

動的に構成される競合層を持つ自己組織化マップ

日大生産工 ○千葉 憲太郎 日大生産工 山内 ゆかり

1 まえがき

人間の脳を模したものとしてニューラルネットワークというものがある。自己組織化マップ(SOM:Self-Organizing Map)とはニューラルネットワークの一種であり、教師なしでクラスタリングを行うことができる。人が理解しにくい高次元のデータを2次元や3次元に表すことで、理解を促すことができる。

SOMは競合層と入力層の二層構造になっている。競合層とは複数のベクトル値を持つユニットが規則的に並んでおり、それぞれのユニットが隣同士リンクで繋がっている。このリンクは本来のSOMの学習では不動である。

本研究では競合層を動的に構成させることによって、構成されるネットワークへの複雑性の適用と、SOMの学習性能の向上を図る手法を提案する。

2 SOM

SOMではデータ同士の距離をもとにそのユニットがどれ程学習するかを決めるが、今回はユニット間の最短経路長をユニット間の距離とする。学習については(1)の式を用いる。¹⁾

$$n(t) = n(t-1) + h(win, m, t)(xn(t-1) - n(t-1)) \quad (1)$$

n は競合層内のそれぞれのユニットの番号である。 h は近傍関数である。 xn 勝者ユニットである。近傍関数については(2)(3)(4)の式である。

$$h(win, m, t) = c \cdot \exp\left(-\frac{dis^2}{\alpha^2}\right) \quad (2)$$

C は α と dis の範囲を調整するためのパラメータである。

$$dis = L_{win, n} \quad (3)$$

L は勝者ユニットと学習するユニットの最短経路長である。

$$\alpha = 1 - \frac{t}{T} \quad (4)$$

以上の式をすべてのユニットに適応させる。

4 提案手法

本研究では競合層内のリンクを張り替える。張り替え方法については図(1)(2)に示す。

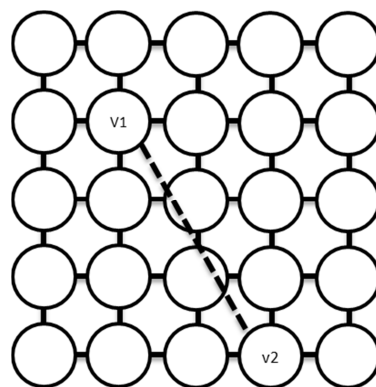


図1. リンク張り替え図1

リンクの張り替えは入力データの中からランダムに1つデータを選ぶ。それを入力した際の第1勝者ユニットと第2勝者ユニットの距離が L の半数以上離れていれば結合する。図1だと、 $V1$ が第1勝者ユニットであり、 $V2$ が第2勝者ユニットであるので、そのユニット同士を結合する。

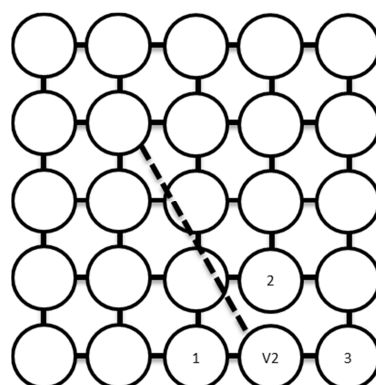


図2. リンク張り替え図2

第2勝者ユニットとそのユニットに繋がっているユニットで、最も類似度が低いユニット間のリ

リンクを切断する。図2だと、第2勝者ユニットであるV2に1,2,3のユニットが繋がっている。その中で2のユニットが最も類似度が低いとしたら、V2と2のリンクを切断する。

SOMでは似ているユニット同士の距離を近づけていくものなので、学習の過程で似ている者同士を繋げてしまうとどのような影響があるかを分析する。

5 実験及び評価方法

今回、実験に使用するSOMは225個のユニットを有し、3次元データを扱うものとする。データはそれぞれ(赤)G(緑)B(青)の値となり、この3つのデータで色を表現する。

入力データは0~255の乱数を3個、1000セット用意する。学習する際に1000セットのデータの中からランダムに選択し入力する。

ネットワークの特徴を示す指標としてのL(平均頂点間距離)を用いる。Lとはあるノードから別のノードへ行くときに通過するリンクの平均値である。

以上のSOMを用いて、学習を5000回行う。リンクの張り替えは1000回までの間に、100回ごと行う。したがって、残りの4000回は通常の学習を行う。学習後、1000セットのデータをすべて入力し、出力されたデータをもとに評価を行う。

SOMの評価として、 Q_e (量子化誤差)²⁾、 T_e (トポロジー誤差)³⁾、 U (ニューロン利用率)⁴⁾を用いる。 Q_e と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。

$$Q_e = \frac{1}{N} \sum_{j=1}^N \|x_j - \bar{w}_j\| \quad (5)$$

$\bar{w}_j$ 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に近いほどデータ全体の特徴を反映していることを示す。

T_e とは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。

$$T_e = \frac{1}{N} \sum_{j=1}^N u(x_j) \quad (6)$$

N は入力データの総数である。また、入力データ x_j に対する第1勝者と第2勝者の勝者が互いに1近傍以内なら $u(x_j)$ は0、それ以外なら $u(x_j)$ は1となる。つまり、 T_e 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。

U とは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。

$$U = \frac{1}{nm} \sum_{i=1}^{nm} u_i \quad (7)$$

もし、ニューロン i が1度以上勝者になったら $u_i = 1$ 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。

6 実験結果

実験を従来のSOMと提案SOMそれぞれに10回ずつ行い、その平均値を表1に示す。

表1. 評価表

	Q_e	T_e	U	L
従来SOM	0.278	0.222	0.95	10.025
提案SOM	0.362	0.152	0.95	7.227

従来SOMに比べ提案SOMの T_e と L が減少し、 Q_e が増加、 U が変わらないという結果となった。 T_e の減少はリンク張り替えの操作が第1勝者と第2勝者を結合することによって、近傍が類似関係を表現しているからだと思われる。

7 まとめ

T_e が減少したが、 Q_e が増加してしまった。 Q_e を減少させるために、他のリンクの追加方法を考えたい。

今回は入力データが乱数であったので、このような結果となった。そこで、他の振る舞いを見るために特徴のあるデータを用いた実験も行いたい。

「参考文献」

- 1) “自己組織化マップ”
<http://www.sist.ac.jp/~kanakubo/research/neuro/selforganizingmap.html>
(最終アクセス：2012/10/19)
- 2) T. Kohonen, Self-Organizing Maps, Springer, vol. 30, 1995.
- 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.