A NEURAL NETWORK FOR BLIND ACOUSTIC SIGNAL SEPARATION

B. Freisleben\textsuperscript{1}, C. Hagen\textsuperscript{2} and M. Borschbach\textsuperscript{1}

\textsuperscript{1}Department of Electrical Engineering and Computer Science (FB12), University of Siegen
Hölderlinstr. 3, D-57068 Siegen, Germany
Tel: (49)-271-7403268, Fax: (49)-271-7402532
E-Mail: freisleb@informatik.uni-siegen.de

\textsuperscript{2}Department of Computer Science (FB20), University of Darmstadt
Julius-Reiber-Str. 17, D–64293 Darmstadt, Germany
Tel: (49)-6151-807180, Fax: (49)-6151-807190
E-Mail: hagen@iti.informatik.th-darmstadt.de

Abstract

In this paper, a two–layer neural network is presented that organizes itself to perform blind source separation. A nonlinear learning rule is proposed which allows to extract the unknown independent source signals out of their linear mixtures. The convergence behaviour of the network is investigated empirically. It is demonstrated that the network is capable of performing blind source separation for a set of acoustic signals, including sounds of a train, a rooster, a chirp and laughter.
1. Introduction

The goal of blind source separation (BSS) [3] is to extract the individual, unobservable, statistically independent source signals out of given linear noisy mixtures of them. BSS is performed as part of independent component analysis (ICA) [2], which is a novel signal processing technique useful in a variety of applications areas, such as antenna array processing, communications, speech processing and medical image processing.

From the mathematical point of view, BSS is a method to find a separation matrix which transforms the mixture signals into statistically independent signals being good estimations of the original source signals. There are some numerical algorithms [2, 4] using higher order statistics for estimating the separation matrix. Recently, several neural networks have been proposed in which both unsupervised and supervised learning rules are used (see [5] for a survey) to perform BSS. The unsupervised neural approaches are based on the observation that introducing a nonlinear function in well known neural networks for principal component analysis (PCA) [6, 7] allows to perform BSS if the input vector and the distribution of the source signals meet some conditions, in order to deal with the higher order moments required.

In this paper, an unsupervised BSS learning rule for a two-layer neural network is presented. The learning rule is hierarchical in the sense that output unit $i$ receives contributions from all units $j$ with $j < i$. The convergence behaviour of the neural network is demonstrated in practice. It will be shown that the network is capable of performing BSS for a set of acoustic source signals, including sounds of a train, a rooster, a chirp and laughter.

The paper is organized as follows. Section 2 briefly reviews the mathematical background of BSS and ICA. Section 3 presents the neural network for performing BSS. In section 4 simulation results are presented. Section 5 concludes the paper and outlines areas for further research.

2. Blind Source Separation

Fig. 1 illustrates the steps of a complete ICA algorithm.

The first box represents the unobservable (linear) transformation of the unknown source signal vector $\mathbf{s}$ into the observed signal vector $\mathbf{u} = \mathbf{M} \cdot \mathbf{s}$. The $m \times n$-mixing matrix $\mathbf{M}$ is assumed to have full column rank, i.e. $\mathbf{u}$ has the dimension $m$, with $m \geq n$.

The second box represents a prewhitening step which transforms $\mathbf{u}$ into $\mathbf{x} = \mathbf{D}^{-1} \cdot \mathbf{F}^T \cdot \mathbf{u}$, such that $\mathbf{x}$ has as its covariance matrix $\mathbf{C}_{xx} = E(\mathbf{x}\mathbf{x}^T)$ the identity matrix $\mathbf{I}_n$. 
Figure 1: Illustration of the ICA Algorithm

Prewitening simplifies the BSS task and can be achieved by any non-neural or neural PCA algorithm; the columns of the matrix $F$ are the $n$ principal eigenvectors of $C_{xx}$, and the diagonal matrix contains the square roots of the $n$ largest eigenvalues.

The third (dashed) box represents the BSS part: finding an (orthogonal) separation matrix $B$ to transform $x$ into an output vector $y$ whose components have a maximal degree of statistical independence (measured by so called contrast functions [2]). Each output $y_i$, $i = 1, \cdots, n$ can be regarded as an approximation of one of the source signals $\pm s_j$, $j = 1, \cdots, n$.

The fourth box represents the last step missing to complete ICA, namely the estimation of the mixing matrix, the so called ICA matrix $H$, whose columns form the ICA basis. The output of the last step is the estimated observed mixture vector $\hat{u} = H \cdot y$.

It should be mentioned that all BSS algorithms only work under the assumption that the distributions of the source signals are non-Gaussians, except at most one, because a linear mixture of Gaussians is again a Gaussian and cannot be separated. In many practical situations, the source signals are either sub-Gaussians or super-Gaussians, i.e. distributions with densities flatter or sharper than that of Gaussians, respectively. For example, acoustic signals are often super-Gaussians [6, 7]. Therefore, the existing approaches proposed for performing BSS can only be applied either to sub-Gaussians or to super-Gaussians.

3. A Neural Network for BSS

The neural network proposed for performing BSS consists of an input and an output layer, each consisting of $n$ units. Both layers are fully connected to each other.

The connection weights are the $n$-dimensional vectors $w_i$, $i = 1, \cdots, n$ which form the columns of the $n \times n$ matrix $W$. The output of the $i$-th output unit $y_i$, $i = 1, \cdots, n$
is
\[ y_i = w_i^T x \quad \text{or} \quad y = W^T x \] (1)
where \( x = (x_1, \ldots, x_n)^T \) is the input vector and \( y = (y_1, \ldots, y_n)^T \) is the output vector.

The weight vectors \( w_i, i = 1, 2, \ldots, n \) are updated according to the learning rule
\[
\dot{w}_i(t + 1) = w_i(t) + \alpha(t) (w_i(t)^T x(t))^3 \left( I - \sum_{j=1}^{i-1} w_j(t) w_j(t)^T \right) x(t) \\
w_i(t + 1) = \frac{\dot{w}_i(t + 1)}{|\dot{w}_i(t + 1)|} \tag{2} 
\]

In addition to the general assumptions of ICA we require for this network to perform BSS that the input vector \( x \) is already prewhitened and that the source signals have super-Gaussian distributions.

4. Experimental Results

The network has been implemented in C on a DEC Alpha workstation [1] under Digital UNIX, and a number of simulations have been carried out to investigate its behaviour in practice.

In the BSS example presented in the following, 4 independent sound signals, sampled at a rate of 8 kHz were used (see Fig. 2, column (a)). The signals represent sounds of a chirp, a train, a laughter, and a rooster (from top to bottom), they are stationary and have super-Gaussian distributions.

The 4-dimensional input vectors for the neural network, which represent 12912 samples of the linear mixtures, were constructed by the orthogonal mixing matrix \( M \) shown in Fig. 3. Since \( M \) has been chosen to be orthogonal, a prewhitening step is not necessary. The mixtures are shown in column (b) of Fig. 2.

In the training mode, the set of 12912 input vectors was presented 3 times to the network. The weights of the connections were initially set to random values taken from the interval \([-1, 1]\). An initial learning rate \( \alpha(0) = 0.8 \) (see equation (2)) has been used and successively decreased after each simulation step according to \( \alpha(1000) = 0.005 \cdot \alpha(0) \).

The convergence behaviour is illustrated in Fig. 3, where the x-axis indicates the number of simulation steps (a simulation step is equivalent to updating the set of weights in response to the presentation of an input vector), and the y-axis indicates the error \( d \) resulting after each simulation step. The error \( d \) represents the distance of the weight matrix to an idealized separation matrix with no difference between the separated and the original signals; it has been introduced in [2].
Figure 2: Source Signals (a), Linear Mixtures (b), and Separated Signals (c)

\[
\begin{pmatrix}
0.11311 & 0.134005 & -0.934862 & -0.308678 \\
0.726009 & 0.085785 & 0.302136 & -0.611772 \\
-0.327223 & -0.803917 & 0.009125 & -0.496547 \\
-0.594173 & 0.573062 & 0.186182 & -0.532817
\end{pmatrix}
\]

Figure 3: Mixing Matrix \( \mathbf{M} \) (left) and Convergence Behaviour (right)
After the weight vectors have converged to the columns of the separating matrix, the network is capable of separating the original source signals from the mixtures used as the input, as shown in column (c) of Fig. 2. The acoustic quality of the separated signals is quite good. It is impossible for humans to distinguish them from the original signals.

5. Conclusions

In this paper we have presented a self-organizing neural network which performs blind acoustic signal separation. Assuming that the input vectors are prewhitened and the source signals are all super-Gaussians, the convergence behaviour of the network was demonstrated by simulation results.

There are several areas for future research, such as investigating the suitability of using other network architectures or nonlinear unsupervised learning rules for performing BSS, and finding methods for separating signals with both sub- and super-Gaussian distributions.

References


