

論文の内容の要旨

論文題目 胸腺教育機構に基づく人工免疫系の研究
—自己・非自己の識別とパターン識別—

氏名 井川 数志

近年、人工免疫系 (Artificial Immune Systems, AIS) が計算科学, 工学の分野で注目されている。人工免疫系とは, 何らかの問題を解決するために, 理論免疫学や免疫系で観測される免疫応答, 原理やモデルから想起された計算手法である。その中でも, 従来から盛んに研究, 議論されているものの1つにネガティブ選択アルゴリズム (Negative Selection Algorithm) がある。変化検知や異常検知 (Change/Anomaly Detection) で研究され, さらなる発展が期待されている。ネガティブ選択アルゴリズムは, ある種の問題には適していると言われる一方で, 以下の3つの理由からパターン識別 (Classification) には不向きとも指摘される。

- 疑似細胞獲得過程が非効率であること。
- 過学習を防ぐ仕組みをもっていないこと。
- 疑似細胞がクラスに関して断片的な情報しかもっていないこと。

このため, 識別問題への適用については, 現在も議論の渦中にある。実際に, 複数のクラスで構成される一般的な識別問題への適用例は少なく, 汎用的な枠組みは存在していない。

そこで本研究では, ネガティブ選択アルゴリズムを一般的なパターン識別に適用できるように拡張し, 新しい人工免疫系, Artificial Negative Selection Classifier (ANSC) を提案する。従来, 識別問題に不向きであるとされていたネガティブ選択アルゴリズムを用いて, 汎用的かつノイズ耐性も有する識別器を構築できることを示す。そして, 上述した3つの問題点を解決し, ネガティブ選択アルゴリズムが識別問題にも有効な手段であることを示す。本研究は新しい識別器の確立とともに, 人工免疫系の議論に一石を投じる具体的研究の1つである。

本論文の構成は以下のとおりである。

第2章では, 免疫系と人工免疫系について概説する。まず, 免疫系の概要を述べ, 適応免疫系についてパターン認識の観点から解説する。そして, 人工免疫系において, 免疫系をモデル化する方法を述べる。そのため, 免疫系の機能を用いて問題を解決するためのフレームワーク, 細胞の抽象モデル化について記す。そして, 人工免疫系の現状を調べ, 本研究に至る研究の流れを明らかにする。

第3章では, ネガティブ選択アルゴリズムを用いた人工免疫系 (Artificial Negative Selection Classifier, ANSC) を導入する。胸腺教育機構における自己・非自己の確立から得られたネガティブ選択アルゴリズムをパターン識別用に拡張する。特に, 各細胞に可変

認識閾値を導入する。ANSCはこの拡張されたアルゴリズムと、未知のデータを識別する判定アルゴリズムから構成される。そして、人工データ集合を用いてANSCの挙動を視覚的に確認し、それらのデータ集合に対する統計的な特性を明らかにする。さらに、ノイズに対する考察を行い、ノイズの影響を緩和する方法を提案する。この章の最後には、可変・固定認識閾値による結果を比較・検討し、可変認識閾値の重要性を考察する。

第4章では、実世界データ集合を用いて、ANSCの特性を考察する。ここでは、ANSCの識別精度を他の識別器のものと比較し、ANSCの独自性を述べる。また、統計的観点から考察を行い、人工免疫系の識別器であるArtificial Immune Recognition Systemとの比較・検討を行う。さらに、本論文で導入した可変認識閾値の利点について考察する。

第5章では、文書の識別を行う。文書としてHTML文書集合を用いる。期待情報量を用いて文書から特徴ベクトルを抽出し、識別を行う。そして、ベイズ推定の結果と比較・検討する。また、特徴ベクトルの次元とANSCの挙動についての考察も行う。

第6章では、人工リンパ球の形状を超楕円型に拡張する。まず、人工リンパ球の拡張方法を述べ、その変異過程に回転・拡大縮小操作を導入する。そして、その挙動を確認し、特性を考察する。

第7章は結論である。