

類似度を考慮した近傍学習による適応型自己組織化マップ

日大生産工(院) ○中倉 昌哉 日大生産工 山内 ゆかり

1 まえがき

情報技術の急速な発達に伴い、データマイニングは重要な課題となっている。データマイニングにおいて、クラスタリングは代表的な分析方法の一つであり、様々な分野で研究されている。近年、自己組織化マップ (Self-Organizing Map: SOM) [1]を利用したクラスタリングに関する研究が注目されている。SOMは1982年にKohonen氏によって提案された教師なしニューラルネットワークであり、教師信号がなくても自分自身で内部構造を変化できることから脳の自己組織化過程を簡略化したモデルであるとも言われている。また、SOMはニューロン間のつながりである位相的順位を保持できるという利点を持ち、類似性のあるデータを分類するのに適している。

本研究では、SOMでの近傍学習において類似度を考慮することにより、より効率的な自己組織化を行うことを目指す。

2 Lazy Self-Organizing Map

SOMの一種として、怠けニューロンを含んだ自己組織化マップ (Lazy Self-Organizing Map: LSOM) [2]が存在する。LSOMの大きな特徴は、3種類のニューロン (働きニューロン、怠けニューロン、優柔不断なニューロン) を持つ点である。それぞれに学習率係数を用いることにより個性を持たせている。

LSOMは入力層と競合層の2層構造であり、競合層ではM個のニューロンが2次元格子状に配列されている。それぞれのニューロンは重みベクトル $w_i=(w_{i1}, w_{i2}, \dots, w_{id},)$ ($i=1,2,\dots,M$) を持ち、初期値は0から1の範囲でランダムに与えられる。さらに、ランダムに選ばれた $(p \times M)$ 個のニューロンは、怠けニューロンとしての集合 S_{lazy} に分類される。

従来のSOMと同様に勝者ニューロン c を決定するとき、 c が S_{lazy} に含まれている場合、 c は S_{lazy} から取り除かれる。このとき、入力データ x_j から最も遠く、 S_{lazy} に存在しないニューロン f が S_{lazy} に入るように選ばれる。

つまり、勝者ニューロン c になった怠けニューロンは働きニューロンになり、もう一方のニューロン f が怠けニューロンとなる。図1に様子を示す。

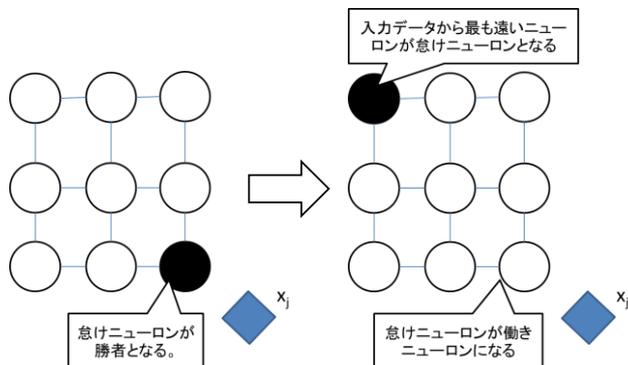


図1 怠けニューロンが勝者ニューロンとなったときの働きニューロンとの交代方法

その後は、ニューロンの性格と怠けニューロンの割合を考慮して、ニューロンの重みを更新していく。

3 提案手法および実験方法

LSOMでは、マップ上の距離を用いて、怠けものニューロンを判断している。SOMでの学習の後半において、参照ベクトルのマップ上の距離ではなく、類似度が離れているものを学習対象から除外する方が、データの局所構造を維持した学習が行えると考える。よって近傍内であったとしても類似度が低い場合は学習を行わない、適応型近傍学習を提案する。

本実験では、図2に示すような2次元の入力データを用いてシミュレーションを行う。入力データ数は1000個、2つのクラスタの他に4つの離れた入力データを持ち、全てのデータはランダムに選択される。

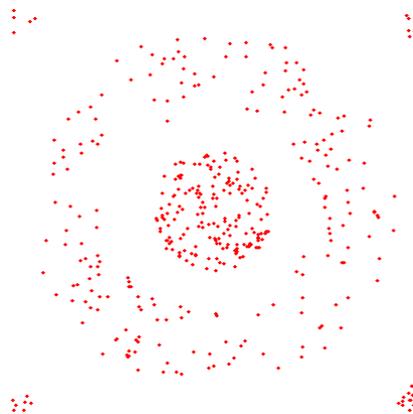


図2 入力データ

本実験の学習パラメータは、ノード数 $N=256$ 、学習回数20000回とする。

学習回数が15000回を超えた場合、類似度を考慮して学習することとする。

学習能力を数値的な方法で比較するために以下の利用頻度の高い3種類の測定値を用いる。

3.1 量子化誤差 $Q_e[1]$

量子化誤差 Q_e はそれぞれの入力ベクトルとその勝者との距離の平均を計算した値である。

$$Q_e = \frac{1}{N} \sum_{j=1}^N \|x_j - w_j\| \quad (1)$$

w_j は入力データ x_j に対する勝者ニューロンの重みベクトルである。従って、 Q_e は0に近いほど入力状態に近いことを示す。

3.2 トポロジー誤差 $Te[3]$

トポロジー誤差 Te はSOMがどのくらい入力データのトポロジーを保存できているかを示す値である。

$$Te = \frac{1}{N} \sum_{j=1}^N u(x_j) \quad (2)$$

N は入力データの総数である。また、入力データ x_j に対する1番目の勝者と2番目の勝者が互いに1近傍以内なら $u(x_j)$ は0、それ以外なら1となる。つまり、 Te は0に近いほど入力データのトポロジーを保存できていることを示す。

3.3 ニューロン利用率 $U[4]$

ニューロン利用率 U は1度以上勝者になったニューロンの割合を示す値である。

$$U = \frac{1}{nm} \sum_{i=1}^{nm} u_i \quad (3)$$

もし、ニューロン i が1度以上勝者となったなら、 u_i は1となり、1度も勝者にならなかったなら0となる。つまり、 U は1に近いほどより多くのニューロンを有効利用できており、不活性ニューロンが少ないことを示す。

4 実験結果

表1に学習結果から計算された3種類の測定値をそれぞれ示す。

表1 入力データに対する3種類の測定値

	Q_e	Te	U
従来SOM	0.010603	0.172000	0.842419
提案手法	0.011286	0.126000	0.90625

実験結果から従来のSOMと比較して、 Te と U にお

いて良い結果を出すことができた。しかし、 Q_e では測定値の悪化が見られた。

図2に学習後のデータと競合層のユニットの配置を示す。

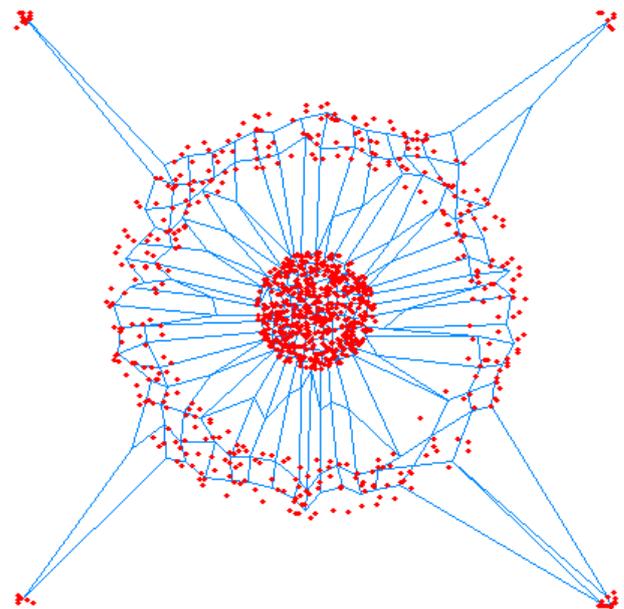


図2 学習結果

5 まとめ

本研究では、SOMでの近傍学習において類似度を考慮することにより、トポロジー誤差が低く、ユニット活用率の高い自己組織化を実現することができた。しかし量子化誤差は拡大するという結果となった。これらの指標は近傍学習の違いにより独立ではなく、全てにおいて良い値を得ることは難しいと考えられる。

今後の課題として各試行によるばらつきの提言や、パラメータの調整による最適値の解析等が挙げられる。

「参考文献」

- [1] T. Kohonen, *Self-Organizing Maps*, Berlin, Springer, vol. 30, 1995.
- [2] 原口卓 松下春奈 西尾芳文, 怠け者の割合を考慮したLazy Self-Organizing Mapとその振る舞い, 電子情報通信学会技術研究報告. NLP, 非線形問題 108(174), 5-10, 2008-07-24
- [3] K. Kiviluoto, "Topology Preservation in Self-Organizing Maps," Proc. of International Conference on Neural Networks, pp. 294-299, 1996.
- [4] Y. Cheung and L. Law, "Rival-Model Penalized Self-Organizing Map," *IEEE Trans. Neural Networks*, vol. 18, no. 1, pp. 289-295, 2007.