Information-Theoretic Metric Learning の分類精度向上を目的とした計量行列の学習手法

1X11C099-0　馬賀 嵩士
指導教員　後藤 正幸

1　研究背景・目的

本研究では，近年ますます重要性が増しつつある電子データに対する分析技術のうち，自動分類問題を対象とする．自動分類問題とは，カテゴリが既知の学習データから分類規則を学習し，カテゴリが未知の入力データのカテゴリを推定する問題である．この問題に対するアプローチの 1 つにデータ間の距離尺度を用いる方法があるが，分類精度が距離尺度に大きく依存することが知られている．そこで，距離尺度にマハラノビス距離を用い，分類に適した距離構造を求めるメトリックラーニングとよばれる手法が存在する．

本研究では，メトリックラーニングの手法のうち，Information-Theoretic Metric Learning (以下 ITML)[1] に着目する．ITML では，学習データのペアに着目し，同一カテゴリのデータペア間の距離を一定値より小さく，別カテゴリのデータペア間の距離を一定値より大きくするような距離構造を学習する．このとき，学習後の識別境界を曖昧にしているデータペアが存在する可能性がある．しかしながら，従来の ITML では全てのデータペアを同等に扱っているため，識別境界の明確化が十分になされない可能性がある．

そこで本研究では，ITML を対象とし，識別境界を明確化するようなデータペアを効率的に選択する方法を提案する．このようなデータペアのみを用いて ITML の学習を行うことで，識別境界がより明確となるような距離構造が学習され，分類精度が向上すると考えられる．提案手法の有効性をベンチマークデータを用いた実験により示す．

2　準備

全データ数が N 個の学習データ集合を $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$，$G$ 個の離散カテゴリ集合を $\mathcal{C} = \{c_g\}_{g=1}^{G}$ とする．$\boldsymbol{x}_i$ は d 次元特徴ベクトル，$y_i \in \mathcal{C}$ は $\boldsymbol{x}_i$ が属するカテゴリとする．また，同一カテゴリに属するデータペア $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の集合を $\mathcal{S}$，別カテゴリに属するデータペア $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の集合を $\mathcal{D}$ とする．

メトリックラーニングは一般に，距離尺度としてマハラノビス距離を仮定し，その計量行列を所望の制約条件のもとで求める方法を指す．いま，計量行列を $A \in \mathbb{R}^{d\times d}$ としたとき，任意の 2 点 $\boldsymbol{x}_i, \boldsymbol{x}_j$ 間のマハラノビス 2 乗距離は式 (1) で定義される．

$$d_A(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} A (\boldsymbol{x}_i - \boldsymbol{x}_j). \tag{1}$$

ただし，T はベクトルの転置を表す．

3　Information-Theoretic Metric Learning

ITML は集合 $\mathcal{S}$ に含まれるデータペアのマハラノビス距離を定数以下に，集合 $\mathcal{D}$ に含まれるデータペアのマハラノビス距離を定数以上にするような計量行列を学習する．この際，目標とする行列 A_0 を定め，A_0 との距離を最小化するように計量行列 A を学習することで過学習の発生を抑制している．A_0 として，経験的に単位行列 I または分散共分散行列の逆行列 Σ^{-1} が用いられる．ここで，行列 A と A_0 間の距離は，式 (2) で表される logdet 情報量により定義する．

$$D_{\ell d}(A, A_0) = \mathrm{tr}(AA_0^{-1}) - \mathrm{logdet}(AA_0^{-1}) - d. \tag{2}$$

ただし，$\mathrm{tr}(\cdot)$ は行列のトレースを示し，$\mathrm{logdet}(\cdot)$ は行列式の対数値を示す．距離の公理を満たすためには，計量行列が半正定値となる必要があるが，logdet 情報量を用いることにより，計量行列の半正定値性を保証するための固有値分解が不要になり，行列のランクを保持できるという利点がある．以下に，ITML における最適化問題を示す．

$$\underset{A}{\text{minimize}} \quad D_{\ell d}(A, A_0) \tag{3}$$

$$\text{s.t.} \quad \mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}) \le u, \quad (i,j) \in \mathcal{S}, \tag{4}$$

$$\mathrm{tr}(A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}) \ge \ell, \quad (i,j) \in \mathcal{D}, \tag{5}$$

$$A \succeq 0. \tag{6}$$

式 (3) は A と A_0 間の logdet 情報量の最小化であり，これが正則化の役目を担っている．また，式 (4) は同一カテゴリに属するデータペア間の距離を u 以下に，式 (5) は別カテゴリに属するデータペア間の距離を ℓ 以上にするための制約条件である．ただし，u, ℓ はデータ間の距離の分布から決定される任意の正の定数とする．式 (6) は行列 A が半正定値対称行列であることを示す．

式 (3)–(6) では，以下のアルゴリズムにより最適解を求める．**Step2** における α は式 (3)–(6) の最適化問題におけるラグランジュ関数の双対変数，β は制約を満たす最小の更新幅を表している．また，式 (4)–(5) を不等式制約として扱うために用いるパラメータを λ_{ij} とする．

Step1) 初期値を $A = A_0$，$\lambda_{ij} = 0$ とする．

Step2) データペア $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を 1 つ選択し，$p = (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} (\boldsymbol{x}_i - \boldsymbol{x}_j)$ とする．

Step2-1) $(i,j) \in \mathcal{S}$ のとき，式 (7) により α, β を計算．

$$\alpha = \min\left(\lambda_{ij}, \frac{1}{p} - \frac{1}{u}\right), \qquad \beta = \frac{\alpha}{1 - \alpha p}. \tag{7}$$

Step2-2) $(i,j) \in \mathcal{D}$ のとき，式 (8) により α, β を計算．

$$\alpha = \min\left(\lambda_{ij}, \frac{1}{\ell} - \frac{1}{p}\right), \qquad \beta = -\frac{\alpha}{1 + \alpha p}. \tag{8}$$

Step3) $\lambda_{ij} = \lambda_{ij} - \alpha$ と更新する．

Step4) 式 (9) に従って，計量行列 A を更新する．

$$A = A + \beta A(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} A. \tag{9}$$

Step5) 収束条件を満たすまで，**Step2–4** を繰り返す．

□

上述の通り，ITML では計量行列の半正定値性を保証するための固有値分解が不要であり，計算量の面で他のメトリックラーニングの手法と比べ有利である．いま，カテゴリ数を G，入力データの次元数を d とすると，1 回の繰り返しは $O(Gd^2)$ の計算量で実行することができる．

4　提案手法

4.1　概要

他のカテゴリの領域に近いデータは，カテゴリ間の境界付近に存在しており，分類のための学習によって重要である可能性が高い．このようなデータをそのデータが属するカテゴリの中心に近付けるように学習を行うことで，識別境界の明確化が図れると考えられる．しかし，従来手法では，学習に

使われるすべてのデータペアが同等に扱われており，識別境界を明確化するような学習を妨げている可能性がある．

そこで，本研究では，$\mathcal{S}$ に含まれるデータペアのうち，カテゴリ間の境界付近に存在するデータを含むデータペアを $\overline{\mathcal{S}}$ の要素として選択し，学習に用いる方法を提案する．これにより，カテゴリごとの識別が容易になり，結果として分類精度が向上すると考えられる．

4.2 データペアの選択

他のカテゴリの領域に近いデータを，そのデータが属するカテゴリの中心に近付けることが目的である．そこで，集合 $\mathcal{S}$ に着目し，他のカテゴリの中心に近いデータとカテゴリの中心に近いデータをペアとすることで，上記の目的を果たすことを考える．

データペアの選択では，カテゴリ c_g の平均ベクトル $\boldsymbol{\mu}_g$ を求める．カテゴリ c_g に属し，$\boldsymbol{\mu}_g$ の近傍にある κ 個のデータを探索し，それらを集合 $\mathcal{N}_g = \{\boldsymbol{n}_g^i\}_{i=1}^{\kappa}$ とする．$\boldsymbol{n}_g^i$ とカテゴリ c_g に属する他の全データとのペアの集合を $\mathcal{P}_g^i = \{(\boldsymbol{n}_g^i, \boldsymbol{x}_j)\}_{j=1}^{N}$ $(y_j = c_g)$ とする．これを $i = 1, \ldots, \kappa$ について行い，カテゴリ c_g における中心付近のデータと他データとのデータペア集合 $\mathcal{P}_g = \cup_{i=1}^{\kappa} \mathcal{P}_g^i$ を得る．さらにこれを全カテゴリについて行い，中心付近のデータと他データとのデータペア集合 $\mathcal{P} = \cup_{g=1}^{G} \mathcal{P}_g$ を得る．ここで，$\mathcal{P}$ に属する全ての $\boldsymbol{x}_j$ に対し，他カテゴリの平均 $\boldsymbol{\mu}_g (c_g \neq y_j)$ との距離の最小値を L_j とする．L_j が小さいものから θ 個のデータペアを集めた集合を $\overline{\mathcal{S}}$ とする．ただし，集合 $\mathcal{N}_g, \mathcal{P}$ を求める際に用いる距離尺度は，式 (1) において $A = A_0$ としたものを用いるものとする．$\kappa = 3$，$\theta = 5$ とした場合の例を図 1 に示す．図 1 では，破線で囲まれた θ 個のペアが集合 $\overline{\mathcal{S}}$ の要素となる．

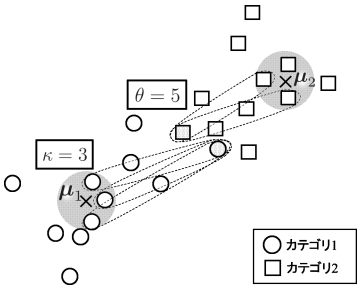

図 1. 提案アルゴリズムのイメージ

4.3 提案アルゴリズム

提案手法のアルゴリズムを以下に示す．

Step1) カテゴリ c_g の平均ベクトル $\boldsymbol{\mu}_g$ を求める．
Step2) $\boldsymbol{\mu}_g$ を用いて，集合 $\mathcal{N}_g = \{\boldsymbol{n}_g^i\}_{i=1}^{\kappa}$ を求める．
Step3) $\boldsymbol{n}_g^i$ について，$\mathcal{P}_g^i$ を求める．
Step4) **Step3** を $i = 1, \ldots, \kappa$ について行い，$\mathcal{P}_g$ を得る．
Step5) **Step1–4** を全カテゴリについて行い，$\mathcal{P}$ を得る．
Step6) $\mathcal{P}$ に属するすべての $\boldsymbol{x}_j$ について L_j を求める．
Step7) $\mathcal{P}$ の要素のうち，L_j が小さい θ 個のデータペアを $\overline{\mathcal{S}}$ の要素とする．
Step8) 集合 $\overline{\mathcal{S}}, \mathcal{D}$ を用いて ITML を実行する．

□

このアルゴリズムにより，$\overline{\mathcal{S}}$ に属するデータペアは，θ 個生成されることになる．

5 実験

5.1 実験条件

提案手法の有効性を示すため，ベンチマークデータセットに対して分類実験を行い，提案手法の分類誤り率の評価を行った．分類誤り率は以下の式 (10) により求める．

$$\text{分類誤り率} = 1 - \frac{\text{正しく分類されたテストデータ数}}{\text{テストデータ数}} \quad (10)$$

実験では，UCI 機械学習レポジトリのベンチマークデータセット 2 種類 (Iris, Bal) を用いた．また，ここでは，$A_0 = I$，Σ^{-1} のそれぞれの場合について，従来手法および提案手法を適用し，実験を行った．集合 $\mathcal{S}, \mathcal{D}$ については，共に $20G^2$ 個のデータペアをランダムに抽出した．パラメータ κ, θ は予備実験により以下の表 1 の通り設定した．

表 1. データセットの概要とパラメータ

データ	次元	カテゴリ	学習	テスト	$A_0 = I$		$A_0 = \Sigma^{-1}$	
セット名	数	数	データ数	データ数	κ	θ	κ	θ
Iris	4	3	120	30	5	20	5	20
Bal	4	3	500	125	5	30	6	30

実験では，結果は 100 回の試行の平均値とする．

5.2 実験結果と考察

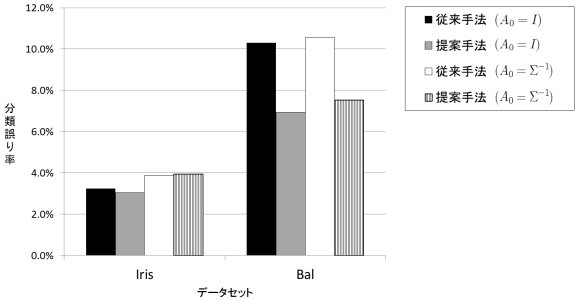

図 2. 分類誤り率の実験結果

実験結果を図 2 に示す．Bal データセットについては，A_0 に I，Σ^{-1} のどちらを用いた場合においても提案手法の分類精度が向上していることがわかる．提案手法では，他のカテゴリの中心に近いデータが識別境界の明確化に寄与していると考えているが，これはデータの分布が球状に近く，かつカテゴリ間の分布の重なりがそれほど大きくないという状況を暗黙のうちに仮定している．そのため，このような特徴を持つデータである Bal データセットでは精度の向上が見られたと考えられる．しかしながら，Iris のようにもともと分類精度が高いデータセットは，カテゴリ間の境界が比較的明確であり，学習データの選択が精度の向上にそれほど大きくは寄与しないことがわかった．また，平均値は外れ値に影響を受けやすいため，外れ値を多く含む場合に対しては精度の向上が図れないことが想定される．

6 まとめと今後の課題

本研究では，ITML における学習データペアの選択アルゴリズムを提案した．また，評価実験によって，一部のデータセットにおいて分類精度が向上することが確認できた．今後の課題としては，集合 $\mathcal{D}$ に対するデータペア選択アルゴリズムの提案，パラメータの定義の改良，およびその自動決定アルゴリズムなどが挙げられる．

参考文献

[1] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-Theoretic Metric Learning,” *Proc, the 24th International Conference on Machine Learning*, pp. 209–216, 2007.