

タンパク質および遺伝的アルゴリズムの基礎学習

田中 裕也

1 前月からの課題

- タンパク質の基礎学習
- GA, DGA の基礎学習
- コインの公開実験

2 タンパク質の基礎学習

2.1 タンパク質とは

人間の体内で様々な働きをしているタンパク質は, 20 種類のアミノ酸の組み合わせで構成される. そのため, タンパク質には, その組み合わせにより, 膨大な種類が存在し, 各々が独自の性質を保持している. Fig. 1 にアミノ酸の構造を示す. タンパク質を構成する 20 種類のアミノ酸は, 側鎖と呼ばれる原子団の違いによってのみ異なり, 2 つのアミノ酸のアミノ基とカルボキシル基が縮合反応を起こすことでペプチドが形成される. ペプチドが複数つながったものをポリペプチドといい, これがタンパク質を成す.

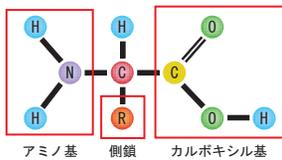


Fig. 1 アミノ酸の構造

2.2 タンパク質の構造

Fig. 2 に示すようにタンパク質の構造は, 一次構造から四次構造まで存在する. 一次構造は, アミノ酸配列を示し, 基本的な 3 次元構造は, 三次構造に当てはまる. 三次構造中に存在する規則正しいパターンが二次構造にあたり, 主なものとして α ヘリックスや β シートなどが存在する. 四次構造とは, 三次構造が複数集まって構成される構造である.

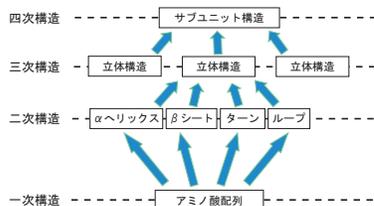


Fig. 2 タンパク質の階層構造

3 GA, DGA の概要

3.1 GA

GA とは, 生物の進化を工学的に模倣した学習的アルゴリズムである. 自然界における生物の進化過程におい

ては, ある世代を形成している個体の集合, すなわち母集団の中で, 環境への適合度の高い個体が高い確率で生き残るように選択される. さらに交叉や突然変異によって, 次の世代の母集団が形成されていく. これらの選択, 交叉, 突然変異といった遺伝的操作を繰り返すことによって最適解を得ることができる.

3.2 DGA

DGA は, 母集団を複数のサブ母集団に分割し, 各サブ母集団内で独立して GA の遺伝的操作を行うアルゴリズムである. サブ母集団に各プロセッサを割り振り, 計算をさせることで計算負荷を分割することができる. また, 各サブ母集団の間で移住と呼ばれる個体の交換を行うため, 多様性が維持され, 早熟収束によって局所解へ収束するといった問題を解消することができる.

4 コインの公開実験

本実験では, 1 人がコイン 5 枚を投げ, 表の出たコインの枚数を数える操作を 1 回試行として全部で 20 回試行を行った. なお, 実験を行った人数は 20 人である. Fig. 3 にコインの表がでた確率の最大値, 平均値, 最小値の推移を示す. 今回の実験より, コインの表の出た確率は, 試行回数を増やせば増やすほど理論値の 0.5 に近い値を示すことが確認された. これより, 実験において試行回数を増やすことは, 信頼性を向上させる上で重要であることが分かった.

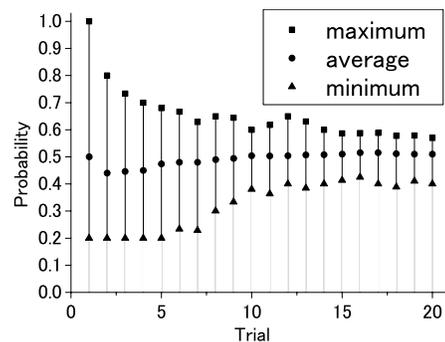


Fig. 3 コインの表がでた確率の推移

5 翌月への課題

- onemax を対象とした自作 sGA の検証とその拡張
- タンパク質の構造解析法と予測法
- ga2k のパラメータの検討