

近傍学習と競合学習を融合した教師あり SOM

日大生産工(院) ○小松 泰士 日大生産工 山内 ゆかり

1. まえがき

データマイニングにおいてデータの構造や関係性を視覚化することは隠れた情報を解釈するうえで重要である。データの視覚化としては主成分分析や多次元尺度構成法といった多変量解析の手法がよく利用されている。他方でデータを視覚化するニューラルネットワークとして自己組織化マップ (Self-Organizing Map: SOM)[1]がある。SOMはKohonenが提案した代表的な教師なし学習法の一つであり、高次元データを2次元という低次元空間の競合層に配置することでその関係性を視覚的に把握することが可能である。

SOMは教師なしの過程を経て形成されることが一般的ではあるが、もし学習段階でクラス情報が考慮されるのであれば、データの分類制度の著しい改良が可能なが分かっている。同じくKohonenによって教師データを入力次元に加えてSOMと同じアルゴリズムで学習する教師ありSOM[2]が提案されている。教師ありSOMでは参照ベクトルの次元数に応じて教師ラベルの学習への影響が変わってしまう問題がある。

そこで、参照ベクトルと教師ラベルに付随する重みベクトルを別々に分け、双方の類似度に応じて競合学習を行う対向伝搬学習ネットワーク (Counter Propagation Network: CPN)[3]がNielsonによって提案された。教師ラベルの持つ重みを独立することで参照ベクトルの次元数の影響を受けずに学習可能となった。

SOM, 教師ありSOMの学習アルゴリズムでは勝者ベクトルからの距離に応じて学習率を変化させることで、競合層マップ上にクラス別の集合を生成する。しかし、クラスの境界領域となったノードの持つベクトルが不鮮明となり、どのクラスにも該当しない不要ノードとなり、認識率が低下すると考えられる。

本研究ではSOMの競合層の分散度の抑制及び認識制度の向上を目的とし、参照ベクトルと属性値情報の競合学習を可能にした属性値SOMを提案する。実験では手書き文字データMNISTを使用し、認識率及び競合層のラベルの分散度を従来のSOMと比較し、提案手法の有効性を示す。

2. 従来研究

ここでは、従来研究のアルゴリズムについて説明する。

2.1. 自己組織化マップ

自己組織化マップ (Self-Organizing Map: SOM)[1]とはKohonenが提案した教師なし学習アルゴリズムである。高次元データを2次元格子状マップに可視化することができる。

以下に一般的なSOMのアルゴリズムを示す。

- A) 競合層のノード番号 i の参照ベクトル m_i をランダムな値で初期化する。
- B) 訓練データからランダムに一つ選び、入力ベクトル x を作成する。
- C) 式(1)により層の参照ベクトル m_i と入力ベクトル x のユークリッド距離を計算し、もっとも類似している(差の小さい)勝者ノード番号 c を算出する。

$$c = \underset{i}{\operatorname{argmin}} \{ \|x - m_i\| \} \quad (1)$$

- D) 勝者ノード c と競合層ノード i 間の近傍関数 h_{ci} を式(2)で求める。 a_t と σ_t はそれぞれ学習率と学習範囲であり、式(3)によって求められる。 a_0 と σ_0 は初期の値を示しており、現在の学習回数 t と総学習回数 T を用いることで学習回数に応じて徐々に減少する。

$$h_{ci} = a_t \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma_t^2}\right) \quad (2)$$

$$a_t = a_0 \cdot \left(1 - \frac{t}{T}\right) \quad (3)$$

$$\sigma_t = \sigma_0 \cdot \left(1 - \frac{t}{T}\right)$$

- E) (B)で求めた近傍を用いて式(4)により競合層の参照ベクトル m_i の更新を行う。

$$m_i(t+1) = m_i(t) + h_{ci}(t) \cdot (x_t - m_i(t)) \quad (4)$$

- F) (B)~(E)の処理を総学習回数 T を満たすまで繰り返し行う。

SOMのアルゴリズムでは、近傍関数 h_{ci} により勝者ノード c からの距離に応じて学習率 a_t が減少する。そのため、似ている参照ベクトルがそれぞれ近くに集まり、結果として1つの2次

元マップ上に別々のクラスの集合を形成することができる。

図.1の例で示すように競合層のクラスとクラスの境界部分には中間パターンが生成される。このことにより訓練データが少ない場合でも複数のパターンを補完することができる。しかし、生成される境界ベクトルの所属クラスが曖昧になったしまう場合、認識が困難になる問題がある。また、クラスの配置が散らばるほど、認識率に悪影響が出ると考えられる。

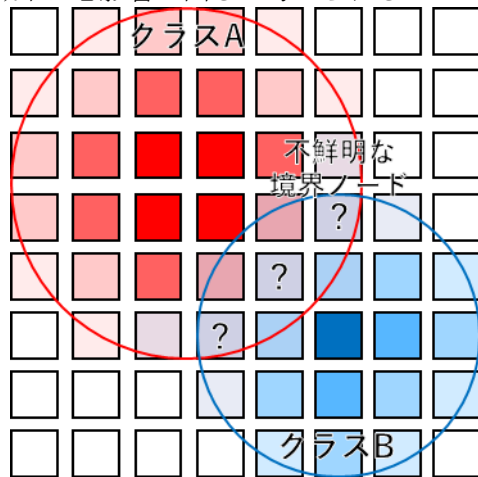


図1 クラスの境界・識別不明ノードの例

2.2. 教師あり自己組織化マップ

本稿で提案する以前1984年頃、すでにクラス情報を用いた自己組織化マップは存在しており、教師ありSOM[2]と呼ばれる。

通常の学習データのベクトル x_s とクラス情報をOne-hotベクトル化した x_u に分けて作成し、 x_s と x_u を連結させて入力ベクトル x を作成する。クラス情報を含んだ x_u により、同じクラスのベクトルは同じに、異なったクラスのベクトルは異なったものとしてSOMの学習で扱われるため、競合層のクラス分離が促進される。未知データ x の認識では x_s 部分のみが競合層の対応するベクトルと比較され、勝者のベクトルを決定する。

この手法では x_s の次元数に応じて入力ベクトル x に対するクラス情報 x_u の次元数割合が変化する。そのため x_s の次元数が膨大になればなるほど、クラス情報の影響度が小さくなる問題がある。

2.3. 対向伝搬学習ネットワーク

対向伝搬学習ネットワーク (Counter Propagation Network: CPN)[3]とはHecht Nielsenが提案した素晴らしいアルゴリズムである。ネットワーク構造を図.2に示す。入力層と第1の結合荷重で結びつけられたKohonen層と第2の結合荷重で結びつけられたGrossberg層と競合層の3層で構成される。

Kohonen層には通常の学習データベクトル x を入力として与え、Grossberg層には学習データのクラス情報 y を入力として与える。Kohonen層の重み r とGrossberg層の重み s の更新式を式(5)に示す。更新にはそれぞれ異なる学習率 a_x と a_y を設定する。

$$\begin{aligned} r_i &= a_x(x - r_i) \\ s_i &= a_y(y - s_i) \end{aligned} \quad (5)$$

教師あり自己組織化マップとは異なり、それぞれ別の過程で参照ベクトルの更新を行う。未知データ x の認識ではGrossberg層の入力をすべて0として計算する。

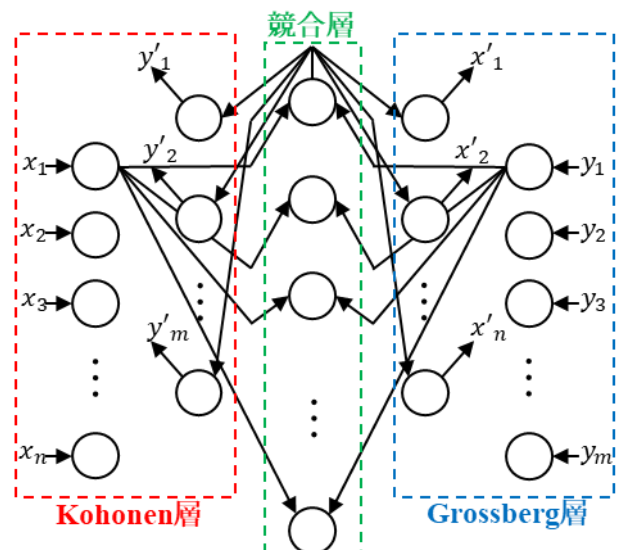


図2 CPNのネットワーク構造図

3. 提案手法

本研究では従来のSOMの競合層の参照ベクトル m_i に加えて、クラス(属性値)情報 L の近傍学習を加えた属性値SOMを提案する。属性値情報 L とは、CPNのGrossberg層のように各競合層ノードがどのクラスに属しているかを示す値である。属性値情報 L はクラスごとに算出し、最も属性値情報の高いクラスを該当する競合層ノードのクラスとみなす。

属性値SOMは従来のSOMのアルゴリズムA)~C)は同一であり、以降の参照ベクトルの更新を行うD)~F)に改良を加えている。以下に提案手法である属性値SOMの競合層の参照ベクトル m_i とクラス(属性値)情報 $L(i)$ の更新アルゴリズムを示す。

D) 式(6)により競合層ノード i 番目の属性値情報 $L(i)$ を更新する num とは入力データの属性値番号を示す。属性値学習範囲 σ'_t はSOMと同様に学習範囲を表すが、参照ベクトルの学習範囲 σ_t とは別の初期値 σ'_0 を用いて式(7)から求める。 γ は属性値の減衰率を示す。

$$L_{num}(i) = \gamma \cdot L_{num}(i) + \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma_t'^2}\right) \quad (6)$$

$$\sigma_t' = \sigma_0' \cdot \left(1 - \frac{t}{T}\right) \quad (7)$$

E') 近傍関数 h_{ci} を式(8)により求める．学習率 a_t は式(9)を用いる W は学習率 a_t に対する重みであり，式(10)により求める．

$$h_{ci} = a_t \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma_t'^2}\right) \quad (8)$$

$$a_t = a_0 \cdot \left(1 - \frac{t}{T}\right) \cdot W \quad (9)$$

$$W = \frac{L_{num}(i)}{\max_j L_j(i)} \quad (10)$$

F') E')で求めた近傍 h_{ci} を用いて競合層ノードの参照ベクトル m_i をSOMと同様の式(4)により求める．

G') (B)~(F')の処理を総学習回数 T を満たすまで繰り返し行う．

未知の入力ベクトル x の識別を行う場合，参照ベクトル m_i を参照して競合層の勝者ノード番号 c を求め，勝者ノードの持つ属性値情報 $L(c)$ が最も大きいクラスを出力とする．

提案手法の属性値SOMでは，競合層ノードが持つクラス情報の重みを属性値情報 L として与え，属性値情報 L を用いて競合層ノードの学習率 a_t を変化させている．このことにより，過去に学習した競合層ノードの分類ラベルによって入力データが同じラベルであれば強く学習し，反対に違うラベルであれば弱く学習する．入力と競合層のクラス情報を比較して学習箇所を変化させることで，競合層のクラスの分類の境界をはっきりし，クラスの散らばりを抑えることにより，学習精度の向上が期待できる．

図3に従来のSOMと提案する属性値SOMの学習率の違いを示す．左側の従来のSOMでは勝者ノード c からの距離に応じて均一に学習率が変化している．周辺の学習度合いを考慮しないため，クラスの境界ノードが生成されやすくなっている．それに対し，右側の提案する属性値SOMでは入力データのクラスに応じて学習率を変化させ，過去の学習により，入力と同じクラスAを持つ競合層ノードは通常の学習を行い，異なるクラスBとCを持つノードは学習率が低下していることが分かる．

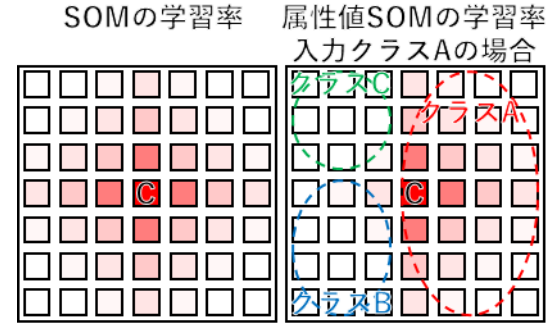


図3 従来手法と提案手法の学習率の違い

4. 実験

実験には手書き文字データMNISTを用いる．MNISTとは0~9の数字手書き文字を784次元で表現したデータと該当クラスを含むデータセットである．表1に本実験で用いた実験パラメータを示す．

実験は，従来手法であるSOMと教師ありSOMとCPN，提案手法である属性値SOMの4種類の学習アルゴリズムで行う．属性値SOMは減衰率 γ が場合とない場合の両方で行う．

評価手法はMNISTのテストデータ認識率とクラス分散度によって行う．クラス分散度とは2次元の競合層マップ上でどれだけ各クラスが分散したかを示す数値である．競合層に分布する各クラスのXY座標からクラス面積を求め，マップ全体の総面積で割ることによって各クラスがマップ全体に占める割合(クラス分散度)を計算する．

表1 実験パラメータ

実験環境	
訓練データ数	60000
テストデータ数	10000
競合層サイズ	12
競合層ノード数	144
学習回数	60000
学習パラメータ	
SOM学習率 α_0	0.3
SOM学習範囲 σ_0	4.0
属性値ラベル学習範囲 σ'_0	1.0
属性値ラベル減衰率 γ	0.99
CPN学習率 a_x	0.1
CPN学習率 a_y	0.0001

5. 結果及び考察

本実験は100回ずつ全てのアルゴリズムで実験を行い、平均を結果としている。表2にアルゴリズムごとのMNISTのテストデータの認識率を示す。教師ありSOMをTSOM、属性値SOMをASOM、減衰率ありの属性値SOMを γ -ASOMとして表示している。

表2 クラス別文字認識率

	CPN	SOM	TSOM	ASOM	γ -ASOM
0	0.877	0.941	0.938	0.951	0.958
1	0.812	0.981	0.978	0.987	0.982
2	0.786	0.882	0.884	0.869	0.914
3	0.675	0.819	0.820	0.852	0.881
4	0.841	0.707	0.711	0.787	0.845
5	0.773	0.808	0.798	0.783	0.860
6	0.813	0.939	0.935	0.942	0.955
7	0.718	0.856	0.855	0.865	0.883
8	0.503	0.780	0.777	0.806	0.852
9	0.447	0.711	0.731	0.749	0.835
平均	0.725	0.844	0.845	0.861	0.898

提案手法であるASOM、 γ -ASOMは従来のTSOMに比べてそれぞれ1.6%、5.3%認識率が改善した。これは属性値(教師)情報を競合層の学習率の変化に使い、クラスの境界が生成されにくくなることで競合層のノードを効率的に利用できたからだと考えられる。また、属性値の更新に減衰率 γ を追加することにより、複数クラスの属性値を持つノードを減少させたからだと考えられる。

次に表3にクラス別競合層の分散度を示す。

表3 クラス別マップ分散度

	SOM	TSOM	ASOM	γ -ASOM
0	0.174	0.141	0.188	0.262
1	0.137	0.136	0.158	0.167
2	0.373	0.353	0.248	0.393
3	0.309	0.341	0.190	0.283
4	0.361	0.308	0.256	0.331
5	0.501	0.466	0.376	0.538
6	0.204	0.204	0.222	0.329
7	0.292	0.227	0.222	0.287
8	0.338	0.318	0.279	0.306
9	0.283	0.284	0.314	0.369
平均	0.297	0.278	0.245	0.326

提案手法であるASOMでは従来のTSOMに比べ、クラスの散らばりを平均3.3%抑制した。これは競合層ノードの学習に属性値情報反映させることでクラスの境界が明確になったためだと考えられる。反対に減衰率を与えた場合、全体で4.8%クラスの散らばりが増大した。これは減衰率により、一つのノードが持つクラス情報が一意に決定するため、クラス間の境界のクラス情報分離し、競合層のクラス情報の分布が滑らかにならなかったからだと考えられる。

6. まとめ

本研究ではSOMの競合層の分散度の抑制及び認識制度の向上を目的とし、参照ベクトルと属性値情報の双方の競合学習を可能にした属性値SOMを提案した。実験では手書き文字データMNISTを使用し、認識率及び競合層のラベルの分散度で評価を行った。

従来手法のSOMに比べて提案手法の属性値SOMでは属性値情報 L によって競合層の学習率 a_t を制御することによりクラス間の不鮮明なノードを削減し、クラスの散らばりを3.3%抑えることができた。更に、属性値情報 L に学習過程で減衰率を加えることにより、競合層ノードが持つクラス分類を特定のクラスに際立たせることで、認識率を5.3%改善できた。

本実験の提案手法では減衰率の導入によって競合層上のクラスの分散度と認識精度がトレードオフの関係になった。SOMアルゴリズムは認識精度だけでなく2次元マップ上のクラス分布の可視化が特徴なため、双方の改善が可能にすることが今後の課題である。また、SOMには近傍学習によってデータセットに不足した特徴を補う利点がある。データセットに偏りがあった場合でも本提案が有効に働くが実験を試みたいと考える。

参考文献

- [1] Teuvo Kohonen, "Self-Organized Formation of Topologically Correct Feature Maps," Biological Cybernetics, vol.52, no.3, (1982) pp.59-69.
- [2] Teuvo Kohonen, "自己組織化マップ(改訂版)," 丸善出版株式会社, (2012) pp.221-223
- [3] R.H. Nielsen, "Counter propagation networks," Appl. Opt. vol.26, (1987) pp.4979-4983