ベイズリスク推定を目指す誤分類尺度の分析

友利　宥也†　　ア　デイビッド†　　渡辺　秀行††　　大崎　美穂†　　片桐　滋†
† 同志社大学　　　†† 国際電気通信基礎技術研究所

1. はじめに

パターン認識の分類器設計における究極の目標は，最小分類誤り確率状態（ベイズリスク）の達成である．しかし，ベイズリスクの推定を行う為には分類器の（クラス境界）表現能力（結果的にクラスモデルサイズ）の最適化が必要である．モデルサイズが大き過ぎるとベイズリスクを過小推定し，モデルサイズが小さ過ぎると過大推定してしまう．

本研究では最小分類誤り(MCE: Minimum Classification Error)学習法[1]で定義されている誤分類尺度に注目し，ベイズリスク推定を行うための誤分類尺度が満たすべき条件について分析を行う．

2. 誤分類尺度

入力標本$\boldsymbol{x}\in\boldsymbol{X}$が与えられた時，これを J 個のクラス$(C_1,\ldots,C_J)$のうち一つに分類する問題を考える．ここで$\boldsymbol{X}$は無限個の入力標本集合である．分類は以下の分類規則によって行われる：

$$C(\boldsymbol{x})=C_k \ \text{ iff } \ k=\arg\max_j g_j(\boldsymbol{x};\boldsymbol{\Lambda}) \tag{1}$$

ここで$g_j(\boldsymbol{x};\boldsymbol{\Lambda})$は識別関数であり，入力標本の$C_j$に対する帰属度を表す．また$\boldsymbol{\Lambda}$は学習対象である分類器モデルのパラメータである．

誤分類尺度は識別関数$g_j(\boldsymbol{x};\boldsymbol{\Lambda})$を用いて式(2)のように定義される：

$$d_y(\boldsymbol{x};\boldsymbol{\Lambda})=-g(\boldsymbol{x};\boldsymbol{\Lambda})+\log\left[\frac{1}{J-1}\Sigma_{\mathrm{j,j\neq y}}\exp\left(\eta g_j(\boldsymbol{x};\boldsymbol{\Lambda})\right)\right]^{1/\eta} \tag{2}$$

$\eta\to\infty$において次式のように簡単化される．

$$d_y(\boldsymbol{x};\boldsymbol{\Lambda})=-g_y(\boldsymbol{x};\boldsymbol{\Lambda})+\max_{j,j\neq y} g_j(\boldsymbol{x};\boldsymbol{\Lambda}) \tag{3}$$

誤分類尺度は標本の分類正誤の程度を示す尺度であり，正負により誤分類と正分類を示している．本研究では式(4)のように勾配のノルムで正規化した幾何マージン型尺度を用いる．

$$D_y=\frac{d_y(\boldsymbol{x},\boldsymbol{\Lambda})}{\|\nabla_x d_y(\boldsymbol{x},\boldsymbol{\Lambda})\|} \tag{4}$$

3. パルツェン推定

ベイズリスク Rは以下の式のように定義される：

$$R(\Lambda)=\sum_{y=1}^{J}P(C_y)\int_X 1\left(D_y(\boldsymbol{x};\boldsymbol{\Lambda})>0\right)p\left(\boldsymbol{x}\middle|C_y\right)dx \tag{5}$$

この R は，式(6)のように誤分類尺度空間の正領域における積分によって置き換えることができる[2]：

$$R(\Lambda)=\sum_{y=1}^{J}P(C_y)\int_0^\infty p\left(z\middle|C_y\right)dz \tag{6}$$

ここで，Rを求めるために確率密度$p\left(z\middle|C_y\right)$を推定しなければならないが，誤分類尺度空間に写像された分類結果に対し式(7)のようにパルツェン窓を適用することでその推定を行うことができる：

$$R(\Lambda)=\frac{1}{\mathrm{N}}\sum_{y=1}^{J}\sum_{n=1}^{N_y}\int_0^\infty\frac{1}{h_y}\varphi\left(\frac{\mathrm{z}-\mathrm{d_y}(x_n^y;\Lambda)}{h_y}\right)dz \tag{7}$$

ここでφはパルツェン窓関数であり，窓幅 h_yはクラス毎に交差確認最尤推定法等で決定することができる．また，クラス境界付近の標本のスパース性を考慮し，各標本の窓幅に関して重み係数を付与することで可変窓幅のパルツェン推定を行うこともできる [3].

4. 誤分類尺度上のクラス境界とベイズリスクの関係

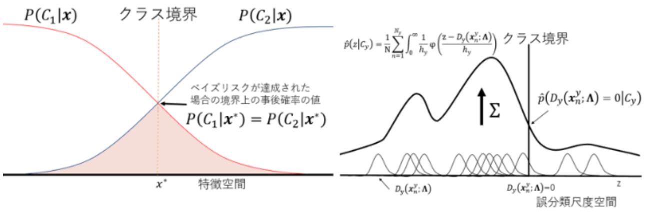

図 1　標本特徴空間と誤分類尺度上のクラス境界.

図 1 に，2 クラス分類問題に対するベイズリスク状態を図解する．標本特徴空間において，クラス境界上の事後確率は両クラス共に等しい．また，この標本特徴空間におけるクラス境界上の標本は，誤分類尺度空間においては全て$D_i(\boldsymbol{x};\boldsymbol{\Lambda})=0$の位置に写像される（$i=1,2$）．従って，次式

$$P\left(C_i\middle|D_i(\boldsymbol{x};\boldsymbol{\Lambda})\right)=\frac{p(D_i(\boldsymbol{x};\boldsymbol{\Lambda})|C_i)P(C_i)}{\sum_{i=1}^{2}p(D_i(\boldsymbol{x};\boldsymbol{\Lambda})|C_i)P(C_i)} \tag{8}$$

を経て，ベイズリスクが達成されている場合の必要条件として，誤分類尺度上の$D_i(\boldsymbol{x};\boldsymbol{\Lambda})=0$において以下の式が成り立つ：

$$p(D_1(\boldsymbol{x};\boldsymbol{\Lambda})=0|C_1)P(C_1)=p(D_2(\boldsymbol{x};\boldsymbol{\Lambda})=0|C_2)P(C_2) \tag{9}$$

ここで両辺それぞれの前項は，誤分類尺度上のパルツェン推定によって求められる．従って，その推定精度が高ければ，この式(9)を満たす学習結果が目指すベイズリスク状態を含むことが期待される．

謝辞：本研究の一部は，科研費（番号：26280063）及び私学研究基盤形成支援事業「ドライバ・イン・ザ・ループ」の支援を受けて行われた．

参考文献：

[1] B.-H. Juang and S. Katagiri, ``Discriminative learning for minimum error classification", IEEE Trans. Signal Processing, vol. 40, no. 12, pp. 3043-3054, Dec. 1992.
[2] E. McDermott and S. Katagiri, ``A derivation of minimum classification error from the theoretical classification risk using Parzen estimation", Comp. Speech Lang., vol. 18, pp. 107-122, April 2004.
[3] D. Ha, ``Class Model Size Optimization using Large Geometric Margin Minimum Classification Error Training for Prototype-Based Classifiers", Doshisha University, Master thesis, 2016.