第 28 卷 第 5 期
2002 年 9 月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 28, No. 5
Sept., 2002

综述与评论

独立成分分析方法综述[1)]

杨竹青 李 勇 胡德文
(国防科技大学机电工程与自动化学院 长沙 410073)
(E-mail: dwhu@nudt.edu.cn)

摘 要 对 ICA 方法的原理和应用进行了综述. 首先,概要叙述 ICA 的产生背景和发展前景,简要介绍和评述了 ICA 的定义、分类以及算法. 然后,对 ICA 在语音信号分离、生物医学信号处理、金融数据分析、图像噪声消除以及人脸识别等方面的实际应用进行了讨论.

关键词 独立成分分析方法,投影法,定点算法,盲源信号分离
中图分类号 TP14

INDEPENDENT COMPONENT ANALYSIS: A SURVEY

YANG Zhu-Qing LI Yong HU De-Wen
(College of Mechatronics Engineering and Automation,
National University of Defense Technology, Changsha 410073)
(E-mail: dwhu@nudt.edu.cn)

Abstract The principle and applications of the Independent Component Analysis (ICA) are surveyed. Firstly, the background and the development prospects of ICA are described and the definition, classification and algorithms of ICA are briefly introduced and evaluated. Then, it is discussed the ICA's applications such as in speech signal separation, biomedical signal processing, financial data analysis, image denoising, face recognition and so on.

Key words ICA, projection pursuit, fixed-point algorithm, blind source separation

1 引言

假想一下,在一个房间里的不同位置放着两个麦克风,同时有两个人说话. 两个麦克风能同时记录下两个时间信号,如果仅用这两个记录的信号来估计出原来的两个语音信号,那将是一件非常有意义的事情,这也就是所谓的"鸡尾酒会"问题.

独立成分分析法(ICA)最初是用来解决"鸡尾酒会"问题. 由于主成分分析(PCA)[1~3]

1) 国家基础研究重大项目前期研究专项基金、国家"863"计划基金、中国科学院模式识别国家重点实验室基金、中国科学院青年科学家创新小组基金资助
收稿日期 2001-04-28 收修改稿日期 2001-10-08

和奇异值分解(SVD)[4]是基于信号二阶统计特性的分析方法，其目的用于去除图像各分量之间的相关性，因而它们主要应用于图像数据的压缩；而 ICA 则是基于信号高阶统计特性的分析方法，经 ICA 分解出的各信号分量之间是相互独立的. 正是因为这一特点，使 ICA 在信号处理领域受到了广泛的关注. 随着近年来在 ICA 方面研究兴趣的增加，使它在许多领域也有了非常有趣的应用.

2 独立成分分析法

2.1 ICA 定义

ICA[5~10]是近几年才发展起来的一种新的统计方法. 该方法的目的是，将观察到的数据进行某种线性分解，使其分解成统计独立的成分. 最早提出 ICA 概念的是 Jutten 和 Herault[5]，当时他们对 ICA 给出了一种相当简单的描述，认为 ICA 是从线性混合信号里恢复出一些基本的源信号的方法.

为了给 ICA 下一个严格的定义[6,7]，这里需要使用一个隐藏的统计变量模型

$$\boldsymbol{x} = A\boldsymbol{s} \tag{1}$$

式(1)中的统计模型称为独立成分分析，或者 ICA 模型，它表示被观察到的数据是如何由独立成分混合而产生的. 独立成分是隐藏的变量，意味着它不能直接被观察到，而且混合矩阵也被假设为未知的. 所有能观察到的仅仅只是随机向量 $\boldsymbol{x}$，必须估计出 A 和 $\boldsymbol{s}$，而且必须在尽量少的假设条件下完成它.

ICA 的出发点非常简单，它假设成分是统计独立的，而且还必须假设独立成分是非高斯分布的. 统计独立的概念将在下面给出定义，为了简单起见，还得假设未知的混合阵为方阵. 如果能计算出 A 的逆 W，这样独立成分可由下式得到

$$\boldsymbol{s} = W\boldsymbol{x} \tag{2}$$

式(1)中的 ICA 模型存在如下的两个不确定性因素：(i)不能确定独立成分的方差；(ii)不能确定独立成分的顺序.

ICA 方法与盲源信号分离 (BSS)方法非常接近. 这里“source”指的是原始信号即独立成分，像“鸡尾酒会”问题上的说话者；“blind”指我们对混合阵几乎未知，对原始信号进行很少的假设. 给定 M 个混合信号，ICA 能同时估计出 M 个成分[11]或 $K \leqslant M$ 个成分[12]，ICA 有可能是 BSS 中用得最广泛的一种方法. 在许多实际应用中，模型中都含有噪声. 但为了简单起见，在模型中我们将忽略噪声的影响，有噪声的 ICA 见文献[13,14].

由于统计独立是 ICA 方法的前提，在开始讲述 ICA 模型估计的方法前，首先将给出独立的确切定义.

2.2 独立性定义

学术上，独立性的定义由概率密度[15]来定义. 如果定义两个随机变量 y_1和 y_2是独立的，当且仅当联合概率密度可按下式分解：$p(y_1, y_2) = p_1(y_1)p_2(y_2)$该定义可扩展到 n 个随机变量，这种情况下联合密度是 n 个随机变量的乘积. 该定义对独立的随机变量可衍生一个重要的特性：给定两个函数 h_1和 h_2，总是有 $E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\}$由于随机变量的概率密度一般都未知，从而从概率的角度来度量独立存在着一定的难度，因此我们建议采用上面这种方法. 而在文献[16]中，则是从另外一种角度来理解独立的概念. 对两个随机变量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$，如果 $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = E[\boldsymbol{xy}] - E[\boldsymbol{x}]E[\boldsymbol{y}] = 0$，那么 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 不相关；如果 $E[x^p y^q] -$

$E[x^p]E[y^q]=0$, p 和 q 对任何整数都成立,那么 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 统计独立.

从上面的推导可知,如果 x 和 y 独立,那么它们一定不相关;相反,如果 x 和 y 不相关,则并不意味着它们是独立的. 因为独立即意味着不相关,因此总是给定独立成分的不相关估计. 这样不仅减少了参数数目,而且简化了问题. 由此可得出,这两种对独立的定义方法是相通的. 只不过文献[16] 依赖于对相关性的定义,它是理解独立这个概念更直观、更简便的一种方法.

2.3 ICA 和投影法

ICA 和投影法是分离混合物为单个独立成分的两个相关的方法. 投影法(Projection Pursuit)[17~20]也是统计学里发展起来的一种方法,它的目的是为多维数据找到有意义的投影. 这种方法可用来对数据进行最优化,例如密度估计. 在最基本的 1-D 投影法里,尽量找到这样的方向使数据在该方向上的投影具有有意义的分布.

一般来说,所有非高斯测量及相应的 ICA 算法都可称作投影法的算法,特别是,投影法允许独立成分比原来的变量少. 假设独立成分所张成的空间没有充满高斯噪声,计算非高斯投影的方向,实际上是估计的独立成分;当所有的非高斯方向都被找到时,所有的独立成分实际上也已经被估计完了,这种程序可称为投影法和 ICA 的混合. 然而我们应看到,投影法并没有关于独立成分的数据模型和假设. 如果 ICA 模型成立,ICA 非高斯测量的优化将产生独立成分;如果模型不成立,那么所得到的为投影法的方向.

在文献[16]中投影法的假设条件是:任何有限个源信号的线性混合信号是高斯分布的,而源信号则是非高斯分布的. 投影法抽取信号的方法是:从线性混合信号里找到一种变换,这样抽取的是非高斯信号. 显然,统计独立假设是暗含在源信号是非高斯假设里的,因此投影法和 ICA 是基于相同的假设,且这种假设是与物质世界的时空观一致的. 不同的是,投影法一次只能抽取一个信号,而 ICA 可同时抽取一系列信号.

2.4 ICA 的分类

近年来 fMRI 成像技术日趋成熟,ICA 是一个从其他生理和非自然成分中,决定与任务相关的激活区的方法. 它的每个成分由一个固定的三维空间分布的脑体素和一个相关的激活时间序列组成,ICA 的两种互为补充方法(sICA 和 tICA)可以将一个图像序列分解成一系列图像和相应的一系列时变的图像幅度.

tICA[11]产生了一系列独立的时间序列和相应的一系列不受限制的图像. 独立时间序列被抽取,是将图像(相对每个时间点)放在 $\boldsymbol{x}$ 的列中,N 个像素中的每一个作为独立的麦克风或混合信号,每个混合信号由 T 个时间点组成.

sICA[21]产生了一系列相互独立的图像和相应的一系列不受限制的时间序列. 代替将图像放在 $\boldsymbol{x}$ 的列中,sICA 将每幅图像放在 $\boldsymbol{x}$ 的行中. 在这种情况下,每幅图像的像素值相对其它的图像来说是独立的.

严格的来讲,无论是 sICA 还是 tICA,都是在牺牲物理上不可能实现的形式来满足对成分独立的限制条件. stICA[22]则将独立成分和它们相应的信号放在同等重要的位置上,它同时最大化空间和时间上的独立度,即 stICA 在相互独立的图像和相互独立的时间序列上采取了一种折衷的方法.

3 ICA 估计原理

当前估计 ICA 模型的主要方法有非高斯的最大化、互信息的最小化、最大似然函数估

计(ML).

3.1 非高斯的最大化

在大多数经典的统计理论里，随机变量被假设为高斯分布. 概率论里一个经典的结论——中心极限定理表明，在某种条件下，独立随机变量的和趋于高斯分布，独立随机变量的和比原始随机变量中的任何一个更接近于高斯分布.

为简单起见，假设所有独立成分都有相同的分布. 为了估计其中的一个独立成分，考虑 $\boldsymbol{x}_i$ 是 $y=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=\sum_i w_i x_i$ 的线性组合，这里 $\boldsymbol{w}$ 是一个待定的向量. 如果 $\boldsymbol{w}$ 是 A 的逆中的一行，这个线性组合实际上将等于一个独立成分. 问题是怎样利用中心极限定理来确定 $\boldsymbol{w}$? 实际上，不能确切的确定 $\boldsymbol{w}$，因为并不知道矩阵 A. 但是可以找到一个很接近的估计，这也就是 ICA 估计的基本原理. 将变量进行一下变换，定义 $\boldsymbol{z}=A^{\mathrm{T}}\boldsymbol{w}$，则有 $y=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{w}^{\mathrm{T}}A\boldsymbol{s}=\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}$，$y$ 是 s_i 的一个线性组合，其权重由 z_i 给出. 因为两个独立随机变量的和比原始的变量更接近高斯分布，所以 $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}$ 是比任何一个 s_i 更接近高斯分布. 因此可把 $\boldsymbol{w}$ 看作是最大化非高斯 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$ 的一个向量，这样的一个向量对应于 $\boldsymbol{z}$，则有 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{z}^{\mathrm{T}}\boldsymbol{s}$ 等于其中的一个独立成分！

最大化 $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}$ 的非高斯性，即可得到一个独立成分. 实际上，在 n 维空间最优化非高斯向量 $\boldsymbol{w}$ 有两个局部最大点，相应的每个独立成分有两个即 s_i 和 $-s_i$. 为找到几个独立成分，需要找到所有的局部最大点. 这一点并不困难，因为不同的独立成分是不相关的.

直接地讲，估计 ICA 模型的关键是非高斯度量. 在 ICA 估计里为了使用非高斯性，因此对随机变量的非高斯性必须有一个定量的测量标准. 现将非高斯性的测量方法综述如下.

1) Kurtosis

经典的测量非高斯方法是 kurtosis 或称 4 阶累计量[23]，y 的 kurtosis 被定义为

$$\mathrm{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2 \tag{3}$$

实际上，因为我们假设 y 是单位方差，等式右边可简化为 $E\{y^4\}-3$. 对一个高斯分布 y，它的 4 阶矩等于 $3(E\{y^2\})^2$. 对一个高斯随机变量来说，它的 kurtosis 等于零；但对大多数非高斯随机变量，它的 kurtosis 不等于零. Kurtosis 有正也有负，下高斯随机变量具有负的 kurtosis，而上高斯随机变量则具有正的 kurtosis. 非高斯性的测量可以用 kurtosis 的绝对值或 kurtosis 的平方，值为零的是高斯变量，大于零的为非高斯变量. 存在 kurtosis 为零的非高斯变量，但这种情况相当的少.

kurtosis 或它的绝对值，由于能从理论上用来作为解决 ICA 问题时的最优化准则及计算和理论上的简单性，因此已经广泛地用于 ICA 的非高斯的测量及其相关领域. 计算上简单，是由于 kurtosis 能用采样数据的 4 阶矩简单地进行估计；理论分析上简单，是由于下面的线性特性，即如果 x_1 和 x_2 是两个独立的随机变量，那么

$$\mathrm{kurt}(x_1 + x_2) = \mathrm{kurt}(x_1) + \mathrm{kurt}(x_2) \tag{4}$$

和

$$\mathrm{kurt}(ax_1) = a^4\mathrm{kurt}(x_1) \tag{5}$$

成立. 这里 a 是一个标量，这些性质由定义很轻易地得到证明.

2) 负熵(Negentropy)

第二个非常重要的非高斯测量方法是负熵，它是基于信息理论上熵的概念. 随机变量的熵可解释为给定观察变量的信息度，越随机，熵越大. 实际上，在一些简单的假设条件下，熵就是指随机变量的代码长度，这在信息论[24]里有介绍.

离散的随机变量 Y 的负熵 H 被定义为

$$H(Y) = -\sum_i P(Y = a_i)\log P(Y = a_i) \tag{6}$$

这里 a_i 指 Y 的可能值. 这是个很好的定义,可扩展到连续的随机变量和向量,这种情况下称为微熵. 随机向量 $\boldsymbol{y}$ 的密度 $f(\boldsymbol{y})$ 的微熵 H 被定义为

$$H(\boldsymbol{y}) = -\int f(\boldsymbol{y})\log f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{7}$$

信息理论一个基本的结论是,在所有具有等方差的随机变量中,高斯变量的熵最大. 这意味着熵能用来作为非高斯性的测量,分布明显的集中于某个值的熵很小. 非高斯性测量中,高斯变量应该为零,而它总是非负. 有人对熵的定义作了修改,称为负熵,定义如下:

$$J(\boldsymbol{y}) = H(\boldsymbol{y}_{\text{gauss}}) - H(\boldsymbol{y}) \tag{8}$$

这里 $\boldsymbol{y}_{\text{gauss}}$ 是一个高斯随机向量,与 $\boldsymbol{y}$ 有相同的协方差. 由于上面提到的特性,负熵总是非负的. 它为零的条件是当且仅当 $\boldsymbol{y}$ 是高斯分布.

使用负熵的问题是计算起来非常困难. 因此,采用负熵的近似是非常有用的.

3) 负熵的近似

如上所述,负熵的估计是很困难的,因此必须采取一些近似. 这里介绍一些有较好特性的近似,它将在 ICA 方法中使用到. 近似负熵古典的方法是使用高阶矩,例如

$$J(\boldsymbol{y}) \approx \frac{1}{12}E\{\boldsymbol{y}^3\}^2 + \frac{1}{48}\text{kurt}(\boldsymbol{y})^2 \tag{9}$$

随机变量 $\boldsymbol{y}$ 被假设为零均值、单位方差. 然而,这种近似的有效性非常有限. 特别是,这种近似对非鲁棒性非常敏感. 为了避免这种问题,采用另外一种近似,这种近似是基于最大熵原理,下面给出一些性能较好的比较函数. 一般来说,可得到如下的近似

$$J(\boldsymbol{y}) \approx k_i[E\{G_i(\boldsymbol{y})\} - E\{G_i(\boldsymbol{v})\}]^2 \tag{10}$$

这里 k_i 是一些正的常数,$\boldsymbol{v}$ 是零均值、单位方差的高斯变量. $\boldsymbol{y}$ 被假设为零均值、单位方差的变量,函数 G_i 是一些非二次函数. 注意,即使在这种情况下,这种近似也是不精确的. 在我们仅用非二次函数 G 的情况下,这种近似变成了

$$J(\boldsymbol{y}) \propto [E\{G(\boldsymbol{y})\} - E\{G(\boldsymbol{v})\}]^2 \tag{11}$$

在式(11)中明显的是基于矩的近似,如果 $\boldsymbol{y}$ 是对称的,例如取 $G(\boldsymbol{y}) = \boldsymbol{y}^4$,就能准确地得到式(11),即基于 kurtosis 的近似

$$G_1(u) = \frac{1}{a_1}\log\cos a_1 u, \quad G_2(u) = \exp(-u^2/2) \tag{12}$$

这里 $1 \leqslant a_1 \leqslant 2$ 为一些适合的常数. 这样近似得到了负熵,它给出了古典的 kurtosis 和负熵在非高斯性测量上的一种很好的折衷. 它们概念上简单,计算起来快速,而且有很好的统计特性,尤其是鲁棒性. 因此,在 ICA 方法中我们建议使用这些比较函数.

3.2 互信息的最小化

对 ICA 估计的另一种方法,是基于信息理论的最小化互信息. 利用熵的概念,我们定义 m 个随机变量 $\boldsymbol{y}_i(i=1,\cdots,m)$ 的互信息 I 如下:

$$I(y_1, y_2, \cdots, y_m) = \sum_{i=1}^{m} H(\boldsymbol{y}_i) - H(\boldsymbol{y}) \tag{13}$$

在随机变量间互信息是对相关性的一种自然测量. 它总是非负的,当且仅当变量是统计独立的时候它才为零. 因此,互信息考虑了变量的整个相关性结构,而不像 PCA 和其他相关的方法一样,仅仅考虑了协方差.

互信息能解释为熵的代码长度，当 y_i 的代码单独给出时，$H(y_i)$ 给出了代码长度；当 y 作为一个随机向量编码时（例如，所有的成分以同样的代码编码），$H(y)$ 给出了代码长度. 互信息表明，代码的减少，是通过对整个向量编码而不是分离成分而得到的. 总的来说，通过对整个向量编码可得到较好的代码. 然而，如果 y_i 是独立的，它们相互间不提供任何信息，也能独立地对变量进行编码而并不增加代码长度.

互信息一个重要的特性是，可以对线性变换 $y=Wx$ 进行如下转换

$$I(y_1,y_2,\cdots,y_n)=\sum_i H(y_i)-H(x)-\log|\det W| \tag{14}$$

现在，如果我们限制 y_i 非相关和单位方差，则 $E\{yy^{\mathrm{T}}\}=WE\{xx^{\mathrm{T}}\}W^{\mathrm{T}}$ 表明：$\det I=1=(\det WE\{xx^{\mathrm{T}}\}W^{\mathrm{T}})=(\det W)(\det E\{xx^{\mathrm{T}}\})(\det W^{\mathrm{T}})$，这意味着 $\det W$ 是一个常数，而且由于 y_i 单位方差，熵和负熵区别仅在于一个符号，因此得到

$$I(y_1,y_2,\cdots,y_n)=C-\sum_i J(y_i) \tag{15}$$

这里 C 是一个并不依赖 W 的常数，表明负熵和互信息之间基本的关系.

在信息论中，既然互信息是随机变量独立性的测量量度，那么可用它来作为找到 ICA 变换的准则. 在 ICA 的定义中，随机向量 x 在 $y=W^{\mathrm{T}}x$ 中是一个可逆变换，一旦 W 被确定了，成分 s_i 的互信息也就被最小化了.

很明显找到一个可逆矩阵 W 最小化互信息，相当于找到了负熵最大化的方向. 更精确地，它等同于找到 1 维子空间，这些子空间的投影有最大的负熵. 严格地说，当估计互不相关时，通过最小化互信息来估计 ICA 模型，相当于最大化非高斯估计的和. 不相关这个约束条件在这里实际没必要，但为了大量简化计算，人为地采用了这种简单的形式.

在随机变量间，互信息是对相关性的一种自然测量. 它总是非负的，当且仅当变量是统计独立的时候它才为零. 互信息考虑了变量的整个相关性结构，而不像 PCA 和其它相关的方法一样，仅仅考虑了协方差.

3.3　最大似然函数估计（ML）

估计 ICA 一个非常普遍的方法是最大似然估计. 它与信息原理紧密相关，本质上它与最小化互信息是相同的. 在无噪声的 ICA 模型中可以直接定义似然函数[25]，然后用最大似然函数的方法来估计 ICA 模型. 如果 $W=(w_1,\cdots,w_n)^{\mathrm{T}}$，等于矩阵 A^{-1}，对数似然函数采取如下形式[21]

$$L=\sum_{t=1}^{T}\sum_{i=1}^{n}\log f_i(w_i^{\mathrm{T}}x(t))+T\log|\det W| \tag{16}$$

这里 f_i 指的是 s_i 密度函数（这里假设已知），$x(t)(t=1,\cdots,T)$ 是 x 的实现. $\log|\det W|$ 来源于古典规则，该规则为了线性转换随机变量和它们的密度. 一般来说，对任何具有密度 p_i 的随机向量 x 和任何矩阵 W，$y=Wx$ 的密度由信息原理给出.

另外一个相当于最大似然函数估计的方法是从神经网络的观点得到的[13,26]，在具有非线性输出的神经网络中，它是基于最大化输出熵（或信息流）. 假设 x 是该神经网络的输入，其输出为 $g_i(w_i^{\mathrm{T}})$，这里 g_i 是一个非线性标量函数，w_i 是神经元的权向量，最大化输出的熵，我们得到

$$L_2=H(g_1(w_1{}^{\mathrm{T}}x),\cdots,g_n(w_n{}^{\mathrm{T}}x)) \tag{17}$$

有些作者，像文献[22,27]已经证明了这个神奇的结论，即最大化网络熵原理. 相当的条件是，要求用在神经网络中 g_i 的非线性选择为累积分布函数相对应的密度 f_i. 如果 g_i 选择得

合适,就能估计出 ICA 模型.

最大似然函数估计要求 f_i的密度必须估计准确[23,28,29],在任何情况下,如果关于独立成分特性的信息不准确,ML 估计将给出完全错误的结论.因此,在使用 ML 估计时必须小心.相反,使用合理的非高斯测量将不会产生该类问题.

4 ICA 算法

当前 ICA 算法简单地可分为两类:第一类[6,11,24,30~34],最大和最小化一些相关准则函数,这类算法的优点是对任何分布的独立成分都适合,但它们要求非常复杂的矩阵运算或张量运算,计算量非常大;第二类[7,10,15,33,34],基于随机梯度方法的自适应算法,该类算法优点是能保证收敛到一个相应的解,但其主要问题是收敛速度慢,且其收敛与否很大程度上依靠学习速率参数的正确选择.

近两年来又出现了一种快速 ICA 算法(FastICA),该算法是基于定点递推算法得到的.它对任何类型的数据都适用,同时它的存在对运用 ICA 分析高维的数据成为可能,目前我们正用该算法对 fMRI 数据进行分析.对已经存在的 ICA 算法来说,FastICA 有许多优良的特性.

· 在 ICA 数据模型的假设下,FastICA 收敛速度是 3 次的(或至少是 2 次的).而普通的 ICA 算法收敛速度仅仅是线性的.FastICA 算法的收敛速度之快,已经通过仿真试验得到了证实(见文献[35]).

· 与基于梯度的算法相比,它不需要选择步长.这表明该算法易于适用.

· FastICA 算法直接找到了任何非高斯分布的独立成分,通过使用一个非线性函数 g.相对其它许多算法来说,它们首先必须进行概率密分布函数的估计,然后才相应地进行非线性的选择.

· FastICA 算法的性能能通过一个合适的非线性函数 g 而使其达到最优.特别是,能得到具有鲁棒的或最小方差的算法,详细请见文献[10].

· 独立成分能一个一个的估计,这在探索性数据分析里非常有用,如果仅需估计一些独立成分时,它可极大地减少计算量.

· FastICA 算法具有很多神经算法里的优点:并行的,分布的,计算简单,要求内存小.而随机梯度法却只有在环境改变很快时才显出其优越性.

Bell 的 ICA 算法[11]采用 fMRI 数据在一台 Alphaserver2100 数字计算机上运行,需要用时大约 60 分钟;而同样的数据用 Comon[6]的 ICA 算法在同样的计算机上进行计算,则耗时达 390 分钟之久.可见,适当的选择算法会大大地提高实验效率.随着新的解剖和功能图像方法的来临,从活人的大脑收集大量数据已成为了可能.与之相适应,未来的 ICA 算法预计将向省时、省内存方面发展.

5 ICA 的应用

ICA 的主要应用是特征提取[3,34]、盲源信号分离[23,32]、生理学数据分析[21,22,36~39]、语音信号处理[40]、图像处理[41]及人脸识别[42]等.在这部分,我们综述一下 ICA 的主要应用范例.

4.1　在脑磁图(MEG)中分离非自然信号

脑磁图是一种非扩散性的方法.通过它,活动或者脑皮层的神经元有很好的时间分辨率和中等的空间分辨率.作为研究和临床的工具使用 MEG 信号时,研究人员面临着在有非自然信号的情况下提取神经元基本特征的问题.干扰信号的幅度可能比脑信号的幅度要高,非自然信号在形状上像病态信号.在文献[36]中,作者介绍了一种新的方法(ICA)来分离脑活动和非自然信号.这种方法是基于假设:脑活动和非自然信号(像眼的运动或眨眼或传感器失灵)是解剖学和生理学上的不同过程,这种不同反映在那些过程产生的磁信号间的统计独立性上.在这之前,人们用脑电图(EEG)信号进行过试验[37],相关的方法见文献[43].

试验结果表明,ICA 能很好地从 MEG 信号里分离出眼运动及眨眼时的信号,还能分离出心脏运动、肌肉运动及其它非自然信号.FastICA 算法是一个很合适的算法,因为非自然信号的去除是一个交互式的方法,研究者可以很方便地选择他所想要的独立成分的数目.除了减少非自然信号外,ICA 还能分解激活区[38],使我们直接访问基本的脑功能成为可能.这一点在神经科学的研究领域将很可能起非常重要的作用,我们也正从事将 ICA 运用到 fMRI 数据分析这方面的工作.

4.2　在金融数据中找到隐藏的因素

将 ICA 用在金融数据中是一个探索性的工作.在这个应用中存在许多情况(并行的时间序列),例如流通交易率或每日的股票成交量,这里存在一些基本的因素,ICA 可以揭示一些仍隐藏着的驱动机制.在近年来的证券研究中,人们发现 ICA 是对 PCA 的一种补充工具,它允许数据的基本结构能更轻易地观察得到.在文献[44]中,将 ICA 用在了不同的问题上,属于同一个销售链的商店的现金流量,尽量找到对现金流量数据有影响的一些基本因素.对独立成分的假设有可能不现实,例如假期和年度的变化,顾客购买力的变化,政府和经营策略(像广告)等等因素,通通假设它们之间是相互独立的.通过 ICA,利用现金流量时间序列数据,能分离出一些基本的影响因素和它们的权重,并且以此还能对商店进行分组.对于试验和解释,详细情况请参见文献[44].

4.3　自然图像中减少噪声

第三个例子是为自然图像找到 ICA 过滤器.它是基于 ICA 分解,从被高斯噪音污染的自然图像中去掉噪声.文献[45]采用了一些数字的自然图像,向量 $\boldsymbol{x}$ 代表了图像窗口的像素(灰度)值.注意,相对前面的两个应用,这次考虑的不是多值的时间序列或图像随时间而改变,相反元素 $\boldsymbol{x}$ 已经由图像窗口的位置固定不变了.采样窗口采样的是随机位置,窗口的2-D结构在这里并不重要,一行一行的扫描整幅图像使其变成像素值的向量.实验结果发现,没有经过边界的模糊及锐化操作,窗口的大部分噪声被去掉了,详细的情况参见文献[45].

当前去噪声方式有许多,例如先作 DFT 变换,然后在作低通滤波,最后作 IDFT 恢复图像[46],这种方式不是很有效.较好的方法是近年来发展起来的小波收缩方法[46](它用到了小波变换)和中值滤波[46].但这些对图像统计量来说并没有很好的优越性.近年来又发展了一种统计原理的方法,叫稀疏代码收缩法[45],该方法与独立成分分析法非常接近.

4.4　人脸识别

人脸识别从 20 世纪 70 年代开始一直是一个很活跃而且很重要的研究领域,当时比较常用的方法是主成分分析(PCA)和本征脸.后来,Bartlett 和 Sejnowski 提议用 ICA 来表示人脸.

将 ICA 运用到人脸识别,随机变量为训练的图像. $\boldsymbol{x}_i$表示一个人脸的图像. 用 m 个随机变量来构造一个训练图像集$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m\}$,这些随机变量被假设为 n 个未知独立成分 $\boldsymbol{s}_1,\cdots,\boldsymbol{s}_n$的线性组合. 采用前面所讲过的矩阵的记法:$X=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_m)^{\mathrm{T}}$,$S=(\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_n)^{\mathrm{T}}$,则有 $X=AS$. 从这个表达式可看出,每个图像 $\boldsymbol{x}_i$由 $\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_n$与 $a_{i1},\cdots,a_{in}$的线性组合来表示. 因此,混合矩阵 A 也称特征矩阵,可看作是所有训练图像的特征. 与 PCA 相比,ICA 有如下几个优点: 1) ICA 是从训练信号里去高阶统计量的相关性,而 PCA 则只对二阶统计量去相关性; 2) ICA 基向量比 PCA 基向量在空间上更局部化,而局部特征对人脸表示很重要; 3) 实践证明,ICA 基向量识别精度比 PCA 要高. 为此,ICA 可以作为模式识别分类的一个预处理步骤.

4.5 图像分离

我们曾用 FastICA 算法将三幅混合图像进行了成功的分离. 仿真结果表明,原图像与分离出来的图像十分相似,而且每次迭代的次数不超过 15 次,计算量非常小. 下一步,我们的的工作是对快速定点算法进行改进,争取在节省内存方面取得一定的成效.

4.6 语音信号处理

ICA 最经典的应用是"鸡尾酒会"问题. 在 n 个麦克风记录的 n 个声音源中,通常仅仅希望得到其中感兴趣的一个声音源,而把其他的声音源视为噪声. 如果仅一个麦克风,我们可以用普通的去噪声方法来去噪声,例如,线性滤波,小波或稀疏码收缩方法. 当然,这种去噪声的方法不是很令人满意. 我们能利用多个麦克风来收集更多的数据,以便更有效的去噪声. 因为在现场麦克风的位置是任意的,而且混合过程也未知,为此必须实行盲估计. 采用的方法就是,盲源信号分离中的一种,即 ICA 方法.

4.7 远程通信

最后,提一下另外一个很有潜力的应用——远程通信. 实时通信的应用例子是,在 CDMA 移动通迅[48]里,从有其他用户干扰的信号里分离用户自己的声音. 这个问题从某种意义上说,在 CDMA 数据模型中预先给出了一些优先信息. 但是需估计的参数数目很大,因此选定某种合适的信号分离方法,它考虑了这种优先信息,从而产生了比传统估计方法更优越的性能.

6 结论

独立成分分析法是一种多用途的统计方法,它在语音信号分离、生物医学信号处理、金融数据分析、图像消噪、远程通信、人脸识别等方面的应用成果充分显示了 ICA 的特点及非常重要的应用价值. 由于 ICA 是近年来才出现的,它的理论和算法还不太成熟,因此本文所讨论的很多内容还需要进一步补充和完善. 目前 ICA 在国外发展得较快,而国内则刚刚起步,写此文的目的是为了提高国内读者对 ICA 理论及应用研究的兴趣.

参 考 文 献

1 Oja E. Principal Components,Minor Components, and linear neural networks. *Neural Networks*, 1992, **5**(6):927～935

2 Oja E. The nonlinear PCA learning rule in independent component analysis. *Neurocomputing*, 1997, **17**(1): 25～45

3 焦李成. 神经网络的应用与实现. 西安:西安电子科技大学出版社,1996

4 孙即祥. 数字图像处理. 石家庄:河北教育出版社,1993

5 Jutten C, Herault J. Independent component analysis versus PCA. In: Proceeding of European Signal Processing

Conf, 1988. 287～314

6 Comon P. Independent component analysis——A new concept? *Signal Processing*, 1994, **36**(3):287～314

7 Jutten C, Herault J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 1991, **24**(1):1～10

8 Oja E. The nonlinear PCA learing rule in independent component analysis. *Neurocomputing*, 1997, **17**(1):25～46

9 Hyvärinen A. Survey on independent component analysis. *Neural Computing Surveys*, 1999, **2**(4):94～128

10 Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 1999, **10**(3):626～634

11 Bell A J, Sejnowski T J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 1995, **7**:1129～1159

12 Porrill J, Stone J. Independent component analysis for signal separation and dimension reduction. In: Technical Report 124, Psychology Department, S10 2UR University, Sheffied, England

13 Nadal J-P, Parga N. Non-linear neurons in the low noise limit: A factorial code maximizes information transfer. *Network*, 1994, **5**(4):565～581

14 Hyvarinen A. Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood. *Neurocomputing*, 1998, **22**(1-3):49～67

15 Hyvarinen A, Oja E. Independent component analysis: a tutorial. *Neural Networks*, 2000, **13**(4-5):411～430

16 Lee Te-Won. Independent Component Analysis: Theory and Applications. Boston: Kluwer Academic Publishers, 1998

17 Friedman J H, Tukey J W. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. of Computers*, 1974, **c-23**(9):881～890

18 Friedman J H. Exploratory projection pursuit. *J. of the American Statistical Association*, 1987, **82**(397):249～266

19 Huber P J. Projection pursuit. *The Annals of Statistics*, 1985, **13**(2):435～475

20 Jones M C, Sibson R. What is projection pursuit? *Journal of the Royal Statistical Society*, Serise. A, 1987, **150**:1～36

21 McKeown, MJ, Jung TP, Makeig S *et al.* Spatially independent activity patterns in functional MRI data during the stroop color-naming task. In: Proceedings of the National Academy of Sciences, 1998, **95**(6):803～810

22 Stone J V, Porrill J, Porter NR, *et al.* Spatiotemporal ICA of fMRI Data. In: Computational Neuroscience Report 202, IEEE International workshop on Biologically Motiviated Vision, 2000

23 Cardoso J-F, Laheld B H. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 1996, **44**(12):3017～3030

24 章照止，林须端. 信息论与最优编码. 上海：上海科学技术出版社，1993

25 Pham D-T, Garrat P, Jutten C. Separation of a mixture of independent sources through a maximum likelihood approach. In: Proceedings of Europen Signal Processing Conf, Elsevier Science Publishers, 1992. 771～774

26 Cover T M, Thomas J A. Elements of Information Theory. New York: Wiley, 1991

27 Hyvärinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 1997, **9**(7):1483～1492

28 Amari S-I, Cichocki A, Yang H H. A new learning algorithm for blind source separation. In: Advances in Neural Information Processing Systems 8. Cambridge, MA: MIT Press, 1996. 757～763

29 Lee T-W, Girolami M, Sejnowski T J. Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 1999, **11**(2):417～441

30 Hyvarinen A. Independent component analysis by minimization of mutual information. In: Technical Report A46, Helsinki university of Technology, Laboratory of Computer and Information Science, 1997

31 Girolami Mark, Fyfe Colin. Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms. NIPS * 96 workshop

32 Oja E, Yang Howard Hua, Amari S-I. Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information. *Neural Computation*, 1997, **9**(7):1457～1482

33 Thomas Heinz Mathis, Hoff P von Joho Marcel. Blind separation of mixed-kurtosis signals using an adaptive threshold nonlinearity. In: Proc. International Conference on Independent Component Analysis and Blind Signal Separation, Helsinki, Finland, 2000. 221～226

34 Karhunen J, Hyvarinen A, Vigario R *et al.* Applications of neural blind separation to signal and image processing.

In: Proceedings of the IEEE 1997 International Conference on Acoustics, Speech, and Signal Processing, Munich, Germany, 1997, to appear. 53

35 Giannakopoulos X, Karhunen J, Oja E. Experimental comparison of neural ICA algorithms. In: Proc. Int. Conf. on Artificial Neural Networks (ICANN'98), Skövde, Sweden, 1998. 651~656

36 Vigário R, Jousmäki V, Oja E *et al*. Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In: Advances in Neural Information Processing Systems 10, MIT Press, 1998. 229~235

37 Vigário R. Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. Clin. Neurophysiol*, 1997, **103**(3):395~404

38 Vigário R, Särelä J, Oja E. Independent component analysis in wave decomposition of auditory evoked fields. In: Proc. Int. Conf. on Artificial Neural Networks (ICANN'98), Skövde, Sweden, 1998. 287~292

39 Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, **381**(6583):607~609

40 Torkkola K. Blind separation for audio signals-are we there yet? In: Proc. Int. Workshop on Independent Component Analysis and Signal Separation(ICA'99), Aussois, France: 1999. 239~244

41 Bell A J, Sejnowski T J. The' independent components' of natural scenes are edge filters. *Vision Research*, 1997, **37**(23):3327~3338

42 Yuen P C, Lai J H. Face representation using independent component analysis. *Pattern Recognition*, 2001, **34**(3): 545~553

43 Makeig S, Bell A J, Jung T-P *et al*. Independent component analysis of electroencephalographic data. In: Advances in Neural Information Processing Systems 8, MIT Press, 1996. 145~151

44 Mallat S G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on PAMI*, 1989, **11**(7):674~693

45 Hyvärinen A. Sparse code shrinkage: Denoising of nongaussian data by maximum likelihood estimation. *Neural Computation*, 1999, **11**(7):1739~1768

46 Gonzalez R, Wintz P. Digital Image Processing. Addison-Wesley Publishing Company, 1987

47 Donoho D L, Johnstone I M, Kerkyacharian G *et al*. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society*, Server. B, 1995, **57**:301~337

48 Ristaniemi T, Joutsensalo J. On the performance of blind source separation in CDMA downlink. In: Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99), Aussois, France, 1999. 437~441

49 Cardoso J-F. Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing*, 1997, **4**(4):112~114

50 Pearlmutter B A, Parra L C. Maximum likelihood blind source separation: A context-sensitive generalization of ICA. In: Advances in Neural Information Processing Systems, Volume 9, MIT Press, 1997. 613~619

51 Kiviluoto K, Oja E. Independent component analysis for parallel financial time series. In: Proc. ICONIP'98, Volume 2, Tokyo, 1998. 895~898

52 Hyvärinen A, Oja E. Independent component analysis by general nonlinear Hebbian-like learning rules. *Signal Processing*, 1998, **64**(3):301~313

53 Luenberger D G. Optimization by Vector Space Methods. Publisher: John Wiley & Sons, 1969

54 Cichocki A, Bogner R E, Moszczynski L *et al*. Modified herault-jutten algorithms for blind separation of sources. *Digital Signal Processing*, 1997, **7**(2):80~93

55 Hyvärinen A. The fixed-point algorithm and maximum likelihood estimation for independent component analysis. *Neural Processing Letters*, 1999, **10**(1):1~5

56 Ziehe A, Muller K, Nolte G *et al*. Artifact reduction in magnetoneurography based on time-delayed second-order correlations. *IEEE Trans. Biomedicine Engineer*, 2000, **47**(1):75~87

杨竹青 1995年考入河北大学自动控制系,2002年于国防科技大学获硕士学位.

李 勇 1997年考入国防科技大学自动控制系,现为硕士研究生.

胡德文 博士,教授,博士生导师.主要从事系统辨识、神经网络、图像处理、脑科学等研究.