H-001

学習を伴った近傍平均パターンを利用する識別法
Classification Using Neighbor Averaging Pattern with Learning

堀田 政二 †
Seiji Hotta

1. はじめに

近傍平均パターンを利用した識別法 (Local Averaging Classifier, LAC) [1, 2] に，学習ベクトル量子化 (LVQ) [3] のアイデアを取り入れることで，高次元パターンに頑健な識別を少数の学習パターンで実現する手法を提案する．具体的には，平均パターンを計算するのに用いた全ての k 近傍参照ベクトルを，エラー率が低下するように Relevance LVQ [4] の学習則を GLVQ [5] で置き換えた GRLVQ [6] により更新する．手書き数字データである USPS を用いた実験により提案手法の有効性を示す．

2. LAC と GRLVQ

はじめに LAC [1, 2] と GRLVQ [6] を概説する．

2.1 LAC

クラス $j(j = 1, ..., C)$ に属す d 次元の第 i パターンを $\boldsymbol{x}_i^{(j)} = (x_{i1}^{(j)} \cdots x_{id}^{(j)})^\top (i = 1, ..., n_j)$ で表す．LAC では，未知パターン $\boldsymbol{q} = (q_1 \cdots q_d)^\top$ が与えられた場合，はじめに $\boldsymbol{q}$ の k 近傍学習パターンを各クラスで何らかの距離関数 $d(\boldsymbol{q}, \boldsymbol{x}_i^{(j)})$ を用いて求める．次に，各クラスで選ばれた k 近傍パターンの平均パターンを

$$\boldsymbol{m}^{(j)} = \frac{1}{k}\sum_{l=1}^{k} \boldsymbol{x}_l^{(j)} \quad (j = 1, ..., C) \tag{1}$$

で求める．識別では求めた平均ベクトルと未知パターンとの距離が最小となるクラスを識別結果として出力する:

$$\omega = \arg\min_j d(\boldsymbol{q}, \boldsymbol{m}^{(j)}) \tag{2}$$

LAC は $k = 1$ のとき最近傍決定則 (NN) と等価となり，全てのクラスで $k = n_j$ であれば最小距離識別法と等価となる．実験により LAC は NN よりも識別率が向上することが示されており [1]，特に次元が高いパターンに対して有効であることが示されている [2].

2.2 GRLVQ

Hammer らは RLVQ [4] の学習則を GLVQ [5] で置き換えることで，各次元に重み付けを行い，分類に有用な次元を強調してさらなる分類性能の向上を目指した一般化適合学習ベクトル量子化 (GRLVQ) [6] を提案した．以下に GRLVQ を概説する．GRLVQ では，ベクトルの各次元に対する重みとして $\boldsymbol{\lambda} = (\lambda_1 \cdots \lambda_d), \lambda_i \geq 0$ を導入し，参照ベクトルと重みの更新を繰り返す．このときパターン間の非類似度は以下の重み付きユークリッド距離により計算される:

$$d^\lambda(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_\lambda^2 = \sum_{i=1}^{d} \lambda_i (x_i - y_i)^2 \tag{3}$$

GRLVQ では，GLVQ と同様の評価関数を用いて学習が定義される．すなわち，入力パターン $\boldsymbol{x}$ と同じクラスの最近傍参照ベクトルを $\boldsymbol{c}^{(1)}$，異なるクラスの最近傍参照ベクトルを $\boldsymbol{c}^{(2)}$ とし，それらとの重み付きユークリッド距離値を $d_1^\lambda = \|\boldsymbol{x} - \boldsymbol{c}^{(1)}\|_\lambda^2$，$d_2^\lambda = \|\boldsymbol{x} - \boldsymbol{c}^{(2)}\|_\lambda^2$ とする．GRLVQ では，評価関数として

$$\mu(\boldsymbol{x}) = \frac{d_1^\lambda - d_2^\lambda}{d_1^\lambda + d_2^\lambda} \tag{4}$$

を考え，GLVQ と同様に式 (4) を最急降下法で最小化する．したがって参照ベクトルの更新は以下のように与えられる [6]:

$$\boldsymbol{c}^{(1)} \leftarrow \boldsymbol{c}^{(1)} + \alpha \frac{\partial f}{\partial \mu} \frac{d_2^\lambda}{(d_1^\lambda + d_2^\lambda)^2} (\boldsymbol{x} - \boldsymbol{c}^{(1)}) \tag{5}$$

$$\boldsymbol{c}^{(2)} \leftarrow \boldsymbol{c}^{(2)} - \alpha \frac{\partial f}{\partial \mu} \frac{d_1^\lambda}{(d_1^\lambda + d_2^\lambda)^2} (\boldsymbol{x} - \boldsymbol{c}^{(2)}) \tag{6}$$

ただし，$0 < \alpha < 1$ は学習率である．GRLVQ では同じクラスに属する参照ベクトルとの距離が小さく，異なるクラスに属する参照ベクトルとの距離が大きくなるように $\boldsymbol{\lambda}$ も更新する．最急降下法に基づく重みの更新は $\lambda_i \leftarrow \lambda_i - \epsilon \frac{\partial S}{\partial \lambda_i}$ で与えられるため，結局以下のように更新すればよい:

$$\lambda_i \leftarrow \lambda_i - \epsilon \frac{\partial f}{\partial \mu} \left(\frac{d_2^\lambda (x_i - c_i^{(1)})^2 - d_1^\lambda (x_i - c_i^{(2)})^2}{(d_1^\lambda + d_2^\lambda)^2} \right) \tag{7}$$

ここで ϵ は α よりも小さい正の実数である．最後に，$\boldsymbol{\lambda}$ の要素の合計値が 1 となるように $\lambda_i = \lambda_i / \sum_{j=1}^{d} \lambda_j$ と規格化する．GRLVQ の分類精度は GLVQ と同程度かそれ以上になることが実験で示されている [6]．また GRLVQ では分類に不要な特徴値に対する重みが 0 となることから，データによっては次元削減も同時に可能となる．

3. GRLLAC

GRLVQ は NN に基づく手法であるが，識別率が高い LAC と組み合わせれば識別率が向上すると期待できる．したがって，ここでは GRLVQ と LAC を組み合わせた GRLLAC を提案する．GRLLAC では，はじめに式 (3) を利用して $\boldsymbol{x}$ の k 近傍参照ベクトルを各クラスで求める．次に求めた k 近傍参照ベクトルの平均ベクトルを

$$\boldsymbol{m}^{(j)} = \frac{1}{k}\sum_{l=1}^{k} \boldsymbol{c}_l^{(j)} \tag{8}$$

で求めた後，各平均ベクトルとの距離を再び式 (3) で計算する．この結果，$\boldsymbol{x}$ と同じクラスに属する平均参照ベ

† 東京農工大学　Tokyo University of Agriculture and Technology.

表 **1. USPS** でのエラー率 **[%]** と標準偏差.

n_j	NN	LVQ1	GRLVQ	LAC	GRLLAC	SVM
10	22.7 ± 1.5	9.5 ± 0.5	$\mathbf{7.9 \pm 0.3}$	19.3 ± 0.9	$\mathbf{7.7 \pm 0.3}$	22.1 ± 1.3
50	12 ± 0.8	7.5 ± 0.3	7.6 ± 0.4	9.3 ± 0.7	$\mathbf{6.5 \pm 0.3}$	11.6 ± 0.9
100	9.4 ± 0.6	6.7 ± 0.4	7.2 ± 0.3	7.3 ± 0.4	$\mathbf{6.1 \pm 0.2}$	8.6 ± 0.5
150	8.1 ± 0.5	6.4 ± 0.4	6.6 ± 0.3	6.6 ± 0.4	$\mathbf{5.6 \pm 0.3}$	8.1 ± 0.4
200	7.5 ± 0.4	6.1 ± 0.3	6.3 ± 0.3	6.0 ± 0.3	$\mathbf{5.4 \pm 0.2}$	7 ± 0.4
250	6.9 ± 0.3	6 ± 0.2	5.9 ± 0.3	5.9 ± 0.4	$\mathbf{5.3 \pm 0.2}$	6.6 ± 0.3
complete set	5.5			4.4		4.2 [7]

クトル $\boldsymbol{m}^{(1)}$ と，クラスの異なる最近傍平均参照ベクトル $\boldsymbol{m}^{(2)}$ を求めることができる．また，それらとの重み付き距離を d_1^λ，d_2^λ とする．GRLLAC では $\boldsymbol{m}^{(j)}(j = 1, 2)$ を直接修正することができないので，$\boldsymbol{m}^{(j)}$ を計算するために用いた全ての k 近傍参照ベクトル $\boldsymbol{c}_l^{(j)}(l = 1, ..., k)$ に対して以下の更新を行う:

$$\boldsymbol{c}_l^{(1)} \leftarrow \boldsymbol{c}_l^{(1)} + \alpha \frac{\partial f}{\partial \mu} \frac{d_2^\lambda}{(d_1^\lambda + d_2^\lambda)^2} (\boldsymbol{x} - \boldsymbol{c}_l^{(1)}) \tag{9}$$

$$\boldsymbol{c}_l^{(2)} \leftarrow \boldsymbol{c}_l^{(2)} - \alpha \frac{\partial f}{\partial \mu} \frac{d_1^\lambda}{(d_1^\lambda + d_2^\lambda)^2} (\boldsymbol{x} - \boldsymbol{c}_l^{(2)}) \tag{10}$$

この修正により $\boldsymbol{m}^{(1)}$ は $\boldsymbol{x}$ に近づき，$\boldsymbol{m}^{(2)}$ は $\boldsymbol{x}$ から遠ざかる．一方，各次元の重みの更新では平均ベクトルと入力パターンとの差分を利用して以下のように更新する:

$$\lambda_i \leftarrow \lambda_i - \epsilon \frac{\partial f}{\partial \mu} \left(\frac{d_2^\lambda (x_i - m_i^{(1)})^2 - d_1^\lambda (x_i - m_i^{(2)})^2}{(d_1^\lambda + d_2^\lambda)^2} \right) \tag{11}$$

ただし $m_i^{(j)}$ はベクトル $\boldsymbol{m}^{(j)}$ の第 i 要素を表す．最後に，$\boldsymbol{\lambda}$ の要素の合計値が 1 となるように $\lambda_i = \lambda_i / \sum_{j=1}^d \lambda_j$ と規格化する．$k = 1$ のとき本手法は GRLVQ と一致する．

4. 実験

手書き数字パターンの USPS を用いた実験結果を示す．USPS は 7291 枚の学習パターンと 2007 枚のテストパターンからなる．実験では，学習パターンをランダムに半分にわけ，一方を参照ベクトルの初期値，もう一方を学習のための学習パターンとして用いた．参照ベクトルの個数を 10 から 250 個に変化させながら学習を行い，テストパターンでエラー率を評価した．この実験を各個数あたり独立に 10 回ずつ行い，エラー率の平均と標準偏差を求めた．なお，16×16 ピクセルの画素値を一次元に並べたものを特徴ベクトルとした．

表 1 に結果を示す．比較として NN, LVQ1, GRLVQ, LAC, SVM (ガウシアンカーネル) の結果も示す．GRLVQ と GRLLAC の学習率は $\alpha = 0.03, \epsilon = \alpha/10$ とし，最大学習回数を 100 回とした．GRLLAC では k と参照ベクトルとして学習パターンに対するエラー率が最小となったものを用いた．GRLVQ ベースの識別器以外では距離としてユークリッド距離を用いた．さらに，LAC と SVM では，テストパターンに対する最良のエラー率を与えるパラメータでの結果を表に示す．表から，GRLLAC が他の手法よりも低いエラー率を達成しているのが分かる．また，GRLLAC では，全ての学習データを用いた NN のエラー率 (5.5%) と同程度のエラー率を約 1/3 のデータ数で達成できた．

5. おわりに

近傍平均パターンを利用した識別法 (LAC) と GRLVQ を組み合わせた GRLLAC を提案し，手書き数字データ USPS を用いた実験により提案手法の有効性を示した．提案手法により，全学習パターンの約 1/3 のデータ数で全ての学習データを用いた最近傍決定則のエラー率と同程度のエラー率を達成することができた．提案手法を画像以外のパターンに適用することや，Distance Metric Learning [8] との比較が今後の課題である．

参考文献

[1] Mitani, Y., Hamamoto, Y.: Classifier design based on the use of nearest neighbor samples. Proc. of 15th ICPR. **2** (2000) 773–776

[2] 堀田ら: 未知パターンとカテゴリカル k 近傍平均パターンとの距離に基づくパターン識別. 信学論. **Vol. J88-D-II**(8) (2005) 1357–1366

[3] Kohonen, T.: Self-organizing maps. 2nd Ed. Springer-Verlag, Heidelberg. (1995)

[4] Bojer, T., *et al.*: Relevance determination in learning vector quantization. ESANN’01. (2001) 271–276

[5] Sato, A., Yamada, K.: Generalized learning vector quantization. NIPS. (1995) 423–429

[6] Hammer, B., Villmann, T.: Generalized relevance learning vector quantization. Trans. Neural Networks. **15** (2002) 1059–1068

[7] Schölkopf, B. *et al.*: Comparing support vector machines with gaussian kernels to radial basis function classifiers. A.I. Memo No. 1599. Massachusetts Institute of Techology. (1996)

[8] Yang, L.: Distance metric learning: A comprehensive survey. Ph.D. Durvey. (2006)