

On self-organizing maps with inputs taking values in inner product space

(内積空間における入力をもつ自己組織化マップの状態保存について)

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. 基本的な自己組織化マップモデル

本報告は Kohonen 型アルゴリズム [6] として知られている自己組織化マップモデルにおける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、モデルの順序化、整列化の形成過程における状態保存性に注目して考察する。

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、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 (model function, reference function) 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) 学習プロセスとして以下の2つを仮定する。

学習プロセス L_A

(a) 学習範囲:

$$I(m_k, x_k) = \left\{ i^* \in I \mid \|m_k(i^*) - x_k\| = \inf_{i \in I} \|m_k(i) - x_k\| \right\} \\ (m_k \in M, x_k \in X), \\ N_1(i) = \{j \in I \mid d(j, i) \leq \varepsilon\} \quad (i \in I).$$

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & \text{if } i \in \bigcup_{i^* \in I(m_k, x_k)} N_\varepsilon(i^*), \\ m_k(i) & \text{if } i \notin \bigcup_{i^* \in I(m_k, x_k)} N_\varepsilon(i^*), \end{cases} \quad k = 0, 1, 2, \dots$$

学習プロセス L_m

(a) 学習範囲:

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

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

$$N_\varepsilon(i) = \{j \in I \mid d(j, i) \leq \varepsilon\} \quad (i \in I).$$

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & \text{if } i \in N_\varepsilon(J(m_k, x_k)), \\ m_k(i) & \text{if } i \notin N_\varepsilon(J(m_k, x_k)), \end{cases} \quad k = 0, 1, 2, \dots$$

上記の2つの学習プロセスは、初期ノードの値によっては、学習反復の初期段階において学習結果に違いが生ずるが、多くの場合、それ以降の反復においては、あまり差が現れない。

2. \mathbb{R} 値ノード, 1次元ノード配列モデル

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。

(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。

(iv) 以下の学習プロセスの一方を仮定する。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:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & \text{if } i \in \bigcup_{i^* \in I(m_k, x_k)} N_1(i^*), \\ m_k(i) & \text{if } i \notin \bigcup_{i^* \in I(m_k, x_k)} N_1(i^*), \end{cases} \quad k = 0, 1, 2, \dots$$

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

(a) 学習範囲:

$$J(m_k, x_k) = \min \left\{ i^* \in I \mid |m_k(i^*) - x_k| = \inf_{i \in I} |m_k(i) - x_k| \right\} \\ (m_k \in M, x_k \in X), \\ N_1(i) = \{j \in I \mid |j - i| \leq 1\} \quad (i \in I).$$

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m_{k+1}(i) = \begin{cases} (1 - \alpha)m_k(i) + \alpha x_k & \text{if } i \in N_1(J(m_k, x_k)), \\ m_k(i) & \text{if } i \notin N_1(J(m_k, x_k)), \end{cases} \quad k = 0, 1, 2, \dots$$

今, n 個のノード $1, 2, \dots, n$ があり, 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. 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, $m_0(1), m_0(2), \dots, m_0(n)$ のなかで x_0 と最も近いものを選び, その値に対応するノード i^* とその周囲のノード i に対して学習

$$m_1(i) = (1 - \alpha)m_0(i) + \alpha x_0$$

が適用され, それ以外のノードに対しては学習が適用されず, $m_1(i) = m_0(i)$ となる. 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, 逐次にノードの更新がおこなわれ, 同時にモデル関数 m_1, m_2, m_3, \dots が逐次に生成される.

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. 実際, 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. また, これらの性質を利用することにより, 多くの実問題へ応用されている.

3. モデルの吸収状態について

次の定理は, 自己組織化マップモデルにおけるモデル関数の単調性保存に関する基本的な結果である.

Theorem 1 学習プロセス L_A を仮定する. モデル関数 m_1, m_2, m_3, \dots 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.

Theorem 2 学習プロセス L_m を仮定する. モデル関数 m_1, m_2, m_3, \dots に関して, 次が成り立つ.

- (i) 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.
- (ii) 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, このような状態を自己組織化マップモデルの吸収状態と呼ぶことにする. モデルの吸収状態として準凸性, 準凹性などがある [5].

4. 内積空間上の値をもつ1次元ノード配列モデルの場合

ここでは, 以下の自己組織化マップモデルを考える.

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

- (i) ノード集合 $I = \{1, 2, \dots, n\}$, $d_I(i, j) = |i - j|$.
- (ii) ノード値空間 V は内積 $\langle \cdot, \cdot \rangle$ をもつ内積空間とする.
- (iii) $x_0, x_1, x_2, \dots \in X \subset V$ は入力列とする.
- (iv) 学習プロセス L_m

(a) 学習範囲:

$$J(m, x) = \min \left\{ i^* \in I \mid \|m(i^*) - x\| = \inf_{i \in I} \|m(i) - x\| \right\}$$

$$(m \in M, x \in X),$$

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

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m'(i) = \begin{cases} (1 - \alpha)m(i) + \alpha x & \text{if } i \in N_1(J(m, x)), \\ m(i) & \text{if } i \notin N_1(J(m, x)), \end{cases} \quad k = 0, 1, 2, \dots$$

Condition $S_{\text{inn}}(i)$ (non-positive inner product property for m)

ノード $i, i+1, i+2$ に対して

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

が成り立つ.

Theorem 3 モデル

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

において, 学習過程 $L_m(\varepsilon = 1)$ を仮定する. m を任意のモデル関数として, x を任意の入力とする. m' を m の入力 x により更新されたモデル関数とする. $i \neq J(m, x) - 3, J(m, x) + 1$ に対して

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

が成り立つならば

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

が成り立つ.

この性質は, [5] における 2次元配列モデルにおける結果と同様の議論により証明される.

更に以下の条件を導入する.

Condition $S_{\text{dis}[+]}(i)$

ノード $i, i+1, i+2$ に対して

$$\|m(i+1) - m(i)\| \leq \|m(i+2) - m(i)\|$$

が成り立つ.

Condition $S_{\text{dis}[-]}(i)$

ノード $i, i-1, i-2$ に対して

$$\|m(i-1) - m(i)\| \leq \|m(i-2) - m(i)\|$$

が成り立つ.

$$\begin{aligned} & \|m(i+2) - m(i)\|^2 - \|m(i+1) - m(i)\|^2 \\ &= \|m(i+2) - m(i+1)\|^2 - 2\langle m(i+2) - m(i+1), m(i) - m(i+1) \rangle \end{aligned}$$

より, 前述の条件の間には以下の関係がある.

Theorem 4

- (i) $S_{inn}(i)$ ならば $S_{dis[+]}(i)$.
- (ii) $S_{inn}(i-2)$ ならば $S_{dis[-]}(i)$.

Theorem 5 モデル

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

において、学習過程 $L_m(\varepsilon = 1)$ を仮定する。 m を任意のモデル関数として、 x を任意の入力とする。 m' を m の入力 x により更新されたモデル関数とする。

- (i) $i \neq J(m, x) - 3, J(m, x) + 1$ のとき、 m に対して $S_{dis[+]}(i)$ が成り立つならば m' に対して $S_{dis[+]}(i)$ が成り立つ。
- (ii) $i \neq J(m, x) - 1, J(m, x) + 3$ のとき、 m に対して $S_{dis[-]}(i)$ が成り立つならば m' に対して $S_{dis[-]}(i)$ が成り立つ。

上記の $S_{dis[+]}(i)$ および $S_{dis[-]}(i)$ が学習により保存される特性は、 [5] における 2次元配列モデルにおける結果と同様の議論により証明される。

5. 1次元ノード配列, \mathbb{R}^2 -値ノードの場合について

ここでは、以下のモデルを仮定する。

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

- (i) ノード集合 $I = \{1, 2, \dots, n\}$. ただし、距離を $d_I(i, j) = |i - j|$ によって定義する。
- (ii) モデル関数 $m : I \rightarrow \mathbb{R}^2$. ここで、各 $x = (x_1, x_2), y = (y_1, y_2)$ に対して、 $\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ とする。
- (iii) 入力 $x_0, x_1, x_2, \dots \in X \subset \mathbb{R}^2$.
- (iv) 学習プロセス L_m (1次元配列, $\varepsilon = 1$)

(a) 学習範囲:

$$J(m, x) = \min \left\{ i^* \in I \mid \|m(i^*) - x\| = \inf_{i \in I} \|m(i) - x\| \right\}$$

$$(m \in M, x \in X),$$

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

(b) 学習率: $0 \leq \alpha \leq 1$.

(c) 更新後の値:

$$m'(i) = \begin{cases} (1 - \alpha)m(i) + \alpha x & \text{if } i \in N_1(J(m, x)), \\ m(i) & \text{if } i \notin N_1(J(m, x)), \end{cases} \quad k = 0, 1, 2, \dots$$

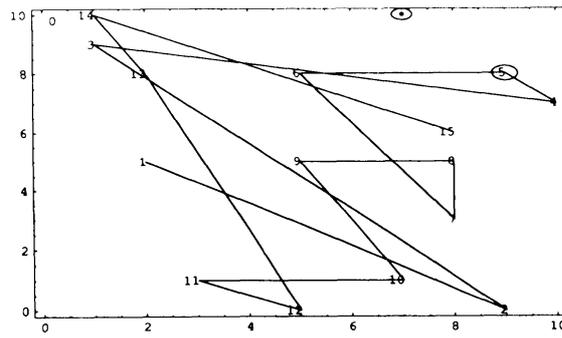


図 1: m_0

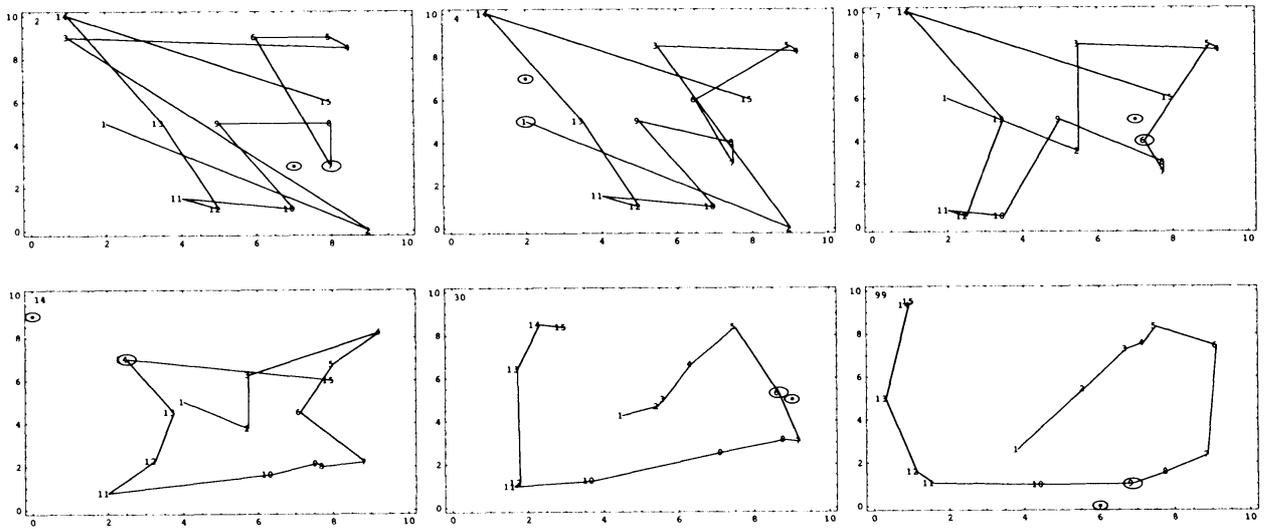


図 2: 左上から右に向かって $m_2, m_4, m_7, m_{14}, m_{30}, m_{99}$

以下の1次元ノード配列, \mathbb{R}^2 -値ノードをもつ自己組織化マップモデルの数値例について考察する.

Example 1 以下の15個のノードをもつ自己組織化マップモデルを考える.

- (1) ノード集合 $I = \{1, 2, 3, \dots, 15\}$.
- (2) 初期モデル関数

$$m_0 = [(2, 5), (9, 0), (1, 9), (10, 7), (9, 8), (5, 8), (8, 3), (8, 5), (5, 5), (7, 1), \\ (3, 1), (5, 0), (2, 8), (1, 10), (8, 6)]$$

各ノードの初期値と初期入力 x_0 を図示すると図1の様になる.

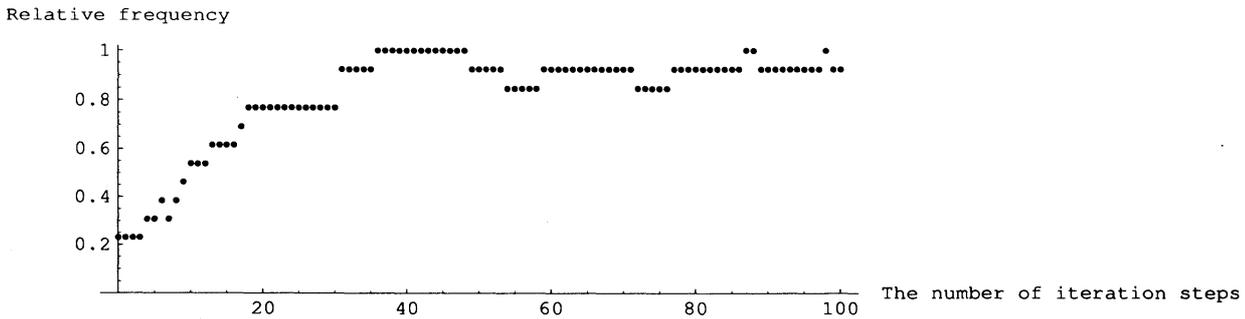


図 3: 条件 $S_{inn}(i)$ の成立頻度

(3) 入力として, $\{0, 1, 2, \dots, 10\} \times \{0, 1, 2, \dots, 10\}$ 上の離散一様分布に従い, 発生させた.

$$x = (7, 10), (5, 2), (7, 3), (10, 8), (2, 7), (0, 0), (8, 2), (7, 5), (8, 3), (4, 4), \\ (10, 1), (8, 2), (6, 4), (7, 2), (0, 9), (7, 2), (6, 6), (4, 10), (1, 3), (2, 8), \dots$$

(4) 学習プロセス L_A を仮定する. (学習率: $\alpha = \frac{1}{2}$)

このとき, 上の学習プロセスにより, モデル関数は以下の様に更新される (図 2).

$$m_1 = [(2, 5), (9, 0), (1, 9), (8.5, 8.5), (8, 9), (6, 9), (8, 3), (8, 5), (5, 5), (7, 1), (3, 1), \\ (5, 0), (2, 8), (1, 10), (8, 6)]$$

$$m_2 = [(2, 5), (9, 0), (1, 9), (8.5, 8.5), (8, 9), (6, 9), (8, 3), (8, 5), (5, 5), (7, 1), (4, 1.5), \\ (5, 1), (3.5, 5), (1, 10), (8, 6)]$$

$$m_3 = [(2, 5), (9, 0), (1, 9), (8.5, 8.5), (8, 9), (6.5, 6), (7.5, 3), (7.5, 4), (5, 5), (7, 1), \\ (4, 1.5), (5, 1), (3.5, 5), (1, 10), (8, 6)]$$

.....

ここで, 上記の学習プロセスに対して, 条件 $S_{inn}(i)$ が成り立つノードの相対頻度

$$\frac{\text{The number of the elements of } \{i \mid \langle m(i) - m(i+1), m(i+2) - m(i+1) \rangle \leq 0\}}{n - 2}$$

(ただし n はノードの個数)

を求めると図 3 のようになる. □

$S_{dis[+]}[m, m', S \rightarrow S]$ ($S_{dis[-]}[m, m', S \rightarrow S]$) をモデル関数 m とその更新されたモデル関数 m' の両方に対して, 条件 $S_{dis[+]}(i)$ ($S_{dis[-]}(i)$) が成り立つようなノード i の個数とし, $S_{dis[+]}[m, m', S \rightarrow N]$ ($S_{dis[-]}[m, m', S \rightarrow N]$) をモデル関数 m に対して $S_{dis[+]}(i)$ ($S_{dis[-]}(i)$)

が成り立つが、その更新されたモデル関数 m' に対しては $S_{\text{dis}[+]}(i)$ ($S_{\text{dis}[-]}(i)$) が成り立たないようなノード i の個数とする。また $S_{\text{dis}[+]}[m, m', N \rightarrow S]$ ($S_{\text{dis}[-]}[m, m', N \rightarrow S]$) をモデル関数 m に対して $S_{\text{dis}[+]}(i)$ ($S_{\text{dis}[-]}(i)$) が成り立たないが、その更新されたモデル関数 m' に対しては $S_{\text{dis}[+]}(i)$ ($S_{\text{dis}[-]}(i)$) が成り立つようなノード i の個数とし、 $S_{\text{dis}[+]}[m, m', N \rightarrow N]$ ($S_{\text{dis}[-]}[m, m', N \rightarrow N]$) をモデル関数 m とその更新されたモデル関数 m' の両方に対して、条件 $S_{\text{dis}[+]}(i)$ ($S_{\text{dis}[-]}(i)$) が成り立たないようなノード i の個数とする。

Example 2 例1のモデルに対して、図4, 5はこれらの値の更新による推移を表したものである。□

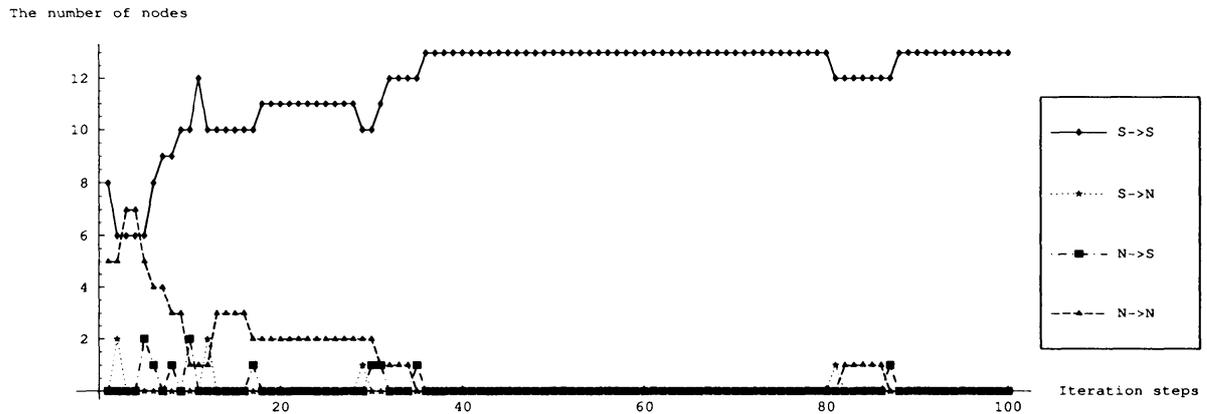


図4: Transitions of $S_{\text{dis}[+]}[m, m', S \rightarrow S]$, $S_{\text{dis}[+]}[m, m', S \rightarrow N]$, $S_{\text{dis}[+]}[m, m', N \rightarrow S]$ and $S_{\text{dis}[+]}[m, m', N \rightarrow N]$

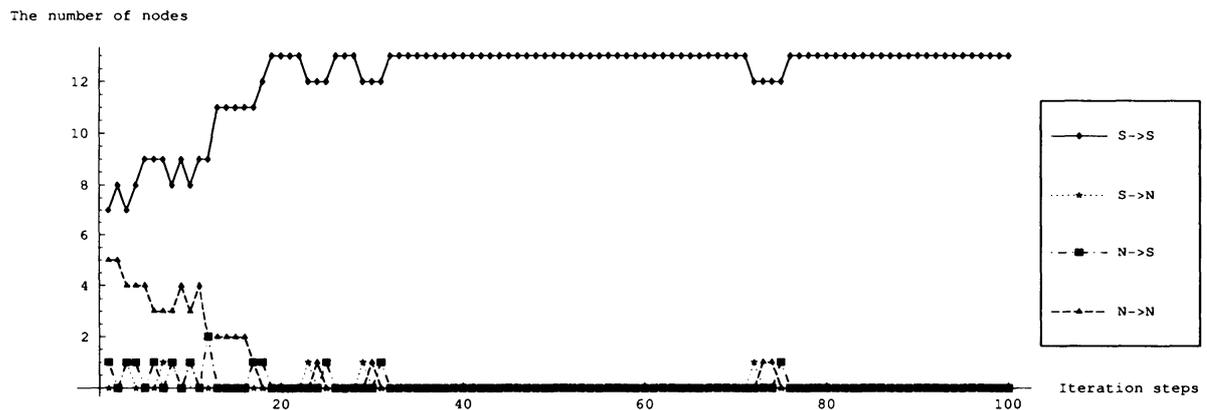


図5: Transitions of $S_{\text{dis}[-]}[m, m', S \rightarrow S]$, $S_{\text{dis}[-]}[m, m', S \rightarrow N]$, $S_{\text{dis}[-]}[m, m', N \rightarrow S]$ and $S_{\text{dis}[-]}[m, m', N \rightarrow N]$

参考文献

- [1] E. Erwin, K. Obermayer and K. Schulten, *Self-organization maps: stationary states, metastability and convergence rate*, Bio. Cybern., 67 (1992), pp. 35–45.
- [2] E. Erwin, K. Obermayer and K. Schulten, *Self-organization maps: ordering, convergence properties and energy functions*, Bio. Cybern., 67 (1992), pp. 47–55.
- [3] W. Fujiwara, E. Itou, M. Hoshino, I. Kaku, A. Sakusabe, M. Sasaki and H. Kosaka, *A study on the effective method of external inspecting using a neural network approach*, Proceedings of 6th ICIM (2002), pp. 369–375
- [4] M. Hoshino, and Y. Kimura, *Absorbing states and quasi-convexity in self-organizing maps*, to appear in *J. Nonlinear Convex Anal.*, Vol.10, No.3 (2009)
- [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] T. Kohonen, *Self-Organizing Maps, Third Edition*, Springer, 2001.
- [7] W. Takahashi, *Nonlinear Functional Analysis*, Yokohama publishers, 2000.
- [8] K. Tanaka, *凸解析と最適化理論*, 牧野書店, 1994.
- [9] 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.