

# 境界曖昧性に基づく分類問題最適化の実験的評価

## Experimental Evaluation of Boundary-Uncertainty-Based Classifier Optimization

蔭山昌幸 David Ha 千田将大 大崎美穂 片桐滋  
 Masayuki Kageyama David Ha Masahiro Senda Miho Ohsaki Shigeru Katagiri  
 同志社大学  
 Doshisha University

### 1. はじめに

ベイズ境界性の性質を基に分類境界のベイズ境界性尺度を定義し、学習された分類器（パラメータ）群から最適な分類器を選択する手法、Bayes Boundary-ness based Selection (BBS) 法の研究が行われている<sup>[1]</sup>。本稿では、この BBS 法のベイズ境界性を再定式化した手法の評価実験を行う。

### 2. ベイズ境界性

理想的な最小分類誤り確率、ベイズエラーの状態においては、分類決定が不確実である性質<sup>[2]</sup>、ベイズ境界性 (Bayes Boundary-ness) が成り立つ。

BBS 法ではこの性質を基に、学習によって推定された境界が理想的な分類境界（ベイズ境界）にどれだけ近い性質を有しているかを尺度値として評価する。この尺度値は分類境界上とその近傍を含めた分類境界付近の極狭い範囲における標本を用いて計算される。ここで境界近傍の標本も用いるのは、有限個の標本を扱う場合、境界上に標本が確実に存在することが保証されていないためである。

### 3. 再定式化された BBS 法

この手法では  $J$  個のクラス分類問題に関して、各分類境界を構成する 2 つのクラスラベルペア  $\{k, l\}$  に注目する。このペアの集合を  $I^*$  とし、対応する分類境界  $B_{kl}(\Lambda)$  のベイズ境界性尺度推定値  $U_{kl}(\Lambda)$  を計算する。各分類境界に関して境界近傍標本の候補集合  $B^*(k, l)$  を選択し、それらの標本数のばらつきによる尺度値への重みを  $P(k, l)$  とすると、全体のベイズ境界性尺度値  $U(\Lambda)$  を次のように導かれる。

$$U(\Lambda) = \sum_{\{k,l\} \in I^*} P(k, l) U_{kl}(\Lambda) \quad (1)$$

候補集合  $B^*(k, l)$  内の各標本  $x$  の近傍標本を分類器の識別関数値に基づく 1 次元空間上に写像し、カーネル関数を用いて境界近傍標本を選択する。標本  $x$  に対する分類境界への重みを  $w$  とすると、ベイズ境界性尺度値  $U_{kl}(\Lambda)$  は各標本  $x$  に関する尺度値  $U(x; \Lambda)$  を用いて次のように導かれる。

$$U_{kl}(\Lambda) = \sum_{x \in B^*(k, l)} -U(x; \Lambda) * w \quad (2)$$

従来の BBS 法では各標本  $x$  に関するベイズ境界性尺

度値  $U(x; \Lambda)$  を定義するのにシャノンエントロピーを用いていたが、本手法では次式の関数を用いて尺度値を定式化する。

$$U(x; \Lambda) = 1 - |2 * P(C_m | x; \Lambda) - 1| \quad (3)$$

式中の  $P(C_m | x; \Lambda)$  は標本  $x$  に対する事後確率の値であり、実際の操作では、選択した近傍標本群を用いた計算による推定値を採用する。

### 4. 評価実験

評価実験には、Wine Quality White 等の 7 種のデータセットを使用した。いずれも似た結果が得られたがその 1 例を示す (図 1)。なお、用いた分類器はサポートベクターマシン (Support Vector Machine) であり、そのハイパーパラメータである正則化係数  $C$  は実験で得た一定の値とし、カーネル幅である  $\gamma$  は 2 ずつ累乗していく。比較として十分収束した交差検証 (CV) 法 (5 fold) を採用した。

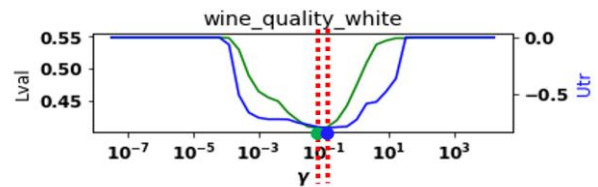


図 1. 実験結果. Wine Quality White は、クラス数は 3 で、次元数は 11、標本数は 3961 である。

図中、左側の縦軸は CV 法における検証用データに対する分類誤り率の推定値  $Lval$ 、右側の縦軸は本手法によるベイズ境界性尺度の推定値  $Utr$ 、横軸は各学習ステップにおける  $\gamma$  の値を表す。緑色の曲線は  $Lval$  の値の推移を、青色の曲線は  $Utr$  の値の推移を示す。 $Lval$  の最小値 (緑色の丸) と  $Utr$  の最小値 (青色の丸) は非常に近く、本手法がベイズエラーを推定する上で有用な手法であることが判断できる。

**謝辞:** 本研究は科研費 (18H03266) の支援を受けた。

### 参考文献

- [1]. D. Ha, et al.; J of Sig Process Sys, pp. 135–151, 2020.
- [2]. K. Fukunaga; “Introduction to Statistical Pattern Recognition (2nd Ed.),” Academic Press (1990).