum_j \left(\frac{d(\text{center}_k, x)}{d(\text{center}_j, x)}\right)^{2/(m-1)}}.$$

which differs from the k-means objective function by the addition of the membership values uij and the fuzzifier m. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships wij and hence, fuzzier clusters. In the limit m = 1, the memberships wij converge to 0 or 1, which implies a crisp partitioning. In the absence of experimentation or domain knowledge, m is commonly set to 2. The basic FCM Algorithm, given

ISSN: 2277-9655
Impact Factor: 1.852

n data points (x1, . . ., xn) to be clustered, a number of c clusters with (c1, . . ., cc) the center of the clusters, and m the level of cluster fuzziness with, The basic process of Fuzzy c means clustering's are

1 Choose a number of clusters.
2 Assign randomly to each point coefficients for being in the clusters.
3 Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than , the given sensitivity threshold) :
4 Compute the centroid for each cluster, using the formula above.
5 For each point, compute its coefficients of being in the clusters, using the formula above

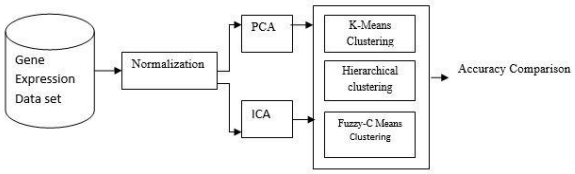

Figure 1: Proposed Modal

Proposed Model

Gene expression microarrays provide a snapshot of all the transcriptional activity in a biological sample. Unlike most traditional molecular biology tools, which generally allow the study of a single gene or a small set of genes, microarrays facilitate the discovery of totally novel and unexpected functional roles of genes. The power of these tools has been applied to a range of applications, including discovering novel disease subtypes, developing new diagnostic tools, and identifying underlying mechanisms of disease or drug response. However, this technology necessarily produces a large amount of data, challenging us to interpret it by exploiting modern computational and statistical tools . We have downloaded benchmark dataset from UCI repository shown in table 1. In our proposed model we are subjecting the data to undergo normalization that is done using various techniques like Min-max normalization, Z-score normalization and normalization by decimal scaling. Min-max normalization performs a linear transformation on the original data. Suppose that mina and maxa are the minimum and the maximum values for attribute A. Min-max normalization maps a value v of A-v in the range (0, 1) by computing:

$$v' = \frac{v - \min_a}{(\max_a - \min_a)}$$

Once data is normalized its feature is extracted using PCA and ICA using algorithm 1 and 2 respectively. This data set having reduced dimension is passed through different clustering technique such as K Means, Hierarchical and Fuzzy c means technique. Their output is validated using b index using algorithm 3. Result shows PCA is better than ICA.

Algorithm 1: PCA

Step 1.Get some data.
Step 2.Substract the mean
Step 3. Calculate the covariance matrix.
Step 4. Calculate the eigenvectors and eigenvalues of the covariance Matrix.
Step 5. Choosing components and forming a feature vector

Algorithm 2: ICA

Step 1. Centering
The most basic and necessary preprocessing is to center x, i.e. subtract the mean vector m = E{x} so as to make x a zero-mean variable
Step 2. . After estimating the mixing matrix A with centered data, we can complete the estimation by adding the mean vector of s back to the centered estimates of s. The mean vector of s is given by $A^{-1}m$, where m is the mean that was subtracted in the preprocessing.
Step 3: for i=1:n
W=random vector;
Orthogonalize initial vector w in terms of the previous components;
Normalize w;
While w(not converged)
Approximation of negonntrophy of $w^T x$
Orthogonalize w in terms of the previous components;
Normalize;
End while;
W(:,i)=w
End for
S=W*X, return s;

Algorith 3: Davies-Bouldin Index

The Davies-Bouldin criterion is based on a ratio of within-cluster and between-cluster distances. The Davies-Bouldin index is defined as

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\{D_{i,j}\},$$

where $D_{i,j}$ is the within-to-between cluster distance ratio for the *i*th and *j*th clusters. In mathematical terms,

$$D_{i,j} = \frac{\left(\bar{d}_i + \bar{d}_j\right)}{d_{i,j}}.$$

$\bar{d}_i$ is the average distance between each point in the *i*th cluster and the centroid of the *i*th cluster. $\bar{d}_j$ is

ISSN: 2277-9655
Impact Factor: 1.852

the average distance between each point in the *i*th cluster and the centroid of the *j*th cluster. $d_{i,j}$ is the Euclidean distance between the centroids of the *i*th and *j*th clusters.

The maximum value of $D_{i,j}$ represents the worst-case within-to-between cluster ratio for cluster *i*. The optimal clustering solution has the smallest Davies-Bouldin index value.

Experimental Evaluation

Proposed model is tested using different data set Breast cancer Wisconsin, Primary Tumours,,, Parkinson's tele monitoring ecoli data and hepatitis data set. High dimension data is at as a curse to the computation. These high dimension can be reduced using PCA and ICA.

Table 1: Data Set description

Data Set	Number of Instances	Number of Attribute	Number of class
Breast cancer	1484	8	10
Ecoli data	72	7129	2
Parkinson's telemonitoring	60	7129	2
Primary tumours	197	581	4
hepatitis	45	4026	2

After normalization and feature extraction dimension of data is reduced is shown in Table 2.

Table 2: Feature Extracted using PCA and ICA

Data Set	PCA	ICA
Breast cancer	1484X5	1484X4
Ecoli data	72X	72X
Parkinson's telemonitoring	60	7129
Primary tumours	197	581
hepatitis	45	4026

Feature extracted data is use to find cluster using different clustering techniques whose output are shown in figure 2-6. Validation of clustering is done using b index shown in table 3. Table 3 clearly depicts that PCA is better then ICA for gene expression data set.

Table 3: B Index : of different clustering technique using PCA and ICA

Data Set	CLUSTERING					
	HIERACHICAL		K MEANS		FUZZY C MEANS	
	PCA	ICA	PCA	ICA	PCA	ICA
Breast cancer	1.134	4.020	229.8	8.442	6906 2.7	1.0977
Ecoli data	2.755	8.106	9.412	27.01	142.1	26364 39.9
Parkinson's telemonitoring	0.954	2.009	18753	7.271	1.149	12827.4
Primary tumours	1.759	8.080	24978	15.48	0.238	383191.4
hepatitis	6.887	5.856	7.063	8.083	6.63	19.114

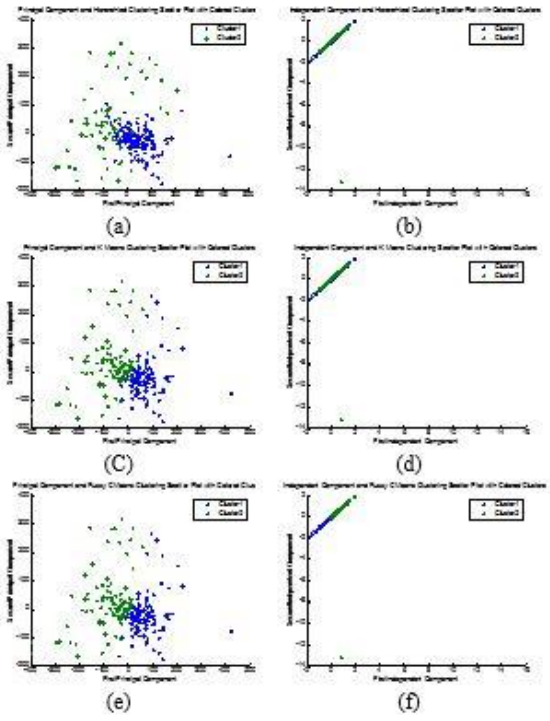

Figure 2: Breast cancer Clustering a) PCA Hierarchical b) ICA Hierarchical c) PCA KMeans d) ICA K Means e) PCA FCM f)ICAM

ISSN: 2277-9655
Impact Factor: 1.852

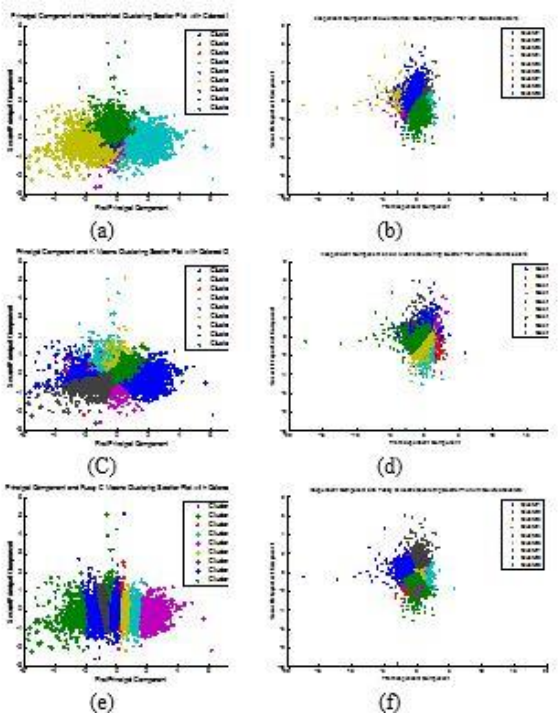

Figure 3: ecoli Clustering a) PCA Hierarchical b) ICA Hierarchical c) PCA KMeans d) ICA K Means e) PCA FCM f)ICA FCM

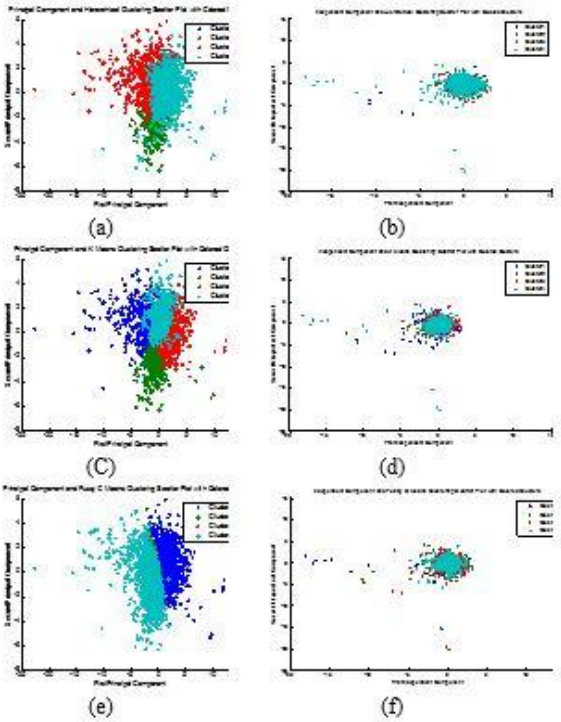

Figure 4:Parkinson's Tele monitoring Clustering a) PCA Hierarchical b) ICA Hierarchical c) PCA KMeans d) ICA K Means e) PCA FCM f)ICA FCM

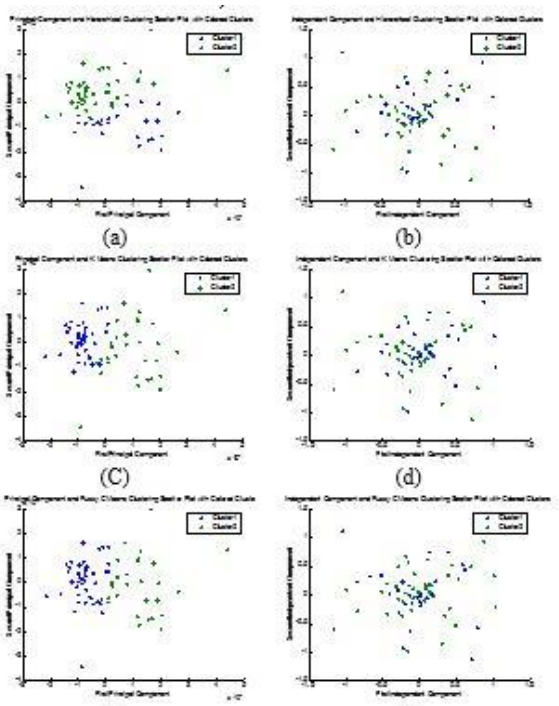

Figure 5: Primary Tumour data Clustering a) PCA Hierarchical b) ICA Hierarchical c) PCA KMeans d) ICA K Means e) PCA FCM f)ICA FCM

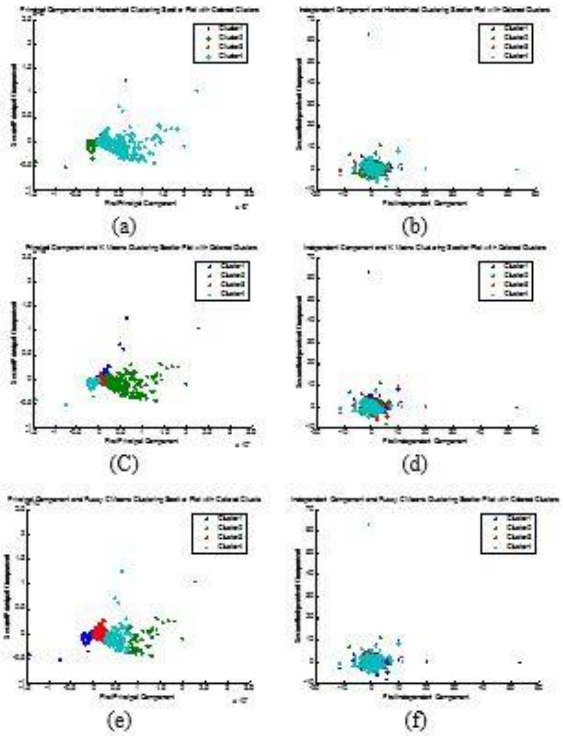

Figure 6: Hepatitis data set Clustering a) PCA Hierarchical b) ICA Hierarchical c) PCA K Means d) ICA K Means e) PCA FCM f)ICA FCM

ISSN: 2277-9655
Impact Factor: 1.852

Conclusion

If we compare all the data sets we observe that PCA outperforms ICA in all the above data sets.ICA is unable to find clear clusters using the raw data . A careful look on scatter plot of ICA is very important to find out how PCA outperforms ICA. PCA minimizes the covariance of the data; on the other hand ICA minimizes higher-order statistics such as fourth-order cummulant (or kurtosis), thus minimizing the mutual information of the output. Specifically, PCA yields orthogonal vectors of high energy contents in terms of the variance of the signals, whereas ICA identifies independent components for non-Gaussian signals.ICA thus possesses two ambiguities :First, the ICA model equation is underdetermined system ;one cannot determine the variances of the independent Components .Second, one cannot rank the order of dominant components.

References

[1] *Butte A: The use and analysis of microarray data. Nat Rev Drug Discov 2002, 1:951-960.*

[2] *Ando T, Suguro M, Hanai T, Kobayashi T, Honda H, Seto M: Fuzzy neural network applied to gene expression profiling for predicting the prognosis of diffuse large B-cell lymphoma. Jpn J Cancer Res 2002, 93:1207-1212.*

[3] *Brown M, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M, Haussler D: Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc Natl Acad Sci USA 2000, 97:262-267.*

[4] *Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999, 286:531-537.*

[5] *Mukherjee S, Tamayo P, Mesirov JP, Slonim D, Verri A, Poggio T: Support vector machine classification of microarray data. Technical Report No. 182, AI Memo 1676 MIT, Cambridge: Massachusetts Institute of Technology; 1999.*

[6] *Ben-Dor A, Shamir R, Yakhini Z: Clustering gene expression patterns. J Comput Biol 1999, 6:281-297.*

[7] *Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998, 95:14863-14868.*

[8] *Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS: A gene expression map for Caenorhabditis elegans. Science 2001, 293:2087-2092.*

[9] *Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan E, Dmitrovsky E, Sander ES, Golub TR: Interpreting patterns of gene expression with self-organizing maps: method and application to hematopoietic differentiation. Proc Natl Acad Sci USA 1999, 96:2907-2912.*

[10] *Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: Systematic determination of genetic network architecture. Nat Genet 1999, 22:281-285.*

[11] *Kaminski N, Friedman N: Practical approaches to analyzing results of microarray experiments. Am J Respir Cell Mol Biol 2002, 27:125-132.*

[12] *Bussermaker HJ, Li H, Siggia ED: Regulatory element detection using correlation with expression. Nat Genet 2001, 27:167-174.*

[13] *Friedman N, Linial M, Nachman I, Pe'er D: Using Bayesian networks to analyze expression data. J Comput Biol 2000, 7:601-620.*

[14] *Lazzeroni L, Owen A: Plaid models for gene expression data. Statistica Sinica 2002, 12:61-86.*

[15] *Segal E, Battle A, Koller D: Decomposing gene expression into cellular processes. In Proceedings of the Eighth Pacific Symposium on Biocomputing: January 3-7 2003 Edited by: Altman RB, Durker AK, Hunter L, Jung TA, Klein TE. Kauai, Hawaii: World Scientific Publishing Company;; 2003:89-100.*

[16] *Segal E, Barash Y, Simon I, Friedman N, Koller D: From promoter sequence to expression. In Proceedings of the Sixth Annual International Conference on Research in Computational Molecular Biology: April 18-21 2002 Edited by: Myers G, Hannenhalli S, Saukoff D, Istrail S, Pevzner P, Waterman M. Washington DC: ACM Press; 2002:263-272.*

[17] *Jolliffe IT: Principle Component Analysis. New York: Springer-Verlag; 1986.*

[18] *Alter O, Brown PO, Botstein D: Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 2000, 97:10101-10106.*

[19] *Misra J, Schmitt W, Hwang D, Hsiao L, Gullans S, Stephanopoulos G, Stephanopoulos G: Interactive exploration of microarray gene expression patterns in a reduced dimensional space. Genome Res 2002, 12:1112-1120.*

[20] *Jutten C, Herault J: Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. Signal Processing 1991, 24:1-10.*

ISSN: 2277-9655
Impact Factor: 1.852

[21] Makeig S, Bell AJ, Jung TP, Sejnowski TJ: Independent component analysis of electroencephalographic data. In Proceedings of the Advances in Neural Information Processing Systems: November 27-December 2 1995; Denver, Colorado Edited by: Touretzky D, Mozer M, Hasselmo M. Cambridge (MA): MIT Press; 1996:145-151.

[22] Vigario R: Extraction of ocular artifacts from EEG using independent component analysis. Electroenceph Clin Neurophysiol 1997, 103:395-404.

[23] Vigario R, Jousmaki V, Hamalainen M, Hari R, Oja E: Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In Proceedings of the Advances in Neural Information Processing Systems: Dec 1-6 1997; Denver, Colorado Edited by: Jordan M, Kearns M, Solla S. Cambridge (MA): MIT Press; 1998:229-235.

[24] Stone JV, Porrill J, Porter NR, Wilkinson ID: Spatiotemporal independent component analysis of event-related fMRI data using skewed probability density functions. NeuroImage 2002, 15:407-421.

[25] Back AD, Weigend AS: A first application of independent component analysis to extracting structure from stock returns. Int J Neural Syst 1997, 8:473-484.

[26] Kiviluoto K, Oja E: Independent component analysis for parallel financial time series. In Proceedings of the Fifth International Conference on Neural Information Processing: October 21-23 1998 Edited by: Kearns M, Solla S, Cohn D. Kitakyushu, Japan: MIT Press; 1999:895-898.

[27] Hyvärinen A, Karhunen J, Oja E: Independent Component Analysis. New York: John Wiley & Sons; 2001.

[28] Amari S: Natural gradient works efficiently in learning. Neural Comput 1998, 10:251-276.

[29] Bell AJ, Sejnowski TJ: An information-maximization approach to blind separation and blind decovolution. Neural Comput 1995, 7:1129-1159.

[30] Cardoso JF: High-order contrasts for independent component analysis. Neural Comput 1999, 11:157-192.

[31] Hyvärinen A: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 1999, 10:626-634.

[32] Lee TW, Girolami M, Sejnowski TJ: Independent component analysis using an extended Infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput 1999, 11:417-441.

[33] Burel G: Blind separation of sources: a nonlinear neural algorithm. Neural Networks 1992, 5:937-947.

[34] Harmeling S, Zieche A, Kawanabe M, Muller K: Kernel feature spaces and nonlinear blind source separation. Proceedings of the Advances in Neural Information Processing Systems: December 3-8 2001; Vancouver, British Columbia, Canada Edited by: Dietterich TG, Becker S, Ghahramani Z. Cambridge (MA): MIT Press; 2002:761-768.

[35] Hyvärinen A, Pajunen P: Nonlinear independent component analysis: existence and uniqueness results. Neural Networks 1999, 12:429-439. Liebermeister W: Linear modes of gene expression determined by independent component analysis. Bioinformatics 2002, 18:51-60