7M-04

共分散記述子のための計量学習によるパターン認識

松澤　知己 †　　レラトー　レイサ ‡　　伊藤　栄祐 †　　加藤　毅 †

† 群馬大学大学院理工学専攻 ‡ 産総研ゲノム情報研究センター

1　はじめに

近年，コンピュータビジョンにおいて，**共分散記述子**による画像表現が注目されている [1]．共分散記述子は正定値行列であり，共分散記述子を使った解析には，アファイン不変距離，対数ユークリッド距離など，これまで様々な距離関数が試されてきた．一方，通常のベクトル空間を使ったパターン認識では，データから距離計量を学習することにより汎化性能が向上することが確かめられている [2]．本研究では，共分散記述子によるパターン認識の性能向上をはかるため，訓練用データから正定値錐上の距離を学習する方法を開発した．

正定値錐上の計量学習は本研究が初ではなく，ITML [2] を拡張した方法が 2015 年にすでに提案されていた [1]．しかし，その学習アルゴリズムは致命的な誤りを含んでいた．本稿では，その誤りを正すとともに，3 節において損失関数のさらなる一般化を行う．4 節では，最適化アルゴリズムにおいて，筆者らが新たに見つけた 2 つの発見を述べる：(i) その各反復を $O(n^3 + Ln)$ (n および L の定義は後述) の計算コストに抑えられる；(ii) 乱択戦略の導入により，反復回数を劇的に減少できる．5 節では，パターン認識実験を通して，計量学習によって汎化性能が顕著に向上することを示す．記法は，文献 [3] 参照．

2　正定値錐上のパラメトリックな距離計量関数

従来，2 つの半正定値行列 $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \mathbb{S}^n_+$ の距離計量関数として，

$$D_{\boldsymbol{\Phi}}(\boldsymbol{X}_1, \boldsymbol{X}_2) := \|\boldsymbol{\Phi}(\boldsymbol{X}_1) - \boldsymbol{\Phi}(\boldsymbol{X}_2)\|_{\mathrm{F}}^2 .$$

であらわされるような距離関数がしばしば検討されてきた．写像 $\boldsymbol{\Phi} : \mathbb{S}^n_{++} \to \mathbb{R}^{n\times m}$ には，$\boldsymbol{\Phi}(\boldsymbol{X}) := \mathrm{logm}(\boldsymbol{X})$，$\boldsymbol{\Phi}(\boldsymbol{X}) := \boldsymbol{X}^p$，$\boldsymbol{\Phi}(\boldsymbol{X}) := \mathrm{chol}(\boldsymbol{X})$ などが選ばれてきた．本研究では，次のパラメトリックな距離計量を導入する：

$$D_{\boldsymbol{\Phi}}(\boldsymbol{X}_1, \boldsymbol{X}_2; \boldsymbol{W}) := \left\langle \boldsymbol{W}, (\boldsymbol{\Phi}(\boldsymbol{X}_1) - \boldsymbol{\Phi}(\boldsymbol{X}_2))(\boldsymbol{\Phi}(\boldsymbol{X}_1) - \boldsymbol{\Phi}(\boldsymbol{X}_2))^\top \right\rangle . \tag{1}$$

ただし，$\boldsymbol{W} \in \mathbb{S}^n_+$ はこの距離関数のパラメータである．もし $\boldsymbol{W}$ が正定値で，$\boldsymbol{\Phi}$ が全単射なら，この距離関数 $D_{\boldsymbol{\Phi}}(\cdot,\cdot;\boldsymbol{W}) : \mathbb{S}^n_{++} \times \mathbb{S}^n_{++} \to \mathbb{R}$ は，非負性，確定性，対称性，三角不等式を満たすので，**距離の公理**に従うことになる．距離関数 (1) において，$\boldsymbol{\Phi}(\boldsymbol{X}) = \mathrm{logm}(\boldsymbol{X})$ とおくと，[1] で提案されていた距離関数に等しくなる．

3　計量学習問題の定式化

本節では，距離関数 (1) におけるモデルパラメータ $\boldsymbol{W}$ の学習方法を提案する．多クラス分類問題を例にとって述べる．訓練用データが $(\boldsymbol{X}_1, \omega_1), \ldots, (\boldsymbol{X}_\ell, \omega_\ell) \in \mathbb{S}^n_+ \times \mathbb{N}_{n_c}$ のように得られていたとする．ただし，$\boldsymbol{X}_i$ は第 i 訓練用例題の共分散記述子で，ω_i はそのクラスラベルである．n_c はクラス数である．これら ℓ 個の訓練用例題に対し，K 個のペアを $(i_1, j_1), \ldots, (i_K, j_K) \in \mathbb{N}_\ell \times \mathbb{N}_\ell$ を考え，各ペアに対して次の制約を与える：

$$D_{\boldsymbol{\Phi}}(\boldsymbol{X}_{i_k}, \boldsymbol{X}_{j_k}; \boldsymbol{W}) \begin{cases} \le b_{\mathrm{ub}}\xi_k & \text{if } \omega_{i_k} = \omega_{j_k}, \\ \ge b_{\mathrm{lb}}\xi_k & \text{if } \omega_{i_k} \ne \omega_{j_k}. \end{cases} \tag{2}$$

ただし，$\xi_k = 1$ のとき，b_{ub} および b_{lb} は，それぞれ，同じクラス間の距離の上限，異なるクラス間の距離の下限を定める定数となる．ここで，$\omega_{i_k} = \omega_{j_k}$ なる $k \in \mathbb{N}_K$ には，$(y_k, b_k) = (+1, b_{\mathrm{ub}})$，$\omega_{i_k} \ne \omega_{j_k}$ なる $k \in \mathbb{N}_K$ には，$(y_k, b_k) = (-1, b_{\mathrm{lb}})$ とおく．この制約を満たす中で，$\boldsymbol{W}$ をなるだけ単位行列に近く，ξ_k をなるだけ 1 に近くなるように最適化問題を定式化すると，

$$\begin{aligned} \min \quad & \mathrm{BD}_\varphi((\boldsymbol{W}, \boldsymbol{\xi}), (\boldsymbol{I}, \boldsymbol{1})) \\ \text{wrt} \quad & \boldsymbol{W} \in \mathbb{S}^n_{++}, \quad \boldsymbol{\xi} = [\xi_1, \ldots, \xi_K]^\top \in \mathbb{R}^K_{++}, \\ \text{subject to} \quad & \forall k \in \mathbb{N}_K, \quad y_k D_{\boldsymbol{\Phi}}(\boldsymbol{X}_{i_k}, \boldsymbol{X}_{j_k}; \boldsymbol{W}) \le y_k b_k \xi_k \end{aligned} \tag{3}$$

のようになる．ただし，$\mathrm{BD}_\varphi(\cdot,\cdot) : (\mathbb{S}^n_{++} \times \mathbb{R}^K_{++}) \times (\mathbb{S}^n_{++} \times \mathbb{R}^K_{++}) \to \mathbb{R}_+$ は，**ブレグマンダイバージェンス** (Bregman divergence, BD) である．$\mathrm{BD}_\varphi((\boldsymbol{W}, \boldsymbol{\xi}), (\boldsymbol{I}, \boldsymbol{1}))$ は $(\boldsymbol{W}, \boldsymbol{\xi}) = (\boldsymbol{I}, \boldsymbol{1})$ の時に限り，0 になり，$(\boldsymbol{W}, \boldsymbol{\xi})$ が $(\boldsymbol{I}, \boldsymbol{1})$ から離れるほど値が大きくなる．$\mathrm{BD}_\varphi(\cdot,\cdot)$ は，連続微分可能で狭義凸な実数値関数 (**シード関数**と呼ぶ) $\varphi : \mathbb{S}^n_{++} \times \mathbb{R}^K_{++} \to \mathbb{R}$ を使って，

$$\mathrm{BD}_\varphi(\boldsymbol{\Theta}, \boldsymbol{\Theta}_0) = \varphi(\boldsymbol{\Theta}) - \varphi(\boldsymbol{\Theta}_0) - \langle \nabla\varphi(\boldsymbol{\Theta}_0), \boldsymbol{\Theta} - \boldsymbol{\Theta}_0 \rangle$$

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.

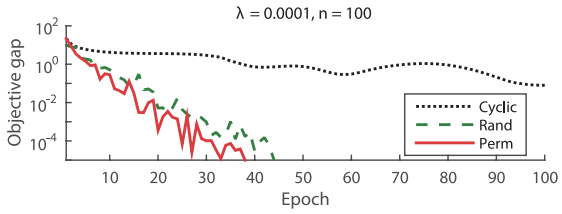

図 1　**最適解への収束の様子．**確率的ダイクストラ法の適用により Objcetive gap が急速に縮まっていることが観測される．

と定義される．ただし，$\boldsymbol{\Theta}, \boldsymbol{\Theta}_0 \in \mathbb{S}^n_{++} \times \mathbb{R}^K_{++}$ である．シード関数は，連続微分可能で狭義凸であれば，任意に定義できる．シード関数の定義によって，$(\boldsymbol{W}, \boldsymbol{\xi})$ の $(\boldsymbol{I}, \boldsymbol{1})$ からの逸脱の測り方が変わり，これによって異なる学習結果を生む．本研究では，次のシード関数は $\varphi(\boldsymbol{W}, \boldsymbol{\xi}) := \varphi_{\mathrm{r}}(\boldsymbol{W}) + \sum_{k=1}^K c_k \varphi_{\mathrm{l}}(\xi_k)$ の形式で与えられるとした．シード関数 φ の第 1 項 $\varphi_{\mathrm{r}} : \mathbb{S}^n_{++} \to \mathbb{R}$ は $\varphi_{\mathrm{r}}(\boldsymbol{W}) := -\mathrm{logdet}(\boldsymbol{W})$ とおき，第 2 項の $\varphi_{\mathrm{l}} : \mathbb{R}_{++} \to \mathbb{R}$ には，次の 3 種類の関数を試した：$\varphi_{\mathrm{is}}(\xi_k) := -\log(\xi_k)$, $\varphi_{\mathrm{l2}}(\xi_k) := \xi_k^2/2$, $\varphi_{\mathrm{e}}(\xi_k) := (\log \xi_k - 1)\xi_k$．1 番目のシード関数 φ_{is} から生成される BD は，[1] で用いられた目的関数と等価になる．

4　最適化アルゴリズム

4.1　巡回的ダイクストラ法

最適化問題 (3) は，$(\mathbb{S}^n_{++} \times \mathbb{R}^K_{++})$ 上の点 $(\boldsymbol{I}, \boldsymbol{1})$ から，K 個の半平面の共通集合へのブレグマン射影 (Bregman projection, BP) を求める問題とみることができる．このような問題を解く方法として，ダイクストラ法 (Dykstra algorithm) [4] を利用できる．ダイクストラ法は，$(\boldsymbol{W}_0, \boldsymbol{\xi}_0) = (\boldsymbol{I}, \boldsymbol{1})$ から開始する反復法で，第 t 反復において，巡回的に半空間を一つ選んで，$(\boldsymbol{W}_{t-1}, \boldsymbol{\xi}_{t-1})$ からその半空間の境界に射影して，$(\boldsymbol{W}_t, \boldsymbol{\xi}_t)$ を得る．この射影を繰り返すと，最適解に収束することが理論的に保証されている．反復 t において，半空間の境界 $\langle \boldsymbol{A}_k, \boldsymbol{W} \rangle = b_k \xi_k$ に射影する問題は，$\boldsymbol{W}_{t-1}^{-1} + \delta y_k \boldsymbol{A}_k \succ \boldsymbol{O}$ を満たす範囲内で，スカラー変数 δ の一変数非線形方程式

$$\left\langle \boldsymbol{A}_k, (\boldsymbol{W}_{t-1}^{-1} + \delta y_k \boldsymbol{A}_k)^{-1} \right\rangle = b_k \nabla \varphi_{\mathrm{l}}^*(\nabla \varphi_{\mathrm{l}}(\xi_{t-1}^{-1}) - \delta y_k b_k / c_k) \tag{4}$$

を解く問題に帰着される．ただし，φ_{l}^* は φ_{l} の凸共役 (convex conjugate) である．文献 [1] では，スペクトラル分解や逆行列計算も不要な閉形式で解を与えてしまったが，これはウッドベリー公式の誤用であり，実際には文献 [1] の更新式ではこの方程式の解にならない．閉形式で求まらないとすると，ニュートン法などの数値算法を頼らざるを得ない．(4) の左辺には $n \times n$ 行列の逆行列を含んでいるため，ナイーブに数値算法にかけてしまうと，計算負荷の高い $n \times n$ 行列の逆行列を収束するまで計算し続けることになる．すると，数値算法内部での反復回数を L とすると，1 回の BP に $O(Ln^3)$ かかることになる．一方，本研究では，$O(n^3 + Ln)$ の計算量で方程式 (4) の根を見つけるアルゴリズムを発見した (詳細は [3] 参照)．

4.2　確率的ダイクストラ法の提案

近年，正則化損失最小化の手法として，確率的勾配法が注目を集めている．確率的勾配法では，例題を巡回的に選ぶのではなく，ランダムに選択することで，早く最適解に収束することが実験的に確かめられている．一方，従来のダイクストラ法は K 個の半空間を巡回的 (**Cyclic**) に選んでいた．本研究では，ダイクストラ法において，K 個の半空間を巡回的に選ぶのではなく，ランダムに選んだり (**Rand**)，1 エポックごとに K 個の半空間の順番を無作為に入れ替える (**Perm**) ことで，飛躍的に最適解への収束が早くなることを発見した (図 1 参照)．半空間の選び方を乱択するように改良したダイクストラ法を**確率的ダイクストラ法**と呼ぶことにする．

5　パターン認識実験

提案法の性能を評価するために，ETH-80，Brodatz texture dataset の 2 つのデータセットで多クラス分類のベンチマークをとった．計量学習無しの場合，それぞれ最高で 79.29%，94.40% であったのに対し，計量学習ありでは，それぞれ 82.90%，96.35% になった．よって，提案法により，汎化性能が向上することが確認された．実験条件と結果の詳細は [3] 参照のこと．

謝辞: 本研究は JSPS 科研費 26249075, 40401236 の助成を受けたものである．

参考文献

[1] Zhiwu Huang et al. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *ICML*, pp. 720–729, 2015.

[2] J. V. Davis et al. Information-theoretic metric learning. In *ICML*, pp. 209–216, New York, NY, USA, 2007. ACM.

[3] T. Matsuzawa et al. Stochastic dykstra algorithms for metric learning on positive semi-definite cone. Technical report, arXiv, 2016. Submitted. Temporary ID: submit/1448442.

[4] Y. Censor and S. Reich. The dykstra algorithm with bregman projections. *Comm. in Applied Anal.*, Vol. 2, pp. 407–419, 1998.

Copyright ©2016 Information Processing Society of Japan.
All Rights Reserved.