

On variable learning rate in self-organizing maps on general input space

(可変学習率をもつ一般入力型自己組織化マップについて)

秋田県立大学 システム科学技術学部 星野満博 (Mitsuhiro Hoshino)
Faculty of Systems Science and Technology, Akita Prefectural University

1. 自己組織化マップモデル

本報告は Kohonen 型アルゴリズム [7] として知られている自己組織化マップモデルにおける学習による順序化, 整列化とモデル関数の性質に関する一つの理論的考察である. 自己組織化マップモデルにおけるノードの配列とノードの値との間に現れるある種の規則性について, ノード値の順序化, 整列化の形成過程と状態クラスの閉性に注目する. 自己組織化マップは非常に実用的であり広範囲に応用例を有し, アルゴリズムも非常にシンプルであるが, その数学的構造はあまり明らかではない.

本報告では, 一般的に状態クラスに厳密な意味での閉性をもたないような学習近傍や学習ルールをもつモデルにおいて, 可変学習による閉性の保存, 整列化について考察する.

本報告では, 自己組織化マップモデルをノード, ノードの値, インプット, 学習プロセスの 4 つの要素によって, 以下の様に定義する.

$$(I, V, X, \{m_k(\cdot)\}_{k=0}^{\infty})$$

- (i) I をすべてのノードの集合とする. I は, 距離 d をもつある距離空間の加算部分集合とする.
- (ii) 各ノードは, それぞれ 1 つの値をもつ. V をノードの値の空間とする. V はノルム空間であると仮定する. V におけるノルムを $\|\cdot\|$ とする. $m(i)$ をノード i の値として, その対応 $m: I \rightarrow V$ をモデル関数と呼ぶことにする. また, M をモデル関数の全体, $m_0: I \rightarrow V$ を初期モデル関数とする.
- (iii) $X \subset V$ を入力集合とする. $x_0, x_1, x_2, \dots \in X$ を入力列とする.
- (iv) 学習プロセス

$$m_{k+1}(i) = (1 - \alpha_{m_k, x_k})m_k(i) + \alpha_{m_k, x_k}x_k \quad (1)$$

ここで, α_{m_k, x_k} は, $0 \leq \alpha_{m_k, x_k} \leq 1$ を満たす学習率を表す.

例えば, n 個のノード $1, 2, \dots, n$ がある場合を考える. そのそれぞれに対してノードの値 $m_0(1), m_0(2), \dots, m_0(n)$ が与えられているとする. このとき, 入力とこれに伴う学習により各ノードの値が更新される. $x_0 \in X$ が入力されたならば,

$$m_1(i) = (1 - \alpha_{m_0, x_0})m_0(i) + \alpha_{m_0, x_0}x_0$$

により, ノードの値を更新する. インプット x_1, x_2, x_3, \dots に対して, これを繰り返すことにより, モデル関数 m_1, m_2, m_3, \dots が逐次に生成され, ノードの更新が更新される.

このような学習を十分な回数, 繰り返したとき, モデル関数において, 単調性等, 各ノードの値の配列にある種の規則性が現れることがある. 様々なノード集合, ノードの値の空間, 学習方法において, 単調化等の幾つかの興味深い現象が現れる.

2. \mathbb{R} 値ノード, 1次元ノード配列モデルと状態クラスの閉性

ここでは, 最も単純な自己組織化マップモデルである, \mathbb{R} 値ノード, 1次元ノード配列の場合について述べる.

$$(\{1, 2, \dots, N\}, V \subset \mathbb{R}, X \subset \mathbb{R}, \{m_k(\cdot)\}_{k=0}^{\infty})$$

- (i) 有限個のノードを仮定する. $I = \{1, 2, \dots, N\} \subset \mathbb{N}$.
- (ii) ノード値の空間を \mathbb{R} (ユークリッドノルム) とする. $m_0 = [m_0(1), m_0(2), \dots, m_0(n)]$ などと記すことにする.
- (iii) $x_0, x_1, x_2, \dots \in X \subset \mathbb{R}$ を入力列とする.

基本的な学習プロセスとして次のタイプが挙げられる. この学習プロセスに対して以下の基本的な性質が成り立つ.

Theorem 1 1次元入力型自己組織化マップ

$$(\{1, 2, \dots, N\}, V \subset \mathbb{R}, X \subset \mathbb{R}, \{m_k(\cdot)\}_{k=0}^{\infty})$$

において, 次の学習プロセスを仮定する.

学習プロセス L_A (1次元配列, \mathbb{R} -値ノード, $\varepsilon = 1$)

(a) 学習範囲:

$$I(m_k, x_k) = \left\{ i^* \in I \mid |m_k(i^*) - x_k| = \inf_{i \in I} |m_k(i) - x_k| \right\}$$

$$(m_k \in M, x_k \in X),$$

$$N_1(i) = \{j \in I \mid |j - i| \leq 1\} \quad (i \in I).$$

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & (i \in \bigcup_{i^* \in I(m_k, x_k)} N_1(i^*) \text{ のとき}) \\ m_k(i) & (i \notin \bigcup_{i^* \in I(m_k, x_k)} N_1(i^*) \text{ のとき}) \end{cases}$$

$$k = 0, 1, 2, \dots$$

このとき、モデル関数に対して以下が成り立つ。

- (i) モデル関数 m_k が I 上で単調増加であるならば、モデル関数 m_{k+1} も I 上で単調増加である。
- (ii) モデル関数 m_k が I 上で単調減少であるならば、モデル関数 m_{k+1} も I 上で単調減少である。
- (iii) モデル関数 m_k が I 上で狭義単調増加であるならば、モデル関数 m_{k+1} も I 上で狭義単調増加である。
- (iv) モデル関数 m_k が I 上で狭義単調減少であるならば、モデル関数 m_{k+1} も I 上で狭義単調減少である。

ここでの単調増加性、単調減少性のように、モデル関数が一度その状態になると、その状態が保存されるという意味において、このような状態のクラスを自己組織化マップモデルの閉じた状態クラスと呼ぶことにする。

3. 穴あき型学習近傍と可変型学習

1次元入力型自己組織化マップ ($\{1, 2, \dots, N\}, V \subset \mathbb{R}, X \subset \mathbb{R}, \{m_k(\cdot)\}_{k=0}^\infty$) において、以下の学習プロセスを導入する。

学習プロセス L_{E_1} (1次元配列, \mathbb{R} -値ノード, $\varepsilon = 1$)

(a) 学習範囲:

$$I(m_k, x_k) = \left\{ i^* \in I \mid |m_k(i^*) - x_k| = \inf_{i \in I} |m_k(i) - x_k| \right\}$$

$$(m_k \in M, x_k \in X),$$

$$N_1(i) = \{j \in I \mid |j - i| \leq 1\} \quad (i \in I).$$

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & (i \in \left(\bigcup_{i^* \in I(m_k, x_k)} N_1(i^*) \right) \setminus I(m_k, x_k) \text{ のとき}) \\ m_k(i) & (i \notin \left(\bigcup_{i^* \in I(m_k, x_k)} N_1(i^*) \right) \setminus I(m_k, x_k) \text{ のとき}) \end{cases}$$

$$k = 0, 1, 2, \dots$$

上記の学習範囲では、ノード i^* の近傍から i^* を除いた穴あき型近傍を用いる。近傍として $(\cup_{i^* \in I(m_k, x_k)} N_1(i^*)) \setminus I(m_k, x_k)$ の代わりに $\cup_{i^* \in I(m_k, x_k)} (N_1(i^*) \setminus \{i^*\})$ を用いることもある。

L_{E_1} を用いたとき、 L_A を用いた場合と異なり、単調性は保存されるとは限らない。しかし、以下のような可変学習率を導入することにより単調性を保存することができる。

Theorem 2 1次元入力型自己組織化マップ

$$(\{1, 2, \dots, N\}, V \subset \mathbb{R}, X \subset \mathbb{R}, \{m_k(\cdot)\}_{k=0}^{\infty})$$

と学習プロセス $L_{E_1}(\varepsilon = 1)$ 及び、以下の可変学習率 $\alpha_{m,x}$ を用いることを仮定する。

$$I(m_k, x_k) = \left\{ i^* \in I \mid |m_k(i^*) - x_k| = \inf_{i \in I} |m_k(i) - x_k| \right\},$$

$$i_m = \min I(m_k, x_k),$$

$$i_M = \max I(m_k, x_k)$$

とする。ここで、上の $\min I(m_k, x_k), \max I(m_k, x_k)$ は、自然数の順序の意味での最小値、最大値を意味する。与えられた α_0 ($0 < \alpha_0 \leq 1$) に対して、 α_1 を

$$m(i_M + 1) = m(i_m - 1) \text{ のとき } \alpha_1 = \alpha_0,$$

$$m(i_M + 1) \neq m(i_m - 1) \text{ のとき}$$

$$\alpha_1 = \min \left\{ \alpha_0, \frac{\min\{|m_k(i_m) - m_k(i_m - 1)|, |m_k(i_M + 1) - m_k(i_M)|\}}{|m_k(i_M + 1) - m_k(i_m - 1)|} \right\}$$

によって定める。ここで、 $i_m = 1$ のとき、 $m_k(i_m - 1) = x$ とし、 $i_M = N$ のとき、 $m_k(i_M + 1) = x$ とする。上記の α_1 に対して、学習率を $0 \leq \alpha_{m,x} \leq \alpha_1$ を満たすようにとる。

このとき、モデル関数に関して、次が成り立つ。

- (i) モデル関数 m_k が I 上で単調増加であるならば、モデル関数 m_{k+1} も I 上で単調増加である。
- (ii) モデル関数 m_k が I 上で単調減少であるならば、モデル関数 m_{k+1} も I 上で単調減少である。

このとき、学習率が小さいので整列化するまでに時間が掛かることに注意する。応用上は、固定学習率等を用いて十分な回数の学習をさせ、十分整列化した後に、上記の可変学習率を用いた学習プロセスを用いることにより、整列化と吸収性を保つことが可能となる。また、学習範囲 $\cup_{i^* \in I(m_k, x_k)} (N_1(i^*) \setminus \{i^*\})$ を用いた場合についても同様の結果が得られている。

4. 一般入力型自己組織化マップと可変型学習

1次元配列ノード, 一般入力型自己組織化マップ

$$(\{1, 2, \dots, n\}, V, X, \{m_k(\cdot)\}_{k=0}^{\infty}) \quad (2)$$

を次のように定義する.

- (i) $I = \{1, 2, \dots, n\}$. $d_I(i, j) = |i - j|$, $i, j \in I$.
- (ii) $m : I \rightarrow V$. V は内積 $\langle \cdot, \cdot \rangle$ をもつノルム空間.
- (iii) 入力 $x_0, x_1, x_2, \dots \in X \subset V$.
- (iv) 学習プロセス L_m ($\varepsilon = 1$)

(a) 学習範囲:

$$j(m, x) = \min\{i^* \in I \mid \|m(i^*) - x\| = \inf_{i \in I} \|m(i) - x\|\}, \quad m \in M, x \in X,$$

$$N_1(i) = \{j \in I \mid d_I(i, j) \leq 1\};$$

(b) 学習率: $0 \leq \alpha \leq 1$;(c) 学習: $i \in N_1(j(m, x))$ のとき $m'(i) = (1 - \alpha)m(i) + \alpha x$ とし, $i \notin N_1(j(m, x))$ のとき $m'(i) = m(i)$ とする.

上述の (2) において, 高々2つのノードを除く, すべてのノードに対して, ノードの値の条件

$$\langle m(i) - m(i+1), m(i+2) - m(i+1) \rangle \leq 0 \quad (3)$$

は, 学習の前後で保存されることがわかっている ([6]). 次のように可変学習率を用いることにより, すべてのノードに対して, この条件を保存することが可能となる.

Theorem 3 一般入力型自己組織化マップ

$$(\{1, 2, \dots, n\}, V, X, \{m_k(\cdot)\}_{k=0}^{\infty})$$

において, $L_m(\varepsilon = 1)$ と次の可変学習率を仮定する. ここで, モデル関数 m と入力 x に対して, 次のように定義された可変学習率 $\beta_{m,x}^+$ を用いて m を更新する.

$$\langle m(j(m, x) + 2) - m(j(m, x) + 1), m(j(m, x) + 3) - m(j(m, x) + 2) \rangle \geq 0$$

かつ

$$\langle x - m(j(m, x) + 1), m(j(m, x) + 3) - m(j(m, x) + 2) \rangle \geq 0$$

のとき

$$\beta_{m,x}^+ = \frac{\langle m(j(m, x) + 2) - m(j(m, x) + 1), m(j(m, x) + 3) - m(j(m, x) + 2) \rangle}{\langle x - m(j(m, x) + 1), m(j(m, x) + 3) - m(j(m, x) + 2) \rangle}$$

とし、その他のとき

$$\beta_{m,x}^+ = 1$$

とする。 $\beta_{m,x}^-$ を

$$\langle m(j(m,x) - 2) - m(j(m,x) - 3), m(j(m,x) - 1) - m(j(m,x) - 2) \rangle \geq 0$$

かつ

$$\langle m(j(m,x) - 2) - m(j(m,x) - 3), m(j(m,x) - 1) - x \rangle \geq 0$$

のとき

$$\beta_{m,x}^- = \frac{\langle m(j(m,x) - 2) - m(j(m,x) - 3), m(j(m,x) - 1) - m(j(m,x) - 2) \rangle}{\langle m(j(m,x) - 2) - m(j(m,x) - 3), m(j(m,x) - 1) - x \rangle}$$

とし、その他のとき $\beta_{m,x}^- = 1$ とする。 $0 < \beta_0 < 1$ を満たす定数 β_0 に対して、学習率 $\alpha_{m,x}$ は

$$0 < \alpha_{m,x} < \min\{\beta_0, \beta_{m,x}^+, \beta_{m,x}^-\}$$

を満たすものとする。このとき、学習の前後で、すべてのノードに対して

$$\langle m(i) - m(i+1), m(i+2) - m(i+1) \rangle < 0 \quad (4)$$

が保存される。

参考文献

- [1] M. Cottrell and J.-C. Fort, *Étude d'un processus d'auto-organisation*, Annales de l'Institut Henri Poincaré, 23(1) (1987), pp.1–20 (in French)
- [2] E. Erwin, K. Obermayer, and K. Schulten, *Convergence properties of self-organizing maps*, In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, Artificial Neural Networks, Amsterdam Netherlands Elsevier (1991), pp.409–414
- [3] E. Erwin, K. Obermayer and K. Schulten, *Self-organization maps: stationary states, metastability and convergence rate*, Bio. Cybern., 67 (1992), pp. 35–45.
- [4] E. Erwin, K. Obermayer and K. Schulten, *Self-organization maps: ordering, convergence properties and energy functions*, Bio. Cybern., 67 (1992), pp. 47–55.
- [5] M. Hoshino and Y. Kimura, *Ordered states and probabilistic behavior of self-organizing maps*, Nonlinear Analysis and Optimization (Shimane, 2008), Yokohama Publishers, pp. 31–44
- [6] M. Hoshino and Y. Kimura, *State preserving properties in self-organizing maps with inputs in an inner product space*, Nonlinear Analysis and Convex Analysis (Tokyo, 2009), pp. 65–76.
- [7] T. Kohonen, *Self-Organizing Maps, Third Edition*, Springer, 2001.
- [8] P. L. Zador, *Asymptotic quantization error of continuous signals and the quantization dimension*, IEEE Trans. Inform. Theory, vol. IT-28, No. 2, March (1982), pp. 139–149.