1M-04

経験ベイズ法を用いた連続値ボルツマンマシンのハイパーパラメータ推定

加藤 航太 † 安田 宗樹 ‡
山形大学 † 山形大学 ‡

1 序論

ボルツマンマシン (Boltzmann Machine) 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り，統計的機械学習で用いられるモデルの一つである．ボルツマンマシンのパラメータの適切な値を推定する問題を，ボルツマンマシンの機械学習 (Boltzmann Machine Learning:BML)[2] と呼び，BML は機械学習や統計力学の分野で積極的に研究されている．本研究では，ボルツマンマシンのパラメータを推定するのではなく，ボルツマンマシン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）を推定する手法を提案する．ハイパーパラメータの推定は，正則化された BML の正則化係数の決定など様々な事に応用することができる．

本研究は，離散値の確率変数 $S_i \in \{-1, +1\}$ を用いた離散値のボルツマンマシンのハイパーパラメータを推定する先行研究 [1] を，連続値ボルツマンマシンへ拡張することを目的とする．通常の BML では，観測データ点の集合が与えられたときに，パラメータを条件とした観測データ点の集合に対するモデルの尤度関数を考え，この尤度関数を最大にするパラメータの値を推定値とする最尤推定という手法を用いて行われる．これに対し，本研究の目的であるハイパーパラメータの推定では，ハイパーパラメータを条件とした周辺尤度関数（本研究では経験ベイズ尤度関数と呼ぶ）を最大化することでハイパーパラメータを推定する．この手法を経験ベイズ法 [3][4] と呼ぶ．通常の BML と本研究の経験ベイズ法の関係を図 1 に示す．経験ベイズ尤度関数には解析が困難な多重積分が含まれるため最大化の計算は容易でない．本研究では，統計力学的解析手法であるレプリカ法 [5][6] とプレフカ展開 [7] を用いることで計算の困難性を回避し，単純かつ高速な経験ベイズ尤度関数の最大化の手法を提案する．

2 連続値ボルツマンマシンにおける経験ベイズ法

n 個の連続値の確率変数 $\boldsymbol{S} := \{S_i \in (-\infty, +\infty) \mid i = 1, 2, ..., n\}$ を持つ全結合ボルツマンマシンを

$$P(\boldsymbol{S} \mid h, \lambda, \boldsymbol{J}) := \frac{1}{Z(h, \lambda, \boldsymbol{J})} \exp\left(h \sum_{i=1}^{n} S_i - \frac{\lambda}{2} \sum_{i=1}^{n} S_i^2 + \sum_{i<j} J_{ij} S_i S_j\right)$$

と定義する．ここで，$h \in (-\infty, +\infty)$，$\lambda \in (0, +\infty)$，$\boldsymbol{J} := \{J_{ij} \in (-\infty, +\infty) \mid i < j\}$ であり，$Z(h, \lambda, \boldsymbol{J})$ は分配関数である．パラメータを支配する事前分布をそれぞれ

$$P_{\mathrm{prior}}(h \mid H) := \delta(h - H)$$

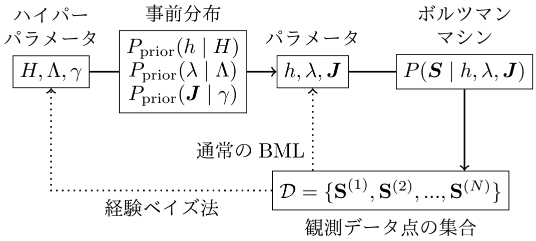

図 1: 経験ベイズ法によるハイパーパラメータ推定の図解

$$P_{\mathrm{prior}}(\lambda \mid \Lambda) := \delta(\lambda - \Lambda)$$
$$P_{\mathrm{prior}}(\boldsymbol{J} \mid \gamma) := \prod_{i<j} P_{\mathrm{prior}}(J_{ij} \mid \gamma)$$
$$P_{\mathrm{prior}}(J_{ij} \mid \gamma) := \sqrt{\frac{n}{2\pi\gamma}} \exp\left(-\frac{nJ_{ij}^2}{2\gamma}\right), \quad \gamma > 0$$

と定義する．ここで $\delta(x)$ はディラックのデルタ関数，H, Λ, γ はハイパーパラメータであり，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 [3][4] を用いて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 個の観測データ点の集合 $\mathcal{D} := \{\mathbf{S}^{(\mu)} \in (-\infty, +\infty)^n \mid \mu = 1, 2, ..., N\}$ が与えられたとき，経験ベイズ尤度関数を

$$L_{\mathrm{EB}}(H, \Lambda, \gamma) := \frac{1}{nN} \ln \left[\prod_{\mu=1}^{N} P(\mathbf{S}^{(\mu)} \mid h, \lambda, \boldsymbol{J})\right]_{h,\lambda,\boldsymbol{J}} \tag{1}$$

と定義する．ここで

$$[\cdots]_{h,\lambda,\boldsymbol{J}} := \int d\boldsymbol{J} \int dh \int d\lambda (\cdots) \times P_{\mathrm{prior}}(\lambda \mid \Lambda) P_{\mathrm{prior}}(h \mid H) P_{\mathrm{prior}}(\boldsymbol{J} \mid \gamma)$$

である．式 (1) の経験ベイズ尤度関数をハイパーパラメータに対して最大化することでハイパーパラメータの推定値を得る．

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 (1) の経験ベイズ尤度関数の多重積分は解析が困難である．そこで，統計力学的解析手法を用いて計算を進める．統計力学的解析手法であるレプリカ法 [5][6] の考え方を利用し経験ベイズ尤度関数を変形すると，

$$L_{\mathrm{EB}}(H, \Lambda, \gamma) = \frac{1}{nN} \ln \lim_{x \to -1} \Psi_x^{\mathrm{gauss}}(H, \Lambda, \gamma)$$

$$\Psi_x^{\mathrm{gauss}}(H, \Lambda, \gamma) := \exp\left\{nNHM - \frac{1}{2}nN\Lambda V\right.$$

Hyperparameter estimation in continuous Boltzmann machine using empirical Bayes method
† Kota Kato, Yamagata University
‡ Muneki Yasuda, Yamagata University

$$+\frac{\gamma N^2(n-1)}{4}C_2 - F_x(H,\Lambda,\gamma)\Big\}$$

となる [1]．ここで M, V, C_k は観測データ点の集合から計算される定数であり，

$$F_x(H,\Lambda,\gamma) := -\ln\int d\mathcal{S}_x \exp\{-E_x(\mathcal{S}_x;H,\Lambda,\gamma)\}$$

はレプリカ拡張系での自由エネルギーである．ここで，

$$E_x(\mathcal{S}_x;H,\Lambda,\gamma) := -H\sum_{i=1}^{n}\sum_{a=1}^{\tau_x} S_i^{\{a\}} + \frac{\Lambda}{2}\sum_{i=1}^{n}\sum_{a=1}^{\tau_x}(S_i^{\{a\}})^2 - \frac{\gamma N}{n}\sum_{i<j} d_{ij}\sum_{a=1}^{\tau_x} S_i^{\{a\}}S_j^{\{a\}} - \frac{\gamma}{2n}\sum_{i<j}\sum_{a=1}^{\tau_x}(S_i^{\{a\}}S_j^{\{a\}})^2 - \frac{\gamma}{n}\sum_{i<j}\sum_{a<b} S_i^{\{a\}}S_j^{\{a\}}S_i^{\{b\}}S_j^{\{b\}}$$

であり，$a, b \in \{1, 2, ..., \tau_x\}$ はレプリカ番号，$\tau_x := xN$，$\int d\mathcal{S}_x := \prod_{i=1}^{n}\prod_{a=1}^{\tau_x}\int dS_i^{\{a\}}$ である．

この自由エネルギーを計算するために自由エネルギーの双対系であるギブス自由エネルギーを導出する．ギブス自由エネルギーは

$$G_x(m,v,H,\Lambda,\gamma) := -Hn\tau_x m + \frac{1}{2}\Lambda n\tau_x v + \mathop{\mathrm{extr}}_{\mu,\epsilon}\Big[\mu n\tau_x m - \frac{1}{2}\epsilon n\tau_x v - \ln\int d\mathcal{S}_x \exp\left(-E_x(\mathcal{S}_x;\mu,\epsilon,\gamma)\right)\Big]$$

として得られる．ギブス自由エネルギーはこのままでは計算量の問題により解析できないので，ギブス自由エネルギーの近似を統計力学的解析手法であるプレフカ展開 [7] を用いて得る．プレフカ展開を用いると，ギブス自由エネルギーの近似は，ギブス自由エネルギーを $\gamma = 0$ の周りでテイラー展開し，2 次項まで計算することで得ることができる．ギブス自由エネルギーの 2 次近似は

$$\frac{G_x(m,v,H,\Lambda,\gamma)}{nN} \approx \frac{1}{nN}G_x(m,v,H,\Lambda,0) + \phi_x^{(1)}(m,v)\gamma + \phi_x^{(2)}(m,v)\gamma^2$$

となる．このギブス自由エネルギーの近似から式 (1) の経験ベイズ尤度関数の近似

$$L_{\mathrm{EB}}(H,\Lambda,\gamma) \approx HM - \frac{1}{2}(\Lambda V + \ln 2\pi) - \mathop{\mathrm{extr}}_{m,v}\Big[Hm - \frac{1}{2}\Lambda v - e(m,v) + \Phi(m,v)\gamma + \phi_{-1}^{(2)}(m,v)\gamma^2\Big] \tag{2}$$

を得る．ここで，$\Phi(m,v) := \phi_{-1}^{(1)}(m,v) - \frac{(n-1)N}{4n}C_2$，$e(m,v) := -\frac{1}{2}(1+\ln(v-m^2))$ であり，$\phi_x^{(1)}(m,v)$，$\phi_x^{(2)}(m,v)$ はプレフカ展開の展開係数である．

式 (2) の経験ベイズ尤度関数の近似を最大化することでハイパーパラメータの推定値はそれぞれ

$$\hat{H} = \frac{M}{V-M^2} - \left(\frac{\partial\phi_{-1}^{(1)}(m,v)}{\partial m}\gamma + \frac{\partial\phi_{-1}^{(2)}(m,v)}{\partial m}\gamma^2\right)\Bigg|_{m=M,v=V}$$

$$\hat{\Lambda} = \frac{1}{M^2-V} + \left(\frac{\partial\phi_{-1}^{(1)}(m,v)}{\partial v}\gamma + \frac{\partial\phi_{-1}^{(2)}(m,v)}{\partial v}\gamma^2\right)\Bigg|_{m=M,v=V}$$

$$\hat{\gamma} = \arg\max_{\gamma}\Big[-\Phi(M,V)\gamma - \phi_{-1}^{(2)}(M,V)\gamma^2\Big]$$

として得られる．

4　数値実験

事前分布の真のハイパーパラメータを H_{true},J_{true},Λ_{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ボルツマンマシンを作成する．この実験では $J := \sqrt{\gamma}$，$\alpha := N/n$ という表記法を用いる．このボルツマンマシンから実験に用いるデータセット $\mathcal{D}$ 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と推定値を比較し，推定の精度を確認する．

$n = 300, \alpha = 0.5, H_{\mathrm{true}} = 0, \Lambda_{\mathrm{true}} = 10$ で 100 回の実験を行った結果の平均値を図 2 に示す．図から今回の手法を用いることでハイパーパラメータを正しく推定できていることがわかる．

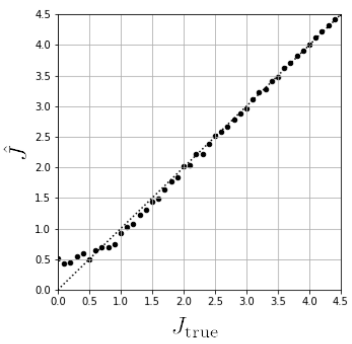

図 2: J_{true}(横軸) と $\hat{J}$(縦軸) の散布図.

5　まとめと今後の課題

連続値ボルツマンマシンにおける経験ベイズ法を考え，統計力学的解析手法であるレプリカ法とプレフカ展開を用いて経験ベイズ尤度関数を解析することで，単純かつ高速なハイパーパラメータの推定アルゴリズムを提案した．

実験により α の値が適切でないと精度の高い推定ができないことが分かった．今後の課題として，この α を決定する方法を考える必要がある．今回の実験は理論のモデルと生成モデルが同じ状況で行ったが，理論のモデルと違う生成モデルを用いた実験を行うことも重要である．また，事前分布として別の分布を使用した場合の拡張や，他のタイプのボルツマンマシンへの拡張も行っていきたい．

参考文献

[1] M. Yasuda and T. Obuchi, Journal of Physics A: Mathematical and Theoretical, 53,1,014004 (2019).
[2] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognitive Science **9**, 147 (1985).
[3] D. J. C. MacKay, Neural Computation **4**, 415 (1992).
[4] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).
[5] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* (Singapore, World Scientific, 1987).
[6] H. Nishimori, *Statistical Physics of Spin Glass and Information Processing -Introduction-* (Oxford University Press, 2001).
[7] T. Plefka and J. Phys. A, Mathematical and General **15**, 1971 (1982).