多クラス分類のための Top-k 多重カーネル学習

Top-k multi-kernel learning for multiclass classification

竹内勇気 Yuki Takeuchi　富井和彦 Kazuhiko Tomii　加藤毅 Tsuyoshi Kato

群馬大学大学院理工学府
Graduate School of Science and Technology, Gunma University

1　はじめに

パターン認識の性能は，特徴と損失関数の選択によって決まる．本論文では，この 2 つの要素を同時に改良する新しい試みを報告する．

多くの応用において，用いる特徴セットには複数の選択肢がある．それらの中で最適な特徴はしばしば不明である．多重カーネル学習 (**MKL**) [1] の枠組みは，特徴セットの選択に対する有力な解を提供する．MKL では，各特徴セットを等価の表現であるカーネルで表現し，カーネルを線形結合したものでパターン識別器の学習を行う．MKL は線形結合における重み係数とパターン識別器のパラメータの同時最適化を実現する枠組みである．MKL の枠組みは，単一カーネルでの正則化リスク最小化が確実に可能であることを仮定している [1]．

近年の多クラス分類に用いられるベンチマークデータセットではクラス数が増大するにつれ，**Top-k 誤分類率**が重視されるようになった．従来，多クラス分類器の学習には **Max** ヒンジ損失が広く用いられてきたが，これは Top-k 誤分類率を最適化するための損失関数ではなかった．これに対し，Lapin ら [2] は **Top-k 誤分類率**を最適化できるように **Top-k ヒンジ損失**

$$\Phi(\boldsymbol{z}\,;y) := \frac{1}{k}\max\left(0, \sum_{p=1}^{k}(\boldsymbol{z} - z_y\mathbf{1} + \mathbf{1} - \boldsymbol{e}_y)_{[p]}\right)$$

を設計した．ただし，$\boldsymbol{z} \in \mathbb{R}^m$ は予測スコア，$y \in \{1,\dots,m\}$ はクラスラベル，m はクラス数である．しかし，本研究では，Lapin らが導出した理論に重大な誤りがあるために正則化リスク最小化が不可能なことを突き止めた [3]．正則化リスク最小化が実現できなければ，Top-k ヒンジ損失を MKL の枠組みに導入することもできない．

著者らは，Lapin ら [2] の理論的誤りを訂正することで，Top-k 誤分類率から定義されるリスクと特徴セットの組み合わせの同時最適化を実現する新しい MKL アルゴリズムを開発した．

2　Top-k 多重カーネル学習 (Top-k MKL)

MKL では，M 個の $n\times n$ カーネル行列 $\boldsymbol{K}^1,\dots,\boldsymbol{K}^M$ を，重み係数 $\boldsymbol{\beta} := [\beta_1,\dots,\beta_M]^\top \in \boldsymbol{\Delta}$ を使って，

$$\boldsymbol{K}(\boldsymbol{\beta}) = \sum_{j=1}^{M}\beta_j\boldsymbol{K}^j \tag{1}$$

のように統合する．$\boldsymbol{\Delta} \subseteq \mathbb{R}_+^M$ は確率単体である．統合したカーネル行列 $\boldsymbol{K}(\boldsymbol{\beta})$ の正則化リスクは

$$\Omega(\boldsymbol{\beta}) := \max_{\boldsymbol{A}\in\mathbb{R}^{m\times n}} D(\boldsymbol{A};\boldsymbol{K}(\boldsymbol{\beta})) \tag{2}$$

と表される [1]．ただし，$D(\boldsymbol{A};\boldsymbol{K})$ は，

$$D(\boldsymbol{A};\boldsymbol{K}) := -\frac{\lambda}{2}\langle\boldsymbol{A},\boldsymbol{AK}\rangle - \frac{1}{n}\sum_{i=1}^{n}\Phi^*(-\boldsymbol{\alpha}_i\,;y_i)$$

とする．y_i は第 i 例題のクラスラベルとし，$\Phi^*(\cdot;y_i)$ は $\Phi(\cdot;y_i)$ の凸共役とする．$\boldsymbol{\alpha}_i$ は $\boldsymbol{A}$ の第 i 列である．正則化リスク $\Omega(\boldsymbol{\beta})$ を最小化するには，導関数

$$\frac{\partial\Omega(\boldsymbol{\beta})}{\partial\beta_j} := -\frac{\lambda}{2}\left\langle\boldsymbol{A}(\boldsymbol{\beta}),\boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{K}^j\right\rangle \tag{3}$$

が必要となる [1]．ただし $\boldsymbol{A}(\boldsymbol{\beta}) \in \operatorname*{argmax}_{\boldsymbol{A}\in\mathbb{R}^{m\times n}} D(\boldsymbol{A}\,;\boldsymbol{K}(\boldsymbol{\beta}))$ とする．よって，目的関数の値を計算するには，勾配法で最小解を探索するときにも，任意の $\boldsymbol{K}(\boldsymbol{\beta})$ に対して，$D(\boldsymbol{A}\,;\boldsymbol{K}(\boldsymbol{\beta}))$ を最大化する $\boldsymbol{A}$ を求めるアルゴリズムが必要となる．Lapin ら [2] も $D(\boldsymbol{A}\,;\boldsymbol{K})$ について議論しているが，凸共役 $\Phi^*(\cdot;y_i)$ を誤って導出してしまったために Lapin らのアルゴリズムでは最適な $\boldsymbol{A}$ を見つけられない．著者らは，Lapin らの導出した実行可能領域が真の領域より狭いために最適解に到達できないことを発見し，正しい実行可能領域で $D(\boldsymbol{A}\,;\boldsymbol{K})$ を最大化するアルゴリズムを開発した [3]．本研究では，そのアルゴリズムと SimpleMKL [1] を組み合わせることで Top-k ヒンジ損失に基づくリスク最小化を行う MKL アルゴリズムを新たに開発した．

3　実験

重みなしで統合したカーネルと提案法について複数の公開データセットを使って比較実験を行ったところ，表 1 に示すように，提案法の Top-k 誤分類率が統計的有意に減少することを確認した．

表 1　平均 Top-k 誤分類率 (%)

Coil 100	$k=2$	$k=3$	$k=4$	$k=5$
重みなし	3.887	2.473	1.849	1.441
Top-k MKL	**3.328**	**2.113**	**1.532**	**1.242**

参考文献

[1] A. Rakotomamonjy et al. SimpleMKL. *JMLR*, 9(Nov):2491-2521, 2008.

[2] M. Lapin et al. Top-k multiclass SVM. In *NIPS 2015*, pages 325-333, 2015.

[3] 竹内 勇気ら．Top-k SVM 学習のための双対座標上昇法．In *FIT2017*，第 2 分冊，pages 269-270, Sept 2017.