

課題名 (タイトル) :

大規模遺伝子ネットワーク推定プログラムの研究開発

利用者氏名 : ○宮野 悟, 玉田 嘉紀

所属 : 社会知創成事業 次世代計算科学研究開発プログラム

次世代生命体統合シミュレーション研究推進グループ データ解析融合研究開発チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

筆者らは、スーパーコンピュータ「京」のためのソフトウェア開発プロジェクト「次世代生命体統合シミュレーションソフトウェアの研究開発」において、理化学研究所次世代計算科学研究開発プログラム・データ解析融合チームの一員として、プロジェクトの研究開発課題の一つとして掲げている大規模遺伝子ネットワーク推定プログラムの研究開発を行っている。筆者らが開発研究している大規模遺伝子ネットワーク推定プログラム SiGN は、ノンパラメトリック回帰によるベイジアンネットワーク、状態空間モデル、グラフィカルガウシアンモデル、ベクトル自己回帰モデルを遺伝子ネットワークモデルとして利用し、遺伝子発現データなどから遺伝子発現の依存関係を表す遺伝子ネットワークを推定するためのものである。そのうちベイジアンネットワークを利用したものを SiGN-BN、状態空間モデルは SiGN-SSM、グラフィカルガウシアンモデル及びベクトル自己回帰モデルを利用したものを SiGN-L1 と呼んでいる。それぞれのモデルのパラメータ推定及びネットワークの構造推定は非常に計算に時間がかかるため、スーパーコンピュータを利用した大規模計算が欠かせない。本研究では、これらの遺伝子ネットワーク推定法に対して「京」を利活用する新規アルゴリズム・ソフトウェアを開発し、既存アルゴリズムが適用不可能であった大規模遺伝子ネットワーク推定を可能にすることが研究課題のひとつである。高精度・大規模な遺伝子ネットワークの推定が可能になることにより細胞内での遺伝子の発現の依存関係を予測することが可能になり、薬剤作用機序解明や新規薬剤標的遺伝子の同定などが可能になることが期待される。

「京」は 2011 年度より試験利用がはじまり 2012 年度も引き続き「京」実機を用いてアルゴリズム・ソフトウェアの開発が可能であったため開発に必要な

計算の大部分は「京」を用いた。しかし利用時間が限られている等制約があったため RICC を補助的に利用した。

2. 具体的な利用内容、計算方法

まず、ベイジアンネットワークを用いる SiBN-BN のうち、ヒト全ゲノムに適用可能な Neighbor Node Sampling & Repeat (NNSR) アルゴリズムの動作確認を行った。NNSR アルゴリズムは NNS 法により抽出した遺伝子セットに対し並列にサブネットワークを推定することにより大規模な遺伝子ネットワーク推定を可能にしたアルゴリズムである (Tamada et al., 2011)。「京」対応のため富士通製乱数ライブラリなどを使用するようにコードを変更したため、過去の実験結果との比較が必要となったため RICC を用いて実施した。

次に L1 正則化法を用いた遺伝子ネットワーク推定を行う SiGN-L1 の実行テストを、RICC を用いて実施した。SiGN-L1 は L1 正則化法を用いるいくつかのアルゴリズム・モデルに対応するがここでは構造方程式モデルに基づき個人ネットワークのプロファイリングが可能な Network Profiler (Shimamura et al., 2011) と呼ばれるアルゴリズムの並列版の実装を行い、その動作確認を行った。

3. 結果

SiGN-BN NNSR アルゴリズムの動作確認の結果を図 1 に示す。過去に Tamada et al. (2011) で用いたシミュレーションデータを用い 256 並列 (コア) で計算を行った。その結果両者に差がないことが確認された。

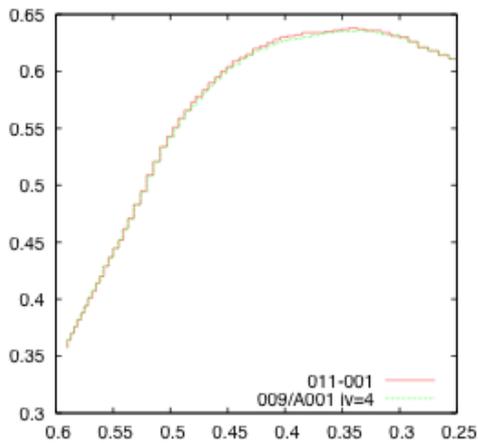


図 1. 閾値を動かした際の sensitivity/specificity の値. 横軸が sensitivity, 縦軸が specificity を表す. 緑の破線が過去の計算結果. 赤の実線が変更後のコードの RICC での実行結果.

SiGN-L1 については開発初期段階での動作確認であったが新たに行った MPI 版の実装において 256 並列での実行を確認している. SiGN-L1 に関しては, RICC の利用はこの初期開発段階で利用にとどまり, 大部分は「京」の実機を利用した. したがって, ここに特筆すべきデータなどは特にない.

#### 4. まとめ

RICC を用いて SiGN-BN NNSR アルゴリズムと SiGN-L1 Network Profiler の動作確認を行った. 両者は「京」で開発を継続し, 現在, 筆者らのウェブページ (<http://sign.hgc.jp/>) 上で公開している.

#### 5. 今後の計画・展望

本年度がプロジェクトの最終年度であるため RICC を用いた研究としては本年度で終了である. SiGN は「京」利用プロジェクトである HPCI 戦略プログラム分野 1「予測する生命科学・医療および創薬基盤」や新学術領域研究「システムの統合理解に基づくがんの先端的診断, 治療, 予防法の開発」で利用予定であるためアルゴリズム・ソフトウェアの研究開発自体はそらのプロジェクトにて継続予定である.

#### 6. 参考文献

Tamada, Y., Imoto, S., Araki, H., Nagasaki, M., Print, C., Charnock-Jones, D.S., and Miyano, S. (2011). Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers, *IEEE/ACM*

*Transactions on Computational Biology and Bioinformatics*, **8** (3), 683-697.

Shimamura, T., Imoto, S., Shimada, Y., Hosono, Y., Niida, A., Nagasaki, M., Yamaguchi, R., Takahashi, T., Miyano, S. (2011). A novel network profiling analysis reveals system changes in epithelial-mesenchymal transition, *PLoS ONE*, **6** (6), e20804.

平成 24 年度 RICC 利用研究成果リスト

【国際会議、学会などでの口頭発表】

玉田 嘉紀, 遺伝子発現データからの大規模遺伝子ネットワーク推定, 定量オミックスワークショップ @ 大阪大学 バイオ関連多目的研究施設 (大阪府吹田市) (Jan. 22, 2013). 口頭発表.

玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 斉藤 あゆむ, 長崎 正朗, 井元 清哉, 宮野 悟, SiGN-BN: ページアンネットワークによる大規模遺伝子ネットワーク推定プログラム, ISLiM ソフトウェア研究開発報告会 @ 東京大学武田ホール (東京都文京区) (Jan. 10-11, 2013). ポスター発表.

Tamada, Y., SiGN: Large-scale Gene Network Estimation Software for K Computer and HGC Supercomputer System, 生命医薬情報学連合大会 @ タワーホール船堀 (東京都江戸川区) (Oct. 14-17, 2012). ポスター発表.

Tamada, Y., Large-scale Gene Network Estimation with K Computer, 第 50 回 日本生物物理学会年会 @ 名古屋大学 東山キャンパス (愛知県名古屋市) (Sep. 23, 2012). 口頭発表.

玉田 嘉紀, 島村 徹平, 山口 類, 新井田 厚司, 斉藤 あゆむ, 井元 清哉, 長崎 正朗, 宮野 悟, スーパーコンピュータを用いた遺伝子ネットワーク推定による癌予後予測および薬剤応答パスイメージ予測, 第 12 回 東京大学 生命科学シンポジウム @ 東京大学 山上会館 (東京都文京区) (Jun. 30, 2012). ポスター発表.