

自己組織化マップによる判別問題の解法

1 はじめに

多変量解析の重要なテーマである判別問題は、医療診断や植物の分類など現実的な応用問題が多いため、積極的に研究がなされてきた。本研究では、項目毎にクラスタリングを行いクラスタ毎に異なった値を割り振り（正規化）、そのデータを用いて自己組織化マップを適用するという手法を提案する。

2 自己組織化マップとは

自己組織化マップ (Self-Organizing Maps:SOM)[1] は、大脳皮質の視覚野をモデル化したニューラルネットワークであり、教師無し学習手法の一つである。

SOM は入力層、出力層の二つの層で構成される。入力層は入力ベクトルの数に対応するニューロンを持ち、出力層はニューロンが格子状に配置される。

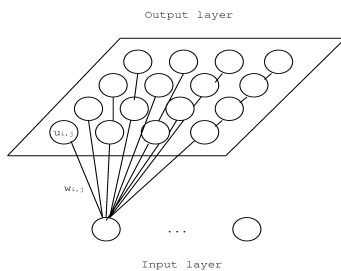


図 1: SOM の構造

結合荷重ベクトルと入力ベクトルとの距離を求める。そして、その距離がもっとも最小とする結合荷重ベクトルをもつニューロンを求め、勝者とする。それから、その勝者となったニューロンと、その近傍（ニューロン間の距離において）にあるニューロンの結合荷重ベクトルを入力ベクトルに近付けるように値を変化させる。これを学習とみなし、これを繰り返すことで最終的な出力である特徴マップを生成する。

3 クラスタリングによる正規化

項目毎にデータから数値を取り出し、正規化用として SOM でクラスタリングを行う。そして、できたクラスタを結合荷重の大きさ順に値を割り振り、元の値と置き換えることで正規化を行う。但し、正規化用 SOM の結合荷重ベクトルについては乱数ではなく、ベクトルの要素 $x_{0j}, x_{1j}, \dots, x_{(N-1)j}$ をソートしたものから、サンプリングし初期値として与える。こうすることにより、収束が早くなり、総学習回数を少なくすることができる。

4 実験

提案手法の有効性を確認するために、アヤメの分類問題と肝臓病診断問題、ワインの分類問題を行った。

アヤメの分類問題とは、花卉とがくの幅と長さから、3種類のをやめに分類する問題である。肝臓病診断問題とは、血液検査のデータから、健康体、急性肝炎、慢性肝炎、肝硬変、肝臓癌という 5 つの病状を判別する問題である。ワインの分類問題とは 13 種類のワインの成分が与えられ、ワインをある 3 銘柄に判別する問題である。実験の結果は表 1,2,3 のようになった。

表 1: アヤメの分類問題の判別結果

	問題数	認識数	誤認識数
setosa	50	50	0
versicolor	50	49	1
virginica	50	49	1

表 2: 肝臓病診断問題の判別結果

	問題数	認識数	判別率 (%)
健康体	115	115	100
急性肝炎	41	32	78.0
慢性肝炎	32	21	65.6
肝硬変	65	52	84.6
肝臓癌	69	58	85.5
全体	322	282	87.6

表 3: ワインの分類問題の判別結果

	問題数	認識数	判別率 (%)
銘柄 1	59	59	100
銘柄 2	71	66	93.0
銘柄 3	48	48	100

アヤメの分類問題では誤認識数 2 であり、優秀な結果であると考えられる。また肝臓病診断問題では、全体で 87.3 % と他の手法より格段に優れた結果が得られた。さらにワインの分類問題では、実値を扱う SOM と比べ、非常に優れた結果が得られた。

5 まとめ

本論文では、SOM によるクラスタリングを用いた正規化法によって特徴抽出を行い、そのデータを SOM に用い判別問題へと適用するという手法を提案した。検証実験を行い、その有効性を示すことができた。

参考文献

- [1] T. コホネン：自己組織化マップ改訂版、シュプリンガー・フェアラーク東京 (2001)