

新しいモデリング・ツールとしての Kohonen ネットワーク

豊橋技術科学大学 船津公人

funatsu@tutkie.tut.ac.jp

はじめに

Kohonen の自己組織化ネットワーク[1]は教師なしニューラルネットワーク (NN) 手法の一種であり、高次元の入力データをトポロジカルな隣接関係を保持したまま二次元平面上に配置された出力層ニューロンに投影する、一種の非線形写像の能力を持つ。これらの性質から Kohonen ネットワークによる写像 (以降、Kohonen マッピング、また結果として得られる写像を Kohonen マップと呼ぶことにする) は、データ・モデリングの前処理であるデータ分類・情報圧縮に有効な新しい手法として注目されている。化学の分野でも、物理化学パラメータを用いた反応の分類[2,3]が Kohonen ネットワークによって行われるなど、その注目度は高い。

一方、Kohonen ネットワークによるデータ間の隣接関係を保存した写像は、他の非線形写像の方法にない特徴であるが、それは同時にデータに含まれない空白部分を著しく圧縮する形での写像であるため、Kohonen マップ上では入力空間における実際の距離関係が表現されないという問題点ともなっている。これは異なるモデルを構築すべきデータの集団 (クラス) が入力データ中に複数含まれる場合、Kohonen マップからはその識別ができない、ということの意味しており、モデリング・ツールとして Kohonen マッピングを用いる際に大きな問題となる。

そこで我々は、この Kohonen マッピングの短所を補う方法として、出力層ニューロンの重みベクトル集合に対してパターン認識の手法である主成分分析とポテンシャル関数法[4,5]の適用を検討した[6]。これら二つの手法はデータ中のクラスタの存在を認識する重要な要素である、データの分布位置と分布密度の二つを近似的に表現する手法で、Kohonen マップ上で入力空間の様子を表現する手法として妥当である。

ここでは実際に、Italian Olive Oil の産地分類問題[7]に対して本手法を応用した結果を検討することで新しいモデリング・ツールとしての Kohonen ネットワークの可能性を検討する。

Kohonen ネットワーク

Kohonen ネットワークは図 1 に示すように、完全結合で結ばれた入力層と出力層の二層から構成される。入力層は入力データの次元数と同じ m 個の入力層ニューロンによって構成され、入力ベクトルの各要素を出力層の全ニューロンに伝える。出力層は一般に二次元平面上の格子点上に配置された出力層ニューロンの集合からなり、それぞれが入力ベクトルと

同じ m 次元の重みベクトルを持つ。初期段階では、重みベクトルは最初、乱数によって初期化される。

Kohonen ネットワークの学習は、参照ベクトルの選択、類似性マッチング、重みベクトルの修正、の 3 つの操作を順次繰り返すことによって行なう。

参照ベクトルの選択は、学習の対象となる入力ベクトル群の中から一つ、その学習ステップで参照する入力ベクトル x をランダムに選び出す操作である。類似性マッチングは、出力層上の全ニューロンの中から、参照ベクトルに最も類似する重みベクトルを持つ勝者ニューロン c を選び出す操作である。類似性の尺度には、入力ベクトル x と重みベクトル w との間のユークリッド距離が一般に用いられる。

勝者ニューロン c が判別されたなら、次に勝者ニューロン c とその近隣のニューロンの重みベクトルを入力ベクトル x に近づくように、重みベクトルの修正を行なう。修正の範囲は勝者ニューロン c を中心としたトポロジカル距離 r_{max} の範囲に属するニューロンに対して行われる (図 2)。ただし、学習範囲 $r_{max}(t)$ は式(2)に示すように学習の繰り返し回数 t の増大とともに減少する関数である。重みベクトルの修正量は勝者ベクトルからのトポロジカル距離 r の関数である近傍関数 $h(r, r_{max}(t))$ と学習率関数 $a(t)$ によって決定する。学習率関数 $a(t)$ は式(3)に示すような学習の繰り返し回数 t の増大とともに減少する関数であり、近傍関数は式(4)に示すようなトライアングル型の近傍関数が用いられることが多い (この他、レクトアングル型、メキシカンハット型等の関数を用いる場合もあり、関数型によって学習の結果が異なる場合がある) これらを組み合わせた重みの修正式は式(5)のようになる。

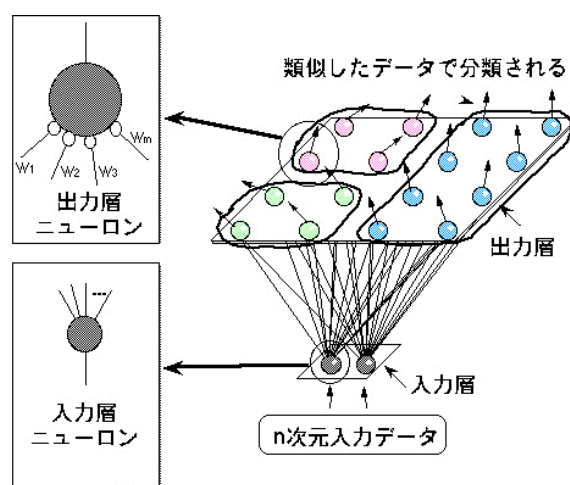


図 1 Kohonen ネットワーク

図9はKohonenマップ上でのポテンシャル表面を示したものである。ポテンシャル表面では独立した山の形で表わされている部分がクラスタにあたる。斜めからの図なので判別が難しいが、おおよそ4つのクラスタがほぼ実際の産地の分類の結果に類似した形で分布している。

3.3で述べたようにポテンシャル関数法では、そのポテンシャル表面の極小点(図中では谷間として表わされる)を探索することで、各クラスタ間の境界を識別することができる。図10は図9で示したポテンシャル表面の極大点にあたるニューロンから、最近隣ニューロンとの間にポテンシャル表面の極小点が存在するかどうかを順次探索し、Kohonenマップ上におけるクラスタの自動識別を行なった例である。右上のクラスタに小さなサブ・クラスタが存在していることを除けば、ほぼ産地別の分類の結果と同様の分類結果になっている。

考察

Kohonenマップ上での重みベクトルの挙動をU-Matrix法、主成分分析手法、ポテンシャル関数法の三つの手法を用いて把握する方法について検討した。

隣接するニューロン間のローカルな距離関係を用いて、データ中で明確に違いがある部分を視覚化するU-Matrix法は、明らかなクラスタを認識するためには有効な手法であることが再確認された。しかし、グローバルな観点からみたデータの挙動を表わしたものではないので、判別された各クラスタ間の関係を理解できないという問題点は残ったままである。

一方、主成分分析手法による結果は、他の二つの手法とは性格が異なり、マップ上でのデータの分布の様子をグローバルな観点から検討することができるという特徴を持っている。また、主成分方向を与えるローディングベクトル p を検討することで、クラスタを特徴づける因子を見つけたす手掛かりを得ることができるという点は、モデリングのツールとして用いる上で重要な要素である。

ポテンシャル関数法は代表サンプルの獲得と自動クラスタ認識を同時に行なうことができる手法であり、Kohonenネットワークと併用することで高次元データへの応用が可能となった。ここでは紹介しなかったが、有機反応の自動分類への応用をすでに行なっている[11]。

今回紹介した手法を組み合わせることによって、

Kohonenマップ上でのデータの分布が明らかになり、従来困難だったKohonenマップ上での明確なクラスタ分類が可能になった。これはKohonenネットワークによる排他的な分類が可能になったことを意味する。これらの結果から、Kohonenネットワークは、モデリングを効率的に行なうためのツールとして、さらに有用性を高めるものと期待される。

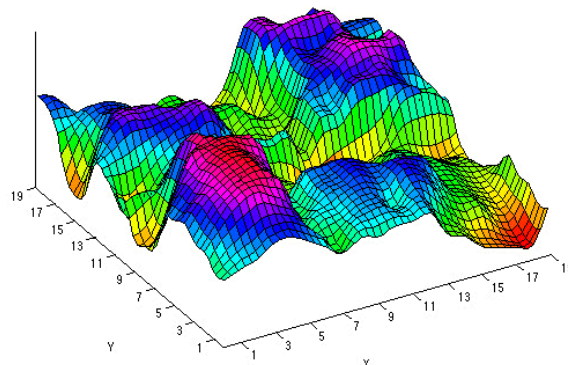


図9 ポテンシャル表面

参考文献

- [1] T.Kohonen, Self-Organizing Maps, Springer - Verlag:Berlin(1995).
- [2] L.Chen, J.Gasteiger; J.Am.Chem.Soc., 119, 4033-4042(1997).
- [3] H.Satoh, O.Sacher, T.Nakata, L.Chen, J.Gasteiger, K.Funatsu; Submitted.
- [4] D.Coomans, D.L.Massart; Anal.Chim.Acta., 43, 215-224(1981).
- [5] D.Coomans, D.L.Massart; Anal.Chim.Acta., 43, 225-239(1981).
- [6] 葛西哲郎,船津公人; 第20回情報化学討論会ポスターセッション(1997).
- [7] M.Forina, C. Armanino; Annali di Chimica, 72, 127-143(1982).
- [8] A.Ultsch,et.al.; Proc.Transputer Anwender Treffen/World Transputer Congress TAI/WTC 93 Aachen, Springer-Verlag: New York, 194-203(1993).
- [9] Helsinki Univ. of Tech.,Lab.of Comp. and Info. Sci., (ftp://cochlea.hut.fi or 130.233.168.48).
- [10] J.Zupan, J.Gasteiger; Neural Networks for Chemists, VCH (1993).
- [11] H.Satoh, K.Funatsu, T.Nakata; 214th ACS National Meeting (1997).
- [12] 佐藤寛子,船津公人,中田忠; 第73回日本化学会秋季年会(1997).

ふなつ きみと FUNATSU, Kimito

有機合成設計や反応予測を特に情報化学的観点から扱うためのアルゴリズムの研究とシステム化を行なっている。また、有機化合物の自動構造解析システムの開発や三次元構造活性相関のための手法の研究開発も進めている。

連絡先 〒441 豊橋市天伯町字雲雀ヶ丘 1-1 豊橋技術科学大学知識情報工学系
電話 0532-44-6879