Techniques & Applications

Independent component analysis: an introduction

James V. Stone

Independent component analysis (ICA) is a method for automatically identifying the underlying factors in a given data set. This rapidly evolving technique is currently finding applications in analysis of biomedical signals (e.g. ERP, EEG, fMRI, optical imaging), and in models of visual receptive fields and separation of speech signals. This article illustrates these applications, and provides an informal introduction to ICA.

Independent component analysis (ICA) is essentially a method for extracting individual signals from mixtures of signals. Its power resides in the physically realistic assumption that different physical processes generate unrelated signals. The simple and generic nature of this assumption ensures that ICA is being successfully applied in a diverse range of research fields.

Despite its wide range of applicability, ICA can be understood in terms of the classic 'cocktail party' problem, which ICA solves in an ingenious manner. Consider a cocktail party where many people are talking at the same time. If a microphone is present then its output is a mixture of voices. When given such a mixture, ICA identifies those individual signal components of the mixture that are unrelated. Given that the only unrelated signal components within the signal mixture are the voices of different people, this is precisely what ICA finds. (In practice, ICA requires more than one simultaneously recorded mixture in order to find the individual signals in any one mixture.) It is worth stressing here that ICA does not incorporate any knowledge specific to speech signals; in order to work, it requires simply that the individual voice signals are unrelated.

On a more biological note, an EEG signal from a single scalp electrode is a mixture of signals from different brain regions. As with the speech example above, the signal recorded at each electrode is a mixture, but it is the individual components of the signal mixtures that are of interest (e.g. single voice, signal from a single brain region). Finding these underlying 'source' signals automatically is called 'blind source separation' (BSS), and ICA the dominant method for performing BSS. A critical caveat is that most BSS methods require at least as many mixtures (e.g. microphones, electrodes) as there are source signals.

ICA in context: related methods

The goal of decomposing measured signals, or variables, into a set of underlying variables is far from new (we use the terms 'variable' and 'signal' interchangeably here). For example, the literature on IQ assessment describes many methods for taking a set of measured variables (i.e. sub-test scores) and finding a set of underlying competences (e.g. spatial reasoning). In the language of BSS, this amounts to decomposing a set of signal mixtures (sub-test scores) into a set of source signals (underlying competences). In common with the IQ literature, many fields of research involve identifying a few key source signals from a large number of signal mixtures. Techniques commonly used for this data reduction (or data mining, as it is now known) are principal component analysis (PCA), factor analysis (FA), linear dynamical systems (LDS).

The most commonly used data reduction methods (PCA and FA) identify underlying variables that are uncorrelated with each other. Intuitively, this is desirable because the underlying variables that account for a set of measured variables should correspond to physically different processes, which, in turn, should have outputs that are uncorrelated with each other. However, specifying that underlying variables should be uncorrelated imposes quite weak constraints on the form these variables take. It is the weakness of these constraints which ensures that the factors extracted by FA can be rotated (in order to find a more interpretable set of factors) without affecting the (zero) correlations between factors. Most factor rotation methods yield a statistically equivalent set of uncorrelated factors. The variety of factor rotation methods available is regarded with some skepticism by some researchers. This is because it is sometimes possible to use factor rotation to obtain new factors which are easily interpreted, but which are statistically no more significant than results obtained with other factor rotations. By contrast, any attempt to rotate the factors ('independent components') extracted by ICA would yield non-independent factors. Thus, the independent components of ICA do not permit post-ICA rotations because such factors are statistically independent, and are therefore uniquely defined. Independent components can also be obtained by making use of the observation that individual source signals tend to be less complex than any mixture of those source signals [1,2].

Statistical independence

ICA is based on the assumption that source signals are not only uncorrelated, but are also 'statistically independent'. Essentially, if two variables are independent then the value of one variable provides absolutely no information about the value of the other variable. By contrast, even though two variables are uncorrelated, the value of one variable can still provide information about the value of the other variable (see Box 1). ICA seeks a set of statistically independent signals amongst a set of 'signal mixtures', on the assumption that such statistically independent signals are derived from different physical processes (Box 2). The objective of finding such a set of statistically independent signals is achieved by maximizing a measure of the 'joint entropy' of the extracted signals.

What is it good for?

ICA has been applied in two fields of research relevant to cognitive science: analysis of biomedical data and computational modelling. One of the earliest biomedical applications of ICA

 1364-6613/02/$ – see front matter PII: S1364-6613(00)01813-1

Box 1. Independent and uncorrelated variables

If two variables (signals) x and y are related then we usually expect that knowing the value of x tells us something about the corresponding value of y. For example, if x is a person's height and y is their weight then knowing the value of x provides some information about y. Here, we consider how much information x conveys about y when these variables are uncorrelated and independent.

Uncorrelated variables

Even if x and y are uncorrelated then knowing the value of x can still provide information about about y. For example, if we define $x = \sin(z)$ and $y = \cos(z)$ (where $z = 0 \ldots 2\pi$) then x and y are uncorrelated (Fig. Ia; note that noise has been added for display purposes). However, the variables $x^2 = \sin^2(z)$ and $y^2 = \cos^2(z)$ are (negatively) correlated; as shown in Fig. Ib, which is a graph of x^2 versus y^2. Thus, knowing the value of x^2 (and therefore x) provides information about y^2 (and therefore about y), even though x and y are uncorrelated. For example, in Fig. Ia, if $x = 0.5$ then it can seen that either $y \approx -0.7$ or $y \approx 0.7$; so that knowing x provides information about y.

Correlated variables

If two variables are correlated then knowing the value of one variable provides information about the corresponding value value of the other variable. For example, the variables in Fig. Ib are negatively correlated ($r = -0.962$), and if the x-axis variable is 0.4 then it can seen that the corresponding y-axis variable is approximately 0.3.

Independent variables

If two signals are independent then knowing the value of one signal provides absolutely no information about the corresponding value of the other signal. For example, if two people are speaking at the same time then knowing the amplitude of one voice at any given moment provides no information about the value of the other voice at that moment. In Fig. Ic, each point represents the amplitudes of two voices at a single moment in time; knowing the amplitude (x value) of one voice provides no information about the amplitude (y value) of the other voice.

Maximum entropy distributions

If two signals are plotted against each other (as x and y in Fig. I) then this approximates the joint 'probability density function' (pdf) of the signals. For signals with bounded values (e.g. between 0 and 1), this joint pdf has 'maximum entropy' if it is uniform (as in Fig. Id). Note that if a set of signals has a maximum entropy pdf then this implies that the signals are mutually independent, but that a set of independent signals does not necessarily have a pdf with maximum entropy (e.g. Fig. Ic).

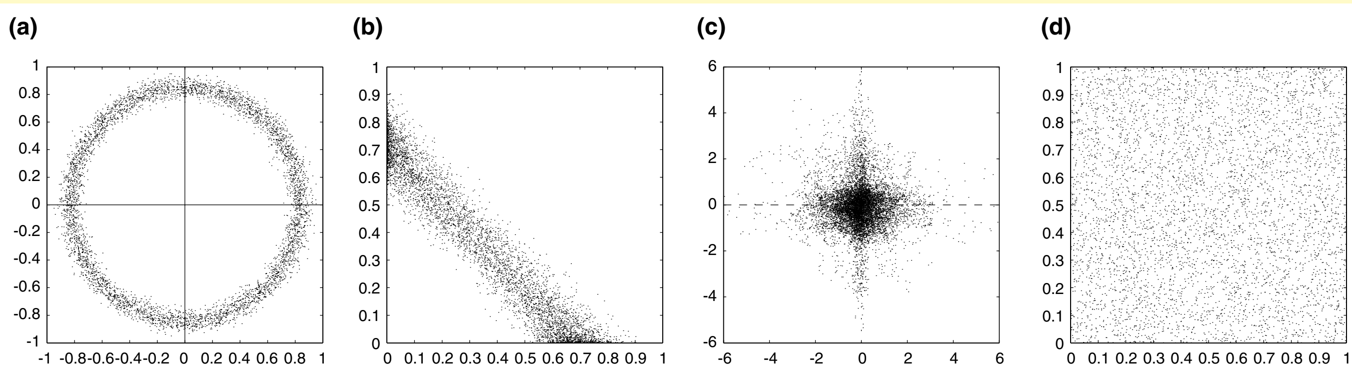

Fig. I

TRENDS in Cognitive Sciences

Box 2. ICA in a Nutshell

The general strategy underlying ICA can be summarized as follows.

(1) It is assumed that different physical processes (e.g. two speakers) give rise to unrelated source signals. Specifically, source signals are assumed to be statistically independent (see Box 1).

(2) A measured signal (e.g. a microphone output) usually contains contributions from many different physical sources, and therefore consists of a mixture of unrelated source signals. [Note that most ICA methods require at least as many simultaneously recorded signal mixtures (e.g. microphone outputs) as there are signal sources (e.g. voices)].

(3) Unrelated signals are usually statistically independent, and it can be shown that a function g of independent signals have 'maximum entropy' (see Box 3). Therefore, if a set of signals with maximum entropy can be recovered from a set of mixtures then such signals are independent.

(4) In practice, independent signals are recovered from a set of mixtures by adjusting a separating matrix W until the entropy of a fixed function g of signals recovered by W is maximized [where g is assumed to be the cumulative density function (cdf) of the source signals, see Box 3]. The independence of signals recovered by W is therefore achieved indirectly, by adjusting W in order to maximize the entropy of a function g of signals recovered by W; as maximum entropy signals are independent, it can be shown that this ensures the estimated source signals recovered by W are also independent (see Boxes 1 and 3).

Box 3. Analysis of biomedical data

EEG

ICA has been used to recover ERP temporal independent components (tICs) associated with detection of visual targets [a]. In this case, each electrode output is a temporal mixture (analagous to a microphone output in Box 4, Fig. Ia). The signal recorded at each electrode is a mixture of tICs, and temporal ICA (tICA) is used to recover estimates of these temporal independent components.

fMRI

ICA was used to recover spatial independent components (sICs) from fMRI data [b]. In this case, the fMRI image recorded at a given time is a mixture (analagous to a face image in the middle column of Box 4, Fig. Ib). The fMRI image recorded at each time point is a mixture of sICs, and spatial ICA (sICA) is used to recover estimates of these sICs (Fig. I).

Optical imaging

Recently, ICA has been used in conjunction with optical recording from the brain of a sea slug *Tritionia diomeda* [c]. A 448 photodiode-detector array recorded neural activity over an 8-second time period (see Fig. II). The close proximity between neurons ensured that each detector recorded a mixture of action potentials from many neurons. Temporal ICA (tICA) was used to decompose the outputs of detectors into a set of temporally independent components (tICs).

References

a Makeig, S. *et al.* (1997) Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10979–10984

b McKeown, M. *et al.* (1998) Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop color-naming task. *Proc. Natl. Acad. Sci. U. S. A.* 95, 803–810

c Brown, G. *et al.* (2001) Independent components analysis at the neural cocktail party. *Trends Neurosci.* 24, 54–63

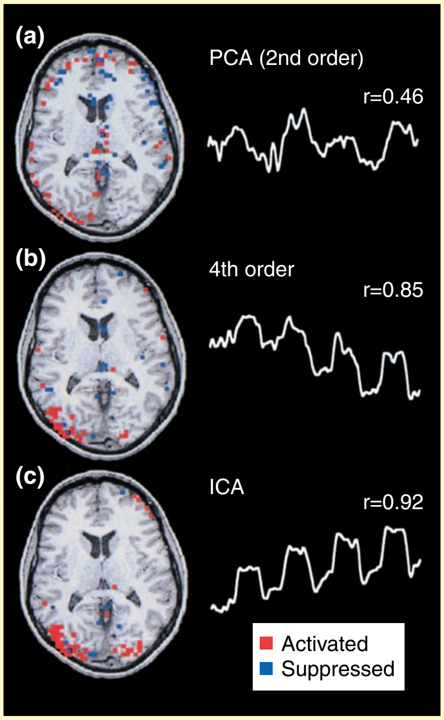

Fig. I. (a) Brain activation recovered by principal component analysis (PCA), and its associated time course. (b) Brain activation recovered by a fourth order projection pursuit method, and its associated time course. (c) Brain activation recovered by ICA, and its associated time course. The correlation (r) between each extracted time course and the temporal sequence of visual stimulation used is indicated in each case, and of these three methods, it can be seen that ICA gives the highest value of r. (Reproduced, with permission, from Ref. b.)

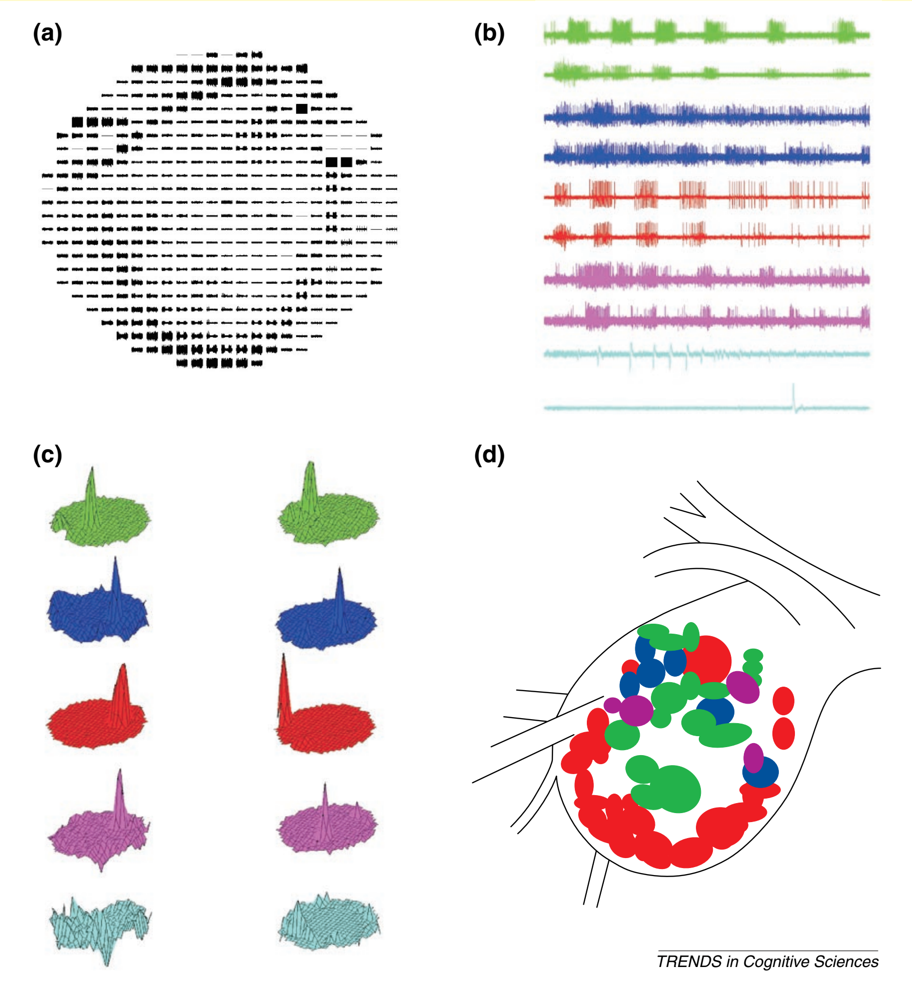

Fig. II. (a) Mean 448-pixel image from array of 448 photodiode-detector array over 8 seconds (sampling rate 1kHz). (b) Temporal independent components (tICs) were classified into five colour-coded classes. Each tIC class identifies a set of neurons with similar firing patterns. One tIC class (light blue, at bottom) identifies signal artifacts. (c) The relative contribution of each detector to each colour-coded tIC, as denoted by height. Only tICs corresponding to single neurons defined spatially localized clusters. By contrast, tICs associated with signal artifacts defined spatially distributed detectors (bottom). (d) Spatial positions of neurons associated with each colour-coded tIC class. (Reproduced, with permission, from Ref. c.)

involved analysis of EEG data, where ICA was used to recover signals associated with detection of visual targets [3] (see Box 3). In this case, the output sequence of each electrode is assumed to consist of a mixture of temporal independent components (tICs), which are extracted by temporal ICA (tICA). Another application of tICA is in optical

Box 4. Computational modelling and applications

Speech separation

Given a set of $N = 5$ people speaking in a room with five microphones, each voice s_i contributes differentially to each microphone output x_j [a]. The relative contribution of the five voices to each of the five mixtures is specified by the elements of an unknown 5×5 mixing matrix A (see Fig. Ia). Each element in A is defined by the distance between each person and each microphone. The output of each microphone is a mixture x_j of five independent source signals (voices) $\mathbf{s} = (s_1,\ldots,s_5)$ (echoes and time delays are ignored in this example). ICA finds a separating matrix W which recovers five independent components $\mathbf{u}$. These recovered signals $\mathbf{u}$ are taken to be estimates of the source signals $\mathbf{s}$. (Note that ICA re-orders signals, so that an extracted signal u_i and its source signal s_i are not necessarily on the same row.)

Face recognition

Fig. Ib shows how ICA treats each photograph X as a mixture of underlying spatial independent components S [b]. It is assumed that these unknown spatial independent components are mixed together with an unknown mixing matrix matrix A to form the observed photographs X. ICA finds a separating matrix W which recovers estimates U of the spatial independent components S. Note how the estimated spatial independent components U contain spatially localized features corresponding to perceptually salient features, such as mouth and eyes.

Modelling receptive fields

ICA of images of natural scenes (Fig. Ic) yields spatial independent components which resemble edges or bars [c] These independent components are similar to the receptive fields of neurons in primary visual areas of the brain (also see [d]).

References

a Bell, A. and Sejnowski, T. (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159

b Bartlett, M. (2001) *Face Image Analysis by Unsupervised Learning* (Vol. 612), Kluwer Academic Publishers

c Hyvarinen, A. *et al.* (2001) Topographic independent component analysis. *Neural Comput.* 13, 1527–1574

d Bell, A.J. and Sejnowski, T.J. (1997) The 'independent components' of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338

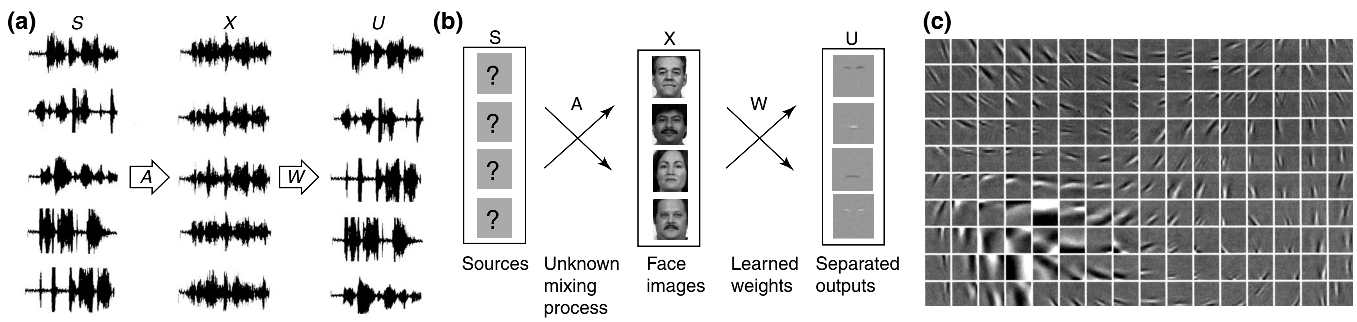

Fig. I

TRENDS in Cognitive Sciences

imaging, where it has been used to decompose the outputs of the photodetector array used to record from neurons of a sea slug [4]. Functional magnetic resonance imaging data has also been analysed using ICA [5]. Here, the fMRI brain image collected at each time point is treated as a mixture of spatial independent components (sICs), which are extracted by spatial ICA (sICA). Note that sICA and tICA make use of the same core ICA method; it is just that tICA seeks temporally independent sequences in a set of temporal mixtures, whereas sICA seeks spatially independent images in a set of image mixtures (Box 3).

The recent growth in interest in ICA can be traced back to a (now classic) paper [6], in which it was demonstrated how temporal ICA could be used to solve a simple 'cocktail party' problem by recovering single voices from voice mixtures (see Box 4). From a modelling perspective, it is thought that different neurons might encode independent physical attributes [7], because this ensures maximum efficiency in information-theoretic terms. ICA provides a powerful method for finding such independent attributes, which can then be compared to attributes encoded by neurons in primary sensory areas. This approach has been demonstrated for primary visual neurons [8] and spatial ICA has also been applied to images of faces (see Box 4).

Model-based vs data-driven methods

An ongoing debate in the analysis of biomedical data concerns the relative merits of model-based versus data-driven methods [9]. ICA is an example of a data-driven method, inasmuch as it is deemed to be exploratory (other exploratory methods are PCA and FA). By contrast, conventional methods for analysing biomedical data, especially fMRI data, rely on model-based methods (also known as parametric methods), such as the 'general linear model' (GLM). For example, with fMRI data, the GLM extracts brain activations consistent with the specific sequence of stimuli presented to a subject.

The term 'data driven' is misleading because it suggests that such methods require no assumptions regarding the data. In fact, all data-driven methods are based on certain assumptions, even if these assumptions are generic in nature. For example, ICA depends critically on the assumptions that each signal mixture is a combination of source signals which are independent and non-gaussian. Similarly, the data-driven methods FA and PCA are based on the assumption that underlying source signals (factors and eigenvectors, respectively) are uncorrelated and gaussian.

Box 5. The nuts and bolts of ICA

The output of a microphone is a time-varying signal $x_i = \{x_i^1, x_i^2, ...\}^t$, which is a linear mixture of N independent source signals (i.e. voices). Each mixture x_i contains a contribution from each source signal $s_j = \{s_j^1, s_j^2 ...\}^t$. The relative amplitude of each voice s_j at the microphone is related to the microphone–speaker distance, and can be defined as a weighting factor A_{ij} for each voice. If $N = 2$ then the relative contribution of each voice s_j to a mixture x_i is,

$$\begin{aligned} x_i &= (s_1 A_{1i}) + (s_2 A_{2i}) \\ &= (s_1, s_2)(A_{1i}, A_{2i})^t \\ &= \mathbf{s} A_{.i}, \end{aligned}$$

where $\mathbf{s}=(s_1,s_2)$ is a vector variable in which each variable is a source signal, and A_i is the ith column of a matrix A of mixing coefficients. If there are $N = 2$ microphones then each voice s_j has a different relative amplitude (defined by A_{ij}) at each microphone, so that each microphone records a different mixture x_i:

$$\begin{aligned} (x_1, x_2) &= (\mathbf{s}A_{.1})(\mathbf{s}A_{.2}) \\ &= \mathbf{s}(A_{.1}, A_{.2}) \\ &= \mathbf{s}A, \end{aligned}$$

where each column of the mixing matrix A specifies the relative contributions of the source signals $\mathbf{s}$ to each mixture x_i. This leads to the first equation in most papers on ICA:

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

where $\mathbf{x} = (x_1, x_2)$ is a vector variable, and x_1 and x_2 are signal mixtures.

The matrix A defines a linear transformation on the signals $\mathbf{s}$. Such linear transformations can usually be reversed in order to recover an estimate $\mathbf{u}$ of source signals $\mathbf{s}$ from signal mixtures $\mathbf{x}$,

$$\mathbf{s} \approx \mathbf{u} = \mathbf{x}W,$$

where the separating matrix $W = A^{-1}$ is the inverse of A. However, the mixing matrix A is not known, and cannot therefore be used to find W. The important point is that a separating matrix W exists which maps a set of N mixtures $\mathbf{x}$ to a set of N sources signals $\mathbf{u} \approx \mathbf{s}$ (this exact inverse mapping exists, but numerical methods find a close approximation to it).

Given that we want to recover an estimate $\mathbf{u} = \mathbf{x}W$ of the source signals $\mathbf{s}$ and that the latter are mutually independent then this suggests that W should be adjusted so as to make the estimated source signals $\mathbf{u}$ mutually independent. This, in turn, can be achieved by adjusting W to maximize the entropy of $\mathbf{U} = g(\mathbf{u}) = g(W\mathbf{x})$, where the function g is assumed to be the cumulative density function (cdf) of $\mathbf{s}$ (see Box 2).

If a set of signals $\mathbf{x}$ is mapped to another set $\mathbf{U}$ then the entropy $H(\mathbf{U})$ of $\mathbf{U}$ is given by the entropy of $\mathbf{x}$ plus the change in entropy ΔH induced by the mapping from $\mathbf{x}$ to $\mathbf{U}$, denoted $(\mathbf{x} \to \mathbf{U})$,

$$\begin{aligned} H(\mathbf{U}) &= H(\mathbf{x}) + \Delta H(\mathbf{x} \to \mathbf{U}) \\ &= H(\mathbf{x}) + \Delta H(\mathbf{x} \to g(W\mathbf{x})) \end{aligned}$$

As the entropy of the mixtures $H(\mathbf{x})$ is fixed, maximizing $H(\mathbf{U})$ depends on maximizing the change in entropy ΔH associated with the mapping $(\mathbf{x} \to \mathbf{U})$. The mapping from $\mathbf{x}$ to $\mathbf{U} = g(W\mathbf{x})$ depends on two terms, the cumulative density function g and the separating matrix W. The form of g is fixed, which means that maximizing $H(\mathbf{U})$ amounts to maximizing ΔH by adjusting W. In summary, a matrix W that maximizes the change in entropy induced by the mapping $(\mathbf{x} \to \mathbf{U})$ also maximizes the joint entropy $H(\mathbf{U})$. The change in entropy ΔH induced by the transformation $g(W\mathbf{x})$ can be considered as the ratio of infinitesimal volumes associated with corresponding points in $\mathbf{x}$ and $\mathbf{U}$. This ratio is given by the expected value of $\ln|J|$, where $|.|$ denotes the absolute value of the determinant of the Jacobian matrix $J = \partial\mathbf{U}/\partial\mathbf{x}$. Omitting mathematical details, this yields:

$$H(\mathbf{U}) = H(\mathbf{x}) + \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{N} \ln g_j'(u_j^{(k)}) + \ln|W|,$$

where the summation over n samples (plus $\ln|W|$) is an estimate of the expected value of $\ln|J|$, and the function $g_j' = \partial g_j / \partial u_j$ is the probability density function (pdf) of the ith source signal. Note that the entropy $H(\mathbf{x})$ is a constant, and can therefore be ignored because it does not affect the value of W that maximizes $H(\mathbf{U})$.

Given that $\mathbf{u} = W\mathbf{x}$, the derivative $\nabla H(\mathbf{U})$ of $H(\mathbf{U})$ with respect to W is

$$\nabla H(\mathbf{U}) = W^{-1} + \psi(\mathbf{u})\mathbf{x}^t,$$

where

$$\psi(\mathbf{u}) = \left[\frac{g'(u_1)}{g''(u_1)}, ..., \frac{g'(u_N)}{g''(u_N)}\right],$$

where $g''(u_j)$ is the second derivative of g.

Recall that we want to adjust W so that it maximizes $H(\mathbf{U})$. The standard method for adjusting W is gradient ascent. This consists of iteratively adding a small amount of the gradient $\nabla H(\mathbf{U})$ of $H(\mathbf{U})$ to W,

$$W_{new} = W_{old} + \eta\nabla H(\mathbf{U}),$$

where η is a learning rate. The function $H(\mathbf{U})$ can be maximized by gradient ascent using either Eqn (9), or using the more efficient 'natural gradient' [a]. Alternatively, if both the function $H(\mathbf{U})$ and its derivative $\nabla H(\mathbf{U})$ can be evaluated (as here) then a second order technique, such as conjugate gradient, can be used to obtain solutions relatively quickly.

The exposition above refers to temporal ICA (tICA) of speech signals. Spatial ICA (sICA) of images can be achieved by concatenating rows or columns of an image to produce a one-dimensional image signal x_i. Each image x_i is considered to be a mixture of independent source images (see Fig. Ib, Box 4). A set of N images is then represented as an N-dimensional vector variable $\mathbf{x}$, as in Equation 3. Spatial ICs can then be recovered using exactly the same method as described for speech signals above.

Reference

a Amari, S. (1998) Natural gradient works efficiently in learning. *Neural Comput.* 10, 251–276

Data-driven methods therefore implicitly incorporate a generic, or weak, model of the type of signals to be extracted. The main difference between such weak models and that of (for example) GLM is that GLM attempts to extract a specific signal that is a best fit to a user-specified model signal. For example, using fMRI, this user-specified signal often consists of a temporal sequence of zeros and ones corresponding to visual stimuli being switched on and off; a GLM (e.g. SPM) would then be used to identify brain regions with temporal activations which correlated with the timing of this visual switching. By contrast, data-driven methods extract a signal that is consistent with a general

ICA Resources

Books

- Bartlett, M.S. (2001) Face Image Analysis by *Unsupervised Learning* (International Series on Engineering and Computer Science), Kluwer Academic Publishers
- Girolami, M. (1999) *Self-Organising Neural Networks: Independent Component Analysis and Blind Source Separation*, Springer-Verlag
- Girolami, M., ed. (2000) *Advances in Independent Component Analysis. Perspectives in Neural Computing*, Springer-Verlag
- Hyvarinen, A. *et al.* (2001) *Independent Component Analysis*, John Wiley & Sons
- Lee, T.W. (1999) *Independent Component Analysis: Theory and Application*s, Kluwer Academic Publishers
- Roberts, S. and Everson, R., eds (2001) *Independent Component Analysis: Principles and Practice*, Cambridge University Press

Mailing list

http://tsi.enst.fr/~cardoso/icacentral/mailinglist.html

Annual conference

Abstract of the Third International Conference on Independent Component Analysis and Signal Separation, San Diego, California 9–13 December 2001, ICA2001, http://ica2001.org.

Demonstrations and software

A good place to start is: http://www.cnl.salk.edu/~tewon/ica_cnl.html

ICA code for 2D images, and demonstrations: http://www.shef.ac.uk/~pc1jvs

type of signal, where the signal type is specified in terms of the general statistical structure of the signal. Thus, the models implicit in data-driven methods are generic because they attempt to extract a type of signal, rather than a best fit to a specific model signal. The difference between model-based and data-driven methods is one of degree, and involves the relative specificity of the statistical assumptions associated with each methods.

This suggests that both classes of methods can be used, depending on the specificity of the hypothesis being tested. If there are sound reasons for hypothesizing that a specific signal will be present in the data (e.g. corresponding to a sequence of visual stimuli) then a model-based method might be preferred. Conversely, if it is suspected that a model-based method would not extract all signals of interest then an exploratory data-driven technique is appropriate. On a pragmatic note, evaluating the statistical significance of data-driven methods tends to be more difficult than that of model-based methods. Box 5 gives some of the mathematical details of ICA.

Conclusion

ICA represents a novel and powerful method, with applications in computational neuroscience and engineering. However, like all methods, the success of ICA in a given application depends on the validity of the assumptions on which ICA is based. In the case of ICA, the assumptions of linear mixing and independence appear to be physically realistic; which is perhaps why it has been successfully applied to many problems. However, these assumptions are violated to some extent by most data sets (see [9]). Whilst reports of ICA's successes are encouraging, they should be treated with caution. Much theoretical work remains to be done on precisely how ICA fails when its assumptions (i.e. linear mixing and independence) are severely violated.

Acknowledgements

Thanks to N. Hunkin, A. Hyvarinen and M. McKeown for comments on this paper.

References

1 Hyvarinen, A. (2001) Complexity pursuit: separating interesting components from time series. *Neural Comput.* 13, 883–898
2 Stone, J. (2001) Blind source separation using temporal predictability. *Neural Comput.* 13, 1559–1574
3 Makeig, S. *et al.* (1997) Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10979–10984
4 Brown, G. *et al.* (2001) Independent components analysis (ica) at the neural cocktail party. *Trends Neurosci.* 24, 54–63
5 McKeown, M. *et al.* (1998) Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop color-naming task. *Proc. Natl. Acad. Sci. U. S. A.* 95, 803–810
6 Bell, A. and Sejnowski, T. (1995) An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159
7 Barlow, H. (1989) Unsupervised learning. *Neural Comput.* 1, 295–311
8 Hyvarinen, A. *et al.* (2001) Topographic independent component analysis. *Neural Comput.* 13, 1527–1574
9 Friston, K. (1998) Modes or models: a critque of independent component analysis. *Trends Cogn. Sci.* 2, 373–375

James V. Stone
Psychology Department, Sheffield University, Sheffield, UK S10 2TP.
e-mail: j.v.stone@sheffield.ac.uk