数理解析研究所講究録
第 835 巻 1993 年 1-10

ボルツマンマシンの最適化問題への応用

広島大学大学院工学研究科　奥原浩之 (Koji OKUHARA)
広島大学工学部第二類　尾崎俊治 (Shunji OSAKI)

1. はじめに

ニューラルネットワーク (*N.N.*) は人間の脳神経回路を模したモデルである. その歴史は約半世紀前, 神経細胞をモデル化した形式ニューロン [1] に始まる. 次いで, 学習のアルゴリズムとして *Hebb* 則 [2] が提案され, それらを用いた階層型 *N.N.* (*Perceptron*)[3] が, 学習によりパターンを識別する能力を実現した. 以来, 生体の情報処理そのものを解き明かそうとする流れと, その優れた特長である並列処理と学習を工学的に応用しようとする流れが一緒になり研究が進められている.

現在, 種々の *N.N.* が提案されており, 最適化問題への適用としては, 巡回セールスマン問題には相互結合型 *N.N.* である Hopfield 型 *N.N.* (*H.N.N.*) が有効であることが知られている. しかし, *H.N.N.* はあらかじめ与えられた系のエネルギーを最小あるいは極小にする状態を並列処理で実現するもの

で，学習の概念は用いられていない．

本研究ではボルツマンマシンを多次元 (mD) 最大エントロピー法 (MEM)[4] との比較により解釈する．更にボルツマンマシンを mD MEM に適用するについて考察する．ボルツマンマシンはニューラルネットワーク ($N.N.$) の一種である．MEM はスペクトル解析の一手法である．一次元 ($1D$)MEM には格子型回路を用いた $Burg$ 法と呼ばれるアルゴリズムがある．しかし mD MEM には，まだ有効なアルゴリズムが確立されていない．そこで本論文ではボルツマンマシンを多次元スペクトル解析として解釈すると共に，ボルツマンマシンが mD MEM に有効であることを示す．

本論文の構成は，2 章でボルツマンマシンについて概説し，3 章でボルツマン分布と正規分布との関係を決定する．4 章で MEM を概説し，5 章でボルツマンマシンを多次元スペクトル解析として解釈すると共に，ボルツマンマシンが mD MEM に有効であることを示す．

2. ボルツマンマシンの定式化

ボルツマンマシンを構成するユニットは n 個あり，相互に結合していて自己帰還がなく，しかも対称であるとする．それぞれのユニットの動作規則は確率的に 2 値をとり多入力一

出力である．そして状態は非同期に更新するものとする．j 番目のユニットの出力を $u_j, j=1,2,\cdots,n$ とすると i 番目のユニットへの入力 U_i は，

$$U_i = \sum_{j=1}^{n} w_{ij}u_j - \theta_i \tag{1}$$

で与えられる．ここで，w_{ij} は j 番目のユニットから i 番目のユニットへの結合荷重，θ_i は i 番目のユニットのしきい値である．ネットワークの結合荷重は自己帰還がなく対称としているので，

$$w_{ii} = 0 \tag{2}$$

$$w_{ij} = w_{ji} \qquad (i \neq j) \tag{3}$$

である．入力が U_i のとき出力 u_i がそれぞれ 1 および 0 となる確率を $P(u_i=1;U_i)$ および $P(u_i=0;U_i)$ で表すと各ユニットは次の確率

$$P(u_i=1;U_i) = f(\frac{U_i}{T}) \tag{4}$$

$$\begin{aligned} P(u_i=0;U_i) &= 1 - P(u_i=1;U_i) \\ &= f(-\frac{U_i}{T}) \end{aligned} \tag{5}$$

で 1 または 0 のうちどちらかを出力する．ここで T はネットワークに与えられた温度である．$f(x)$ は非線形関数で通常，次のシグモイド関数

$$f(x) = \frac{1}{1+\exp(-x)} \tag{6}$$

が用いられる．ネットワークのパターンを $\mathbf{u} \equiv (u_1, u_2, \cdots, u_n)$ で表す．このとき，ネットワークのエネルギー関数 $E(\mathbf{u})$ が次のように，

$$E(\mathbf{u}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}u_iu_j + \sum_{i=1}^{n}\theta_iu_i \tag{7}$$

と定義できる．$w_{ii}=0$ より式(7)において改めて $-2\theta_i$ を w_{ii} とおくことにより，

$$E(\mathbf{u}) = -\sum_{i=1}^{n}\sum_{j\geq i}^{n} w_{ij}u_iu_j \tag{8}$$

を得る．これはユニットの出力 u_i が 1 または 0 の 2 値であり，${u_i}^2 = u_i$ とできるからである．

式(4)または式(5)にしたがって状態変化を繰り返すと，各状態がネットワークの結合荷重と温度によって決まる確率で出現する定常状態に収束する．この確率分布はエネルギー関数を用いて，次のようなボルツマン分布

$$B(\mathbf{u}) = Q\exp(-\frac{E(\mathbf{u})}{T}) \tag{9}$$

で表せる．ここで，Q は規格化定数である．T を小さくするにしたがい系の動作は決定論的になる．そこでエネルギーを最小にする状態を実現するためには焼き鈍し(*simulated anneal*

を行う．

,

3. ボルツマン分布と正規分布との関係

式 (9) を簡単にするために,

$$W_{Tij} = -\frac{w_{ij}}{T}. \tag{10}$$

とする. W_{Tij}を要素とする行列を$\mathbf{W}_T$とする. そのとき, 式 (9) は,

$$P(\mathbf{u}) = Z^{-1}\exp(-\mathbf{u}\mathbf{W}_T\mathbf{u}/2), \tag{11}$$

とできる.

次に平均ベクトルμ, 分散-共分散行列$\mathbf{\Sigma}$である多次元正規分布 $P(\mathbf{x}) \equiv N(\mu, \mathbf{\Sigma})$ の特性関数$\varphi_{\mathbf{x}}(j\mathbf{u})$を考える.

$$\varphi_{\mathbf{x}}(j\mathbf{u}) = \varphi_{\mathbf{x}'}(j\mathbf{u})\exp(-\frac{E(\mathbf{u})}{T}) \tag{12}$$

ここで$\varphi_{\mathbf{x}'}(j\mathbf{u})$は, $P(\mathbf{x}') \equiv N(\mu, \mathbf{\Sigma} - \mathbf{W}_T)$ の特性関数である. $B(\mathbf{u})$ は式 (12) の右辺第 2 項を規格化したもので,

$$\varphi_{\mathbf{x}'}(j\mathbf{u})B(\mathbf{u}) = \int_{\mathbf{x}} e^{j\mathbf{u}\mathbf{x}}(\sum_{\mathbf{u}'}\varphi_{\mathbf{x}'}^{-1}(j\mathbf{u}')e^{j\mathbf{u}'\mathbf{x}})^{-1}d\mathbf{x} \tag{13}$$

で与えられる. ここで u の集合と u′の集合は一致している.

,

4. 最大エントロピー法

今, 時系列 $X(t) = x_i$ の取り得る値は連続である. $x_1, x_2, \cdots, x_n$ の結合確率関数を $p(\mathbf{x})$ とする. もし, $p(\mathbf{x})$ が *Toeplitz* 行

列 Γ_m を用いて，$N(m, \Gamma_m)$ と表せるなら，時系列のエントロピーとして次式が定義できる．

$$H = \int_{-f_n}^{f_n} \log S_x(f)\, df \tag{14}$$

ここで $f_n = 1/(2\Delta t)$ は $Nyquist$ 周波数であり，Δt はサンプリング周期である．$S_x(f)$ はパワースペクトル密度であり，次の $Wiener - Khintchine$ の関係式

$$C_x(k) = \int_{-f_n}^{f_n} S_x(f) e^{2\pi f k \Delta t}\, df \tag{15}$$

を満たしている．ここで $C_x(k)$ は自己相関関数である．

1D MEMは式 (15) の制約のもと 式 (14) を最大にする最適化問題（M）であり，

$$\textbf{maximize} \qquad H = \int_{-f_n}^{f_n} \log S_x(f)\, df \tag{16}$$

$$\textbf{subjectto} \qquad C_x(\tau) = \int_{-f_n}^{f_n} S_x(f) e^{2\pi f \tau}\, df. \tag{17}$$

のように定式化される．

次に時系列 $X(t, \mathbf{z}), z_i, i = 1, 2, \cdots, m$ を考える．この時，周波数-波数パワースペクトル密度は，

$$S_x(f, \mathbf{v}) = \int_{\mathbf{r}} P_x(f, \mathbf{r}) e^{-j2\pi \mathbf{v}\mathbf{r}}\, d\mathbf{r} \tag{18}$$

となる．ここで $\mathbf{v}$ は空間ベクトル，$\mathbf{r}$ はラグ空間である．クロススペクトル $P_x(f, \mathbf{r})$ は，

$$P_x(f, \mathbf{r}) = \int_{\tau} C_x(\tau, \mathbf{r}) e^{-j2\pi f \tau}\, d\tau \tag{19}$$

で与えられる．$C_x(\tau, \mathbf{r})$ は時空自己共分散関数であり，

$$C_x(\tau, \mathbf{r}) = E[[x(t+\tau, \mathbf{z}+\mathbf{r}) - m_x][x(t, \mathbf{z}) - m_x]^*] \tag{20}$$

で定義される．ここで m_x は，

$$m_x = E[x(t, \mathbf{z})] \tag{21}$$

である．先と同様にして時系列のエントロピーとして次式

$$mH = \int_{\mathbf{v}} \log S_x(f, \mathbf{v})\, d\mathbf{v} \tag{22}$$

が定義できる．制約は，式 (18) を逆フーリエ変換した，

$$P_x(f, \mathbf{r}) = \int_{\mathbf{v}} S_x(f, \mathbf{v}) e^{j2\pi \mathbf{v}\mathbf{r}}\, d\mathbf{v} \tag{23}$$

である．

$mD\ MEM$ は，式 (23) の制約のもと式 (22) を最大にする最適化問題（mM）であり，

$$\textbf{maximize} \qquad mH = \int_{\mathbf{v}} \log S_x(f, \mathbf{v})\, d\mathbf{v} \tag{24}$$

$$\textbf{subjectto} \qquad P_x(f, \mathbf{r}) = \int_{\mathbf{v}} S_x(f, \mathbf{v}) e^{j2\pi \mathbf{v}\mathbf{r}}\, d\mathbf{v}. \tag{25}$$

のように定式化される．このような $S_x(f, \mathbf{v})$ は，

$$S_x(f, \mathbf{v}) = (\sum_{\mathbf{r}'} \lambda_{\mathbf{r}'} e^{j2\pi \mathbf{v}\mathbf{r}'})^{-1} \tag{26}$$

で与えられる．ここで $\mathbf{r}'$ はサンプリングした $\mathbf{r}$ であり，$\lambda_{\mathbf{r}'}$ は式 (25) の制約に対するラグランジュ乗数である．これより，

$$P_x(f,\mathbf{r}') = \int_{\mathbf{v}} e^{j2\pi\mathbf{v}\mathbf{r}'}(\sum_{\mathbf{r}'}\lambda_{\mathbf{r}'}e^{j2\pi\mathbf{v}\mathbf{r}'})^{-1}d\mathbf{v} \tag{27}$$

となる.

,

5. ボルツマンマシンの mD MEM への適用について

ボルツマンマシンの定常状態は式 (13) のような $B(\mathbf{u})$ を与える. mD MEM のクロススペクトル $P_x(f,\mathbf{r}')$ は式 (27) で与えられる. これらを比較することで,

$$\mathbf{r} = \mathbf{u} \tag{28}$$

$$2\pi\mathbf{v} = \mathbf{x} \tag{29}$$

$$\lambda_{\mathbf{r}'} = (2\pi\varphi_{\mathbf{x}'}(j\mathbf{u}'))^{-1} \tag{30}$$

がいえる. つまり $N.N.$ のパターン $\mathbf{u}$ はラグ空間にあたり, $\mathbf{x}$ は波数ベクトルとなっていることがわかる. 更にこれらを用いて,

$$P_x(f,\mathbf{u}) = \varphi_{\mathbf{x}'}(j\mathbf{u})B(\mathbf{u}) \tag{31}$$

$$S_x(f,\mathbf{x}) = 2\pi(\sum_{\mathbf{u}}\varphi_{\mathbf{x}'}^{-1}(j\mathbf{u})e^{j\mathbf{x}\mathbf{u}})^{-1} \tag{32}$$

が導かれる. ところが $\varphi_{\mathbf{x}'}(j\mathbf{u})$ は簡単には決定できない. そこで式 (31) より $\varphi_{\mathbf{x}'}(j\mathbf{u})$ を消去することにより,

$$S_x(f,\mathbf{x}) = 2\pi(\sum_{\mathbf{u}}(B(\mathbf{u})/P_x(f,\mathbf{u}))e^{j\mathbf{x}\mathbf{u}})^{-1} \tag{33}$$

を得る.

以上より，周波数 f，波数ベクトル $\mathbf{x}$，クロススペクトル $P_x(f,\mathbf{u})$ を与えれば，周波数-波数パワースペクトル $S_x(f,\mathbf{x})$ はボルツマンマシンの状態 $\mathbf{W}_T$によって決定する．

,

6. 結論

本研究では並列処理と学習という特長をもつボルツマンマシンをスペクトル解析の一手法である $mD\ MEM$との比較により解釈した．更にボルツマンマシンを $mD\ MEM$に適用することについて考察した．その結果，ボルツマンマシンはクロススペクトル等から計算することにより，与えられる確率分布を獲得することで，多次元スペクトル解析に適用できることが示された．もしこの仕組みが明らかになると生体との関わりがどうなのか興味深い．

今後は獲得したい確率分布がどのようにして計算され与えられるかを求め，ボルツマンマシン多次元最大エントロピー法へ適用するアルゴリズムを提案する予定である．

,

Reference

[1] McCulloch, W.S., and Pitts, W.: "A logical calculus of the ideas immanent in nervous activity", *Bullet. Math. Biophysics* **5**, pp.

115-133 (1943).

[2] Hebb, D.O.: *The Organization of Behavior*, Wiley, New York (1949).

[3] Rosenblatt, F.: "The perceptron: A probabilistic model for infomation storage and organization in the brain", *Psychol. Rev.* **65**, pp. 386-408 (1958).

[4] Haykin, S. (ed.), *Nonlinear Methods of Spectral Analysis*, Topics in Applied Physics **34**, Springer-Verlag, Berlin, pp. 230-232 (1979).