6M-02

経験ベイズ法の統計力学的解析の一般化と性能検証

高橋隼汰 † 安田宗樹 ‡
山形大学大学院理工学研究科 † 山形大学大学院理工学研究科 ‡

1．はじめに

マルコフ確率場 (Markov random field (MRF)) は確率変数同士の関連性をグラフによって示しているグラフィカルモデルである．その中でボルツマンマシン (Boltzmann machine (BM)) は MRF の最も単純な場合の一つである．MRF は確率変数やその相互作用に対応するパラメータによって確率を表現でき, MRF のパラメータが既知でない場合にそのパラメータを推定する問題を MRF の機械学習と呼ぶ.

この MRF のパラメータを推定するのではなく，MRF のパラメータを支配する事前分布のパラメータ (ハイパーパラメータ) を推定する手法が先行研究 [1] において提案されている．ハイパーパラメータの推定は正則化係数に役立てることができる．先行研究では，ハイパーパラメータを条件とした周辺尤度関数 (経験ベイズ尤度関数) を最大化することで推定が行っている．この手法を図 1 に示している経験ベイズ法 [2][3] と呼ぶ．しかし，経験ベイズ尤度関数には多重積分が含まれるため計算が困難となる．この問題を解決するために，統計的力学的解析手法であるレプリカ法 [4][5] とプレフカ展開 [6] を用いた近似計算が行われている．しかし, 先行研究では離散値の確率変数 $x_i \in \{-1, +1\}$ を用いた全結合 BM とした限定的な条件で行われている. 本研究では, 任意の標本空間と任意のグラフ構造をもった BM に適用できるように，先行研究の解析手法を拡張し，アルゴリズムを導出することを目的とする．さらに，提案アルゴリズムのハイパーパラメータの推定性能を数値実験を通して検証する.

2．BM と経験ベイズ法

n 個のノード $V := \{1, 2, ..., n\}$ からなる無向グラフ $G(V, E)$ を考える．ここで，E は頂点 i, j を繋ぐ無向リンク $\{i, j\}$ の集合である．ノードに対応させた標本空間 $\mathcal{X}$ の確率変数 $\boldsymbol{x} = \{x_i \in \mathcal{X} \mid i \in V\}$ を用いて，BM は次の確率モデルで定義される.

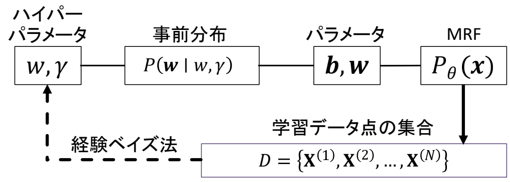


図 1: 経験ベイズ法によるハイパーパラメータ指定の図解

$$P_\theta(\boldsymbol{x}) := \frac{1}{Z_\theta} \exp\left(\sum_{k=1}^{K} b_k \sum_{i \in V} x_i + \sum_{\{i,j\} \in E} w_{i,j} x_i x_j\right) \quad (1)$$

ここで，$\boldsymbol{b} = \{b_k \in \mathbb{R} \mid k = 1, 2, ..., K\}$，$\boldsymbol{w} = \{w_{i,j} \in \mathbb{R} \mid \{i, j\} \in E\}$ であり，Z_θ は規格化定数であり，パラメータ $\boldsymbol{b}, \boldsymbol{w}$ をまとめて θ と表記する．また，式 (1) 中の $\phi_i(x) \in \mathbb{R}$ は任意の関数である．相互作用パラメータの事前分布が以下のガウス分布であると仮定する.

$$P(\boldsymbol{w}) := \prod_{\{i,j\} \in E} \mathcal{N}\left(w_{i,j} \mid w, \gamma\right)$$

$$\mathcal{N}(w_{i,j} \mid w, \gamma) := \frac{1}{\sqrt{2\pi\gamma}} \exp\left(-\frac{(w_{i,j} - \mu)^2}{2\gamma}\right) \quad (2)$$

ここで，w, γ はハイパーパラメータであり，パラメータ $\boldsymbol{b}$ に関しては事前分布を設定しないとする．このハイパーパラメータを推定することが本研究の目的となる.

経験ベイズ法を用いてパラメータ $\boldsymbol{b}$ とハイパーパラメータを推定するために，経験ベイズ尤度関数を定義する．N 個の学習データ点の集合 $D := \{\mathbf{x}^{(\mu)} \in \mathcal{X}^n \mid \mu = 1, 2, ..., N\}$ が与えられたとき，経験ベイズ尤度関数を

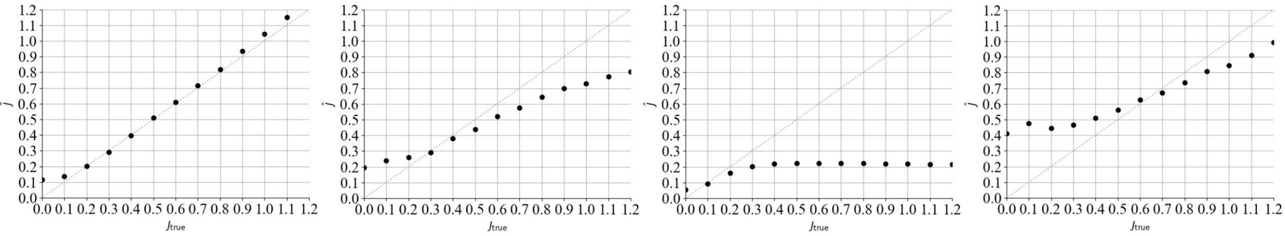
$$\ell_{\mathrm{EB}}(h) := \frac{1}{nN} \ln \int_{-\infty}^{+\infty} d\boldsymbol{w} P_\theta(D) P(\boldsymbol{w}) \quad (3)$$

と定義する．ここで，$P_\theta(D) := \prod_{\mu=1}^{N} P_\theta(\mathbf{x}^{(\mu)})$ であり，パラメータ $\boldsymbol{b}$ とハイパーパラメータをまとめて h と表記する.

3．統計力学的解析手法を用いた経験ベイズ尤度関数の解析

式 (3) の経験ベイズ尤度関数の多重積分は計算が困難である．そこで，統計力学的解説手法であるレプリカ法

Generalization and performance verification of statistical mechanical analysis for empirical Bayes method
† Syunta Takahasi; Graduate School of Science and Engineering, Yamagata University
‡ Muneki Yasuda; Graduate School of Science and Engineering, Yamagata University



(a) 全結合 BM　**(b)** 正方格子 BM　**(c)** 全結合 BM　**(d)** 正方格子 BM

図 2: J_{true} に対する $\hat{J}$ の結果

を用いる．レプリカ法の考えより，r を自然数とした規格化定数 Z_θ^r を定義し，Z_θ^r を r 個のレプリカとして考え計算を行う．r を実数として扱い，式 (3) を変形すると，

$$\ell_{\mathrm{EB}}(h) := \frac{1}{nN} \ln \lim_{r\to -1} \Gamma_r(h) \tag{4}$$

$$\Gamma_r(h) := \exp\left(nN \sum_{k=1}^{K} b_k M_k + \frac{|E|N^2\gamma}{2} C_2 - F_r(h) \right) \tag{5}$$

となる．ここで M_k, C_k は定数であり，

$$F_r(h) := -\ln \sum_{X_r} \exp\left(-E_r(X_r; h)\right) \tag{6}$$

はレプリカ拡張系での自由エネルギーである．

式 (6) の自由エネルギーには相互作用の項が含まれているため，このままでは計算困難である. そこで自由エネルギーの双対の関係であるギブス自由エネルギーを導出し，ギブス自由エネルギーを計算することで自由エネルギーを求める．ギブス自由エネルギーにも相互作用項が含まれるため，統計力学的解決手法であるプレフカ展開を用いて近似を行う．プレフカ展開は解析が簡単な状況を中心とし近似を行う手法であり，今回はギブス自由エネルギーの相互作用項が消失させると計算が容易になるため，相互作用項が消失するようにをテイラー展開し，2 次項まで計算することでギブス自由エネルギーの近似を計算する. このギブス自由エネルギーの近似から式 (3) の経験ベイズ尤度関数の近似を得る．この経験ベイズ尤度関数の近似を最大化することでパラメータ $\boldsymbol{b}$ とハイパーパラメータの推定値を得ることが出来る.

4．数値実験

式 (2) の事前分布の真のハイパーパラメータを w_{true}, γ_{true} とし，この事前分布から得たパラメータを用いて実験に用いる全結合の BM と正方格子の BM を作成する．この実験では $J := \sqrt{\gamma/N}$, $\alpha := N/n$ という表記法を用いる．これら BM から実験に用いるデータセット D を生成し，提案した手法を用いてハイパーパラメータの推定値を得る．真のハイパーパラメータ J_{true} と推定値 $\hat{J}$ を比較し，推定の精度を確認する.

$x_i \in \{-1, 0, +1\}$, $n = 300, \alpha = 0.4$, $\boldsymbol{b} = 0$, $w_{\mathrm{true}} = 0$ と設定を行い，100 回の実験を行った結果の平均値を図 2(a)(b) に示す．また，図 (c)(d) には $x_i \in \{0, +1\}$ に変更したものを示す. 図 2(a) より，$0.2 < J_{\mathrm{true}} < 0.8$ のとき，$\hat{J}$ は J_{true} とよく一致している．図 2(b) より，$\hat{J}$ は J_{true} の値を下回ってしまう. 今回の手法において BM のグラフ構造は大きくハイパーパラメータの推定値に影響を与えることが分かる．また，図 (c)(d) より，確率変数の値がゼロ対称でない場合，推定性能が大きく劣化することが分かる.

5．まとめ

先行研究の統計力学的解析 [1] を拡張し、任意の標本空間とグラフ構造をもつ BM に適用可能な経験ベイズ法を導出し，更に，提案法の推定性能を数値実験を通して調べた．その結果，提案アルゴリズムは，標本空間がゼロ対称であり，かつ，グラフ構造が蜜である場合に有効であることが分かった．そして，そうでない場合は推定性能が大きく劣化した．この性能劣化を低減することが今後の課題である.

謝辞

本研究は科研費（21K11778）の助成を受けたものである.

文献

[1] M. Yasuda and T. Obuchi: Journal of Physics A: Math. Theor. 53 014004(2019)

[2] MacKay D J C Neural Comput. 4 415–47(1992)

[3] Bishop C M Pattern Recognition and Machine Learning (Berlin: Springer,2006)

[4] Mezard M, Parisi G and Virasoro M Spin Glass Theory and Beyond: an Introduction to the Replica Method and Its Applications (Singapore: World Scientific,1987)

[5] Nishimori H Statistical Physics of Spin Glass and Information Processing—Introduction (Oxford: Oxford University Press,2001)

[6] Plefka T J. Phys. A: Math. Gen. 15 1971–8(1982)