

索引層を用いた SOM の学習高速化アルゴリズム

Quick Learning Algorithm of Self-organizing Map using Index Layer

佐々木 英史
Hidemi Sasaki

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

Self-Organizing Map (SOM) is a powerful tool for data visualization. But searching for the Best Match Unit (BMU) on competition layer of the SOM is quite time consuming when the large number of neurons are set on the layer. In this paper we propose a quick learning algorithm using an Index Layer that is an additional competition layer. The algorithm was implemented and tested with a synthesized dataset and several real dataset. They were resulted that our approach successfully works to shorten the computational time to 1/10 approximately in the best case.

1. 序論

自己組織化地図 (Self-Organizing Map, SOM) は、実データ空間中の近接関係を保持した写像を行なう教師なし学習の1つであり、多次元データを可視空間に非線形写像することができる[デイホフ 92][徳高 99]。従来の SOM の基本学習アルゴリズムは以下の通りである。

- 学習データの提示
- 競合層上のユニット全てに対して類似度の計算を行い、類似度の一番大きいユニット (Best Match Unit, BMU) を探索
- BMU の近傍空間に位置するユニットの重みを学習データに近づける。

これらの操作を、指定した終了条件を満たすまで繰り返すことで学習が行なわれる。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、類似度計算を行なう必要があり、訓練時間の多くを BMU の探索が占める。これは、サンプル数や競合層のニューロン数の増大に伴い、より顕著なものとなる。そのため、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多な処理時間が必要となる。

本研究では、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 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するとともに、その性能について評価・検証を行った。

2. 索引層による学習高速化アルゴリズム

本手法では、入力層と競合層の間に、新たに索引層を設けた。その概要を図 1 に示す。索引層を用いた学習高速化アルゴリズムは以下の通りである。

- 学習データの提示
- 索引層上で仮の BMU (BMU') を探索
- 競合層上で、BMU' の下に位置する範囲 child のみを対象として、BMU を探索

- BMU の近傍空間に位置するユニットに対し、その重みを訓練データに近づけると共に、索引層のユニットの重みをその下に位置する近傍ユニットの重みに近づける。

改良法では競合層ユニット全てに対して類似度を求める必要がないことから、BMU の探索時間を短縮することができる。しかしながら、新たに導入した索引層での重み調節によるオーバーヘッドが生じることも明らかである。従って、その実効性能も計算機実験によって検証する必要がある。

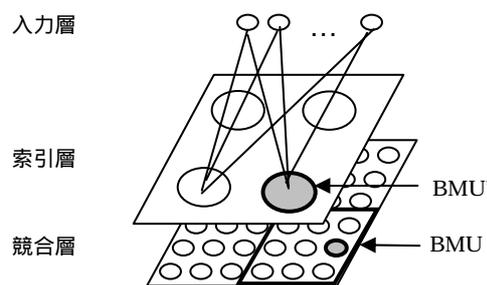


図 1 索引層を用いた SOM アーキテクチャ

3. 合成データによる従来法との比較実験

本手法の有用性を確認するために、二次元格子データの復元問題に対し、従来法と改良法との性能比較を試みた。実験には、 50×50 の二次元格子点データ 2500 件をデータセットとして用いた。学習に用いたパラメータはマップサイズ 50×50 、学習回数 3000 回、距離関数はユークリッド距離、索引層サイズ 5×5 、初期学習率 0.4、初期近傍サイズ 21×21 とした。

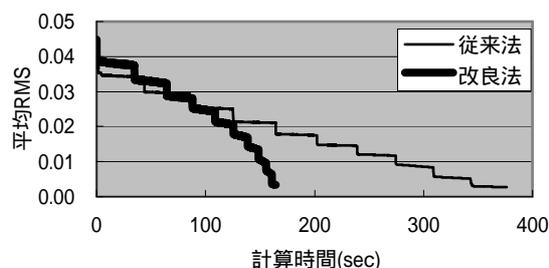


図 2 二次元格子点データ復元問題の計算時間と RMS

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

図 2 に RMS と計算時間について従来法と改良法で比較を行った結果を示す。従来法および改良法によって最終的に得られたマップの RMS 値はそれぞれ 0.0027、0.0033 となり、ほぼ同等な写像がえられた。一方、計算時間はそれぞれ 376.6 秒、164.438 秒となり、約 1/2 に短縮することができた。以上のことから、索引層を導入した改良法は、得られるマップの精度も従来法のものと比較して遜色がなく、なおかつ計算時間の短縮が可能であると考えることが示された。

4. 薬物の構造類似性マッピングへの応用

治験薬構造データベース MDDR[MDL01]に記載されている Dopamine Antagonist 活性を持つ化合物 1364 件に対し、化学物質のトポロジカルな構造プロフィールの記述表現として提案されている TFS (Topological Fragment Spectra)[Takahashi98]を用い、上述の索引層を有する SOM による構造類似性マッピングを試みた。化学構造の TFS 表現による入力パターン次元数は 165 であり、マップサイズは 40×40、学習回数は 1000 回、類似度の計算のための評価関数にはユークリッド距離を用いた。また、初期学習率は 0.4、初期近傍サイズは 19×19、索引層のサイズは 8×8 として実験をおこなった。図 3 に得られたマップを示す。

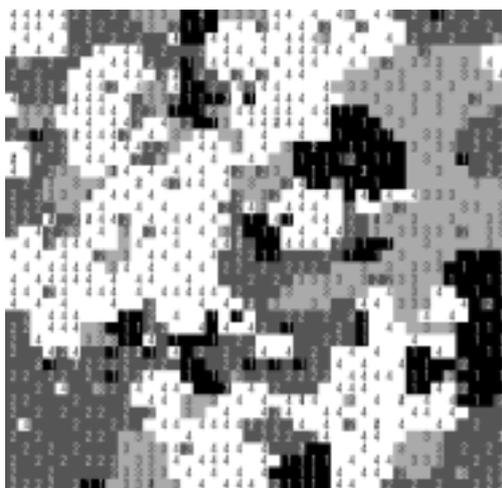


図 3 構造類似性マッピングの結果(索引層サイズ 8×8)

図 3 の結果は、各化合物の活性クラス Dopamine (D1~D4) Antagonists ごとに異なる色を用いて応答ニューロンを表示している。この結果から、同じ活性クラスに属する化合物群が互いに近傍に写像され、活性クラスごとにクラスター(あるいはサブクラスター)を形成していることがわかる。

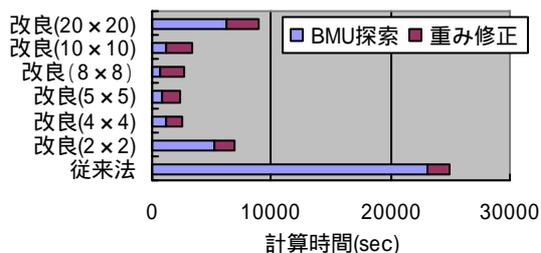


図 4 索引層サイズと学習時間の関係 ()内は索引層サイズ

図 4、表 1 に索引層のサイズと学習時間の関係を表す。従来法では 1000 回の学習回数に対して 25010 秒(BMU 探索時間

23123 秒、重み修正 1885 秒)であった。一方、サイズ 8×8 の索引層を用いた改良法では、同様に 1000 回の学習回数で処理時間が 2632 秒(BMU 探索 716 秒、重み修正 1914 秒)となり、従来法の約 1/10 となった。

表 1. 各手法における BMU 探索、重み修正時間

	従来法	2×2	4×4	5×5
BMU 探索	23123	5192	1102	770
重み修正	1885	1765	1462	1552
		8×8	10×10	20×20
BMU 探索		716	1135	6173
重み修正		1914	2162	2786

5. 従来法、改良法での計算時間比較

次に、データの数や次元数が異なる複数の実データ集合に対して従来法と改良法で学習を行い、計算効率(=従来法での訓練時間/改良法での訓練時間)がどのように変化するかを比較検討した。これらの結果をまとめて図 5 に示す。実験はすべて、マップサイズ 20×20、近傍空間サイズ 11×11、初期修正量 0.4、ユークリッド距離規則の条件下で学習を行なった。

この結果から、データ集合のサイズが大きくなるほど、計算効率が向上している傾向が読み取れる。このことは、索引層を導入した本手法が大容量・多変量データ空間の可視化に有効であることを示している。

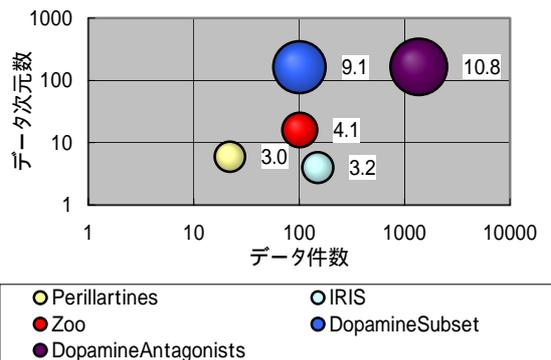


図 5 データセットの次元数・件数が訓練時間に及ぼす影響

6. おわりに

本研究では、SOM に索引層を加えることで、学習時間が従来法と比べ 1/3 ~ 1/10 と大幅に短縮できることを示した。

今後は、初期マップの効率的な生成法の工夫や Batch Learning SOM 等の併用による高速化に向けて引き続き検討を重ねたい。

参考文献

[デイホフ 92] J. デイホフ, ニューラルネットワークアーキテクチャ入門, 森北出版, 1992
 [徳高 99] 徳高平蔵, 岸田悟, 藤村喜久郎, 自己組織化マップの応用, 多次元情報の 2 次元視覚化, 海文堂出版, 1999
 [MDL01] MDL Information Systems, Inc., MDL Drug Data Report, Ver 2001.
 [Takahashi98] Y. Takahashi, H. Ohoka, Y. Ishiyama, In "Advance in Molecular Similarity", 2, (Eds, R Carbo & P. Mezey), JAI Press, Greenwich, CT, 1998, pp.93-104