

超高速ネットワークシミュレータで用いる 線虫の遺伝子調節ネットワークの記述

Description of Worm Gene Regulatory Network for High Speed Network Simulator

鈴木 義実[†] 中山 伸一[†] 伊藤 将弘[‡] 真栄城 哲也[†]
Yoshimi Suzuki Shin-ichi Nakayama Masahiro Ito Tetsuya Maeshiro

我々は、遺伝子制御ネットワークの予測のために高速専用計算機を開発し、運用している [1, 2] (図 1)。この計算機は、遺伝子群、遺伝子間の制御関係、遺伝子およびその他の物質の初期量で表現された遺伝子制御ネットワークをシミュレーションする。高速性が1つの特徴であり、シミュレーションはソフトウェアを介さずに直接ハードウェアで実行されるため、その実行速度は、同じ遺伝子制御ネットワークを最新の計算機上で既存のシミュレーションソフトウェアを用いてシミュレーションした場合の1万倍以上である。この高速性を利用して、遺伝子制御ネットワークの候補を複数生成し、それぞれの妥当性をシミュレーションによって検証する、遺伝子制御ネットワークの予測手法を提案した [2]。

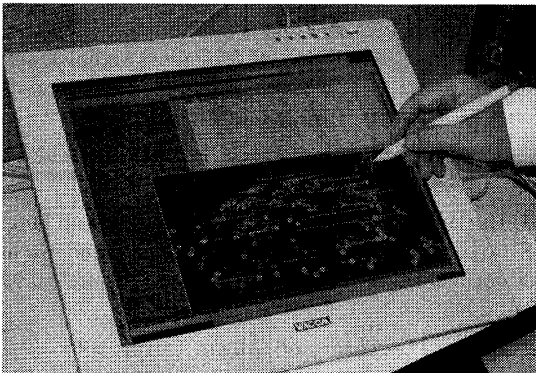


図 1: Starpack のユーザインタフェース。ネットワークの3次元可視化、遺伝子の情報や現時点での量を表示できる。

Starpack は、シミュレーション対象の全遺伝子の量を個別に保持し、遺伝子間の制御関係が電子回路によって実装された複数のプロセッサが、関連する遺伝子の量を並列動作によって変化させることで、遺伝子制御ネットワークをシミュレーションする。プロセッサは単純な回路で構成されることから、大量のプロセッサが実装できるため高いスケーラブル性を持ち、従って大規模な遺伝子ネットワークを高速にシミュレーション可能な計算機が実現できる。

シミュレーションを実行するには、対象である遺伝子調節ネットワークを記述し、その記述を基に Starpack へ入力するデータの生成や編集、そして出力を行う。記述方法としては、

知識表現手法に、Starpack のアーキテクチャおよび実行に由来する表現の特性を付加した手法を用いる。

知識表現の基本的な手法として、意味ネットワークやフレーム、ルールベース等、様々な手法が提案されている。これらの手法に共通するのは、全てがグラフモデルである点である。バイオ情報学で一般的に使用されるオントロジーや、SBML [4] に代表されるマークアップ言語は、これらグラフ理論 [5] またはグラフモデルに基づいている。グラフモデルは、2つの要素間の関係を表すのに適しているが、3つ以上の要素間の関係である多項関係を表すことができない。これは、グラフの定義による制約である。ノードの集合を V とした場合、グラフ G は $G = V \times V$ と定義される。これはノード間の接続関係を表す隣接行列であり、リンクの集合を表すとも捉えることができる。

遺伝子制御ネットワークをネットワークの形で表現する場合、遺伝子をノード、制御関係をリンクと表すのが一般的である。例えば、遺伝子 A が発現し合成された蛋白質が、遺伝子 B の転写因子として機能し、遺伝子 B を発現させる場合、単純な表現では遺伝子 A を表すノードから遺伝子 B を表すノードへ矢印でつなぐ (図 2)。また、制御関係としてアクティベータとリプレッサーに区別も必要である。さらには、転写因子としての影響度や、他の転写因子が同時に存在する時にしか機能しない等の他の因子との関係の記述も必要になる。複数の因子が同時に存在しなければ発現しない場合や、どれか1つでも存在すれば発現する場合、さらには複数の蛋白質が複合体として結合した後、転写因子として働く場合もある [3]。これらは、転写因子を表すノード間の関係であり、制御される遺伝子毎に条件が異なる。

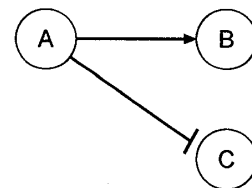


図 2: 遺伝子 A が遺伝子 B と C を制御する例。遺伝子 B に対してはアクティベータ、遺伝子 C に対してはリプレッサーとして機能する。矢印の終端記号によって区別する。

ここで挙げた制御関係を示す関係を表すリンクと、転写因子間の機能的な関係を表すリンクは、性質の異なる。前者は、ノード間の関係という遺伝子制御ネットワークで表す関係である。一方、後者は、遺伝子制御ネットワークで表す関係の関係

[†]筑波大学, Univ. of Tsukuba

[‡]立命館大学, Ritsumeikan Univ.

という2重の関係であり、グラフ理論の枠組みでは表せない関係である。

さらに、遺伝子には配列、イントロンおよびエクソン、合成される蛋白質の機能といった属性情報の記述が必要である。これらの情報は従来のグラフモデルに基づく表現手法で記述可能である。遺伝子と同様に、転写制御関係についても、転写制御の強度等の属性情報を記述する必要がある。しかし、転写制御関係は要素間の関係であり、グラフ理論の枠組みでは記述できない。これは、転写因子間の関係の場合と同様である。これらの記述は例外的な規則を導入することで表面的には対応可能だが、記述を全体的に捉える際に内包する矛盾が表面化し内容の一貫性が保てなくなることで、さらには記述の自由度も制限されるといった問題がある。

既存の表現手法の別の問題として、表現できる関係が2項関係に制限されることである。遺伝子間の制御関係は2つの遺伝子の間だけでなく、3つ以上の関係や自己を制御する一項関係も多く見られる。しかし、このような多項関係は既存の手法では厳密には表現できない。また、一見、表現できているようでも、多項関係と密接に関連する双対性を扱えないため、やはり全体としての記述に矛盾が生じる。

本研究では、従来の表現手法よりも表現能力の高い記述方法[6]を、遺伝子制御ネットワークを表現するために改良して用いる。この表現手法は、ハイパーグラフを拡張したものである。グラフ G は $G = V \times V$ と定義される。ここで、 V はノードの集合である。一方、ハイパーグラフ H は $H = V \times E$ と定義される。なお、 E はリンクの集合である。ノードとリンクの行列であることから、リンクはノードの部分集合を表しているとも見なせる。従って、多項関係が表現できる。このハイパーグラフ表現を、ノードとリンクをどちらもノードで表す2部グラフ表現に変換して、遺伝子制御ネットワークを表現する(図3)。

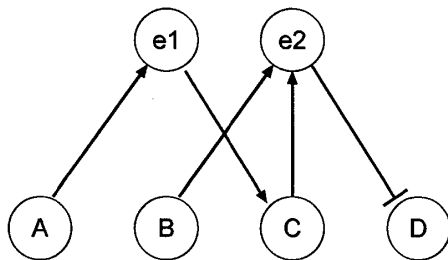


図3: 遺伝子Aが遺伝子Cを制御し、遺伝子BとCが遺伝子Dを制御する例。遺伝子Cに対してはアクティベータ、遺伝子Dに対してはリプレッサーとして機能する。上部のノードe1とe2は制御関係を表す。遺伝子B, C, Dは3項関係にある。

さらに、ハイパーグラフに対する拡張として、関係の属性、関係間の関係や他の要素を表現するために、ノードが概念、関係、属性のどれを表しているかを示す視点情報を各ノードに付加している。このように視点を導入することで、転写制御の強度は制御関係を表すノード(図3のノードe1とe2)に接続されるノードに記述可能であり、制御関係間の関係も同様に記述できる。

本研究の記述手法は、遺伝子制御ネットワークをシミュレートするための専用システム Starpack に用いることが主要な目的であるため、Starpack のアーキテクチャに依存した記述部

分がいくつかある。主要なものとして挙げられるのは、シミュレーション時の初期値の設定であり、これは絶対値、他の遺伝子に対する相対値、計算式を用いた記述がある。もう1つは、ネットワークへの外部入力の設定であり、主に数学関数を使う。

既存の記述言語として、SBML等が挙げられるが、標準的な言語は存在しない。さらに、これらの記述言語は生物に関して広範囲かつ一般化されており、本研究の用途には不要な項目が多く、かつ表現力が不十分であり、対応に要する労力も見合わない。また、我々の記述言語からこれらの記述言語への変換は容易であり、既存の言語による記述が必要になった場合には、視点を固定することによる表現内容の欠落を許容すれば対処可能である。

記述方法の有効性を検証するため、線虫 *C.elegans* の細胞分裂に関わる既知の遺伝子制御ネットワーク [7, 8] を記述したところ、問題は見当たらなかった。現在は、ここで報告した記述方法と超高速ネットワークシミュレータ Starpack を用いて、遺伝子制御ネットワークの推測を行っている。

なお、本研究の一部は科学研究費補助金 21510214 号による。

参考文献

- [1] H. Hemmi, T. Maeshiro, K. Shimohara, "New computing system architecture for scientific simulations — Non CPU-oriented methodology", *Frontiers of Computational Science*, Springer, 2007.
- [2] T. Maeshiro, H. Hemmi, K. Shimohara, "Ultra-fast genome wide simulation of biological signal transduction networks — Starpack", *Frontiers of Computational Science*, 177–180, Springer, 2007.
- [3] M. Ptashne, A. Gann, *Genes & Signals*, Cold Spring Harbor Laboratory Pr, 2001.
- [4] M. Hucka et. al, "The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models", *Bioinformatics*, 9:524–531, 2003.
- [5] C. Berge, *The Theory of Graphs*, Dover, 2001.
- [6] T. Maeshiro, T. Morimoto, K. Shimohara, Y. Fujiwara: "Knowledge infrastructure for bioinformatics based on Homogenized Bipartite Model", *Intl. For. Info. Docum.*, 25:1–18, 2000.
- [7] Eisenmann, D. M., Wnt signaling (June 25, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.7.1, <http://www.wormbook.org>
- [8] <http://www.wormbase.org>