

I - 68

Boltzmann Machineの確率論による基礎考察

(株)地崎工業 技術開発部 正会員 須藤 敦史

1. はじめに

離散変数等の組み合わせ最適化問題は、多数の解候補の中から特殊な組み合わせを見つけ出す確定的問題であり、最適解の探索は極めて難しいため、遺伝的アルゴリズム(GA)¹⁾、Simulated Annealing(SA)²⁾、モンテカルロ法を基本とした方法³⁾など、問題を確率的に考えた組み合わせ最適化手法が用いられてきている。

GAは、生物の進化過程を工学的にモデル化したものであり、解候補を遺伝子の配列と見なし、その交配や突然変異等を探索アルゴリズムに取り入れたものである。SAは金属結晶が徐々に冷却された場合に安定した分子構造に落ち着く点に注目したアルゴリズムである。また、モンテカルロ法を基本とした方法は、解候補をランダムに作成し、より良い解集合の分布範囲を徐々に限定してゆき、サンプルに最適解が選択される確率を高めてゆく手法である。

一方、人間の脳における情報処理を模擬したNeural Networkの原型にも、Hopfield Model⁴⁾やBoltzmann Machine⁵⁾というアルゴリズムがある。しかし、Neural Networkが注目される中でこれらの手法の組合せ最適化問題への適応は検討なされていない。そこで本研究では、このHopfield ModelとBoltzmann Machineの組合せ最適化問題への適用性の考察を確率論を基本として試みている。

2. Neural Network

Neural Networkは脳の情報処理を模擬し、工学に活用しようとするものである。人間の神経細胞における情報伝達の仕組みは、『シナプス結合を通じて前の細胞からの情報を入力し、細胞の膜電位が変位する。膜電位がある値<しきい値>を越えると、細胞は興奮しインパルスが生じる。そして、次の細胞へ情報を伝達する。』この神経細胞をモデル化し、これらを組み合わせてネットワークとして構成したものがNeural Networkである。

2.1 神経細胞のモデル化

いま、1つの神経細胞が n 個の細胞から入力を受けるとする。前細胞の出力を X_i ($1 < i < n$)、それぞれのシナプス結合の強さ(重み)を W_i で表すと、 i 番目の細胞からの入力は $W_i X_i$ と表される。各細胞からの入力の総和 $\sum_i W_i X_i$ がしきい値 θ を越えたとき、細胞は興奮(発火)する。

$$u = \sum_i W_i X_i - \theta \quad (1) \quad Z = f(u) \quad (2)$$

これらをモデル化したものをFig.1に示す。ここで u は膜電位、 $f(u)$ は出力関数と呼ぶ。入力 X_i と出力の取りうる値としては、実数値の場合と2値(0, 1)の場合がある。ここで2値モデルを以下に示す。

$$Z = f(u) = \begin{cases} 1 (u \geq 0) \\ 0 (u < 0) \end{cases} \quad (3)$$

2.2 階層型と相互結合型

Neural Networkをネットワークの構造で分類すると、階層型と相互結合型に分類できる。階層型は入力側から出力側へ向けての情報伝達のみ可能である。また層間の結合はなく同一層内の情報伝達も考えない(Fig.2)。相互結合型は各ニューロンが相互に結合するネットワークであり情報伝達は一般的に双方向となる。(Fig.3)

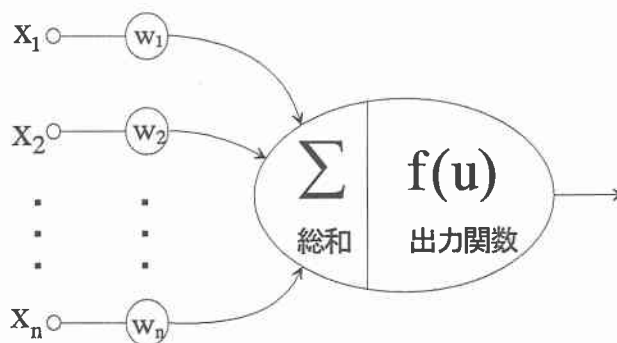


Fig. 1 ニューロンのモデル化

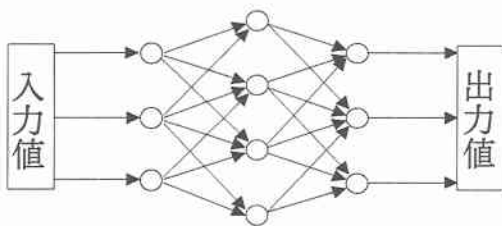


Fig. 2 階層型ネットワーク

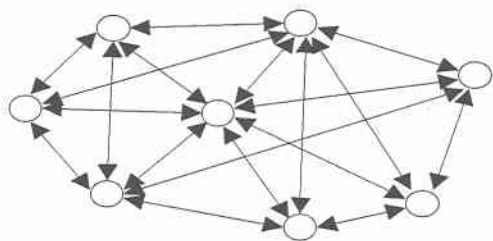


Fig. 3 相互結合型ネットワーク

階層型は入力側から出力側へ一方方向に流れ、状態の変化が行われるのに対して、相互結合型は初期状態から出発し、安定な平衡状態に達するまでネットワーク間の結合状態が継続していくものである。

2.3 学習

Neural Network の学習はシナプス間の結合荷重値を変化させることで定義しており、教師あり学習と教師なし学習に分類することができる。教師あり学習は主に階層型に用いられ、入力データと教師データを用いて学習させるものである。これより学習後のネットワークは未知な入力データに対しても適切な出力を得る。代表例としてバックプロパゲーション学習が上げられる。まず入力により出力層ニューロンが発火する。しかし学習初期では、ニューロンの発火度合いが小さかったり、他のニューロンが大きく発火したりして誤差を生じる。この誤差を以下に示す基準を用いて、重みを変化させることにより最小化させるものである。

$$(\text{誤差}) = \{\text{教師信号 (希望出力)}\} - (\text{実際の出力}) \quad (4)$$

一方、教師なし学習は相互結合型に用いられており、ネットワークに入力データを与えるだけで、選択的(自動的に)反応するネットワークを形成する。自己により組織形成を行うことから自己組織化と呼ばれ、代表例としてヘッブ型学習がある。これは隣り合う2つのニューロンが互いに発火すれば、両者を結ぶ重みを大きくすることを基本としている。このようにNeural Networkは、この学習アルゴリズムや結合パターンの選択によって、様々な問題に適用することができる。加えて、Boltzmann Machineは組み合わせ最適化に適していると言われ、相互結合型教師なし学習アルゴリズムを用いたNeural Networkの一手法である。

3. Hopfield Model

相互結合型ネットワークに2つの制約条件を付加したものが、Hopfield Modelである。第1の制約はニューロン間の結合の強さは対称、第2の制約は各ニューロンは完全に非同期的に動作する。 N 個のニューロンで構成されるネットワークにおいて、時刻 t に i 番目のニューロンが $N-1$ 個のニューロンから受ける信号の総和 $u_i(t)$ は以下となる。

$$u_i(t) = \sum_j^{N-1} W_{ij} X_j(t) + \theta_i \quad (5)$$

$$\begin{cases} W_{ij}: \text{ニューロン } i, j \text{ 間の重み} \\ X_j: \text{ニューロン } j \text{ の出力} \\ -\theta_i: \text{ニューロン } i \text{ のしきい値} \end{cases}$$

このとき、時刻 $t+1$ におけるニューロン i の出力 $X_i(t+1)$ は、 $u_i(t)$ が正のとき1、負のとき0に更新され、状態を変化させる。ここでHopfield Modelはエネルギー関数というを導入し時刻 t において以下となる。

$$E(t) = -\frac{1}{2} \sum_i^{N-1} \sum_j^{N-1} W_{ij} X_i(t) X_j(t) - \sum_i^{N-1} \theta_i X_i(t) \quad (6)$$

このエネルギー関数はニューロン i とその前細胞ニューロン j との出力値で決定され、ネットワークの平衡状態はエネルギー関数の最小値(または局所解)である。

4. Boltzmann Machine

Boltzmann Machineは、Hopfield Modelを拡張(確率判定の導入)によって、組合せ最適化問題を解くことが可能となったといえる。Hopfield Modelと同様にネットワークが信号の総和 $u_i(t)$ を式(5)で定義し、Boltzmann Machineの出力は2値のみとなる。よって、2値を0, 1とすると1を出力する確率は以下と

なる（当然 0 を出力する確率は $1 - P$ ）。

$$P[X_i(t+1)=1]=\frac{1}{1+\exp\{-u_i(t)/T\}} \tag{7}$$

この出力判定のために確率を使用しているのが Boltzmann Machine の特徴といえる。ここで T は温度パラメータである。(7)式はシグモイド関数と呼ばれ Fig.4 に示すように $u \rightarrow \infty : P \rightarrow 1, u \rightarrow -\infty : P \rightarrow 0$ に収束する。また、この関数は温度 T が大きいと u の値に鈍感になり、 T を小さくすると敏感になる。 $T \rightarrow 0$ の極限においては Hopfield Model の階段関数となる。ここで、温度 T を小さくすることによって

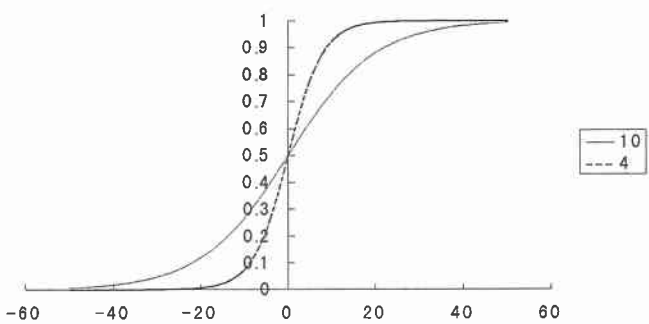


Fig. 4 シグモイド関数

$x_i = 1$ となる確率が高くなり、エネルギー関数の値が決定論的になる。また、温度 T が大きいとエネルギー関数はなだらかになるため確率的になる。この T を徐々に下げると安定したネットワークが構成できる。次にネットワーク全体がある状態となる実現確率は以下となる。

$$Q(E) = \frac{1}{Z} \exp\left(\frac{-E}{T}\right) \tag{8}$$

$Q(E)$: 実現確率
 Z : 確率分布規格化の定数

ここで Boltzmann Machine においてもエネルギー関数の増減を考える。温度 T を高い値に設定すると、エネルギー関数は増加を許されることから関数上を広い範囲で動くことができ、最小値を取りそうな場所を探索する。そして徐々に温度 T を下げてゆくことで、状態をエネルギー関数の最小値へと導くのである。

5. Boltzmann Machine の確率的検討

5.1 巡回セールスマン問題

組合せ最適化問題は、全ての組合せを求めれば最適解が得られる。しかし、変数の数が増加するにつれて組合せ総数は指数的に増加し、全ての組合せを求めることは不可能となる。たとえば組合せ最適化問題の代表例である“巡回セールスマン問題”は、ある決められた都市を訪問するとき、その全ての都市を1回だけ訪問し、かつ総距離が最短となるような都市訪問順序を求める問題である。この問題では都市数が決定変数となり、Fig.5に示すように訪問ルート数は増加する。このような問題はNP完全と呼ばれ、

都市数	ルート数	都市数	ルート数
3	1	10	181440
4	3	11	1814400
5	12	12	19958400
6	60	13	239500800
7	360	14	3113510400
8	2520	15	43589145600
9	20160		

Fig. 5 巡回セールスマン問題のルート数

すべての組み合わせを作成し、その中から最適解を求めるしかなかった。N が小さいときは、容易に求められるが、N が大きくなるとすべての組み合わせを探索することはできず、解を求めることができなくなる。

本論文では 10 都市の問題に Hopfield Model と Boltzmann Machine を適用することにより、基本的特性の考察を試みる。

5.2 ネットワーク表現

ニューラルネットワークで解く場合問題をネットワークの形で表現する必要がある。そこで問題は 10 都市であるため、 $10 \times 10 = 100$ 個のニューロンによりネットワークを表現し、かつ以下に示すの 3 つの制約がある。

- ・条件 1 : 各都市を 1 回だけ訪問 → それぞれの行に “1” は 1 つだけ存在
- ・条件 2 : 同時に複数回を訪問することができない → それぞれの列に “1” は 1 つだけ存在
- ・条件 3 : 全経路は短い方がよい。

上記条件を満たす場合に 0 となるエネルギー関数を作成する。条件 1 より 10 都市の場合を考えると、 j, X を用いると式(9)となり、同様に条件 2 の場合は式(10)となる。

$$E_1 = \sum_{X=1}^{10} \left(\sum_{j=1}^{10} x(X, j) - 1 \right)^2 \quad (9)$$

$$E_2 = \sum_{j=1}^{10} \left(\sum_{X=1}^{10} x(X, j) - 1 \right)^2 \quad (10)$$

次に条件3を i, X, Y を用いて表すと以下となる。

$$E_3 = \frac{1}{2} \sum_{i=1}^n \sum_{X=1}^n \sum_{Y=1}^n d(X, Y) x(X, i) \{x(Y, i+1) + x(Y, i-1)\} \quad (11)$$

以上より a, b, c はエネルギー関数 E の定数とし、式(12)、(13)比較して重み w_{ij} 、しきい値 θ_i を決定する。

$$E = aE_1 + bE_2 + cE_3 \quad (12)$$

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j - \sum_{i=1}^n \theta_i x_i \quad (13)$$

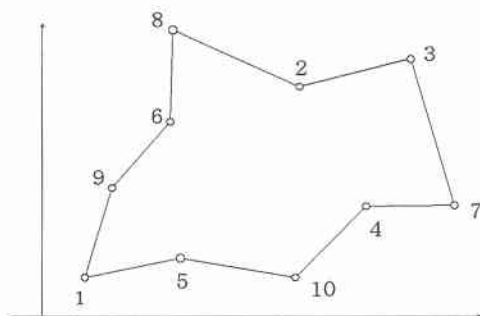


Fig. 6 10都市の位置

5.4 解析結果

Hopfield Model と Boltzmann Machine により Fig.6 に示すような、10 都市の巡回セールスマン問題を解き、結果を比較してみる。

Table 5 10都市の座標

1(1,1)	2(7,8)	3(11,8)
4(9,3)	5(4,2)	6(5,6)
7(12,3)	8(4,10)	9(2,4)
10(7,1)		

ここで、各初期状態は乱数により決定し、状態の数字は、乱数の取り出し回数を示している。また、この問題における最適解は Fig.6 に示した線分である。

(1) Hopfield Model による結果

時間 : K=100

初期状態	訪問順序
13	5-1-2-3-4-7-6-8-9-10
394	1-7-2-3-4-10-6-8-9-5
758	6-7-2-3-4-5-1-8-9-10
1352	6-7-2-3-4-5-1-9-10-8

(2) Boltzmann Machine による結果

温度変化関数 : $TEMP = 200 \exp(-0.00002K)$

時間 : K=450000

初期状態	訪問順序
13	8-2-3-7-4-10-5-1-9-6
394	9-6-8-2-3-7-4-10-5-1
758	1-5-10-4-7-3-2-8-6-9
1352	3-2-8-6-9-1-5-10-4-7

6. まとめ

Hopfield Model は収束速度は速いが、エネルギー関数の増加を許容されないことから、正解値に達することができずにそれぞれ異なった状態(局所解)に収束した。一方、Boltzmann Machine は、エネルギー関数の増加が許される(確率的探索)ため、局所解に陥ることがなく正解値が得られた。しかし、状態が収束するに要する時間は非常に大きい結果となった。加えて、Boltzmann Machine はプログラムの作り手に関わらず、正解値が得られる可能性が高いため、組合せ最適化問題には適している。しかし温度関数、エネルギー関数の係数 a, b, c 設定に経験的なものが要求される。

参考文献

- 1) D.E.Goldberg: Genetic Algorithm in Search, Optimization and Machine Learning. Addison-Wealy, 1983.
- 2) E.Aarts and J.Korst: Simulated Annealing and Boltzmann Machines - A Stochastic Approach to Combinatorial Optimization and Neural Computing, John Wiley, 1989.
- 3) 須藤敦史・星谷 勝・宮沢和樹: 遺伝的要素を考慮したインホ-タス・ザッ・リンクによる離散型変数を有するシステムの最適化, 土木学会論文集, No.519/ I -32, pp.223-232, 1995.
- 4) J.J.Hopfield: Neural Network and Physical Systems Emergent Collective Computational Abilities, Proc. of National Academy of Science USA, pp.2254-2258, 1982.
- 5) D.H.Ackley et al: A Learning Algorithm for Boltzmann Machines, Cognitive Science, 9-1, pp.147-169, 1983.