

ビリーフプロパゲーション ～統計力学の視点から～

樺島 祥介*

Yoshiyuki Kabashima

Abstract: ビリーフプロパゲーション (BP) 法と類似した枠組みとして物理学では平均場近似法という計算技法が知られている。近年、特に近似計算の文脈で両枠組みの類似性に着目した研究が増えているが、実験や観測等、計算量爆発の問題が生じない代替手段により間接的にであっても結果の検証が可能な物理モデルを対象としてきた平均場近似法の研究は BP 法のそれとは一味違うこともまた事実である。特に、要素数を無限大にした極限で解析的な取り扱いが容易になる「平均場モデル」に関する研究は大自由度系に対する物理学的接近法を最も特徴付けるものであろう。本稿では、物理学の視点を際立たせるため、平均場モデルの典型例である連想記憶モデルへの BP 法の適用を考察する。それ自体は情報処理としてさほど意味のあるモデル解析ではないが、i) 計算論的には計算爆発を引き起こすと考えられる完全 2 部グラフに対する BP 法が物理的考察によって高々要素数の自乗程度の計算量まで削減され、ii) しかも要素数を増やすほど計算結果が厳密（と予想される）解に近づく例となっており、より実際的な問題の解決法に対して何らかのヒントを与えるかもしれない。

1 連想記憶モデル

まず、本稿で考察するホップフィールドの連想記憶モデルを導入しよう [1, 2]。各成分が 2 値 ± 1 で表される P 個の N 次元パターンベクトル

$$\xi^\mu = (\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu)^T \quad (1)$$

($\mu = 1, 2, \dots, P$) を考える。ただし、各成分 ξ_l^μ は ± 1 が等確率で独立に生成されているものとする。

ホップフィールドはヘブ則

$$J_{lk} = \begin{cases} \frac{1}{N} \sum_{\mu=1}^P \xi_l^\mu \xi_k^\mu & \text{for } l \neq k \\ 0 & \text{for } l = k \end{cases} \quad (2)$$

によりシナプス結合を定義することでこれらのパターンを N 個の 2 値ニューロン $S_{l=1,2,\dots,N} = \pm 1$ からなるニューラルネットワークに「記憶」させることが出来ることを示した [1]。もう少し具体的に言うと、パターン数 P が N に比例するある臨界値以下の場合、ニューロンをあるパターン (例えば ξ^1) に近い状態に置けば

$$S_l^{t+1} = \text{sign} \left(\sum_{k \neq l} J_{lk} S_k^t \right) \quad (3)$$

に従い繰り返し状態更新を行うことで記憶パターンに近いベクトルに状態を収束させることが出来る。これはヒ

*東京工業大学大学院知能システム科学専攻 〒226-8502 横浜市緑区長津田町 4259 tel. 045-924-5643, e-mail kaba@dis.titech.ac.jp, Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama 226-8502, Japan

トが提示された情報を手がかりに記憶を「連想」する過程を彷彿とさせる。つまり、高度な情報処理に見えるヒトの記憶や連想と似たような仕組みがこれくらい簡単な数理モデルにより構成できるという訳である。

連想記憶の研究ではより高機能な記憶や脳との対応など数多くの興味深い進展があるのだが、ここで必要なものは BP 法を適用するための統計モデルという側面だけなので式 (3) がコスト関数

$$\begin{aligned} \mathcal{H}(S|\{\xi^\mu\}) &= - \sum_{l>k} J_{lk} S_l S_k \\ &\sim - \frac{1}{2} \sum_{\mu=1}^P \left(\frac{\xi^\mu \cdot S}{\sqrt{N}} \right)^2 \end{aligned} \quad (4)$$

の最急降下として表現されることを言及するに留める。

2 ボルツマン分布とグラフ表現

つまり、連想記憶モデルにおける記憶の連想は与えられたパターンの組 $\{\xi^\mu\}$ により定まるコスト関数 (4) の局所最小化に他ならない。ただし、脳などの生体における対応を考える場合外界のノイズの影響を考えなければならない。そのような効果の特徴付けるものとして式 (4) に関する最小化の代わりにしばしばボルツマン分布

$$P(S|\{\xi^\mu\}) \propto \exp[-\beta \mathcal{H}(S|\{\xi^\mu\})] \quad (5)$$

によるモデル化がなされる [2]。ここで $\beta > 0$ はその逆数が絶対温度に対応するパラメータであり、外界のノイ

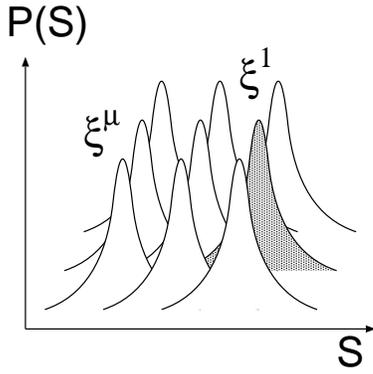


図 1: 連想記憶モデルにおけるボルツマン分布の概念図。ズの強さを表す。この定式化ではニューロンの状態 S のボルツマン分布に関する平均値が理想状態を表す。大規模統計モデルの平均値計算には一般に計算爆発が生じる。そこで、以下ではニューロン数 N 、パターン数 P を比 $\alpha = P/N \sim O(1)$ としたまま大きくした極限（つまり厳密計算が爆発する局面）でこれを BP 法により近似的に評価することを考える。

ただし、通常の意味での平均値を求める訳ではない。なぜなら、連想記憶モデルは各パターンをコスト関数の局所最小状態として記憶しているからである。ボルツマン分布に関する真の平均値を計算するとそれぞれの局所最小状態が混ざり合った意味のないものが求まってしまう。本来のモデルが局所最小状態を利用した記憶の表現をしている以上、記憶パターンに対応したボルツマン分布の局所最大付近のみの重みでの平均計算が必要となるのである（図 1）。

そのため次のような手続きを取る。例えば、ニューロンがパターン ξ^1 に近いボルツマン分布の局所最大に近い状態にあるとする。ガウス積分を用いると

$$P(S|\{\xi^\mu\}) = \int d\phi P(S|\{\xi^\mu\}, \phi) P(\phi|\{\xi^\mu\}) \quad (6)$$

ただし、

$$\begin{aligned} P(\phi|\{\xi^\mu\}) &\propto \exp\left[-\frac{N\phi^2}{2\beta}\right] \\ &\times \sum_S \exp\left[\frac{\beta}{2} \sum_{\mu=2}^P \left(\frac{\xi^\mu \cdot S}{\sqrt{N}}\right)^2 + \phi \xi^1 \cdot S\right] \\ &\equiv \exp\left[-N \left(\frac{\phi^2}{2\beta} + g(\phi|\{\xi^\mu\})\right)\right] \quad (7) \end{aligned}$$

$$P(S|\{\xi^\mu\}, \phi) \propto \exp\left[\frac{\beta}{2} \sum_{\mu=2}^P \left(\frac{\xi^\mu \cdot S}{\sqrt{N}}\right)^2 + \phi \xi^1 \cdot S\right] \quad (8)$$

のように表現される。状態がパターン ξ^1 に近くオーバーラップ $M_1 = (1/N)\xi^1 \cdot S$ が $O(1)$ の値を取る場合には $N \gg 1$ に対して補助（潜在）変数 ϕ に関する積分 (6)

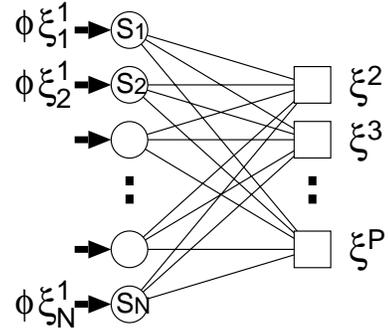


図 2: 連想記憶モデルのグラフ表現。

は局所的に被積分関数 (7) に関する鞍点法によって評価できる。つまり、補助変数 ϕ は式 (7) を最大にする確定的な値に定まり、これがパターン ξ^1 に対応するボルツマン分布の局所最大を特徴付けるのである。

そこで、 ϕ が与えられたとして条件付確率 (8) によるニューロン活動度 S の平均評価を BP 法により行おう。そのために $P(S|\{\xi^\mu\}, \phi)$

$$P(S|\{\xi^\mu\}, \phi) \propto e^{\phi \xi^1 \cdot S} \prod_{\mu=2}^P \exp\left[\frac{\beta}{2} \left(\frac{\xi^\mu \cdot S}{\sqrt{N}}\right)^2\right] \quad (9)$$

のように各パターン ξ^μ 毎の積に分解し、この依存関係をビリーフネットワークで表現する。各パターン ξ^μ は全てのニューロンと関係を持つためこのネットワークはニューロン活動度 S_1, S_2, \dots, S_N とパターン ξ^2, \dots, ξ^P が相互に全て結合した完全 2 部グラフとなる（図 2）。また、連想されるパターン ξ^1 はニューロン活動度に対する事前確率として導入される。

3 ビリーフプロパゲーション

補助変数 ϕ が与えられたとしてパターン ξ^1 の連想を表す $P(S|\{\xi^\mu\}, \phi)$ での平均を BP 法により計算しよう。そのために、ビリーフネットワーク上でのニューロン活動度とパターン間の依存関係を表す条件付確率

$$P(\xi^{\mu(\geq 2)}|S) = Y_\mu^{-1} \exp\left[\frac{\beta}{2} \left(\frac{\xi^\mu \cdot S}{\sqrt{N}}\right)^2\right] \quad (10)$$

および、事前確率

$$P(S|\xi^1, \phi) = \prod_{l=1}^N P(S_l|\xi_l^1, \phi) \equiv \prod_{l=1}^N \frac{\exp[\phi \xi_l^1 S_l]}{2 \cosh \phi} \quad (11)$$

を定義する。ただし、 Y_μ は式 (10)、を確率とするための規格化定数であり、 S に依存しない定数となることに注意。これを用いるとボルツマン分布はベイズの公式を用いて

$$P(S|\{\xi^\mu\}, \phi) = Z^{-1} P(S|\xi^1, \phi) \prod_{\mu=2}^P P(\xi^\mu|S) \quad (12)$$

と表現される． Z は規格化定数である．

さて，連想記憶モデルのピリーフネットワークは多数のループを含むので厳密計算の意味での BP 法は成り立たないのであるが，近似計算法としての BP 法（と結果として一致する計算法）は以下のように構成される．統計力学ではこのような近似法をベータ近似 [3] という．

ベータ近似はより一般的な近似計算法の枠組みである平均場近似の一つとして位置づけられる．平均場近似の基本方針は大自由度分布として定義されている確率分布 (10)，(11) および (12) を周辺化し，大自由度分布間の関係性を 1 体分布関数間のそれに縮約することである．ベータ近似ではあるニューロン活動度 S_l へのあるパターン $\xi^{\mu(\geq 2)}$ の影響を評価するために直接的な影響をあらゆる条件付確率 (10) の表現は近似なしに取り入れる一方， $\xi^{\nu \neq \mu}$ から決まる他のニューロン活動度 $S_{k \neq l}$ の結合分布を周辺分布の積で近似する．具体的には $\mu \geq 2$ として周辺化

$$\begin{aligned} & P(\xi^\mu | S_l, \{\xi^{\nu \neq \mu}\}, \phi) \\ &= \sum_{S_{k \neq l}} P(\xi^\mu | S) \prod_{k \neq l} P(S_k | \{\xi^{\nu \neq \mu}\}, \phi) \end{aligned} \quad (13)$$

により，条件付確率 (10) を S_l に関する 1 変数関数に縮約する．直感的には，近似なしに取り入れる影響の範囲を直接関係している条件付確率だけでなく間接に繋がっているものまで広げていけば近似の精度は向上しそうである．このような観点からベータ近似を一般化した枠組みはクラスター変分法 (CVM) [4] として知られている¹．

さて，周辺化された条件付確率 (13) が得られるとベイズの公式を用いることで任意のパターンの組から作られる各ニューロン活動度 S_l の 1 体分布 (= 近似された周辺分布) を求めることは容易である．特に ξ^μ 以外のパターンから作られる 1 体分布関数は

$$\begin{aligned} & P(S_l | \{\xi^{\nu \neq \mu}\}, \phi) \\ &= \alpha_{\mu l} P(S_l | \xi_l^1, \phi) \prod_{\nu \neq \mu} P(\xi^\nu | S_l, \{\xi^{\sigma \neq \nu}\}, \phi) \end{aligned} \quad (14)$$

と評価されるが，式 (13) および (14) は閉じた方程式をなしていることに注意しよう．ただし， $\alpha_{\mu l}$ は規格化定数である．そこで (原理的には) これを解くことで，近似的な周辺分布

$$P(S_l | \{\xi^\mu\}, \phi)$$

¹ちなみに歴史的に CVM の定式化には日本人が多大な貢献をしている．近年，CVM を BP の枠組みに翻訳した [5] などの引用が国内でも増えてきたが，その元ネタの多くは [4] など国内研究者の業績なので日本人の仕事の方をもっと引用しましょう．

$$= \alpha_{\mu l} P(S_l | \xi_l^1, \phi) \prod_{\mu=2}^P P(\xi^\mu | S_l, \{\xi^{\nu \neq \mu}\}, \phi) \quad (15)$$

が得られる．式 (13)，(14) および (15) はまさに計算機科学で (ループ-) ピリーフプロパゲーションと呼ばれるものである．

さて，以上の関係式は関数方程式として表現されているため具体的な計算を行う際にいささか不便である．連想記憶の問題ではニューロン活動度 S_l が ± 1 の 2 状態に限定されているため

$$P(\xi^\mu | S_l, \{\xi^{\nu \neq \mu}\}, \phi) \propto \frac{1 + \hat{m}_{\mu l} S_l}{2} \quad (16)$$

$$P(S_l | \{\xi^{\nu \neq \mu}\}, \phi) = \frac{1 + m_{\mu l} S_l}{2} \quad (17)$$

と表すことにより，パラメータ $m_{\mu l}$ $\hat{m}_{\mu l}$ 間の非線形方程式として表現した方が都合が良い．すると式 (13)，(14) は

$$\hat{m}_{\mu l} = \frac{A_+ - A_-}{A_+ + A_-} \quad (18)$$

$$m_{\mu l} = \tanh \left(\phi \xi_l^1 + \sum_{\nu \neq \mu} \tanh^{-1} \hat{m}_{\nu l} \right) \quad (19)$$

ただし，

$$\begin{aligned} A_+ &= \sum_{S_{k \neq l}} \exp \left[\frac{\beta}{2} \left(\frac{\xi_l^\mu}{\sqrt{N}} + \Delta_{\mu l} \right)^2 \right] \\ &\quad \times \prod_{k \neq l} \left(\frac{1 + m_{\mu k} S_k}{2} \right) \end{aligned} \quad (20)$$

$$\begin{aligned} A_- &= \sum_{S_{k \neq l}} \exp \left[\frac{\beta}{2} \left(-\frac{\xi_l^\mu}{\sqrt{N}} + \Delta_{\mu l} \right)^2 \right] \\ &\quad \times \prod_{k \neq l} \left(\frac{1 + m_{\mu k} S_k}{2} \right) \end{aligned} \quad (21)$$

$$\Delta_{\mu l} \equiv \frac{\sum_{k \neq l} \xi_k^\mu S_k}{\sqrt{N}} \quad (22)$$

となる．求めるべき平均値はこの非線形方程式の解を用いて

$$m_l = \tanh \left(\phi \xi_l^1 + \sum_{\mu=2}^P \tanh^{-1} \hat{m}_{\nu l} \right) \quad (23)$$

となる．

ただし，ピリーフネットワークが完全 2 部グラフで表現されることから A_+ ， A_- の「厳密」計算にはニューロン数 N に関する指数乗程度の計算が必要となる．そのため，近似計算としての BP 法であるとはいえ，計算論的にはその実行は困難となる．

4 大自由度極限での BP 法

さて、前節ではノイズ存在化での連想記憶モデルの復号状態の近似的評価が BP 法により可能であるがピリフネットが完全 2 部グラフとなるため実行が困難となることを示した。ところが、パターン $\xi^\mu = 1, 2, \dots, P$ の統計的独立性に着目し中心極限定理をうまく利用することで必要な計算量を高々 $O(N^2)$ 程度にまで削減することが可能となる。このようにシステムの統計的特徴を利用して、効率的なベータ近似を構成する手法を統計力学ではサウレス - アンダーソン - パルマー (TAP) の接近法 [6, 7, 8] と呼ぶ²。

個別システムの特徴 (主に大自由度性と統計的対称性) を最大限に利用して、効率的なアルゴリズムを構成する TAP の方法は統計力学的接近法を特徴的に表す近似計算法である。逆に言えば、最悪評価による厳密な議論に基づき広いクラスの問題に対して普遍的に成立するアルゴリズムの性質を究明するという情報科学や計算機科学で馴染み深い視点は統計力学では一般に希薄である。

具体的には以下のように考える。まず、 N が十分大きなとき $\frac{\xi_l^\mu}{\sqrt{N}}$ は微量である。そこで、 A_+ 、 A_- の評価ではこれに関する最低次の寄与のみを考えれば十分であろう。また、仮定より ξ^μ と $\xi^{\nu \neq \mu}$ は統計的に独立である。ということは $\{\xi^{\nu \neq \mu}\}$ から定まる平均値 $m_{\mu l}$ と ξ^μ は統計的に独立である。そのため、 $S_{k \neq l}$ が分布 $P(S_{k \neq l} | \{\xi^{\nu \neq \mu}\}) = \prod_{k \neq l} \left(\frac{1 + m_{\mu k} S_k}{2} \right)$ に従うとき式 (22) で定める $\Delta_{\mu l}$ は中心極限定理により平均

$$\langle \Delta_{\mu l} \rangle_\mu \equiv \frac{\sum_{k \neq l} \xi_l^\mu m_{\mu k}}{\sqrt{N}} \quad (24)$$

分散

$$1 - q_{\mu l} \equiv 1 - \frac{1}{N} \sum_{k \neq l} m_{\mu k}^2 \quad (25)$$

の正規分布に従うと考えてよい。更に、十分大きな N に対しては $m_{\mu l}^2/N$ の寄与は小さいので重なり $q_{\mu l}$ の要素依存性は無視して

$$q_{\mu l} \simeq q_\mu = \frac{1}{N} \sum_{l=1}^N m_{\mu l}^2 \quad (26)$$

として良いであろう。また、パターン生成法の対称性と大数の法則から q_μ の値にはパターンについての依存性もなくなり、求めるべき平均値 m_l を用いて

$$q_{\mu l} \simeq q_\mu \simeq q = \frac{1}{N} \sum_{l=1}^N m_l^2 \quad (27)$$

として良いことが期待される。

以上を認めると、計算論的に困難である A_+ および A_- の評価 (20) および (21) における $S_{k \neq l}$ での和は $\Delta_{\mu l}$ に関するガウス積分により置き換えられ式 (18) は

$$\hat{m}_{\mu l} = \frac{\beta}{N(1 - \beta(1 - q))} \xi_l^\mu \sum_{k \neq j} \xi_k^\mu m_{\mu l} \quad (28)$$

となる。各添え字対 μl に対し、これを評価するのに必要な計算量は高々 $O(N)$ であるのでパターン数がニューロン数と同じオーダーであるとしても一更新当たり高々 $O(N^3)$ の計算量で BP 更新を行うことが可能である。

さらに、 $\mu = 2, 3, \dots, N$ に関して $\hat{m}_{\mu l}$ が典型的に $O(N^{-1/2})$ であることを使うと、もう一段計算量を削減することが可能となる。その鍵は、求めるべき平均値 m_l と $m_{\mu l}$ のずれは式 (23) と (19) を比較すれば分かるように微小であり、テイラー展開を用いて

$$\begin{aligned} m_{\mu l} &\simeq m_l - \frac{\partial m_l}{\partial \hat{m}_{\mu l}} \hat{m}_{\mu l} \\ &= m_l - (1 - m_l^2) \hat{m}_{\mu l} \end{aligned} \quad (29)$$

と表現できることにある。これを認めると、式 (29) と (28) からパターンとニューロン活動度の添え字を持つ $\hat{m}_{\mu l}$ がニューロン活動度のみの添え字を持つ m_l で表現することが可能となり、BP 法の計算コストは一更新当たり高々 $O(N^3)$ から $O(N^2)$ へと削減される。

実際に大きな N に関して式 (28) と (29) から $\hat{m}_{\mu l}$ を m_l で表現するにはそれなりに注意深い計算が必要となるが式 (28) がパターン ξ^μ に関する射影演算子

$$P_\mu \equiv \frac{1}{N} \xi^\mu (\xi^\mu)^T = \frac{1}{N} (\xi_l^\mu \xi_k^\mu) \quad (30)$$

および、単位行列 I を用いて

$$\hat{\mathbf{m}}_\mu = \frac{\beta}{1 - \beta(1 - q)} \left(P_\mu - \frac{1}{N} I \right) \mathbf{m}_\mu \quad (31)$$

と表現できることに気づくと効率的に計算できる。ただし、 $\hat{\mathbf{m}}_\mu = (\hat{m}_{\mu l})$ 、 $\mathbf{m}_\mu = (m_{\mu l})$ とした。

少々長い計算になるので結果のみを記すと

$$\hat{\mathbf{m}}_\mu = \beta \left(P_\mu - \frac{1}{N(1 - \beta(1 - q))} I \right) \mathbf{m} \quad (32)$$

となる。これを式 (23) に代入すると m_l に関する閉じた方程式

$$\begin{aligned} m_l &= \tanh \left(\phi \xi_l^1 + \beta \sum_{k \neq l} \left(\frac{1}{N} \sum_{\mu=2}^P \xi_l^\mu \xi_k^\mu \right) m_k \right. \\ &\quad \left. - \frac{\alpha \beta^2 (1 - q)}{1 - \beta(1 - q)} m_l \right) \end{aligned} \quad (33)$$

²現在では TAP という名称が定着しているが歴史的に見ると東北大学の守田氏と堀口氏が同じ結果を TAP[7] に先立って与えている [6]。

が得られる．ただし， $\alpha \equiv P/N$ である．

しかしながら，ここでこれまでの計算は「補助変数 ϕ が与えられた」として進んできたことを思い出そう．実際はこの変数をうまく決めてやらなくてはならない．それにもやはりニューロン数 N が大きいことを利用する．具体的には N が大きい場合式 (6) における ϕ の積分を鞍点評価し式 (7) の指数にある

$$\frac{\phi^2}{2\beta} + g(\phi|\{\xi^\mu\}) = \frac{\phi^2}{2\beta} - \frac{1}{N} \ln \sum_S \exp \left[\frac{\beta}{2} \sum_{\mu=2}^P \left(\frac{\xi \cdot S}{\sqrt{N}} \right)^2 + \phi \xi^1 \cdot S \right] \quad (34)$$

を最大にする補助変数 ϕ^* の寄与のみで近似するのである． m_l が平均値

$$\frac{\sum_S S_l \exp \left[\frac{\beta}{2} \sum_{\mu=2}^P \left(\frac{\xi \cdot S}{\sqrt{N}} \right)^2 + \phi \xi^1 \cdot S \right]}{\sum_{S'} \exp \left[\frac{\beta}{2} \sum_{\mu=2}^P \left(\frac{\xi \cdot S'}{\sqrt{N}} \right)^2 + \phi \xi^1 \cdot S' \right]} \quad (35)$$

(の近似値) を表していることに注意すると鞍点条件から ϕ^* はこの近似の範囲で

$$\phi^* = \frac{\beta}{N} \sum_{l=1}^N \xi_l^1 m_l \quad (36)$$

となることが分かる．これを方程式 (33) に代入し大きな N に対して無視できる項を除いて式 (2) を用いて整理すると最終的に

$$m_l = \tanh \left(\beta \sum_{k \neq l} J_{lk} m_k - \frac{\alpha \beta^2 (1-q)}{1 - \beta(1-q)} m_l \right) \quad (37)$$

(ただし， $q = (1/N) \sum_{l=1}^N m_l^2$) を得る．これを(連想記憶モデルにおける) TAP 方程式 [9] と呼ぶ．

この方程式が近似計算法としてどれくらい優れているかを調べる前にその形式的な特徴を見ておこう．まず，右辺の \tanh の中身を計算するコストは各ニューロンあたり $O(N)$ である．全部でニューロンは N 個あるので 1 更新あたりの計算コストは $O(N^2)$ である．また，これまではパターン ξ^1 を連想していると仮定し， ξ^1 をそれ以外のパターン $\xi^{\mu(\geq 2)}$ と区別して扱ってきたが式 (37) ではそういった「特別待遇」はなく全てのパターン ξ^μ が平等に扱われていることに注意しよう．

それどころか，パターン ξ^μ の情報すら必要ではなくそれから構成される結合 J_{lk} の情報のみがあればよい．ただし，複数のパターンを結合 J_{lk} の形に変換して記憶するという連想記憶モデルの目的を考えれば，その変換された表現から原情報を引き出すためのアルゴリズムとしてこれは当然満たすべき要件である．

最後に \tanh の中身で対角項 m_l に比例する項が差し引かれていることに注意しよう．これは一般にオンサーガー反跳項と呼ばれ，グラフにループが存在することにより生じる各ニューロンの自分自身への影響を表していると解釈される [7]．非対角項からの寄与 $\beta \sum_{k \neq l} J_{lk} m_k$ のみではこの効果が入ってしまう．そこで，これを実効的に消すために差し引かなければならないのである．この「引き算」が発見法的ではなく系統的な計算法から導出されるところが興味深い．オンサーガー反跳項の形はシステムの統計的特長と密接に関係しており例えば結合 J_{lk} の生成規則を変更すると異なる形になる．

5 解析解との比較

計算論的に困難な(ルーピー)BP法から，大自由度性とシステムの統計的特徴を利用することで必要計算量の少ない TAP 方程式を導いた．この計算法はどの程度優れた近似法となっているのであろうか．そのためには厳密計算による正解との比較を行うのが一番手っ取り早い．

しかしながら，厳密計算は計算量的に困難であり，そもそも比較すべき正解を導くのが難しい．そうであるからこそ，近似計算法をあれこれ考えていたのであった．それにも関わらず正解と比較することが果たして可能なのであろうか．

そういった状況でも，大自由度性を利用することで論理的に意味のありそうな解答を与えるのが統計力学である．具体的には以下のような比較を行う．

BP法やTAP方程式で求まるのは個々のニューロン活動度の平均値 m_l である．これは沢山 (N 個) あるニューロンに付随する微視的平均量である．これを正確に求めるのは難しい．ところが N が十分大きな場合(つまり無限大とした場合)，システム全体に平均的に依存する巨視的平均量，例えば連想しているパターン ξ^1 とニューロン活動度の平均値 m とのオーバーラップ

$$M = \frac{\xi^1 \cdot m}{N} = \frac{1}{N} \sum_{l=1}^N \xi_l^1 m_l \quad (38)$$

はレプリカ法というBP法やTAP法とは全く異なる手法により評価することが出来る．レプリカ法の根幹にはまだその正当性が明らかにされていない部分も残されている．しかしながら，連想記憶モデルのようにシステムの構成要素間にある種の統計的対称性を持つ平均場モデルに対してはこれまでの経験から巨視的平均量に対して正しい評価を与えると強く予想されている．

そこで，ニューロン数 $N = 10000$ ，パターン数 $P = 500$ の系に対して TAP 方程式により得られた解のオー

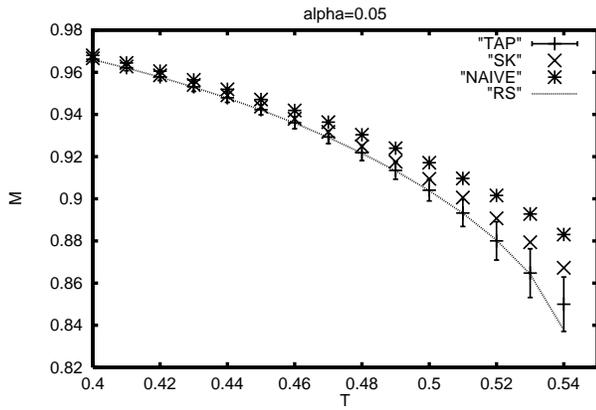


図 3: 種々の平均場近似とレプリカ法による予想の比較。パターンの統計性を適切に用いた TAP 方程式 (37) はニューロン数 $N = 10000$ 程度でレプリカ法による予想と非常に良い一致を示す。

パーラップ (38) とレプリカ法による予想との比較を行った (図 3)。また、他の近似手法として TAP 方程式でオンサーガー反跳項を無視した近似に対応するナイーブ平均場近似

$$m_l = \tanh \left(\beta \sum_{k \neq l} J_{lk} m_k \right) \quad (39)$$

さらに、システムの統計性による TAP 方程式の相違を示すため反跳項として結合 J_{kj} が添え字 kl 毎に独立な正規分布に従うとした際に得られる SK モデルの TAP 方程式

$$m_l = \tanh \left(\beta \sum_{k \neq l} J_{lk} m_k - \alpha \beta^2 (1 - m_l^2) m_l \right) \quad (40)$$

による結果も同時に比較した。図 3 より臨界的な状況以外では正しい TAP 方程式はレプリカ法による (正解であると強く予想されている) 結果と非常に良い一致を示していることが分かる。

6 おわりに

本稿では、統計力学においてピリーフプロパゲーション法がどのような文脈で登場し、それをういてどのような解析が行われているのかを紹介するため連想記憶モデルへの適用例を示した。

ここで示した接近法を敢えて「統計力学の」ということにすればその特徴は、問題クラス全体に対して普遍的に成立する「面」としての成果を最初から狙うのではなく、まずは個別問題に備わる特徴を利用することでその問題に対する深い理解とその特徴を活かした強力な計算技法の開発を目標とする「点」の発想にあると言える。

この点の個数を徐々に増やし、それらをつないで「網」を形成し最終的に面を包み込もうという作戦である。

もちろん、はじめから面的に深い成果が得られればそれに越したことはない。しかしながら、線形系や少数自由度系など例外的な場合を除き多体問題に対して普遍的に成立し、かつ強力な結果を導くことは難しい。ケーススタディに基づく統計力学の点的接近法は $O(10^{23})$ の構成要素からなる非線形大自由度系との約 1 世紀にわたる格闘の末に得られた現実的対処法であるのかもしれない。

最後に宣伝を一つ。本稿で述べたような統計力学の接近法をより体系的に情報科学に導入し、また逆に情報科学の手法によって統計力学に新たな展開を促そうという研究プロジェクトが特定領域研究「確率的情報処理への統計学的アプローチ」(領域代表: 東北大学田中和之氏)として今秋から 3 年半の予定で発足する。このプロジェクトを通して大規模な確率モデルに対する基礎理論の体系化とその定着を包括的に進めていく計画である。非専門家へのチュートリアルも含めた公開シンポジウムの開催をプロジェクト期間中毎年度予定しているので今後の動向に注目して頂ければ幸いです³。

参考文献

- [1] Hopfield, J.J., 1982, Proc. Nat. Acad. Sci. **79**, 2554.
- [2] Amit, D.J., Gutfreund, H., and Sompolinsky, H., 1985, Phys. Rev. A **32**, 1007.
- [3] Bethe, H.A., 1935, Proc. R. Soc. London, Ser. A, **151** 552.
- [4] T. Morita., 1979, Physica 98A, 566; Special issue on cluster variation methods, 1994, Prog. Theor. Phys. Suppl. **115**.
- [5] Yedidia, J.S., Freeman, W.T., and Weiss, Y., 2000, in NIPS **13**, 689.
- [6] Morita, T. and Horiguchi, T., 1976, Solid State Communications, **19**, 833.
- [7] Thouless, D.J., Anderson, P.W., and Palmer, R.G., 1977, Phil. Mag. **35**, 593.
- [8] Kabashima, Y., and Saad, D. 1998, Europhys. Lett. **45**, 668.
- [9] Kabashima, Y., and Saad, D. 2001, in Advanced Mean Field Method (MIT Press), 668.

³ ちなみに平成 14 年度は 12 月 1 日 (日) ~ 4 日 (水) に東京で開催する予定である。