

複数情報源に対する主成分分析

諏訪 恭平[†] 富岡 亮太[†] 矢入 健久[†] 鹿島 久嗣[†]

[†] 東京大学

あらまし データ解析の多くの場面において、データは数十ないし数百、数千といった非常に高い次元のベクトルとして表現される。高次元データを人間が解釈するために、あるいは予測モデル構築の前処理として、主成分分析などの教師なし次元削減手法がしばしば用いられる。さらに、実際のデータ解析では、画像データ、音声データ、テキストデータなどの異種混合データを統合して扱う必要が生じるが、複数の異なるデータを統合して解析するための方法は、特に教師つき学習の文脈では近年盛んに研究が行われているものの、教師なし学習の文脈ではまだ少ない。そこで本論文では、次元削減法の代表的手法である主成分分析を、各情報源の重要度パラメータを導入した定式化を与えることによって複数情報源に拡張し、主成分の推定と情報源の重要度の推定を順に繰り返すような効率のよい最適化アルゴリズムを与える。また、観測データからの物体地図復元タスクを題材に、提案手法の有効性を確認する。

キーワード 次元削減, (カーネル) 主成分分析, 情報統合, グループラッソ, マルチカーネル学習

Principal component analysis for multiple information sources

Kyohei SUWA[†], Ryota TOMIOKA[†], Takehisa YAIRI[†], and Hisashi KASHIMA[†]

[†] The University of Tokyo

Abstract In many data analysis applications, data are represented as high-dimensional feature vectors. Dimensionality reduction techniques such as principal component analysis are often used to interpret such high-dimensional data, or to pre-process the data for predictive modeling. Recently, there have been substantial research work on selective integration of heterogeneous data sources, including images, voices, and texts, for supervised learning, while not much work has been done in the context of unsupervised learning. In this paper, we propose an extension of the principal component analysis for multiple information sources by introducing importance of each information source, and give an efficient optimization algorithm that alternates subspace estimation and importance estimation. Experimental evaluation on map building from sensor data shows the effectiveness of the proposed method.

Key words Dimension reduction, (Kernel) principal component analysis, Information integration, Group lasso, Multiple kernel learning

1. はじめに

データ解析の多くの場面において、データは数十ないし数百、数千などといった非常に高い次元のベクトルとして表現される。高次元データを人間が解釈するために、あるいは、分類や回帰などといった予測のための解析の前処理として、高次元データからデータの本質的な特徴を表す低次元表現を抽出するための次元削減手法がしばしば用いられる。

次元削減の代表的な手法のひとつは主成分分析 [1] であり、高次元空間におけるデータをよく近似する低次元空間への線形写像を求める。複雑な分布をもったデータの場合には、非線形の次元削減手法が用いられることもある。非線形の次元削減法としては、先の主成分分析をカーネル法を用いることによって非線形化したカーネル主成分分析 [2] をはじめ、ラプラス固有写

像 [3]、多次元スケーリング [4] など数多くの手法が提案されている。

実際のデータ解析においては、異なる種類の様々なデータが与えられることがある。例えば、マルチメディアデータを対象とした解析では、画像データ、音声データ、テキストデータなど全く異なる形式をもった異種混合データを扱う必要が生じる。また、ロボットの制御などを行う場合、加速度センサーや距離センサーなどの各種センサーデータに加え、画像データなどを統合して用いることが重要になる。

このように複数の異なるデータを統合して解析するための方法は、統計学や機械学習の分野で近年盛んに研究が行われている。特に、回帰や分類などといった、機械学習でいうところの教師つき学習の枠組みでは、グループラッソ [5] やマルチカーネル学習 [6], [7] などを皮切りに、様々な学習アルゴリズムがマルチ

カーネル化されている [8]~[10]. これらのアプローチでは、モデルにおいて重要度の高い情報源に関する部分は相対的に大きくモデルの出力に影響し、重要でない情報源に関する部分の影響は相対的に小さくなるようにする. そのために各情報源の重要度を表すパラメータが導入され、その重要度はデータをもとに自動的に調整されるようになっている. 一方で、教師なし次元削減などのいわゆる教師なし学習の文脈では、いくつかの研究 (例えば [11]~[13]) を除いては、情報源の重要度を考慮したデータ統合の試みはほとんど行われていない.

以上をふまえ、本論文では、非線形次元削減法の代表的手法であるカーネル主成分分析を、複数情報源 (カーネル主成分分析の文脈では複数のカーネル関数) に拡張することを試みる. 基本的なアプローチは、主成分分析の目的関数として用いられる次元削減後のデータの分散に加え、グループラッソ [5] で用いられる 2-ノルムの 1-ノルムを学習のペナルティ項 (正規化項) として用いることで、分散の最大化に寄与しない情報源にかかる重みを小さくするというものである. さらに、グループラッソからマルチカーネル学習 [14] への導出にならう、このカーネル法版を導くことによって手法の非線形化を行う. この導出によって、各情報源の重要度として解釈できるパラメータが現われ、全体としては、主成分を示すパラメータと、各情報源の重要度パラメータの両方をもった最適化問題となる. また、そのアルゴリズムは、各ステップで主成分を上から順に 1 つずつ、各情報源の重みとともに求めていくような繰り返しアルゴリズムとなる.

提案手法の有効性を検証するため、観測データからの地図作成問題 [15] を題材にした人工データによる数値実験を行う. これは、高次元の観測データを 2 次元に次元削減することで、各物体の相対位置を復元するタスクであり、その復元精度によって手法の良さを測ることにする. 実験は 2 種類行い、1 つ目の実験では複数の相補的な情報源があるときに、これを適切に組み合わせることで、単一の情報源や、複数の情報源の等重みでの単純な組み合わせを用いた場合よりも高い精度で物体の配置図を復元できることを示す. また、2 つ目の実験では、1 つの有用な情報源に加え、有用でないランダムな情報源を追加していても、提案手法はノイズに影響されることなく、正しい情報源を選び出すことで、推定精度を高く保つことができることを示す.

2. レビュー：主成分分析

まずは、提案手法である複数情報源に対する (カーネル) 主成分分析のベースとなる、単一情報源に対する (カーネル) 主成分分析について確認する.

2.1 主成分分析とカーネル主成分分析

それぞれが D 次元であるようなデータを N 個あり、これらを並べて得られるデザイン行列を $\mathbf{X} \in \mathbb{R}^{N \times D}$ とする (なお、簡単のため \mathbf{X} は中心化されているものとする). また、データを一次元空間に写像する単位ベクトルを $\mathbf{w} \in \mathbb{R}^D$ とする. 主成分分析の目的関数は、写像後の分散

$$J(\mathbf{w}) \equiv \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} \quad (1)$$

であり、これを制約

$$\|\mathbf{w}\|_2 \equiv \mathbf{w}^\top \mathbf{w} \leq 1 \quad (2)$$

の制約のもとで最大化するような \mathbf{w} を求める. 最大値を与えるような \mathbf{w} は、以下の固有値問題に対する最大固有値をもつ単位固有ベクトルとなる.

$$\mathbf{X}^\top \mathbf{X} \mathbf{w} = \lambda \mathbf{w} \quad (3)$$

最大固有ベクトルを $\mathbf{w}^{(1)}$ とすると、これによるデータの 1 次元表現は $\mathbf{y}^{(1)} \equiv \mathbf{X} \mathbf{w}^{(1)}$ と与えられる. $d (< D)$ 次元空間に写像する場合には、固有値の大きい方から d 個の固有ベクトル ($\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(d)}$) をとり、これらによるデータの d 次元表現は $(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(d)}) \equiv (\mathbf{X} \mathbf{w}^{(1)}, \mathbf{X} \mathbf{w}^{(2)}, \dots, \mathbf{X} \mathbf{w}^{(d)})$ で与えられる.

カーネル主成分分析の場合には

$$\mathbf{w} \equiv \mathbf{X}^\top \boldsymbol{\alpha}$$

とおくことで、パラメータを $\boldsymbol{\alpha}$ に置き換えた以下の固有値問題を解く.

$$\mathbf{K} \boldsymbol{\alpha} = \lambda \boldsymbol{\alpha} \quad (4)$$

ここで、 $\mathbf{K} \equiv \mathbf{X} \mathbf{X}^\top$ はカーネル行列と呼ばれる. カーネル行列はデータ間の類似度を表す任意の半正定値行列としてとることができる. また、正規化条件は $\boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha} = 1$ となっていることに注意する. 固有値問題 (4) の固有値の大きい方から d 個の固有ベクトルを $(\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \dots, \boldsymbol{\alpha}^{(d)})$ とすれば、これらによるデータの d 次元表現は $(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(d)}) \equiv (\mathbf{K} \boldsymbol{\alpha}^{(1)}, \mathbf{K} \boldsymbol{\alpha}^{(2)}, \dots, \mathbf{K} \boldsymbol{\alpha}^{(d)})$ で与えられる.

2.2 主成分の逐次抽出

(カーネル) 主成分分析においては、対応する固有値問題の大きい方から d 個の固有値に対応する固有ベクトルを取り出すことで、 d 次元空間での射影を求めることができる.

ここでは一度に d 個の固有ベクトルを取り出すのではなく、ひとつずつ取り出していく方法を紹介しておく. これは、提案手法が通常の主成分分析を拡張する際の基本となる.

固有値問題 (3) の解のうち、最大の固有値を持つ固有ベクトルを $\mathbf{w}^{(1)}$ とすると、この写像ベクトルによる射影は $\mathbf{y}^{(1)} \equiv \mathbf{X} \mathbf{w}^{(1)}$ 、また、これをもとの空間でみると、 $\mathbf{y}^{(1)} \mathbf{w}^{(1)\top}$ であるため、その寄与分をもとのデザイン行列 \mathbf{X} から取り除くと、新たなデザイン行列

$$\mathbf{X}^{(2)} \equiv \mathbf{X} - \mathbf{y}^{(1)} \mathbf{w}^{(1)\top} \quad (5)$$

が得られる.

第 2 軸を求めるためには、今度は固有値問題 (3) のデザイン行列を新たなデザイン行列 $\mathbf{X}^{(2)}$ で置き換えた固有値問題に対する最大固有ベクトル $\mathbf{w}^{(2)}$ を求め、その寄与分を差し引く. このように、 $d-1$ 次元目までの寄与分を取り除いたデザイン行列 $\mathbf{X}^{(d)}$ に対する最大固有ベクトル $\mathbf{w}^{(d)}$ の抽出と、その寄与分の除去によるデザイン行列の更新

$$\mathbf{X}^{(d+1)} \equiv \mathbf{X}^{(d)} - \mathbf{y}^{(d)} \mathbf{w}^{(d)\top} \quad (6)$$

を繰り返すことで、軸をひとつずつ求めることができる.

カーネル主成分分析の場合には、(6) 式は

$$\mathbf{K}^{(d+1)} \equiv \mathbf{K}^{(d)} - \mathbf{y}^{(d)} \mathbf{y}^{(d)\top}$$

となる。

3. 提案法：複数情報源に対する主成分分析

本章では複数の情報源がある場合、すなわち、複数のデザイン行列が与えられた場合に対して主成分分析を拡張することを考える。

3.1 情報源選択の定式化

N 個の対象に対して、 M 種類の情報源から、それぞれ D_1, D_2, \dots, D_M 次元のデータが得られたものとする。つまり、 M 個のデザイン行列 $\mathbf{X}_1 \in \mathbb{R}^{N \times D_1}, \mathbf{X}_2 \in \mathbb{R}^{N \times D_2}, \dots, \mathbf{X}_M \in \mathbb{R}^{N \times D_M}$ が得られたとする。これらを並べて $\mathbf{X} \equiv (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M)$ とかく。また、これら M 個のデザイン行列それぞれに対する写像ベクトルを $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M$ とする。そしてこれらをを並べたものを $\mathbf{w} \equiv (\mathbf{w}_1^\top, \mathbf{w}_2^\top, \dots, \mathbf{w}_M^\top)^\top$ と書くことにする。

単純に、各対象に対して $\sum_{m=1}^M D_m$ 次元の特徴ベクトルが与えられたと考えれば、前節で紹介した通常的主成分分析を適用すればよい。

しかし、ここで相対的に重要な情報源、あるいは、役に立たない情報源があった場合に、前者をより重視した次元削減を行いたい。つまり、有用な情報源については相対的に \mathbf{w}_m のノルムが大きく、逆に役に立たない情報源については \mathbf{w}_m のノルムを 0 に近づけるようにしたい。

教師付き学習の文脈では、このような情報源選択を実現するための方法として、グループラッソ [5] で用いられる正則化が知られている。グループラッソでは、通常用いられる 2-ノルムの 2 乗和で定義される正則化項 $\sum_{m=1}^M \|\mathbf{w}_m\|_2^2$ の代わりに、2-ノルムの 1 乗和 $\sum_{m=1}^M \|\mathbf{w}_m\|_2$ を正則化項として用いる。特徴選択を行う際に用いられるパラメータの 1-ノルムを用いた正則化 (L₁-正則化) とのアナロジーでこれを解釈すれば、この正則化項は各情報源に対応するパラメータの 2-ノルムの 1-ノルムであるため、情報源レベルでのパラメータ取舍選択が起こることになる。

この考え方を我々の問題設定に導入すると、主成分分析におけるノルム制約 (2) は、複数情報源の場合においては

$$\sum_{m=1}^M \|\mathbf{w}_m\|_2 \leq 1 \quad (7)$$

であるが、これにかわり、

$$\sum_{m=1}^M \|\mathbf{w}_m\|_2 \leq 1 \quad (8)$$

を用いればよさそうである。従って、最適化問題としては、制約 (8) のもとで、目的関数 (1) を最大化すればよいことになる。

3.2 アルゴリズム

3.2.1 目的関数の下界の導出

我々の最適化問題に対するラグランジアンは

$$L(\mathbf{w}, \lambda) = \mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} - \lambda \left(\sum_{m=1}^M \|\mathbf{w}_m\|_2 - 1 \right)$$

と書くことができる。これは原点で微分不可能であるので、直接の最適化が困難になる。そこで、相加相乗平均から導かれる次の関係

$$\|\mathbf{w}_m\|_2 = \min_{\beta_m \geq 0} \frac{1}{2} \left(\frac{\|\mathbf{w}_m\|_2^2}{\beta_m} + \beta_m \right)$$

を用いることで、

$$L(\mathbf{w}, \lambda) = \max_{\beta \geq 0} \tilde{L}(\mathbf{w}, \beta, \lambda)$$

ただし

$$\begin{aligned} \tilde{L}(\mathbf{w}, \beta, \lambda) &\equiv \mathbf{w} \mathbf{X}^\top \mathbf{X} \mathbf{w} - \lambda \left(\sum_{m=1}^M \frac{1}{2} \left(\frac{\|\mathbf{w}_m\|_2^2}{\beta_m} + \beta_m \right) - 1 \right) \quad (9) \end{aligned}$$

のように書き換えることができる。ただし、 $\beta \equiv (\beta_1, \beta_2, \dots, \beta_M)$ は、下界をつくるために導入された新たなパラメータベクトルである。これが、各情報源の (非負の) 重要度として働く。なお、通常的主成分分析における制約 (7) は、これらがすべての m について等しい ($\beta_m = 1$) である場合に対応している。

さて、上の議論は $L(\mathbf{w}, \lambda)$ の下界が $\tilde{L}(\mathbf{w}, \beta, \lambda)$ によって与えられることを意味している。下界によって、本来 2-ノルムの 1 乗であったところが 2 乗になっており、 β の存在を除いては通常的主成分分析と同じような目的関数となっているところに注意する。従って最適化の戦略としては、この下界 (9) を \mathbf{w} と β について交互に最大化することで局所最適解を得る。

3.2.2 パラメータ \mathbf{w} の最適化

まずは、 β を固定して \mathbf{w} について最大化することを考える。

$$\tilde{\mathbf{w}}_m \equiv \frac{\mathbf{w}_m}{\sqrt{\beta_m}} \quad (10)$$

とおくと、 \tilde{L} は

$$\begin{aligned} \tilde{L}(\tilde{\mathbf{w}}, \lambda) &= (\tilde{\mathbf{w}}_1^\top, \tilde{\mathbf{w}}_2^\top, \dots, \tilde{\mathbf{w}}_M^\top) \left(\sqrt{\beta_1} \mathbf{X}_1, \sqrt{\beta_2} \mathbf{X}_2, \dots, \sqrt{\beta_M} \mathbf{X}_M \right)^\top \\ &\quad \cdot \left(\sqrt{\beta_1} \mathbf{X}_1, \sqrt{\beta_2} \mathbf{X}_2, \dots, \sqrt{\beta_M} \mathbf{X}_M \right) \left(\tilde{\mathbf{w}}_1^\top, \tilde{\mathbf{w}}_2^\top, \dots, \tilde{\mathbf{w}}_M^\top \right)^\top \\ &\quad - \lambda \left(\sum_{m=1}^M \frac{1}{2} (\|\tilde{\mathbf{w}}_m\|_2^2 + \beta_m) - 1 \right) \\ &= \tilde{\mathbf{w}}^\top \tilde{\mathbf{X}}(\beta)^\top \tilde{\mathbf{X}}(\beta) \tilde{\mathbf{w}} - \frac{\lambda}{2} \left(\|\tilde{\mathbf{w}}\|_2^2 + \sum_{m=1}^M \beta_m - 2 \right) \quad (11) \end{aligned}$$

のように書き直すことができる。尚、ここで、

$$\begin{aligned} \tilde{\mathbf{w}} &\equiv \left(\tilde{\mathbf{w}}_1^\top, \tilde{\mathbf{w}}_2^\top, \dots, \tilde{\mathbf{w}}_M^\top \right)^\top \\ \tilde{\mathbf{X}}(\beta) &\equiv \left(\tilde{\mathbf{X}}_1(\beta_1), \tilde{\mathbf{X}}_2(\beta_2), \dots, \tilde{\mathbf{X}}_M(\beta_M) \right) \\ &\equiv \left(\sqrt{\beta_1} \mathbf{X}_1, \sqrt{\beta_2} \mathbf{X}_2, \dots, \sqrt{\beta_M} \mathbf{X}_M \right) \end{aligned}$$

とおいた $\tilde{\mathbf{X}}_m(\beta_m) \equiv \sqrt{\beta_m} \mathbf{X}_m$.

目的関数 (11) の $\tilde{\mathbf{w}}$ についての最大化は、固有値問題

$$\tilde{\mathbf{X}}(\beta)^\top \tilde{\mathbf{X}}(\beta) \tilde{\mathbf{w}} = \frac{\lambda}{2} \tilde{\mathbf{w}}$$

の最大固有ベクトル $\tilde{\mathbf{w}}^{(1)}$ によって達成される.

カーネル主成分分析の場合には,

$$\tilde{\mathbf{w}} \equiv \tilde{\mathbf{X}}(\beta)^\top \tilde{\boldsymbol{\alpha}} \quad (\tilde{\mathbf{w}}_m \equiv \tilde{\mathbf{X}}_m(\beta)^\top \tilde{\boldsymbol{\alpha}})$$

とおくことによって、固有値問題

$$\tilde{\mathbf{K}}(\beta) \tilde{\boldsymbol{\alpha}} = \frac{\lambda}{2} \tilde{\boldsymbol{\alpha}}$$

が得られる. ここで,

$$\tilde{\mathbf{K}}(\beta) \equiv \sum_{m=1}^M \beta_m \mathbf{K}_m \quad (12)$$

$$\mathbf{K}_m \equiv \mathbf{X}_m \mathbf{X}_m^\top$$

とおいた. 例によって、各カーネル行列 \mathbf{K}_m は任意の半正定値行列としてとることができる. また、固有値問題に現れるカーネル行列 (12) は、それぞれのカーネル行列を β の各要素で重みつき和をとったものとなっている. したがって、 β の各要素は各カーネル行列の重要度であるとみることができる.

3.2.3 パラメータ β の最適化

次に、 $\tilde{\mathbf{L}}$ を \mathbf{w} を固定し β について最大化することを考える. 目的関数 (9) を最大化するような β_m は,

$$\beta_m = \|\mathbf{w}_m\|_2 \quad (13)$$

のように簡単に求めることができる. ここで、最大化する前の現在の β_m を新たに β_m^{old} 、最大化後の β_m を β_m^{new} と書くことにすると、(13) 式は

$$\beta_m^{new} = \sqrt{\beta_m^{old}} \|\tilde{\mathbf{w}}_m\|_2$$

とできる. また、これはカーネル主成分分析の場合には,

$$\beta_m^{new} = \beta_m^{old} \sqrt{\tilde{\boldsymbol{\alpha}}^\top \mathbf{K}_m \tilde{\boldsymbol{\alpha}}}$$

となる.

3.2.4 第2主成分以降の抽出

前節の繰り返し最適化を収束するまで繰り返すことによって、第一主成分が抽出できる. 2.2 節で述べたように、その寄与分をデータのデザイン行列 (もしくはカーネル行列) から差し引いたうえで、上記繰り返しアルゴリズムを適用することによって第2主成分が抽出できる. (以下同様のことを繰り返すことによって、任意の数の主成分を抽出できる.)

d 番目の主成分への写像ベクトルによって得られるデータの第 d 主成分 $\mathbf{y}^{(d)}$ は $\mathbf{w}^{(d)}$ のノルムを 1 に正規化したのちに

$$\begin{aligned} \mathbf{y}^{(d)} &\equiv \sum_{m=1}^M \mathbf{y}_m^{(d)} \\ \mathbf{y}_m^{(d)} &\equiv \mathbf{X}_m \mathbf{w}_m^{(d)} \end{aligned} \quad (14)$$

によって得られる. $\mathbf{y}_m^{(d)}$ は $\mathbf{y}^{(d)}$ のうち、 m 番目の情報源による寄与分である.

第 d 主成分を現在のデザイン行列 $\mathbf{X}_m^{(d)}$ ($m = 1, 2, \dots, M$) から差し引くことで、

$$\begin{aligned} \mathbf{X}_m^{(d+1)} &\equiv \mathbf{X}_m^{(d)} - \mathbf{y}^{(d)} \mathbf{w}_m^{(d)\top} \\ &= \mathbf{X}_m^{(d)} - \sqrt{\beta_m^{(d)}} \mathbf{y}^{(d)} \tilde{\mathbf{w}}_m^{(d)\top} \end{aligned} \quad (15)$$

となる. ここで $\beta_m^{(d)}$ は d 番目の主成分についての、各情報源の重要度を表す. (つまり、主成分ごとに各情報源の重要度は異なる.)

カーネル主成分分析の場合には、

$$\mathbf{w}_m^{(d)} = \beta_m^{(d)} \mathbf{X}_m^{(d)\top} \tilde{\boldsymbol{\alpha}}^{(d)}$$

より

$$\|\mathbf{w}^{(d)}\|_2 = \sqrt{\sum_{m=1}^M \beta_m^{(d)2} \tilde{\boldsymbol{\alpha}}^{(d)\top} \mathbf{K}_m^{(d)} \tilde{\boldsymbol{\alpha}}^{(d)}}$$

であることに注意して、まず $\tilde{\boldsymbol{\alpha}}^{(d)}$ を $\tilde{\boldsymbol{\alpha}}^{(d)} := \tilde{\boldsymbol{\alpha}}^{(d)} / \|\mathbf{w}^{(d)}\|_2$ によって正規化する. そして第 d 主成分 (14) は

$$\begin{aligned} \mathbf{y}_m^{(d)} &= \mathbf{X}_m \mathbf{w}_m^{(d)} \\ &= \beta_m^{(d)} \mathbf{X}_m \mathbf{X}_m^{(d)\top} \tilde{\boldsymbol{\alpha}}^{(d)} \\ &= \beta_m^{(d)} \left(\mathbf{K}_m - \sum_{\ell=1}^{d-1} \mathbf{y}_m^{(\ell)} \mathbf{y}_m^{(\ell)\top} \right) \tilde{\boldsymbol{\alpha}}^{(d)} \end{aligned}$$

によって得られる. 最後の変形では、再帰式 (15) から導かれる

$$\mathbf{X}_m^{(d+1)} \mathbf{X}_m^\top = \mathbf{K}_m - \sum_{\ell=1}^d \mathbf{y}_m^{(\ell)} \mathbf{y}_m^{(\ell)\top}$$

を用いた.

そして、 $\mathbf{K}_m^{(d+1)} \equiv \mathbf{X}_m^{(d+1)} \mathbf{X}_m^{(d+1)\top}$ とすると、やはり (15) 式より、新しいカーネル行列は

$$\mathbf{K}_m^{(d+1)} = \left(\mathbf{I} - \beta_m^{(d)} \tilde{\boldsymbol{\alpha}}^{(d)} \mathbf{y}^{(d)\top} \right)^\top \mathbf{K}_m^{(d)} \left(\mathbf{I} - \beta_m^{(d)} \tilde{\boldsymbol{\alpha}}^{(d)} \mathbf{y}^{(d)\top} \right)$$

によって得られる.

4. 数値実験

提案手法である複数情報源に対する情報源の自動選択機構を備えた主成分分析の有効性を検証する. 情報源の自動選択機能とは、すなわち、

(1) 相補的に働く情報源の適切な組み合わせ

(2) 有用でない情報源の切り捨て

を行うことができるということであり、提案手法が本当にこれらの機能を持っているかどうかを確かめるために、観測データからの地図復元のタスクを題材に数値実験を行う.

4.1 観測データからの地図作成タスク

実験で扱う地図作成タスクとは、2次元のユークリッド空間上に配置された物体の相対位置 (地図) を、複数の観測点からの観測 (例えば、ある一定距離内にそれぞれの物体が見える (1) が

否 (0) かなど) を元に復元する問題である [15]. 物体個数に対応する数のデータ (それぞれ観測点数だけの次元をもつ) を 2 次元に次元削減することにより, 地図が推定できることが知られており, 正しい地図を用いて地図の復元精度を評価することができる.

具体的には, 一辺の長さが 2.5 である正方形の中に $N = 50$ 個の物体を一樣にランダムに配置する. 同様に, $D = 500$ 点の観測点もランダムに設定する. いずれもその座標は不明であるとする. ここから各々 50×500 のデザイン行列 X_m を作成し, そこから 50×50 の各カーネル行列 K_m を作成する.

与えられたカーネル行列に対して, 2 次元ユークリッド空間への次元削減を適用し, これを推定された物体の配置図とする. 推定された配置図の評価は, 正しい配置図と比較することにより行うが, スケーリングや向きが実際の地図とは異なるため, 直接の比較を行うことはできない. そこで, 次元削減の性能評価法として, 任意の 3 点の位置関係の正誤を考える. すなわち, 3 つの物体 A, B, C に対し, 推定された配置図と真の配置図において, 時計回りの順序が同じであるかどうかを判定する. この判定は, ベクトル AB とベクトル AC の外積の符号の正負が一致するかどうかによって行うことができる. これを 3 つの物体点の組み合わせすべてについて考え,

$$(\text{誤差}) = \frac{(\text{外積の符号が一致しなかった数})}{(\text{3 つの物体点の組み合わせの総数})} \quad (16)$$

によって推定配置図の精度を測る. なお, 推定配置図は反転している (つまり鏡像の関係は区別できない) 場合があるため, ひとつの主成分の正負を入れ替えたものと比較して誤差の小さいものを評価値として用いる.

4.2 実験 1: 相補的に働く情報源の統合

まずは, 提案手法によって複数の相補的な情報源からのデータを組み合わせることによって, 地図作成の精度が高くなることを確認する. ここでは, 3 つの情報源を仮定する. それぞれの情報源から得られる 3 つのデザイン行列 X_1, X_2, X_3 はそれぞれ次のように作成される. 各情報源はそれぞれある特定の方向 (ランダムに生成された 2 次元の単位ベクトル) をもっているものとする. そして, 各観測点において, それぞれの物体位置のその軸への射影と観測点の距離が観測されるものとする. このように作られた 3 つのデザイン行列から, 3 つの線形カーネル行列 $K_1 \equiv X_1 X_1^T, K_2 \equiv X_2 X_2^T, K_3 \equiv X_3 X_3^T$ を作成し, これらを用いる.

比較手法としては以下の 3 つを用いた.

- 提案手法
- 3 つのカーネル行列の平均をとったカーネル $K = \frac{1}{3}K_1 + \frac{1}{3}K_2 + \frac{1}{3}K_3$ を用いてカーネル主成分分析を行ったもの
- 3 つのカーネル行列をそれぞれ単独で用いてカーネル主成分分析を行ったとき, 最も誤差が小さいもの

表 1 および図 1 はデータの生成と配置図作成を 50 回繰り返して, 各手法の平均誤差と標準偏差を示したものである. これより, 提案手法の方が平均的に誤差が小さいことがわかる. また, 平均カーネルに比べ提案手法のほうが標準偏差が小さく, これ

は 3 つの情報源に対する 3 つの軸の取り方によっては, 平均カーネルにおける等重みでの組み合わせではうまく地図を復元できない場合があるが, 提案手法ではそのような都合の悪い場合でも適切に重みを調整し, 結果として安定して良い推定精度を達成していることがわかる.

	誤差	標準偏差
提案手法	0.294	0.062
平均カーネル	0.327	0.111
単独カーネル	0.459	0.017

表 1 各手法による誤差と標準偏差 (実験 1)

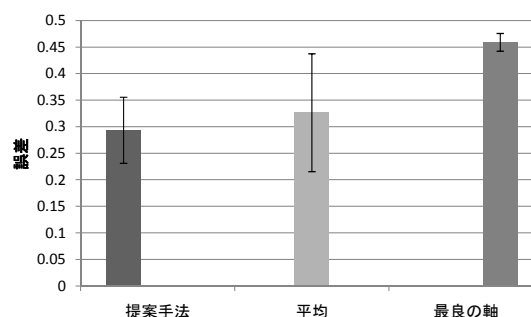


図 1 各手法による誤差と標準偏差 (実験 1)

5. 実験 2: 有用でない情報源の切り捨て

次に, ひとつの有用な情報源と, 複数の有用でない情報源 (ノイズ) が存在するときに, 提案手法が正しい情報源のみを選択し, ノイズの影響を抑えることができるかどうかを検証する.

まず, 有用な情報源から得られるデザイン行列 X_1 は, 各観測点から 0.55 の範囲内にある物体については対応する要素を 1, そうでない場合には 0 として作成した. そして, このデザイン行列から線形カーネル $K_1 = X_1 X_1^T$ を作成した. また, ノイズの行列は, 半正定値であるような 50×50 のランダム行列とした.

ノイズ行列の数を増やしていったときの提案手法と平均カーネル (全てのカーネル行列を等重みで足し合わせたもの) による予測精度の変化を調べたものが, 表 2 および図 2 である. 平均カーネルの性能はノイズ行列の増加とともに悪化していくが, 提案手法の場合はノイズ行列の数が増加しても, ノイズが無い場合とほぼ変わりなくほぼ一定の精度を保っていることがわかる. また, 実際に提案手法で最終的に得られた各情報源の重要度を確認したところ, ノイズ行列に対応する重要度はほぼ 0 に収束していた. これらの結果から, 提案手法が有用な情報のみをうまく選択できていることがわかる.

ノイズ行列の数	0	1	3	5	10
提案手法	0.291	0.278	0.273	0.295	0.282
平均カーネル	0.291	0.376	0.445	0.466	0.476

表 2 各手法による誤差と標準偏差 (実験 2)

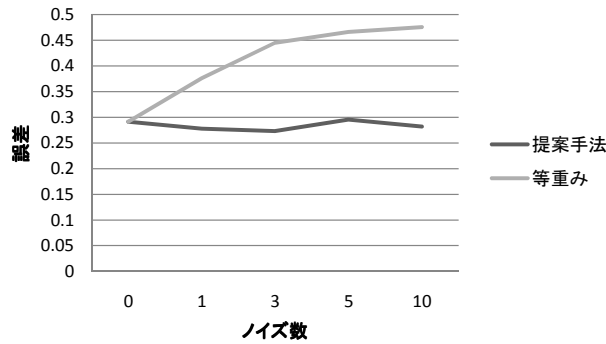


図 2 各手法による誤差と標準偏差 (実験 2)

文 献

- [1] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol.24, pp.417–441, 1933.
- [2] B. Schölkopf, A. Smola, and K.-R. Müller, “Kernel principal component analysis,” In *Lecture Notes in Computer Science*, vol.1327, 1997.
- [3] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural computation*, vol.15, no.6, pp.1373–1396, 2003.
- [4] M.A.A. Cox and T.F. Cox, “Multidimensional scaling,” *Handbook of data visualization*, pp.315–347, 2008.
- [5] M. Yuan and Y. Lin, “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B*, vol.68, pp.49–67, 2006.
- [6] F.R. Bach, G.R.G. Lanckriet, and M.I. Jordan, “Multiple kernel learning, conic duality, and the SMO algorithm,” In *Proceedings of the 21st International Conference on Machine Learning*, 2004.
- [7] G.R.G. Lanckriet, N. Cristianini, P. Bartlett, L.E. Ghaoui, and M.I. Jordan, “Learning the kernel matrix with semi-definite programming,” *The Journal of Machine Learning Research*, vol.5, pp.27–72, 2004.
- [8] F. Yan, J. Kittler, K. Mikolajczyk, and A. Tahir, “Non-sparse multiple kernel learning for fisher discriminant analysis,” In *Proceedings of the 9th Ninth IEEE International Conference on Data Mining*, 2009.
- [9] J. Ye and S.J. Chen, “Learning the kernel matrix in discriminant analysis via quadratically constrained quadratic programming,” In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007.
- [10] A. Argyriou, M. Herbster, and M. Pontil, “Combining graph laplacians for semi-supervised learning,” In *Advances in Neural Information Processing Systems 18*, 2005.
- [11] B. Zhao, J.T. Kwok, and C. Zhang, “Multiple kernel clustering,” In *Proceedings of the 9th SIAM International Conference on Data Mining*, 2009.
- [12] Y. Lin, T. Liu, and C. Fuh, “Multiple kernel learning for dimensionality reductions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- [13] P. Rai, “Co-regularized spectral clustering with multiple kernels,” In *NIP 2010 Workshop: New Direction Multiple Kernel Learning*, 2010.
- [14] Z. Xu, R. Jin, H. Yang, I. King, and M.R. Lyu, “Simple and efficient multiple kernel learning by group lasso,” In *Proceedings of the 27th International Conference on Machine Learning*, 2010.
- [15] T. Yairi, “Map building without localization by dimensionality reduction techniques,” In *Proceedings of the 24th International Conference on Machine Learning*, 2007.