ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Paper / Case Study
Available online at: www.ijarcsms.com

Performance Analysis of Hybrid (SVM+ICA) Method for two class dataset

P. Malathi[1]
PG Scholar
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore - India

B. Kirthika[2]
PG Scholar
Department of Electronics and Communication Engineering
Kumaraguru College of Technology
Coimbatore - India

Abstract: In this paper, a new hybrid (SVM+ICA) dimensionality reduction method is proposed to improve the classification accuracy and to represent the data in a better way. Standard dimensionality reduction methods are classified into stand-alone and hybrid methods. The hybrid approach combines both supervised and unsupervised criterion. This paper uses supervised approach as Support Vector Machines (SVM) and unsupervised approach as Independent Component Analysis (ICA). SVM uses structural risk minimization to find decision boundaries to improve the performance and ICA improves the performance by the maximum independence among features. The orthogonal projection is used to minimize the redundancy and to improve the dimensionality reduction. The kNN classifier is used to classify the data samples and the classification accuracy is calculated. Experiments are performed for various two class datasets and the performance of hybrid (SVM+ICA) method is compared with the stand-alone methods (SVM, ICA).

Keywords: Hybrid dimensionality reduction, Structural risk minimization, Projection, Data independence, support Vector Machines, Independent Component Analysis.

I. INTRODUCTION

Advances in data collection and storage capabilities during the past decade have lead to an information overload in most sciences. There is an increasing demand for high dimensional data for applications such as electrocardiogram signal analysis and gene expression analysis for cancer detection. Dimensionality reduction is the process of reducing the number of random variables and it transforms the data in a high-dimensional space to lower dimensions space. This is performed in order to improve classification accuracy and to reduce the computational complexity. Dimensionality reduction process is classified into standalone and hybrid methods. The standalone method uses either supervised or unsupervised approach. The supervised approach needs prior knowledge of class assignment for training data and unsupervised approach do not need knowledge about class assignment. The hybrid method integrates both supervised and unsupervised methods inorder to improve classification accuracy and to represent the data in a better way. There are many issues challenging the efficiency of the hybrid approach. There are many supervised methods such as Linear Discriminant Analysis (LDA)[2] and the projection in this is obtained by maximizing the variance between classes and minimizing the variance within class to achieve better accuracy and it is extended to kernel Discriminant Analysis (KDA). There are three issues in this method such as small sample size, common mean and robustness problem. These problems are alleviated by some other supervised methods such as null space LDA [3], Discriminative Common Vector (DCV), Orthogonal Centroid Method (OCM) [4]. There are some unsupervised methods such as Principal Component Analysis (PCA), Regression model [5] and intrinsic data geometry preservation to provide robustness and better data representation capability. PCA is extended to kernel PCA (kPCA) [6] to perform nonlinear dimensionality reduction. Regression approaches can be classified as unsupervised dimensionality reduction method and the response is not identical to the class assignment. Existing hybrid approaches are asymmetric principal and discriminant analysis (APCDA) [7], ICA augmented by LDA [8], discriminant nonnegative matrix factorization (DNMF) [9]. The proposed hybrid (SVM+ICA) method improves the classification performance.

II. Support Vector Machines

Support Vector Machine (SVM) is a supervised learning method that analyse data and recognize patterns and it is used for classification and regression. In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. In addition to perform linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

A support vector machine constructs a hyperplane or set of hyper-planes in a high or infinite dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. In order to perform nonlinear classification, Kernel trick is used. For SVM if a training data of D is given, a set of n points is of the form,

$$D= \{(x_i, y_i) \mid x_i \in R^p, y_i \in \{-1, 1\}$$

where y_i is either 1 or -1 indicating the class to which the point x_i belongs. The hyperplane is given by,

$$w^T x + b = 0$$

where w is the normal value to the hyperplane and b represents the bias value. The margin of SVM is denoted by,

$$\text{Margin}=\frac{2}{||W||}$$

The margin has to be maximized in order to reduce the error.

III. Independent Component Analysis

In signal processing, independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents by assuming that the subcomponents are non-Gaussian signals and that they are all statistically independent of each other. ICA is a special case of blind source separation.ICA finds the independent components (also called factors, latent variables or sources) by maximizing the statistical independence of the estimated components. We may choose one of many ways to define independence, and this choice governs the form of the ICA algorithm. The two broadest definitions of independence for ICA are Minimization of mutual information and Maximization of non-Gaussianity. Typical algorithms for ICA use centering (subtract the mean to create a zero mean signal), whitening (usually with the eigenvalue decomposition) and dimensionality reduction as preprocessing step in order to simplify and reduce the complexity of the problem for the actual iterative algorithm. Whitening and dimension reduction can be achieved with principal component analysis or singular value decomposition. Whitening ensures that all dimensions are treated equally a priori before the algorithm is run.

Well-known algorithms for ICA include infomax, FastICA and JADE, but there are many others. In general, ICA cannot identify the actual number of source signals, a uniquely correct ordering of the source signals, nor the proper scaling (including sign) of the source signals. The nongaussianity of ICA is measured by the Kurtosis, Negentropy. ICA is used in applications such as mixed signal separation and dimensionality reduction. Fast ICA package is properly used. Linear independent component analysis can be divided into noiseless and noisy cases, where noiseless ICA is a special case of noisy ICA. Nonlinear ICA should be considered as a separate case.

IV. SVM+ICA METHOD

This section presents an effective linear hybrid dimensionality reduction method based on Support Vector Machine (SVM) and Independent Component Analysis (ICA), referred to as SVM plus ICA (SVM+ICA), to maintain high classification accuracy in lower dimensional space that is less sensitive to noise. Since SVM+ICA is not based on LDA, it does not suffer from the S3 or common mean problems inherited from the LDA criteria. This hybrid method provides better performance compared to methods such as SVM, PCA, LDA and ICA. This method is referred as SVM + ICA method.

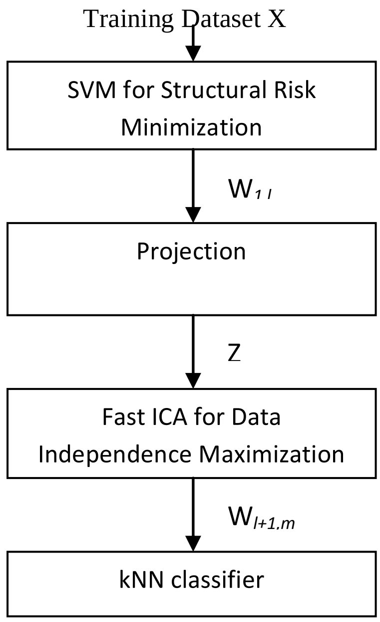

Fig.1. Block Diagram of proposed SVM+ICA Method

Fig.1. provides a block diagram of the proposed linear SVM+ICA method. It consists of four components namely structural risk minimization, projection, independence maximization, and kNN classifier. In Fig.1 X= {xi Є Rn, "i} represents a training data set of dimension n, which is to be reduced to another set, S, of dimension, m, where m << n, using the projection matrix, of m mapping column vectors constructed from the SVM+ICA process. The structural risk minimization component generates the first l mapping vectors of W, denoted as W1,l and the data independence maximization component yields the other m-l vectors of W, denoted as Wl+1,m. Z is the projected dataset from X based on W1,l and it is fed to the data independence maximization component to find Wl+1,m.The weight vector from independence maximization is given to the kNN classifier and the classification accuracy is computed.

A. *Structural Risk Minimization*

Structural risk minimization (SRM) is an inductive principle of use in machine learning. The SRM principle addresses this problem by balancing the model's complexity against its success at fitting the training data. SVM minimizes structural risk so as to offer projection with better generalization capability to improve classification/estimation performance for unknown samples. Since maximum margin among features provides better data representation to improve classification performance. The structural risk minimization step generates W1,*l* whose column vectors represent the direction of the decision surface in classification. The concept of using the decision surface as mapping vectors for the supervised dimensionality reduction is not new. LDA is the first-of-a-kind that adopts this idea. The only difference between LDA and SVM is the different objective functions; they utilize to obtain the decision surface.

The structural risk-based dimensionality reduction shows equal or better classification accuracy than LDA or kDA, since LDA can only obtain a decision boundary identical to the one from SVM when there exist sufficiently large number of observations for effective representation of the internal structure of data.SVM provides the decision surface satisfying minimum structural risk by maximizing the separation margin through constrained quadratic problem. The dual problem for binary classification is given by,

$$\alpha^* = \text{argmin}_\alpha\{ (1/2)\alpha^T Q\alpha - \alpha^T 1\}$$

$$\text{s.t.} \sum \alpha_i y_i = 0,\ 0 \le \alpha_i \le C$$

where α_i represents the Lagrange multiplier and x_i is the data sample and y_i Є {-1, 1} is the class index and N is the number of samples in the training set.1 represents the column vector consist of N many 1s. C is the relaxation parameter to tolerate certain level of empirical error in decision margin so as to generalize the decision boundary for arbitrary input. $Q = [q_{ij}]$ is an N*N matrix where $qij=y_i\, y_j\, (x_i, x_j)$, i, j Є {1… N}. Since w is the core information of the decision making process, we utilize w as part of the overall linear mapping, W, in the proposed SVM+ICA framework. Although b also plays an important role as for classification purpose, it is not as essential as w since it does not affect the direction of the decision surface. Since yi takes on two values is defined for two-class datasets, resulting in W1, l = w where l = 1. The decision boundaries by both LDA and SVM should converge to the optimal when there exists sufficient amount of clean data with unbiased data distribution. Using structural risk as a dimensionality reduction criterion, we expect that mappings from structural risk minimization by SVM are more robust than LDA due to the generalization capability. The weight vector is given by,

$$w = \sum \alpha_i^* y_i x_i \in R^n$$

B. Projection

The projection component, as an intermediate step in the SVM+ICA framework, allows for mapping vectors derived from structural risk minimization and independence maximization to achieve minimum correlation. It does so by projecting the given data X onto the subspace satisfying $W^T_{1,l}\, \mathbf{x} = \mathbf{0}$,yielding the projected data, *Z*, such that the subsequent independence maximization process based on *Z* is least affected or correlated with the previous structural risk minimization process. After the projection procedure, the projected data, *Z*, would lose information along the direction of $W_{1,l,}$ which indicates that decision information through $W_{1,}l$ is no longer valid in the projection subspace. Therefore, the projection guarantees that any mapping vectors from structural risk minimization,$W_{1,l}$,and independence maximization, $Wl_{+1,m}$,are uncorrelated since $W_{l+1,m} \perp W_{1,l}$.The pair-wise orthogonality is also represented by $W_{1,l} \perp W_{l+1,m}$ or equivalently $W^T_{l+1,m} W_{1,l} = \mathbf{0}$. The projection onto the subspace, orthogonal to the decision hyperplane from structural risk minimization, W1,l, is formulated as a constrained optimization problem as follows:

$$z^* = argmin\|x-z\|^2$$

$$\text{s.t}\ W^T_{1,l}\, z = 0$$

where z represents the projected data onto the subspace orthogonal to W1,l and parallel to the decision hyperplane. The projected data Z is given as an input to the independent component analysis for calculating $W_{l+1,\, m.}$

C. Data Independence Maximization

ICA can be considered a variant of projection pursuit. In the formulation of projection pursuit, no data model or assumption about independent components is made. Computing the nongaussian projection pursuit directions, we effectively estimate the independent components. When all the nongaussian directions have been found, all the independent components have been estimated. If the ICA model holds, optimizing the ICA nongaussianity measures produce independent components; if the model does not hold, then what we get are the projection pursuit directions.ICA offers projection which maximizes independence among features with better data representation and it play an important role in classification performance improvement. As unsupervised dimensionality reduction component in the proposed SVM+ICA framework, independence maximization is applied over the projected data, *Z*. Independence maximization searches for a linear non-orthogonal coordinate system whose axes are determined by both the second and higher order statistics of the original data. Since independence maximization is known as a method providing better data representation than other conventional techniques such as PCA, higher classification accuracy is expected, leading to the adoption of independence maximization in the proposed hybrid dimensionality reduction framework.

To find mappings which maximize independence, the approximated negative entropy criterion introduced also referred to as FastICA, due to well-justified statistical theory and computational efficiency. The FastICA algorithm involves two sequential processes, the one unit (weight vector) estimation and the decorrelation among weight vectors. The weight vector from one unit process is given as,

$$w_i^+ = E\{z\, g(w_i^T z)\} - E\{g'(w_i^T z)\}\, w_i$$

where w_i^+ is temporal approximation of independent components with i Є {l+1…m} and g(u) is the derivative of non quadratic function and g (u) = tanh (au) and g' (u) = $a\,sech^2(au)$.

The purpose of decorrelation is to keep different weight vectors from converging to the same maximum. The weight vector is given by,

$$W_{l+1,m} = W^+_{l+1,m}\,[(W^+_{l+1,m}{}^T\, W^+_{l+1,m})^{-1/2}]^T$$

where $W_{l+1,m}$ represents decorrelated mappings based on $W^+_{l+1,m}$ from independence maximization.

D. k-NN ALGORITHM

k-nearest neighbors (k-NN) is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions).The k-nearest neighbor algorithm is a method used for classification and regression in pattern recognition. Dimensionality reduction is performed prior to the k-NN algorithm inorder to avoid the curse of dimensionality effect. This paper uses the k-NN algorithm as classifier and the classification accuracy is computed and it is the performance metric of dimensionality reduction method due to its non-parametric nature. By this classification, several parameters are calculated and it is compared with the existing methods.

V. EXPERIMENTAL RESULTS AND ANALYSIS

We used a dataset named Diffusion Large B cell Lymphoma (DLBCL).This two class dataset contains diagnostic classes such as diffuse large B-cell lymphoma (DLBCL) and Follicular lymphoma (FL). The no of genes in this dataset is 5470 and no of samples are 77. The no of features in this dataset is reduced to 96 for the proposed SVM+ICA method. The kNN Algorithm is used as a classifier and the performance metrics such as classification accuracy, precision, sensitivity and specificity is calculated in Table I and Table II. Experimental results shows that the proposed SVM +ICA method outperforms compared to SVM, ICA. The proposed method has very high accuracy and it is approximately 94%. The accuracy is calculated from the confusion matrix. From the confusion matrix, the other parameters such as specificity, sensitivity are calculated. This method is efficient compared with the existing methods.

Table I

Comparison of SVM+ICA method with existing methods

Method	**Accuracy** (%)	**Sensitivity** (%)	**Specificity** (%)
SVM	75.32	75.22	66.66
ICA	86.23	93.54	65.38

SVM+ICA	94.82	93.54	75

Table II

Performance of proposed method for DLBCL dataset

Method	Positive Predicted Value (%)	Negative Predicted Value (%)
SVM	96.66	17.39
ICA	86.5	80.9
SVM+ICA	92.06	78.9

VI. CONCLUSION

This paper proposed a SVM plus ICA hybrid dimensionality reduction algorithm. The algorithm provides projection that minimizes SVM-based structural risk in supervised manner and maximizes ICA-based data independence in an unsupervised manner. This hybrid approach combines the advantages of both SVM and ICA. SVM+ICA offer projection vectors as a mapping from observations to reduced-dimensional space. The hybrid algorithm gives linear projection to obtain important features for better classification performance. This approach is performed for two class datasets and the experimental results shows that the dimensionality (no of features) is reduced for various two class datasets. This approach is extended for various multiclass datasets and for SVM a one-against-all technique is used.

References

1. Sangwoo Moon and Hairong Qi ,"Hybrid Dimensionality Reduction Method Based on Support Vector Machines and Independent Component Analysis", IEEE Transactions on Neural networks and Learning Systems,vol. 23, no. 5, may 2012.
2. A.M.Martinez and A.C.Kak, "PCA versus LDA,"IEEE Trans.Pattern Anal.Mach.Intell. vol.23, no. 2, pp. 228-233, Feb. 2001.
3. L.-F. Chen, H.-Y. M. Liao, M.-T. Ko, J.-C. Lin, and G.J. Yu, "Anew LDA-based face recognition system which can solve the small sample size problem," Pattern Recognit., vol. 33, no. 10, pp.1713–1726, 2000.
4. H. Park, M. Jeon, and J. B. Rosen, "Lower dimensional representation of text data based on centroids and least squares," BIT Numerical Math vol. 43, no. 2, pp. 427–448, 2003.
5. H. Wold, "Estimation of principal components and related models by iterative least squares," in Multivariate Analysis. New York: Academic, 1966, pp. 391–420.
6. B. Scholkopf, A. Smola, and K. Muller, "Nonlinear component analysis as a Kernel Eigenvalue problem," Neural Comput., vol. 10, no. 5, pp, 1299–1319, 1998.
7. X.Jiang, "Asymmetric Principal component and discriminant analyses for pattern classification," IEEE Trans.Pattern Anal. Mach. Intell., vol. 31, no. 5, pp.931-937, May 2009.
8. K. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," IEEE Trans. Neural Netw, vol. 18, no. 2, pp. 530–541, Mar. 2007.
9. S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," IEEE Trans. Neural Netw. vol. 17, no. 3, May 2006.