

解 説

ニューラルネットワーク理論と符号

三 村 和 史

広島市立大学情報科学部*

Theory of Neural Networks and Coding

Kazushi Mimura

Faculty of Information Sciences, Hiroshima City University*

概要

ニューラルネットワークの解析に適用されている統計力学的な手法は、確率モデルで記述できる課題に対して性能評価だけにとどまらず実用的なアルゴリズム開発へも役立っている。本稿では、ニューラルネットワークの解析が、情報処理の課題へどのように適用されているかについて有歪圧縮を例として解説する。

1. はじめに

ニューラルネットワークのひとつである Hopfield モデル¹⁾の解析に適用されて以来、統計力学的な解析手法はパーセプトロンの容量評価^{2,3)}や学習理論⁴⁾など様々な情報処理の問題に適用されている。特に通信路符号化の分野で、Sourlas 符号をはじめとして LDPC 符号など性能評価にとどまらず、効率的な復号化アルゴリズムも開発されるなど活発に研究が進められている⁵⁻¹¹⁾。情報源符号化¹²⁾においても有歪圧縮などの研究がなされ、Shannon が示した理論的な性能限界を達成する有歪圧縮法がいくつか発見された¹³⁻¹⁵⁾。有歪圧縮では、理論的な性能評価が先行して実用的なアルゴリズムの開発が遅れていたが、最近になって高速な符号化アルゴリズムが発見され話題をよんでいる¹⁶⁾。

本稿では、具体的な例として有歪圧縮を具体的な例として取り上げ、ニューラルネットワークの解析に用いられている統計力学的手法がどのように情報通信の分野に適用されているかについて解説する。

2. 情報源符号化

まず、レート歪理論を簡単に紹介しよう。有限な原情報アルファベット \mathcal{J} をもつ無記憶情報源を考える。この情報源からの長さ M の系列 $\mathbf{J} = {}^t(J^1, \dots, J^M) \in \mathcal{J}^M$ について考える。符号器 \mathcal{F} は長さ M の原情報

$\mathbf{J} \in \mathcal{J}^M$ を長さ N の圧縮符号 $\xi = \mathcal{F}(\mathbf{J}) \in \mathcal{X}^N$ に変換する。復号器 \mathcal{G} は圧縮符号 $\xi = {}^t(\xi_1, \dots, \xi_N)$ を用いて、再び長さ M の復元情報 $\hat{\mathbf{J}} = \mathcal{G}(\xi) \in \hat{\mathcal{J}}^M$ に戻す。この場合、圧縮率は $R = N/M$ であり、情報圧縮ではいつも $N < M$ なので $R < 1$ となっている。圧縮率は符号化レート（または単にレート）とも呼ばれる。

歪測度は、原情報アルファベット \mathcal{J} と復元情報アルファベット $\hat{\mathcal{J}}$ の組から非負実数への写像 $d: \mathcal{J} \times \hat{\mathcal{J}} \rightarrow [0, \infty)$ である。多くの場合、原情報のアルファベット \mathcal{J} は復元情報のアルファベット $\hat{\mathcal{J}}$ と同じである。以降、 $\mathcal{J} = \hat{\mathcal{J}} \equiv \{-1, 1\}$ とし歪測度は、

$$d(J_\mu, \hat{J}_\mu) = \begin{cases} 0, & J_\mu = \hat{J}_\mu \\ 1, & J_\mu \neq \hat{J}_\mu \end{cases} \quad (1)$$

の Hamming 距離とした場合について考える。通常、情報は 0,1 の 2 値で表現されることが統計力学的な解析に便利のように、より対称性の高い $-1, 1$ の表現を利用することが多い。このようににおいても、一般性を失うことはない。この Hamming 距離では、期待値が $E[d(\mathbf{J}, \hat{\mathbf{J}})] = P[\mathbf{J} \neq \hat{\mathbf{J}}]$ とビット誤り率となっている。ここで、 E, P はそれぞれ期待値と確率を表す。2 つの系列 $\mathbf{J} = (J^\mu), \hat{\mathbf{J}} = (\hat{J}^\mu) \in \mathcal{J}^M$ について、

$$d(\mathbf{J}, \hat{\mathbf{J}}) = \frac{1}{M} \sum_{\mu=1}^M d(J^\mu, \hat{J}^\mu) \quad (2)$$

* 〒 731-3194 広島県広島市安佐南区大塚東 3-4-1

と歪測度を拡張して、原情報と復元情報の平均歪を $D = E[d(\mathbf{J}, \hat{\mathbf{J}})]$ と定義する。ここで、符号化レート (圧縮率) と許容歪の組 (R, D) が到達可能であるとは、圧縮符号長 N が十分に大きい極限で平均歪が許容歪以下となる、すなわち $E[d(\mathbf{J}, \hat{\mathbf{J}})] \leq D$ を満たすような符号器と復号器の組 $(\mathcal{F}, \mathcal{G})$ が存在することをいう。到達可能なレート歪対 (R, D) の集合を到達可能域という。レート歪関数 $R(D)$ とは、与えられた許容歪 D に対して、レート歪対 (R, D) の到達可能域の R の下限を与える関数のことである。

一般に、具体的に与えられた個々の情報源や歪測度に対して、レート歪関数を解析的に求めることは困難である。以降、簡単のため原情報アルファベットは $P[J^\mu = 0] = P[J^\mu = 1] = 1/2$ とビットの偏りが無い場合について考えよう。このような偏りが無い 2 元定常無記憶情報源で、歪測度に Hamming 距離を用いる場合に関しては、

$$R(D) = 1 - h_2(D) \tag{3}$$

と解析的にレート歪関数が与えられる。ここで、 $h_2(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ である。これは、Shannon 限界と呼ばれる。一方、符号器と復号器の組 $(\mathcal{F}, \mathcal{G})$ を指定したうえで、与えられた許容歪 D に対して (R, D) の到達可能域の R の下限を与える関数はレート歪特性と呼ばれる。符号器と復号器 $(\mathcal{F}, \mathcal{G})$ の性能の良さはレート歪特性が Shannon 限界にどれくらい近いかで判断される。さらに、実用的な観点では更に符号化や復号化に必要な計算量が、符号長の多項式程度であることが重要である。

情報圧縮の模式図を図 1 に示す。有歪圧縮では、復号器 \mathcal{G} を先に構成しておくのが一般的である。このとき符号器 \mathcal{F} は、復号したときに最も歪が小さくなるような圧縮符号を探せばよいので単に、

$$\mathcal{F}(\mathbf{J}) = \operatorname{argmin}_{\mathbf{s} \in \mathcal{X}^N} d(\mathbf{J}, \mathcal{G}(\mathbf{s})), \tag{4}$$

とすればよい。

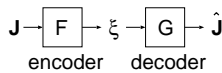


Fig. 1 情報圧縮の模式図. 符号長 M の原情報 \mathbf{J} を符号器 \mathcal{F} によって、符号長 $N (< M)$ の圧縮符号 \mathbf{s} へ符号化する。復号器 \mathcal{G} は圧縮符号 \mathbf{s} から元の符号長 M の復元情報 $\hat{\mathbf{J}}$ に戻す。

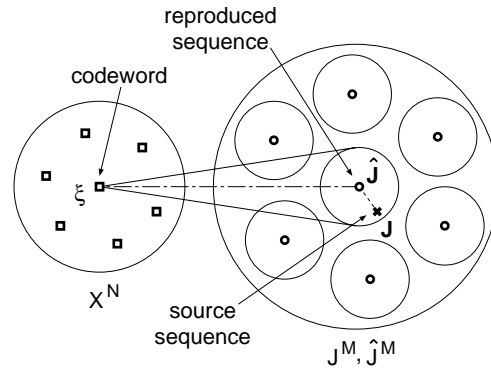


Fig. 2 ランダム符号化の模式図。すべての圧縮符号 (印) に対してランダムに復元情報 (印) を予め決めておく。符号化は、原情報 (×印) との歪が最も小さい復元情報を探して、その最適な復元情報に対応する圧縮符号を選ぶ。

3. ランダム符号化法

ランダム符号化法は、Shannon が通信路符号化定理の証明に用いた誤り訂正符号であるが、有歪圧縮に対しても同様の考え方が適用できる。

復号化： 圧縮符号の長さを N ビットとする。復号器は、この 2^N 通りある圧縮符号それぞれに対応する M ビットの復元情報を予めランダムに決めておいて、この対応関係を保存しておく。復号は、単にこの対応関係を参照することによって行う。

符号化： 一方、符号器は復号器に保存されている対応関係にある復元情報の中から、原情報と最も歪が小さくなるものを探す。その最良の復元情報に対応する圧縮符号を出力する。

この様子を Fig. 2 に示す。このランダム符号化法のレート歪特性は、Shannon 限界を達成できるが、残念ながら復号器で 2^N 個の対応関係を保存しておく必要がある上に、符号器でも 2^N 回の比較が必要であるため、現実的な符号化法とはいえない。

保坂は、パーセプトロンの容量計算と同様にレプリカ法によってもレート歪関数が導出できることを示した¹⁷⁾。解析結果は、復号したとき歪が最小となる圧縮符号が複数あることが示唆する。圧縮符号と復元情報の対応関係をランダムに決めているので、これは圧縮符号がばらばらに存在してその平均は 0 になることに

対応すると考えられる．このため，圧縮符号の平均が 0 となることが，Shannon 限界を達成するために必要であると予想される．

4. パーセプトロンと通信

Kabashima は，最も単純な学習機械のひとつであるパーセプトロンが通信に利用できることを指摘した．さらに，Hosaka らはこの考えを応用して有歪圧縮法を提案し，レート歪特性が Shannon 限界を達成することを示した¹³⁾．まず，パーセプトロンを簡単に紹介して，それがどのように通信に利用できるかについて説明しよう．

4.1 パーセプトロン

パーセプトロンとは，図 3 のようなニューラルネットワークの基本素子である．パーセプトロンは，結合荷重を適切に変更することにより与えられた入出力関係を学習するモデルとなっている． N 次元の結合荷重を $\mathbf{s} = {}^t(s_1, \dots, s_N) \in \mathcal{X}^N$ とする．結合荷重を $\mathcal{X} = \{-1, 1\}$ のように 2 値に限定したパーセプトロンは，特にイジングパーセプトロンと呼ばれている．パーセプトロンは， N 次元の入力ベクトル $\mathbf{x} = {}^t(x_1, \dots, x_N) \in \mathbb{R}^N$ と結合荷重ベクトル \mathbf{s} の内積 $\mathbf{s} \cdot \mathbf{x}$ によって，

$$y = f(\mathbf{s} \cdot \mathbf{x}) \quad (5)$$

を出力する．この関数 f を出力関数という．ここで，閾値は入力ベクトル \mathbf{x} の要素うちのひとつを 1 とおくと導入できるので省略した．例えば， $x_N = 1$ とおくと s_N が閾値の役割をはたす．

いま， M 個の入出力例 $\{(\mathbf{x}^1, J^1), \dots, (\mathbf{x}^M, J^M)\}$ について考えよう．パーセプトロンの学習とは，すべての入出力例について，

$$J^\mu = f(\mathbf{s} \cdot \mathbf{x}^\mu) \quad (6)$$

を満たす結合荷重 \mathbf{s} を求めることである．以降，情報通信を考えるので， $\mathbf{s} \in \mathcal{X} \equiv \{-1, 1\}$ ， $J^\mu \in \mathcal{J} \equiv \{-1, 1\}$ ， $f: \mathbb{R} \rightarrow \mathcal{J} = \{-1, 1\}$ というように，2 値の場合に限定して考えることにする．

4.2 パーセプトロンの容量

では，入出力例を正しく再現できる例題数 M はどれくらいであろうか．素子数に応じて正しく再現できる例題数には上限がある．この上限を素子数 N で割ったものをパーセプトロンの容量という．

容量は，与えられた入出力例を全て正しく再現できる結合荷重ベクトル \mathbf{s} の数を求めることで評価できる．先ほどの M 個の入出力例 $\{(\mathbf{x}^1, J^1), \dots, (\mathbf{x}^M, J^M)\}$

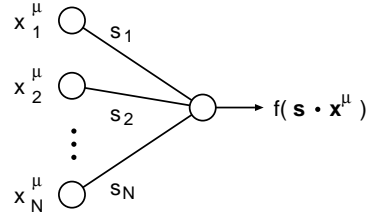


Fig. 3 パーセプトロン．結合荷重ベクトル \mathbf{s} を調整することによって，入出力関係を学習する．

について考えよう．これらは確率変数であるとする．まず最初に，さらにこれらのうち μ 番目の入出力例 (\mathbf{x}^μ, J^μ) ひとつだけに限定して考えよう．いま，

$$\Theta(x) = \begin{cases} 1, & x > 0, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

という階段関数を用いると， $\Theta[J^\mu f(\mathbf{s} \cdot \mathbf{x}^\mu)]$ は，出力関係が正しく (\mathbf{x}^μ, J^μ) となるときに 1 となり，間違ってしまうときに 0 になる関数であることがわかる．これより，ある結合荷重ベクトル \mathbf{s} を持つパーセプトロンが入出力例を全て正しく再現するときは，

$$\frac{1}{M} \sum_{\mu=1}^M \Theta[J^\mu f(\mathbf{s} \cdot \mathbf{x}^\mu)] = 1, \quad (8)$$

が成り立つことになる．素子数が N のイジングパーセプトロンでは，結合荷重ベクトル $\mathbf{s} \in \{-1, 1\}^N$ は 2^N 通りが考えられる．この 2^N 通りの結合荷重ベクトルのうち，(8) 式の値が 1 になる結合荷重ベクトルの数は，

$$\mathcal{N}_G = \sum_{\mathbf{s} \in \{-1, 1\}^N} \delta\left(\frac{1}{M} \sum_{\mu=1}^M \Theta[J^\mu f(\mathbf{s} \cdot \mathbf{x}^\mu)] - 1\right), \quad (9)$$

と評価できる．この数は Gardner 体積と呼ばれている．ここで， δ はデルタ関数である．結合荷重ベクトルの要素当たりのエントロピー S_G は対数をとって，

$$S_G = \frac{1}{N} \langle \ln \mathcal{N}_G \rangle_{\mathbf{J}, \mathbf{x}}, \quad (10)$$

と求めることができる．ここで，個々の入出力例の実現値に依存しない議論とするために，それらに関する平均 $\langle \cdot \rangle_{\mathbf{J}, \mathbf{x}}$ をとっている．組合せ論的な議論から，例題数を $M = \alpha N$ と一定に保っておく．容量は，エントロピーが 0 になるような最小の比 α で求められることになる．このエントロピーは次の恒等式，

$$\ln Z = \lim_{n \rightarrow 0} \frac{\partial Z^n}{\partial n} \quad (11)$$

を使って対数関数の期待値計算をべきの期待値計算に置き換えることで計算できる。この恒等式を使った計算方法はレプリカ法と呼ばれている。自然数の n に対して、 $n \rightarrow 0$ の極限をとることの数学的な正当性は示されていないというものの、これまでなされてきたレプリカ法による結果は実験を実によく説明できている。

解析では鞍点法を適用するために、素子数と例題数を比 α を一定に保っておいて $N, M \rightarrow \infty$ としてエントロピー S_G を計算する。出力関数が $f(x) = \text{sgn}(x)$ と符号関数のとき $S_G = 0$ となる最小の α を求めると、単調イジングパーセプトロンの容量 $\alpha_c = 0.83$ が得られる³⁾。

4.3 パーセプトロンを用いた情報通信

では、次にどのようにパーセプトロンを通信に利用するかについて説明する。いま、 M ビットの情報 $\mathbf{J} = {}^t(J^1, \dots, J^M) \in \{-1, 1\}^M$ の通信方法について考えよう。情報通信にはノイズがつきものなので、その対策として送信側では誤りを訂正するために何らかの符号化を施して情報を冗長にしておくことが多い。この符号化に、パーセプトロンを利用する。まず、送信側と受信側の両方で、素子数が N 個のパーセプトロンと、同じ乱数生成用の関数を予め用意しておく。その上で次のようにする。

送信：

1. M 個の入力ベクトル $x^1, \dots, x^M \in \mathbb{R}^N$ をランダムに生成する。
2. 入出力関係が $(x^1, J^1), \dots, (x^M, J^M)$ となるように結合荷重ベクトル $s \in \mathcal{X}^N$ を見つける (学習に対応する)。
3. 結合荷重ベクトル $s \in \{-1, 1\}^N$ と入力ベクトルの生成に使った乱数のタネを送信する。

受信：

1. 受信した乱数のタネを使って、送信側と同じ M 個の入力ベクトル x^1, \dots, x^M を生成する。
2. 受信した結合荷重ベクトル s を使って、パーセプトロンで復号する。

これは乱数のタネを除くと、 M ビットの情報 \mathbf{J} を N ビットの情報 s への変換を表している。

ところで、パーセプトロンが正しく学習できる入出力関係の数には上限 (パーセプトロンの容量) があるので、送信者はいつでも全ての入出力関係 $(x^1, J^1), \dots, (x^M, J^M)$ を満たすように、結合荷重ベ

クトル $s \in \mathcal{X}^N$ を決められる訳ではない。パーセプトロンを用いた通信では、パーセプトロンの容量を超えないようにするためや、誤り訂正の能力を決めるために、素子数 N の調節が必要である。

4.4 パーセプトロンを用いた情報圧縮

パーセプトロンを用いた情報通信で、容量のことを気にせずに任意にパーセプトロンの素子数 N を小さくするとどうなるであろうか。 $N < M$ となるように N を選べば、データ圧縮と同じである。比 $R = N/M$ を圧縮率 (符号化レート) と呼ぶ。ただし、すべての入出力関係を正しく学習できないかもしれないので、できるだけ多くの入出力関係を再現するように結合荷重ベクトルを決めるようにする。このため、データは復号しても元通りになるとは限らない。このような不可逆な圧縮を有歪圧縮という。パーセプトロンを用いた有歪圧縮は、次のようにするとよい。

符号化：

1. M 個の入力ベクトル $x^1, \dots, x^M \in \mathbb{R}^N$ をランダムに生成する。
2. できるだけ入出力関係が $(x^1, J^1), \dots, (x^M, J^M)$ となるような結合荷重ベクトル $s \in \mathcal{X}^N$ を見つける。
3. 結合荷重ベクトル $s \in \{-1, 1\}^N$ と入力ベクトルの生成に使った乱数のタネを送信する。

復号化：

1. 受信した乱数のタネを使って、送信側と同じ M 個の入力ベクトル x^1, \dots, x^M を生成する。
2. 受信した結合荷重ベクトル s を使って、パーセプトロンで復号する。

通信のときと比べて、符号化の手順 2 を少し変更しているが後は同じである。このように、パーセプトロンの結合荷重ベクトル s が圧縮符号の役割をはたすことがわかる。上記の復号化法を \tilde{G} とおく。この符号化は、(4) 式の符号化と同じである。

レート歪特性は、次のようにして評価できる¹³⁾。まず、許容歪 D を指定する。原情報と復元情報の歪 $d(\mathbf{J}, \tilde{G}(s))$ が許容歪 D 以下になるような最小の圧縮率 R を求めればよい。圧縮符号 (結合荷重ベクトル) s の長さを $N (= RM)$ としたとき、原情報と復元情報の歪がちょうど許容歪と等しくなるような圧縮符号 s の数 $\mathcal{N}_C(D)$ を求めてみる。このような圧縮符号がひとつでも存在すれば、レート歪対 (R, D) は到達可能である。先程と同様に、圧縮符号数 $\mathcal{N}_C(D)$ は、

$$\mathcal{N}_C(D) = \sum_{\mathbf{s} \in \{-1,1\}^N} \delta\left(D - d(\mathbf{J}, \tilde{\mathcal{G}}(\mathbf{s}))\right), \quad (12)$$

と評価できる．圧縮符号 1 ビット当りのエントロピー $S_C(D)$ は対数をとって， $S_C(D) = N^{-1} \ln \mathcal{N}_C(D) > \mathbf{J}, \mathbf{x}$ となる．容量はエントロピーが 0 になるような最小の圧縮率 $R_C(D)$ によって，レート歪特性 $R = R_C(D)$ を求めることができる．レート歪特性は，圧縮率 R と許容歪 D の関係を表しているのだから，ある圧縮率 R を与えたときの最小の歪 D を与える特性と読み替えることもできる．

ところで歪を許容しない場合について考えると，この有歪圧縮法は全ての入出力関係を正しく再現する結合荷重ベクトルを探すことになるので，パーセプトロンの容量評価問題と全く同じ問題のはずである．実際，原情報と復元情報の歪が，

$$d(\mathbf{J}, \tilde{\mathcal{G}}(\mathbf{s})) = \frac{1}{M} \sum_{\mu=1}^M \left(1 - \Theta[J^\mu f(\mathbf{s} \cdot \mathbf{x}^\mu)]\right), \quad (13)$$

で与えられることに注意すると，

$$\mathcal{N}_C(0) = \mathcal{N}_G, \quad (14)$$

となることがわかる．これは，許容歪が $D = 0$ のときの最小の圧縮率 $R_C(0)$ が，用いるパーセプトロンの容量の逆数 $1/\alpha_c$ と一致することを示す（逆数になるのは， R の定義を M/N ではなく N/M としたためである）．単調イジングパーセプトロンの容量は $\alpha_c = 0.83$ なので，単調イジングパーセプトロンを用いた有歪圧縮法では許容歪を 0 にしようとするとき，最小の圧縮率は $R_C(0) = 1/\alpha_c \simeq 1.2$ となり，かえって圧縮符号の方が長くなってしまふ．Shannon 限界は (3) 式のレート歪関数より $R(0) = 1 < R_C(0)$ となるので，単調イジングパーセプトロンを用いた有歪圧縮法ではレート歪特性が Shannon 限界を達成できないことは明らかである．

5. 学習機械を用いた有歪圧縮

ここまでの議論では，情報通信や情報圧縮に用いるモデルはパーセプトロンに限定していたが，同じ長さの適当な可変パラメータ \mathbf{s} を含む関数 $g(\mathbf{x}; \mathbf{s})$ を用いても同じ議論が成り立つ．初めてパーセプトロンを用いた有歪圧縮法を提案した Hosaka らが用いた非単調イジングパーセプトロンによる方法のほか¹³⁾，学習の統計力学で詳しく議論されてきた木構造 3 層パーセプトロンも有歪圧縮への適用が Mimura らによってなされてお¹⁵⁾，これらの結果を簡単に紹介する．

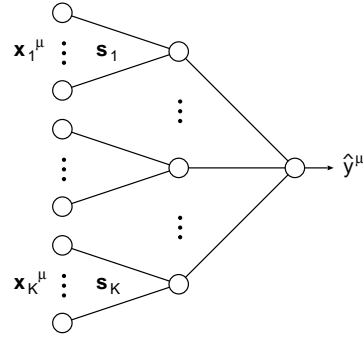


Fig. 4 木構造 3 層パーセプトロン．出力素子の出力が中間素子の出力の多数決で決まるものをコミティマシン，積で決まるものをパリティマシンという．

5.1 非単調パーセプトロン復号器

Hosaka らは，次のような偶関数の出力関数

$$f(x) = \begin{cases} 1, & |x| < k, \\ -1, & \text{otherwise,} \end{cases} \quad (15)$$

をもつ非単調イジングパーセプトロンを復号器として用いた有歪圧縮を提案して，レプリカ法によってレート歪特性を解析的に評価した¹³⁾．出力関数のパラメータ k を最適化すると，レート歪特性は Shannon 限界を達成できること示されている．

レプリカ法による解析では，圧縮符号となりえる符号の間のオーバーラップ（方向余弦）が求められる．この量は，Edwards-Anderson (EA) オーダーパラメータと呼ばれている．いま， μ 番目の入力パターン \mathbf{x}^μ に対する非単調イジングパーセプトロンの出力を $\mathcal{G}_{NP}^\mu(\mathbf{s}) \equiv f(\mathbf{s} \cdot \mathbf{x}^\mu)$ とすると出力関数 f が偶関数なので，

$$\mathcal{G}_{NP}^\mu(-\mathbf{s}) = \mathcal{G}_{NP}^\mu(\mathbf{s}), \quad (16)$$

となっていることがわかる．このため，圧縮符号 \mathbf{s} に対して必ず $-\mathbf{s}$ という符号も同じ歪を持つ，つまり原点を対称に圧縮符号が存在するので，EA オーダーパラメータは 0 となる．

5.2 コミティツリー復号器

Mimura らは非単調パーセプトロン復号器を，図 4 のような木構造 3 層パーセプトロンへ拡張した有歪圧縮法について性能評価を行なった．3 層パーセプトロンは，各中間素子が全ての入力素子と結合した完全結合 3 層パーセプトロンと，各入力素子がそれぞれひとつの中間素子にのみ結合している木構造 3 層パーセブ

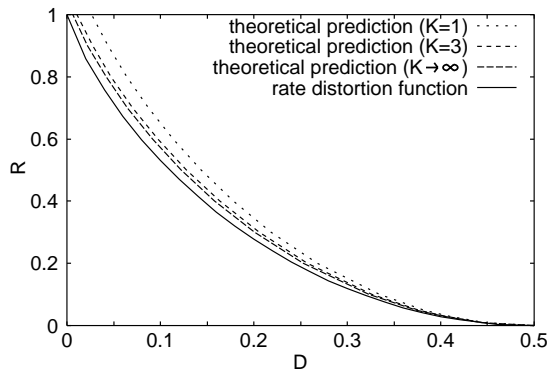


Fig. 5 コミティマシンを用いた有歪圧縮法のレート歪特性．中間素子数が増えると性能はよくなるが，中間素子数が無限大の極限でも Shannon 限界は達成できない．

トロンがよく解析されている．結合荷重が圧縮符号の役割をはたすので，完全結合 3 層パーセプトロンは結合荷重の数が多くなりすぎるため情報圧縮には適していない．このため，木構造 3 層パーセプトロンについてのみ考える．

入力素子を N ，中間素子数を K とする．各中間素子は N/K 個の入力素子と結合し，出力素子は全ての中間素子を結合荷重 1 の一定値で結合している．中間素子の出力の多数決によって出力を決める出力素子を採用した木構造 3 層パーセプトロンをコミティツリー（またはコミティマシン）という．

μ 番目の入力パターン x^μ に対するコミティツリーの出力を，

$$G_{CT}^\mu(s) \equiv \text{sgn} \left[\sum_{l=1}^K \text{sgn} \left(\sqrt{\frac{K}{N}} s_l \cdot x_l^\mu \right) \right], \quad (17)$$

とする．ここで， $x^\mu = {}^t(x_1^\mu, \dots, x_K^\mu)$ ， $s^\mu = {}^t(s_1^\mu, \dots, s_K^\mu)$ とし，中間素子の出力関数は符号関数 sgn とした．この場合は，非単調イジングパーセプトロンのときのような圧縮符号の対称性はなく，EA オーダパラメータも全ての K で 0 にはならない．

図 5 にコミティツリーを用いた有歪圧縮法のレート歪特性を示す．中間素子数 $K = 1$ のときは，単調イジングパーセプトロンを用いた場合と同じである．中間素子数を増やすとレート歪特性がよくなるのがわかるが，中間素子数を $K \rightarrow \infty$ の極限でもレート歪関数（Shannon 限界）には達しないことがわかる．

5.3 パリティツリー復号器

そこで，パリティツリー（またはパリティマシン）と

呼ばれる中間素子の出力の積によって出力を決める出力素子を採用した木構造 3 層パーセプトロンを復号器に使ってみよう． μ 番目の入力パターン x^μ に対するパリティツリーの出力は，

$$G_{PT}^\mu(s) \equiv \prod_{l=1}^K \text{sgn} \left(\sqrt{\frac{K}{N}} s_l \cdot x_l^\mu \right), \quad (18)$$

である．コミティマシンの場合と同様に，中間素子の出力関数は符号関数 sgn とした．

図 6 にパリティツリーを用いた有歪圧縮法のレート歪特性を示す．このときもちろん，中間素子数 $K = 1$ のときは単調イジングパーセプトロンを用いた場合と同じである．中間素子数が $K \geq 2$ では，レート歪特性はレート歪関数に一致することが解析的に示される．AT 不安定性を計算すると，中間素子数が $K = 2$ のときだけレプリカ対称解が不安定な領域がすこしあることがわかる．

中間素子数が $K \geq 2$ のとき Shannon 限界を達成するが，このときも EA オーダパラメータは 0 になっている．パリティツリーの出力は，

$$G_{PT}^\mu(-s) = (-1)^K G_{PT}^\mu(s), \quad (19)$$

となるので，中間素子数 K が偶数個なら非単調パーセプトロンの場合と同様な対称性があることがわかる．奇数個ならこのような対称性は一見なさそうに見える．しかし，圧縮符号 $s = {}^t(s_1, s_2, s_3, \dots, s_K)$ のうち複数個のブロックの符号を反転させた符号，例えば， ${}^t(-s_1, -s_2, s_3, \dots, s_K)$ は s と復号したとき同じ歪となる．このような符号は全部で 2^{K-1} 個あるので，それらの和は 0 となり EA オーダパラメータはやはり 0 になる．図 7 にこの様子を示す．図 7 (a) は，非単調パーセプトロンを用いたとき，圧縮符号 $s, -s$ が同じ復元情報を出し示す．図 7 (b) のように，単調パーセプトロンではこの対称性はない．図 7 (c) の $K = 2$ パリティツリーのときは (a) と同様である．図 7 (d) は $K = 3$ パリティツリーのときの様子を示す． (s_1, s_2, s_3) ， $(-s_1, -s_2, s_3)$ ， $(s_1, -s_2, -s_3)$ ， $(-s_1, s_2, -s_3)$ ，という 4 つの符号が同じ復元情報を出し示す．

6. LDGM 符号

Murayama らは，パーセプトロンの代わりに LDPC 符号で使うような疎行列を用いた有歪圧縮法を提案した¹⁴⁾．用いる疎行列 A は，1 が各行に K 個，各列に C 個しかなく，その他の要素は 0 の $M \times N$ 行列である．復号化を次の線形変換

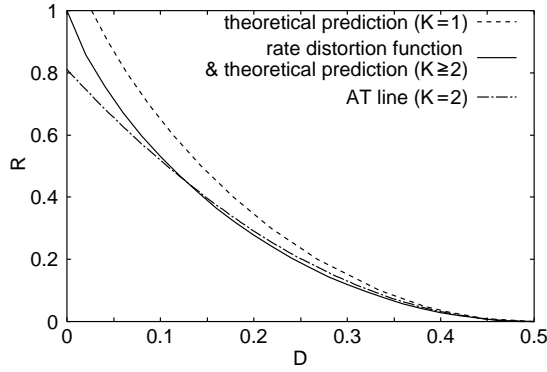


Fig. 6 パリティマシンを用いた有歪圧縮法のレート歪特性．中間素子数が $K \geq 2$ 以上で，Shannon 限界を達成する．中間素子数が $K = 2$ のときだけレプリカ対称解が不安定な領域（レート歪み関数と AT 線で囲まれた細い領域）がある．

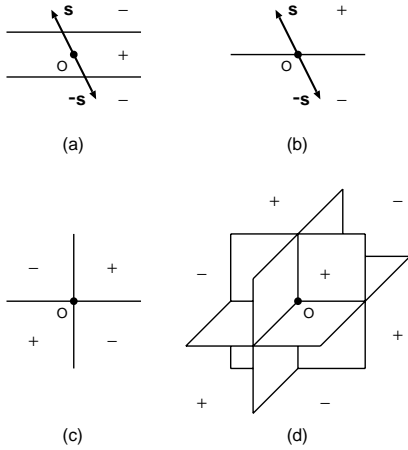


Fig. 7 復元情報の各ビットと圧縮符号の関係．記号 $+$ 、 $-$ はそれぞれ復元情報のビットが 1 、 -1 であることを表す．圧縮符号の集合 $\{-1, 1\}^N$ は，互いに直交する法ベクトルをもつ K 枚の超平面で分割される．(a) 非単調パーセプトロン．(b) 単調パーセプトロン．(c) $K = 2$ パリティツリー．(d) $K = 3$ パリティツリー．EA オーダパラメータは (b) の場合だけ 0 にならない．

$$\mathcal{G}_{LDGM}(s) = \mathcal{A}s \pmod{2}, \quad (20)$$

で定義する．ただし，符号は全て $\{0, 1\}$ の 2 値表現とする．符号化は，(4) 式の符号化と同じである．この符号は，LDGM (Low Density Generator Matrix) 符号と呼ばれている． $K \ll N$ を保ったまま， $K, C \rightarrow \infty$

の極限をとると，LDGM 符号のレート歪特性は Shannon 限界を達成できることが示されている¹⁴⁾．また，Shannon 限界を達成するこの極限の場合も EA オーダパラメータ（関数）は 0 （デルタ関数）となる． K, C が有限のときでも，レート歪特性はかなり Shannon 限界に近い¹⁴⁾．

7. アルゴリズム

では，これらの有歪圧縮法において，圧縮符号はどのようにして求めればよいであろうか．パーセプトロン，3層パーセプトロン，疎行列を用いた有歪圧縮方法の符号化は，全て (4) 式のように 2^N 個ある圧縮符号の候補 s から復号情報の歪 $d(\mathbf{J}, \mathcal{G}(s))$ が最も小さくなる圧縮符号を探すというものである．これには指数関数的な計算量が必要であり実行が困難なので，実用のためには計算量を減らす工夫が必要となる．

そこで，まず逆温度と呼ばれるパラメータ β を導入して，次の Boltzmann 分布

$$p(s|\mathbf{J}) \propto e^{-\beta d(\mathbf{J}, \mathcal{G}(s))}, \quad (21)$$

を考えよう．原情報 \mathbf{J} が与えられたときの圧縮符号が s となる事後確率が，この Boltzmann 分布で与えられると考えるのである．このとき，歪 $d(\mathbf{J}, \mathcal{G}(s))$ の最小化は，Boltzmann 分布 $p(s|\mathbf{J})$ の最大化と等価となっている．(4) 式の符号化は，

$$\mathcal{F}(\mathbf{J}) = \operatorname{argmax}_{s \in \{-1, 1\}^N} p(s|\mathbf{J}), \quad (22)$$

と書き直すことができるので，事後確率最大化 (Maximum A Posteriori: MAP) 推定とよばれている．このまま実行しようとする，MAP 推定の計算量もやはり指数関数的である．

一方，圧縮符号を探索する範囲を絞るために事後確率 $p(s|\mathbf{J})$ の周辺事後確率

$$p(s_l|\mathbf{J}) = \sum_{s \setminus s_l} p(s|\mathbf{J}), \quad (23)$$

を使う方法もある．符号器を $\mathcal{F}(\mathbf{J}) = {}^t(\mathcal{F}^1(\mathbf{J}), \dots, \mathcal{F}^M(\mathbf{J}))$ としたとき，周辺事後確率によって圧縮符号を要素ごとに，

$$\begin{aligned} \mathcal{F}^\mu(\mathbf{J}) &= \operatorname{argmax}_{s_l \in \{-1, 1\}} p(s_l|\mathbf{J}) \\ &= \operatorname{sgn} \langle s_l \rangle_{p(s_l|\mathbf{J})}, \end{aligned} \quad (24)$$

と求める方法は周辺事後確率最大化 (Maximizer of Posterior Marginals: MPM) 推定と呼ばれている．ここで， $\langle \cdot \rangle_{p(s_l|\mathbf{J})}$ は分布 $p(s_l|\mathbf{J})$ での期待値を表す．

MPM 推定では、単に $p(s_l = -1|J)$ と $p(s_l = 1|J)$ の 2 つの確率を比較して、 $p(s_l = -1|J)$ の方が大きければ -1 として、 $p(s_l = 1|J)$ の方が大きければ 1 とするだけで済む。これは、 s_l の期待値の符号をとることと同じ操作である。

逆温度は次のように設定する。デルタ関数の積分表示 $\delta(x) = \int_{-\infty}^{\infty} \frac{d\beta}{2\pi} e^{\beta x}$ より、(12) 式の状態数 $N_C(D)$ は Boltzmann 分布 $\exp[-\beta d(J, \bar{G}(s))]$ の分配関数の形となっていることがわかる。この逆温度に対応するパラメータ β はレプリカ解析ではオーダパラメータとして取り出せる。ピリーフプロパゲーションを実行するとき用いる逆温度は、エントロピーが $S_C(D) = 0$ となるときのオーダパラメータ β の値を用いればよい。逆温度を $\beta \rightarrow \infty$ とすると、歪が最小とならない圧縮符号の事後確率は 0 になるので、MPM 推定の結果は MAP 推定の結果と一致する。

MPM 推定では、探索範囲を限定できる代わりに (23) 式の周辺事後確率には、 2^{N-1} 個の項の和を計算する必要があるため、やはりそのままでは MPM 推定も指数関数的な計算量が必要になってくるのが問題となる。このため、ピリーフプロパゲーションを適用して周辺事後確率の期待値計算を近似的に行なうことで、符号化の計算量を減らすことが試みられている^{16,18)}。さらに、デンス系では中心極限定理を適用して、計算量を減らすなどの工夫が必要である。

しかしながら、ここでもまた問題が起こる。例えば、非単調パーセプトロンを用いた方法では、圧縮符号 s と同じ歪となる符号 $-s$ が必ず存在するために、周辺事後確率は必ず $p(s_l = -1|J) = p(s_l = 1|J) = 1/2$ となってしまう。このため、期待値は $\langle s_l \rangle_{p(s_l|J)} = 0$ となってしまう、圧縮符号の推定がうまくできない¹⁷⁾。他の方法でも、EA オーダパラメータが 0 であれば同じ問題が起こることになる。

7.1 慣性項を導入したピリーフプロパゲーション

この問題を解決するためには、ピリーフプロパゲーションの改良が必要であった¹⁶⁾。ピリーフプロパゲーションは、確率分布 $p(s|J)$ が、

$$p(s|J) = \prod_{\mu} e^{\phi_{\mu}(s_{\mu}|J^{\mu})}, \quad (25)$$

のような形で与えられるとき、高速に各変数 s_l の期待値 $m_l = \sum_{s \setminus s_l} s_l p(s|J)$ を反復計算によって近似的に求める方法である。しかし、EA オーダパラメータが 0 となる有歪圧縮法では、真の期待値が 0 であると考えられるので、単純にピリーフプロパゲーションを適用しても期待値は 0 が得られるだけであらう。な

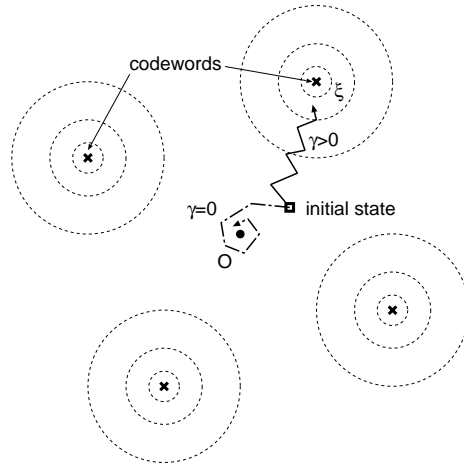


Fig. 8 慣性項を導入したピリーフプロパゲーションのダイナミクス (模式図)。圧縮符号の推定値 (印) の推移を表す。破線は、圧縮符号を復元した情報と原情報の歪の等高線であり、圧縮符号のところがくぼんでいる。圧縮符号の平均は 0 なので、慣性項がない $\gamma = 0$ のとき、圧縮符号の推定値は原点へ移動する (一点鎖線)。慣性項がある $\gamma > 0$ のとき、圧縮符号の推定値は、過去の状態と同じ状態を取ろうとするため、ランダムに与えられる初期値に依存してどこか特定の圧縮符号の付近へと移動することができる (実線)。この模式図の圧縮符号の分布は $K = 3$ パリティツリーを用いた場合に対応する。

んらかの方法で、特定の圧縮符号の情報だけを増やす必要があるのである。

Murayama は、ピリーフプロパゲーションに慣性項を導入して、LDGM 符号の高速な符号化アルゴリズムの構成に成功した¹⁶⁾。Murayama の方法は、理論が予測する性能をほぼ実現する符号化アルゴリズムとして注目を集めている。慣性項は、メッセージの更新式を次のようにすることで導入される。すなわち、

$$\begin{aligned} \rho_{\mu l}^t &= \prod_{\nu \in \mathcal{M}(l) \setminus \mu} \hat{\rho}_{\nu l}^t(s_l), \quad (26) \\ \hat{\rho}_{\mu l}^{t+1}(s_l) &= q^{t+1}(s_l) \sum_{\mathbf{s}^{\mu} \setminus s_l} e^{\phi_{\mu}(\mathbf{s}^{\mu}|J^{\mu})} \prod_{j \in \mathcal{L}(\mu) \setminus l} \rho_{\mu j}^t(s_j), \quad (27) \end{aligned}$$

とする。反復ステップ t のメッセージを利用して計算した期待値 m_l^{t+1} は、

$$m_i^{t+1} = q^{t+1}(s_i) \sum_{s^\mu} e^{\phi_\mu(s^\mu | J^\mu)} \prod_{j \in \mathcal{L}(\mu) \setminus i} \rho_{\mu j}^t(s_j), \quad (28)$$

となる．圧縮符号の l 番目のビットは， $\xi_l = \text{sgn}(m_i^{t+1})$ とMPM推定できる．ここで， $\mathcal{L}(\mu)$ はクリーク s^μ に含まれる変数 s_l の添え字 l の集合で， $\mathcal{M}(l)$ は変数 s_l を含むクリーク s^μ の添え字 μ の集合である．また， $q^{t+1}(s_i)$ がヒューリスティックに導入された慣性項であり，

$$q^{t+1}(s_i) \equiv e^{s_i \tanh^{-1} \gamma m_i^t}, \quad (29)$$

と定義されている．パラメータ $0 \leq \gamma \leq 1$ は慣性項の大きさを表し， $\gamma = 0$ とすると慣性項のない場合に返る．

この慣性項は，各反復ステップでの期待値を変化させないように働く．これによって，図8の模式図のように反復式の初期値に依存して，どれか特定の圧縮符号に関する情報が増えることが期待できる．慣性項の大きさ γ の選び方については，現在のところ理論的指針はない．また，慣性項のないビリーブプロパゲーションと同様に反復計算が収束しないこともあることに注意されたい．

Hosakaらは，非単調パーセプトロンの有歪圧縮法の符号化アルゴリズムを慣性項を導入する方法を用いて構成し¹⁸⁾，この場合も理論的な性能評価に近い符号化に成功しているほか，誤り指数を求めている¹⁹⁾．

パリティツリーを用いた有歪圧縮法についても，慣性項を導入すると高速な符号化アルゴリズムが構成できることが予測される．パリティツリーの場合は中間素子数 K に応じて，復号したときに最も歪が小さくなる符号が 2^{K-1} 個あることが保障されるので，中間素子数が大きくなるほど反復式がより少ない回数で収束したり，反復解の振動が起こる割合を減らすことができる可能性がある．

8. おわりに

容量評価問題に関するニューラルネットワーク理論と有歪圧縮の関係について説明した．解析に用いられる統計力学的手法は，既に誤り訂正符号，有歪圧縮，暗号，画像修復といった広範囲な情報処理の課題に対して適用されている．確率モデルで記述できる課題に対して，統計力学的手法には性能評価だけにとどまらず，情報処理で重要となる処理過程などに適用できる様々な有用な解析手法が用意されている．有歪圧縮の問題についても，よりよい符号化アルゴリズムの開発のた

めには，その符号化の処理過程などの解析が不可欠である．また，理論限界を達成するための必要条件など興味深い問題がたくさん残されており，その発展が期待されている．

謝辞 執筆の機会を与えて下さいました山口大学の庄野逸先生をはじめ，本号を分筆されている諸先生方に感謝致します．

参考文献

- 1) Hopfield, J. J. (1982): *Proceeding of the National Academy of Sciences*, **79**, pp. 2554.
- 2) Gardner, E. (1988): *Journal of Physics A: Mathematical and General*, **21**, pp. 257.
- 3) Krauth, W., Mezard, M. (1989): *Journal of Physics France*, **37**, pp. R1.
- 4) Watkin, T.L.H., Rau, A., Biehl, M. (1993): *Reviews of Modern Physics*, **65**, pp. 499.
- 5) Surlas, N. (1989): *Nature*, **339**, pp. 693.
- 6) Nishimori, H., Wong, K. Y. M. (1999): *Physical Review E*, **60**, pp. 132.
- 7) Kabashima, Y., Saad, D. (1999): *Europhysics Letters*, **60**, pp. 132.
- 8) Kabashima, Y., Murayama, T., Saad, D. (2000): *Physical Review Letters*, **60**, pp. 132.
- 9) Yoshiyuki, K., Saad, D. (2004): *Journal of Physics A: Mathematical and General*, **37**, pp. R1.
- 10) 樺島祥介 (2004): 数理科学, 11月号, pp. 48.
- 11) 樺島祥介 (2002): 物理の世界『学習と情報の平均場理論』, 岩波書店.
- 12) Murayama, T. (2002): *Journal of Physics A: Mathematical and General*, **35**, pp. L95.
- 13) Hosaka, T., Kabashima, Y., Nishimori, H. (2002): *Physical Review E*, **66**, pp. 066126.
- 14) Murayama, T., Okada, M. (2003): *Journal of Physics A: Mathematical and General*, **36**, pp. 11123.
- 15) Mimura, K., Okada, M. (2005): *preprint* (e-print arXiv cond-mat/0508598).
- 16) Murayama, T. (2004): *Physical Review E*, **69**, pp. 035105.
- 17) 保坂忠明 (2005): 博士論文, 東京工業大学.
- 18) Hosaka, T., Kabashima, Y. (2005): *preprint* (e-print arXiv cs.IT/0509086).
- 19) Hosaka, T., Kabashima, Y. (2005): *Journal of the Physical Society of Japan*, **74**, pp. 488.