

k 次伝播 SOM

5P-04

柳田 卓郎 三浦 孝夫
法政大学工学部電気電子工学科

塩谷 勇
産能大学経営情報学部

概要

本稿では、自己組織化マップ（以下 SOM）の学習の拡張について述べる。従来の SOM のネックであるパラメータの依存度を軽減するため、クラス情報を学習の際に伝播させることを提案する。

1 動機と解法

SOM を使ったデータの分類において困難なものは、パラメータ値の設定である。特別な決まりごとが無いため、その決定に関するほとんどが勘や先人達の経験則に基づいている。上手く値を決めないと結果に大きな影響が出るにも関わらず、このような曖昧な定義しかなされていないものに依存しすぎていると、研究の精密さを欠くことになる。そこで本研究では少しでもパラメータに対する依存度を軽減し、学習方法をより明確にするために従来の SOM の試験データの中にクラス情報を追加し、学習率係数を学習近傍の外側に試験者が任意に決めた k 回だけ伝播する手段を実現する。k 次伝播を取り入れることで、これまで不明瞭だった「学習近傍の近傍」との関係の明かにできる。入力データに与えられるクラス数は既知のものであり、データのラベルインデックスとともにクラス情報が格納される。クラスが伝播してくるたびに配列内の対応した部分がインクリメントされて、比例配分を行うものとする。

2 SOM の拡張

SOM とは多次元の入力データを二次元平面上に可視化するもので、競合学習により自己組織化を行う。今回の研究では SOM を分類の手段として利用するため SOM によって振り分けられたマップを、学習ベクトル量子化（以下 LVQ）を併用して一致率をはかることでその評価を行う。従来の SOM における問題点は、ユニットの学習が与えられた学習近傍の内側だけで終わってしまうことである。今の状態では勝者ベクトルの最近傍ユニットは影響を受けるが、その「近傍の近傍」にかかる影響が考慮されていない。この問題を解決するために学習近傍の外にも勝者の影響を伝播させると共に、SOM アルゴリズムにクラス情報を加えることで以下のような式になる。

$$mi(t+1) = mi(t) + \sum hcix - mi(t) + class[a, b, \dots, n]$$

k ji dempa SOM

Takuro Yanagida, Takao Miura
Hosei University, Dept. of Elec. and Elec. Eng.
Kajino-cho 3-7-2, Koganei, Tokyo, JAPAN
Isamu Shioya
Sanno University, Dept. of Management and Informatic
Kamikasuya 1563, Isehara, Kanagawa, JAPAN

3 クラスの伝播

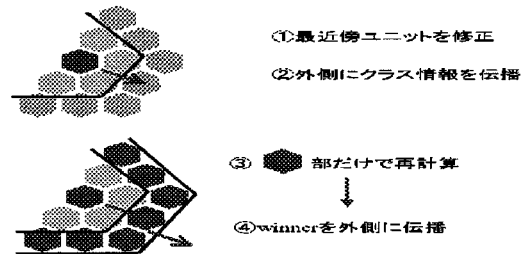


図 1: クラスの伝播

クラスの伝播の流れは図 1 のようになる。まず最初に勝者ユニットが決定し、その最近傍ユニットのベクトル修正が行われる。さらに外側のユニットに勝者のクラス情報とパラメータが伝播され、周囲にも影響が及ぶ。これを k 回繰り返して学習を進めていく。クラス情報は配列に格納し比例配分して表現するものとして、最終的に最も値の大きいものが、そのユニットのクラスとなる。

ex class[a,b,c] の形で表す winner_class=A → class[1,0,0] 今、影響を受ける最近傍ユニットのひとつがクラス B の情報を持っていたとすると、
class[1,1,0] → (比例配分) → class[0.5,0.5,0]
比例配分は以下のように表すものとする。[a,b,c] → [a/(a+b+c), b/(a+b+c), c/(a+b+c)]

4 実験

4.1 準備

本研究では動物の分類データ 17 次元 2000 件（クラスは 6 つ）と、波形パターンデータ 21 次元 5000 件（クラスは 3 つ）の 2 つのデータを用意する。どちらも欠損値はなく、動物データは 0 ~ 1 の間の値をとりパラメータはそれぞれ「翼の有無」や、「胎生」など動物の特性を現すものである。また、波形パターンデータは実数値をとり、負の値の使用も許されている。この 2 つを用意した理由は動物データは標準化されていて、パターンマッチングに近い分類手法であるのに対して、波形パターンデータは実数値であることから、違った手法の分類にどこまで対応しているのかをみる事が出来ると考えたからである。

4.2 手段

実験は伝播数を 0 から 100 まで間隔をあけてとっていくものとし、学習の回数もその都度変えていく。出力された SOM マップとマップ生成の際に得たコードブックベクトルを使用し、検定を行う。検定は LVQ を併用した一致率により行うものとし、グラフ化して評価を行う。評価の際に作成するグラフは、縦軸が一致率、横軸が伝播回数として、伝播回数と一致率の関係について調べることにする。k 次伝播と SOM は共にほぼ同じ条件で実験をする。

4.3 結果

SOM マップの出力結果を図 2 に示し、一致率のグラフは図 3 に示す。図 3 のグラフより、伝播回数が増えるほど一致率が下がっている事がわかる。同様の実験を SOM を用いて行った際の結果と比較してみると waveform については伝播回数、実行時間、一致率ともに SOM に劣っているといえるが、zoo の伝播回数の最初から 10 回までの間に SOM 以上の一致率を見せている部分がある。

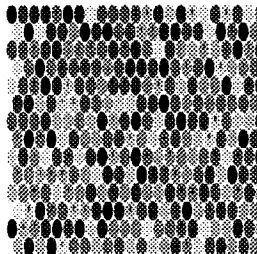


図 2: 出力結果

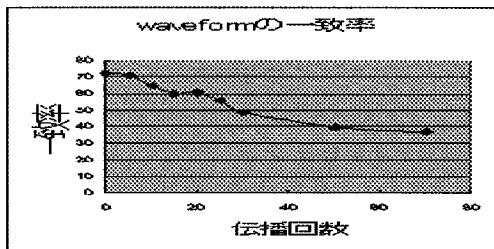


図 3: 一致率のグラフ

5 評価

これらの結果から過度の伝播は実験に良い影響を与えないと考えることが出来る。また、従来の SOM に比べてマップが乱れているように見えるが、k 次伝播の場合ユニットのクラス情報はアナログ的に考えることが出来、A と B の差が微

妙であったとしても最終的に値の大きかったほうを勝者として扱ってしまうことからこのように考えると考える。

6 結び

今回の拡張によってパラメータに大きく依存した分類からクラス情報によるものに近づけた。SOM に比べると実行時間が長くなりすぎる上に伝播回数によるぶれが新たに発生したともいえるので、これらについての問題がのこっている。

参考文献

- [1] T コホネン: 自己組織化マップ, シュプリンガーフェアークラーク東京 (1996)