

論文内容の要旨

論文題目 A Memetic Algorithm for Reconstructing Gene Regulatory Networks from Expression Profile (Memetic アルゴリズムを用いた遺伝子制御ネットワークの推定)

氏名 ノマン ナシムル

近年の分子生物学の急速な進歩は、多くの新領域、複合領域の応用研究の発展を促した。新しい研究領域の一つに、生物有機体の全てをシステムティックに説明しようとするシステム生物学がある。システム生物学は、新しい生物システムの設計と、現存生物の制御を可能とする事を研究の目的としている。目的を達成するためには、現存生物の構造の特徴と、その振る舞いを明らかにするという基礎研究が必要になる。そのような特徴と振る舞いは、遺伝子の携帯に必要な全ての情報を含むゲノムと相互作用をする、分子の構成要素間の複雑な制御ネットワークによって決定される。この制御ネットワークとネットワークの働きのメカニズムを解明することが、システム生物学における重要な課題である。

現在の所、多くの生体内作用の中心である遺伝子の制御関係は、プロモーター遺伝子によるシス作用制御と、その他の遺伝子の転写因子によるトランス作用制御の組合せであると考えられている。現象論的モデルを用いてシス作用を単純化すれば、遺伝子間の全てのトランス作用は「遺伝子制御ネットワーク」としてモデル化することができる。しかしながら、この生体制御ネットワークは動的かつ高次の非線形的性質を持っており非常に複雑であるため、このネットワークで表現される生体内要素やその依存関係、分子レベルでの相互作用の解明が進んでいなかった。

ところが最近、DNA マイクロアレイやオリゴヌクレオチドチップなどの測定のための新技術が開発され、複数の遺伝子発現の時系列変化や定常状態データを高速かつ同時に測定する事が可能になった。ゲノム規模でトランスクリプトームを観測することにより、ゲノム活性の構造と動的な変化について大域的な視点から考察を行うことができる。観測によって得られた大量の実験データにより、特定の生化学プロセスにおける特定遺伝子の活動の包括的な理解が可能になるのみならず、遺伝子制御ネットワークの構造も理解することが可能になると考えられている。従って、大量の生物学な観測データをもとに計算モデルをたて同定を行うシミュレーションにより、遺伝子ネットワークの転写制御を明らかにする手法の開発が必要となっている。

用いられる計算モデルはネットワークの接続関係や速度定数、生化学濃度のようなパラメータを考慮にいれたものである。GRN のモデルとして利用できるものに生化学系理論モデル化と生化学系の分析のための一般化されたフレームワークに基づく S-system モデルがある。このモデルは生物学的妥当性と数学的柔軟性の双方の利点を持つ。高次元問題を扱うために、元のモデルは結合した形態からネットワークの推定において有効である分離問題に分

解される。この問題に対して進化論的計算手法を用いると、モデルの動的な性質と観測されたシステム応答のノイズの優位水準に優れていることが判明している。そのため進化論的計算手法を **S-system** を用いた **GRN** 推定に適用することが期待されている。

現在研究されている **S-system** を用いた **GRN** 推定での課題は、(i)疎なトポロジーアーキテクチャを検知すること、(ii)有意水準のノイズにより破損している限られた量の遺伝子発現データから **kinetic** パラメータを推定すること、(iii)**GRN** を同定するために大域的最適解を探索する能力のある効率的な最適化方法を考案すること、の三つである。本論文では、これらの課題に対処するために分解 **S-system** モデルを利用した **GRN** 推定のためのミメティック方法を提案している。

最初に、有用であるとされる進化アルゴリズム(EA)である標準的な **Differential Evolution (DE)** を、交叉に基づく局所探索を用いて性能強化を行った。また **DE** に近傍探索を組み込み性能向上を狙った。局所探索のためのヒューリスティクスを組み込むことは大域的最適化のための効果的な進化アルゴリズムにとって非常に有効であることが広く知られており、近傍探索を組み込んだ EA は **Memetic Algorithm (MA)** として有名である。

広範囲のベンチマーク問題で実験を行った結果、提案手法である拡張版 **DE** は従来の **DE** より良いか、同等の性能を持つことが分かった。また、他の有名な進化アルゴリズムとの性能比較も行った。連続ランドスケープ発生器によって生成されるランダム問題で、元のアルゴリズムに対する優越性が確認された。さらに、実問題に対する適合性を、高次非線形、多峰性、だまし要素を含む **S-system** 最適化問題に適用することによって調査した。その結果、実問題の最適化では元の **DE** と比較して優越性があることが示された。

次に、この効率的かつロバストな最適化アルゴリズムを用いて、分離 **S-system** 形態で表される生化学ネットワーク内の転写制御関係を推定するためのミメティックアルゴリズムを設計した。提案アルゴリズムは、ロバストな転写制御関係の同定、正確なキネティックパラメータの推定、ネットワークのアーキテクチャの獲得、そして効率的な計算などの課題の解決を目標として設計した。アルゴリズムの核である、最適化エンジンは **memDE** と呼ばれる前述の拡張した最適化アルゴリズムを用いて実装した。さらに、拡張アルゴリズムでは骨格ネットワークを効率的に獲得するための局所探索を組み込んでいる。局所探索では、より疎なネットワークを獲得するために各世代からの最良個体とランダムに選ばれた個体の周辺の上り探索を行う。また、ネットワーク内の最もロバストな制御関係と正確なキネティックパラメータを同定するために、アルゴリズムを2倍の最適化用に設計した。

ターゲットとするネットワークの最適パラメータの組合せを探索するとき、解候補を評価するための測度が必要となる。最もよく用いられている評価基準は数値計算によって算出される遺伝子発現レベルと観測されたシステムのダイナミクスの差である平均二乗誤差(**MSE**)である。ネットワークの骨格構造を獲得するために、適合度基準に基づく **MSE** のための新しいペナルティ項の提案を行う。本提案では、ネットワークのパラメータの組み合わせ最適化の性能向上のために、伝統的な **MSE** の代わりに、適合度基準に基づく **AIC** を用いる。実際には、適合度評価関数で **AIC** のペナルティ項に **Complexity** 項の追加を行い、疎なネットワークの構造を有するモデルの選択を容易にした。

提案手法の評価として、推定に用いるデータの量、ノイズのレベルなど様々な次元・特徴において、提案手法の適合性を調べる実験を行った。どの推定方法もネットワークトポロジーと制御パラメータを獲得することができたが、その精度は発現データの量とノイズのレベルに制限されてしまう。既存の適合度関数と比較して、提案手法は正しい発現データ量やノイ

ズのレベルに対してロバスト性が高く，どのような状態でもネットワークトポロジーの同定と正確なパラメータを推定するのに適していることが分かった．提案アルゴリズムの異なる要素が，ロバストさ，効率，正確な制御パラメータの推定に必要であることを示すためにアルゴリズムの実験分析を行った．さらに既存の GRN 推定アルゴリズムに比べて，提案した推定アルゴリズムは，計算量について効率的，ノイズに対してロバスト，より少ないサンプルサイズで推定可能であることが分かった．最後に，提案方法を 2 つの実データに適用したところ，主要な制御遺伝子の基本的なネットワークが獲得された．

今回提案した拡張アルゴリズムによる遺伝子ネットワークアーキテクチャとパラメータの推定の性能評価は，数十の遺伝子から構成される中型ネットワークを推定する事で行った．しかし数百，数千の遺伝子を含むネットワークを推定するにはさらなる強化，拡張，改良が必要となると考えられる．本論文では今後の課題についての方向付けを与えて結論を出す．