

# 遺伝的アルゴリズムによるニューラルネットの最適化と 膜貫通セグメント領域の予測\*

5P-06

水田 智史<sup>†</sup>弘前大学理工学部<sup>‡</sup>杉山 佳明<sup>§</sup>弘前大学理学研究科<sup>¶</sup>清水 俊夫<sup>||</sup>

弘前大学理工学部

## 1 はじめに

ニューラルネットは、複数のニューロンと呼ばれる処理ユニットとそれらを相互に接続する結合線から構成されており、その動作は、ニューロンの数、ニューロン間の結合形態、ニューロン間の結合の強度、およびニューロンの閾値によって決まる。これらの中で、結合の強度と閾値については誤差逆伝播法や共役勾配法等により効率良く求めることが可能であるが、必要なニューロンの数とそれらの結合形態については系統的に求めることは難しく、試行錯誤により求めるのが一般的であった。

そこで我々は、ニューロンの数と結合形態を含めたこれらすべてを遺伝的アルゴリズム (GA) を用いることによって統一的かつ系統的に求めるための手法の一つを開発し、いくつかの論理演算問題に適用して良好な結果を得た [1, 2]。

本研究では、そこで得られた手法を改良し、膜たんぱく質のアミノ酸配列から膜貫通セグメント領域を予測するという具体的な問題に適用することを検討する。

## 2 本研究の手法

本研究では層状の順方向性ニューラルネットのみを対象とし (図 1)、回帰ネットワークについては考えない。このネットワークの構成要素の中で、入力層と出力層に属するニューロンの数は扱う問題によって決められているので、隠れ層に属するニューロン (隠れニューロン) の数、ニューロン間の結合形態、すべての結合線の強度、そしてすべてのニューロンの閾値を

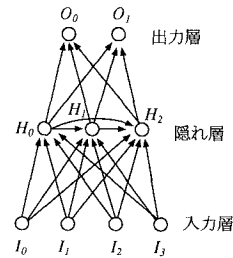


図 1: 層状順方向性ニューラルネット

GA により決定することになる。そのためにもまず個々のネットワークを染色体として表現し、それらに GA を適用することによって最適な個体を求める。

なお、本研究では

$$f = N_c + \frac{N_w^{max}}{N_h \cdot N_w^{max} + N_w}$$

を適応度関数とし、 $f$  を最大にするネットワークを最適と定義する。ここで、 $N_c$  は教師信号と実際の出力値とが一致した数、 $N_w, N_h$  はそれぞれ結合線、隠れニューロンの数である。また、 $N_w^{max}$  は  $N_h$  を固定したときの結合線の最大数を示す。

GA は一般に

- 選択
- 交叉
- 突然変異

の 3 つの操作から構成されるが、本研究の手法では隠れニューロンの数によって染色体の長さが異なるので、単純に交叉処理を施すことができない。そのため、先の研究 [1, 2] では選択と突然変異のみによって GA を実行した。

ここで、突然変異としては

- 隠れニューロンの付加、削除
- 結合線の付加、削除

\*Neuralnet Optimization by Genetic Algorithms and Prediction of Transmembrane Segment

<sup>†</sup>Satoshi Mizuta

<sup>‡</sup>Faculty of Science and Technology, Hirosaki University

<sup>§</sup>Yoshiaki Sugiyama

<sup>¶</sup>Graduate School of Science, Hirosaki University

<sup>||</sup>Toshio Shimizu

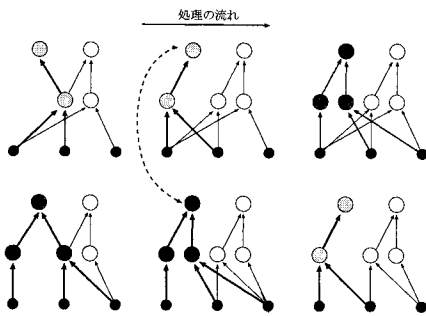


図 2: 交叉

- 結合線の強度の修正
- ニューロンの閾値の修正

を導入し、選択操作にはルーレット戦略を採用している。先の研究の結果をまとめたものを表 1 に示す。

実験	正解数/全出力数	最適
XOR	4/4	○
6XOR	24/24	○
4XOR-2AND	24/24	○
2XOR-2AND-2OR	24/24	○
4-bit pattern copying	64/64	○
8x8-encoder/decoder	62/64	×
4x4-encoder/decoder	16/16	○

表 1: 参考文献 [1, 2] の結果のまとめ

本研究では新たに交叉を施す手法を開発し、GA の操作に含めることが可能となった。先に述べたように、本手法では染色体の長さが個体によって異なるため、通常の交叉のように単純に染色体をある点で切断して交換することはできない。そこで、染色体レベルで交叉を行うのではなく、ニューラルネットの構成に立ち戻って考えることにした。

図 2 は本研究で新たに導入した交叉の概要を示したものである。交叉は、2 個のニューラルネットの間で対応する出力ユニットをそっくり交換することで行う。そこでまず、対象とする出力ユニットに関する部分を結合線を含めてコピーする。その後、それらを 2 個のニューラルネットの間で交換する。こうすることで、他の出力ユニットの出力値に影響を与えることなく交換することが可能となる。

交叉を新たに導入した結果、先の結果で 2 ビットのエラーを出力していた 8x8-encoder/decoder の実験において、エラーの出力が 0 になった。しかし、最適なネットワーク構成では隠れニューロンは 3 個である

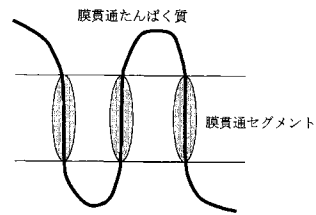


図 3: 膜貫通たんぱく質

のに対し、得られた結果では 8 個の隠れニューロンが存在し、今のところ最適な構造を得ることはできていない。

### 3 膜貫通セグメント領域の予測

膜たんぱく質は約 20 種のアミノ酸の 1 次元配列で構成されており、細胞膜内に存在して膜内外の物質のやりとりにおいて重要な役割を果たしている (図 3)。ここでは、膜貫通セグメント領域が特徴的なアミノ酸配列パターンを有しているものと考え、アミノ酸配列をニューラルネットによって解析して膜貫通セグメント領域を予測する。これは、アミノ酸の 1 次元配列から膜貫通たんぱく質の立体構造を予測するという問題につながるもので、バイオインフォマティクスの分野において重要な問題として取り上げられているものである。

その際、ニューラルネットの入力データとなるアミノ酸配列をどのように符号化するかということが重要な問題となるが、現在、2 進数による単純な符号化や、個々のアミノ酸の性質やそれぞれの相互関係などを採り入れた符号化など、様々な方法を用いて実験を進めている。

### 参考文献

- [1] S. Mizuta, T. Sato, D. Lao, M. Ikeda and T. Shimizu, "Structure Design of Neural Networks Using Genetic Algorithms", *Complex Systems*, **13** (2001) 161-175.
- [2] S. Mizuta and T. Shimizu, "Designing Neural Networks by Genetic Algorithms", *Proceedings of the International Conference on Artificial Intelligence, IC-AI'2001* (2001) 674-680.