

変分ベイズ法における確定的アニーリングと ハイパーパラメータの部分最適化について

On Deterministic Annealing and Partial Optimization of Hyperparameters for Variational Bayes Algorithm

永田賢二* 片平健太郎† 岡ノ谷一夫‡ 岡田真人§
Kenji Nagata Kentaro Katahira Kazuo Okanoya Masato Okada

Abstract: Variational Bayes (VB) algorithm is widely used as an approximation method of Bayesian learning. In a recent study, the deterministic annealing VB algorithm has been proposed to overcome the local optimal problem. In this study, we propose a new deterministic annealing method and a partial optimization method of hyperparameter for VB algorithm by introducing two types of temperature parameters to variational free energy. We also apply the proposed methods to Gaussian mixture model to show the effectiveness of the proposed methods.

Keywords: Variational Bayes algorithm, Deterministic Annealing, Optimization of Hyperparameter, Gaussian Mixture Model

1 まえがき

混合正規分布や隠れマルコフモデルなどの、階層構造をもつ確率モデルにおいて、ベイズ学習が、最尤推定法や事後確率最大化法よりも、汎化性能に優れることが知られている [11]。ベイズ学習を計算する際に、複雑な多重積分を計算することが必要になり、この操作の困難さから、様々な近似手法が提案されている。その中でも、変分ベイズ (VB) 法は、Expectation Maximization (EM) アルゴリズムと同様の反復計算により事後分布についての積分計算ができ、その計算量の少なさから、広く用いられている [1]。また、理論的にも、確率的複雑

さの漸近挙動を解析することで、VB 法がベイズ学習の良い近似を与えることが明らかにされている [10]。

しかしながら、VB 法のアルゴリズムは勾配型であるため、得られた解が局所最適解にトラップされてしまうといった問題がある。この問題に対して、変分ベイズ法における確定的アニーリングや量子アニーリングが提案され、その有効性が確認されている [5][8]。これらの手法では、最適化の対象となる事後分布に温度パラメータを導入し、温度を用いてアニーリングスケジュールを制御することで局所最適解の問題に対処している。

一方で、事後分布への温度パラメータの導入については、変分ベイズ法に限らず、様々な応用で用いられている。事後分布は尤度関数と事前分布の積により構成されているため、部分的に温度パラメータを導入するアプローチが提案されている。その一例として、Hara らは、LDPC 符号における復号のダイナミクスを観測する際に、事前確率に温度パラメータを導入することで交換モンテカルロ法を実装している [3]。また、ベイズ学習において周辺尤度を計算する際に、事後分布の尤度関数の部分にのみ温度パラメータを導入する方法が提案されている [6]。

そこで本研究では、変分ベイズ法において、尤度関数と事前分布の 2 つの項にそれぞれ温度パラメータを導入

*東京大学 大学院新領域創成科学研究科, 〒 277-8561 千葉県柏市柏の葉 5-1-5, e-mail nagata@mns.k.u-tokyo.ac.jp, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan

†科学技術振興機構 ERATO 岡ノ谷情動情報プロジェクト, 東京大学 大学院新領域創成科学研究科, e-mail katahira@mns.k.u-tokyo.ac.jp

‡Japan Science Technology Agency, ERATO, Okanoya Emotional Information Project, Graduate School of Frontier Sciences, The University of Tokyo.

§科学技術振興機構 ERATO 岡ノ谷情動情報プロジェクト, 理化学研究所 脳科学総合研究センター, e-mail okanoya@brain.riken.jp Japan Science Technology Agency, ERATO, Okanoya Emotional Information Project, RIKEN Brain Science Institute.

§東京大学 大学院新領域創成科学研究科, 理化学研究所 脳科学総合研究センター, e-mail okada@k.u-tokyo.ac.jp Graduate School of Frontier Sciences, The University of Tokyo, RIKEN Brain Science Institute.

し、その枠組みから、事前分布のハイパーパラメータの部分最適化を組み合わせた確定的アニーリング法を提案する。また、それらの有効性を混合正規分布モデルに適用することで、実験的に検証する。

2 ベイズ学習

観測されたデータ集合を $X^n = \{x_1, \dots, x_n\}$ とし、それらを生成している確率分布を推定するために、モデルパラメータ θ をもつ確率モデル $p(x|\theta)$ を考える。ベイズ学習では、データ集合 X^n が与えられたもとのパラメータ θ の分布をあらわすベイズ事後分布を次式により導入する。

$$p(\theta|X^n) = \frac{1}{Z(X^n)} p(X^n|\theta) \varphi(\theta) \quad (1)$$

ここで、 $p(X^n|\theta)$ は尤度関数、 $\varphi(\theta)$ はモデルパラメータの事前分布であり、 $Z(X^n)$ は規格化定数である。このベイズ事後分布を用いて、学習モデル $p(x|\theta)$ を期待値計算することにより予測分布を構成し、データを生成している真の確率分布を推定する。すなわち、

$$p(x|X^n) = \int p(x|\theta) p(\theta|X^n) d\theta \quad (2)$$

近年、ベイズ学習において、学習の性質を解明するための理論的研究が精力的に行われ、その結果、ベイズ学習は、最尤推定法やMAP推定法に比べ、汎化能力に優れることが明らかにされている [11]。しかしながら、一般にベイズ事後分布が複雑になってしまい、予測分布の期待値計算が容易に実行できないといった問題があるため、様々なベイズ学習の近似手法が考案されている。その一つが変分ベイズ法であり、少ない計算量で実行できることで広く応用されている。

3 変分ベイズ法

観測データ集合 X^n に対応して、潜在変数の集合 $Y^n = \{y_1, \dots, y_n\}$ を導入し、それに合わせて確率モデルを $p(x, y|\theta)$ とする。変分ベイズ法では、潜在変数とパラメータの同時事後分布 $p(Y^n, \theta|X^n)$ を以下の変分事後分布 $q(Y^n, \theta)$ で近似することを考える。

$$q(Y^n, \theta) = Q(Y^n) r(\theta) \quad (3)$$

近似の際の評価関数として、変分事後分布 $q(Y^n, \theta)$ と事後分布 $p(Y^n, \theta|X^n)$ との間のカルバック情報量

$$D(q||p) = \sum_{Y^n} \int q(Y^n, \theta) \log \frac{q(Y^n, \theta)}{p(Y^n, \theta|X^n)} d\theta \quad (4)$$

を考え、これを最小にする変分事後分布を見つけることが目的となる。式 (4) は、カルバック情報量の性質

$D(q||p) \geq 0$ と、ベイズ事後分布の式 (1) を同時事後分布 $p(Y^n, \theta|X^n)$ に拡張したものを考えることで、以下の不等式が導かれる。

$$\begin{aligned} -\log Z(X^n) &\leq \sum_{Y^n} \int q(Y^n, \theta) \log q(Y^n, \theta) d\theta \\ &\quad - \sum_{Y^n} \int q(Y^n, \theta) \log p(X^n, Y^n, \theta) d\theta \\ &\equiv F(q) \end{aligned} \quad (5)$$

ここで、 \sum_{Y^n} は潜在変数 Y^n の全ての出方についての和を表す。上の不等式において、左辺は変分事後分布 $q(Y^n, \theta)$ には依存しないため、右辺の変分自由エネルギー $F(q)$ を $Q(Y^n)$ と $r(\theta)$ について最小化することが、最良の変分事後分布を求めることと等価になる。 $Q(Y^n)$ についての最小化をVB-Eステップ、 $r(\theta)$ についての最小化をVB-Mステップと呼ぶ。それぞれのステップにおける更新式は、以下の式 (6), (7) で表される。

$$Q(Y^n) \propto \exp \langle \log p(X^n, Y^n|\theta) \rangle_{r(\theta)} \quad (6)$$

$$r(\theta) \propto \varphi(\theta) \exp \langle \log p(X^n, Y^n|\theta) \rangle_{Q(Y^n)} \quad (7)$$

ここで、 $\langle \cdot \rangle_{r(\theta)}$ 、 $\langle \cdot \rangle_{Q(Y^n)}$ はそれぞれ、確率分布 $r(\theta)$ 、 $Q(Y^n)$ での期待値を表す。これらの更新式は、別の分布での期待値計算を含んでいるために、アルゴリズムとして反復計算を必要とする。その操作はEMアルゴリズムのそれと類似しており、EMアルゴリズムと同様に少ない演算量で実行可能であるといった利点が知られている。しかしながら、勾配型のアルゴリズムであるゆえ、局所解の問題もある。この問題に対して、変分ベイズ法における確定的アニーリングが考案されている。確定的アニーリングでは、ベイズ事後分布に対して下記のように逆温度パラメータ β を導入することにより構成されている。

$$p(Y^n, \theta|X^n, \beta) \propto \{p(X^n, Y^n|\theta) \varphi(\theta)\}^\beta \quad (8)$$

4 提案手法

前節で説明したように、従来のアニーリングでは、温度パラメータを事後分布全体に作用させることにより提案されていた。本研究では、尤度関数と事前分布それぞれに温度パラメータを導入する枠組みを提案する。また、その枠組みから、変分ベイズ法におけるハイパーパラメータの部分最適化を組み込んだ改良型の確定的アニーリング法を提案する。

まずはじめに、ベイズ事後分布に2つの逆温度パラメータ β_1, β_2 を導入した事後分布を考える。

$$p(Y^n, \theta|X^n, \beta_1, \beta_2) \propto p(X^n, Y^n|\theta)^{\beta_1} \varphi(\theta)^{\beta_2} \quad (9)$$

この事後分布に対して、第3節と同様の解析を行うことで、以下の変分自由エネルギーを導出することができる。

$$\begin{aligned}
F_{\beta_1, \beta_2}(q) &= \sum_{Y^n} \int q(Y^n, \theta) \log q(Y^n, \theta) d\theta \\
&\quad - \beta_1 \sum_{Y^n} \int q(Y^n, \theta) \log p(X^n, Y^n | \theta) d\theta \\
&\quad - \beta_2 \int r(\theta) \log \varphi(\theta) d\theta \quad (10)
\end{aligned}$$

また、この変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ の最小化を考えることで、VB-E ステップ、VB-M ステップでの更新式を、以下のように導出できる。

$$\begin{aligned}
Q(Y^n) &\propto \exp \langle \beta_1 \log p(X^n, Y^n | \theta) \rangle_{r(\theta)} \\
r(\theta) &\propto \varphi(\theta)^{\beta_2} \exp \langle \beta_1 \log p(X^n, Y^n | \theta) \rangle_{Q(Y^n)}
\end{aligned}$$

これらの式は、 $\beta_1 = 1.0$ 、 $\beta_2 = 1.0$ のときに、従来の変分ベイズ法の更新式と同じになる。

変分ベイズ法では、変分自由エネルギー $F(q)$ が収束するまで、E ステップと M ステップを交互に繰り返すことで、 $F(q)$ を最小化する。本研究で提案する手法では、目的関数として温度パラメータ付き変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ を適用しアニーリングを制御するための β_1, β_2 に関するループを付加することでアルゴリズムを構成することが出来る。

従来の確定的アニーリング付きの変分ベイズ法では、 $\beta_1 = \beta_2 = \beta$ とした上で、一変数の温度パラメータを制御することで構成していた。本研究で提案する手法では、 β_1, β_2 を別々に制御するように構成する。そのアルゴリズムを以下に示す。

[提案アルゴリズム]

1. β_1, β_2 の初期値を設定する。
2. パラメータについての変分事後分布 $r(\theta)$ のハイパーパラメータを初期化する。
3. 以下の VB-EM ステップを収束するまで繰り返す。
VB-E ステップ:

$$Q(Y^n) \propto \exp \langle \beta_1 \log p(X^n, Y^n | \theta) \rangle_{r(\theta)}$$

VB-M ステップ:

$$r(\theta) \propto \varphi(\theta)^{\beta_2} \exp \langle \beta_1 \log p(X^n, Y^n | \theta) \rangle_{Q(Y^n)}$$

4. β_1 を増加させる。
5. $\beta_1 = 1.0$ ならばステップ 6 へ、さもなければステップ 3 へ戻る。
6. β_2 を増加させる。
7. ステップ 3 と同様に、VB-EM ステップを収束するまで

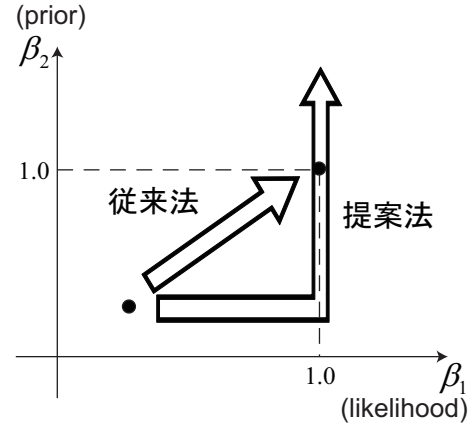


図 1: 提案アルゴリズムにおける温度パラメータのスケジューリング。アルゴリズム [VB03] のみ、目標とする温度パラメータが 1.0 とは限らず、変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ を最小にする β_2 が目標となる。

繰り返し、収束時の結果と変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ の値を保持する。

8. $\beta_2 < \bar{\beta}_2$ ならばステップ 6 へ戻り、さもなければ、保持した変分自由エネルギーの最小値に対応する変分事後分布をアルゴリズムの結果とし、終了する。

図 1 は、確定的アニーリング法における従来法と提案手法での温度パラメータのスケジューリングを表したものである。このアルゴリズムの設定により得られる効果について、以下で議論する。

アルゴリズム前半(ステップ 3-5)では、尤度関数における温度パラメータを制御しており、従来のアニーリングと同様の効果が期待できる。混合正規分布モデルや隠れマルコフモデルなどのモデルにおいて、ベイズ事後分布が複雑になってしまう主な原因として、学習モデルがもつ階層性が挙げられる。その特徴から、モデルとパラメータが一对一に対応せず、尤度関数のピークがパラメータ空間上で一点ではなく、ある低次元多様体上に広がりをもつように構成されてしまう。 β_1 によるアニーリングは、その複雑化した尤度関数に着目し、その部分にのみアニーリングを適用した手法になっており、階層的な構造を持つ学習モデルにおいて、その効果が期待される。

一方で、アルゴリズム後半(ステップ 6-8)では、尤度関数の温度パラメータを固定するため、アニーリング効果はさほど期待できない。代わりに、事前分布のハイパーパラメータの部分最適化を行うことができる。

尤度関数に温度パラメータを導入した場合、モデルの

構造を変化させているため温度パラメータ付きの変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ から、学習として最適な温度パラメータを見つけることはできない。その一方で、 β_2 のみを変化させる過程を考えると、温度パラメータの変化に対して、尤度関数の変化はなく、事前分布のみが変化する構造になっている。そのため、変分自由エネルギー $F_{\beta_1, \beta_2}(q)$ を最小にする温度パラメータ β_2 を探索することで、事前分布の最適化を行うことができる。また、変分ベイズ法でよく用いられる混合指数型分布や隠れマルコフモデルでは事前分布 $\varphi(\theta)$ に共役事前分布を用いることが要請され、その結果、 $\varphi(\theta)^{\beta_2}$ は、もとの事前分布のハイパーパラメータを変化させた分布となる。一般に事前分布のハイパーパラメータを最適化することは、ベイズ学習において重要な課題であるものの、すべてのハイパーパラメータにおいて最適解を探索することは容易でない。本提案手法における、 β_2 のアニーリング過程は、各ハイパーパラメータの比を一定にした上での一次元最適化問題になっており、ハイパーパラメータ最適化を考える際に有用であると考えられる。

以上の性質をまとめると、提案手法では、まず尤度関数についての温度パラメータ β_1 を変化させることで解を安定化させ、その後事前分布についての温度パラメータ β_2 を変化させることでハイパーパラメータについても最適化を行うことによって、従来のアニーリング法よりも最良な解を探索できることが期待できる。

5 混合正規分布モデルによる検証

提案手法の有効性を検証するために、混合正規分布モデルにおける変分ベイズ法を実験的に行った。この節では、その結果を示す。

5.1 混合正規分布モデル

本実験では、有効性の検証として、混合正規分布モデルを扱う。データ x を M 次元とし、 K 個の正規分布からなる混合正規分布モデルの式は、

$$p(x|\theta) = \sum_{k=1}^K a_k \mathcal{N}(x; \mu_k, S_k^{-1}) \quad (11)$$

で与えられる。ここで、 $\mathcal{N}(x; \mu, S^{-1})$ は、平均が μ 、分散共分散行列が S^{-1} の M 次元正規分布である。潜在変数 y を含めた形は以下のように表せる。

$$p(x, y|\theta) = \prod_{k=1}^K \{a_k \mathcal{N}(x; \mu_k, S_k^{-1})\}^{y^k} \quad (12)$$

ここで、 y^k は、データ x が k 番目の正規分布から生成されていたならば $y^k = 1$ となり、そうでなければ $y^k = 0$

をとる潜在変数である。このモデルのパラメータは、各正規分布の混合比 $a = \{a_k\}_{k=1}^K$ 、平均 $\mu = \{\mu_k\}_{k=1}^K$ 、精度行列 $S = \{S_k\}_{k=1}^K$ である。各パラメータの事前分布は、共役事前分布を用いた。具体的には、混合比 a に関する事前分布は、Dirichlet 分布:

$$\varphi(a) = \text{Dir}(a; \phi_0) \propto \prod_{k=1}^K a_k^{\phi_0 - 1}, \quad (13)$$

各正規分布について、 S_k が与えられたもとの μ_k の事前分布は正規分布:

$$\varphi(\mu_k | S_k) = \mathcal{N}(\mu_k; \nu_0, (\xi_0 S_k)^{-1}), \quad (14)$$

S_k の事前分布は Wishart 分布:

$$\begin{aligned} \varphi(S_k) &= \mathcal{W}(S_k; \eta_0, B_0) \\ &\propto |S_k|^{(\eta_0 - M - 1)/2} \exp\left\{-\frac{1}{2} \text{Tr}(S_k B_0)\right\} \end{aligned} \quad (15)$$

とする。この場合の事前分布のハイパーパラメータは、 $\{\phi_0, \nu_0, \xi_0, \eta_0, B_0\}$ となる。ここで、事前分布に関する温度パラメータ β_2 を導入すると、事前分布は、その分布型は変わらず、各ハイパーパラメータは以下のようになる。

$$\begin{aligned} \phi'_0 &= \beta_2(\phi_0 - 1) + 1, & \nu'_0 &= \nu_0 \\ \xi'_0 &= \beta_2 \xi_0, & \eta'_0 &= \beta_2(\eta_0 - M - 1) + M + 1 \\ B'_0 &= \beta_2 B_0 \end{aligned}$$

このことから、 $\beta_1 = 1.0$ の場合、異なる事前分布を用いた場合の変分ベイズ法と等価であるため、提案手法において、事前分布のハイパーパラメータの最適化を行うことができることが伺える。

次に、尤度関数についての温度パラメータ β_1 を導入した変分ベイズ法におけるアルゴリズムを議論する。VB-M ステップについて、パラメータ a の変分事後分布 $r(a)$ は Dirichlet 分布:

$$r(a) = \text{Dir}(a; \{\phi_k\}_{k=1}^K), \quad (16)$$

各正規分布について、 S_k が与えられたもとの μ_k の変分事後分布は正規分布:

$$r(\mu_k | S_k) = \mathcal{N}(\mu_k; \bar{\mu}_k, (\xi_k S_k)^{-1}), \quad (17)$$

S_k の変分事後分布は Wishart 分布:

$$r(S_k) = \mathcal{W}(S_k; \eta_k, \bar{B}_k) \quad (18)$$

で与えられる。ここで、それぞれの変分事後分布のハイパーパラメータは

$$\phi_k = \beta_1 n_k + \phi'_0, \quad \bar{\mu}_k = \frac{\beta_1 n_k \bar{x}_k + \xi'_0 \nu'_0}{\beta_1 n_k + \xi'_0},$$

$$\begin{aligned}\xi_k &= \beta_1 n_k + \xi'_0, & \eta_k &= \beta_1 n_k + \eta'_0, \\ \bar{\mathbf{B}}_k &= \mathbf{B}'_0 + \beta_1 \bar{\mathbf{C}}_k + \frac{\beta_1 n_k \xi'_0}{\beta_1 n_k + \xi'_0} (\bar{\mathbf{x}}_k - \boldsymbol{\nu}'_0) (\bar{\mathbf{x}}_k - \boldsymbol{\nu}'_0)^T, \\ n_k &= \sum_{i=1}^n \bar{y}_i^k & \bar{\mathbf{x}}_k &= \frac{1}{n_k} \sum_{i=1}^n \bar{y}_i^k \mathbf{x}_i, \\ \bar{\mathbf{C}}_k &= \sum_{i=1}^n \bar{y}_i^k (\mathbf{x}_i - \bar{\mathbf{x}}_k) (\mathbf{x}_i - \bar{\mathbf{x}}_k)^T\end{aligned}$$

となる．また，VB-E ステップについて，潜在変数における変分事後分布 $Q(Y^n)$ についての y_i^k の期待値 \bar{y}_i^k は以下ようになる．

$$\begin{aligned}\bar{y}_i^k &= \langle y_i^k \rangle_{Q(Y^n)} = \frac{\exp(\beta_1 \gamma_i^k)}{\sum_{j=1}^K \exp(\beta_1 \gamma_i^j)}, & (19) \\ \gamma_i^k &= \Psi(\phi_k) - \Psi\left(\sum_{j=1}^K \phi_j\right) - \frac{1}{2} \log |\bar{\mathbf{B}}_k| \\ &+ \frac{1}{2} \sum_{i=1}^M \Psi\left(\frac{\eta_k + 1 - i}{2}\right) - \frac{M}{2\xi_k} \\ &- \frac{\eta_k}{2} \text{Tr}\left(\bar{\mathbf{B}}_k^{-1} (\mathbf{x}_i - \bar{\boldsymbol{\mu}}_k) (\mathbf{x}_i - \bar{\boldsymbol{\mu}}_k)^T\right)\end{aligned}$$

ここで， $\Psi(\cdot)$ は，ディガンマ関数を表す．変分ベイズ法の場合，事前分布として共役事前分布を用いることで，変分事後分布も事前分布と同型の確率分布になるという利点が知られているが，温度パラメータ β_1, β_2 を導入した場合にもその利点は生かされていることがわかる．

本研究では，提案手法の有効性を検証するために以下の3つのアルゴリズムにおける挙動を比較した．

1.[改良アニーリング VB 法]

提案アルゴリズムにより，変分ベイズ法を行う．具体的には，尤度関数における温度パラメータ β_1 についてアニーリングを行い，その後，事前分布における温度パラメータ β_2 について変化させ，事前分布の部分最適化を行う手法である．その際の温度パラメータ β_1, β_2 の初期値は， $\beta_i(0) = 0.01$ とした．温度パラメータ β_1, β_2 の更新式は， $\beta_i(t+1) = \frac{2\beta_i(t)}{1+\beta_i(t)}$ とした [2]．この更新式は， $t \rightarrow \infty$ において $\beta_i = 1.0$ に漸近するように構成されている．終了条件として， $t = 10$ において， $\beta_1(t) = 1.0$ となるようにした．さらに， β_2 における $\beta_2 > 1.0$ の領域における更新式は， $\beta_2(t+1) = 1.25 \times \beta_2(t)$ として15回の更新を行った．

2.[従来のアニーリング VB 法]

Katahira らの提案したアニーリング法を用いた変分ベイズ法を行う．その際の温度パラメータ β の制御は [改良アニーリング VB 法] と同様に設定した．具体的には，初期値を $\beta(0) = 0.01$ とし，更新式は， $\beta(t+1) = \frac{2\beta(t)}{1+\beta(t)}$ とした．また，終了条件として， $t = 10$ において， $\beta(t) = 1.0$

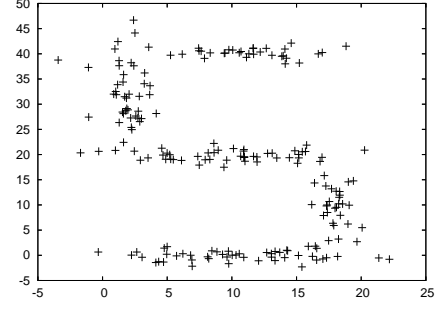


図 2: 検証に用いた学習データのプロット図

となるようにした．

3.[アニーリングなし VB 法]

アニーリングを用いず，従来の変分ベイズ法を行う．

以上の3つのアルゴリズムのうち，アニーリングを用いた変分ベイズ法では，全ての正規分布を完全に同じにするといった局所解が多く得られてしまった．このような状況を回避するために，各温度パラメータでの収束先として正規分布の平均ベクトル $\bar{\boldsymbol{\mu}}_k$ の値が同じになる k_1, k_2 が存在した場合，次の温度パラメータでの初期値として， $\bar{\boldsymbol{\mu}}_{k_1}$ と $\bar{\boldsymbol{\mu}}_{k_2}$ に微小な乱数を加えたものを設定するようにアルゴリズムを変更した．

5.2 人工データによる検証

まず始めに，学習データ X^n を，5個の二次元正規分布からなる混合正規分布モデルから人工的に生成した場合で検証を行った．すなわち， $M = 2$ である．図 2 は，用いた学習データのプロット図である [4][5]．学習データの個数 n は 200 とした．この学習データに対して， $K = 5$ の混合正規分布モデルで変分ベイズ学習を行った．事前分布 $\varphi(\theta)$ のハイパーパラメータは，

$$\begin{aligned}\phi_0 &= 1.0, & \boldsymbol{\nu}_0 &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, & \xi_0 &= 0.01, \\ \eta_0 &= M + 1 = 3.0, & \mathbf{B}_0 &= \mathbf{I}_M\end{aligned}$$

とした．ここで， \mathbf{I}_M は $M \times M$ の単位行列を表す．また，各手法における変分事後分布 $r(\theta)$ のハイパーパラメータの初期値は， $\boldsymbol{\nu}_k$ については，学習データの平均と分散に基づいてランダムに設定し， $\bar{\mathbf{B}}_k$ に関しては， $\bar{\mathbf{B}}_k = \mathbf{B}_0$ に設定した．その他のハイパーパラメータについては， $n_k = n/K$ として，

$$\phi_k = n_k + \phi'_0, \quad \xi_k = n_k + \xi'_0, \quad \eta_k = n_k + \eta'_0$$

と設定した．

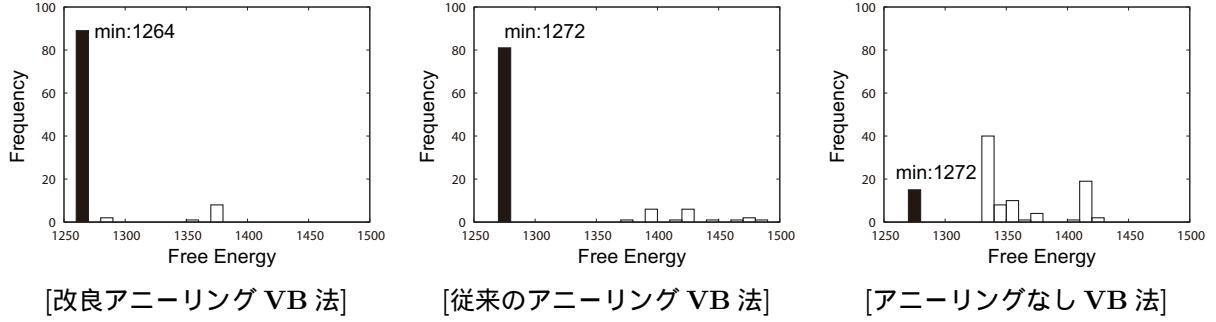


図 3: 人工データでの，各アルゴリズムにおける変分自由エネルギーのヒストグラム

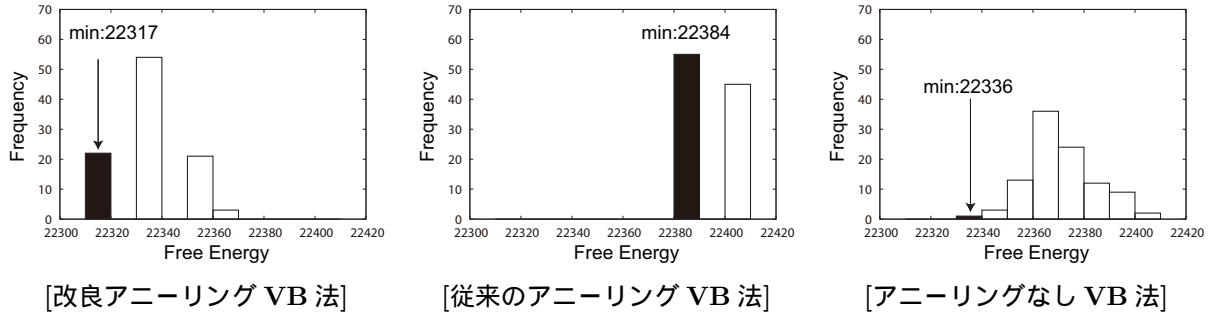


図 4: 実データでの各アルゴリズムにおける変分自由エネルギーのヒストグラム

5.2.1 実験結果

これらのアルゴリズムについて，初期値の取り方について100回のシミュレーションを行った．図3は，各アルゴリズムにおける収束先での変分自由エネルギーのヒストグラムを表したものである．それぞれの図において，変分自由エネルギーが最小となるヒストグラムを黒塗りしてある．この結果を見ると，[従来のアニーリングVB法]では，局所解にトラップされる頻度が[アニーリングなしVB法]と比較してかなり小さく，最適解に到達することが多く，改善されていることが分かる．一方，[改良アニーリングVB法]での結果は，変分自由エネルギーをより最小にする最適解が得られ，その解へ到達する頻度は[従来のアニーリングVB法]と同等である．このことから，提案手法である[改良アニーリングVB法]が従来の手法と比較して，有効であることがわかる．

5.3 実データによる検証

本研究では，提案アルゴリズムの有効性の検証のため，The Saccharomyces Cerevisiae Morphological Database (SCMD)[9]で公開されている出芽酵母の必須遺伝子破壊株における形態データに混合正規分布モデルによる変分ベイズ法を適用した．このデータは，501次元のデータであり，その個数は4718個である．本研究では，501

次元のうち，論文[7]で着目している4次元を用いた．つまり， $n = 4718$ ， $M = 4$ という設定でシミュレーションを行った．この学習データに対して， $K = 10$ の混合正規分布モデルで変分ベイズ学習を行った．事前分布 $\varphi(\theta)$ のハイパーパラメータは，

$$\begin{aligned} \phi_0 &= 1.0, \quad \nu_0 = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \xi_0 = 1.0, \\ \eta_0 &= M + 1 = 5.0, \quad \mathbf{B}_0 = \mathbf{I}_M \end{aligned}$$

とした．各手法における変分事後分布 $r(\theta)$ のハイパーパラメータの初期値の設定は，先の実験と同様にし，温度パラメータのスケジュールに関しても，初期条件，更新式は同じにし，終了条件のみ， $t = 15$ において $\beta(t) = 1.0$ となるように変更した．

5.3.1 実験結果

人工データの場合と同様に，初期値の取り方について100回のシミュレーションを行った．図4は，各アルゴリズムにおける変分自由エネルギーのヒストグラムを表したものである．それぞれの図中において，変分自由エネルギーの最小値におけるヒストグラムを黒塗りして表示している．この結果を見ると，[従来のアニーリングVB法]の結果は安定してはいるものの，変分自由エネルギーの値を[アニーリングなしVB法]と比較すると，その結果の全てが局所解であることがわかる．それに対

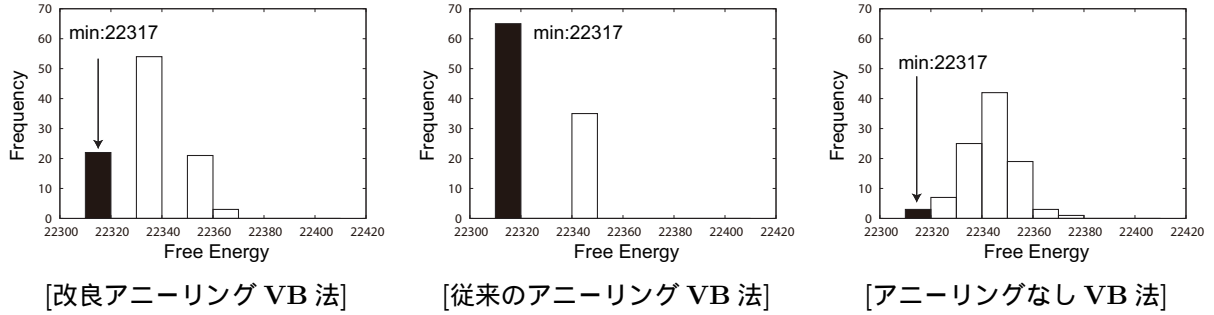


図 5: 実データにおける，最適な事前分布を用いた場合の各アルゴリズムにおける変分自由エネルギーのヒストグラム

し，[改良アニーリング VB 法] では [従来のアニーリング VB 法] よりも小さな変分自由エネルギーが得られており，提案手法の有効性がみてとれる．しかしながら，[改良アニーリング VB 法] の中で得られた結果をみると，最適解に到達する頻度は，局所解にトラップされる頻度に比べ小さく，安定した結果が得られていないことも分かる．この点に関しては，次章でより深く考察を行う．

6 考察

本章では，本研究で提案した手法に関連した考察を行う．

(1) 本研究で行ったシミュレーションの結果として，アルゴリズム [VB01] および [VB02] は解として全ての正規分布を同じにするという局所解にトラップされる現象が多く観測された．これは，式 (19) から分かる通り，温度パラメータ β_1 が小さい領域では，各々の k について， \bar{y}_i^k の値が等しくなっていくために起こるものと考えられる．この現象は，アルゴリズムの性質上，混合正規分布に限らず，混合指数型分布や隠れマルコフモデルなど，潜在変数を導入した変分ベイズ法ならば，共通に現れる現象と思われる．

本実験では，そのような現象を回避するために，次の温度パラメータの初期値に微小な乱数を加えることを行ったが，乱数を加える際の条件や，乱数の種類や大きさなど，設定に恣意性が残る．また，Katahira らは，温度パラメータとして小さい値を用いずにある程度大きな値からアルゴリズムを開始する方法を用いており [5]，Sato らは，各温度パラメータ上での VB アルゴリズムを収束させずに途中で温度パラメータを増加させることを行っている [8]．しかしながら，これらの方法では，初期値の影響が残ってしまう可能性があり，また，開始する温度パラメータの値や，各温度パラメータ上での VB アルゴリズムのステップ数など，やはり設定に恣意性が

残ると思われる．

このように，変分ベイズ法におけるアニーリングを構築する際，全ての要素を等しくする局所解にトラップされる問題は深刻であり，どの学習モデルでも一般的に現れるものと考えられる．その問題に対応したアルゴリズムや設定を考えることは今後の課題である．

(2) 本研究で考えたアルゴリズムの性質について考察する．提案手法では，温度パラメータを 2 つ導入し，まず尤度関数についての温度パラメータを変化させ，その後，事前分布についての温度パラメータを変化させることで，アニーリング効果だけでなく，事前分布のハイパーパラメータの部分最適化の効果も目指している．人工データによる実験では，その両方の効果が現れたことが確認できたが，実データによる実験においては，局所解にトラップされる結果が多く確認され，アニーリング効果が弱くなってしまっていた．

この点について，より深く考察するために，実データにおける追加実験を行った．図 5 は，先の実験における [改良アニーリング VB 法] で求められる最適な温度パラメータ β_2 に基づいて，事前分布を最適化し直した場合において，3 つのアルゴリズムを比較したものである．これらの結果を比較すると，[従来のアニーリング VB 法] が，最も頻繁に最適解に到達することが確認できる．このことから，[改良アニーリング VB 法] では，ハイパーパラメータの部分最適化を追加したものの，アニーリング効果としては，[従来のアニーリング VB 法] に比べ弱まる場合が存在することが分かった．また，従来のアニーリング VB 法についても，事前分布が最適でない場合には，アニーリングを用いることでより悪い局所解にトラップされてしまうことも分かった．

このように，2 つの温度パラメータのスケジューリングにはより最適な方法があるものと考えられる．これを確立することは非常に重要であり，今後の課題として取

り組む必要がある。

7 おわりに

本研究では、変分ベイズ法において、2種類の温度パラメータを導入することで、確定的アニーリングの拡張と、事前分布のハイパーパラメータの部分最適化法を提案し、その有効性を混合正規分布モデルでのシミュレーションにより示した。今後の課題は、温度パラメータのスケジューリングなどの設定法の確立、及び、その他のモデルや問題に適用することなどが挙げられる。

謝辞

本研究は、科学研究費補助金 18079003, 20240020, 及び、20650019 の援助を受けた。また、The *Saccharomyces Cerevisiae* Morphological Database (SCMD) で公開されているデータを用いてシミュレーションを行った。SCMD データベースは本研究での使用の目的に限り、東京大学から無償で提供されている。

参考文献

- [1] H. Attias, “Inferring Parameters and Structure of Latent Variable Models by Variational Bayes,” in Proc. 15th Conf. on Uncertainty in Artificial Intelligence, pp.21-30, 1999.
- [2] Z. Ghahramani, G. E. Hinton, “Variational Learning for Switching State-Space Models,” Neural Computation, vol.12, pp. 831-864, 2000.
- [3] S. Hara et. al. “LDPC Decoding Dynamics from a PCA Viewpoint,” Interdisciplinary Information Sciences, Vol. 13, No. 1, pp. 43-48, 2007.
- [4] 樺島祥介, 上田修功, “平均場近似・EM法・変分ベイズ法,” 計算統計 I - 確率計算の新しい手法 統計科学のフロンティア 11, 岩波書店, 2003.
- [5] K. Katahira, K. Watanabe, and M. Okada, “Deterministic Annealing Variant of Variational Bayes Method,” Journal of Physics: Conference Series, Vol. 95, 012015, 2008.
- [6] Y. Ogata, “A Monte Carlo Method for an Objective Bayesian Procedure,” Ann. Inst. Stat. Math., Vol. 42, No. 3, pp. 403-433, 1990.
- [7] Y. Ohya et.al., “High-dimensional and large-scale phenotyping of yeast mutants,” Proc. Natl. Acad. Sci. USA, Vol. 102, No. 52, pp. 19015-19020, 2005.
- [8] I. Sato et. al., “Quantum Annealing for Variational Bayes Inference,” in Proc. 25th Conf. on Uncertainty in Artificial Intelligence, 2009.
- [9] “The *Saccharomyces Cerevisiae* Morphological Database (SCMD),” <http://yeast.gi.k.u-tokyo.ac.jp/datamine/>.
- [10] K. Watanabe, and S. Watanabe, “Stochastic Complexities of Gaussian Mixtures in Variational Bayesian Approximation,” The Journal of Machine Learning Research, Vol. 7, pp. 625-644, 2006.
- [11] S. Watanabe, “Algebraic Analysis for Non-Identifiable Learning Machines,” Neural Computation, vol. 13, pp. 899-933, 2001.