

進化型自己組織化マップにおける ノード領域の自動調整法の検討

A Study on Automatic Adjustment of Node Area in Evolving Self-organization Map

坂根 公和 串田 淳一 原 章 高濱 徹行
Kimikazu Sakane Junichi Kushida Akira Hara Tetsuyuki Takahama
広島市立大学大学院情報科学研究科知能工学専攻
Email:ksakane@ints.info.hiroshima-cu.ac.jp,
Email:{kushida,ahara,takahama}@hiroshima-cu.ac.jp

Abstract—In this study, the method of classifying data or extracting the regularity of the data is proposed in order to acquire useful information from data. We pay our attention to clustering and visualization. We improve Evolving Self-organizing maps (ESOM) which is one of the clustering techniques proposed by Kasabov. Since the node area is always constant in ESOM, there is a problem that a parameter for node area must be determined for each data set. Therefore we propose a method to automatically adjust the parameter from the quantization error of the node to solve this problem.

I. はじめに

データから有益な情報を得るために、データを分類したり、データから規則性を抽出する方法が提案されている。本研究ではこのような方法のうち、クラスタリングと可視化に着目した。クラスタリングは教師なしでデータを自動的に分類する方法であり、可視化は分類の様子を視覚的に表示することによって人間を支援する方法である。

本研究ではクラスタリング手法の1つである Kasabov の提案した進化型自己組織化マップ (Evolving Self-Organizing Maps :ESOM)[3][4] のノード領域の問題点を解決するためにノード領域の自動調整の研究を行った。

II. 進化型自己組織化マップ

A. ESOM の概要

ESOM は SOM (Self-Organizing map)[1] や GNG(Growing Neural Gas)[2] と同様に教師なし学習を行う。ESOM は初め何もない状態から学習を始めるため、ネットワークの大きさを決める必要がない。また、SOM や GNG よりも短い学習時間で学習を終えることが可能であり、精度も高い。ESOM は各ノードから一定の領域に入力データが入力されると、ノードとその接続先が入力データに近づくように学習

を行う。逆に全てのノードの領域の外に入力データが入力されるとその入力データが新しいノードになる。

しかし、従来の ESOM ではこのノード領域を決定するパラメータが常に一定であるため、入力データのクラスタ毎で広がり異なる入力データに対してはパラメータを決定することが困難であるという問題がある。

B. ESOM のアルゴリズム

初めネットワークには何もない状態から始め、基本的に入力データが既存のノードに一致していないときに新しいノードを増やしていく。最新の入力データを x 、既存のノードのセットを $W = \{w_1, w_2, \dots, w_N\}$ とする。

1. 以下の条件を判定する。

$$\|w_i - x\| \leq \epsilon, \exists w_i \in W, \quad (1)$$

ここで ϵ はノード領域の半径を表すパラメータである。

この条件式は入力データが各ノードの領域に含まれるかを判別する条件式である。

- 2a. 式 (1) を満たすとき
- 3a. 全てのノードの活性値 a を以下の式で更新する。

$$a_i = e^{-2\|x - w_i\|^2/\epsilon^2} \quad (2)$$

a_i は w_i への近接の度合いを表している。

- 4a. 次にノードの重みベクトルと接続強度を更新する。この更新では入力データに一番近いノードとそのノードに接続したノードを更新する。ノードの重みベクトルの更新は以下の式で行う。

$$\Delta w_i = \gamma h_{i,b}(x)(x - w_i), i \in \Omega(b) \quad (3)$$

$$\Omega(i) = \{j | s(i, j) > 0\}, \forall j \in [1, N], i \neq j \quad (4)$$

b は入力データに一番近いノードで、 γ は学習率を表している。 b に接続したノード全て

に対して更新を行うのではなく、式 (4) を満たすノードのみを更新する。ただし、 $s(i, j)$ はノード i と j の接続強度を表している。また、 $h_{i,b}(x)$ はノード i とノード b の近傍関数を表しており、ESOM では、近傍関数は以下の式となる。

$$h_{i,b}(x) = a_i(x) / \sum_k a_k(x) \quad (5)$$

以上のことをまとめると重みベクトルの更新式は以下の式になる。

$$\Delta w_i = \begin{cases} \gamma \sum_k \frac{a_k}{a_i} (x - w_i) & \text{if } i \in \Omega(b) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

接続強度の更新を行う。接続強度の更新は以下の式で更新する。

$$s(i, j) = \beta s(i, j) + (1 - \beta) a_i a_j \quad (7)$$

β は忘却定数を表し、値は通常 0.8 を用いる。式 (7) を用いて、一番近いノードとそれに接続したノード全ての接続を更新していく。ここまでが (1) 式を満たすときの動作である。

- 2b. 式 (1) を満たさないとき
3b. 新しいノードを以下の式で追加する。

$$w_{N+1} = x \quad (8)$$

新しいノードを追加したあとに全てのノードの中で w_{N+1} に一番近いノード s_1 と二番目に近いノード s_2 を見つける。そして、新しいノードを s_1, s_2 それぞれに接続する。新しく接続する場合は接続強度を 1 とする。

- 4b. s_1 と s_2 が未接続であるかどうかを調べ、未接続であれば接続する。そうでなければ s_1 と s_2 との接続強度を式 (7) で更新する。ここまでが式 (1) を満たさないときの動作である。
5. これらを繰り返し、学習時間 T_P 毎に一番弱い接続強度の接続を削除する。そして、終了条件を満たすまで 1. に戻る。

III. 提案手法

従来の ESOM ではノード領域を決定するパラメータが常に一定であるため、入力データのクラスタ毎で広がり異なる入力データに対してはパラメータを決定することが困難である。この問題点を解決するために、ノード領域を決定するパラメータ ϵ を自動調整する方法を提案する。

ϵ は大きいとノード領域が広いと、ノードは更新されやすく追加されにくい。逆に小さいとノード領域が狭いため、ノードは更新されにくく追加されやすい。この性質を利用して自動調整を行う。

本論文では 1 度学習したノード全体から平均量子化誤差を算出する。それを基準として、ノード全体の平均量子化誤差よりも各ノードの平均量子化誤差が大き

いと ϵ を減少させ、ノード全体の平均量子化誤差よりも各ノードの平均量子化誤差が小さいと ϵ を増加させるという方法を提案する。これは以下の式でまとめられる。

$$\epsilon_i = \begin{cases} \epsilon_i - \alpha(Q_i - \bar{Q}) & (\|w_i - x_b\| \leq \epsilon_i - \alpha\Delta Q) \\ \|w_i - x_b\| & (\text{otherwise}) \end{cases}$$

Q_i は w_i に最も近いデータの集合とノード w_i との平均量子化誤差を表している。 $\Delta Q = Q_i - \bar{Q}$ である。 \bar{Q} はノード全体の平均量子化誤差を表している。 x_b は w_i に最も近いデータの集合の中で最も w_i に近いデータである。また、 α は学習係数である。

IV. 実験

従来の ESOM と提案手法の ESOM で、 ϵ の値を 0.3, 0.4, 0.5, 0.6, 0.7 と変化させ、10 回繰り返して学習する。そのときのノード数と平均量子化誤差 (MQE) を比較する。また、そのときの入力データとノードデータを可視化させた。なお、使用したデータは四次元データの *Iris* データである。なお、*Iris* データは 3 つのクラス *Iris-Setosa*、*Iris-Versicolor*、*Iris-Virginica* に分類されている。

また、 ϵ 以外のパラメータはそれぞれ、 $\alpha = 0.5, \beta = 0.2, \gamma = 0.05, T_P = 20$ とし、実験した。

ノード数と平均量子化誤差の結果を表 I に示す。 ϵ の値を 0.3, 0.4, 0.5 としたときのデータを可視化する。また、従来の ESOM の結果をそれぞれ図 1 から図 3 に示し、提案手法の ESOM の結果をそれぞれ図 4 から図 6 に示した。

表 I
従来の ESOM と提案手法の ESOM の比較

ϵ	従来の ESOM		提案手法の ESOM	
	ノード数	MQE	ノード数	MQE
0.3	100	0.045487	124	0.031466
0.4	80	0.086918	100	0.063969
0.5	54	0.170431	76	0.123127
0.6	44	0.220274	57	0.186956
0.7	32	0.251270	44	0.226892

表 I から、ノード数と平均量子化誤差の比較を行う。

従来の ESOM と提案手法の ESOM の両方共 ϵ が大きくなるに連れてノード数が減少し、平均量子化誤差が増加していることが分かる。 ϵ が同じ値の場合の例として、 ϵ が 0.5 の場合で比較する。従来の ESOM ではノード数は入力データの 3 割がノードになっている。それに対して、提案手法の ESOM ではノード数は入力データの半分がノードになっている。これは ϵ の値が減少して、ノードが追加されやすくなったことを意味している。また、ノードの増加により、平均量子化誤差が減少している事が分かる。

次に可視化したデータの比較を行う。 ϵ が 0.3, 0.4 のときは従来の ESOM と提案手法の ESOM の両方共大きな変化は見られない。しかし、 ϵ が 0.5 のときは従来の方

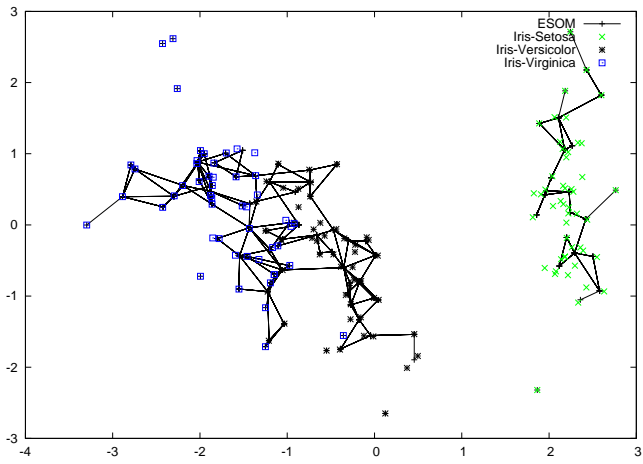


図 1. 従来の ESOM($\epsilon = 0.3$)

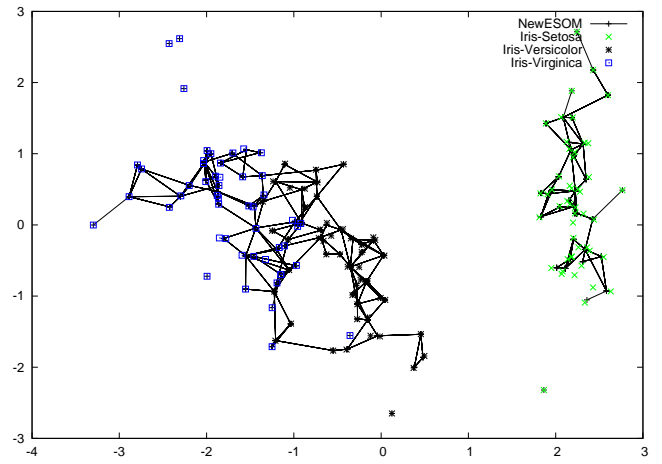


図 4. 提案手法の ESOM(初期 $\epsilon_i = 0.3$)

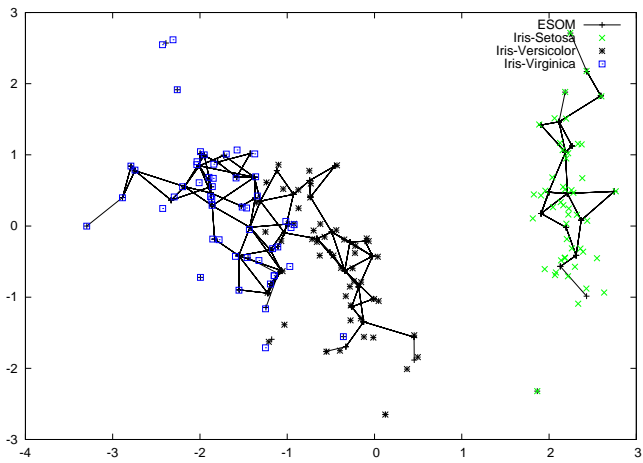


図 2. 従来の ESOM($\epsilon = 0.4$)

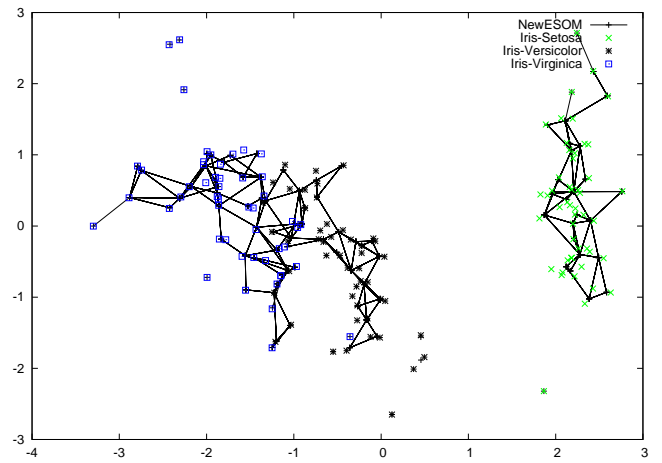


図 5. 提案手法の ESOM(初期 $\epsilon_i = 0.4$)

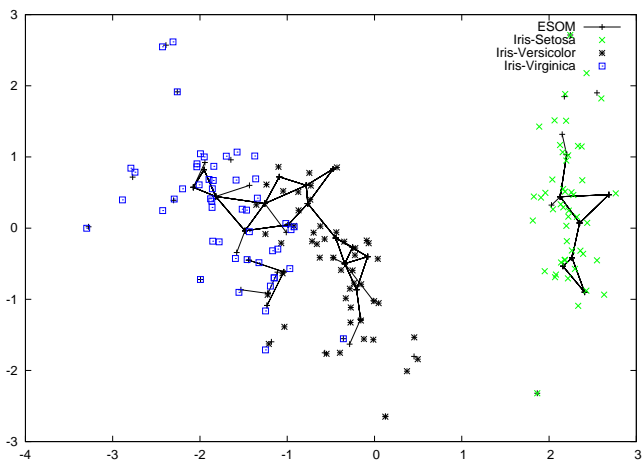


図 3. 従来の ESOM($\epsilon = 0.5$)

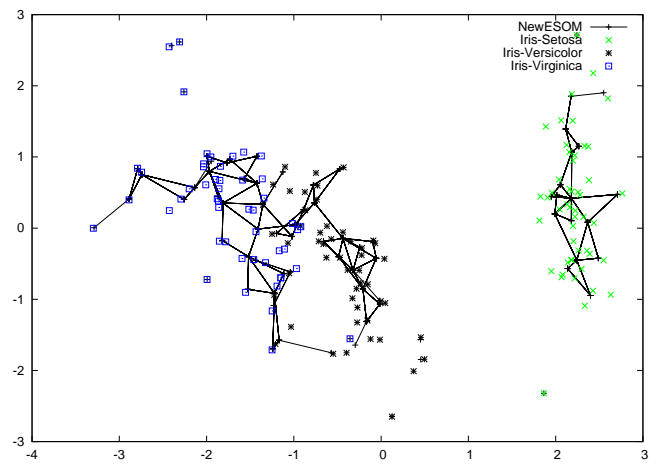


図 6. 提案手法の ESOM(初期 $\epsilon_i = 0.5$)

は接続が少ないが提案手法の方はまだ接続が残っていることが分かる。また、*Iris-Versicolor* と *Iris-Virginica* の2クラス間の接続が少ないことが分かる。このことから、*Iris-Versicolor* と *Iris-Virginica* のクラス分類しやすいと言える。

V. まとめ

ESOMではノード領域が常に一定であるため、入力データのクラス毎で広がり異なる入力データにはノード領域を決定するパラメータを適切に決定することが困難という問題点がある。本研究ではこれを解決するためにノード領域を自動調整する方法を研究した。本論文では平均量子化誤差を基準として、各ノードの平均量子化誤差の善し悪しから、ノード領域を自動調整する方法について提案した。

実験結果からデータの広がりに対応して、ノード数が増え量子化誤差を減らすことができたことが分かった。しかし、本論文で提案したノード領域の自動調整法はあらかじめ初期値を決めておかなければならず、ノードのほとんどは初期値に依存してしまうため、ESOMの問題を解決したとはいえない。今後の課題として、パラメータそれぞれの調整と入力データに対して、最も適している初期値を決定する理論が必要であると考えられる。

参考文献

- [1] 坂和 正敏, 田中 雅博, "ニューロコンピューティング入門", 1997
- [2] Bernd Fritzke, "A Growing Neural Gas Network Learns Topologies," *Advances in Neural Information Processing Systems* 7:pp.625-632, 1995
- [3] Da Deng, N. Kasabov, "On-line pattern analysis by evolving, self-organizing maps," *Neurocomputing* 51, pp.87-103, 2003
- [4] Da Deng, N. Kasabov, "An Algorithm to Evolve Self-Organizing Maps from On-line Data Stream," *Proceedings of IJCNN'2000*, Vol.VI, IEEE Press, NewYork, pp.3-8, 2003

問い合わせ先

〒731-3194

広島市安佐南区大塚東三丁目4番1号

広島市立大学大学院情報科学研究科知能工学専攻

知能システム研究室

坂根 公和