

6F-2

## 所属性質問を利用した区分的線形識別規則の追加学習

天元 宏      工藤 峰一      外山 淳      新保 勝

北海道大学工学部

## 1 はじめに

パターン認識系において、訓練サンプル数が限られるという問題に対処するために各種の質問をパターン認識系に導入する試みが行なわれている [1,2]。特に、所属性質問を利用して仮想的に無限の訓練サンプルを得ることで識別規則を追加学習することは、識別規則に依らず有効な方法である。従来の試み [2] では、識別規則として部分クラス法 [3] を用いていたため、識別規則の構成に必要な計算量が膨大になるという難点があった。本研究では、より計算量の少ない区分的線形識別規則を用いてこの方法を実現することを考える。また、効率良く所属性質問を行うために、与えられたパターンと識別境界の距離を計算する効率的なアルゴリズムを提案する。さらに、本研究の有効性を英文字パターンの認識実験において確認する。

## 2 所属性質問を利用した追加学習

計算論的学習理論の分野における「質問を用いた学習パラダイム」は Angluin が提案し、定式化した新しい学習方法である [4]。所属性質問とは、学習アルゴリズムが示す任意のサンプルに対して、それがどのクラスに属するかを問う質問である。計算論的学習理論においてはオラクル（神託）がその問いに答える。

パターン認識系において所属性質問を実現する場合、ユーザーがオラクルの役割を果たすことができる。しかし、このためには、特徴空間上の任意の点に対応するパターンを可視パターンとしてユーザーに示す必要がある。ところが、一般に特徴空間から観測空間への写像は困難である。そこで、所属性質問を疑似的に実現するために以下のようなアルゴリズムを用いる [2]。

1. 訓練サンプル集合を用いて初期の識別規則を構成する。

2. 人工パターン生成器を用いて観測空間にパターンを発生させ、それを特徴空間上の点に写像する。
3. その点が所属性質問の採用基準を満たすまでステップ2を繰り返す。
4. 選んだ点に対応する観測空間上のパターンをユーザーに示し、所属性質問を行なう。
5. ユーザーはそのパターンを見て所属を答える。
6. その点を訓練サンプル集合に追加し、識別規則を更新する。
7. 現在の識別規則が十分な精度に達したと判断されるまで、ステップ2~6を繰り返す。この判断は追加サンプルに関する誤識別の割合で決定する。

本来の意味での所属性質問が実現可能であるならば、特徴空間上の領域中の任意の点に対して質問を行ない、訓練サンプルとは無関係に完全な学習を行なえる。それに対して本研究では、任意の点ではなくランダムに発生した点のうち識別境界付近のものに関してのみ所属性質問を行なうという点で、所属性質問の疑似的実現にとどまっている。

## 3 区分的線形識別規則

Sklansky and Michelotti による区分的線形識別規則の構成方法 [5] を用いる。この方法では、始めに訓練サンプルをクラスタリングし、その代表点のみを使用する。次に、異なるクラスに属する代表点間で互いに  $k$  番目まで最近となる対を見つけ、それらの対をなるべく多く切断するような方法で超平面を作る。

## 4 所属性質問を行なう点の採用基準

人工的に生成する点のうち、現在推定している識別境界とのユークリッド距離が十分小さい数  $\delta$  以下のものを新しい訓練サンプルとして採用し、ユーザーに対して所属性質問を行う。

Supplementary Learning of Piecewise Linear Classifiers  
Using Membership Queries.  
Hiroshi Tenmoto, Mineichi Kudo, Jun Toyama and Masaru Shimbo.  
Faculty of Engineering, Hokkaido University.  
Kita 13, Nishi 8, Sapporo 060, Japan.  
E-mail: tenmo@huie.hokudai.ac.jp

## 5 識別境界とサンプル点の距離計算

区分的線形識別規則の識別境界は、複数の超平面で構成されるため、全体の識別境界はかなり複雑な形状となる。この境界と任意のサンプル点の距離を計算するには、超平面のすべての組合せの共通部分に対してサンプル点を正射影し、その点が実際に識別境界上にあるかどうかを判定し、境界までの最小距離を求める。

**procedure** { サンプルと識別境界の距離を計算する }

**input:** サンプル  $\mathbf{x}$ ,

超平面の集合  $\mathcal{H} = \{H_i\} (i = 1, \dots, m)$

**output:** 距離  $d(\mathbf{x}, \mathcal{H})$

**begin**

$X := \phi;$

**for**  $k:=1$  **to**  $\min(m, N)$  **begin**

**for**  $\forall (H_{i_1}, H_{i_2}, \dots, H_{i_k})$  **begin**

$\mathbf{x}^* := \{ \mathbf{x} \text{ の } \bigcap_{j=1}^k H_{i_j} \text{ への正射影点} \};$

**if** {  $\mathbf{x}^*$  が識別境界上 } **then begin**

$X := X \cup \{ \mathbf{x}^* \};$

{  $H_{i_1}, H_{i_2}, \dots, H_{i_k}$  をマーク };

**end**

**end**

**if** {  $H_1, H_2, \dots, H_m$  がすべて

マークされている } **then break;**

**end**

$d := \min_{\mathbf{x}^* \in X} \| \mathbf{x}^* - \mathbf{x} \|;$

**return**  $d$

**end**

ただし、 $N$  は特徴数を示す。

## 6 実験

ETL-3[6] の英文字データに対して認識実験を行なった。訓練サンプルは各クラス 100 個、テストサンプルには訓練サンプルとは別に各クラス 100 個を用いた。特徴ベクトルは Characteristic Loci[7] により得られた 81 次元データを KL 展開して上位から 30 個を用いた。所属性質問の採用基準にあたる  $\delta$  は 0.01 とした。

結果の詳細は当日述べる。

## 7 問題点

区分的線形識別規則が識別境界付近の訓練サンプルの追加に対して敏感であるため、認識率は大局的には向上

するものの、局所的には大きく変動する。そのため、どの時点で追加学習を打ち切るのが良いのかの判断がしにくいという問題がある。

## 8 まとめ

所属性質問を利用して区分的線形識別規則を追加学習する方法を提案した。また、これを計算機上の実装し、実際に英文字データに対して学習実験を行なった。その結果、この方法を用いることにより単純にサンプルを追加するよりも認識率の向上が見込めることを確認した。

本方法での認識率は識別規則の性能に大きく依存しているため、今後の課題として、識別規則の構成方法をより追加学習に適した形に改良することが考えられる。

## 文献

- [1] M.Kudo and M.Shimbo, Supplementary Learning of Discrimination Rules Using Oracles and Queries in Concept Learning, *Advances in Structural and Syntactic Pattern Recognition*, Edited by H.Bunke, World Scientific, 1992, 141-150.
- [2] M.Kudo, K.Mizukami, Y.Nakamura and M.Shimbo, Realization of Membership Queries in Character Recognition, submitted to *Pattern Recognition*.
- [3] M.Kudo and M.Shimbo, Optimal subclasses with dichotomous variables for feature selection and discrimination, *IEEE Transactions on Systems, Man, and Cybernetics*, 19(1989), 1194-1199.
- [4] D.Angluin, Queries and Concept Learning, *Machine Learning*, 2(1988), 319-342.
- [5] J.Sklansky and L.Michelotti, Locally Trained Piecewise Linear Classifier, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(1980), 101-110.
- [6] ETL3, ETL character database, Electrotechnical Laboratory / Japanese Technical Committee for Optical Character Recognition, 1993.
- [7] H.A.Glucksman, Classification of Characteristic Loci, *Proceedings of the IEEE World Congress on Computational Intelligence*, 1967, 138-141.