

## 自适应多目独立成分分析

余 堃<sup>1,2</sup>, 蒲红梅<sup>3</sup>, 郑方伟<sup>1</sup>, 周明天<sup>1,2</sup>

(1. 电子科技大学计算机科学与工程学院 成都 610054; 2. 四川师范大学四川省软件重点实验室 成都 610066;  
3. 重庆通信学院第4系 重庆 沙坪坝区 400035)

**【摘要】**通过分析LCNN的学习方程,发现Lagrange约束项的物理本质是有监督学习的下降速率,提出了自适应LCNN(ALCNN)算法,避开了病态矩阵的问题,并将学习矩阵和独立成分求解复杂性都降到了 $O(n)$ 。

**关键词** 独立成分分析; 盲源分离系统; 鸡尾酒会问题  
中图分类号 TN911.7 文献标识码 A

## Adaptive Independent Component Analysis under Multisensing

SHE Kun<sup>1,2</sup>, PU Hong-mei<sup>2</sup>, ZHENG Fang-wei<sup>1</sup>, ZHOU Ming-tian<sup>1,2</sup>

(1. School of Computer, Univ. of Electron. Sci. & Tech. of China Chengdu 610054;  
2. Software Key Lab. of Sichuan Province, Sichuan Normal University Chengdu 610066;  
3. 4th Department, Chongqing Communication College Shapingba Chongqing 400035)

**Abstract** In this paper, LCNN equation is investigated carefully and the inward nature of constraints, which was the down speed of supervised learning, is discovered. At the end, adaptive LCNN (ALCNN) is proposed, which not only can solve ill-conditioned matrix, but also the computing complexities of learning matrix and independent components are symplified to  $O(n)$ .

**Key words** independent component analysis; blind source separation; cocktail party problem

独立成分分析(Independent Component Analysis, ICA)源于文献[1]提出的语音和一维信号的盲源分离系统(Blind Source Separation, BSS)。“H-J算法”在假设源信号间相互独立的情况下,对此进行了求解。文献[2]利用ICA解释了著名的“鸡尾酒会问题”,即“一个人能够在充满噪声的酒会中分离出混合的声音信息,并能识别出说话人以及在交叉对话中的一些重要信息”。国际上有4个研究小组提出了用目标函数统计独立方法(如最大化熵<sup>[2-3]</sup>、最小互信息<sup>[4]</sup>、峭度绝对值最大<sup>[5]</sup>、最大负熵<sup>[3]</sup>等)解决ICA问题,它们的工作现在统称为“BSAO”方法<sup>[6]</sup>。BSAO方法问题,来自于需要所有后验知识的统计平均和后验概率分布,但在现实环境中是很难确定的。另外,全部后验信息的统计平均损失了几何信息,因此求解是不精确的。

### 1 盲源分离系统(BSS)和ANN方法

人脑对于全新的东西,如何判断其是否有用,以及如何吸纳其有用成分、抛弃无用垃圾?现在的

认识认为存在一个无监督的学习过程作用于这种全新的信号,其数学描述为:

$$As=x \quad (1)$$

式中  $s$  是  $n$  维未知的待估计源;  $x$  是  $m$  维观测矢量;  $A=[a_{ij}]$  是已知的典型列满秩实数转换矩阵,  $A \in R^{m \times n}$ 。BSS问题是仅仅知道观测的信号  $x$ , 需要分离出源信号  $s$  的各个成分  $s_i$ , 因此, 解决此类问题的系统统称为盲源分离系统(BSS)。

文献[7-8]指出无论是有监督、无监督学习, 还是前馈型、反馈型, 所有的人工神经网络(ANN)问题都等价于Lagrange最优化问题, 即:

$$\varepsilon=C+\lambda(x-As) \quad (2)$$

式中  $\lambda$  是常数向量;  $C$  是目标函数。求解  $\min(\varepsilon)$ , 式(1)转换成式(2)中的Lagrange约束项, 实际上要求  $\min(C)$  和  $\min(x-As)$  联合收敛。

求解此类问题有很多方法, 目前最成功的方法来源于模拟生物神经网络的人工神经网络模型(ANN)的无监督学习, 即“原始数据进, 知识留, 垃圾出”。知识沉淀(即表达)在神经元之间的权矩阵上。

文献[7]引入动力学方程对式(2)求解，得到：

$$dg/dt = -d\varepsilon/dv = 0 \quad (3)$$

式中  $v=f(g)$ ，很多情况下  $g$  就是网络的权值  $W_{ij}$ 。对于ICA，目标函数是最大独立性测度的负值，问题转为最小优化(最大化的负值就是最小化)问题，以熵最大化为例，式(3)转变为式(4)：

$$dW_{ij}/dt = dH(W)/dW_{ij} = 0 \quad (4)$$

式(3)和(4)不等式的成立取决于  $v=f(g)$  是一个单调上升函数，这是ANN最常用的模型。如果  $v$  不是单调上升的，解空间会出现混沌和分叉，文献[9-10]利用混沌神经网络对此类问题逐步给出了解决的方案。

## 2 基于Helmholtz自由能的ICA

### 2.1 问题的提出和解决方案

式(4)是目前流行的BSAO方案的Infomax方法，它需要全部后验信息的联合熵，在不知道联合后验概率密度分布的情况下，所有假设都是不准确的。“鸡尾酒会问题”应该是大脑无监督学习和有监督学习并发学习的表现。文献[7]指出在恒定温度下的Helmholtz自由能  $\Omega = E - TH$ ，等同于ANN的Lyapunov函数，即：

$$L = E - TH \quad (5)$$

因此统一了最小化有监督学习能量  $E(Y)$  和最大化输出熵  $H(W)$  的无监督学习模型，即最小化自由能使得基于最小化神经元发放率能量的有监督学习和基于最大化独立基(熵最大)的无监督学习达到适当的平衡。

利用式(2)的Lagrange最优化方法，可得到Lagrange约束神经网络(LCNN)技术，该技术主要使用先验概率，为方便，把最小化转化成最大化，即：

$$\max_{\lambda, A} L(s, A) = T_0 H(s) - \lambda E(s, A) \quad (6)$$

考虑两个约束，并令  $T_0=1$ ，有：

$$\begin{cases} \mathbf{x} = \mathbf{A}\mathbf{s} \\ \sum_{i=1}^n s_i = 1 \end{cases} \quad (7)$$

加入  $\lambda$  约束，得：

$$\lambda E(s, A) = \lambda^T (\mathbf{A}\mathbf{s} - \mathbf{x}) + (\lambda_0 - 1) (\sum_{i=1}^n s_i - 1) \quad (8)$$

$$L(s, A) = -\sum_{i=1}^n s_i \ln(s_i) -$$

$$\sum_{j=1}^m \lambda_j (\sum_{i=1}^n \alpha_{ji} s_i - x_j) - (\lambda_0 - 1) (\sum_{i=1}^n s_i - 1) \quad (9)$$

求得的黎曼空间的Hebb学习规则为：

$$\mathbf{A}(k+1) = \mathbf{A}(k) + \Delta \mathbf{A} = (\mathbf{I} + \eta \lambda \mathbf{X}) \mathbf{A}(k) \quad (10)$$

式(10)表面上看是学习矩阵  $\mathbf{A}$  的迭代，实际上隐含了Lagrange约束  $\lambda$  的生成问题。 $\lambda$  与环境有关，Lagrange约束  $\lambda$  的学习算法由下面的方程给出：

$$\lambda(k+1) = \lambda(k) + \eta (\mathbf{I} - \lambda(k) \lambda(k)) \Delta \lambda \quad (11)$$

$$\Delta \lambda = \mathbf{J}^{-1} \Delta \mathbf{x} \quad (12)$$

式中  $\mathbf{J} = \mathbf{x}\mathbf{x}^T - \mathbf{A}\mathbf{D}\mathbf{A}^T$ ； $\mathbf{D} = \text{diag}(s_i)^n = 1$ 。

### 2.2 LCNN与BSAO的关系

根据式(10)和(11)建立的LCNN框架如图1所示。

(1) 信息流向  $\lambda \Rightarrow [\mathbf{A}] \Rightarrow \mathbf{s} \Rightarrow [\mathbf{A}] \Rightarrow \mathbf{x} \Rightarrow \mathbf{W} \Rightarrow \mathbf{u}$ ， $\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$ ，是BSAO方法；而  $\lambda \Rightarrow [\mathbf{A}] \Rightarrow \mathbf{s}$  是LCNN方法的无监督学习， $\mathbf{s} \Rightarrow [\mathbf{A}] \Rightarrow \mathbf{x}$  是LCNN的有监督学习。(2) BSAO操作在后验  $\mathbf{x}$  概率上，LCNN使用先验  $\mathbf{s}$  的联合概率之上(联合熵最大)。式(12)意味着  $\lambda$  的学习会遇到病态矩阵问题，LCNN的计算复杂度是  $O(n^2)$ 。(3) BSAO的学习矩阵  $\mathbf{W}$  把谱能量数据空间变为密度空间，使得分离的独立成分的概率密度趋于一致，输出  $\mathbf{u}$  与各成分统计不相关，在所有后验信息的整体平均意义上有  $\mathbf{W} = \mathbf{A}^{-1}$ ；LCNN在联合熵  $H(\mathbf{s})$  最大条件下分解  $\mathbf{s}$  的联合概率密度。例如，对于遥感图像信息，LCNN不需要其他像素的信息，多个传感器获得的同一像素的信息就能分解该像素的独立成分，既较好地体现了各像素的空间特性，也能有效地对各个像素互不相关地并发分解，开创了单像素独立成分模型的先河。因此LCNN方法也称为多目(或者多通道)独立成分分析技术。

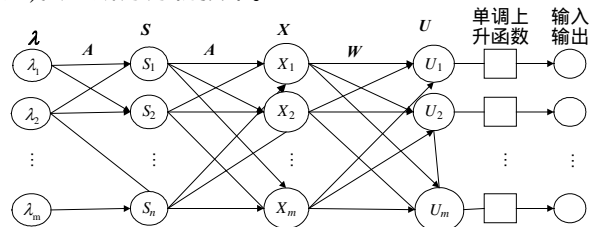


图1 LCNN与BSAO的关系

## 3 自适应LCNN

由式(12)知，后验数据的变化确定约束条件的变化，体现了BSAO的环境自适应性，然而这种反馈引来了病态矩阵求逆的复杂性。通常在LCNN计算之前，应先作白化处理，但白化并不一定能消除病态矩阵。有必要从另一个角度来理解Lagrange约束。

$\mathbf{A}\mathbf{s} - \mathbf{x}$  的有监督变化，实际上就是  $\Delta \mathbf{x}$  的变化，对式(9)进行  $\Delta \mathbf{x}$  求导，并利用动力学方程，有：

$$\frac{\partial(\Delta \mathbf{x})}{\partial t} = -\frac{\partial L}{\partial(\Delta \mathbf{x})} = \lambda \Rightarrow \Delta \mathbf{x}(k+1) = \Delta \mathbf{x}(k) + \eta \Delta \lambda \quad (13)$$

可令：

$$\lambda = \eta \Delta \mathbf{x} = \eta (\mathbf{A}\mathbf{s} - \mathbf{x}) \quad (14)$$

式(13)说明,  $\lambda$  本质是有监督学习的耻降速率。 $\lambda$  的选择至关重要, 如果  $\lambda$  太大, 有监督学习收敛很快, 效果不太好。也许正是有监督学习比无监督学习收敛得快, 造成了病态矩阵问题。将式(14)代入式(10), 则有:

$$A(k+1) = (I + \eta (As - x)x^T)A(k) \quad (15)$$

式(13)~(15)中,  $\eta$  为常数。

至此, 式(15)中不再出现Lagrange约束项, 学习矩阵在  $O(n)$  时间段收敛。这种方法称为自适应LCNN(Adaptive LCNN, ALCNN)。ALCNN最主要的特征如图1和图2所示。

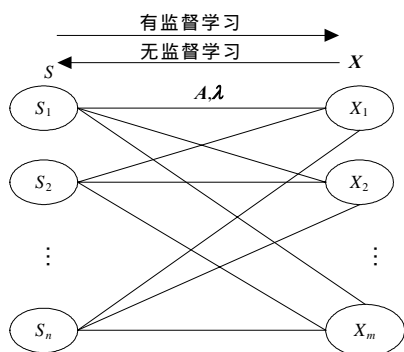


图2 ALCNN的ICA方法

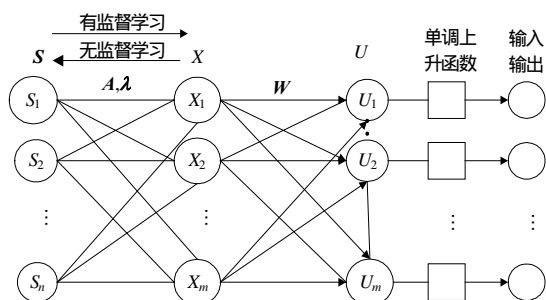


图3 ALCNN与BSAO的关系

图2~3的ALCNN部分是无向图, 包含了前馈和反馈两个并行的神经网络结构, 前馈网络是ALCNN的有监督学习网, 反馈网络是无监督学习网。ALCNN融合了有监督和无监督的并行学习过程。

### 3 小 结

ALCNN不仅学习矩阵收敛快, 而且把图1的先无监督学习、后有监督学习的串行收敛, 改为了有监督学习蕴含在无监督学习中的并发方式, 可以更好地解释“鸡尾酒会问题”。人在酒会中一方面利用

无监督学习对每一个听到的混合音素进行分离; 在音素独立成分分离中, 使用有监督学习的联想机制进行误差校正, 从而更好地识别说话人和内容。这是同时进行的, 不需要顺序操作。

根据BSAO方法的解释, 人在酒会中收集所有听到的信息, 提取相近的分布的独立信息, 分离混合的声音。因此, 这种分离是不精确的。

标准LCNN方法的解释是, 人在酒会中一方面利用无监督学习对每一个听到的混合音素进行分离; 另一方面使用有监督学习的联想机制对分离的独立信息进行识别和校正, 判定其内容, 这是顺序进行的。

### 参 考 文 献

- [1] HERAULT J, JUTTEN C. Space or time adaptive signal processing by neural network models[C]// Neural Networks for Computing, AIP Conference Proceeding, New York, 1986.
- [2] BELL A J, SEJNOWSKI T J. An information maximization approach to blind separation and blind deconvolution[J]. Neural Computation, 1995, 7: 1129-1159.
- [3] LEE T W, SEJNOWSKI T J. Independent component analysis using an extended infomax algorithm for subGaussian and super-Gaussian sources[J]. Neural Computation, 1999, 11(2): 409-433.
- [4] AMARI Shun-Ichi. New learning in structural parameter space—natural riemannian gradient[J]. Advances in Neural Information Processing Systems, 1997, 9: 127-133.
- [5] HYVARINEN A, OJA E. A fast fixed-point algorithm for independent component analysis[J]. IEEE Trans. on Neural Network, 1999, 10(3): 626-634.
- [6] 斯华龄, 张立明. 智能视觉图像处理[M]. 上海: 上海科技教育出版社, 2002.
- [7] HYVARINEN A, PAJUNEN P. Nonlinear independent component analysis: existence and uniqueness results[J]. Neural Networks, 1999, 12(3): 429-439.
- [8] SZU H. ICA—An enabling tech. for intelligent sensory processing[J]. IEEE Circuits and Systems Newsletters, 1999, 12: 14-41.
- [9] SZU H, Hsu C. Meansfield chaos dynamics eith changing chaos threshold values[C]// Int. Conf. on Neural Networks, Houston, 1997.
- [10] SZU H, HSU C. Chaotic neural networks for information processing[C]// Int'L Conf. CCCI, South Africa, 1998.

编 辑 熊思亮