Independent Component Analysis: Algorithms and Applications

E. Oja[1]

[1] Helsinki University of Technology, Department of Computer Science and Engineering, P.O.Box 5400, FIN-02015 HUT, Finland

Keywords: Independent component analysis, Latent variable models.

1 What is ICA?

Independent Component Analysis (ICA); see Hyvärinen, Karhunen, and Oja (2001) and Cichocki and Amari (2002) is a novel statistical signal and data analysis method. It assumes a statistical model whereby the observed multivariate data $\mathbf{x}$, typically given as a large database of samples, are assumed to be linear or nonlinear mixtures of some unknown latent variables. In the linear noise-free case the model is simply

$$\mathbf{x} = \mathbf{A}\mathbf{s}.$$

The mixing coefficients or elements of matrix $\mathbf{A}$ are also unknown. The latent variables s_i or elements of vector $\mathbf{s}$ are nongaussian and mutually independent, and they are called the independent components of the observed data. By ICA, these independent components, also called sources or factors, can be found. This problem is also called blind source separation because almost everything is unknown. Many practical algorithms exist such as the FastICA algorithm (Hyvärinen and Oja (1997); Hyvärinen (1999)), which is generally considered to be one of the best ones of the existing algorithms and is widely used by practitioners all over the world.

2 Variations

Because almost nothing is assumed known except the sample data itself, ICA is an ideal technique for exploratory data analysis and data mining. However, in many cases there exists prior information on the mixing model or the sources, and it should be taken into account. This is called the semi-blind situation. Särelä and Valpola (2005) developed recently a novel general framework for semi-blind source separation algorithms called Denoising Source Separation (DSS). In this framework, source separation algorithms are constructed around a denoising procedure. The denoising allows for easy and practical guidance of the separation by the prior knowledge that is available.

3 Applications

A prominent application for ICA and DSS is biomedical signal and image analysis. Another signal processing application is in telecommunications, separation of CDMA signals. But this set of methods can be used for many other kinds of data, too; we have started work on applying ICA also on text document data and genomics data in bioinformatics. Another application is multimodal data containing text, images, and Web links, in which ICA enhances the well-known Latent semantic indexing methodology. Yet another application domain is spatiotemporal data, for example long-time measurements of climate factors over the globe. In general, ICA is one of the new tools to be used in linear vector space representations of a wide range of data types, made possible by recent kernel expansions.

In the talk, the basics of ICA, FastICA, and DSS are covered and illustrated by two novel applications: biomedical image analysis and climate data analysis.

References

A. Cichocki and S. Amari (2002). *Adaptive Blind Signal and Image Processing*. J. Wiley, New York.

A. Hyvärinen, J. Karhunen and E. Oja (2001). *Independent Component Analysis*. J. Wiley, New York.

A. Hyvärinen and E. Oja (1997). A fast fixed-point algorithm for independent component analysis. *Neural Comp.*, 9, 1483–1492.

A. Hyvärinen (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks*, 10, 626–634.

J. Särelä and H. Valpola (2005). Denoising source separation. *J. Machine Learning Res.*, in press.