Snowbird, Utah
April 6-9 1999

Nonlinear Blind Source Separation Using Ensemble Learning

Harri Lappalainen[1], Xavier Giannakopoulos[2], and Juha Karhunen[2]

Helsinki University of Technology
[1]Neural Networks Research Centre
[2]Laboratory of Computer and Information Science
P.O.Box 5400, FIN-02015 HUT, Finland
email: Harri.Lappalainen@hut.fi, Juha.Karhunen@hut.fi
http://www.cis.hut.fi

Blind source separation (BSS) and the closely related Independent Component Analysis (ICA) have recently drawn a lot of attention in unsupervised neural learning and statistical signal processing [1, 2, 3]. In the basic linear case, ICA and BSS use the same data model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^{m} s_i(t)\mathbf{a}_i. \tag{1}$$

Here $\mathbf{x}(t)$ is the n-dimensional t:th data vector, $s_1(t), \ldots, s_m(t)$ are the respective m unknown source signals or independent components, and $\mathbf{A}$ is the unknown $n \times m$ mixing matrix. The ICA expansion (1) provides the solution also to the linear BSS problem. This is determined blindly from the data by employing the strong but plausible assumption that the source signals or independent components $s_i(t)$ in (1) must be mutually statistically independent (or as independent as possible). This problem and neural methods for solving it are discussed in more detail for example in [2, 3].

In this work, we consider the respective noisy BSS (and ICA) problem for the general nonlinear data model

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t) \tag{2}$$

where $\mathbf{f}$ is an unknown real-valued n-component vector function and $\mathbf{n}(t)$ a Gaussian noise vector. The basic ICA problem is now to estimate the realizations of the independent components $s_i(t)$ using only observations of the mixtures $x_j(t)$. Usually there exists infinitely many solutions to this nonlinear ICA problem, because the indeterminacies are much more severe than in the linear case [4]. If an ICA solution is found for the nonlinear data model (2), it may be quite different from the solution of the respective nonlinear BSS problem, where one should find the original source signals $\mathbf{s}(t)$.

For solving the general BSS problem described by the nonlinear data model (2), only a few methods have been proposed. In particular, we considered the use of modified self-organizing map (SOM) in [5]. However, this approach is limited mainly to separation of sources having distributions not too far from the uniform distribution. In another paper [6], we used a modification of generative topographic mapping (GTM) for the nonlinear BSS problem. This approach is more general, but requires the knowledge of the source densities. Generally, some regularization procedure must be used for obtaining meaningful results, which are usually at most of fair quality due to the difficulty of the problem. Furthermore, all general nonlinear BSS approaches proposed thus far suffer from the curse of dimensionality.

In this work, we present first results on applying a recently developed, computationally efficient Bayesian method called ensemble learning to nonlinear blind source separation. Ensemble learning, described for example in [7, 8, 9], is an approximation for exact Bayesian analysis. In general, the posterior probability density function (pdf) contains all the relevant information for a specific estimation problem. However, it is a complex high-dimensional function which is difficult to treat exactly. In ensemble learning, the posterior pdf is approximated by a simpler function, and the Kullback-Leibler divergence is then used as the criterion for minimizing the misfit between the actual posterior pdf and its parametric approximation. Generally, ensemble learning yields reliable estimates, avoiding possible problems due to overfitting which may appear in context with maximum a posteriori (MAP) methods.

In a very recent paper [10], we have combined ensemble learning and independent component analysis using a mixture-of-Gaussian model for sources. The resulting new method was applied to the standard linear ICA/BSS problem with good results. This approach resembles in some aspects closely the work done in [11], but is more general.

Here we extend the ensemble learning method described in detail in [10] to the nonlinear data model (2). The data generating mapping (2) is modeled using the well-known multilayer perceptron structure, allowing in principle a reasonably good approximation of any nonlinear mapping. The inverse separating mapping is then found using gradient descent. The first experimental results are encouraging, showing that it is possible to separate the sources for moderately complicated nonlinear mappings without any prior knowledge on the number of sources and their distributions. Basically, the proposed method tries to find out the ICA solution which has most probably generated the data. For mappings which are not too complicated, this usually turns out to be the nonlinear BSS solution.

References

[1] J.-F. Cardoso, C. Jutten, and P. Loubaton (Eds.), *ICA'99 - First Int. Workshop on Independent Component Analysis and Signal Separation*, Aussois, France, January 11-15, 1999.

[2] T.-W. Lee, *Independent Component Analysis - Theory and Applications*, Kluwer, 1998.

[3] E. Oja, J. Karhunen, A. Hyvärinen, R. Vigario, and J. Hurri, "Neural independent component analysis - approaches and applications," in S.-I. Amari and N. Kasabov (Eds.), *Brain-Like Computing and Intelligent Information Systems*, Springer-Verlag, Singapore, 1997, pp. 167-188.

[4] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: existence and uniqueness results", to appear in *Neural Networks*, vol. 12, no. 2, 1999, pp. 209-219.

[5] P. Pajunen, A. Hyvärinen, and J. Karhunen, "Nonlinear blind source separation by self-organizing maps," in S.-I. Amari et al. (Eds.) *Progress in Neural Information Processing (ICONIP-96)*, Springer-Verlag 1996, pp. 1207-1210.

[6] P. Pajunen and J. Karhunen, "A maximum likelihood approach to nonlinear blind source separation," in *Proc. Int. Conf. on Artificial Neural Information Networks (ICANN'97)*, Lausanne, Switzerland, October 1997, pp. 1207-1210.

[7] G. Hinton and D. van Camp, "Keeping neural networks simple by minimizing the description length of the weights," in *Proc. of the COLT'93*, Santa Cruz, California, 1993, pp. 5-13.

[8] D. Barber and C. Bishop, "Ensemble learning for multi-layer networks," in M. Jordan et al. (Eds.), *Advances in Neural Information Processing Systems 10*, MIT Press, 1998, pp. 395-401.

[9] C. Bishop, N. Lawrence, T. Jaakkola, and M. Jordan, "Approximating posterior distributions in belief networks using mixtures," in M. Jordan et al. (Eds.), *Advances in Neural Information Processing Systems 10*, MIT Press, 1998, pp. 416-422.

[10] H. Lappalainen, "Ensemble learning for independent component analysis," in *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France, January 11-15, 1999, pp. 7-12.

[11] H. Attias, "Independent factor analysis," submitted to *Neural Computation*, 1998.

Topic: learning algorithms (or secondarily learning theory).
Preference: both oral and poster are acceptable.
The paper will be represented by Prof. **Juha Karhunen**.