

迷路問題に対する免疫型進化アルゴリズムの適用

2004MM092 山本勇作

指導教員 高見 勲

1 はじめに

本研究では免疫ネットワークを模擬したアルゴリズムによって迷路問題を学習する事を目標としている。人間の優れた情報処理機構の中で免疫ネットワークに着目し、独自の考えをモデル化させ、アルゴリズム化し、プログラムによって検証する。他の免疫ネットワークの研究と大きく違う点は、微分方程式などを使わず代数方程式のみで学習を表現している点や、免疫ネットワークのとらえ方、初期知識を与えない、ペナルティやボーナスといったルールの考案などである。これらの新規性の優れた点として、簡潔な免疫ネットワークで、計算時間が短い事、またペナルティやボーナスによって学習や自律性を表現している点である。

2 免疫ネットワークに基づくモデル化

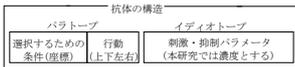


図 1: 抗体の構造

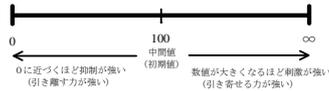


図 2: 濃度パラメータのとりえ方

図1が抗体の構造である。イデオトープとは抗体間関係の度合いを示していて、本研究では濃度という形で表す。その濃度パラメータのとりえ方を図2に示す。この濃度を更新することによって、学習や自律性を表現する。

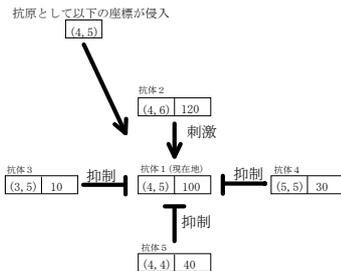


図 3: 抗原と抗体の関係

図3は抗原として侵入してきた座標情報に対して、その抗原の座標情報と抗体のパラトープが一致する抗体を呼び出し、その抗体同士が濃度によってどのような影響を受けるか表した図である。抗体1が現在地であり、この抗体1の濃度より他の抗体が高ければ刺激を受け(引き寄せする力が強い)、低ければ抑制を受ける(引き離す力が強い)ことになる。

具体的にどのような流れで学習するのか提示する。図4は初期の段階の迷路図である。マス上の数字は座標、下の数字はその座標の抗体の濃度である。抗体の選択は確率的に行う。

1. 現在地と周囲の座標情報が抗原として侵入。図4では現在地(2,1)上(2,2)左(1,1)右(3,1)の4つが抗原。
2. 上と左右の濃度が等しいので抗体の選択の確率は同じ。最短ルートである右の確率と、無駄足となってしまふ上と左が全て等しいので、無駄な動きをしてしまふ可能性がある。そこでペナルティを導入して濃度を更新し、最短ルートを進む確率を高くしていく。ペナルティとは一度行った場所(1エピソード内のみ)の濃度を下げる数値である。
3. 移動する際に現在地の濃度の更新を行う。

単に濃度が高い抗体を選択するのではなく、確率的に選択するようにした理由として、局所解に陥ることを回避し、大域的最適解を発見する探索を可能にするため(低い濃度の先に最短ルートがある可能性があるため)である。

ここでポテンシャルという概念を考案する。ポテンシャルとは、抗体の濃度にペナルティなどを加味した数字の合計で、ポテンシャルが高い方に行きやすくなる。ペナルティのみを考えた場合のポテンシャルの式が式(1)である。cは抗体の濃度、pは抗体に働くペナルティ(一定値)である。ペナルティの値は100に近づくほど平均ステップ数が安定したので99とする。

$$U = c - p \quad (1)$$

式(2)は抗体の選択の確率を決める式である。 $P_k(k = 1, \dots, N)$ は抗体kを選択する確率で、周囲の抗体の個数N個分存在する。図4では $N=3$ になる。 $\sum_{i=1}^N U_i$ は現在地の抗体に隣接する抗体のポテンシャルの和である。

$$P_k = \frac{U_k}{\sum_{i=1}^N U_i} \quad (k = 1, \dots, N) \quad (2)$$

式(3)は抗体の濃度の更新を行う式になる。Cは更新後の現在地の抗体の濃度で、 $\sum_{i=1}^M U_i$ は更新前の現在地の抗体と、周囲の抗体のポテンシャルの和である。 $M=N+1$ で図4では $M=4$ になる。

$$C = \frac{\sum_{i=1}^M U_i}{M} \quad (3)$$

3 シミュレーションと考察

図5は最短ルートが1通りしかなく、最短ルート以外は全て行き止まりになっているパターンである。S(1,1)からG(24,16)まで最短ステップ数60の迷路である。この迷路を100エピソード繰り返す。

図6は図5の学習結果のグラフである。エピソード11くらいまでにほとんど学習が完了しているが、時々ステップが多くなっている。その原因としては、選択する抗体は確率で決めているので、濃度が低くなったからといって選択されないわけではないためである。

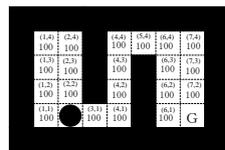


図 4: 迷路1

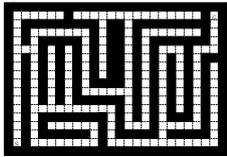


図 5: 迷路2

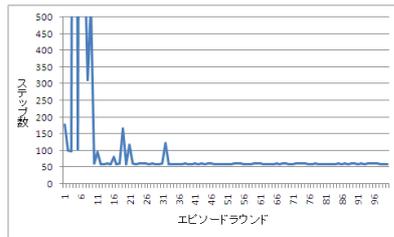


図 6: 迷路2のグラフ

図7は 図5の 迷路 について、座標 (7,16)(11,1)(12,6)(22,15)を 通路にした 迷路である。最短ルートが2通りあり、図5より複雑になっている。S(1,1)からG(24,16)まで最短ステップ数は60である。この迷路を100エピソード繰り返す。

図8は図7の学習結果のグラフである。最短ルートに落ち着くのがエピソード32ぐらいで、しかも図6よりも初期の学習回数がかかり多い。また一度最短ステップに安定してもエピソード37やエピソード46のように突然ステップ数が多くなってしまふ。これらの原因としては、最短ルートが2通りあるため、2つのルートの抗体でループしてしまつて、行き止まりのルートの抗体の濃度と均衡してしまつたためであると考えられる。解決策として、二度と通つてしまつた場所を通行禁止にしたり、ループする事を禁止するルールなど考えられる。しかし、通行禁止にした場所を通らないとゴールに行けない場合デッドロックになつてしまふし、ループを禁止するアルゴリズムが難しい。

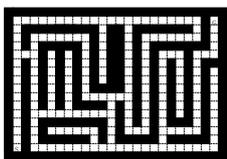


図 7: 迷路3

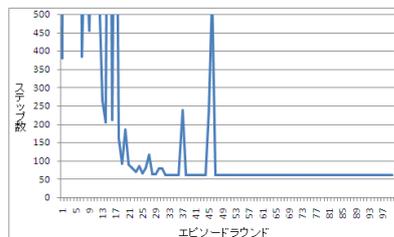


図 8: 迷路3のグラフ

図9は図5などと違い道幅が決まっていないので、最短ルートが複数パターンある場合である。最短ステップは48である。この迷路を200エピソード繰り返す。

図10が図9の学習結果のグラフである。最短ルートが複数パターンあるためペナルティのみでは全く学習していない。そこで方角ボーナスを考案する。

図11は方角ボーナスの例である。この場合ゴールの方向が 方向なので上と右の濃度を上げる。方向ならば上と左になるし、ならば下と左になるし、ならば下と右の濃度を上げる。もし軸上に方角がある場合その方角のみにボーナスとなる。bをボーナスとし、ポテンシャルを次の式(4)のようにする。ボーナスを50とする。

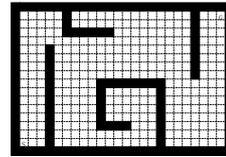


図 9: 迷路4

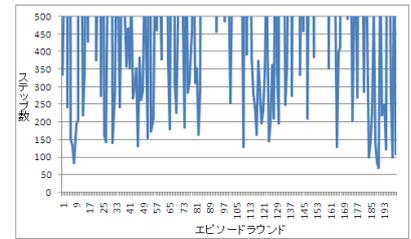


図 10: 迷路4のグラフ

$$U = c - p + b \quad (4)$$

図12が図9に対して、ボーナスとペナルティを併用して検証を行った場合のグラフである。図10のグラフと比べるとかなりステップ数は安定した。しかし、最短ステップ48とはかけ離れているステップ数が多く、学習しているとは言い難い。解決策として、最短ルートを更新した場合、更新したルートの濃度を上げる報酬というルールを考えた。しかし、うまく機能しないので今後更なる検証が必要である。



図 11: 方角ボーナスの説明

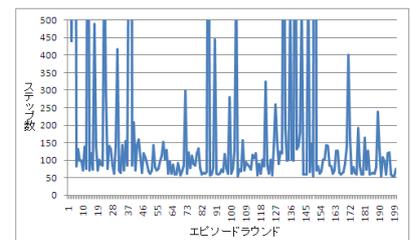


図 12: 迷路4のグラフ(ボーナス採用)

4 あとがき

本研究の成果として、考案した免疫ネットワークのモデルリングや数式で、幅1マスの迷路に関しては効率的な学習が見られた。

今後の課題として、どのシミュレーションでも言えることで、最大ステップ数が1000を超えてしまう場合が多く、最短ステップが60などの迷路からすると、あまりにも大きい数字である。また、迷路によって適用するルールを変えたりアルゴリズムを変えたりすることは汎用性がないので、どんな迷路に対しても適用できるアルゴリズムを考え、より複雑なものへと発展させたい。

参考文献

- [1] 小川長久,大林正直,小林邦和,呉本亮: 免疫回路網式強化学習法, 計測自動制御学会論文集, Vol. 43 No. 6, pp. 525-527 (2007).