

拡張ペアワイズ表現を用いた一般化多変量解析

木村 昭悟[†] 杉山 将^{††} 亀岡 弘和^{†,†††} 坂野 鋭[†]

[†] 日本電信電話 (株) コミュニケーション科学基礎研究所 〒 619-0237 京都府相楽郡精華町光台 2-4

^{††} 東京工業大学 大学院情報理工学研究科 〒 152-8552 東京都目黒区大岡山 2-12-1

^{†††} 東京大学 大学院情報理工学系研究科 〒 113-8656 東京都文京区本郷 7-3-1

E-mail: [†]akisato@ieee.org, ^{††}sugi@cs.titech.ac.jp

あらまし 主成分分析 (PCA)・Fisher 線形判別分析 (FDA)・線形回帰分析 (MLR)・正準相関分析 (CCA) など、多変量解析に基づく次元削減手法、及びカーネルトリックを用いたそれらの非線形拡張が (一般化) 固有値問題を解くことで実現されることは広く知られている。本論文では、まず、上記の多変量解析を統一的な視点で捉えるための枠組として、拡張ペアワイズ表現と呼ぶ共分散行列やグラム行列に関する新しい表現形式を導入した多変量解析の一般化について議論する。この枠組は、上記一連の多変量解析だけではなく、 L_2 ノルム正則化・graph Laplacian・局所線形判別分析 (LFDA) など解に望ましい性質を加える各種方法や、教師情報が不足している状況下での半教師付き次元削減をも、その特殊例として含む汎用的かつコンパクトな表現形式となっている。本論文では、この枠組の中で具体的な解析手法を設計する指針についても提示する。この指針では、上記の標準的な多変量解析手法やその各種拡張を設計のテンプレートとして利用し、それを適切に組み合わせるだけで解析手法を設計できることを示している。さらに、この設計指針に基づいて具体的に生成した新しい多変量解析手法を提案すると共に、その有効性を一般物体認識を題材として示す。

キーワード 多変量解析, 次元削減, ペアワイズ表現, 一般化固有値問題, 正則化, 半教師付き学習

1. はじめに

画像・映像だけでなく、購買履歴・経済データ・地球観測データなどのデータが大量にオンラインで容易に入手できるようになる一方で、その大量さゆえにデータにどのような傾向や性質があるのかを読み取ることが困難になっている。これら大量のデータの中に隠れている構造や情報を抽出するために、多変量解析が強力な道具の 1 つとして活用されている。

標準的な多変量解析手法として、主成分分析 (PCA) [1]・Fisher 判別分析 (FDA) [2]・線形回帰分析 (MLR)・正準相関分析 (CCA) [1] などが知られている [3], [4]。上記の多変量解析手法は、多次元ベクトルで与えられるサンプルの共分散行列もしくはそれを組み合わせた拡大行列の (一般化) 固有値問題として定式化される。

また、非ベクトルサンプル・非線形解析に対応するために、従来の多変量解析にカーネル法を導入したカーネル多変量解析についても、数多くの研究がなされている [4]~[9]。カーネル多変量解析では、共分散行列に代えて所定のカーネルによって決まるグラム行列が用いられ、その (一般化) 固有値問題によって解析が実現される。カーネル法の導入によって必要以上に大きくなった表現能力を抑える正則化についても各種提案されており、代表的な手法として、Ridge 回帰 [10]・Tikhonov 正則化 [11], [12]・graph Laplacian [13] などが知られている。

さらに、局所性保存射影 (LPP) [14] や局所線形判別分

析 (LFDA) [15] に代表される外れ値や多峰性分布に対する頑健性向上や、半教師付き判別分析 [16] や半教師付き正準相関分析 [13], [17] などに代表される半教師付き次元削減などの拡張も各種なされている。

現在のデータマイニングにおけるもう 1 つの特徴は、目的やデータの性質が複雑で多様化している点にある。このような場合、どこにどのような構造が隠れているかが不明確であるため、そもそもどの手法を用いるべきかがはっきりしない点が問題となる。故に、マイニングの実施者がそれぞれの目的に応じてテラーメイドのかつ簡易に様々な解析手法を設計できる状況が望ましい。

上記の議論に基づき、本論文では、まず、拡張ペアワイズ表現と呼ぶ、共分散行列やグラム行列に関する新しい表現形式を導入することにより、既存の多変量解析手法やその拡張を統一的な視点で捉え直す枠組を提供する。この枠組を、本論文では以降、簡単のため、一般化多変量解析と呼ぶこととする。本論文の後半では、一般化多変量解析の枠組の中で所望の解析手法を具体的に設計する指針について議論する。この指針が示すところは、上記の標準的な多変量解析手法やその各種拡張を設計のテンプレートとして利用し、それを適切に組み合わせるだけでも解析手法を設計できるという点にある。この設計指針に基づいて具体的に新しい多変量解析手法を比較的容易に生成することができる。その具体例をいくつか提案すると共に、その有効性を一般物体認識を題材として示す。

2. ベクトルデータに対する多変量解析

多変量解析の対象として、2種類^(注1)のサンプル集合 X と Y を考える．各サンプルはそれぞれ d_x 次元及び d_y 次元の多次元列ベクトルで構成されており、その数はそれぞれ N_x 個及び N_y 個である．同じ下付添字を持つサンプルが共起するサンプル対を表すとすると、各サンプル集合は、以下のように、共起サンプル集合 (X_C, Y_C) と非共起サンプル集合 X_I, Y_I に分類される．

$$\begin{aligned} X &= X_C \cup X_I, \quad Y = Y_C \cup Y_I, \\ X_C &= \{\mathbf{x}_1, \dots, \mathbf{x}_N\} = \{\mathbf{x}_{C,1}, \dots, \mathbf{x}_{C,N}\}, \\ Y_C &= \{\mathbf{y}_1, \dots, \mathbf{y}_N\} = \{\mathbf{y}_{C,1}, \dots, \mathbf{y}_{C,N}\}, \\ X_I &= \{\mathbf{x}_{N_x+1}, \dots, \mathbf{x}_{N_x+N_y}\} = \{\mathbf{x}_{I,1}, \dots, \mathbf{x}_{I,N_x-N_y}\}, \\ Y_I &= \{\mathbf{y}_{N_y+1}, \dots, \mathbf{y}_{N_x+N_y-N}\} = \{\mathbf{y}_{I,1}, \dots, \mathbf{y}_{I,N_y-N}\}. \end{aligned}$$

本論文では、サンプル平均を 0 ベクトルであると仮定する．この仮定により一般性が失われることはない．以下、特に注釈がない場合を除き $N_x = N_y = N$ とする．

これまでに知られている多変量解析の多くは、下記に示す関数を最大化する基底ベクトル $\mathbf{w}^{(opt)}$ を求める問題として定式化される．

$$\mathbf{w}^{(opt)} = \arg \max_{\mathbf{w} \in \mathcal{R}^d} \mathbf{w}^\top \overline{\mathbf{C}} \mathbf{w} (\mathbf{w}^\top \underline{\mathbf{C}} \mathbf{w})^{-1}, \quad (1)$$

ここで、行列 $\overline{\mathbf{C}}$ 及び $\underline{\mathbf{C}}$ は何らかの統計的性質を表現する行列である．例えば、主成分分析 (PCA) では $\overline{\mathbf{C}}$ が X の共分散行列かつ $\underline{\mathbf{C}}$ が単位行列、線形判別分析 (FDA) では $\overline{\mathbf{C}}$ がクラス間共分散行列かつ $\underline{\mathbf{C}}$ がクラス内共分散行列となる．式 (1) に示す最大化は、行列 $\overline{\mathbf{C}}$ に対応する性質を強調し、行列 $\underline{\mathbf{C}}$ に対応する性質を抑制する操作として直感的には理解される．この最大化において、基底ベクトル \mathbf{w} にはスケール任意性が残っているため、この任意性を取り除いた以下の形式が利用される．

$$\begin{aligned} \mathbf{w}^{(opt)} &= \arg \max_{\mathbf{w} \in \mathcal{R}^d} R_1(\mathbf{w}) \quad \text{s.t.} \quad R_2(\mathbf{w}) = 1, \\ R_1(\mathbf{w}) &= \mathbf{w}^\top \overline{\mathbf{C}} \mathbf{w}, \quad R_2(\mathbf{w}) = \mathbf{w}^\top \underline{\mathbf{C}} \mathbf{w}. \end{aligned}$$

この制約付最大化問題を Lagrange 未定乗数法を用いて解くことで、以下の一般化固有値問題をj得る．

$$\overline{\mathbf{C}} \mathbf{w} = \lambda \underline{\mathbf{C}} \mathbf{w}, \quad (2)$$

この一般化固有値問題を解いた結果として得られた固有ベクトルの集合から、固有値の大きい順に r 本選択した固有ベクトル \mathbf{w}_k ($k = 1, 2, \dots, r$) が、当初の多変量解析において求めるべき基底ベクトルとなる．

(注1): 3種類以上のサンプル集合に対しても以降の議論は同様に成り立つが、簡単のため本論文では2種類の場合に限定して説明する．

3. 一般化多変量解析

3.1 拡張ペアワイズ表現

式 (2) で解析の対象となる行列として共分散行列がしばしば用いられる．^(注2)

$$S_{xy} = N^{-1} \sum_{n=1}^N \mathbf{x}_n \mathbf{y}_n^\top = N^{-1} \mathbf{X}^\top \mathbf{Y}.$$

また、局所性保存射影 (LPP) [14] や局所線形判別分析 (LFDA) [15] では、サンプル間の類似性を考慮した共分散行列の拡張として下記の2次統計量を用いられる．

$$S_{Q,xy} = \sum_{n_1=1}^N \sum_{n_2=1}^N Q_{n_1, n_2} (\mathbf{x}_{n_1} - \mathbf{x}_{n_2}) (\mathbf{y}_{n_1} - \mathbf{y}_{n_2})^\top.$$

ただし、行列 Q は任意の $N \times N$ 非負値対称行列とする．ここで、 $N \times N$ 対角行列 D を $D_{n,n} = \sum_{n_2=1}^N Q_{n, n_2}$ 、行列 Q によって決まる $N \times N$ 行列 L_Q を $L_Q = D - Q$ と置くと、2次統計量 $S_{Q,xy}$ は行列 L_Q を用いて以下のように表現できる．

$$S_{Q,xy} = \mathbf{X} L_Q \mathbf{Y}^\top.$$

この表現を、2次統計量 $S_{Q,xx}$ のペアワイズ表現 [16] と呼ぶ．行列 L_Q は一般に Laplacian 行列と呼ばれ、半正定値行列になることが知られている．また、行列 Q がサンプル数 N と同数の頂点数を持つグラフの辺の重みを保持する場合、行列 L_Q は特に graph Laplacian と呼ばれる．

本論文では、このペアワイズ表現に行列 Q と独立な項を追加した以下の表現を新たに導入する．

$$\hat{S}_{Q,xy} = \mathbf{X} L_{Q,1} \mathbf{Y}^\top + L_2. \quad (3)$$

ただし、 $L_{Q,1}$ は行列 Q によって決まる $N \times N$ 半正定値対称行列、 L_2 は行列 Q とは独立に決まる $d_x \times d_y$ 非負値行列である．式 (3) に示す表現を以降、拡張ペアワイズ表現と呼ぶ．

3.2 一般化多変量解析の設計指針

ここで、拡張ペアワイズ表現を用いた、以下の一般化固有値問題で定式化される解析を考える．

$$\overline{\mathbf{C}}_{\overline{Q}} \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix} = \lambda \underline{\mathbf{C}}_{\underline{Q}} \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix}, \quad (4)$$

$$\overline{\mathbf{C}}_{\overline{Q}} = \begin{pmatrix} \hat{S}_{\overline{Q},xx} & \hat{S}_{\overline{Q},xy} \\ \hat{S}_{\overline{Q},yx} & \hat{S}_{\overline{Q},yy} \end{pmatrix}, \quad (5)$$

$$\underline{\mathbf{C}}_{\underline{Q}} = \begin{pmatrix} \hat{S}_{\underline{Q},xx} & \hat{S}_{\underline{Q},xy} \\ \hat{S}_{\underline{Q},yx} & \hat{S}_{\underline{Q},yy} \end{pmatrix}. \quad (6)$$

本論文では、式 (4) で定式化される多変量解析を一般化

(注2): 紙面の制約から、以降では S_{xy} 及びその拡張についてのみ言及する． S_{xx} や S_{yy} についても同様に定義できる．

多変量解析と呼ぶ。

式 (5)(6) に示される行列をどのように設計するかによって、式 (4) に示す一般化多変量解析の性質が決定される。基本的には、通常の変量解析と同様に、一般化多変量解析においても、行列 \overline{C}_Q に対応する性質を強調し、行列 \underline{C}_Q に対応する性質を抑制するという構造は変わらない。しかし、この構造のみでは、目的やデータの性質に応じた解析手法を設計する指針とはならない。

以降では、一般化多変量解析を具体的にどのように設計すれば良いかについて詳説する。まず、一般化多変量解析を構成するテンプレートの役割を担う標準的な多変量解析手法のうち、特に解説が必要と考えられるいくつかの例について述べる。その後、標準的な多変量解析を組み合わせる構成される、一般化多変量解析に含まれる既知の特殊例について、組み合わせ方のテンプレートとして紹介する。さらに、これらの考察から導かれるいくつかの新しい多変量解析手法について述べる^(注3)。

4. 一般化多変量解析の既知例

4.1 局所 Fisher 判別分析 (LFDA)

局所 Fisher 判別分析 (LFDA) [15] は、一般的な FDA の拡張の 1 つとして知られる教師付き次元削減手法であり、外れ値や分布の多峰性などによる判別性能の低下を緩和することができる。LFDA のポイントは、教師なし次元削減手法の 1 つである局所性保存射影 (LPP) [14]^(注4) を源流とする、サンプル間類似度行列 Q を考慮したクラス間分散 $S_Q^{(lb)}$ ・クラス内分散 $S_Q^{(lw)}$ の計算にある。

$$S_Q^{(lb)} = \sum_{n_1=1}^N \sum_{n_2=1}^N Q_{n_1, n_2}^{(lb)} (\mathbf{x}_{n_1} - \mathbf{x}_{n_2})(\mathbf{x}_{n_1} - \mathbf{x}_{n_2})^\top,$$

$$S_Q^{(lw)} = \sum_{n_1=1}^N \sum_{n_2=1}^N Q_{n_1, n_2}^{(lw)} (\mathbf{x}_{n_1} - \mathbf{x}_{n_2})(\mathbf{x}_{n_1} - \mathbf{x}_{n_2})^\top.$$

$Y = \{y_1, \dots, y_N\}$ の各要素が X の各要素の所属クラスを示すインデックスであるとき、LFDA における行列 Q の各要素は以下の式によって計算される。

$$Q_{n_1, n_2}^{(lb)} = \begin{cases} A_{n_1, n_2}(1/N - 1/N_c) & \text{if } y_{n_1} = y_{n_2} = c, \\ 1/N & \text{if } y_{n_1} \neq y_{n_2}, \end{cases}$$

$$Q_{n_1, n_2}^{(lw)} = \begin{cases} A_{n_1, n_2}/N_c & \text{if } y_{n_1} = y_{n_2} = c, \\ 1/N & \text{if } y_{n_1} \neq y_{n_2}, \end{cases}$$

ただし、 N_c はクラス c に属するサンプルの総数である。正方行列 $A = \{A_{n_1, n_2}\}_{n_1, n_2=1}^N$ は任意に設計することができるが、LPP 及び LFDA では局所スケール法 [19] による設計が用いられている。

(注3): 以降の節では、 \overline{C}_Q や \underline{C}_Q の表記の際に類似度行列 \overline{Q} や \underline{Q} が明示的に意識されない場合には、その表記を省略する。

(注4): LPP も一般化多変量解析で記述できるが、詳説は割愛する。

$$A_{n_1, n_2} = \exp \left\{ -\frac{\|\mathbf{x}_{n_1} - \mathbf{x}_{n_2}\|^2}{\sigma_{n_1} \sigma_{n_2}} \right\},$$

$$\sigma_n = \|\mathbf{x}_n - \mathbf{x}_n^{(k)}\|,$$

ただし、 $\mathbf{x}_n^{(k)}$ は \mathbf{x}_n の第 k 近傍のサンプルとする。

上記の内容により、LFDA は式 (5) と式 (6) に示される行列をそれぞれ

$$\overline{C}_Q^{(LFDA)} = S_Q^{(lb)} \text{ and } \underline{C}_Q^{(LFDA)} = S_Q^{(lw)}$$

と設定した一般化多変量解析の特殊例となる。

4.2 正準相関分析 (CCA)

正準相関分析 (CCA) [1] は、複数のサンプル集合に共通して含まれる情報を抽出して統合する次元削減手法の 1 つである。具体的には、2 つのサンプル集合 (X, Y) を線形変換 (w_x, w_y) でそれぞれ変換した際に、変換後の変量の正規化相互相関が最大となるような線形変換 (w_x, w_y) を求める。すなわち、CCA が最大化すべき目的関数 $\rho(w_x, w_y | X, Y)$ は以下ようになる。

$$\rho^{(CCA)}(w_x, w_y | X, Y) = \frac{(\mathbf{X}^\top w_x)^\top (\mathbf{Y}^\top w_y)}{\|\mathbf{X}^\top w_x\| \cdot \|\mathbf{Y}^\top w_y\|} = \frac{w_x^\top S_{xy} w_y}{\sqrt{w_x^\top S_{xx} w_x w_y^\top S_{yy} w_y}}.$$

目的関数の最大化に線形変換 (w_x, w_y) のスケールが影響しないことから、正規化 $w_x^\top S_{xx} w_x = w_y^\top S_{yy} w_y = 1$ を制約条件とした Lagrange 未定乗数法を用いることにより、以下の式を得る。

$$S_{xy} w_y - \lambda S_{xx} w_x = 0,$$

$$S_{yx} w_x - \lambda S_{yy} w_y = 0.$$

上記の内容により、CCA は、式 (5)(6) に示される行列を以下のように設定した一般化多変量解析の特殊例となる。

$$\overline{C}^{(CCA)} = \begin{pmatrix} \mathbf{0} & S_{xy} \\ S_{yx} & \mathbf{0} \end{pmatrix}, \underline{C}^{(CCA)} = \begin{pmatrix} S_{xx} & \mathbf{0} \\ \mathbf{0} & S_{yy} \end{pmatrix}.$$

4.3 線形回帰分析 (MLR)

線形回帰分析 (MLR) は、サンプル y とその線形近似である Wx との二乗誤差を最小にする線形変換 W を求める問題として定式化される。一般的には、MLR は最小自乗法を用いて解を導出するが、本節では一般化固有値問題での定式化を試みる。まず、ランク 1 の変換行列 W を考え、それが 2 つの基底ベクトル w_x と w_y の直積の形に分解できるとする。このとき、最小化すべき二乗誤差 $\epsilon^2(w_x, w_y | X, Y)$ は以下のように計算される。

$$\epsilon^2(w_x, w_y | X, Y) = N^{-1} \sum_{n=1}^N \|\mathbf{y}_n - \alpha w_y w_x^\top \mathbf{x}_n\|^2$$

$$= S_{yy}^2 - 2\alpha w_y^\top S_{xy}^\top w_x + \alpha^2 w_x^\top S_{xx} w_x.$$

これを α について最小化すると，以下の式が得られる．

$$\alpha = (\mathbf{w}_y^\top \mathbf{S}_{xy}^\top \mathbf{w}_x)(\mathbf{w}_x^\top \mathbf{S}_{xx} \mathbf{w}_x)^{-1}.$$

よって二乗誤差 $\epsilon^2(\mathbf{w}_x, \mathbf{w}_y | \mathbf{X}, \mathbf{Y})$ は

$$\begin{aligned} \epsilon^2(\mathbf{w}_x, \mathbf{w}_y | \mathbf{X}, \mathbf{Y}) \\ = \mathbf{S}_{yy}^2 - (\mathbf{w}_y^\top \mathbf{S}_{xy}^\top \mathbf{w}_x)^2 (\mathbf{w}_x^\top \mathbf{S}_{xx} \mathbf{w}_x)^{-1}. \end{aligned}$$

二乗誤差 $\epsilon^2(\mathbf{w}_x, \mathbf{w}_y | \mathbf{X}, \mathbf{Y})$ が非負であること，上式右辺第 1 項が基底ベクトル $(\mathbf{w}_x, \mathbf{w}_y)$ とは独立に決まることから，二乗誤差最小化は以下の目的関数の最大化に置換できる．

$$\rho^{(MLR)}(\mathbf{w}_x, \mathbf{w}_y | \mathbf{X}, \mathbf{Y}) = \frac{\mathbf{w}_x^\top \mathbf{S}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^\top \mathbf{S}_{xx} \mathbf{w}_x \mathbf{w}_y^\top \mathbf{w}_y}}.$$

上式を CCA の目的関数と比較することで，MLR が CCA の特殊例となること，及び解析対象の行列が以下で与えられることがわかる．

$$\overline{\mathbf{C}}^{(MLR)} = \begin{pmatrix} \mathbf{0} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{0} \end{pmatrix}, \quad \underline{\mathbf{C}}^{(MLR)} = \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_y} \end{pmatrix}.$$

4.4 Tikhonov 正則化

多変量解析における一般的な正則化の方法の 1 つとして，基底ベクトルの二乗和を正則化項とする Tikhonov 正則化 [11], [20] が知られている．Tikhonov 正則化を適用した MLR は，特に Ridge 回帰と呼ばれる [10]．この Ridge 回帰を例として，Tikhonov 正則化とその一般多変量解析との関係について詳説する．

Ridge 回帰で最小化すべき計量は，回帰分析で最小化すべき二乗誤差に基底ベクトルの二乗和を正則化項として加えた以下の量である．(注5)

$$\begin{aligned} \epsilon^2(\mathbf{w}_x, \mathbf{w}_y | \mathbf{X}, \mathbf{Y}) \\ = N^{-1} \sum_{n=1}^N \|\mathbf{y}_n - \alpha \mathbf{w}_y \mathbf{w}_x^\top \mathbf{x}_n\|^2 + \delta \|\mathbf{w}_x\|^2 \\ = \mathbf{S}_{yy}^2 - 2\alpha \mathbf{w}_y^\top \mathbf{S}_{xy}^\top \mathbf{w}_x + \alpha^2 \mathbf{w}_x^\top (\mathbf{S}_{xx} + \hat{\delta} \mathbf{I}_{d_x}) \mathbf{w}_x. \end{aligned}$$

ただし， $\hat{\delta} = \delta/\alpha^2$ である．このことから，Ridge 回帰も一般化多変量解析の特殊例となり，

$$\begin{aligned} \overline{\mathbf{C}}^{(Ridge)} &= \begin{pmatrix} \mathbf{0} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{0} \end{pmatrix}, \\ \underline{\mathbf{C}}^{(Ridge)} &= \begin{pmatrix} \mathbf{S}_{xx} + \hat{\delta} \mathbf{I}_{d_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_y} \end{pmatrix}. \end{aligned}$$

となることがわかる．Ridge 回帰と同様の手順を踏むことで，CCA における Tikhonov 正則化 [12] も一般化多変量解析の特殊例となることが示される．

$$\begin{aligned} \overline{\mathbf{C}}^{(Tik)} &= \begin{pmatrix} \mathbf{0} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{0} \end{pmatrix}, \\ \underline{\mathbf{C}}^{(Tik)} &= \begin{pmatrix} \mathbf{S}_{xx} + \hat{\delta}_x \mathbf{I}_{d_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{yy} + \hat{\delta}_y \mathbf{I}_{d_y} \end{pmatrix}. \end{aligned}$$

本節における考察から，式 (3) で示した拡張ペアワイズ表現における行列 L_2 は，基底ベクトルに対する正則化項としての役割を果たすことがわかる．

4.5 半教師付き局所 Fisher 判別分析 (SELF)

半教師付き局所 Fisher 判別分析 (SELF) [16] は，いわゆる半教師付き次元削減の問題に対して，教師なし次元削減法としての PCA と教師付き次元削減法としての LFDA を融合させた手法である．SELF は，一般化多変量解析の設計指針において重要となる以下の 2 点を包含する典型的な例となっている．

- (1) 複数の標準的なテンプレートを組み合わせる．
- (2) 組み合わせる各テンプレート内で計量計算に用いるサンプル集合を変える．

本節では，サンプル集合 \mathbf{Y} がクラス番号を保持する属性データ，及び $N_x > N_y = N$ (すなわち $\mathbf{X}_I \neq \emptyset, \mathbf{Y}_I = \emptyset$) という状況を考える．

少数の教師付きサンプル集合 $(\mathbf{X}_C, \mathbf{Y}_C)$ のみを用いた LFDA では，過適応が生じる可能性がある．SELF では，教師なしサンプル集合 \mathbf{X}_I も含めた PCA を組み合わせることで，この過適応を回避する．具体的には，教師付きサンプル集合のみから計算される LFDA のクラス内・クラス間分散行列 $(\mathbf{S}_{Q,C}^{(lw)}, \mathbf{S}_{Q,C}^{(lb)})$ と，教師なしサンプル集合をも用いた PCA の共分散行列 \mathbf{S}_{xx} とを混合した以下の行列による一般化多変量解析が，SELF となる．

$$\begin{aligned} \overline{\mathbf{C}}_Q^{(SELF)} &= \beta \mathbf{S}_{Q,C}^{(lb)} + (1 - \beta) \mathbf{S}_{xx}, \\ \underline{\mathbf{C}}_Q^{(SELF)} &= \beta \mathbf{S}_{Q,C}^{(lw)} + (1 - \beta) \mathbf{I}_{d_x}. \end{aligned}$$

ただし， β は $0 \leq \beta \leq 1$ を満たすフリーパラメータである． $\beta = 1$ の場合に教師付きサンプル集合 $(\mathbf{X}_C, \mathbf{Y}_C)$ のみを用いた LFDA と一致し， $\beta = 0$ の場合にサンプル集合 \mathbf{X} のみを用いた PCA と一致する．一般に，SELF は，LFDA と PCA の双方の性質を受け継いでおり，教師付きサンプル集合の局所的な構造と，サンプル空間における大域的な分布構造とを，パラメータ β で制御する．

4.6 半教師付き正準相関分析 (SemiCCA)

前節で紹介した SELF の考え方を CCA に適用した手法が，半教師付き正準相関分析 (SemiCCA) である [17], [21]．具体的には，2 つのサンプル集合 (\mathbf{X}, \mathbf{Y}) がいずれもベクトルデータ，かつ $N_x > N, N_y > N$ ($\mathbf{X}_I \neq \emptyset, \mathbf{Y}_I \neq \emptyset$) という状況を考え，共起サンプル集合 $(\mathbf{X}_C, \mathbf{Y}_C)$ のみを用いた CCA と，非共起サンプル集合 $(\mathbf{X}_I, \mathbf{Y}_I)$ をも取り込む PCA とを，拡大行列の内分混合によって組み合わせる．

(注5): 定数 δ を Lagrange 未定乗数と見なすと，この最小化問題は，基底ベクトルの自乗和をある定数以下とする制約付問題と等価となる．

$$\begin{aligned} \overline{\mathbf{C}}^{(sCCA)} &= \beta \begin{pmatrix} \mathbf{0} & \mathbf{S}_{Cxy} \\ \mathbf{S}_{Cyx} & \mathbf{0} \end{pmatrix} + (1-\beta) \begin{pmatrix} \mathbf{S}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{yy} \end{pmatrix}, \\ \underline{\mathbf{C}}^{(sCCA)} &= \beta \begin{pmatrix} \mathbf{S}_{Cxx} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{Cyy} \end{pmatrix} + (1-\beta) \begin{pmatrix} \mathbf{I}_{d_x} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d_y} \end{pmatrix}, \end{aligned}$$

SELF と同様、 $\beta = 1$ の場合に共起サンプル集合 (X_C, Y_C) のみを用いた CCA と一致し、 $\beta = 0$ の場合に X と Y を独立に PCA を行う問題に帰着される。

これまでの議論を踏まえると、 $\underline{\mathbf{C}}^{(sCCA)}$ 第 2 項は Tikhonov 正則化項と見なすことができる。すなわち、SemiCCA で得られる基底は、ノルムがある定数以下になる制約の下で最適化される。一方、 $\overline{\mathbf{C}}^{(sCCA)}$ 第 2 項は、 X と Y の独立性に関する制約を緩め、共起サンプル集合 (X_C, Y_C) のみからでは得られない相関関係を、非共起サンプル集合も含めた全サンプル集合 (X, Y) の同時分布から獲得する効果が期待される。

5. カーネル一般化多変量解析

一般化多変量解析の枠組に含まれるいずれの解析手法も、カーネルトリックを用いて非ベクトルデータ・非線形解析へ拡張できる。以降では、例としてカーネル正準相関分析 (kernel CCA) [4], [8], [9] 及びその拡張について概説する。詳細は各文献に譲る。

5.1 カーネル正準相関分析 (Kernel CCA)

Kernel CCA では、各サンプル (x_n, y_n) にそれぞれ (非線形) 変換 (ϕ_x, ϕ_y) を施した後のベクトル $(\phi_x(x_n), \phi_y(y_n))$ に対して CCA を行い、基底ベクトル (w_x, w_y) を求める。Kernel CCA で最大化すべき目的関数は、変換 $(\phi_x(x_n), \phi_y(y_n))$ に対応するグラム行列 K_x, K_y を用いて以下のように記述される。

$$\rho^{(kCCA)}(w_x, w_y | X, Y, \phi_x, \phi_y) = \frac{\alpha^\top K_{xx} K_{yy} \beta}{\sqrt{\alpha^\top K_{xx}^2 \alpha \beta^\top K_{yy}^2 \beta}}.$$

$$K_{xx} = \{K_x(x_{n_1}, x_{n_2})\}_{n_1, n_2=1}^N,$$

$$K_{yy} = \{K_y(y_{n_1}, y_{n_2})\}_{n_1, n_2=1}^N,$$

$$K_x(x_{n_1}, x_{n_2}) = \phi_x(x_{n_1})^\top \phi_x(x_{n_2}),$$

$$K_y(y_{n_1}, y_{n_2}) = \phi_y(y_{n_1})^\top \phi_y(y_{n_2}),$$

$$w_x = \sum_{n=1}^N \alpha_n \phi_x(x_n), \quad w_y = \sum_{n=1}^N \beta_n \phi_y(y_n),$$

通常の CCA と同様にして、上記の目的関数最大化は、以下の一般化固有値問題を解くことによって行われる。

$$\begin{aligned} \overline{\mathbf{C}}^{(kCCA)} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} &= \lambda \underline{\mathbf{C}}^{(kCCA)} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \\ \overline{\mathbf{C}}^{(kCCA)} &= \begin{pmatrix} \mathbf{0} & K_{xx} K_{yy} \\ K_{yy} K_{xx} & \mathbf{0} \end{pmatrix}, \end{aligned}$$

$$\underline{\mathbf{C}}^{(kCCA)} = \begin{pmatrix} K_{xx}^2 & \mathbf{0} \\ \mathbf{0} & K_{yy}^2 \end{pmatrix}.$$

上記の一般化固有値問題から射影軸 (w_x, w_y) そのものを求めることはできないが、これら射影軸への射影はカーネル関数を用いて以下のように求められる。

$$w_x^\top \phi_x(x) = \sum_{n=1}^N \alpha_n K_x(x_n, x),$$

$$w_y^\top \phi_y(y) = \sum_{n=1}^N \beta_n K_y(y_n, y),$$

5.2 Kernel CCA における正則化

Kernel CCA は、用いるグラム行列のランクが高々 $\min(m_x, m_y)$ となるため、過学習が非常に起こりやすく、通常は何らかの正則化と共に用いられる。最も一般的な正則化の 1 つが、先に紹介した Tikhonov 正則化である [12]。Tikhonov 正則化により、解析に用いられる各行列は以下のように置き換わる。

$$\begin{aligned} \overline{\mathbf{C}}^{(kTik)} &= \begin{pmatrix} \mathbf{0} & K_{xx} K_{yy} \\ K_{yy} K_{xx} & \mathbf{0} \end{pmatrix}, \\ \underline{\mathbf{C}}^{(kTik)} &= \begin{pmatrix} K_{xx}^2 + \delta_x K_{xx} & \mathbf{0} \\ \mathbf{0} & K_{yy}^2 + \delta_y K_{yy} \end{pmatrix}. \end{aligned}$$

もう 1 つの代表的な正則化手法として、graph Laplacian を用いる方法が知られている [13]。Graph Laplacian 正則化を用いることにより、非線形変換前の原空間での平滑化をもたらす効果がある。サンプル間類似度行列 Q_x, Q_y によって決まる Laplacian 行列を L_{Q_x}, L_{Q_y} とすると、graph Laplacian 正則化を用いることにより、解析に用いられる各行列は以下のように置き換わる。

$$\begin{aligned} \overline{\mathbf{C}}^{(Lap)} &= \begin{pmatrix} \mathbf{0} & K_{xx} K_{yy} \\ K_{yy} K_{xx} & \mathbf{0} \end{pmatrix}, \\ \underline{\mathbf{C}}^{(Lap)} &= \begin{pmatrix} K_{xx}^2 + \gamma_x \hat{K}_{xx} & \mathbf{0} \\ \mathbf{0} & K_{yy}^2 + \gamma_y \hat{K}_{yy} \end{pmatrix}, \\ \hat{K}_{xx} &= K_{xx} L_{Q_x} K_{xx}, \quad \hat{K}_{yy} = K_{yy} L_{Q_y} K_{yy}, \end{aligned}$$

5.3 Kernel SELF

一般多変量解析のカーネル拡張を知る上で重要となるのが、SELF のカーネル拡張である kernel SELF [16] である。先に示した通常の SELF と同様、複数の標準的なテンプレートを組み合わせる方法、及び組み合わせる各テンプレート内で計量計算に用いるサンプル集合を変える方法を、kernel SELF を見ることで知ることができる。

Kernel SELF を導出するポイントは、ペアワイズ表現にある。まず、教師付きサンプル集合のみから計算されるクラス間共分散行列 $S_{Q,C}^{(lb)}$ とクラス内共分散行列 $S_{Q,C}^{(lw)}$ 、及び全サンプルから計算される共分散行列 S_{xx} を、それぞれペアワイズ表現に変換する。

$$\begin{aligned} S_{Q,C}^{(lb)} &= X_C L_{Q,C}^{(lb)} X_C^\top, & S_{Q,C}^{(lw)} &= X_C L_{Q,C}^{(lw)} X_C^\top, \\ S_{xx} &= X L_{xx} X^\top, \end{aligned}$$

ここで得られた各 Laplacian 行列 $L_{Q,C}^{(lb)}$, $L_{Q,C}^{(lw)}$ 及び L_{xx} を用いることにより, 解析に用いる各行列を以下のように記述できる.

$$\begin{aligned} \bar{C} &= K_x \{ \beta L_{Q,C}^{(lb)} + (1 - \beta) L_{xx} \} K_x, \\ \underline{C} &= \beta K_x L_{Q,C}^{(lw)} K_x + (1 - \beta) K_x, \end{aligned}$$

このことから, 線形一般多変量解析における共分散行列の内分混合は, カーネル一般多変量解析では (拡張) ペアワイズ表現に変換した後の内分混合として扱うことができることがわかる.

5.4 半教師付き Kernel CCA

SELF を SemiCCA に拡張する方法と同様にして, kernel SELF を拡張することで半教師付き正準相関分析 (kernel SemiCCA) を導出できる. 詳細は省略する.

半教師付き正準相関分析の別の実現方法として, Blashcko らによる方法 [13] が知られている. この方法では, 半教師付き次元削減を実現するために用いる計量計算の方法が SELF や SemiCCA と異なる. 具体的には, 一般化多変量解析における各行列を以下のように置くことで実現される.

$$\begin{aligned} \bar{C}^{(Bla)} &= \begin{pmatrix} \mathbf{0} & K_{XXC} K_{YC} \\ K_{YYC} K_{XC} & \mathbf{0} \end{pmatrix}, \\ \underline{C}^{(Bla)} &= \begin{pmatrix} K_{XXC} K_{XC} & \mathbf{0} \\ \mathbf{0} & K_{YYC} K_{YC} \end{pmatrix}, \\ K_{XXC} &= \{ K_x(\mathbf{x}_{n_1}, \mathbf{x}_{C,n_2}) \}_{n_1=1}^{N_x} \}_{n_2=1}^N, \\ K_{XCX} &= \{ K_x(\mathbf{x}_{C,n_1}, \mathbf{x}_{n_2}) \}_{n_1=1}^N \}_{n_2=1}^{N_x}. \end{aligned}$$

この方法はすなわち, 拡張グラム行列 K_{XXC} の転置積 $K_{XXC} K_{XCX}$ を利用することで, よりサイズの大きなグラム行列の積 K_{XX}^2 に相当する情報を抽出している.

6. 一般化多変量解析の構成例

前節までの議論にて, (1) 一般化多変量解析を設計する上でのテンプレートとしての既出の多変量解析手法, (2) そのテンプレートの組み合わせ方の例としての共分散行列の内分混合, 及び (3) 2 次統計量計算の母集合の切替を提示した. 一般化多変量解析の枠組そのものは非常に広範であり, これらの組み合わせに留まるものではない. しかし, 前節まで紹介したテンプレートとその組み合わせ方を利用するだけでも, いくつかの有用かつ新しい多変量解析手法を生成することができる. 本節では, その例について紹介する.

6.1 局所正準相関分析 (LCCA)

FDA を LFDA に拡張する方法と同様に, CCA にサンプル間類似度行列 Q を導入することで, 外れ値や分布多峰性の影響を一定程度緩和した基底の導出が可能となる. 本稿では, この方法を局所正準相関分析 (LCCA) と呼ぶ. EM アルゴリズムによる混合ガウス分布パラメータの推定と CCA とを組み合わせる方法を用いても類似の効果を期待できるが, あらかじめ混合数を設定する必要がない点, 一般化固有値問題を一度解くだけで最適解が算出できる点において, LCCA がより優れている.

具体的には, サンプル間類似度行列 Q_{xx} , Q_{yy} , Q_{xy} を考慮した 2 次統計量 $S_{Q_{xx,xx}}$, $S_{Q_{yy,yy}}$, $S_{Q_{xy,xy}}$ を各共分散行列に代えて用いた CCA が, LCCA となる.

$$\begin{aligned} \bar{C}_Q^{(LCCA)} &= \begin{pmatrix} \mathbf{0} & S_{Q_{xy,xy}} \\ S_{Q_{yx,yx}} & \mathbf{0} \end{pmatrix}, \\ \underline{C}_Q^{(LCCA)} &= \begin{pmatrix} S_{Q_{xx,xx}} & \mathbf{0} \\ \mathbf{0} & S_{Q_{yy,yy}} \end{pmatrix}, \\ Q &= \begin{pmatrix} Q_{xx} & Q_{xy} \\ Q_{yx} & Q_{yy} \end{pmatrix}, \end{aligned}$$

6.2 正準相関判別分析 (CCDA)

音響信号と画像信号とによって構成される映像信号にメタ情報が付与されている場合などのように, それぞれ性質の異なる共起サンプル集合 (X, Y) の各サンプル対にラベルが付与されている状況を考える.

このとき, 各サンプル対 $(\mathbf{x}_n, \mathbf{y}_n)$ ($n = 1, 2, \dots, N$) を連結したベクトル $(\mathbf{x}_n^\top, \mathbf{y}_n^\top)^\top$ を新しいサンプルと考えて FDA を適用し, 判別軸を求める方法が考えられる. しかし, この方法では, CCA で発見可能な 2 種類のベクトル集合の相関関係を PCA で計算できないことと同様に, 適切な相関関係を計算できないという問題がある. この対処法として, 共起サンプル集合 (X, Y) について CCA を用いた次元削減を行った後に FDA を用いる方法が考えられる. しかし, 2 種類のベクトル集合の相関関係がクラスごとに異なる場合には, その関係性を見出すことが困難である.

そこで, SELF や SemiCCA で活用した共分散行列の内分混合を利用することで CCA と FDA を融合し, クラスごとに異なる相関関係を持つ可能性のある 2 種類のベクトル集合に対する識別的次元削減を可能にする.

一般化多変量解析の枠組における具体的な定式化は以下の通りとなる. $\beta = 1$ の場合に CCA と一致し, $\beta = 0$ の場合に各サンプル対を連結したベクトルを新しいサンプルと考えた FDA と一致する.

$$\begin{aligned} \bar{C}^{(CCDA)} &= \beta \begin{pmatrix} \mathbf{0} & S_{xy} \\ S_{yx} & \mathbf{0} \end{pmatrix} + (1 - \beta) \begin{pmatrix} S_{xx}^{(lb)} & S_{xy}^{(lb)} \\ S_{yx}^{(lb)} & S_{yy}^{(lb)} \end{pmatrix}, \\ \underline{C}^{(CCDA)} &= \beta \begin{pmatrix} S_{xx} & \mathbf{0} \\ \mathbf{0} & S_{yy} \end{pmatrix} + (1 - \beta) \begin{pmatrix} S_{xx}^{(lw)} & S_{xy}^{(lw)} \\ S_{yx}^{(lw)} & S_{yy}^{(lw)} \end{pmatrix}. \end{aligned}$$

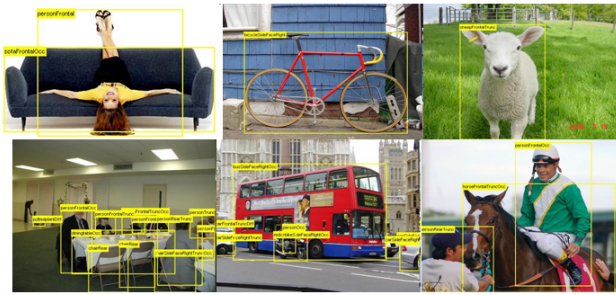


図1 VOC2008 データセットに含まれる画像の例

6.3 半教師付き学習への拡張

上記の2つの手法は、SELFやSemiCCAの手順を参考にすることで、半教師付き次元削減に拡張することができる。LCCAの半教師拡張の手順はSemiCCAのそれとほぼ同様であり、LCCAと局所性保存射影(LPP)[14]を組み合わせる。また、CCA+LDAの半教師拡張については、教師付き次元削減であるFDAに対応する2次統計量を教師付きサンプル集合のみから、教師なし次元削減であるCCAに対応する2次統計量を全サンプル集合から計算することで実現できる。

7. 検証実験

7.1 実験条件

提案法の有効性を検証するために、PASCAL Visual Object Challenge (VOC) 2008及び2009[22]データセットを利用して、主に物体認識に関する実験を行い、その精度を評価した。

PASCAL VOC データセットは、20個のクラスのいずれかに属する「物体」(人物、動物、車両、家具)が含まれる画像からなる、一般物体認識を主目的としたベンチマークである。1枚の画像の中に複数の物体が含まれることもあり、それら複数物体のクラスが同じ場合も異なる場合もある。VOC2008データセットに含まれる学習用画像の例を図1に示す。学習用画像には、物体の位置を示す方形領域とそのクラスがメタ情報が付与されている。しかし、一般の画像認識検索が利用される状況において、上記のようなメタ情報が完全に与えられることは想定しにくい。そこで、本実験においては、全ての処理において物体の存在位置を示す方形領域は用いず、画像に付与されたクラスラベルのみを利用した。また、VOC2008学習用データセットに含まれるラベル付画像5096枚のうち、本実験では、ランダムに選択した4000枚をラベル付き学習用画像 $G^{(L)}$ 、1000枚を評価用画像 $G^{(E)}$ として利用した。

物体認識アルゴリズムは、中山ら[23]が提案した、CCAとカーネル密度推定を組み合わせた方法を用いた。この方法は、(1)画像特徴量集合とラベル特徴量集合とを共起サンプル集合として用いたCCAを用いることにより低次元潜在変数空間を構築し、(2)その潜在変数空

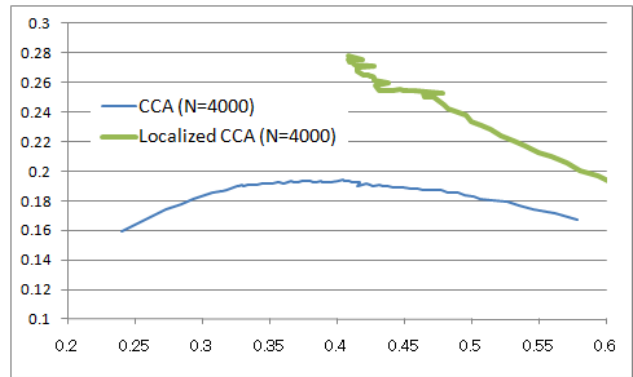


図2 一般物体認識タスクにおける精度評価の結果

間内でのカーネル密度推定により導出された潜在変数の確率密度関数に基づいてラベル割り当てを行う。本実験では、上記2手順のうち第1の手順である潜在変数空間の構築について、その手法を変更して評価検証を行う。画像特徴量はSURF[24]を特徴点抽出・局所特徴量抽出に用いたBag-of-Features(256次元)、ラベル特徴量はラベルの有無を2値で表現するラベル数(=20)と同次元のベクトルを利用した。潜在変数空間の次元は8次元とした。

7.2 実験結果

一般物体認識における評価尺度として、以下で定義される正答率(PR)、再現率(RE)、及びF値(F)を用いる。

$$PR = \frac{\sum_{n=1}^{N_e} tp_n}{\sum_{n=1}^{N_e} (tp_n + fp_n)}, \quad (7)$$

$$RE = \frac{\sum_{n=1}^{N_e} tp_n}{\sum_{n=1}^{N_e} (tp_n + fn_n)}, \quad (8)$$

$$F = 2 \left(\frac{1}{PR} + \frac{1}{RE} \right)^{-1}, \quad (9)$$

ここで、 N_e は評価用画像の数(=500)、 tp_n 、 fp_n 、 fn_n はそれぞれ n 番目の評価用画像に対する認識結果についての正解数、誤検出数、見落とし数を表現する。

図2に実験結果を示す。青太線、緑太線が、それぞれ、潜在変数空間構築にCCA、LCCAを用いた場合の結果である本結果が示すように、サンプル間類似度行列を考慮したLCCAで、それを考慮しないCCAを上回る結果を示している。

8. おわりに

本論文では、拡張ペアワイズ表現と呼ぶ共分散行列やグラム行列に関する新しい表現形式を導入することにより、標準的な多変量解析手法やその正則化・半教師付き拡張を統一的な視点で捉え直す枠組を提供すると共に、これら標準的な多変量解析手法やその正則化・半教師付き拡張をテンプレートとして組み合わせるといふ、新し

い多変量解析手法の設計指針を提示した。また、この設計指針に基づいて、いくつかの有用で新しい多変量解析を具体的に提案した。さらに、一般物体認識ベンチマークを用いた実験によって本手法の有効性を示した。

一般化多変量解析の枠組そのものは非常に広範であり、これらの組み合わせに留まるものではない。カーネル多変量解析におけるカーネルの設計と同様に、拡張ペアワイズ表現に含まれる2つの行列を問題や対象に合わせて設計する、より汎用的な設計指針については、今後の研究課題として最も重要なものの1つである。

謝辞

本研究に対し真摯にご議論いただき有益な御助言を頂いたNTTコミュニケーション科学基礎研究所 大石康智氏及び前田英作 主席研究員に深謝する。また、本共同研究に御支援を頂いたNTTコミュニケーション科学基礎研究所 上田修功 所長、大和淳司 主幹研究員、柏野邦夫 主幹研究員、並びに澤田宏 主幹研究員に感謝する。

文 献

- [1] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, Vol. 24, 1933.
- [2] R. Fisher, "The use of multiple measurements in taxonomic problems," *Annals Eugen.*, Vol. 7, pp. 179–188, 1936.
- [3] T. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley-Interscience, 2003.
- [4] 赤穂, *カーネル多変量解析*, 岩波書店, 2008.
- [5] 福水, *カーネル法入門*, 朝倉書店, 2010.
- [6] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, Vol. 10, no. 5, pp. 1299–1319, 1998.
- [7] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, Vol. 12, no. 10, pp. 2385–2404, 2000.
- [8] S. Akaho, "A kernel method for canonical correlation analysis," in *In Proc. IMPS2001*.
- [9] P. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," in *Proc. ICANN*, Vol. 4, pp. 4614, 2000.
- [10] A. Hoerl, "Application of ridge analysis to regression problem," *Chemical Engineering Progress*, vol. 58, pp. 54–59, 1962.
- [11] A. Tikhonov, "On the stability of inverse problems," *Dokl. Akad. Nauk SSSR*, vol. 39, no. 5, pp. 195–198, 1943.
- [12] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [13] M. Blaschko, C. Lampert, and A. Gretton, "Semi-supervised laplacian regularization of kernel canonical correlation analysis," in *Proc. ECML PKDD 2008*, pp. 133–145.
- [14] X. He and P. Niyogi, "Locality preserving projections," in *Proc. NIPS*, 2003.
- [15] M. Sugiyama, "Dimensionality reduction of multi-modal labeled data by local Fisher discriminant analysis," *JMLR*, Vol. 8, pp. 1027–1061, May 2007.
- [16] M. Sugiyama, T. Idé, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Machine Learning*, vol. 78, no. 1–2, pp. 35–61, 2010.
- [17] A. Kimura, H. Kameoka, M. Sugiyama, T. Nakano, E. Maeda, H. Sakano, and K. Ishiguro, "SemiCCA: Efficient semi-supervised learning of canonical correlations," in *Proc. ICPR2010*, pp. 2933–2936.
- [18] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, Inc., 1990.
- [19] L. Zelnik-manor and P. Perona, "Self-tuning spectral clustering," in *Proc. NIPS2004*, pp. 1601–1608.
- [20] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [21] 木村, 中野, 亀岡, 杉山, 前田, 坂野, "SSCDE: 画像認識検索のための半教師付正準密度推定法," in *画像の認識・理解シンポジウム予稿集*, July 2010, OS8–1.
- [22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results," <http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html>.
- [23] 中山, 原田, 國吉, "大規模 web 画像のための画像アノテーション・リトリバル手法," in *画像の認識・理解シンポジウム予稿集*, July 2009, pp. 55–62.
- [24] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *CVIU*, Vol. 110, no. 3, pp. 346–359, 2008.