

レプリカ交換モンテカルロ法を用いた Mixture of Experts モデルにおけるベイズ推論

松平京介^{†1} 永田賢二^{†2} 本武陽一^{†1} 岡田真人^{†1}

概要：本論文では、マルコフ連鎖モンテカルロ法（MCMC 法）の一種であるレプリカ交換モンテカルロ法（REMC 法）を用いて、Mixture of Experts（ME）モデルにおけるベイズ推論を行う方法を提案する。これまで、ME モデルを扱ったベイズ推論では、主に変分ベイズ法による推定が行われてきた。変分ベイズ法では、計算の都合上、入力空間を分割する考えに基づく ME モデルとしては不自然な確率モデルを導入することや、近似事後分布をガウス分布に近似することで推定が行われる。本論文で提案する手法を用いると、入力空間を分割する考えに基づく ME モデルの本来の確率モデルにおいてベイズ推論を行うことができる。提案手法を計算機シミュレーションで実施し、特異モデルに特有の事後分布の特徴が捉えられることを確認することで、ベイズ推論が正しく行えることを確認した。また、自由エネルギーに基づくモデル選択により正しいモデルが精度よく選ばれることを確認した。

キーワード：ベイズ推論, Mixture of Experts, 正規化ガウス関数ネットワーク, レプリカ交換モンテカルロ法, 特異モデル

Bayesian inference for Mixture of Experts using Replica Exchange Monte Carlo method

KYOSUKE MATSUDAIRA^{†1} KENJI NAGATA^{†2} YOH-ICHI MOTOTAKE^{†1} MASATO OKADA^{†1}

Abstract: In this paper, we propose a method of the Bayesian inference for Mixture of Experts (ME) using Replica exchange Monte Carlo methods (REMC) which is one type of Markov chain Monte Carlo methods (MCMC). Variational Bayes method has been mainly used for estimation of ME. The probabilistic model of ME is constructed based on the idea of dividing input-space. To avoid the difficulty of analytical treatment, the variational Bayesian estimation of ME needs to modify the probabilistic model to unnatural one, which is far from the idea of dividing input-space, or approximate the posterior distribution to the Gaussian distribution. Our methods can deal with the original probabilistic model as ME in the Bayesian inference. We demonstrate experimentally that the characteristics of posterior distribution peculiar to a singular model appear in the posterior distribution. High accuracy of model selection can be performed based on Bayesian free energy.

Keywords: Bayesian inference, Mixture of Experts, Normalized Gaussian network, Replica Exchange Monte Carlo methods, Singular model

^{†1} 現在, 東京大学大学院新領域創成科学研究科
Presently with Graduate School of Frontier Science, The University of Tokyo

^{†2} 現在, 国立研究開発法人産業技術総合研究所, 人工知能研究センター/JST さきがけ
Presently with Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology/JST PRESTO

1. はじめに

非線形関数に従って生成されたデータをモデル化する一つの方法として, 分割統治的なアプローチが考えられる。このような場合の分割統治とは, 入力空間をいくつかの領域に分割し, その領域内のデータを表現する回帰関数を設

計することである。分割統治を行うモデルの一つに ME モデルが存在する。本論文では、ME モデルを用いて回帰を行い、データから領域の分割数や回帰関数のパラメータを推定する問題を考える。

ME モデルは、多層ニューラルネットワークモデルや混合正規分布、隠れマルコフモデルなどと同じく、特異モデルと呼ばれるモデルに属する。特異モデルでは、モデルに階層性を含むことから、モデルとパラメータが 1 対 1 に対応しないため、パラメータ空間において最適なパラメータが一点に定まらず、パラメータ空間において特異点を含む多様体構造をなす。この多様体構造のため、フィッシャー情報行列が縮退してしまう性質を有する。そのため、フィッシャー情報行列の正則性を仮定して作られた理論を特異モデルに適用することができず、例えば、赤池情報量基準 (AIC) やベイズ情報量基準 (BIC) に代表されるガウス分布に依拠した理論が成立しない。近年、特異モデルに対して、ベイズ推定が最尤推定に比べて推定精度の観点で優れていることが明らかにされている [1]。そこで、本論文では MCMC 法の一つである REMC 法を用いて ME モデルにおけるベイズ推論を行う方法を提案する。

これまで、ME モデルにおけるベイズ推論では、主に変分ベイズ法による推定が行われてきた [2], [3], [4], [5]。変分ベイズ法では、計算の都合上、入力空間を分割する考えに基づく ME モデルとしては不自然な確率モデルを導入することや [2], [3], [4]、近似事後分布をガウス分布に近似することによって推定が行われる [5]。入力空間を分割する考えに基づく ME モデルとしては不自然な確率モデルにおいて得られる推定結果は、ME モデルの本来の確率モデルにおいて得られる推定結果とは異なる可能性がある。また、ME モデルは特異モデルであるため、事後分布をガウス分布で近似することも適切ではないと考えられる。本論文での目的は、REMC 法を利用することで、入力空間を分割する考えに基づく ME モデルの本来の確率モデルに対して、近似を用いることなくベイズ推論を実行できる方法を提案することである。以降、入力空間を分割する考えに基づく ME モデルの本来の確率モデルを自然な確率モデルと呼ぶ。提案手法を計算機シミュレーションで実施し、特異モデルに特有の事後分布の特徴が捉えられることを確認することで、ベイズ推定が正しく行えることを確認した。また、自由エネルギーに基づくモデル選択により正しいモデルが精度よく選ばれることを確認した。

2. ME モデルと NGnet

2.1 ME モデル

ME モデルとは、分割統治の原則を元に Jacobs らによって提案されたモデルである [6], [7]。図 1 は ME モデルの入出力関係を表す概念図である。ME モデルは複数のエキスパートネットワークと、1 つのゲートネットワークからな

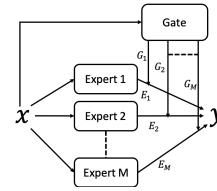


図 1: ME モデルの概念図

Fig. 1 Concept of ME.

る。ゲートが入力空間 (x の空間) を分割し、各分割領域内で各々のエキスパートが出力を担う。入出力関係を表す式を以下で与える。

$$y = \sum_{i=1}^M G_i(x) E_i(x) \quad (1)$$

ここでは、 x は入力、 y は出力、 M は領域数 (エキスパートの数) を表す。エキスパート関数 $E_i(x)$ がゲート関数 $G_i(x)$ で重みづけられ、それらを全領域で足し合わせることで最終的な出力 y が決定する。また、任意の x について $G_i(x)$ は $\sum_{i=1}^M G_i(x) = 1$ を満たしている。仮に、ある j において $G_{i=j}(x) = 1$ となる x の領域があれば、その領域では $G_{i \neq j}(x) = 0$ となり、最終的な出力に関わるのは j 番目のエキスパートだけとなる。この意味で i は領域を指定するインデックスだと考えることもできる。また、ゲート関数 $G_i(x)$ は通常、soft-max 関数と呼ばれる領域間を滑らかに分割するような関数である。

2.2 ME モデルの確率モデル

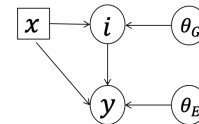


図 2: ME モデルのグラフィカル表現

Fig. 2 Graphical representation of ME.

図 2 は ME モデルのグラフィカルモデルを示したものである [7]。図 2 で、確率変数であるものは丸で囲まれており、確率変数でないものは四角で囲まれている。つまり入力 x は確率変数ではない。図 2 から分かるように、このモデルでは、与えられた入力 x とゲートのパラメータ θ_G から領域 i が指定される。そして、指定された領域 i 、エキスパートのパラメータ θ_E 、入力 x から出力 y が決定される。この時の確率モデルは以下のように定式化できる。

$$p(y, i | x, \theta) = p(y | x, i, \theta_E) p(i | x, \theta_G) \quad (2)$$

ここで、 $\theta = \{\theta_G, \theta_E\}$ である。一般的に $p(i | x, \theta_G)$ はゲート関数 $G_i(x)$ そのものになる。この確率モデルでは、入力 x の値に基づいて領域 i が生成されると考えているので、入力空間を分割するという分割統治の考え方を反映していることがわかる。

2.3 NGnet

NGnet は ME モデルの一種であり、本論文で用いるモデルは NGnet である。以下では入力 x 、出力 y は全て 1 次元のスカラー量と考える。NGnet は ME モデルの一種であるため、入力 x と出力 y の関係は式 (1) に従う。NGnet の具体的なエキスパート関数 $E_i(x)$ とゲート関数 $G_i(x)$ は以下のようになる。

$$E_i(x) = w_i x + b_i + \varepsilon'_i, \quad (3)$$

$$G_i(x) = \frac{\mathcal{N}(x|\mu_i, s_i)}{\sum_{j=1}^M \mathcal{N}(x|\mu_j, s_j)}. \quad (4)$$

ここでのゲート関数 $G_i(x)$ は正規化ガウス関数であり、エキスパート関数 $E_i(x)$ は線形関数である。 $\mathcal{N}(x|\mu_i, s_i)$ は平均 μ_i 、逆分散 s_i のガウス関数を表す。 $\theta_E = \{w_i, b_i\}_{i=1}^M$ はエキスパートのパラメータであり、 $\theta_G = \{\mu_i, s_i\}_{i=1}^M$ はゲートのパラメータである。また、 ε'_i は平均 0、逆分散 s' のガウス分布に従うガウスノイズである。今回推定するパラメータは線形関数の傾き w_i 、切片 b_i 、ゲートのガウス関数の平均 μ_i 、逆分散 s_i の 4 種類である。

式 (2) に従って、NGnet の確率モデルを具体的に定式化する。式 (2) の右辺の項は以下のように定式化できる。

$$p(y_k|x_k, i, \theta_E) = \sqrt{\frac{s'}{2\pi}} \exp\left(-\frac{s'}{2}(y_k - (w_i x_k + b_i))^2\right), \quad (5)$$

$$p(i|x_k, \theta_G) = G_i(x_k). \quad (6)$$

ここでデータは $D = \{X, Y\} = \{x_k, y_k\}_{k=1}^n$ で表す。 k はデータのインデックスであり、 n はデータ数を表す。尤度は以下のように定義する。

$$\begin{aligned} p(Y|X, \theta) &= \prod_{k=1}^n \sum_{i=1}^M p(y_k|i, x_k, \theta_{E,i}), \\ &= \prod_{k=1}^n \sum_{i=1}^M p(y_k|i, x_k, \theta_{E,i}) p(i|x_k, \theta_{G,i}). \end{aligned} \quad (7)$$

3. 推定方法

本論文では、ベイズ推論の枠組みでモデル選択とパラメータの事後分布の推定を行った。ここでのモデル選択とは、適切な領域数を推定することである。モデル選択の指標として、自由エネルギーを用いた。自由エネルギーの計算と事後分布の推定には MCMC 法の一種である REMC 法を用いた。

3.1 ベイズ推論

ベイズ推論とは観測事象（データ D ）から推定したいものを確率的に推論する方法である。今回推定したいものはパラメータ θ と領域数 M であるため、パラメータの事後確率 $p(\theta|D)$ とモデルの事後確率 $p(M|D)$ を考える。ベイズの定理を用いるとパラメータの事後確率 $p(\theta|D)$ は以下のように表される。

$$\begin{aligned} p(\theta|D) &= \frac{p(D|\theta)p(\theta)}{p(D)}, \\ &= \frac{p(Y|X, \theta)p(\theta)}{p(Y|X)}, \\ &\propto p(Y|X, \theta)p(\theta). \end{aligned} \quad (8)$$

ここで、尤度 $p(Y|X, \theta)$ は前節で定義した式 (7) であり、パラメータの事前確率 $p(\theta)$ は 3.2 小節で定義する。同様にして、モデルの事後確率 $p(M|D)$ は以下のように表される。

$$\begin{aligned} p(M|D) &= \frac{p(D|M)p(M)}{p(D)}, \\ &= \frac{p(M)}{p(Y|X)} \int p(Y|X, \theta)p(\theta|M)d\theta, \\ &\propto \int p(Y|X, \theta)p(\theta|M)d\theta. \end{aligned} \quad (9)$$

また、モデル選択の指標である自由エネルギー $F(M)$ は以下のように定義される。

$$\begin{aligned} F(M) &= -\log \int p(Y|X, \theta)p(\theta|M)d\theta, \\ &= -\log p(M|D) + \text{const.} \end{aligned} \quad (10)$$

$F(M)$ を最小化する領域数 M を推定することがモデル選択の目的である。式 (10) より、自由エネルギー $F(M)$ を最小化することは、モデルの事後確率 $p(M|D)$ を最大化することと等価である。自由エネルギーの計算は 3.3 小節で紹介する REMC 法を用いることにより行うことが出来る。

3.2 事前分布

本論文では、事前分布 $p(\theta)$ を以下のように設定した。

$$p(\theta) = \prod_{i=1}^M p(w_i)p(b_i)p(\mu_i|s_i)p(s_i), \quad (11)$$

$$p(w_i) = \mathcal{N}(w_i|0, s'v_0), \quad (12)$$

$$p(b_i) = \mathcal{N}(b_i|0, s'v_0), \quad (13)$$

$$p(\mu_i|s_i) = \mathcal{N}(\mu_i|m_0, \gamma_0 s_i), \quad (14)$$

$$p(s_i) = \mathcal{G}(s_i|\nu_0, \eta_0). \quad (15)$$

$\mathcal{N}$ はガウス分布、 $\mathcal{G}$ はガンマ分布を表す（付録参照）。ハイパーパラメータは、6 種類で $\{s', v_0, m_0, \gamma_0, \nu_0, \eta_0\}$ である。

3.3 REMC 法

REMC 法は統計物理で登場するスピングラスと呼ばれる系を研究する際に提案された手法である [8]。REMC 法を用いることにより、事後分布に現れる正規化定数（統計物理では分配関数と呼ばれる。）や自由エネルギーを計算することできる [9], [10], [11]。ボルツマン分布との対応から、事後分布 $p(\theta|D)$ に逆温度パラメータ β_l を導入する。異なる逆温度を有する複数の系を考え、今回は以下のようにそれぞれの系での確率分布を与える。

$$p_{\beta_l}(\theta|D) \propto \exp(-\beta_l E(\theta))p(\theta). \quad (16)$$

ここで、式 (5), (6), (7) を用いて、 $E(\theta)$ を $E(\theta) = -\log p(Y|X, \theta)$ と定義する。逆温度は $0 = \beta_1 < \beta_2 < \dots < \beta_L = 1$ ($l = 1, 2, \dots, L$) とする。目標とする分布（事後分布）は $\beta_L = 1$ となる最低温の系の分布である。REMC 法は、これら複数の確率分布から同時に独立にサンプリングし、隣り合った温度間のサンプルを確率的に交換するアルゴリズムである。

REMC 法では高温の効果（高温では、 β_l が小さいので、

$E(\theta)$ によらず様々な領域をサンプルできる。) を取り入れることにより, $E(\theta)$ の局所最小解を脱して効率的にサンプリングを行うことができる. REMC 法を用いるもう一つの利点は, モデル選択の指標として用いる自由エネルギー $F(M)$ を REMC 法により得られるサンプルから計算できることである. 具体的に, 自由エネルギーは以下のように各系ごとの $E(\theta)$ の期待値の β 積分に書き直すことができる.

$$F(M) = \int_0^1 \langle E(\theta) \rangle_{p_{\beta}(\theta)} d\beta. \quad (17)$$

各系ごとの $E(\theta)$ の期待値 $\langle E(\theta) \rangle_{p_{\beta}(\theta)}$ は REMC 法を用いて得られるサンプルから即座に計算でき, β の積分も区分求積法などを用いて実行することができる.

4. 数値実験の結果・考察

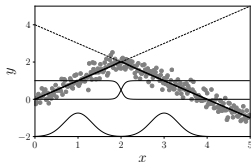


図 3: 人工データの例

Fig. 3 An example of artificial data.

表 1: パラメータの真値
Table 1 the true parameters.

	True	
	$i = 1$	$i = 2$
w_i	1.0	-1.0
b_i	0.0	4.0
μ_i	1.0	3.0
s_i	10.0	10.0

数値実験で用いたデータの一例を図 3 に示す. 入力と出力はともに 1 次元のデータであり, 入力空間 $0.0 \leq x \leq 5.0$ の範囲で等間隔に $n = 250$ 個のデータを生成した. 領域数は $M = 2$ で, パラメータは表 1 のように設定した. 図の点線がエキスパート関数, 実線がゲート関数を表している. 図下部に示したものはゲートのガウス関数である. 領域の切り替わり点は領域ごとのゲートのガウス関数の交点であり, 今回の場合, 切り替わり点は $x = 2.0$ である.

実験条件は次のように設定した. モンテカルロステップ数は 20,000 ステップ, burn-in 期間は 10,000 ステップと設定した. 逆温度 β_l は Nagata らの論文を参考に, 以下の式に従うように設定した [12].

$$\beta_l = \begin{cases} 0.0 & (l = 1) \\ \gamma^{l-L} & (l \neq 1) \end{cases} \quad l = 1, 2, 3, \dots, L \quad (18)$$

今回は $\gamma = 1.7$, 温度数 $L = 20$ とした. また, ハイパーパラメータは $\{s', v_0, m_0, \gamma_0, \nu_0, \eta_0\} = \{16, 0.01, 2.5, 0.05, 5.5, 0.5\}$ と設定した. 今回, 領域数 M に関する事前分布 $p(M)$ は, $M = 1$ から $M = 5$ の範囲までの一様分布であるとして $p(M) = 1/5$ と設定した.

4.1 モデル選択と MAP 解による回帰の一例

図 4 は図 3 に示したデータに対してモデル選択を行なった結果である. モデル選択では, $M = 1$ から $M = 5$ のモデルまでの自由エネルギーを計算し, 自由エネルギーが最小となる M を選択した. 図 4 は自由エネルギー $F(M)$

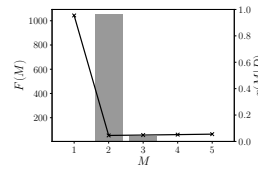


図 4: モデル選択の結果

Fig. 4 An example of result of model selection.

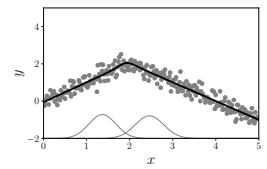


図 5: MAP 解による回帰結果

Fig. 5 An example of result of regression using MAP solution.

表 2: 真値と MAP 解

Table 2 The true parameters and MAP solutions.

	True		MAP	
	$i = 1$	$i = 2$	$i = 1$	$i = 2$
w_i	1.0	-1.0	1.060	-1.006
b_i	0.0	4.0	-0.054	4.026
μ_i	1.0	3.0	1.367	2.461
s_i	10.0	10.0	10.266	9.205

とモデルの事後確率 $p(M|D)$ の計算結果を示したものであり, 図 4 の折れ線が自由エネルギー $F(M)$, 棒グラフがモデルの事後確率 $p(M|D)$ を表している. 自由エネルギー $F(M)$ が最小になる $M = 2$ で, モデルの事後確率 $p(M|D)$ が最大となり, 真のモデルが選択された.

図 5 は, 選択された $M = 2$ のモデルを用いて推定した MAP 解による回帰結果である. 図 5 下部に示したものはゲートのガウス関数である. 表 2 は図 5 の回帰で用いたパラメータの MAP 解と真値である. この結果から, エキスパートパラメータ w, b については精度よく真値が推定できていることがわかる. 一方, 領域の切り替わりを担うゲートパラメータ μ に関しては, 他のパラメータほど推定精度は良くなかった. また, 推定したゲートパラメータ μ はより領域の真の切り替わり点 ($x = 2.0$) の近くに寄る傾向が見られた.

上記の結果は一つのデータのみに対して行なったものであるため, 上記の結果だけでは提案手法のモデルやパラメータの推定精度を評価することができない. 提案手法のモデルやパラメータの推定精度については後の 4.3 小節で議論する.

4.2 事後分布の形状

事後分布の形状を表す一例として図 6 を示す. 図 6 は REMC 法によって得られたサンプルを用いて描いた 2 次元ヒストグラムである. 図 6(a) は $M = 2$ のモデルで推定した場合の w_1, w_2 の事後分布であり, 図 6(b) は $M = 3$ のモデルで推定した場合の w_1, w_3 の事後分布である. $M = 2$ のモデルで推定した場合には, w_1, w_2 の真値の周りにサンプルが分布していた. 一方, $M = 3$ のモデルで推定した場合には, 分布の形状は $M = 2$ の場合の分布とは大きく

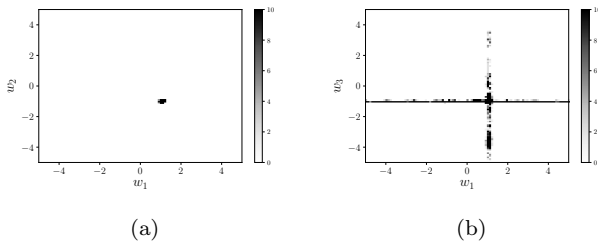


図 6: (a) : $M = 2$ のモデルで推定した場合の w_1, w_2 の事後分布
(b) : $M = 3$ のモデルで推定した場合の w_1, w_3 の事後分布

Fig. 6 (a):Posterior distribution of w_1, w_2 estimating with $M = 2$ model, (b):Posterior distribution of w_1, w_3 estimating with $M = 3$ model.

異なっており、真値付近を交点とする十字の構造が見られた。この十字の構造は特異モデルに特有の事後分布の特徴である。 w_1, w_3 以外のパラメータも考慮して分布を考えると、 $M = 3$ の場合はより複雑な構造をした分布となることが考えられる。

$M = 2$ の場合では、サンプルが真値周りに一点で集まっていたことから事後分布をガウス分布に近似することができると考えられるが、 $M = 3$ の場合では、そのような形状には全くなっておらず事後分布をガウス分布に近似することは難しいと考えられる。つまり、真のモデルと推定に使うモデルが一致している場合でないとガウス分布に近似することは適切でないと考えられる。そのため、ME モデルにおいて変分ベイズ法を行う際に、近似事後分布をガウス分布に近似することも適切ではないと考えられる。また、この結果から、正則モデル（データ数が十分に大きいとガウス分布に収束する性質を持つ）を仮定している AIC や BIC などの情報量基準を用いてモデル選択をすることは適切でないということが分かる [1]。

4.3 50 個のデータに対するモデル選択と MAP 解

表 3: 50 個のデータをモデル選択した結果
Table 3 The results of model selection of 50 data.

M	1	2	3	4	5
選ばれた数	0	50	0	0	0

図 3 に示した人工データについてそれぞれ乱数のシードを変えて 50 個のデータを生成し、それらのデータ全てに対してモデル選択を行った。モデル選択の結果を表 3 に示す。この結果から、50 個すべてのデータに対して真のモデル $M = 2$ が選ばれていることが確認された。これは、本手法によって精度よくモデル選択が可能であることを示す。

次に、50 個のデータそれぞれで MAP 解を計算し、それらの値を用いて描いたヒストグラムを以下の図 7 に示す。図 7(a) - (d) は MAP 解のヒストグラムである。図 7(e) は MAP 解を用いて求めた領域の切り替わり点のヒストグラムである。図 7 の点線は真値を表す。エキスパートパラ

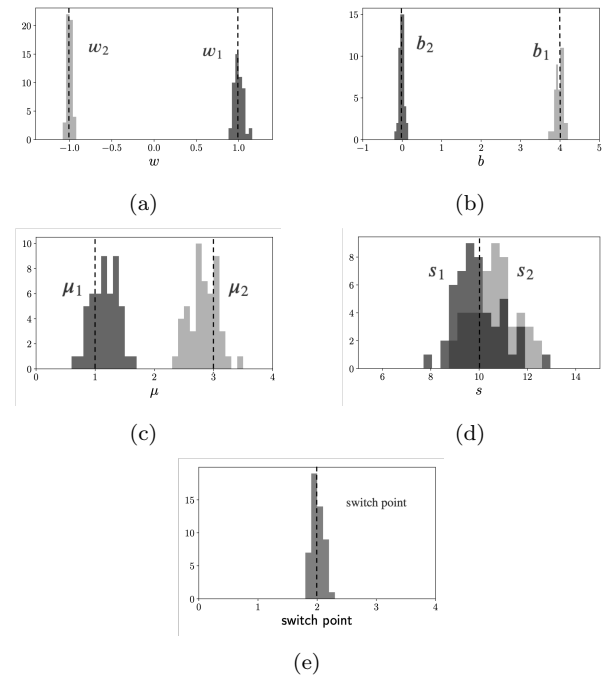


図 7: MAP 解のヒストグラム (a) : w_1, w_2 (b) : b_1, b_2 (c) : μ_1, μ_2 (d) : s_1, s_2 (e) : MAP 解で求めた切り替わり点のヒストグラム

Fig. 7 Histogram of MAP solution of (a): w_1, w_2 (b): b_1, b_2 (c): μ_1, μ_2 (d): s_1, s_2 (e) Histogram of switching points obtained by MAP solution.

メータ w, b の MAP 解は真値を中心に小さなばらつきで推定されていた (図 7(a),(b))。この結果から、ノイズによらず真値を高い精度で安定して推定できることがわかった。一方、ゲートパラメータ μ, s の MAP 解はエキスパートパラメータの MAP 解よりばらつきが大きく (図 7(c),(d))、エキスパートパラメータほどの推定の精度と安定性は望めない。図 7(c) より、 μ_1, μ_2 共に、領域の真の切り替わり点 $x = 2.0$ に寄って推定されやすい傾向があることがわかった。また、ゲートパラメータ自体の推定精度は高くないが、図 7(e) より、領域の切り替わり点は真値を中心に小さなばらつきで推定されており、ノイズによらず真値を高い精度で安定して推定できることがわかった。

5. おわりに

本論文では、REMC 法を用いて、ME モデルの自然な確率モデルにおいてベイズ推論を行う手法を提案した。数値実験では、モデル選択と事後分布の推定を行なった。モデル選択は REMC 法によって計算される自由エネルギーを指標とし、数値実験により精度よく正しいモデルが選ばれていることを確認した。また、REMC 法により推定したパラメータの事後分布に特異モデルに特有の特徴が捉えられることを確認することで、ベイズ推論が正しく行えることを確認した。

本論文で提案する手法を用いると、NGnet の自然な確率モデルにおいて推定を行うことができる。一方、NGnet に

おいて変分ベイズ法を用いて推定を行っている先行研究では、本論文の第2節で紹介した NGnet の自然な確率モデルとは異なる確率モデルを用いている [2], [3], [4]。以下では、その確率モデルの違いについて触れる。

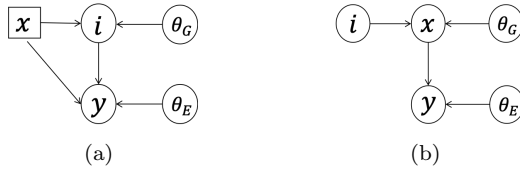


図 8: (a): NGnet モデルの既存のグラフィカル表現 (b): 変分ベイズ法で用いられる NGnet モデルのグラフィカル表現

Fig. 8 (a): “natural” Graphical representation of NGnet, (b): Graphical representation of NGnet used in variational bayesian methods.

図 8 は NGnet のグラフィカルモデルを示したものである。図 8 も図 2 と同様に、確率変数であるものは丸で囲まれており、確率変数でないものは四角で囲まれている。図 8(a) のモデルが NGnet の自然なグラフィカルモデルであり、図 2 で示したものと同一のものである。第 2 節で図 8(a) に示したこのモデルは、入力空間を分割するという分割統治の考え方を反映しているモデルであることを述べた。

図 8(b) は変分ベイズ法で推定する場合に用いられる確率モデルである。図 8(b) から分かるように、このモデルでは、領域 i とゲートのパラメータ θ_G から入力 x が生成される。そして、生成された入力 x とエキスパートのパラメータ θ_E から出力 y が決定される。このモデルでは入力 x は確率変数として考える。この時の確率モデルは以下のように定式化できる。

$$p(x, y, i | \theta) = p(y | x, i, \theta_E) p(x | i, \theta_G) p(i | \theta) \quad (19)$$

このモデルでは NGnet の自然な確率モデルとは違い、領域 i から入力 x が生成されると考えるので、入力空間を分割するという分割統治の考え方を反映していないことがわかる。

変分ベイズ法を用いる際に式 (19) に従うモデル化を行うのは、計算が困難になるのを避けるためである。式 (2) のように、ゲート関数 $G_i(x)$ に等しい $p(i | x, \theta_G)$ が確率モデルに含まれると、変分ベイズ法において必須である近似事後分布を解析的に正規化することが困難である。

NGnet において、式 (19) に従うモデル化を行うと、確率モデルが混合ガウス分布の特殊な場合として定式化される [2], [3], [4]。この確率モデルにおいて、式 (19) の $p(y | x, i, \theta_E)$ と $p(x | i, \theta_G)$ はガウス分布になる。先行研究では、ゲート関数 $G_i(x)$ に混合比パラメータ g_i を導入し、 $p(i | \theta) = g_i$ として定式化している [2], [3], [4]。具体的には以下のようにゲート関数 $G_i(x)$ に混合比パラメータ g_i を導入する。

$$G_i(x) = \frac{g_i \mathcal{N}(x | \mu_i, s_i)}{\sum_{j=1}^M g_j \mathcal{N}(x | \mu_j, s_j)} \quad (20)$$

このようにゲート関数に混合比パラメータを導入する方法は、Xu らによって提案された [13]。

変分ベイズ法では上記のように確率モデルを変えることによって推定が行われているが、確率モデルを変えることで生じる推定結果への影響について言及している研究はなく、今後その影響について調べるのが一つの課題であるといえる。

謝辞 本研究の一部は、科学研究費補助金基盤研究 (A)(課題番号 18H04106)、科学技術振興機構 CREST(課題番号 JPMJCR1761)、さががけ (課題番号 JPMJPR15E8) の補助を受けて実施されたものである。

参考文献

- [1] Watanabe, Sumio. “Algebraic analysis for nonidentifiable learning machines.” *Neural Computation* 13.4 (2001): 899-933.
- [2] 吉本潤一郎, 石井信, 佐藤雅昭: 変分法的ベイズ推定法に基づく正規化ガウス関数ネットワークと階層的モデル選択法. 計測自動制御学会論文集 Vol.39, No.5, 503-512 (2003).
- [3] Ueda, Naonori, and Zoubin Ghahramani. “Bayesian model search for mixture models based on optimizing variational bounds.” *Neural Networks* 15.10 (2002): 1223-1241.
- [4] Baldacchino, Tara, et al. “Variational Bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems.” *Mechanical Systems and Signal Processing* 66 (2016): 178-200.
- [5] Waterhouse, Steve R., David MacKay, and Anthony J. Robinson. “Bayesian methods for mixtures of experts.” *Neural Information Processing Systems* 7 (1996): 351-357.
- [6] Jacobs, Robert A., et al. “Adaptive mixtures of local experts.” *Neural computation* 3.1 (1991): 79-87.
- [7] Yuksel, Seniha Esen, Joseph N. Wilson, and Paul D. Gader. “Twenty years of mixture of experts.” *IEEE transactions on neural networks and learning systems* 23.8 (2012): 1177-1193.
- [8] Hukushima, Koji, and Koji Nemoto. “Exchange Monte Carlo method and application to spin glass simulations.” *Journal of the Physical Society of Japan* 65.6 (1996): 1604-1608.
- [9] Nagata, Kenji, Seiji Sugita, and Masato Okada. “Bayesian spectral deconvolution with the exchange Monte Carlo method.” *Neural Networks* 28 (2012): 82-89.
- [10] Tokuda, Satoru, Kenji Nagata, and Masato Okada. “Simultaneous Estimation of Noise Variance and Number of Peaks in Bayesian Spectral Deconvolution.” *Journal of the Physical Society of Japan* 86.2 (2016): 024001.
- [11] Ogata, Yoshihiko. “A Monte Carlo method for an objective Bayesian procedure.” *Annals of the Institute of statistical Mathematics* 42.3 (1990): 403-433.
- [12] Nagata, Kenji, and Sumio Watanabe. “Asymptotic behavior of exchange ratio in exchange Monte Carlo method.” *Neural Networks* 21.7 (2008): 980-988.
- [13] Xu, Lei, Michael I. Jordan, and Geoffrey E. Hinton. “An alternative model for mixtures of experts.” *Neural Information Processing Systems* 7 (1995): 633-640.