

ANFISを用いたパターン識別における ノイズクラスタリング機構の導入

第 7 グループ 北森 頌規

1. はじめに

ANFIS [1] は、ファジィ If-Then ルールを利用することで高木-菅野ファジィ推論システム (TS-FIS) を導入した、説明可能なニューラルネットワークの一種である。本論文では、ノイズファジィクラスタリングの概念に基づいた多クラス分類のためのロバストな ANFIS モデルを構築する手法を提案する。分類器の学習に際して、教師クラスラベルにおけるノイズの存在を想定し、ノイズ個体による影響を排除することを考慮する。

2. ノイズファジィクラスタリングに基づく ロバストな ANFIS 分類器

多クラス (C クラス) 分類のための ANFIS では、5 層のニューラルネットワークアーキテクチャによってファジィ推論を行う。 n 個体の m 次元入力と C 次元出力のペア $((x_{i1}, \dots, x_{im})^\top, (y_{i1}, \dots, y_{iC})^\top)$ があり、個体 i がクラス c に所属するとき $y_{ic} = 1$ 、それ以外を $y_{ic} = 0$ とするときに、 K 個の Fuzzy If-Then ルールを用いて識別器を構築する。

主成分分析によるロバストな k -means 法 [2] では、個体 i の非ノイズ度とノイズ度を表すファジィメンバシップ u_i とその補数 $1 - u_i$ をそれぞれ利用したロバストな PCA を k -means 法に導入し、クラスター中心 \mathbf{b}_k をロバストに求めるための以下の目的関数の最小化を考える。

$$J_{rkm} = \sum_{i=1}^n \left((1 - u_i)^\theta \gamma + u_i^\theta \sum_{k=1}^K \sum_{i \in G_k} \|\mathbf{x}_i - \mathbf{b}_k\|^2 \right) \quad (1)$$

θ がファジィ度を調整する重みであり、1 に近づけるほどクリスプな分割に近づく。 γ は、各個体とノイズクラスとの距離であり、全個体で共通の値を持つ。

実データの分析では、全個体に正しいラベルを付与することは困難であり、不正確なノイズラベルをもつ個体の影響を排除して分類器構築を行う必要がある。そこで提案法では、多クラス分類のための ANFIS においてノイズロバスト性を導入するアプローチとして、ノイズファジィクラスタリングの概念を導入し、以下の目的関数の最小化を考える。

$$J_{ranfis} = \sum_{i=1}^n \left((1 - u_i)^\theta \gamma + u_i^\theta \sum_{c=1}^C (y_{ic} - \hat{y}_{ic})^2 \right) \quad (2)$$

提案法におけるパラメータ最適化は Layer 5 の回帰係数 \mathbf{r}_c の更新と勾配降下法による Layer 1 の前件部のファジィメンバシップ関数を定めるパラメータ ϕ_{kj} の更新、および個体ごとの非ノイズ度 u_i の更新を繰り返す。

$$\mathbf{r}_c = (W^\top U W)^{-1} W^\top U \hat{\mathbf{y}}_c \quad (3)$$

$$\phi_{kj} \leftarrow \phi_{kj} - \tau \frac{\partial d_i}{\partial \phi_{kj}} \quad (4)$$

$$d_i = \sum_{c=1}^C (y_{ic} - \hat{y}_{ic})^2 \quad (5)$$

$$u_i = \left(1 + \left(\frac{d_i}{\gamma} \right)^{\frac{1}{\theta-1}} \right)^{-1} \quad (6)$$

$\hat{\mathbf{y}}_c$ は、 n 次元ベクトル $\hat{\mathbf{y}}_c = (y_{1c}, \dots, y_{nc})^\top$ である。 W は、 $n \times K$ 行列 $W = \{\bar{w}_{ik}\}$ であり、 U は、 i 番目の対角要素が u_i^θ の $n \times n$ 対角行列である。

3. 数値実験

3 種類のクラス ($C = 3$) からなる 4 次元 ($m = 4$) の特徴量をもつ Iris データの中を、各クラス 40 個体ずつをランダムに抽出した 120 個体 ($n = 120$) からなる訓練データと残りの 30 個体からなるテストデータに分割した。また、提案法におけるノイズロバスト性は、以下のように調べた。訓練データの一部のクラスラベルをランダムに 10% (12 ノイズ個体)、20% (24 ノイズ個体)、30% (36 ノイズ個体)、40% (48 ノイズ個体)、50% (60 ノイズ個体) の 5 種類のノイズ比率を用いて変更し、そのノイズを含む訓練データを用いて ANFIS による分類器を構築した。ファジィ度と学習率、ノイズ感度は、それぞれ $\theta = 2$, $\tau = 0.00001$, $\gamma = 1$ と設定した。

ノイズを含まない元の訓練データとノイズを含む訓練データに対して、従来法の ANFIS 分類器と提案法の ANFIS 分類器を構築した。その分類性能を、訓練データでは表 1、テストデータでは表 2 に示す。

表 1: 訓練データに対する分類性能の比較

ノイズ比率	original	10%	20%	30%	40%	50%
従来法	96.7%	86.7%	76.7%	70.8%	58.3%	48.3%
提案法	97.5%	86.7%	76.7%	67.5%	59.2%	51.7%

表 2: テストデータに対する分類性能の比較

ノイズ比率	original	10%	20%	30%	40%	50%
従来法	100.0%	96.7%	96.7%	90.0%	86.7%	76.7%
提案法	100.0%	100.0%	96.7%	93.3%	90.0%	86.7%

訓練データに対しては、ノイズ比率を考慮すると、分類性能が従来法と提案法どちらのモデルにおいてもほぼ完璧な分類性能 (100% - ノイズ比率) を達成している。一方、テストデータに対しては、提案法においてテストデータの汎化性能が従来法より向上していることがわかる。

4. おわりに

本論文では、ANFIS による分類器のノイズを含むクラスラベルに対するロバスト化を目的として、ノイズファジィクラスタリングの概念を導入する改良を提案し、数値実験により提案法の特性を示した。今後は、クラス比が不均衡であったり、クラス境界の重複があるようなデータに対するノイズ感度を調査していく。

参考文献

- [1] Jang, J.-S. R.: ANFIS: Adaptive-network-based fuzzy inference system, IEEE Trans. on Systems, Man and Cybernetics, 23-3, 665/685 (1993)
- [2] Honda, K., Notsu, A., Ichihashi, H.: Fuzzy PCA-guided robust k -means clustering, IEEE Trans. Fuzzy Systems, 18-1, 67/79 (2010)