

免疫系の抗体遺伝子組換え機構を取り入れた 遺伝的アルゴリズムによる学習の性質

Learning Properties of Artificial Immune System incorporating Antibody Gene Recombination

園田 隆志

Takashi Sonoda

富士ゼロックス株式会社 未来ワーク研究所

Future Work Laboratory, Fuji Xerox Co., Ltd.

The learning properties of antibody gene recombination in the immune system is investigated. Recently, we proposed the model of the immune system incorporating antibody gene recombination and analyzed it using the techniques of statistical mechanics. The analysis showed that gene recombination is effective in the adaptability of learning antigens. We also showed that analysis is in good agreement with results of numerical simulations. This work suggests the possibility of the immune system evolutionary obtained the gene recombination mechanism for development in ability of learning antigen, consequently obtained also antibody diversity. However, this research is based on the simple neural network model, "perceptron" and the adaptability was checked by simulations with the neural network model. In this paper we modify previous model to for genetic algorithm (GA) simulation and prove that the adaptability is satisfied in here. The analysis is in qualitative agreement with results of GA simulation.

1. はじめに

あらゆる抗原を認識するために、免疫系は、体細胞遺伝子の組換えによって多様性を発現している。免疫系は、特定の抗原を認識するように進化し、その過程において、ある時期に、組換えを取り入れ、多様性を増加させたという研究がある[Janeway 1993]。抗原の学習から、抗体の多様化へのステップは、学習機械進化と見ても興味深い。

我々は、先に、この点に注目し、免疫系を並列的な判別学習装置と考え、人工ニューラルネットワークモデルであるパーセプトロンでモデル化し、学習の性質を調べた[Sonoda 2002]。その結果、抗原の学習という点において、免疫系全体の記憶容量には、遺伝子組み替えは影響しないことがわかった。解析結果は、組換えと多様性との関係は明らかにしていないが、組換えにより、各抗体が学習できる抗原数に適応性が現れることが分かった。この研究では、パーセプトロンの内部表現に連続値表現を使っている。その理由は、連続値表現のパーセプトロンは、多くの研究者によって解析され、その振る舞いの理解には蓄積があり、それらの成果が利用できると思ったからである。しかしながら、遺伝子組換えの効果を調べるとい目的のためには、遺伝子のアミノ酸の表現として離散値を用いるのが適当であろう。離散値のパーセプトロンは、記憶容量は求められているが、そこでは素子数が無限大であると仮定されている。我々は、抗体を素子数有限のパーセプトロンと考えているが、そのようなパーセプトロンの振る舞いはまだ明らかにされていない。

そこで、本発表では、組換えを持つ免疫系の連続値表現での振る舞いが、離散表現でも成り立つかを、遺伝的アルゴリズム(GA)により調べ、その結果を報告する。

2. 抗体遺伝子組換え機構をもつ免疫系のモデル

我々はモデル化の基礎に Farmer らの 2 値表現のモデル

連絡先: 富士ゼロックス株式会社 未来ワーク研究所, 〒259-0157 神奈川県足柄上郡中井町境 430, Tel: 0465-80-2479, takashi.sonoda@fujixerox.co.jp

[Farmer 1990]と、Perelson らの組換え機構のモデル[Perelson 1996]を使っている。Farmer らのモデルでは、抗体の遺伝子を 2 値(± 1)表現し、さらに、その表現を、抗体の結合部位のアミノ酸配列とみなしている。さらに、抗原のアミノ酸配列も 2 値表現し、抗体と抗原の結合の親和性は、2 つの配列の相補的ビットの数で表現している。抗体遺伝子の組換え機構は、Perelson らが、GA シミュレーションで利用した方法に基づいている。1 つの抗体遺伝子は 2 つの同じ長さ $l/2$ の遺伝子セグメントから構成されると仮定する。それぞれの遺伝子セグメントは、 m 個の集団から選択される。この集団を、セグメントファミリーと呼ぶことにする。免疫系全体の遺伝子長を N とするとき、 m と抗体の長さ l には、 $N = lm$ という関係がある。

そこで、 μ 番目のセグメント I_i^μ と、 ν 番目のセグメント J_j^ν から構成される $\mu \cdot \nu$ 番目の抗体と、 ε 番目の抗原 ξ_i^ε との間の親和性を、次の式で定義する。

$$M_{\mu\nu}^\varepsilon = - \sum_{i=1}^{l/2} I_i^\mu \xi_i^\varepsilon - \sum_{j=l/2+1}^l J_j^\nu \xi_j^\varepsilon. \quad (1)$$

この親和性(1)がある閾値 θ を超えたときに、 $\mu \cdot \nu$ 番目の抗体は ε 番目の抗原を認識したとする。

3. 免疫系の記憶容量

親和性(1)は、パーセプトロン型の判別関数であることが分かる[Minsky 1969]。パーセプトロンは、統計力学的に解析されており、学習能力は詳細に調べられている。本研究では、抗体をパーセプトロンとみなし解析を進める。内部表現を連続値とするパーセプトロンは、その内部表現の規格化条件から球面(Spherical)パーセプトロンと呼ばれている。一方、離散値のパーセプトロンは、イジング(Ising)パーセプトロンと呼ばれている。

パーセプトロンが、入力パターンを学習し、正しく反応できるパターンの最大数を記憶容量と呼ぶ。パターン数を素子数で正規化した記憶容量比 $\alpha = p/l$ で議論することが便利である。球面パーセプトロンの記憶容量は、Cover が線形判別機の容量として幾何学的に論じ、 $\theta = 0$ のとき、 $\alpha_s = 2$ であることを導いている[Cover 1965]。また、Gardner は、無秩序系の解析のため

に開発されたレプリカ法と呼ばれる統計力学の計算手法を使った議論より、同様の結果を導いている[Gardner 1988]。イジングパーセプトロンには、Cover の幾何学的な考察を拡張することは難しいが、統計力学的手法により $\alpha_1 \approx 0.82$ という値が得られている。この解は、Derrida [Derrida 1991]らによる数値実験の結果とも一致しており、正しい解であると信じられている。

我々は、Gardner の手法を連続値表現の親和性(1)に適用し、組換え機構を持つ免疫系の記憶容量を計算した。 $\mu \cdot \nu$ 番目の抗体の記憶容量を $\alpha_{\mu\nu}$ とし、次の式を得ている。

$$\sum_{\nu} \alpha_{\mu\nu} = \alpha_s / m, \quad \sum_{\mu} \alpha_{\mu\nu} = \alpha_s / m \quad (2)$$

ここで、 α_s は、球面パーセプトロンの計算で得られている記憶容量 2 である。この結果は、各抗体の記憶容量は、単独では評価できず、免疫系全体の記憶容量しか評価できないことを示している。さらに、セグメントファミリーの要素数 m が、十分大きければ、各抗体が学習する抗原数は任意に設定することも簡単な考察から理解できる。(2)の片方、例えば左の式の両辺を、 μ に対して和をとると、免疫系全体の記憶容量は、全素子数の 2 倍であり、組換えが影響を与えないことが分かる。連続値表現では、Cover の考察により、素子数が有限の場合にも、記憶容量は素子数の 2 倍であることが知られている。しかしながら、有限長ではゆらぎがあり、複数のパーセプトロンを、並列に組み合わせた場合、ゆらぎの積み重ねにより記憶容量は低下する。一方、組換えを取り入れた場合、複数のパーセプトロンが組換えによりあたかも 1 つのパーセプトロンのように振る舞い、ゆらぎは減少する。このことはシミュレーションによっても確認している。

さらに、離散値表現に対しても親和性(1)の記憶容量を計算し、(2)の α_s を α_1 と置き換えただけの同様の結果を得た。連続値の場合には有限長パーセプトロンの記憶容量は明らかにされていないが、数値計算では、素子数に比例しないことが示されている[Derrida 1991]。したがって、複数の有限長パーセプトロンの記憶容量と、同じ素子数を持つ 1 つのパーセプトロンの関係は不明であり、ここから記憶容量について評価することは困難である。しかしながら、組換えによりあたかも 1 つのパーセプトロンのように振る舞うことは、記憶容量(2)の結果から期待される。

4. 遺伝的アルゴリズムによるシミュレーション結果

離散値表現の免疫系の記憶容量を、GA シミュレーションで確認する。アルゴリズムは、標準的な教科書(例えば、[Goldberg 1989])に載っているようなシンプル GA である。シミュレーションは、各抗体に等しい数の抗原を割り当て、すべてを学習するという課題で行った。適応度関数は、各抗体が学習できていない抗原数の自乗誤差の和とした。この関数を最小にするように世代交代を進める。交差率は 0.002、変異率は 0.005 である。ただし、セグメント間での交差率は 0.01 としている。セグメント遺伝子の間には、意味のないスペースを含んでいることが知られているが、この部分の交差率を総和したものと表現している。ポピュレーションは 200 とした。1000 回の世代交代を行い、全抗原を学習した試行の回数を、全試行回数(100 回とした)で割った値を、学習率とした。抗原数を変化させ、各抗原数に対して求めた学習率をプロットしたのが図 1 である。全遺伝子長 N を 1600 とし、パーセプトロンの素子数 l を 200(●), 400(○), 800(△)と変化させ学習率を求めた。それぞれ、組換えのある場合(黒記号、Recom.)、組換えのない場合(白記号、Parallel)についてシミュレーションを行っている。

その結果、記憶容量の 0.82 は再現していないが、組換えが学習率の向上に働いていることは確認できる。セグメント数が多

いほど学習率が向上していることが分かる。しかしながら、この結果が、1 つのパーセプトロンのように振る舞うことによるゆらぎの減少に起因するかについては、さらに考察が必要である。

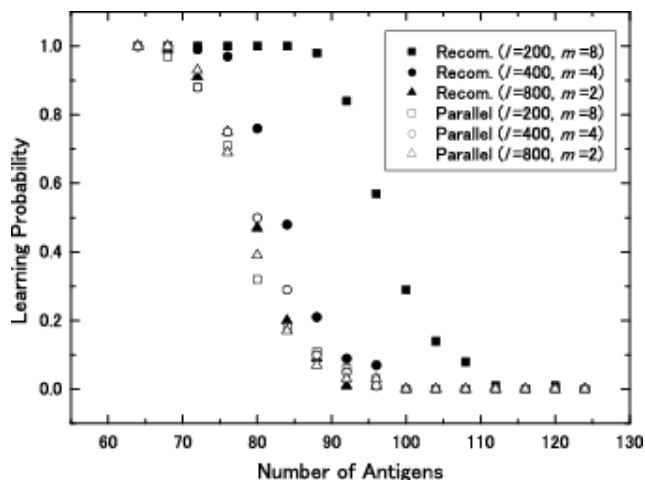


図 1. 組換え機構を持つ免疫系のシミュレーション結果

5. おわりに

連続値表現を持つ免疫系モデルの振る舞いを GA シミュレーションによって調べた。有限長パーセプトロンは、離散値の場合、その振る舞いを解析的には説明できておらず、組換えの効果については、さらなる考察が必要である。しかしながら、シミュレーションは、組換え機構が、学習率に有効に働くことを示した。学習課題に対する GA の改善方法とも捉えることができる。

参考文献

- [Cover 1965] T. M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Transactions on Electronic Computers*, **EC-14** 326-334 (1965)
- [Derrida 1991] B. Derrida, R. B. Griffiths and A. Prugel-Bennet, Finite-size effects and bounds for perceptron models, *Journal of Physics A: Mathematical and General*, **24** 4907-4940 (1991)
- [Farmer 1990] J. D. Farmer, N. H. Packard and A. S. Perelson, The immune system, adaptation and machine learning, *Physica*, **D22** 187-204 (1990)
- [Gardner 1988] E. Gardner, The space of interactions in neural network models, *Journal of Physics A: Mathematical and General*, **21** 257-270 (1988)
- [Goldberg 1989] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley Publishing Company, Reading, 1989)
- [Janeway 1993] C. A. Janeway, Jr., How the immune system recognizes invaders, *Scientific American*, **269** (3) 72-79 (1993)
- [Minsky 1969] M. L. Minsky and S. A. Papert, *Perceptrons* (MIT Press, Cambridge, 1969)
- [Perelson 1996] A. S. Perelson, R. Hightower and S. Forrest, Evolution and somatic learning in V-region genes, *Research in Immunology*, **147** 202-208 (1996)
- [Sonoda 2002] T. Sonoda, Adaptation of learning antigens by gene recombination in the immune system, *Journal of Physics A: Mathematical and General*, **35** 5973-5983 (2002)