相互情報量最大化に基づくモダリティ間の独立成分抽出

赤穂 昭太郎 (PY)
電子技術総合研究所 情報科学部

Multimodal Independent Component Analysis Maximizing Mutual Information

Shotaro Akaho (akaho@etl.go.jp)
Information Science Division, Electrotechnical Laboratory

Abstract

We propose a new method to extract features from different types of inputs. The algorithm is based on the idea of maximizing the mutual information between the outputs of different network modules. Since the amount of the mutual information does not change by nonsingular transformation, there are infinite number of candidates of feature sets. Among those candidates, the algorithm finds a set of 'independent' features.

1 はじめに

脳の中でパターンがどのように表現されているかという問題は脳の情報処理を考える上で基本的かつ重要な問題である．Barlow は factorial code というものを考え, 情報理論的にこの問題を捉えようと試みた [4]. また最近になって, 独立成分分析 (ICA) が, blind source separation などデータ解析手法として注目されるとともに, 脳のコーディング, 特に初期視覚における特徴抽出細胞の形成に関係しているという考えが提案されている [6].

ICA は視覚あるいは聴覚といった個々のモダリティにおける特徴抽出の説明には適しているが, 異なるモダリティ間の関係といった人間の高次機能の説明には十分でない．例えば, 視覚と聴覚の連合を行なう場合に, 個々のモダリティで抽出した独立成分が連合を行なうために必ずしも重要なものとは限らない．それはモダリティ間の関係を無視しているからである．一方, 従来からある単純な相関学習では次元圧縮の観点が欠けており, また, 複雑な関連性の表現には適さない．

複数の入力からの情報圧縮を行なうための多変量解析手法として正準相関分析 (CCA) と呼ばれる手法がある．しかしながら, CCA は変数間の相関を最大にするという基準 (正規分布の仮定に対応する) に基づいて特徴抽出を行なうため, モダリティ間の複雑な関係を記述するのに十分ではなく, また, 得られる特徴も一意的でないことがあるなどの欠点がある．

最近になって, Becker は相互情報量という観点からコーディングの問題を考察し, 離散分布を中心としたアルゴリズムを提案している [5]. 本発表では, この考え方をより進め, 連続的なドメインにおいて, 正準相関分析に相互情報量と ICA の考え方を取り込むという形で定式化を行ない学習アルゴリズムを提案する．

Keywords: 感覚統合, アルゴリズム, 正準相関分析, 独立成分分析 (ICA)

2 従属成分分析 (MICA) の定式化

未知の源信号の組 $(u_i^*(t), v_i^*(t))$, $i = 1, \ldots, n$ を考える．ここで, $u_i^*(t)$ と $v_i^*(t)$ は各時刻 t で, 統計的に従属であるとし, $\boldsymbol{u}^*(t) = (u_1^*(t), \ldots, u_n^*(t))^{\mathrm{T}}$ と $\boldsymbol{v}^*(t) = (v_1^*(t), \ldots, v_n^*(t))^{\mathrm{T}}$ はそれぞれ独立なベクトルだとする．$(\boldsymbol{u}^*(t), \boldsymbol{v}^*(t))$ は未知の分布 $p(\boldsymbol{u}^*, \boldsymbol{v}^*)$ から生成されたものだとする．

観測値 $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ は源信号に未知の非退化行列 V, W をかけて得られるものとする (簡単のため入力と出力の次元は同じだとした).

$$\boldsymbol{x}(t) = V\boldsymbol{u}^*(t), \qquad \boldsymbol{y}(t) = W\boldsymbol{v}^*(t). \tag{1}$$

MICA の目的は観測信号を線形変換して源信号を推定することである．

$$\boldsymbol{u}(t) = A\boldsymbol{x}(t), \qquad \boldsymbol{v}(t) = B\boldsymbol{y}(t). \tag{2}$$

ただし, 源信号の組の順序と信号の振幅に関する自由度が残るが, 源信号の組の対応を正しく復元したい．

損失関数を構成する際に, モダリティ間の従属性と特徴ベクトルの独立性という二つの基準を考える必要がある．まず従属性については, 負の相互情報量を損失関数としてとる．

$$\mathcal{E}_{\mathrm{dep}} = -\sum_{i=1}^{n} K[p(u_i, v_i) \| p(u_i)p(v_i)]. \tag{3}$$

ただし, K は Kullback-Leibler divergence.

一方, 独立性については, ICA での独立性の基準,

$$\mathcal{E}_{\mathrm{ind}} = K[p(\boldsymbol{u}) \| \prod_{i=1}^{n} p(u_i)] + K[p(\boldsymbol{v}) \| \prod_{i=1}^{n} p(v_i)], \tag{4}$$

を採用する.

MICA では, (CCA と異なり) 必ずしも完全に独立という信号が得られるわけではないので一般に $\mathcal{E}_{\rm dep}$ と $\mathcal{E}_{\rm ind}$ の二つの目的関数の最適化問題となる. このための最も単純な方法として, ここではその線形和

$$\mathcal{E}_{\rm tot} = \mathcal{E}_{\rm ind} + \gamma(t)\mathcal{E}_{\rm dep}, \tag{5}$$

を最小化するように学習を行なうことにする (ここで, $\gamma(t)$ は重み係数).

3　確率的最急降下法

A に関する on-line 更新式は,

$$\begin{aligned}\frac{\partial A}{\partial t} &= -\eta(t)\frac{\partial \mathcal{E}_{\rm tot}}{\partial A}A^{\rm T}A \\ &= \eta(t)\Big[I-(1-\gamma(t))\boldsymbol{\varphi}(\boldsymbol{u})\boldsymbol{u}^{\rm T} \\ &\qquad -\gamma(t)\boldsymbol{\xi}(\boldsymbol{u},\boldsymbol{v})\boldsymbol{u}^{\rm T}\Big]A,\end{aligned} \tag{6}$$

で与えられる (B に関する学習則もほぼ同様な形で得られる). ただし,

$$\boldsymbol{\varphi}(\boldsymbol{u}) = -\frac{\partial p(\boldsymbol{u})}{\partial \boldsymbol{u}}, \tag{7}$$

$$\boldsymbol{\xi}(\boldsymbol{u},\boldsymbol{v}) = -\frac{\partial p(\boldsymbol{u},\boldsymbol{v})}{\partial \boldsymbol{u}}, \tag{8}$$

これは Amari[2] によって提案された natural gradient を用いた勾配法になっている.

$p(\boldsymbol{u})$ および $p(\boldsymbol{u},\boldsymbol{v})$ の近似の仕方によりいろいろなアルゴリズムを作ることができる. 脳のモデルとの観点からはできるだけ局所的かつ on-line で計算でできることなどがポイントとなると考えられるが, 次節では統計的な近似法の一つである Gram-Charlier 展開を用いて近似を行なって導いたアルゴリズムの結果 [1] を示す.

4　シミュレーション

3 個の信号が混ざった観測ベクトルに対しアルゴリズムを適用した. 一方の源信号 $\boldsymbol{u}^*$ は (1) $[-1,1]$ 上の一様乱数, (2) $\sin(2\pi 800t+6\cos(2\pi 60t))$, (3) $\sin(2\pi 90t)$, とし, t は 10kHz($\Delta t = 10^{-4}$) でサンプリングした. 対となるもう一方の信号 $\boldsymbol{v}^*$ は $\boldsymbol{v}^* = (|u_1^*|,|u_2^*|,|u_3^*|)^{\rm T}$ にとった. 源信号を $[-1,1]$ 上の一様分布を要素とするランダム行列で変換したものを観測信号とした. 源信号を一様乱数を要素とした乱数で混合した観測信号を図 1 に示す.

重みの値は徐々に独立性が強くなるように $\gamma(t) = 0.2/(t/\Delta t+1)$ でだんだん小さくなるようにスケジューリングを行なった. 学習係数は $\eta(t) = 1000.0$, $\mu(t) = 50.0$ にとった.

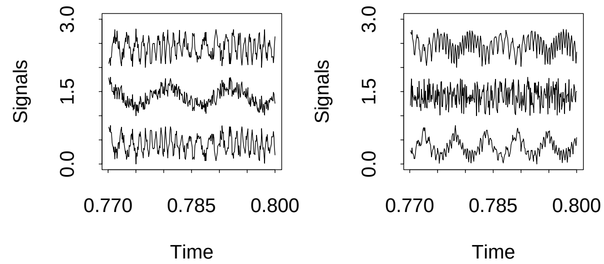

図 1: 混合信号 (左 : x_i, 右 : y_i)

時間区間 $t = [0.77, 0.8]$ における復元信号 $\boldsymbol{u}$, $\boldsymbol{v}$ を図 2 に示す. 左右の対応が (無相関にもかかわらず) 正しく復元されていることがわかる.

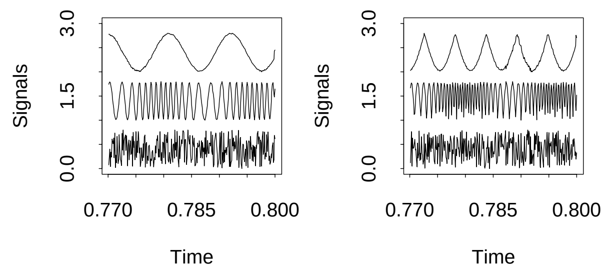

図 2: 復元信号 (左 : u_i, 右 : v_i)

5　おわりに

二つの情報源からの最大相互情報量基準に基づく独立特徴の抽出を行なう手法 MICA の提案を行ない, 簡単な計算機実験により非線形に関連する特徴量の抽出ができることを確認した. 非線形 CCA[3] などとの関連を調べたり, 重み係数を適切に決定する方法, より biologically plausible な学習則の導出などは今後の課題として残されている. また, 実際に生体がどのようなコーディングを行なっているかの検証は生理・心理実験を待つしかない. 一方脳計測データ解析への適用については (1) fMRI とタスクとの関連信号抽出 (2) MEG, EEG など複数データの間の関連, といった応用が考えられる.

参考文献

[1] Akaho, S. et al.: *Proc. of IJCNN* (1999)
[2] Amari, S.: *Neural Computation*, **10**, 251–276 (1998)
[3] Asoh, H. et al.: *Proc. of ICANN*, 713–716 (1994)
[4] Barlow, H. B.: *Neural Computation*, **1**, 295–311 (1989)
[5] Becker, S.: *Network: Computation in Neural Systems*, **7** (1996)
[6] Bell, A.J. et al.: *Vision Research*, **37**, 3327–3338 (1997)