局所的距離学習と適応的重み付け和に基づく分類精度向上に関する研究

1X12C084-4　中野修平
指導教員　後藤正幸

1　研究背景と目的

近年の情報技術の発展に伴い，カテゴリが既知の学習データから分類規則を学習し，カテゴリが未知のデータのカテゴリを推定する自動分類手法の研究が盛んに行われている．自動分類に対するアプローチの一つに，k 近傍法などのデータ間の距離に着目した手法があるが，分類性能が用いる距離尺度に依存することが知られている．その性能向上のため，学習データの統計的特徴を考慮した距離構造を任意の制約条件のもとで学習するメトリックラーニングと呼ばれる手法が提案されている．特に，複数の計量行列を学習し，その線形結合により計量行列を表現するマルチタスク型メトリックラーニングが近年盛んに研究されている．

本研究では，マルチタスク型メトリックラーニング手法の一つであり，SVM などで用いられるマージンの概念とタスクの振り分け方法に確率構造を用いた Large Margin Local Metric Learning (以下 LMLML) [1] に着目する．LMLML では，学習データに対して各タスクへの所属確率を算出し，それをもとに各タスクごとに計量行列の学習を行う．さらに，各データペアに対し両者の各ガウス分布への所属確率をもとに算出した重みを用いることにより，一つの計量行列を定義しその学習を行う．しかしながら，この重み付け法には，学習データに付与されているカテゴリ情報を考慮せずにタスクへの振り分けを行うこと，データペアに対する重み算出の際に，偏った重みを付与してしまう可能性があることの 2 つの問題点が存在する．一方で，カテゴリ情報を用いたタスク分けとデータペアの双方にとって重要なタスクへの重みを大きくすることができれば，これらの問題点を解決できる可能性がある．そこで本研究では，学習した計量行列を用いたタスク振り分け法とデータペアの各分布への所属確率の積による計量行列への重み付け法を提案し，分類精度の向上を図る．UCI 機械学習レポジトリを用いた評価実験により提案手法の有効性を示す．

2　準備

いま，カテゴリが既知である N 個の学習データ集合 $\mathcal{D}=\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ を用いて，カテゴリが未知の新規入力データ $\boldsymbol{x}_t$ の所属カテゴリを推定することを考える．$\boldsymbol{x}_i$, $\boldsymbol{x}_t \in \mathbb{R}^d$ は d 次元の特徴ベクトル，$\mathcal{C}=\{c_1, c_2, \cdots, c_G\}$ を G 個の離散カテゴリ集合とし，$y_i \in \mathcal{C}$ は $\boldsymbol{x}_i$ の所属カテゴリとする．

メトリックラーニングとはデータ間の距離尺度にマハラノビス距離を仮定し，マハラノビス距離における計量行列 $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ を任意の制約条件の下で学習する手法である．ここで，任意の 2 点 $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathbb{R}^d$ の間のマハラノビス距離の 2 乗 $d^2(\boldsymbol{x}, \tilde{\boldsymbol{x}}, \boldsymbol{A})$ を以下の式 (1) で定義する．

$$d^2(\boldsymbol{x}, \tilde{\boldsymbol{x}}, \boldsymbol{A}) = (\boldsymbol{x}-\tilde{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}-\tilde{\boldsymbol{x}}) \tag{1}$$

ただし，T はベクトルの転置を表す．

3　Large Margin Local Metric Learning

3.1　概要

LMLML はマルチタスク型メトリックラーニング手法の一つであり，マージンの概念とタスクの振り分けに混合ガウス分布を用いた手法である．LMLML では，混合ガウス分布を用いて類似した学習データをまとめたものをタスクと呼ぶ．具体的には，データに対して混合ガウス分布の各ガウス分布への所属確率を用いてタスク分けを行い，タスク毎に計量行列を導出することで局所的な距離構造を考慮した学習を行う．

3.2　最適な計量行列の導出法

LMLML では混合数 S の混合ガウス分布に加え，学習データ全体から求めた計量行列 $\boldsymbol{M}_0$ を加えた S+1 個の計量行列行列集合 $\mathcal{M}=\{\boldsymbol{M}_0, \boldsymbol{M}_1, \cdots, \boldsymbol{M}_S\}$ を求める．また，各計量行列の導出には SVM 等で用いられるマージンの概念を用いる．

いま，$\boldsymbol{x}_i$ のガウス分布 s に対する所属確率を $P(s|\boldsymbol{x}_i)$, $\boldsymbol{x}_i$ の生起確率を $P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s^{-1}\right)$ として，データペア $\{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ ごとに重み $w_s(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を用いて算出する一つの計量行列 $\boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を以下の式 (2)–(5) で定義する．ただし，$\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \pi_s$ は各ガウス分布の平均ベクトル，分散共分散行列，混合比，β は $\beta>0$ とする任意の定数を表す．

$$\boldsymbol{M}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \sum_{s=0}^{S} w_s\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\boldsymbol{M}_s \tag{2}$$

$$w_s\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \begin{cases} \beta & s=0 \\ P\left(s|\boldsymbol{x}_i\right)+P\left(s|\boldsymbol{x}_j\right) & otherwise \end{cases} \tag{3}$$

$$P\left(s|\boldsymbol{x}_i\right) = \frac{\pi_s P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s^{-1}\right)}{\sum_{s=1}^{S}\pi_s P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s^{-1}\right)} \tag{4}$$

$$P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s^{-1}\right) = \frac{\exp\left(-\frac{1}{2}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_s\right)^{\mathrm{T}}\boldsymbol{\Sigma}_s^{-1}\left(\boldsymbol{x}_i-\boldsymbol{\mu}_s\right)\right)}{(2\pi)^{\frac{d}{2}}\left|\boldsymbol{\Sigma}_s\right|^{\frac{1}{2}}} \tag{5}$$

すなわち，$\boldsymbol{M}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ は S+1 個の計量行列とそれぞれの対応する重み $w_s(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を用いて導出され，重み $w_s(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の導出には，データペアの各ガウス分布への所属確率の和を用いている．これらを用いることで，最適な計量行列集合 $\mathcal{M}$ は以下の最適化問題を解くことにより導出される．

$$\begin{aligned} \underset{\mathcal{M}, \boldsymbol{\alpha}}{\text{minimize}}\ & \boldsymbol{H}\left(\mathcal{M}, \boldsymbol{\alpha}\right) \\ &= \frac{1}{N}\sum_{\{\boldsymbol{x}_i, \boldsymbol{x}_j\}\in\mathcal{D}} l\left(y_{ij}, d^2\left(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{M}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\right)\right) \\ &\quad + \lambda\sum_{s=0}^{S}\left\|\boldsymbol{M}_s-\alpha_s\boldsymbol{I}\right\|_F \\ \text{subject to}\ & \boldsymbol{M}_s \succeq 0 \quad (s=0,1,\ldots,S) \end{aligned} \tag{6}$$

ここで，$\|\cdot\|_F$ は行列のフロベニウスノルム，$\boldsymbol{\alpha}=\{\alpha_0, \alpha_1, \cdots, \alpha_S\}$ はチューニングパラメータの集合，$\boldsymbol{I}$ は d 次元の単位行列，$\lambda(\geq 0)$ は正則化パラメータ，$\boldsymbol{M}_s \succeq 0$ は $\boldsymbol{M}_s$ が半正定値対称行列であることを示す．また，$l(\cdot)$ はヒンジロス関数，$y_{ij} \in \{1, -1\}$ は $\{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ が同一カテゴリならば 1，別カテゴリならば -1 をとるインジケータ関数である．

これらにより，過学習を抑制しながらデータペア $\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ が同一カテゴリならば近付け，別カテゴリならばマージンよりも遠ざけるように計量行列を学習する．

3.3　最適解導出のアルゴリズム

式 (6) における最適解 $\mathcal{M}, \boldsymbol{\alpha}$ を以下のアルゴリズムにより求める．ここで $\mathcal{L}=\{\boldsymbol{L}_0, \boldsymbol{L}_1, \cdots, \boldsymbol{L}_S\}$ は式 (7) を満たす $\boldsymbol{L}_s \in \mathbb{R}^{d\times d}$ の行列集合，∇ は関数の勾配を表す演算子，ε は更新幅，t は更新回数とする．

$$\boldsymbol{M}_s = \boldsymbol{L}_s^{\mathrm{T}}\boldsymbol{L}_s \tag{7}$$

Step1) EM アルゴリズムにより混合数 S の混合ガウス分布のパラメータ $\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \pi_s$ を推定する (s=1,2,$\cdots$,S).
Step2) t=0 とし，初期値設定として，各 s に対し，${\boldsymbol{L}_s^{(0)}}^{\mathrm{T}}\boldsymbol{L}_s^{(0)} = 0.1 \times \boldsymbol{I}$，$\alpha_s^{(0)}= 0.1$ とする (s=0,1,$\cdots$,S).
Step3) 式 (7) を用いて，$\boldsymbol{M}_s^{(t)}$ を算出する．
Step4) 学習データからランダムに m 個を選択する．
Step5) 式 (3) を用いて $w_s(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を計算する．
Step6) $\mathcal{L}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$ を以下の式 (8), (9) を用いて更新する．

$$\boldsymbol{L}_s^{(t+1)} := \boldsymbol{L}_s^{(t)} + \varepsilon\left(\nabla_{\boldsymbol{L}_s^{(t)}}\boldsymbol{H}\left(\mathcal{M},\boldsymbol{\alpha}\right)\right) \tag{8}$$

$$\alpha_s^{(t+1)} := \alpha_s^{(t)} + \varepsilon\left(\nabla_{\alpha_s^{(t)}}\boldsymbol{H}\left(\mathcal{M},\boldsymbol{\alpha}\right)\right) \tag{9}$$

Step7) 収束条件を満たしていれば終了．さもなくば $t = t+1$ として **Step3** へ戻る． □

4 提案手法

4.1 概要

LMLML では，混合ガウス分布を用いて $P(s|\boldsymbol{x}_i)$ を算出し，その所属確率によりタスク分けを行っている．しかし，混合ガウス分布の推定の際には学習データが持つカテゴリ情報を用いておらず，適切なタスク分けが行えていない可能性がある．他方，式 (3) により，データペア $\{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ における各計量行列の重みを導出する際には，$\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ の各ガウス分布への所属確率の和を用いている．しかし，あるガウス分布に対して，一方のデータの所属確率が極めて高く，もう一方のデータの所属確率が低い場合，そのタスクの重みはある程度の大きさを持ってしまう．そのため，もう一方の所属確率の低いデータにとって重要性が低いにも関わらず，そのタスクが過度に重視されてしまう．これらの問題に対して，本研究では学習の過程で導出される計量行列を用いたタスク分けとデータペアの双方にとって重要な計量行列の重みを強くすることで分類精度の向上を図る．

4.2 改良重み付け法

式 (5) の右辺の分子から，式 (4) で算出される $P(s|\boldsymbol{x}_i)$ は距離尺度に $\boldsymbol{\Sigma}_s^{-1}$ を用いた $\boldsymbol{x}_i$ と $\boldsymbol{\mu}_s$ の距離と解釈できる．しかし，$\boldsymbol{\Sigma}_s^{-1}$ はカテゴリ情報が考慮されておらず，$\boldsymbol{M}_s^{(t)}$ の更新に関わらず一定であるため，学習経過を考慮したタスク分けが行えていない．そこで，$P(s|\boldsymbol{x}_i)$ の算出の際に，距離尺度にカテゴリを考慮し学習した $\boldsymbol{M}_s^{(t)}$ を用いることで，二点の関係性を考慮した距離計量を算出できると考えられる．すなわち，式 (10) のように，$P^{(t)}(s|\boldsymbol{x}_i)$ を改良する．

$$P^{(t)}(s|\boldsymbol{x}_i) = \frac{\pi_s P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{M}_s^{(t)}\right)}{\sum_{s=1}^{S}\pi_s P\left(\boldsymbol{x}_i|\boldsymbol{\mu}_s, \boldsymbol{M}_s^{(t)}\right)} \tag{10}$$

加えて，データペアの双方に重要である $\boldsymbol{M}_s$ を考慮するため，重み $w_s^{(t)}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の算出を $\boldsymbol{x}_i, \boldsymbol{x}_j$ のガウス分布 s への各所属確率の積を用いて式 (11) により行う．

$$w_s^{(t)}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} \beta & s = 0 \\ \frac{P^{(t)}(s|\boldsymbol{x}_i)\times P^{(t)}(s|\boldsymbol{x}_j)}{\sum_{s=1}^{S}P^{(t)}(s|\boldsymbol{x}_i)\times P^{(t)}(s|\boldsymbol{x}_j)} & otherwise \end{cases} \tag{11}$$

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) EM アルゴリズムにより混合数 S の混合ガウス分布のパラメータ $\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s, \pi_s$ を推定する (s=1,2,$\cdots$,S).
Step2) $\boldsymbol{\Sigma}_s$ を用いて $\boldsymbol{\Sigma}_s^{-1}$ を算出する．
Step3) t=0 とし，初期値設定として，各 s に対し，${\boldsymbol{L}_s^{(0)}}^{\mathrm{T}}\boldsymbol{L}_s^{(0)} = \boldsymbol{\Sigma}_s^{-1}$，$\alpha_s^{(0)} = 0.1$ とする (s=1,2,$\cdots$,S). また，s=0 に対して ${\boldsymbol{L}_0^{(0)}}^{\mathrm{T}}\boldsymbol{L}_0^{(0)} = 0.1 \times \boldsymbol{I}$ とする．
Step4) 式 (7) を用いて，$\boldsymbol{M}_s^{(t)}$ を算出する．
Step5) 学習データからランダムに m 個を選択する．
Step6) 式 (10), (11) を用いて $w_s^{(t)}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を計算する．
Step7) $\mathcal{L}^{(t)}$, $\boldsymbol{\alpha}^{(t)}$ を式 (8), (9) を用いて更新する．
Step8) 収束条件を満たしていれば終了．さもなくば $t = t+1$ として **Step4** へ戻る． □

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を用いて分類実験を行った．分類手法は k=3 の k 近傍法，評価指標は以下の式 (12) で定義される分類誤り率を用いた．

$$\text{分類誤り率} = 1 - \frac{\text{正しく分類したテストデータ数}}{\text{テストデータ数}} \tag{12}$$

実験では，UCI 機械学習レポジトリから 3 種類 (wine, letter, ionoshpere) を用いた．また，比較手法として従来手法の LMLML を用いた．データセットの概要を表 1 に示す．

表 1．データセットの概要

データセット名	次元数	カテゴリ数	学習データ数	テストデータ数
wine	13	3	142	36
letter	16	26	14000	6000
ionosphere	34	2	280	70

各パラメータは，予備実験より $\beta = 0.3$, $\gamma = 1$, $\lambda = 15$ とし，混合数 S は 1 から 20 の範囲で各 100 回の実験を行い最良の結果を用いることとした．

5.2 実験結果と考察

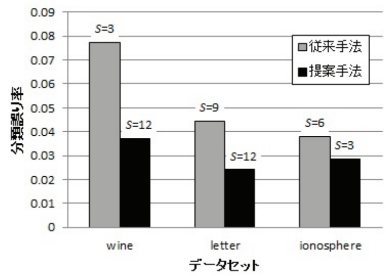

図 1．各データセットにおける実験結果

実験結果を図 1 に示す．全てのデータセットにおいて，提案手法は従来手法と比較し低い分類誤り率を得る事ができた．提案手法では，データの各ガウス分布に対する所属確率の計算に $\boldsymbol{M}_s^{(t)}$ を用いている．これは，各ガウス分布の形状を学習データのカテゴリ情報を用いて推定していることに等しい．そのため，カテゴリ情報をデータの各ガウス分布に対する所属確率に反映させ，より分類に適したタスク分けが行えたと考えられる．また，各タスクに対する重みの導出方法を変更することで，カテゴリが異なるデータペア $\{\boldsymbol{x}_i, \boldsymbol{x}_j\}$ の距離計量の導出に $P(s|\boldsymbol{x}_i)$ と $P(s|\boldsymbol{x}_j)$ の値が共に高いタスクが重視された．以上のことから，従来手法と比較して提案手法の分類精度が向上したと考えられる．

6 まとめと今後の課題

本研究では計量行列を用いた所属確率の導出と所属確率の積を用いた重み付け方法を提案した．また，評価実験により，提案手法が分類精度の面で優れていることを示した．今後の課題として，適切な混合数 S の決定方法などが挙げられる．

参考文献

[1] J. Bohné, Y. Ying, S. Gentric, and M. Pontil, “Large Margin Local Metric Learning”, *Proc. ECCV 2014*, pp. 679–694, 2014.