

# Clonal Selection Algorithm におけるクラスタリングを用いた免疫学的記憶細胞の生成方法の提案

A Proposal of Generating Method of Clustering Based Immunological Memory Cell in Clonal Selection Algorithm

○<sup>1</sup> 鎌田 真,  
○<sup>1</sup> Shin Kamada,

<sup>2</sup> 市村 匠  
<sup>2</sup> Takumi Ichimura

<sup>1</sup> 県立広島大学大学院 総合学術研究科 経営情報学専攻

<sup>1</sup> Graduate School of Comprehensive Scientific Research, Prefectural University of Hiroshima

<sup>2</sup> 県立広島大学 経営情報学部 経営情報学科

<sup>2</sup> Faculty of Management and Information Systems, Prefectural University of Hiroshima

**Abstract:** The clonal selection principle explains the basic features of an adaptive immune response to a antigenic stimulus, according to the idea that only those cells that recognize the antigens are selected to proliferate and differentiate. This paper explains a computational implementation of the clonal selection principle that explicitly takes into account the affinity maturation of the immune response. Especially, the immunological memory cells are generated to record the specific antibodies by using clustering results.

## 1 はじめに

人工免疫システム (Artificial Immune System; AIS) は、多くの生物に備わっている免疫系を応用したシステムであり、パターン認識や最適化などの様々な分野に適用されている。人工免疫システムには、Burnet 氏が提唱したクローン選択理論 (Clonal Selection Theory) を元にしたクローン選択アルゴリズムがある。抗体の増殖過程では主に体細胞超変異 (Somatic Hypermutation; HM) とレセプター編集 (Receptor Editing; RE) の2つの手法があるが、Gao 氏は RECSA と呼ばれる、RE を取り入れたクローン選択アルゴリズム [1] を提案し、これを巡回セールスマン問題 (Traveling Salesman Problem; TSP) に適用し、非常に高い性能があることを示した。

Gao 氏の RECSA では、免疫系本来の機能である免疫学的記憶については実装されていない。市村らは、親和性が高い抗体を記憶細胞に保存する手法として、記憶細胞を導入した免疫学的記憶モデル (Clonal Selection Algorithm with Immunological Memory; CSAIM)[2] を提案した。

本論文では、生成された抗体のクラスタリング結果を用いて記憶細胞を生成する手法を提案する。記憶細胞に変化する抗体は特定の訓練事例に対して反応するように学習され、また、正答できない訓練事例が表れた場合、それに適応する新たな抗体が生成される。提案手法

の CSAIM を文献 [3] の心疾患データベース Coronary Heart Disease Database; CHD DB) に適用し、実験結果をここに報告する。

## 2 クローン選択アルゴリズム

Gao 氏の RECSA の適用事例は TSP であった。我々は提案手法を CHD\_DB に適用するために、抗体の構造及びアルゴリズムを次節に示すように修正した。

### 2.1 抗体の構造

提案手法の適用事例である CHD\_DB の各訓練事例は、表 1 に示す 8 つの入力 (ORIGIN, EDUCATE, TOBACCO, ALCOHOL, SBP, DBP, TC, LVH) と、その出力 (CHD) から構成される。これに対して、抗体を図 1 に示すように、これら 8 つの入力に対する重み  $w_i (i = 1, 2, \dots, k)$  と、出力 (CHD) に対するしきい値  $\theta$  で表す。これをパラトープ  $P = (w_1, \dots, w_k, \theta)$  とする。

### 2.2 Hypermutation と Receptor Editing

HM は、パラトープ  $P = (w_1, \dots, w_k, \theta)$  から  $w_i$  または  $\theta$  を選択し、式 (1) を行う。ここで、 $\Delta w, \Delta \theta$  は  $-\gamma_w \leq \Delta w \leq \gamma_w, -1 \leq \Delta \theta \leq \gamma_\theta$  を満たす。

$$w_i = w_i + \Delta w, \theta = \theta + \Delta \theta \quad (1)$$

RE は、パラトープ  $P = (w_1, \dots, w_k, \theta)$  に対して、図 2 のようにランダムに選択された 2 点間の並びを入

表 1: CHD\_DB の訓練事例の要素

Name	Data Item	Value
ORIGON	National Origin	0=Native-born; 1=Foreign-born
EDUCATE	Education	0=Grade School; 1=High School, not graduate; 2=High School, graduate; 3=College
TOBACCO	Tobacco	0=Never; 1=Stopped; 2=Cigars or Pipes; 3=Cigarettes (<20/day); 4=Cigarettes (≥20/day)
ALCOOHOL	Alcohol	[0,60.7] (oz/mo) (Continuous Value)
SBP	Systolic Blood Pressure	[54,254] (mmHg) (Continuous Value)
DBP	Diastolic Blood Pressure	[35,151] (mmHg) (Continuous Value)
TC	Cholesterol	[83,339] (mg/dl) (Continuous Value)
LVH	Left Ventricular Hypertrophy	0=None; 1=Definite or Possible
CHD	Development of CHD	0=No; 1=Yes

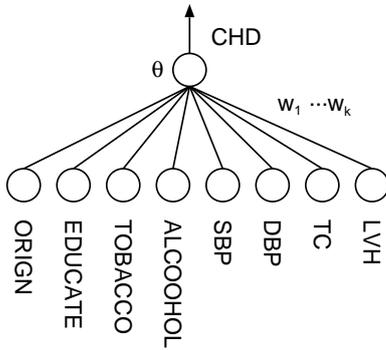


図 1: 抗体の構造

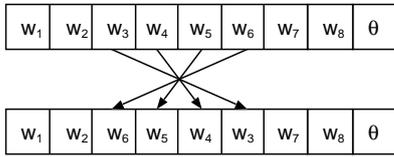


図 2: RE の例

れ替える。

### 2.3 親和性の計算

各抗体の親和性は、以下の式 (4) により計算される。ここで、 $x^P$  は各訓練事例を表し、 $x_{CHD}^P$  は各訓練事例が CHD であるかどうかを表す 0, 1 の変数である。 $tr\_num$  は全ての訓練事例の数である。

$$f(x^P) = \begin{cases} 1 & \text{if } |\sum_{i=1}^k w_i x_i^P - \theta| \geq E_{sim} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$g(x^P) = \begin{cases} 1 & \text{if } f(x^P) = x_{CHD}^P \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$\text{親和性} = \sum_{p=1}^{tr\_num} g(x^P) \quad (4)$$

### 2.4 アルゴリズム

クローン選択アルゴリズムを以下に示す。

Step 1)  $m$  個の抗体 (パトープ) を作成する。 $w_i, \theta$  は  $-\gamma_w \leq w_i \leq \gamma_w, -1 \leq \theta \leq \gamma_\theta$  の範囲でランダムに初期化される。

Step 2) 各抗体に対する親和性を計算し、親和性の高い順に抗体を並び替える。

Step 3) 抗体の生成を次のように行う。

Step 3-1)  $m$  個の抗体の中から、親和性の高い順に  $n(1 \leq l \leq n)$  個の抗体を選択する。

Step 3-2)  $t$  世代に 1 回、 $c(m$  の数の  $r\%$  とする) 個の新しい抗体をランダムに作成し、その中で  $m$  個の抗体よりも親和性が高い抗体があれば、入れ換える。

Step 4) 選択した  $n$  個の抗体を親とする。エリートプールを  $n$  個作成する。

Step 5) 各エリートプール  $l$  で、親の抗体のクローンを  $p_l$  個作成する。

$$p_l = \text{round} \left( \frac{(n-l)}{n} \times Q \right) \quad (5)$$

ここで、 $\text{round}()$  は、小数点第 1 位を四捨五入し、整数にする関数である。 $Q$  は定数である。

Step 6) 作成した各クローンに対して、HM 又は RE を 1:1 の割合で行う。

Step 7) 各エリートプールの中で、もっとも親和性の高いクローンを選択する。

Step 8) 親の抗体ともっとも親和性の高いクローンの親和性を比較し、クローンの方が高ければ、親の抗体をクローンで置き換える。ここで、現在の世

代が，最大世代数  $G_{max}$  に達すれば，プロセスを終了する．そうでなければ Step 3) に戻る．

### 3 免疫学的記憶細胞

HM, RE により生成された  $n$  個の抗体を記憶細胞に保存するアルゴリズムを以下に示す．図 3 はシステムのフローチャートであり，図 4 は生成された抗体のクラスタリング結果から記憶細胞を生成する過程を示している．

次節で示す学習アルゴリズムでは，文献 [2] のマッチングアルゴリズムにより求められる訓練事例の部分集合に対して学習がされる．ここでマッチングアルゴリズムは，訓練事例を  $\vec{d} = (d_1, \dots, d_i, \dots, d_k)$ ，抗体を  $\vec{h} = (h_1, \dots, h_i, \dots, h_k)$ ， $\vec{d}$  の中で値が最も小さい要素を  $d_j$ ， $d'_i = d_i \times \frac{h_j}{d_j}$  ( $d_i \neq 0 \wedge h_i \neq 0$ ) とした時， $\vec{d}$  と  $\vec{h}$  のユークリッド距離が，あらかじめ決められた距離  $\mu_\theta$  以下であれば，一致したと判定するアルゴリズムである．

#### 3.1 学習アルゴリズム

Step 1) 誤答する訓練事例に対して，式 (6)，(7)，(8) により，誤差を求め，重みを更新する．

$$O = \sum_{i=1}^k w_i x_i \quad (6)$$

$$E = \frac{1}{2} \delta^2 = \frac{1}{2} (\theta_q - O)^2 \quad (7)$$

$$w_i = w_i + \eta \delta x_i \quad (8)$$

ここで， $q (1 \leq q \leq IM)$  は  $IM$  個に分類されたクラスタを示す添字であり， $\eta (0.1 < \eta < 1.0)$  は学習率である．

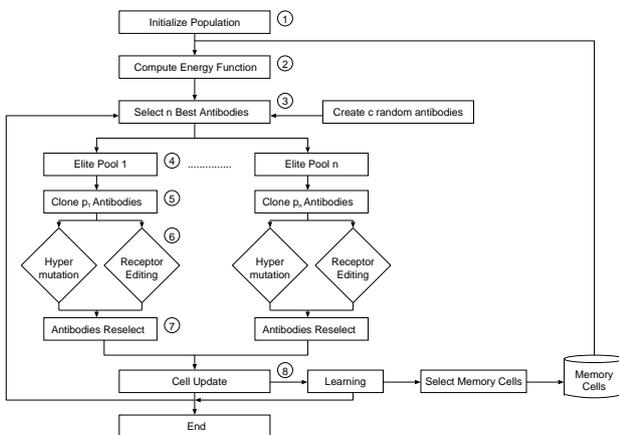


図 3: 記憶細胞を用いた免疫システムのフローチャート

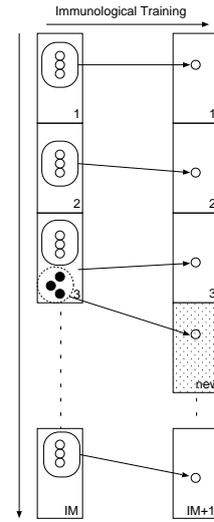


図 4: 記憶細胞の生成

Step 2) Step 1) のプロセスを，訓練事例の部分集合に含まれる全ての事例に対して  $t_{IM}$  回繰り返すか， $E$  があらかじめ決められた定数  $E_{min}$  より小さくなるまで繰り返す．

#### 3.2 記憶細胞生成アルゴリズム

Step 1) 学習を行った  $n$  個の抗体を  $IM$  個のクラスタに分類する．

Step 2) 各クラスタの代表となる抗体に対してマッチングアルゴリズムにより訓練事例の部分集合を定め，学習アルゴリズムにより学習を行う．学習を行った抗体は記憶細胞に保存される．なお，代表となる抗体は，クラスタ内で  $\theta$  が中央値の抗体とする．

Step 3) 次の手順で示す条件を満たす場合， $IM + 1$  個目とする抗体を記憶細胞に保存する．

Step 3-1) 記憶細胞に保存されている抗体の訓練事例に対して，誤答する訓練事例を抽出する．

Step 3-2) 誤答する訓練事例の中で，マッチングアルゴリズムによるユークリッド距離が  $\mu_\theta$  よりも大きい事例の数が，ある値以上ならば，HM, RE によって新しく抗体を生成し，これらの事例を訓練事例の部分集合として定める．

Step 3-3) 生成した抗体に対して，学習アルゴリズムによる学習を行い，正答する訓練事例があれば， $IM + 1$  個目として記憶細胞に保存する．

## 4 実験

文献 [3] の CHD\_DB には、4 つのデータセット、“Train\_A”、“Train\_X”、“Train\_Y”、“Train\_Z” と 1 つのテストデータセット “Test” がある。本論文では、RECSA と提案手法である CSAIM を “Train\_A” に適用し、“Test” を使って分類結果を検証する。“Train\_A” と “Test” の訓練事例数はどちらも 13,000 で、CHD=1 と CHD=0 の割合は 1:1 である。パラメタは  $G_{max} = 100$  ,  $m = 150$  ,  $n = 100$  ,  $Q = 50$  ,  $t = 10$  ,  $c = 10$  ,  $\gamma_w = 1$  ,  $\gamma_\theta = 1$  ,  $E_{sim} = 0.05$  ,  $IM = 10$  ,  $t_{IM} = 50$  ,  $E_{min} = 0.001$  ,  $\eta = 0.1$  ,  $c_{max}^{memory} = \frac{1}{2}n$  ,  $\mu_\theta = 0.3$  とした。

表 2: 実験結果 (正答率)

	Train_A	Test
RECSA	62.3%	58.7%
CSAIM	99.6%	99.4%

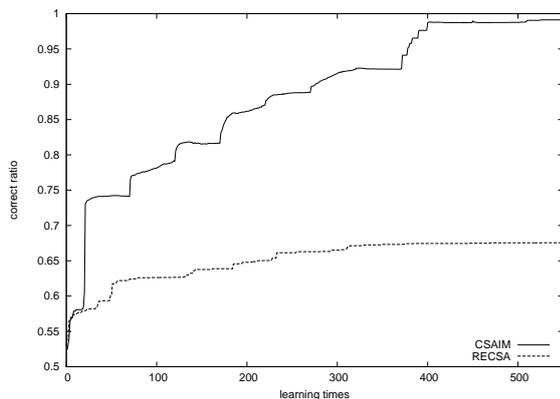


図 5: 正答率

表 2 は “Train\_A” と “Test” に対する正答率であり、図 5、図 6 は各世代ごとに計算した正答率と記憶細胞の数である。図 5 において、RECSA と CSAIM とともに 20 世代付近までは同じような曲線になっているが、CSAIM では図 6 で示すように、記憶細胞が生成される度に特定の訓練事例に対する学習がされ、正答率が上昇している。

## 5 おわりに

市村らは Gao 氏の RECSA に免疫学的記憶を取り入れた CSAIM を提案し、特定の訓練事例に対して反応する記憶細胞を生成するモデルを構築した。

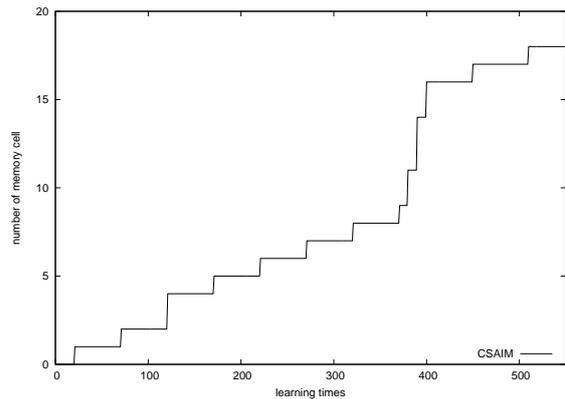


図 6: 生成されている記憶細胞の数

本論文では、クラスタリングの結果を用いて記憶細胞を生成する手法を提案し、RECSA と比べて非常に高い正答率を出すことができた。さらに、記憶細胞を用いたことにより RECSA と比べて計算時間を半分以下にすることができ、CSAIM の有効性を十分に検証できたと言える。今後は、CSAIM を “Train\_X”、“Train\_Y”、“Train\_Z” に対しても適用していく予定である。

## 参考文献

- [1] S. Gao, H. Dai, G. Yang and Z. Tang, “A novel clonal selection algorithm and its application to travelling salesman problem”, IEICE Trans. Fundamentals, Vol.E90-A, pp.2318-2325(2007)
- [2] 市村 匠, 山口 崇志, “種分化を取り入れた免疫システムの提案”, 第 27 回ファジィシステムシンポジウム講演論文集, pp.480-485(2011)
- [3] M. Suka, T. Ichimura and K. Yoshida, “Development of coronary heart disease databases”, Proc. of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES2004), Vol.2, pp.1081-1088(2004)

## 連絡先

〒 734-8558  
 広島市南区宇品東一丁目 1-71  
 県立広島大学 経営情報学部  
 市村 匠  
 E-mail: ichimura@pu-hiroshima.ac.jp