

索引層を用いた SOM の学習高速化: 初期マップ生成アルゴリズムの改良

Quick Learning Algorithm of Self-organizing Map using Index Layer: Rational Initial Mapping

渡邊 旬
Hitoshi Watanabe

高橋 由雅
Yoshimasa Takahashi

豊橋技術科学大学 工学部 知識情報工学系
Department of Knowledge-based information Engineering, Toyohashi University of Technology

In the preceding work we proposed a quick learning algorithm of Self-Organizing Map (SOM) using an Index Layer that is an additional competition layer. And it was shown that the approach successfully works to shorten the computational time to 1/10 approximately in the best case. In this work, aiming to further shorten the computational time for learning of the SOM we have investigated a new algorithm for rational initial mapping based on the FastSOM. Computational experiments suggest that the rational initial mapping allows us to considerably save the total learning time of the SOM.

1. はじめに

Kohonen が提案した SOM(Self-Organizing Map)[コホネン 96]は画像解析、音声認識など幅広い分野で用いられてきたデータ可視化手法の 1 つである。SOM は多次元空間でのデータ間の近接性を保持したまま可視空間上に非線形写像を行うことができるという特徴を持つ。従来法では 1 つの学習データが提示されると、全ての競合層上のユニットに対して、類似度計算を行う必要がある。よって、大規模なデータ集合に対してより詳細なデータ構造の可視化を行うためには、サンプル数や競合層のニューロン数の増大に伴い、多大な処理時間が必要となる。

当研究室の佐々木らはこうした大容量多変量データ空間の可視化への応用を念頭に置き、競合層上での BMU(Best Much Unit)の出現域を大まかに推定するための索引層の利用を基礎とした学習高速化アルゴリズムを提案した(図 1)[佐々木 04]。

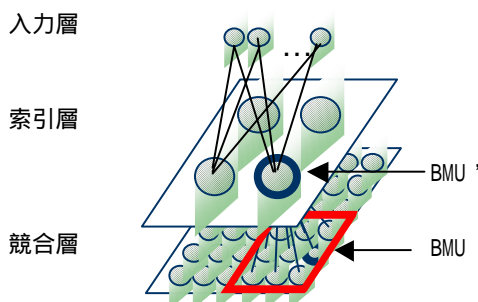


図 1 索引層を用いた SOM アーキテクチャ

本研究では SOM 学習のより一層の高速化を目的とし、合理的な初期マップ生成のための新たなアルゴリズムを提案すると共に、その性能について検討を試みた。

2. 初期マップの生成

学習を行う前に競合層マップに初期重みを与え、初期マップを生成する。初期マップの生成には乱数が用いられるのが一般

連絡先: 高橋 由雅, 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1 豊橋技術科学大学 知識情報工学系, Tel: 0532-44-6878, taka@mis.tutkie.tut.ac.jp

的である。今、学習の進行に伴う生成マップの変化の様子を示した図 2 の例について考える。例えば、初期マップを初めからある程度格子を形成している状態(学習が進んだ状態)、すなわち図 2 の a ではなく c から学習を始めることができれば学習回数を大きく減らすことが可能となる。ここでは、FastSOM[Su 00]の方法を参考に、初期マップ生成のための新たなアルゴリズムについて検討し、その性能を従来法と比較した。

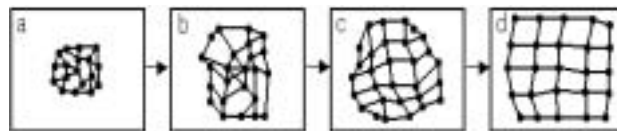


図 2: 学習の経過によるマップ生成の様子

2.1 FastSOM による初期マップ生成アルゴリズム

FastSOM は Mu-Chun Su らによって提案された手法で、以下手順で初期マップの生成を行っている。

入力データ内で距離が最も遠い 2 つのデータを競合層マップの対角端の重みとする。さらにその 2 つのデータから距離が最も遠いデータを競合層の残っている端の片方の重みとする。この 3 つのデータから距離が最も遠いデータを残りの端の重みとする。4 つの端の重みが求まったら、1 辺の両端の重みから最も遠いデータを対辺の重みとして代入していく。これを時計回りに全てのユニットに重みが割り振られるまで繰り返す。

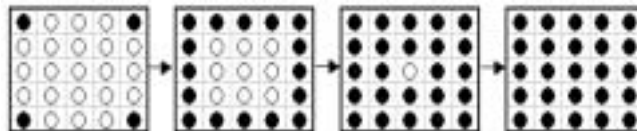


図 3: FastSOM による初期マップ生成アルゴリズム

2.2 本研究で提案する新たなアルゴリズム

以下に本研究で提案するアルゴリズムを示す。

まず、FastSOM と同様に 4 つの端の重みを求める。次に、全ての入力データを平均した値を競合層マップの中心の重みとする。そして、各端点の重みと中心の重みとをその間のユニット数で等間隔分割した値を、各ユニットの重みとする。

最後に、辺に沿って端の重みと重みの間のユニットに、その重みと重みとを等間隔に分割した値を重みとする。

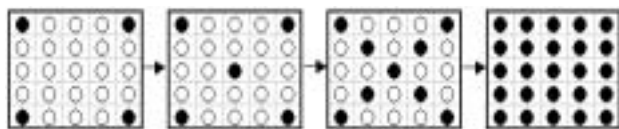


図4: 考案した新たな初期マップ生成アルゴリズム

3. 結果と考察

データセットには動物のデータ(データ数 101,次元数 16)を用いて各アルゴリズムによる初期マップ生成を行った。乱数、FastSOM、改良アルゴリズムによる初期マップの生成結果を学習済みマップと合わせて図5に示す。

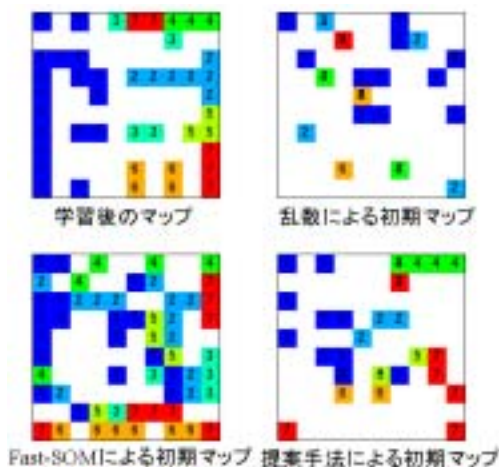


図5: 各アルゴリズムによる初期マップ生成結果

この結果から、乱数による初期マップと比較すると FastSOM、および本研究での改良アルゴリズムによる初期マップのほうが学習後のマップにより近い写像を与えていることが確認できる。

しかしながら、上記の結果は定量的なものではない。そこで、これらを定量的に表現する為に入力データと BMU(Best Much Unit)との誤差値を求めた。図6の結果では、乱数、FastSOM による方法では、収束しているのに対し、提案手法では、最初からなだらかに誤差値が減少していている。これは、すでに収束が行われた後であると考えられる。つまり、初期学習率、初期近傍といった初期パラメータが大きいということが考えられる。

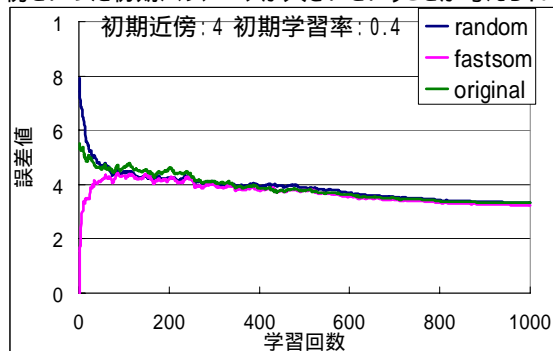


図6: 学習回数による誤差値変動

そこで、初期パラメータを小さくした結果を図7に示す。図7では、3つの方法とも収束までの過程が見てとれる。さらに、若干ではあるが乱数による方法よりも FastSOM、NewAlgorithm の方法のほうが収束するまでの学習回数が少ないことが分かる。これらの結果からも、誤差値の結果は初期パラメータに依存することが分かる。

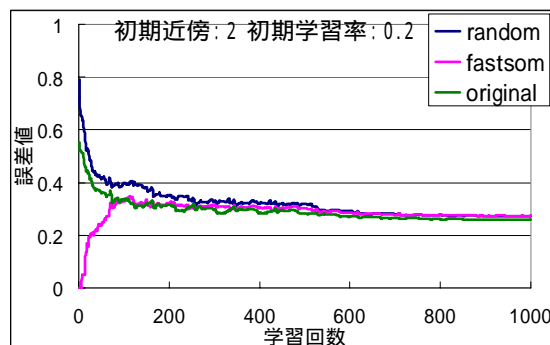


図7: 学習回数による誤差値変動

次に、初期近傍による誤差値の比較を行った。図8の結果より初期近傍が小さくなるにつれて誤差値も小さくなっていることが分かる。さらに、誤差値が飽和するまでの学習回数も近傍が小さくなるにつれて少なくなっていることが分かる。

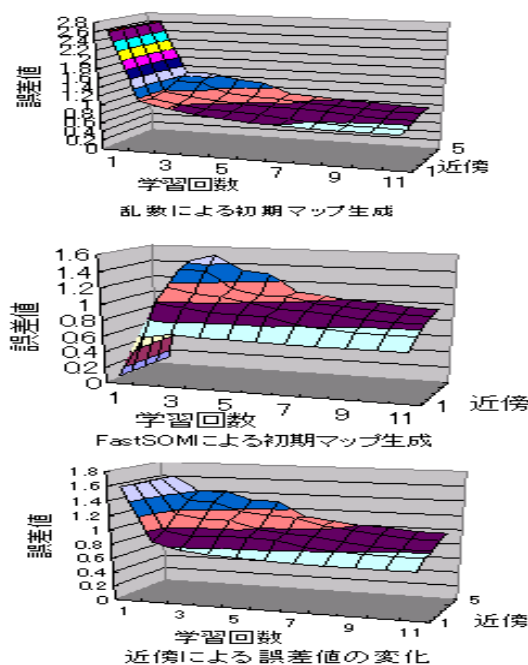


図8: 近傍の変化による誤差値変動

4. まとめ

以上の結果、乱数による初期マップ生成に比べ、FastSOM、および改良アルゴリズムによる初期マップ生成のほうがより学習後の最終マップに近い良好な結果を与えることを検証した。しかし、学習の高速化としては索引層付 SOM での利用に際しての工夫すべき点もあり、引き続き検討を進める必要がある。今後はこれらの点をふまえ、更なる SOM 学習の高速化を試みたい。

参考文献

[コホネン 96] Teuvo Kohonen, シュプリングァーフェアラーク東京, 1996, 479 p (ISBN4-431-70700-X)
 [Su 00] Mu-Chun Su, Hsiao-Te Chang, "Fast Self-Organizing Feature Map Algorithm", IEEE Transactions on Neural Networks, 11(3):721--733, MAY 2000.
 [佐々木 04] 佐々木英史, 高橋由雅, 索引層を用いた SOM の学習高速化アルゴリズム, 人工知能学会, 1F2-04