

交換 Monte Carlo 法を用いた Boltzmann Machine の 標本抽出法の提案

柴田 祐樹^{1,†1,a)} 高間 康史^{†1,b)}

概要: 本稿では積層型 Boltzmann Machine (BM) に対し交換 Monte Carlo 法 (EMC: Exchange Monte Carlo) を適用し標本抽出を高精度に行う方法を提案する。これまでに BM の学習法として Contrastive Divergence 法などいくつかの近似解法が提案されているが、一般的な積層型の BM に対し効果を示したものは著者らの確認する限り見当たらない。それら学習法ではいずれも BM の統計量を計算するための標本抽出法として Gibbs Sampling が用いられるが、この手法の積層型モデルに対する精度の低さが学習法の構築を困難にしていると言える。これに対し本稿では、BM と同系統のモデルに対し適用される EMC と、焼きなまし法の結晶化効果を用いた標本抽出法を提案し、人工データセットにより Gibbs Sampling に対する提案手法の有効性を評価した結果を報告する。

1. はじめに

本稿では Boltzmann Machine (BM: Boltzmann Machine) [7] の学習法構築を念頭に置いた新しい標本抽出法を提案する。特に積層された BM (DBM: Deep BM) に対し、非一様温度分布と交換 Monte Carlo (EMC: Exchange Monte Carlo) 法 [4] の併用が有効であることを示す。

近年多くの神経網モデル (Neural Network) に対する学習法、推論法が決定論的で一方向伝搬型として構築されているが [1,2], BM は結合が双方向性で対称的であり、解析的に扱いやすい確率モデルであるため、依然として学習モデルの振る舞いを考える上で有用であると言える。BM に対する学習に対し、勾配降下法 [7], Contrastive Divergence 法 [6], Gibbs Sampling (GS) [8], 平均場近似 [5] などの利用が提案されてきているが、BM の定式化を崩さずに学習、推論を可能とする手法、ならびに、BM の標本抽出を高精度に行える一般的手法は提案されていない。

BM の学習を困難としている要因として、その内部状態が確率的挙動を示すことが挙げられる。BM に対し提案されたどの学習法も、モデルの適合度合いを測る尺度として尤度を用いるが、尤度の計算には BM の期待値を計算する必要がある。この期待値の計算は組み合わせ爆発の問題から厳密に行うことができず、Monte Carlo 法による標本平均

による近似、あるいは平均場近似を用いることが提案されている。しかしながら、平均場近似ではモデルの正しい分布を捉えることができない。Monte Carlo 法は計算量の増加を許容すれば標本点を増やすことで精度を上げることができるが、計算量以外にも、DBM では積層数が増加するにつれ両端の状態が多数の中間層の確率的挙動による攪乱を受け、互いに独立に振る舞うようになり、標本抽出が困難になるという問題がある。文献 [12] ではこれに対し、温度を非一様に低下させることで DBM の端から端へ状態を伝達する手法を提案しているが、温度の変化は学習過程では母数^{*1}へ吸収されるため、学習法へこの方法をそのまま適用することはできない。

標本抽出を行う Monte Carlo 法の反復中に異なる固定温度による複数のモデルを用いて標本抽出の精度を向上させる手法として、EMC、または並列攪乱法 (PT: Parallel Tempering) [3] と呼ばれる手法がある。EMC は標本を抽出したいモデルと近い母数を持つモデルを複数用意し、それらモデルの状態を一定の規準で確率的に交換することで、それぞれの分布で得意とされる状態遷移を組み合わせ、標本抽出精度を向上させることが可能である。似た手法として焼きなまし法 (SA: Simulated Annealing) [10] では温度を高い状態から低い状態へ局所最適化を行う間徐々に変化させることで大域的最適解の探索を行うが、この方法は一つの最適解を求めることを目的としており、確率モデルの分布を近

¹ 情報処理学会
IPSI, Chiyoda, Tokyo 101-0062, Japan

^{†1} 現在, Tokyo Metropolitan University

^{a)} hshibata@tmu.ac.jp

^{b)} ytakama@tmu.ac.jp

^{*1} 本稿では確率モデルで標本抽出の対象となるものを状態変数、分布を特徴づける変数を母数、解法の挙動を定めるものを制御変数 (Hyper-parameter) と呼ぶ。

似することには適していない。これに対し EMC では、分布の近似が可能である。

本稿では、前述した EMC と、文献 [12] で用いられる温度分布を組み合わせた手法を提案する。評価実験では人工モデルから標本を抽出し、その統計量と理論的に得られる統計量を理論値と比較し、従来手法に比べ多くの状況で優れた結果を示すことを報告する。

1.1 記法と用語

本稿で利用する記法について述べる。標本空間を Ω と表し、その上に定義される確率変数 \mathbf{x} に対してその状態空間を $\mathbf{x}(\Omega)$ とする。また、 $p(\mathbf{x} = x)$ は \mathbf{x} が $x \in \mathbf{x}(\Omega)$ を取る確率を示し、期待値を $\langle \mathbf{x} \rangle = \sum_{x \in \mathbf{x}(\Omega)} p(\mathbf{x} = x)$ と表す。また $\mathbf{x}(\Omega) = \mathbf{v}(\Omega)$ である任意の 2 つの確率変数 \mathbf{x}, \mathbf{v} について $p(\mathbf{x}) = p(\mathbf{v})$ と書いた場合 $\forall x \in \mathbf{x}(\Omega), p(\mathbf{x} = x) = p(\mathbf{v} = x)$ であることを意味する。

文献 [9] で提案されている確率変数に対する添字記法を用いる。この記法において、多次元確率変数のそれぞれの要素を識別するために添字集合が用いられる。ある添字集合 Λ で添字づけられる確率変数の族を $\mathbf{x}_\Lambda = (\mathbf{x}_i | i \in \Lambda)$ などと表す。 Λ の任意の部分集合 $M \subset \Lambda$ に対してその補集合を $M^c = \Lambda \setminus M$ と定義する。ただし、 \setminus は差集合を意味する。このとき、 $p(\mathbf{x}_\Lambda) = p(\mathbf{x}_M, \mathbf{x}_{M^c}) = p(\mathbf{x}_M | \mathbf{x}_{M^c}) p(\mathbf{x}_{M^c})$ が成り立つとする。添字集合が明らかなき場合は省略し、 \mathbf{x}_Λ を単に \mathbf{x} などと表す*2。

任意の x, y に対し、 $\delta(x, y)$ は Kronecker delta を表す。その他集合論の記法と概念について [15] に従う。

2. 関連研究

2.1 Boltzmann Machine

BM は Boltzmann 分布に従う確率モデルの一つである。各変数を識別する添字の集合を Λ 、その状態を表す確率変数を \mathbf{x} 、エネルギー関数を $\phi: \mathbf{x}(\Omega) \rightarrow \mathbb{R}$ 、温度を T としたとき、Boltzmann 分布は式 (1) で定義される。

$$p(\mathbf{x}_\Lambda) = \frac{1}{Z} \exp\left(-\frac{1}{T} \phi(\mathbf{x}_\Lambda)\right) \quad (1)$$

ここで、 Z は正規化定数である。 $E \subset \Lambda^2$ を結合を表す添字の対、 $w: \Lambda^2 \rightarrow \mathbb{R}$ を結合の重み、 $b: \Lambda \rightarrow \mathbb{R}$ を発火率の偏りとしたとき、BM における ϕ は式 (2) で定義される。

$$\phi(\mathbf{x}_\Lambda) = - \sum_{(i,j) \in E} \mathbf{x}_i w_{i,j} \mathbf{x}_j - \sum_{i \in \Lambda} b_i \mathbf{x}_i \quad (2)$$

また、学習中に温度の違いは吸収されるため BM では通常 $T = 1$ とする。状態空間について $\mathbf{x}(\Omega) = \{0, 1\}^\Lambda$ としたものは Binary BM と呼ばれ、多くの研究がこのモデルについてなされている [5-7]。

*2 文献 [9] では添字のみで確率空間を区別するとあるが、本稿では確率変数の記号によっても確率空間を区別するものとする。

データの表現を行う確率変数を \mathbf{v}_Λ としたとき、母数 w, b を勾配降下法で学習するための勾配は次のように書ける。ここでは w についてのみ記載する。

$$\frac{\partial S}{\partial w_{i,j}} = \langle \mathbf{v}_i \mathbf{v}_j \rangle - \langle \mathbf{x}_i \mathbf{x}_j \rangle, (i, j) \in E \quad (3)$$

ただし、 Λ のうち隠れ層でないものを V 、データ分布を $p(\mathbf{v}_V)$ としたとき、 $p(\mathbf{v}_\Lambda) = p(\mathbf{v}_V) p(\mathbf{x}_{V^c} = \mathbf{v}_{V^c} | \mathbf{x}_V = \mathbf{v}_V)$ であるとする。また、 S は負の対数尤度である [12]。

E について、モデルが積層構造を取るように $\Lambda = \sum_{k=1}^K H_k, E = \sum_{k=1}^{K-1} H_k \times H_{k+1}$ ($H \in \mathfrak{P}(\Lambda)^K$) と結合に制限を加えたものを制限 (Restricted) Boltzmann Machine (RBM) と呼び、特に $\exists k, (H_k + H_{k+1}) \cap V = \emptyset$ であるものを ($K \geq 3$ が必要) を DBM と呼ぶ。

式 (3) 中の期待値の厳密計算は困難であるため、近似手法が提案されており、代表的なものに Markov Chain Monte Carlo (MCMC) 法 [11] および Gibbs Sampling (GS) 法 [8]、平均場近似 [5] がある。また MCMC 法の発展として RBM に対し EMC [3, 4] を用いることも提案されている。DBM に対する EMC の適用事例は著者らの確認する限り報告されていない。また、4 層以上の DBM に対する GS の適用事例も確認できないが、これについては評価実験でも示すとおり、DBM に対し GS が有効ではないことが原因であると考えられる。本稿ではこれらのうち MCMC, GS, EMC についてそれぞれ紹介する。

2.2 Monte Carlo 法

Monte Carlo (MC) 法は統計量の厳密算出が困難な分布に対し、標本を用いて統計量を近似する手法である。ここでは特に BM から標本を求めるためによく使われる MCMC 法について説明を行う。

標本を抽出したい元の分布を $p(\mathbf{x})$ としたとき、MCMC では確率変数の系列 $\mathbf{z}_1, \mathbf{z}_2, \dots$ と、標本の抽出が容易な遷移確率 $p(\mathbf{z}_n | \mathbf{z}_{n-1})$ を用い、元の分布を段階的に近似することを考える。このとき、 $p(\mathbf{z}_n | \mathbf{z}_{n-1})$ が式 (4) を満たすとき、系列は元の分布に従う。

$$p(\mathbf{x} = x) = \sum_{x' \in \mathbf{x}(\Omega)} p(\mathbf{z}_n = x | \mathbf{z}_{n-1} = x') p(\mathbf{x} = x'), \quad \forall x \in \mathbf{x}(\Omega), n \in \mathbb{N} \quad (4)$$

この条件は釣り合い条件 (Balance condition) と呼ばれる [11, 13]。有限の標本で系列を近似する場合、初期の分布が $p(\mathbf{x})$ からずれている場合、系列が収束するまでの計算量が増えるが、式 (4) が満たされている限り、元の分布に収束する。

式 (4) を満たす遷移確率として、Metropolis 法と呼ばれる以下のものが提案されている [11]。

$$p(\mathbf{z}_{i+1} = x' | \mathbf{z}_i = x) = A(x', x)q(\mathbf{z}_{i+1} = x' | \mathbf{z}_i = x) + C(x)\delta(x, x') \quad (5)$$

$$A(x', x) = \min\left(1, \frac{p(\mathbf{x} = x')}{p(\mathbf{x} = x)}\right) \quad (6)$$

q は提案分布と呼ばれ、状態の遷移を試みるための遷移確率であり、対称性条件 $\forall x', x, q(\mathbf{z}_{i+1} = x | \mathbf{z}_i = x') = q(\mathbf{z}_{i+1} = x' | \mathbf{z}_i = x)$ を満たす。 q で提案された遷移は、 A により受理するか判断される。 A は Metropolis 規準と呼ばれ、MCMC 法を構築するために広く用いられている [3, 4, 11]。 $C(x)$ は遷移が棄却された場合に元の状態を保つ確率である*3。以上の漸化式を計算し得られた系列から、適当な間隔で標本を抽出することを繰り返し、統計量を計算する。

C は小さければ計算の無駄が少ない。これを小さく、あるいは 0 にする方法が研究されている [8, 13]。GS [8] は部分空間毎に標本を抽出することにより、Metropolis 規準を用いず式 (4) を満たす遷移確率を実現している。GS を適用するには、元の分布の条件付き分布が計算できる必要が有る。条件付き分布が互いに独立に計算可能な場合について計算を並列化することができるという利点もある。

2.3 交換 Monte Carlo 法

EMC [4] は単一の確率変数の系列を用いる通常の MCMC に加え、複数の系列を用意し、各系列を求めながら互いの状態を交換し合うことで、それぞれの系列の特徴を相互に活かしながら標本抽出の質を向上させる手法である。それぞれの系列が近似するモデルは複製と呼ばれ、その分布は元のモデルと母数が異なる。複製間の状態の交換は、式 (5) に基づき行われる。以降上述の定式化を行う。各複製を識別する添字の集合を L 、全複製の状態を記述する確率変数を \mathbf{x}_L としたとき、この分布は次の通り定義される。

$$p(\mathbf{x}_L) = \prod_{l \in L} p(\mathbf{x}_l)$$

ここで、 $p(\mathbf{x}_l)$ は l で識別されるモデルを表し、2 つのモデル $l, l' \in L$ について、 $l \neq l'$ であれば互いに母数は異なる。

\mathbf{x}_L を近似する系列を $\mathbf{z}_1, \mathbf{z}_2, \dots$ としたとき、EMC では次のとおり定義される x から x' への状態変化を考える。

$$x' = (\dots, x_b, \dots, x_a, \dots), x = (\dots, x_a, \dots, x_b, \dots)$$

ただし、 $a, b \in L$ である。この変化では、モデル a とモデル b の状態の入れ替えのみを試みている。この変化を与える提案分布は式 (7) の通りである。

$$q(\mathbf{z}_{i+1} = x' | \mathbf{z}_i = x) = \delta(x'_b, x_a) \delta(x'_a, x_b) \prod_{l \neq a, b} \delta(x'_l, x_l) \quad (7)$$

この提案分布を式 (5) へ代入し遷移確率を得る。式 (6) の分数部分は、 $l \neq a, b$ について因数が約分されるため、式 (8)

*3 式 (5) の両辺の総和を取れば計算可能である。

のようになる。

$$\frac{p(\mathbf{x}_L = x')}{p(\mathbf{x}_L = x)} = \frac{p(\mathbf{x}_a = x_b) p(\mathbf{x}_b = x_a)}{p(\mathbf{x}_a = x_a) p(\mathbf{x}_b = x_b)} \quad (8)$$

本手法は複数の温度を持つモデルを利用可能な点で SA と同様の利点を持つが、温度は固定で状態の交換を行う点で異なる。また、各系列の計算は並列化可能である。標本は元の分布 $p(\mathbf{x}_*)$ に対し、 $p(\mathbf{x}_l) = p(\mathbf{x}_*)$ となる l を用意し、ここから抽出する。

2.4 焼きなまし法と結晶化効果

焼きなまし法 (SA) は、最適化問題中の説明変数で定義された目的関数の最小値を求めるために、Boltzmann 分布と MCMC を利用した手法である [10]。

SA では式 (1) の ϕ に対し、目的関数を設定する。式 (1) の分布は $T \rightarrow 0$ としたときに $\min(\phi(\mathbf{x}))$ を与える \mathbf{x} で高い確率となるため、 T を徐々に 0 に近づけながら MCMC による分布の近似を続けていくことで、実用的な最適解の近似解を得ることが可能である。本手法は最適値を求める手法であり、統計量を近似するための標本抽出を行うことには向かない。

SA は確率分布の最頻値の推定にも利用可能であるが、DBM に対して効果がないと報告されている [12]。また、SA で従来より用いられる全ての状態変数に対し一様な温度を用いる温度低下過程ではなく、入力層など情報発信の起点となる状態変数から順に温度を低下させるといった、非一様な温度分布を用いた温度低下過程を用いることで、最頻値の推定精度を改善できることが報告されている [12]。以降この手法と、この手法を用いるために必要な BM の定式化を説明する。

2.1 節で紹介したのと同様に、 Λ をモデルの各要素を定義する添字の集合、 \mathbf{x}_Λ を Λ で定義される確率変数としたとき、文献 [12] で用いられる BM の分布を次のように書ける。

$$p(\mathbf{x}_\Lambda) = \frac{1}{Z} \exp(-\phi'(\mathbf{x}_\Lambda, T)) \quad (9)$$

Z は正規化定数で、 T は温度であるが SA と異なり、このモデルでは温度が非一様に分布する。 ϕ' は修正エネルギー関数であり*4、次の通り定義される。

$$\phi'(\mathbf{x}_\Lambda, T) = - \sum_{(i,j) \in E} \frac{1}{T_{i,j}} \mathbf{x}_i w_{i,j} \mathbf{x}_j - \sum_{i \in \Lambda} \frac{1}{T_i} \mathbf{x}_i b_i \quad (10)$$

ここで、 E, w, b の定義は式 (2) と同様であるが、 $i (i \in \Lambda)$ 毎に異なる温度 T_i が新たに導入されている。また、 $T_{i,j}$ は $i, j (i, j \in \Lambda)$ 間の温度を示し、 $T_{i,j} = 2T_i T_j / (T_i + T_j)$ で定義される。このとき、 \mathbf{x}_Λ の標本を抽出するために GS で用いる条件付き分布は、式 (11) のように書ける。

*4 文献 [12] とは同等であるが表記上において異なる定式化である。本稿では簡素化のためにこちらを用いる

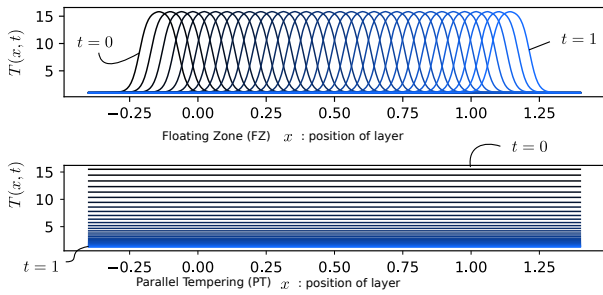


図1 式(14)と式(15)それぞれについて $T_1 = 16, T_0 = 1, \beta = 0.01, \alpha = 0.99, K = 8$ の条件で描かれる曲線の例。

$$p(\mathbf{x}_i = \mathbf{1} | \mathbf{x}_{\Lambda \setminus \{i\}}) = \sigma \left\{ \frac{1}{T_i} b_i + \sum_{(i,j) \in E} \frac{1}{T_{i,j}} w_{i,j} \mathbf{x}_j \right\} \quad (11)$$

ここで、 $\sigma(x) = 1/(1 + \exp(-x))$ は Sigmoid 関数であり、全ての $i \in \Lambda$ について順番に GS を適用しながら系列の更新を行う。式(11)中の条件付き分布はこれら既存研究 [3, 5, 6] で用いられているものと類似した形をしているが、温度が含まれている点で異なる。

この温度の項により、より低い温度の変数の状態が、他の項の分布へ強い影響を与え、それにより、情報の伝搬が温度の低い方から高い方へ制限されることで最頻値の推定精度を上げることが可能であると報告されている [12]。これは材料科学における Floating Zone (FZ) [14] 法により Silicon 単結晶を生成可能なことを例として定性的に説明される。本稿ではこの作用を結晶化効果と呼ぶ。この手法で推論精度の向上は期待できるが、文献 [12] の適用方法では SA と同様の理由で学習法に応用することはできず、また分布の推定に用いることもできない。

3. 提案手法

本稿では、DBM に対し有効な EMC と SA の結晶化効果を用いた標本抽出法を提案する。結晶化効果を利用して DBM 中に比較的大きな結晶粒を生成させる、つまりエネルギー関数の値が低く生起確率の高い標本を優先的に生成させることで標本抽出法の高精度化を狙う。また、異なる固定温度分布を持つ複製のモデルによる EMC を用いることで、SA を用いた場合と異なり分布の近似を可能とする。

提案手法では温度分布を表現するために式(9)を確率分布として採用する。GS を用いて系列を更新する遷移確率を p_u 、状態の交換が起こる遷移確率を p_e 、 p_e による遷移が起こる確率を $c_e (\in [0, 1])$ として、提案手法では以下の遷移確率を用いる。

$$p(\mathbf{z}_k | \mathbf{z}_{k-1}) = c_e p_e(\mathbf{z}_k | \mathbf{z}_{k-1}) + (1 - c_e) p_u(\mathbf{z}_k | \mathbf{z}_{k-1}) \quad (12)$$

c_e は制御変数である。 p_e が選ばれた場合、すべての $l, l' \in L$ について交換を試みるものとする*5。

*5 実装に当たっては状態を入れ替える代わりに、温度分布を入れ替えたほうが記憶域の操作が少なくて済む。

3.1 温度制御

温度について、文献 [12] で提案されている結晶化作用のある温度分布は複雑で一般性が低いため、本稿ではこれを改良した以下の分布を用いる。

$$T(t, k^*) = (T_1 - T_0 + a_f(T_1 - 2T_0))\sigma(-f(k^*, t)) + T_0 \quad (13)$$

ここで、 $t \in [0, 1]$ は制御変数であり、SA では進行度合いに当たるものである。EMC では、温度分布を計算の進捗に合わせて変化させることはしないが、 t を複製毎に変化させることで温度分布の異なるモデルを作成する。 k^* は正規化された層番号にあたり、 $k = 1, 2, \dots, K$ を層番号とするならば、 $k^* = (k - 1)/(K - 1)$ である。 T_1 と T_0 はそれぞれ温度分布の境界値 (最大値、もしくは最小値) に相当する。 $f(k^*, t)$ については、以下の2通りを用いる。

$$f(k^*, t) = c_1 |(1 + 2c_2)t - c_2 - k^*| + c_3, \quad (14)$$

$$f(k^*, t) = -c_3 t \quad (15)$$

a_f は、 f が式(15)で与えられる場合に1、式(14)で与えられる場合に0を取る。式(15)の方は一様温度分布を与えるもので、結晶化作用の効果をみるための比較用である。定数は次の通り定められる： $c_1 = 2(N_L - 1) \ln(1 - \alpha)$ 、 $c_2 = -2c_3/c_1$ 、 $c_3 = \ln(1 - \beta_2)/\beta_2$ 。その他は制御変数であり、 $\alpha (\in (0, 1))$ は分布の曲率を決定し、 β_2 は最大値と最小値の再現精度を決める制御変数である。具体的には $\sigma(-f(1, 1)) = \beta_2$ 、 $\sigma(-f(0, 0)) = 1 - \beta_2$ となるためこれを制御する。図1に曲線の例を示す。

本稿での提案は、式(14)を用いることである。これに対応する温度分布は図1の上部に示されている。この温度分布はケイ素棒を端から端へ順に溶かして全体を単結晶とする、結晶化作用を有する FZ 法の温度分布を模擬している。図中下部は式(15)を用いたもので、提案手法の有効性を示すための比較用分布である。

4. 評価実験

それぞれの手法について、人工モデルから標本を抽出する精度を評価する。用意するモデルは統計量の理論値を得られ、積層モデルで特に標本抽出が困難と考えられる遠距離変数間での共起が、発火模様から明確にわかるモデルとしている。比較を行う手法は式(15)の一様温度分布を用いたもの (PT)、提案する式(14)の温度分布を用いたもの (FZ)、また、単に GS を用いたもの (GS) の3つである。

4.1 人工データセットと評価指標

正解として、次の設定を行った DBM を用意する。

$$w_{i,j} = \begin{cases} w_0, & v(i) = v(j) \\ -w_0/m, & v(i) \neq v(j) \end{cases}, b_i = 0, (i, j) \in E \quad (16)$$

ここで、 $v(i)$ は i 要素の層内での位置を示す。全ての層は

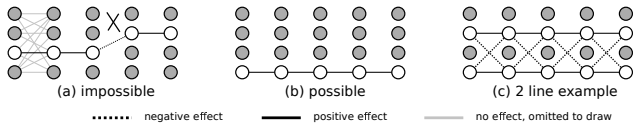


図2 試験用模型と、食い違いがあり生成されない発火模様 (a), 生成される線状発火模様 (b, c) の例。

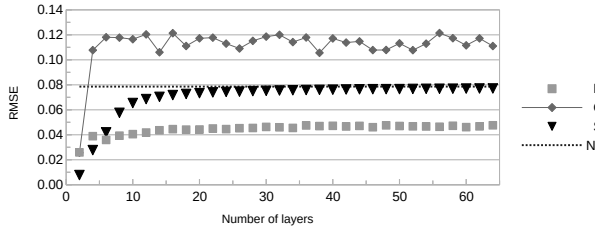


図3 様々な $K = 2, 4, \dots, 64$ に対する各手法の $\tau = 1$ における Root Means Square Error の比較。

同じ数の変数からなる。この模型では、 w_0 が十分に大きい場合に、図 2(b), (c) に示すように、一連の長い発火模様となる状態のみが生成されるようになる。もし図 2(a) の例のように、途中で食い違いが有る場合式 (16) の設定ではエネルギー関数の値が大きくなることからそのような状態が生成される確率は低くなり、結果としてほとんど、もしくは w_0 が十分に大きければ実質的に生成されない。

いま、 K を層の数とし、発火する線の本数を n 、その状態を x としたとき、エネルギー関数は以下の値を取る。

$$\phi(x) = -(K-1)nw_0 + (K-1)n(n-1)w_0/m$$

上式の ϕ の最小値を与える n を n^* としたとき、 $n^* = (m+1)/2$ と求まる。このことから、 m が特に整数で奇数のときは $(m+1)/2$ 本、 m が偶数のときは $m/2$ あるいは $(m+2)/2$ の線が同時に現れることになる。

この人工模型から得られる比較用の分散共分散を算出する。 n^* と層あたりの状態変数の数 H が与えられたとき、食い違い模様が生成される可能性を無視すれば、分散共分散 $\sigma_{i,j} = \langle \mathbf{x}_i \mathbf{x}_j \rangle - \langle \mathbf{x}_i \rangle \langle \mathbf{x}_j \rangle$ は以下の通り求まる。

$$\sigma_{i,j} = \left(\frac{H-2+\delta_{i,j}}{n^*-2+\delta_{i,j}} \right) \left(\frac{H}{n^*} \right)^{-1} - \frac{(n^*)^2}{H^2}, \quad i, j \in \Lambda$$

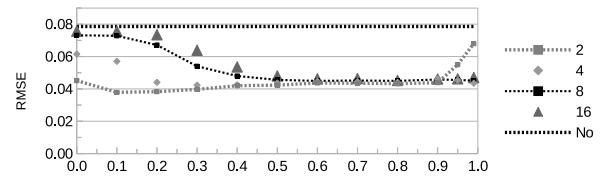
ただし、 $()$ は選択の組み合わせの数を示す^{*6}。各手法により得られる標本で計算した標本分散を $\hat{\sigma}_{i,j}$ として、これに対し評価実験では以下に示される Root Mean Square Error (RMSE) を評価尺度として用いる。

$$\text{RMSE} = \sqrt{\frac{\sum_{i,j \in \Lambda} (\sigma_{i,j} - \hat{\sigma}_{i,j})^2}{|\Lambda|^2}}$$

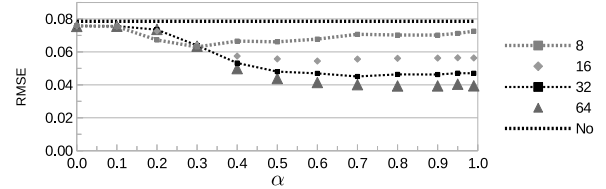
4.2 設定値

本節では評価実験で共通して用いる設定について説明する。常に $\beta_2 = 0.01$, $w_0 = 16$, $c_e = 0.4$ とする。これらの設

^{*6} $\sigma_{i,j}$ は i, j に依存しない。



(a): α と T_1 に関する変化



(b): α と $|L|$ に関する変化

図4 $\tau = 1$ における、 $\alpha = 0, 0.1, \dots, 0.9, 0.95, 0.99$ について FZ の Root Means Square Error の変化。横軸は α を示す。

定値は事前実験より特に実験結果に他の制御変数と比べ有意な影響を与えないことを確認している。 $H = 9$ と $m = 5$ についても共通して用いられるが、これらは状態空間の大きさ（ここでは 84）が複製の数を超える、つまり複製が全状態を記憶することがないように設定されている。

PT は式 (15)、GS は $T_0 = T_1 = 1$ とした式 (15)、FZ は式 (14) を様々な設定値で用い、PT と FZ には EMC を用いる。EMC で使うそれぞれの複製の温度分布は、式 (13) 中の t を複製毎に変えることで設定を行う。複製の数を $|L|$ 、複製の番号を自然数で $l = 1, 2, \dots, |L|$ としたとき、 $t = (l-1)/(|L|-1)$ と計算される。以上の複製の他に元の模型（元模型）が用意されるため、EMC で用いる複製の数は $|L|+1$ であり、交換が起こりうる組み合わせの数は $2|L|-1$ である。元模型の母数は PT, FZ とともに GS で用いられるものと同じであり、統計量を算出するための標本はこの元模型からのみ抽出される。

実験結果の説明において、 $\tau \in [0, 1]$ は正規化された系列の進展度合いを示す。系列の更新は $\tau = 1$ までの間に 2^{17} 回行われる。図中 No と示される破線は RMSE の基準値を示し、これは共分散行列が独立に一様分布する状態変数より算出されたものである。つまり各状態に相関はないため、 $\forall i, j \in \Lambda, \hat{\sigma}_{i,j} = 0$ となる。

その他の設定値について、特に断りがなければ、 $|L| = 32$, $K = 32$, $\alpha = 0.9$, $T_1 = 16$, $T_0 = 1$ が用いられる。全ての RMSE は 32 回の実験の平均値である。

4.3 実験結果

図 3 に、FZ, PT, GS それぞれについて層の数を $K = 2, 4, \dots, 64$ と変化させた場合の RMSE を示す。この結果より、層の数が多くなるにつれ、FZ と PT では差が開いていくことがわかり、GS は $K = 2$ の場合でしか有効に機能していないことがわかる。PT が $K = 2, 4$ の場合において FZ を上回っている。 $K = 2$ (狭義の RBM) に対して PT が機能することが報告されているが [3]、この結果からもその

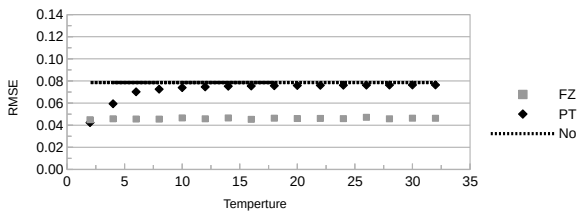


図5 様々な $T_1 = 2, 4, \dots, 32$ に対する各手法の $\tau = 1$ における Root Means Square Error の比較.

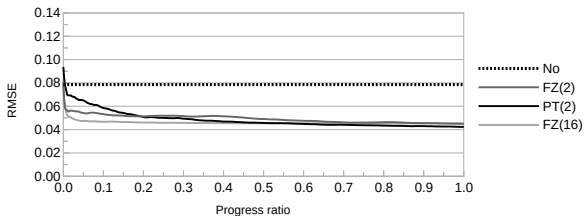


図6 Root Means Square Error の比較. PT は $T_1 = 2$ について, FZ は $T_1 = 2, 16$ について実験を行った.

ことを確認できる.

図4 は $\alpha = 0, 0.1, 0.2, \dots, 0.9, 0.95, 0.99$ について, (a) は $T_1 = 2, 4, 8, 16$, (b) は $|L| = 8, 16, 32, 64$ における FZ の RMSE の変化を示す. この結果より, $\alpha = 0.7$ 程度であれば, 複製の数 ($|L|$) に対し単調に性能は改善し, 温度は結果に対して大きな影響を与えないことがわかる.

図5 に $T_1 = 2, 4, \dots, 32, \tau = 1$ における FZ と PT の RMSE の変化を示す. この結果より, FZ は温度の値域の変化に大きく影響を受けないことがわかり, 一方で PT は影響を受けることがわかる.

図5 の結果では温度が $T_1 = 2$ の場合について PT が FZ を上回るが, これについてより詳細に調べるために, RMSE の収束速度を更新回数 τ を横軸にとり履歴を描いた結果を図6 示す. この結果より, $\tau = 1$ において PT が FZ を僅かに上回ると言えるが, 大きな差は見られない. 一方で FZ は PT にくらべ収束速度が速いことがわかる. さらに, $T_1 = 16$ の設定では同等の RMSE でより速く収束していることがわかる.

5. おわりに

本稿では, DBM の標本抽出法として, SA の結晶化効果と EMC を組み合わせた手法を提案した. 評価実験より, 提案手法は高い収束速度をもち, 設定値に対し敏感でなく, また, ほとんどの場合で比較手法に比べ高い性能を持つことを示した. さらに, 温度分布を一様とした GS や PT では, 積層数が多くなるに連れて高精度な標本の抽出が困難になっていことから, 実際に提案手法が持つと考えられる結晶化効果が DBM に対して有効であると考えられる. 設定に対し敏感でないことは, 母数の適用範囲が広いことを意味するため, 特に母数の変動が大きい学習初期において PT よりも提案手法である FZ が標本抽出法として適してい

るのではないかと考える.

評価実験はここで用意された人工模型についてのみ行われたもので, 結果から得られる結論の一般性や実用性は保証されない. しかしながら, この人工模型を用いた評価方法の一つを提案したことは本研究の貢献の一つであると考えられる. 今後の展望として, 学習法への応用や, 本稿で用いたデータセット以外に対する適合性の調査が挙げられる.

参考文献

- [1] Kaiming, H., Zhang, X., Ren, S. and Jian, S.: *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016).
- [2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Azair, S., Courville, A. and Bengio, Y.: *Generative Adversarial Nets*, Advances in Neural Information Processing Systems 27, 9 pages (2014).
- [3] Cho, K., Raiko, T. and Ilin, A.: *Parallel Tempering is Efficient for Learning Restricted Boltzmann Machines*, Proceedings of the 2010 International Joint Conference on Neural Networks, 8 pages (2010).
- [4] Hukushima. K. and Nemoto. K.: *Exchange Monte Carlo Method and Application to Spin Glass Simulations*, Journal of the Physical Society of Japan, vol. 65, no. 6, pp. 1604–1608 (1995).
- [5] Salakhutdinov, R. and Hinton, G.: *Deep Boltzmann Machines*, Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, pp. 448–445 (2009).
- [6] Tieleman, T.: *Training Restricted Boltzmann Machines Using Approximations to the Likelihood Gradient*, Proceedings of the 25th International Conference on Machine Learning, pp. 1064–1071 (2008).
- [7] Ackley, D. and Hinton, G.: *A Learning Algorithm for Boltzmann Machines*, Cognitive Science, vol. 1, no. 1, pp. 147–169 (1985).
- [8] Geman, S. and Geman, D.: *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721–741 (1984).
- [9] Shibata, H. and Takama, Y.: *Index-Based Notation for Random Variable and Probability Space*, Advanced Computational Intelligence and Intelligent Informatics, vol. 23, no. 4, pp. 715–718 (2019).
- [10] Kirkpatrick, S., Gelatt, C. and Vecchi, M.: *Optimization by Simulated Annealing*, Science, vol. 220, no. 4598, pp. 671–680 (1983).
- [11] Metropolis, N., Rosenbluth, A., Rosenbluth, M. and Teller, A.: *Equation of State Calculations by Fast Computing Machines*, The Journal of Chemical Physics, vol. 21, no. 6, pp. 1087–1092 (1953).
- [12] Shibata, H., Ishikawa, H. and Takama, Y.: *Crystalizing Effect of Simulated Annealing on Boltzmann Machine*, Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 23, no. 3 (2019).
- [13] Sohl-Dickstein, J., Mudigonda, M. and DeWeese, M.: *Hamiltonian Monte Carlo Without Detailed Balance*, Proceedings of the 31 st International Conference on Machine Learning, 8 pages (2014).
- [14] Kaiser, T. and Benz, W.: *Floating-zone growth of silicon in magnetic fields 3, Numerical Simulation*, Journal of Crystal Growth, vol. 230, pp. 164–171 (2001).
- [15] 松坂和夫, 集合・位相入門, 岩波書店, 新装版第2刷 (1968).