Signal Processing 90 (2010) 2314–2318

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

On including sequential dependence in ICA mixture models

Addisson Salazar, Luis Vergara *, Ramón Miralles

Instituto de Telecomunicaciones y Aplicaciones Multimedia, Universidad Politécnica de Valencia, 46022 Valencia, Spain

ARTICLE INFO

Article history:
Received 25 November 2009
Received in revised form
26 January 2010
Accepted 9 February 2010
Available online 11 February 2010

Keywords:
ICA mixtures
Classifiers
Sequential dependence
Hypnograms

ABSTRACT

We present in this communication a procedure to extend ICA mixture models (ICAMM) to the case of having sequential dependence in the feature observation record. We call it sequential ICAMM (SICAMM). We present the algorithm, essentially a sequential Bayes processor, which can be used to sequentially classify the input feature vector among a given set of possible classes. Estimates of the class-transition probabilities are used in conjunction with the classical ICAMM parameters: mixture matrices, centroids and source probability densities. Some simulations are presented to verify the improvement of SICAMM with respect to ICAMM. Moreover a real data case is considered: the computation of hypnograms to help in the diagnosis of sleep disorders. Both simulated and real data analysis suggest the potential interest of including sequential dependence in the implementation of an ICAMM classifier.

1. Introduction

Mixtures of independent components analyzers are progressively recognized as powerful tools for versatile modeling of arbitrary data densities [1–7]. In most cases the final goal is to classify the observed data vector $\mathbf{x}$ (feature) in a given class from a finite set of possible classes. To this aim, the probability of every class given the observed data vector $p[C_k/\mathbf{x}]$ is to be determined. Then the class having maximum probability is selected. Bayes theorem is claimed for the practical computation of the required probabilities, as it allows expressing $p[C_k/\mathbf{x}]$ in terms of the vector observation mass density. Considering K classes we may write

$$p[C_k/\mathbf{x}] = \frac{p[\mathbf{x}/C_k]p[C_k]}{p[\mathbf{x}]} = \frac{p[\mathbf{x}/C_k]p[C_k]}{\sum_{k'=1}^{K} p[\mathbf{x}/C_{k'}]p[C_{k'}]} \tag{1}$$

where the mixture model of $p[\mathbf{x}]$ is evident in the denominator of (1). ICA mixture model (ICAMM) considers that the observations corresponding to a given class k are obtained by linear transformation of vectors having independent components plus a bias term: $\mathbf{x}=\mathbf{A}_k\mathbf{s}_k+\mathbf{b}_k$. Equivalently, this implies that the observation vector in a given class can be expanded around a centroid vector $\mathbf{b}_k$ in a basis formed by the columns of $\mathbf{A}_k$. It is assumed that the basis components are independent so that matrix $\mathbf{A}_k$ is non-singular. When this assumption is becoming invalid, due for example to high dimension of the observation vector, some dimension reduction techniques like classical principal component analysis (PCA) are routinely used. The transforming matrix, the centroid and the marginal probability density functions (which can be arbitrary) of the independent components of $\mathbf{s}_k$ (called sources) define a particular class.

Using standard results from probability theory, we have that $p[\mathbf{x}/C_k] = |\det \mathbf{A}_k^{-1}| p[\mathbf{s}_k]$. On the other hand, algorithms for learning the ICAMM parameters $(\mathbf{A}_k, \mathbf{b}_k, p[\mathbf{s}_k]\ \ k=1\ldots K)$ in supervised or unsupervised frameworks can be found in the given Refs. [1–7]. Therefore, if the classifier has been trained, we can compute the required probabilities using

$$p[C_k/\mathbf{x}] = \frac{|\det \mathbf{A}_k^{-1}| p[\mathbf{s}_k] p[C_k]}{\sum_{k'=1}^{K} |\det \mathbf{A}_{k'}^{-1}| p[\mathbf{s}_{k'}] p[C_{k'}]} \quad \mathbf{s}_k = \mathbf{A}_k^{-1}(\mathbf{x}-\mathbf{b}_k). \tag{2}$$

However, very often, the classes and observations do not appear in a totally random manner, but they exhibit some degree of sequential dependence in time or space domains. This means that the computation of the class

* Corresponding author. Tel.: +34 96 3877308; fax: +34 96 3877919.
E-mail address: lvergara@dcom.upv.es (L. Vergara).

0165-1684/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2010.02.010

probabilities should consider the whole history of observations. Thus, if we define the indexed matrix of observations $\mathbf{X}(n) \equiv [\mathbf{x}(0)\ \mathbf{x}(1) \ldots \mathbf{x}(n)]$ we should compute $p[C_k(n)/\mathbf{X}(n)]$.

We present in this communication (Section 2) a procedure to extend ICAMM to the case of sequential dependence. We call it sequential ICAMM (SICAMM). In Section 3 some experiments are included with simulated and real data, showing that the classification error percentage can be reduced by SICAMM, in comparison with ICAMM.

2. Sequential ICAMM

To compute $p[C_k(n)/\mathbf{X}(n)]$ we start from (1). We assume that, conditional on $C_k(m)$, the observed vectors $\mathbf{x}(m) m = 0 \ldots n$ are independent. This is a key assumption in the classical Hidden Markov Model (HMM) [8] structure which is described in Fig. 1. Statistical dependences between two successive instants are defined by the arrows connecting the successive classes. However, successive observed vectors are not directly connected, i.e., the distribution of every $\mathbf{x}(m)$ is totally defined if we know the corresponding class $C_k(m)$. In particular, this implies that $p[\mathbf{X}(n)/C_k(n)] = p[\mathbf{x}(n)/C_k(n)] \cdot p[\mathbf{X}(n-1)/C_k(n)]$. Using this property and the Bayes's rule, we may write

$$\begin{aligned} p[C_k(n)/\mathbf{X}(n)] &= \frac{p[\mathbf{X}(n)/C_k(n)]p[C_k(n)]}{\sum_{k'=1}^{K} p[\mathbf{X}(n)/C_{k'}(n)]p[C_{k'}(n)]} \\ &= \frac{p[\mathbf{x}(n)/C_k(n)]p[\mathbf{X}(n-1)/C_k(n)]p[C_k(n)]}{\sum_{k'=1}^{K} p[\mathbf{x}(n)/C_{k'}(n)]p[\mathbf{X}(n-1)/C_{k'}(n)]p[C_{k'}(n)]} \\ &= \frac{p[\mathbf{x}(n)/C_k(n)]p[C_k(n)/\mathbf{X}(n-1)]p[\mathbf{X}(n-1)]}{\sum_{k'=1}^{K} p[\mathbf{x}(n)/C_{k'}(n)]p[C_{k'}(n)/\mathbf{X}(n-1)]p[\mathbf{X}(n-1)]} \\ &= \frac{|\det \mathbf{A}_k^{-1}| p[\mathbf{s}_k(n)]p[C_k(n)/\mathbf{X}(n-1)]}{\sum_{k'=1}^{K} |\det \mathbf{A}_{k'}^{-1}| p[\mathbf{s}_{k'}(n)]p[C_{k'}(n)/\mathbf{X}(n-1)]} \end{aligned} \tag{3}$$

where, considering the implicit HMM

$$p[C_k(n)/\mathbf{X}(n-1)] = \sum_{k'=1}^{K} p[C_k(n)/C_{k'}(n-1)] \cdot p[C_{k'}(n-1)/\mathbf{X}(n-1)]. \tag{4}$$

Notice that, although rather artificially, (4) holds even for the case of sequential class independence, in which case we may write $p[C_k(n)/C_{k'}(n-1)] = p[C_k(n)]$ and then

$$p[C_k(n)/\mathbf{X}(n-1)] = \sum_{k'=1}^{K} p[C_k(n)] \cdot p[C_{k'}(n-1)/\mathbf{X}(n-1)] = p[C_k(n)].$$

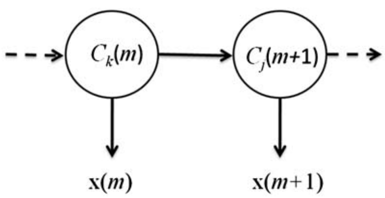

Fig. 1. HMM description of the class-conditionally independence between successive observation vectors.

Thus, using (3) and (4), $p[C_k(n)/\mathbf{X}(n)]$ can be computed from the class transition probabilities $p[C_k(n)/C_{k'}(n-1)]$ and from the last estimates available of the class probabilities $p[C_{k'}(n-1)/\mathbf{X}(n-1)]$.

Therefore a sequential algorithm is proposed to be called SICAMM. Let us describe the algorithm in a more specific form. We assume that the parameters $\mathbf{A}_k, \mathbf{b}_k, p[\mathbf{s}_k]$ $k = 1 \ldots K$ have been previously estimated by means of an ICAMM learning algorithm among the several available in the literature and that the class-transition probabilities are also known or estimated. The following describes the algorithm:

Initialization $n=0$

$$\mathbf{X}(0) = [\mathbf{x}(0)]$$

$$\mathbf{s}_k(0) = \mathbf{A}_k^{-1}(\mathbf{x}(0) - \mathbf{b}_k) \quad k = 1 \ldots K;$$

$$p[C_k(0)/\mathbf{X}(0)] = \frac{|\det \mathbf{A}_k^{-1}| p[\mathbf{s}_k(0)]}{\sum_{k'=1}^{K} |\det \mathbf{A}_{k'}^{-1}| p[\mathbf{s}_{k'}(0)]}$$

For $n=1$ to N

$$\mathbf{X}(n) = [\mathbf{x}(0)\ \mathbf{x}(1) \ldots \mathbf{x}(n)]$$

$$\mathbf{s}_k(n) = \mathbf{A}_k^{-1}(\mathbf{x}(n) - \mathbf{b}_k) \quad k = 1 \ldots K$$

$$p[C_k(n)/\mathbf{X}(n-1)] = \sum_{k'=1}^{K} p[C_k(n)/C_{k'}(n-1)] \cdot p[C_{k'}(n-1)/\mathbf{X}(n-1)]$$

$$p[C_k(n)/\mathbf{X}(n)] = \frac{|\det \mathbf{A}_k^{-1}| p[\mathbf{s}_k(n)]p[C_k(n)/\mathbf{X}(n-1)]}{\sum_{k'=1}^{K} |\det \mathbf{A}_{k'}^{-1}| p[\mathbf{s}_{k'}(n)]p[C_{k'}(n)/\mathbf{X}(n-1)]}$$

Notice that the SICAMM algorithm can be expressed in the form of a sequential Bayesian processor [8]

$$p[C_k(n)/\mathbf{X}(n)] = W_k(n) \cdot p[C_k(n)/\mathbf{X}(n-1)] \quad W_k(n) = \frac{p[\mathbf{x}(n)/C_k(n)]}{p[\mathbf{x}(n)/\mathbf{X}(n-1)]}, \tag{5}$$

where $p[C_k(n)/\mathbf{X}(n-1)]$ is a "prediction" of the current class given the pass history of observations and $W_k(n)$ is an "updating weight" which measures the significance of the current class relative to the significance of the pass history of observations in generating the current observation.

3. Experiments

3.1. Simulations

We have considered a simple scenario similar to the first example included in the classical ICAMM Ref. [1]. Observations are vectors of dimension 2, and the number of classes is also 2. In class 1 the observation vectors are obtained by linearly transforming independent component vectors where both components are obtained from uniform distributions having zero mean and unit variance. The same in class 2, but the distributions are zero mean and unit variance Laplacian. The centroids were selected relatively close, thus $\mathbf{b}_1 = [1\ \ 1]^T$ and $\mathbf{b}_2 = [1.5\ \ 1.5]^T$. We have compared the error percentages in classifying an