

SOM 特徴マップの初期値とクラスタリングの明確化に関する研究

T11M090 百井 伸次

指導教員 三好 力 教授

1. はじめに

データマイニングの一手法として、自己組織化マップ（以下 SOM）という技術がある。SOM で学習を行う際、学習結果に大きく影響を与える要素として特徴マップの初期値がある。特徴マップの初期化方法はこれまで幾つか提案されているが、反復学習において高次元データに対する学習結果の視覚的安定性を得られる手法はない。そこで本研究では、高次元データにも対応する手法を提案する。また、別の問題点として、データ空間内でのデータのグループ化が弱く SOM の学習結果において視覚的分類が困難な場合がある。そこで本研究では、そのようなデータに対しクラスタリングを行い、さらなる視認性の向上を図る。これらの研究により、どのようなデータに対しても視覚的安定性を持った学習結果を得る手法の開発を目指した。

2. 提案手法の概要

2.1 学習データ及び特徴マップへの傾向付与

学習過程を規則化するため、学習データと特徴マップに傾向を与える。そのために、学習データを各次元値を基に辞書的にソートし、学習データの抽出を昇順に行う。また特徴マップも各次元値を基に辞書的にソートすることで、より強い傾向を与えることが出来る。

2.2 Ward 法と郡内平方和によるクラスタ分割

視覚的分類が困難な学習結果に対し、Ward 法を用いて階層化クラスタリングを行う。またクラスタリング結果に客観性を持たせるために、生成されたデンドログラムの切断点の指標を与える。各クラスタ形成時に採択された郡内平方和の増加量最小のものの中で、増加の幅が大きい階層を検索し、その階層を切断点としてクラスタ数を決定する。

3. 各提案手法に対する実験結果および考察

まず 2.1 について、既存手法と学習結果で同一データに対し繰り返し学習実験を行った。4000 回の学習回数の中に勝利ノード選出回数の最も多いベクトルの座標を記録する。100 回繰り返した結果をグラフ化する。

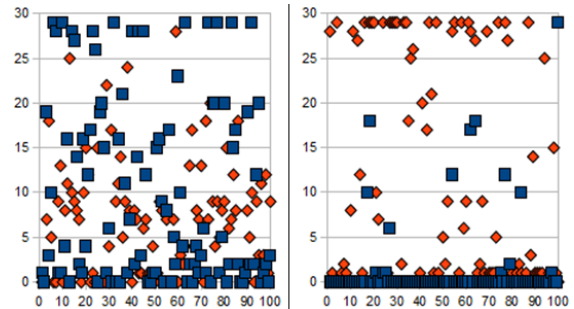


図 1: 既存手法(左)と提案手法(右)の実験結果

図 1 より、X 座標(青点)と Y 座標(赤点)はある程度 0 付近に固まっているものの、全体的に分散傾向にあることがわかる。一方提案手法では X, Y 座標ともほぼ同一箇所に固まっていることがわかる。これにより、反復学習における高次元データの学習結果の視覚的安定性は、既存手法と比べ提案手法の方が高いということが言える。

次に 2.2 について、視覚的分類が困難なデータを用いて分類実験を行う。データは X, Y それぞれ 0~25 の値を持つ 14 個のデータを 100 個用意し、クラスタ数の平均と標準偏差を求める。

表 1: クラスタ数の平均と標準偏差

平均	標準偏差
4.1	0.7

表 1 から、クラスタ数の平均はデータ 14 個に対し 4.1 と、視覚的分類がし易い数に収まったと言える。また標準偏差からクラスタ数のばらつきは少ないと言える。これにより、視覚的分類が困難なデータに対しても、安定して視覚的分類の補助になるようなクラスタ分割が出来ると言える。

4. おわりに

本研究では SOM の初期化手法を改善することで同一データの反復学習に対し視覚的安定性を向上させた。また SOM にクラスタリングを行い、視覚的分類が困難なデータに対する視認性を向上させた。これら二手法により、どのようなデータに対しても視覚的安定性を持った学習結果を得るという目的を達成した。