理工学研究報告書執筆と BLAST に関する調査

永松　秀人

1　研究課題

今月の課題は以下の通りである．

- 昨年の研究のまとめ
- BLAST の調査
- テストプログラミング

2　昨年の研究のまとめ

2.1　理工学研究報告書の執筆

昨年，松本と共同研究を行っていた GA を用いた画像認識に関する研究を，「遺伝的アルゴリズムを用いた画像認識による商品管理システムの構築」というタイトルで理工学研究報告書にまとめ，提出した．

論文で主張した点を以下に述べる．

- 画像認識によって商品管理をすることによって，バーコードが不要になり，これまで扱えなかった特殊な商品も扱える．
- 画像処理モジュールの呼び出しには，SOAP を用いて異種言語間通信を実現している．
- 特徴量データを用いたマッチングでは，識別に用いることができるものと，用いることができないものがあり，それらの最適な組み合わせを見つけることが重要である．
- 遺伝的アルゴリズムを用いて，特徴量の使用，不使用を決定する本手法は，高い精度で識別することが可能である．

2.2　計算力学講演会の準備

昨年の研究のうち，私が担当した GA による画像認識に関して，計算力学講演会で発表を行う．そのため，昨年の研究で問題点が指摘されていた，評価関数の再検討を行っている．

3　BLAST

DNA の塩基配列やタンパク質のアミノ酸配列の決定が一般的になった現在では，新たに決定された配列と相同性（homology：ホモロジー）の高い配列をデータベース中で発見することが，進化・系統分類の解析，タンパク質の機能解析などを考える上で，重要な要素となっている．この相同性検索のための強力なツールが，BLAST（Basic Local Alignment Search Tool）である．

3.1　相同性

遺伝子において相同性とは，共通の祖先遺伝子から由来していることを意味している．しかし，実際には相同性検索の結果得られる配列の類似度を相同性と呼ぶことが多い．

2 本の配列の相同性を計算するには，2 つの配列を要素ごとに対応づけて並べる操作（alignment：アラインメント）により行われる．連続した 5 個のアミノ酸のみが完全に一致している場合も「相同」となり，数百個のアミノ酸のうち 20%程度が一致している場合も「相同」と呼ぶことができる．そのため，前者を局所的相同性（local homology），後者を大域的相同性（global homology）と呼び区別される．Fig. 1 にその例を示す．

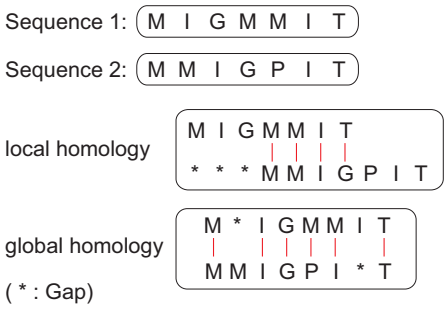

Fig. 1　局所的相同性と大域的相同性

3.2　アラインメント

2 つの配列のアラインメントの結果は，配列要素ごとに定義される類似度のスコア（塩基の一致度，アミノ酸の一致度，あるいは類似度を示す）の和で与えられる．アラインメントでは，配列要素をそのまま対応づけるだけでなく，進化の過程で生じうる配列要素の挿入・欠失を扱うため，ギャップを対応付けることが多い．

Fig. 2　BLAST による検索結果

4　今後の課題

- 昨年の研究成果で計算力学講演会の講演論文執筆
- BLAST のより詳細な調査