平成２２年度特別研究報告書要旨

SOM 特徴マップ初期値の改善による可視化安定性の向上

情報メディア学科
T070470　百井 伸次
指導教員　三好 力 教授

1.　はじめに

自己組織化マップによって学習を行う際、既存技術の多くはその特徴マップが持つ各ベクトルの初期値を乱数で決定している。しかし、それでは同じ入力データに対して学習のたびに違う結果が表れてしまい、結果として安定性に欠けているといえる。それ以外の初期化手法として学習データのベクトルのユークリッド距離を計算し、それを基に特徴マップのベクトルを並び替える方法[1]が提案されている。しかしこの手法の場合全てのベクトルの距離の比較をしなければならないので、計算量が　$O(n!)$　と学習データの分析に多くの手順を踏むことになり、自己組織化マップの持つデータの自動学習という特徴が薄れてしまう問題がある。また特徴マップに一定の法則性を持たせることにより安定化に繋がる可能性も考えられるが、学習の高速化が目的である手法なので出力結果の安定性は重視されていない。そこで本研究では出力結果の安定性の向上を目的とし、学習データの分析を出来るだけ少なくした特徴マップの初期化手法を提案する。

2.　提案手法

以下のような初期化手法を提案する。まずｎ次元の学習データ内の値を無作為抽出し、その中から最小値・最大値を取得する。次にその範囲で乱数を生成し、ｎ次元ベクトルを作成する。そのベクトルをソートしたものを特徴マップの初期値として与える。

この手法では全ての学習データを参照するのではなく無作為抽出した少数のデータのみで最小値・最大値を決定することによって、学習データの分析を少なく抑えている。さらに単に乱数で初期化するのではなく、ソートしたベクトルを初期値とすることで、特徴マップに傾きを持たせている。これにより同じ入力データに対して同じような出力結果になり、安定性が向上すると考えられる。

3.　実験

特徴マップの初期値が乱数決定である既存手法と提案手法のプログラムを作成し、同一の学習データに対して繰り返し学習実験を行った。実験パラメータは、学習データが 3 次元×200 個、特徴マップが 30×30 ノード、学習回数が 4000 回、近傍領域が 20 で学習 50 回毎に 1 減少、学習率係数が 0.01×近傍領域　という値を設定した。この条件で表れた学習結果の図を比較し、同入力データに対する安定性の検証を行った。

4.　結果と考察

図１に既存手法と提案手法の実験結果を示す。左が既存手法、右が提案手法の自己組織化マップの特徴マップである。

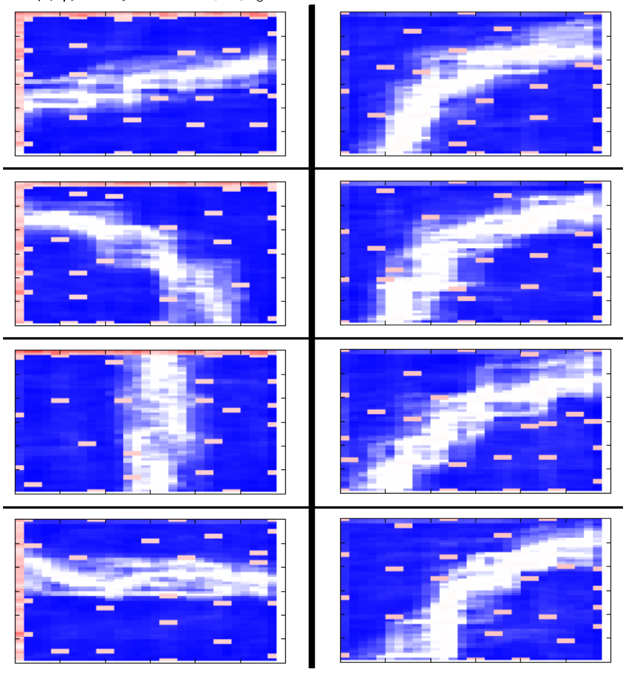

図１　既存手法と提案手法の出力結果

既存手法では同一の学習データにもかかわらず毎回違う特徴マップが生成されており、それにより同一の入力データに対しても毎回違う位置に自己組織化されていることがわかる。一方提案手法では同一の学習データで毎回ほぼ同一の特徴マップが生成され、入力データも毎回同じような位置に自己組織化されていることがわかる。また、提案手法の初期化の計算量は　$O(nlogn)$　となり、既存初期化手法の計算量　$O(n!)$　と比較して大幅に削減できていることがわかる。以上から、今提案手法は学習データの分析を抑えつつ出力結果の安定性が向上した初期化手法であると考えられる。

参考文献

[1]Mu-Chun Su, Hsiao-Te Chang
″Fast Self-Organizing Feature Map Algorithm″
(IEEE Transactions on Neural Networks,
11(3):721--733 , MAY 2000)
[2] 徳高平蔵, 大北正昭, 藤村喜久郎
“自己組織化マップとその応用”(2007)