

正準相関分析への独立成分分析の拡張

Extention of Independent Component Analysis to Canonical Correlation Analysis

赤穂 昭太郎*

Shotaro Akaho

梅山 伸二†

Shinji Umeyama

Abstract: Canonical correlation analysis (CCA) extracts features from multiple information source (multi-modality). However, CCA assumes the joint distribution of data is Gaussian. In this article, we propose MICA (multimodal independent component analysis) which generalizes CCA from information-theoretical viewpoint. MICA extracts independent features which maximize mutual information between the modalities. Mutual information between two variables are approximated by 2-D Gram-Charlier expansion and a stochastic gradient method is derived based on a natural gradient. We also show a simple simulation result.

1 はじめに

独立成分分析 (ICA, independent component analysis) は脳計測データの解析の強力なツールとして、また、脳の初期視覚のモデルとして、脳科学をはじめとして信号処理、統計学などの分野で注目を集めている手法である [6, 8].

ICA は多変量分析の中で、主成分分析や因子分析といった外的基準のない場合の分析法 (教師なし学習) として位置付けられるが、本稿では判別分析や正準相関分析といった外的基準のある場合 (教師あり学習) の手法に対して ICA と同様の拡張を試みる。

二つの多変量情報源が与えられたときに、それらに共通に含まれる情報を線形変換によって抽出する問題を考える。このための手法として、よく知られているのが正準相関分析 (CCA, canonical correlation analysis) である [3]. CCA は線形変換してできた特徴ベクトルの各成分の相関係数が最大になるように変換行列を求める。ただし、特徴量の変換の自由度があるので、それぞれの情報源から得られた特徴ベクトルの共分散行列が単位行列になるという制約をおく (sphering condition)。ところ

が、CCA には以下の二つの問題点がある。

1. 相関係数を最大にするという基準は、二つの情報源の同時分布が多変量正規分布をなす場合には、二つの情報源からの最大相互情報量抽出に等価となる。しかしながら、分布が非正規であったり、抽出された特徴が非線形に関連しているような場合には、望ましい特徴抽出ができないことがある。
2. CCA で得られる解は、必ずしも唯一の解を与えない。例えば、相関が等しい二つの特徴量が得られたとすると、その軸内での回転は全く同じ相関構造となる。従って、sphering condition は特徴を得るための十分な制約とは言えず、特徴を分析したり可視化したりする際には障害になる可能性がある。

これらの問題を解決するために、MICA (Multimodal Independent Component Analysis) とよぶ新しい手法を提案する。MICA では相関係数ではなく相互情報量を直接最大化し、sphering condition の代わりに各特徴が統計的に独立になるようにする。

以下ではまず MICA の枠組を定式化し、そのための最急降下法の学習則を導く。さらに簡単な実験結果を示し、応用について考察する。

2 MICA の定式化

観測信号 $x(t) = (x_1(t), \dots, x_n(t))^T$ および $y(t) = (y_1(t), \dots, y_n(t))^T$ が与えられたとしよう。観測信号は

*電子技術総合研究所 情報科学部 〒 305-8568 茨城県つくば市梅園 1-1-4 tel. 0298-54-5549, e-mail akaho@etl.go.jp

Information Science Division, Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba-shi, Ibaraki 305-8568, Japan

†電子技術総合研究所 情報科学部 〒 305-8568 茨城県つくば市梅園 1-1-4 tel. 0298-54-5837, e-mail umeyama@etl.go.jp

Information Science Division, Electrotechnical Laboratory, 1-1-4 Umezono, Tsukuba-shi, Ibaraki 305-8568, Japan

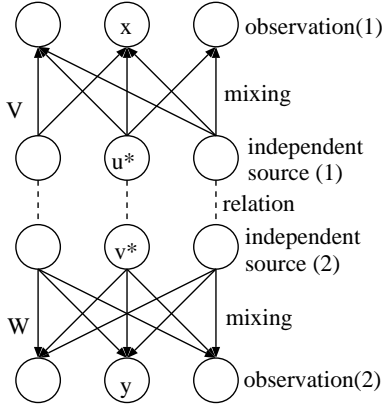


図 1: MICA

強定常で、確率分布 $p(\mathbf{x}, \mathbf{y})$ に従うものとする。

観測信号を線形変換して、相互情報量を最大とするような特徴量 $\mathbf{u}(t) = (u_1(t), \dots, u_n(t))^T$ および $\mathbf{v}(t) = (v_1(t), \dots, v_n(t))^T$ を抽出したい¹。

$$\mathbf{u}(t) = A\mathbf{x}(t), \quad \mathbf{v}(t) = B\mathbf{y}(t). \quad (1)$$

ただし、 $(u_i(t), v_i(t))$ の組は pairwise に統計的に独立であるとする。すなわち、

$$p(\mathbf{u}, \mathbf{v}) = \prod_{i=1}^n p(u_i, v_i)$$

が成り立つとする。

見方を変えるとこれは次のように言い換えることができる。未知の源信号のペア $(u_i(t), v_i(t))$, $i = 1, \dots, n$ があるとする (図1)。 $u_i(t)$ と $v_i(t)$ の間は統計的に依存関係があり、一方ペアの集合 $(u_i(t), v_i(t))$ は互いに統計的に独立であるとする。観測値ベクトル $\mathbf{x} = (x_1, \dots, x_n)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$ は源信号に未知の混合行列 V, W をかけて得られるものとみなし、

$$\mathbf{x}(t) = V\mathbf{u}(t), \quad \mathbf{y}(t) = W\mathbf{v}(t), \quad (2)$$

とする。MICA は $\mathbf{x}(t), \mathbf{y}(t)$ から復元行列 A, B によって源信号を (振幅とペアの順序の自由度を除き) 復元するものと考えることができる。

3 損失関数

MICA は、特徴ベクトルの pairwise な独立性を制約条件として相互情報量を最大にする (負の相互情報量を最小化する) という問題となる。

¹簡単のため観測信号と特徴量の次元は同じであるとする。特徴量の次元が小さい場合は ICA で行なわれている議論を流用できると考えられる

負の相互情報量は、pairwise な独立性から

$$\begin{aligned} \mathcal{E}_{\text{dep}} &= -K[p(\mathbf{u}, \mathbf{v}) \| p(\mathbf{u})p(\mathbf{v})] \\ &= -\sum_{i=1}^n K[p(u_i, v_i) \| p(u_i)p(v_i)]. \end{aligned} \quad (3)$$

となる。ここで、 K は Kullback-Leibler divergence で

$$K[p(z) \| q(z)] = \int p(z) \log \frac{p(z)}{q(z)} dz.$$

一方、pairwise な独立性は、

$$\tilde{\mathcal{E}}_{\text{ind}} = K[p(\mathbf{u}, \mathbf{v}) \| \prod_{i=1}^n p(u_i, v_i)] \quad (4)$$

が 0 であることと等価である。

すると、Lagrange の未定係数 λ を導入し、

$$\tilde{\mathcal{E}}_{\text{tot}} = \mathcal{E}_{\text{dep}} + \lambda \tilde{\mathcal{E}}_{\text{ind}} \quad (5)$$

の最小化問題解くことに帰着されるが、以下で述べる理由から実際には、これを近似した

$$\mathcal{E}_{\text{tot}} = \gamma_t \mathcal{E}_{\text{dep}} + \mathcal{E}_{\text{ind}} \quad (6)$$

という式の最小化を行なう。ただし、

$$\mathcal{E}_{\text{ind}} = K[p(\mathbf{u}) \| \prod_{i=1}^n p(u_i)] + K[p(\mathbf{v}) \| \prod_{i=1}^n p(v_i)] \quad (7)$$

で、これは $\tilde{\mathcal{E}}_{\text{ind}}$ よりも弱い条件 (つまり $\tilde{\mathcal{E}}_{\text{ind}} = 0$ なら $\mathcal{E}_{\text{ind}} = 0$) になっている。

CCA では独立性は sphering condition に対応し、どのような信号が与えられたとしても常にそれを満たす解が存在したが、MICA では、必ずしも完全に独立であるような信号が存在するわけではないので上記の問題は一般に \mathcal{E}_{dep} と $\tilde{\mathcal{E}}_{\text{ind}}$ の二つの目的関数の最適化問題となる。その場合には λ は Lagrange の未定係数としてではなく、二つの目的関数に対する重みを決める係数として働く。本論文でも λ (実際には γ_t) はユーザが決める定数として扱う。また、pairwise に独立な解を求めることは実際上困難であるので $\tilde{\mathcal{E}}_{\text{ind}}$ の代わりに \mathcal{E}_{ind} を用いる。また、アルゴリズムを ICA のものと比較しやすくするために、全体に γ_t/λ を乗じてある。

\mathcal{E}_{tot} をエントロピーを用いて書き直すと、

$$\begin{aligned} \mathcal{E}_{\text{tot}} &= -H(\mathbf{x}) - H(\mathbf{y}) + \log |A| + \log |B| \\ &\quad + (1 - \gamma_t) \sum_{i=1}^n \{H(u_i) + H(v_i)\} \\ &\quad + \gamma_t \sum_{i=1}^n H(u_i, v_i), \end{aligned} \quad (8)$$

となる。以下ではこの定式化に基づいて確率的最急降下法を構成する。

4 損失関数の近似

4.1 無相関化

一般性を失うことなく, $E[u_i] = E[v_i] = 0$, $\text{Var}[u_i] = \text{Var}[v_i] = 1$ を仮定する. そうでない場合にはシフトとスケーリングにより正規化できる.

ここではまず (u_i, v_i) を無相関な変数 (r_i, s_i) に変換する. この操作により, 同時エントロピー $H(u_i, v_i)$ を相関に関係する項とそれ以外の項に分離することができる. 大きな相関は相互情報量を大きくするための十分条件となっており, また, 次節で行なう Gram-Charlier 展開による近似がより簡単になる.

(u_i, v_i) から (r_i, s_i) への変換として

$$\begin{pmatrix} r_i \\ s_i \end{pmatrix} = \begin{pmatrix} c_i^+ & c_i^- \\ c_i^- & c_i^+ \end{pmatrix} \begin{pmatrix} u_i \\ v_i \end{pmatrix} \quad (9)$$

をとる. ここで,

$$\begin{pmatrix} c_i^+ \\ c_i^- \end{pmatrix} = \frac{1}{2} \begin{pmatrix} (1 + \rho_i)^{-1/2} + (1 - \rho_i)^{-1/2} \\ (1 + \rho_i)^{-1/2} - (1 - \rho_i)^{-1/2} \end{pmatrix},$$

$$\rho_i = E[u_i v_i].$$

この変換により (r_i, s_i) は無相関となり, 同時エントロピーは

$$H(u_i, v_i) = H(r_i, s_i) + \frac{1}{2} \log(1 - \rho_i^2), \quad (10)$$

となる.

4.2 2次元エントロピーの近似

無相関な2変数の確率分布に対する Gram-Charlier 展開を4次までで打ち切ったものは次のようになる.

$$p(r_i, s_i) = \phi_i \{ 1 + \frac{1}{3!} (\beta_i^{3,0} h_i^{3,0} + 3\beta_i^{2,1} h_i^{2,1} + 3\beta_i^{1,2} h_i^{1,2} + \beta_i^{0,3} h_i^{0,3}) + \frac{1}{4!} (\beta_i^{4,0} h_i^{4,0} + 4\beta_i^{3,1} h_i^{3,1} + 6\beta_i^{2,2} h_i^{2,2} + 4\beta_i^{1,3} h_i^{1,3} + \beta_i^{0,4} h_i^{0,4}) \},$$

ここで, ϕ_i は r_i と s_i の2次元標準正規分布を表し, $\beta_i^{k,l}$ は次のように定義される(無相関変数の)同時キュムラントである.

$$\beta_i^{k,l} = E[(r_i)^k (s_i)^l] - \beta_0^{k,l}, \quad (11)$$

ただし, $k + l \leq 4$ で

$$\beta_0^{k,l} = \begin{cases} 3 & \text{if } k = 4 \text{ or } l = 4, \\ 1 & \text{if } k = l = 2, \\ 0 & \text{otherwise,} \end{cases}$$

また $h_i^{k,l}$ は (k, l) 次の2変数エルミート多項式で

$$h_i^{k,l} = \frac{(-1)^{k+l}}{\phi_i} \frac{\partial^{k+l}}{\partial r_i^k \partial s_i^l} \phi_i.$$

これを使って r_i と s_i の同時エントロピーは

$$H(r_i, s_i) \simeq H(\beta_i) = 1 + \log 2\pi - \frac{1}{2 \cdot 3!} \left\{ (\beta_i^{3,0})^2 + 3(\beta_i^{2,1})^2 + 3(\beta_i^{1,2})^2 + (\beta_i^{0,3})^2 \right\} - \frac{1}{2 \cdot 4!} \left\{ (\beta_i^{4,0})^2 + 4(\beta_i^{3,1})^2 + 6(\beta_i^{2,2})^2 + 4(\beta_i^{1,3})^2 + (\beta_i^{0,4})^2 \right\}, \quad (12)$$

と近似できる².

4.3 確率的最急降下法

MICA における評価基準を確率的最急降下法により最小化する. A に関する更新式は,

$$\begin{aligned} \frac{\partial A}{\partial t} &= -\eta_t \frac{\partial \mathcal{E}_{\text{tot}}}{\partial A} A^T A \\ &= \eta_t \left[I - (1 - \gamma_t) \varphi^A(\mathbf{u}) \mathbf{u}^T - \gamma_t \xi^A(\mathbf{u}, \mathbf{v}) \mathbf{u}^T + \gamma_t \text{diag}[\bar{\rho} - \psi^A(\mathbf{u}, \mathbf{v})] R^T \right] A, \quad (13) \end{aligned}$$

という形になる (B に関する学習則もほぼ同様な形で得られる). 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. 上記の式において, φ^A は ICA においてもあらわれる項であり, ここでは Yang and Amari[11] によって提案された Gram-Charlier 展開による近似を用いた更新式を用いる.

一方, φ^A 以降の項は MICA 固有の部分であり, $\bar{\rho}$, R , ξ^A , ψ^A はそれぞれ以下の節の (15), (16), (21), (22) で与えられる.

(13) 式は on-line 型あるいは確率的勾配法のアルゴリズムであり, それは勾配を計算した後全体にかかる期待値オペレータ E を除くことによって得られる (E を残すと batch 型の学習則になる).

4.3.1 相関に関する項

(10) 式の中で相関係数に関する項 $(1/2) \log(1 - \rho_i^2)$ の勾配を考えることにより

$$-\left\{ \sum_i \frac{1}{2} \frac{\partial}{\partial A} \log(1 - \rho_i^2) \right\} A^T A = \text{diag}[\bar{\rho}] R^T A \quad (14)$$

² キュムラントの2次式までで打ち切ったが, 3次式を加えるとさらに22個の項が加わる. どこまでを用いるかは必要な近似精度や計算量に依存する

が得られる。ここで

$$\bar{\rho}_i = \frac{\rho_i}{1 - \rho_i^2}. \quad (15)$$

また, R は u と v の共分散行列,

$$R = E[uv^T]. \quad (16)$$

である. ρ_i は R の対角成分 R_{ii} になっている.

4.3.2 相関以外の項

$H(r_i, s_i)$ の近似式 $H(\beta_i)$ の勾配を考えることにより,

$$\frac{\partial H(\beta_i)}{\partial A} A^T A = E \left[f_i \frac{\partial r_i}{\partial A} A^T + g_i \frac{\partial s_i}{\partial A} A^T \right] A \quad (17)$$

となる. ただし,

$$\begin{pmatrix} f_i \\ g_i \end{pmatrix} = \sum_{k,l \in \mathcal{K}} \frac{\partial H(\beta_i)}{\partial \beta_i^{k,l}} \frac{\partial (r_i)^k (s_i)^l}{\partial (r_i, s_i)^T}, \quad (18)$$

ただし \mathcal{K} は (12) 式に含まれる k, l のすべての添え字集合をあらわす. $\partial H(\beta_i) / \partial \beta_i^{k,l}$ は $\beta_i^{k,l}$ の多項式でとなる.

また

$$\left(\frac{\partial r_i}{\partial A} A^T \right)_{ij} = (d_i^+ u_i + d_i^- v_i)(R)_{ji} + c_i^+ u_j, \quad (19)$$

$$\left(\frac{\partial s_i}{\partial A} A^T \right)_{ij} = (d_i^- u_i + d_i^+ v_i)(R)_{ji} + c_i^- u_j, \quad (20)$$

ただし,

$$\begin{aligned} \begin{pmatrix} d_i^+ \\ d_i^- \end{pmatrix} &= \frac{\partial}{\partial \rho_i} \begin{pmatrix} c_i^+ \\ c_i^- \end{pmatrix} \\ &= -\frac{1}{4} \begin{pmatrix} (1 + \rho_i)^{-3/2} + (1 - \rho_i)^{-3/2} \\ (1 + \rho_i)^{-3/2} - (1 - \rho_i)^{-3/2} \end{pmatrix}. \end{aligned}$$

従って, ξ^A, ψ^A の第 i 成分はそれぞれ

$$\xi_i^A = c_i^+ f_i + c_i^- g_i, \quad (21)$$

$$\psi_i^A = (d_i^+ u_i + d_i^- v_i) f_i + (d_i^- u_i + d_i^+ v_i) g_i. \quad (22)$$

4.4 統計量の on-line 推定

前節で導いた学習則は on-line 学習の際には未知のキムラント $\beta_i^{k,l}$ や共分散行列 R に依存した式になっている. これらの値は次のように on-line で推定することができる.

$$\frac{d\beta_i^{k,l}}{dt} = -\mu_t \{ \beta_i^{k,l} - (r_i)^k (s_i)^l + \beta_0^{k,l} \}, \quad (23)$$

$$\frac{dR_{ij}}{dt} = -\mu_t (R_{ij} - u_i v_j), \quad (24)$$

ここで μ_t は学習係数をあらわす.

5 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 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 $10\text{kHz} (\Delta t = 10^{-4})$ でサンプリングした. 対となるもう一方の信号 v は $v = (|u_1|, |u_2|, |u_3|)^T$ にとった. u と v の一部を図 2 に示す. 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.

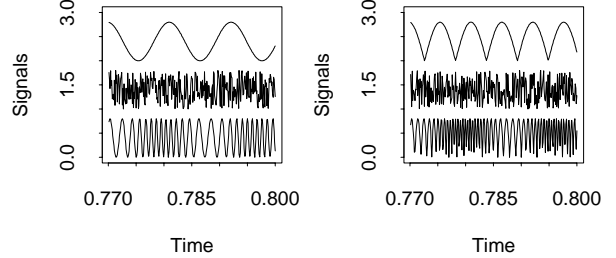


図 2: Original signals (left : u_i , right : v_i)

この信号復元に対して以下の二つのアルゴリズムを比較した.

- (1) Alg-ICA: ICA を x と y に対し独立に行なう,
- (2) Alg-MICA: MICA を行なう

まず, 全体のデータを CCA を用いて sphering を行なった³. u_i と v_i の相関はほぼ 0 となるため, CCA を用いた sphering はこの実験では PCA で各信号をそれぞれ独立に sphering するのと大きな差はない. この前処理を行なった観測信号を図 3 に示す.

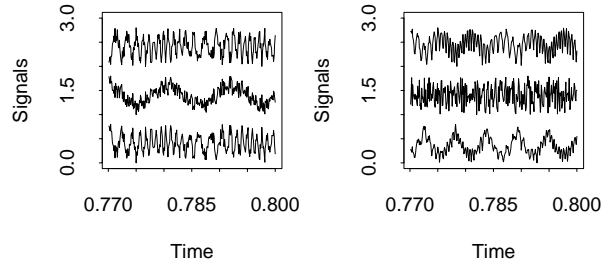


図 3: Mixed signals (left : x_i , right : y_i)

続いて前処理を施した観測信号に対してそれぞれのアルゴリズムを on-line 学習則で実行した. 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.

³これは完全に on-line 学習とするセットアップでは不可能だが, CCA との比較のために行なった.

学習係数は $\eta_t = 1000.0$, $\mu_t = 50.0$ にとった.

時間区間 $t = [0.77, 0.8]$ における復元信号 u, v を図 4 (Alg-ICA) と図 5 (Alg-MICA) に示す. 後者では復元信号の対応がうまくとれていることがわかる.

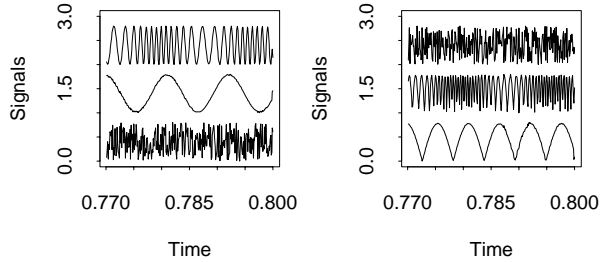


図 4: Recovered signals (Alg-ICA. left : u_i , right : v_i)

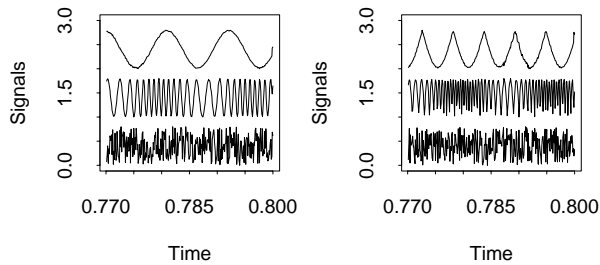


図 5: Recovered signals (Alg-MICA. left : u_i , right : v_i)

以下ではアルゴリズムの定量的な比較を試みる. まず, 独立性の復元の尺度として以下のものがよく用いられる.

$$\mathcal{E}_{ICA} = \sum_{i=1}^n \left(\sum_{j=1}^n \frac{|P_{ij}|}{\max_k |P_{ik}|} - 1 \right) + \sum_{i=1}^n \left(\sum_{j=1}^n \frac{|P_{ij}|}{\max_k |P_{kj}|} - 1 \right), \quad (25)$$

ここで P は混合行列と復元行列の積で, MICA では各情報源ごとに $P^A = AV$ あるいは $P^B = BW$ が対応する. \mathcal{E}_{ICA} は P が置換と振幅の自由度を除いて単位行列に等しい時に 0 となる.

一方, 従属性の尺度として, 以下のような尺度を考える.

$$\mathcal{E}_{MICA} = \sum_{i=1}^n \sum_{j=1}^n \left| \frac{P^A_{ij}}{\max_k |P^A_{ik}|} - \frac{P^B_{ij}}{\max_k |P^B_{ik}|} \right|, \quad (26)$$

これは P^A と P^B がスケーリングを除いて一致した時に 0 となる.

\mathcal{E}_{ICA} をプロットしたものが図 6 である. 独立性だけに特化した Alg-ICA の方が収束の後段では Alg-MICA よりも小さな値になる傾向があるが, それほど大きな差はない. 一方 \mathcal{E}_{MICA} を図 7 にプロットした. 対応の取れていない Alg-ICA では大きな値に収束してしまう.

もちろん, Alg-ICA が正しい対応の所に収束することもあるが, あくまでそれは偶然の一致にすぎず, 信号の個数が増えるにつれてその比率は小さくなると思われる. また, Alg-MICA でもうまく対応がとれない場合もある. それは確率分布を近似していることと, 対応が取れていないで独立な解が (特に γ_t が小さい時) 強いローカルミニマムになってしまうからである. このためには十分ゆっくり γ_t を小さくしていく必要があり, on-line 学習の際には計算時間との trade-off が問題となる.

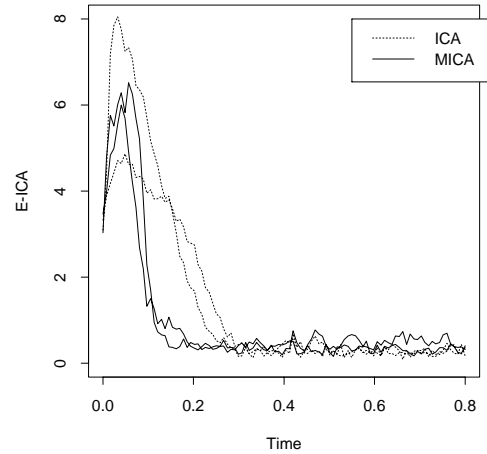


図 6: Error index on independence \mathcal{E}_{ICA} .

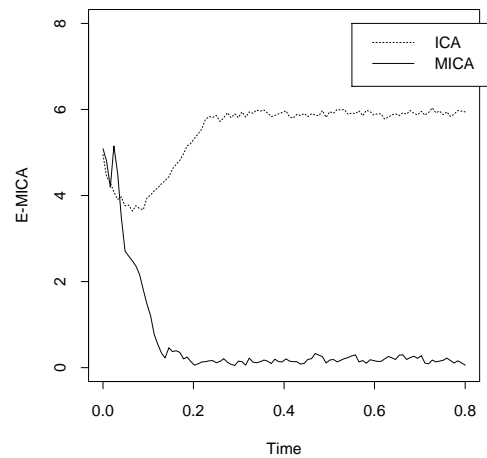


図 7: Error index on dependence \mathcal{E}_{MICA} .

6 MICA の応用の可能性

MICA では CCA における相関を相互情報量に一般化し, sphering condition を独立性として一般化した. 実際の応用の際には目的に応じてその一部だけを採用することも考えられる. MICA の特殊な場合として, 対応成分の相関を最大化するような独立成分を取り出すとい

う観点から、教師付き独立成分分析 (SICA, Supervised ICA) の枠組みを梅山ら [9, 10] が提案している⁴。これは従来の部分空間法や判別分析を拡張したものになっている。アルゴリズムを実際に適用した例として、表情ラベルや個体ラベルのついた顔の画像データをもとにし、これらのラベルに応じた独立な特徴を抽出する試みを行っており、これは固有顔 (Eigen face) などと呼ばれている手法を拡張したもので、可視化や識別課題に対して有効に働くと考えられる。

また、脳計測データ (fMRI) においてタスクを教師信号として、タスクに関連の深い脳画像成分を選択的に取り出そうとする試みも行われており、Jung ら [7] が ICA で行なっている手法よりも高い確度で特徴量を得られることが期待できる。

その他、MICA の場合はたとえば 2 種類の脳計測データ (EEG と MEG) があつたときに、その二つに共通して含まれている成分を取り出すことができれば、余分な artifact を自然に除去することができると考えられる。

一方、最初にも述べたように ICA は初期視覚のモデルとして注目を集めている。また Becker らは脳の高次における情報処理のモデルとして二つの離散情報源の相互情報量を最大化する問題を考えている [5]。MICA はこれらの手法を統合したものとみなすこともでき、脳の低次機能の上位概念として高次機能モデルを構築できる可能性をもつ。

更に工学的応用として、ユーザインタフェースの分野において、複数のモダリティの情報をを用いたマルチモーダルインタフェースの研究が盛んだが、そのようなインタフェースシステムが学習を行なうような枠組 [1] においても有効な特徴抽出の手段となると考えられる。

7 まとめと今後の課題

非正規的な二つの情報源からの最大相互情報量抽出を行なう手法 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。

非線形な特徴を抽出する他のアプローチとしては他に非線形 CCA[4] があり、目的によっては有効と考えられるが、得られる特徴が独立ではなく、また、正規性の仮定が常に残ってしまう。

MICA は独立成分分析と正準相関分析を情報論的に統合した手法と考えられる。しかしながら、現在の所 2 次元分布の近似が 1 次元分布の近似よりも難しいという理由で本来の $\tilde{\mathcal{E}}_{\text{tot}}$ ではなく、 \mathcal{E}_{tot} の最適化で置きか

えてしまっている。これらの解析や重み係数 γ_t の決定法は収束性と関係して重要な問題である。これらは実際のデータへの適用等を含め今後の課題として残されている。

参考文献

- [1] 赤穂, 速水, 長谷川, 吉村, 麻生: E M法を用いた複数情報源からの概念獲得, 電子情報通信学会論文誌, Vol. J80-A, No. 9, pp. 1546-1553, 1997.
- [2] Amari, S.: Natural gradient works efficiently in learning, *Neural Computation*, Vol. 10, pp. 251-276, 1998.
- [3] Anderson, T.W.: *An Introduction to Multivariate Statistical Analysis — Second edition*, John Wiley & Sons, 1984.
- [4] Asoh, H., Takechi, O.: An approximation of Non-linear Canonical Correlation Analysis by Multi-layer Perceptrons, *Proc. of Int. Conf. Artificial Neural Networks*, pp. 713-716, 1994.
- [5] Becker, S.: Mutual Information Maximization: Models of Cortical Self-Organization, *Network: Computation in Neural Systems*, Vol. 7, No. 1, 1996.
- [6] Bell, A.J., Sejnowski, T.J.: The ‘independent components’ of natural scenes are edge filters, *Vision Research*, Vol. 37, pp. 3327-3338, 1997.
- [7] Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M., Iragui, V., Sejnowski, T.J.: Extended ica removes artifacts from electroencephalographic recordings, *Advances in Neural Information Processing Systems 10*, 1998.
- [8] Lee, T.-W.: *Independent Component Analysis — Theory and Applications*, Kluwer, 1998.
- [9] 梅山, 赤穂, 菅生: 教師付き独立成分分析の提案とその顔画像分析への応用について, 信学技報 NC 99-2, 1999.
- [10] 梅山, 栗田, 赤穂: 教師付き独立成分分析を用いた顔の同定について, 信学技報 PRMU 99, 1999.
- [11] Yang, H.-H., Amari, S.: Adaptive Online Learning Algorithms for Blind Separation: Maximum Entropy and Minimum Mutual Information, *Neural Computation*, Vol. 9, pp. 1457-1482, 1997.

⁴SICA では混合行列との相関を考えている (問題としては双対関係にある) ので MICA に完全に含まれるというわけではない