

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 1 日現在

機関番号：82401

研究種目：若手研究(B)

研究期間：2012～2015

課題番号：24700295

研究課題名(和文) ワークスティーリングを用いた分散並列環境における遺伝子制御ネットワーク推定手法

研究課題名(英文) A study of gene regulatory networks analysis on distributed and parallel computers using work steal method

研究代表者

木戸 善之(Kido, Yoshiyuki)

国立研究開発法人理化学研究所・生命システム研究センター・客員研究員

研究者番号：70506310

交付決定額(研究期間全体)：(直接経費) 3,000,000円

研究成果の概要(和文)：本研究では、大規模な遺伝子制御ネットワークを解析するアプリケーションのための、並列計算機上でのワークスティーリング手法のジョブを可能にするジョブスケジューラの研究、開発を行った。大規模な遺伝子制御ネットワークとは人間の全遺伝子、約20000遺伝子の相関を解析すること相互作用や影響などを知ることができる。本研究を通じ、ネットワーク推定として余計なジョブについては枝刈りをするための判断機構、アルゴリズムを動かすためのジョブスケジューラフレームワークを開発した。また大規模な遺伝子解析結果を表示するための可視化アプリケーションについても研究を進めることができた。

研究成果の概要(英文)：In this research, we have developed and implemented a job scheduler framework with work steal mechanism for large scale gene regulatory networks analysis. The analysis applications usually execute on a distributed and parallel computational environment. Because a whole of human genes is a approximate 20,000, therefore a human gene regulatory networks analysis needs large scale computational environment. So, our products aims to reduce a calculation time of gene regulatory networks analysis using work steal mechanism. A work steal mechanism is able to cut off a useless network estimation. We also have modified and improved a network visualization application for large scale gene regulatory networks on a high resolution visualization system (i.e. tiled display wall).

研究分野：分散コンピューティング

キーワード：分散コンピューティング 遺伝子発現ネットワーク推定 ジョブスケジューラ 可視化

1. 研究開始当初の背景

次世代シーケンサーの発展に伴い、遺伝子解析の需要は、研究のみならず医療の分野でも広がりつつある。シーケンサーは遺伝子の発現情報をマイクロアレイに記し、時系列で遺伝子発現の変化を知ることができる。遺伝子解析は、このマイクロアレイから得た情報を元に遺伝子間の発現の相互作用を解析することが生命科学において非常に重要である。なぜなら近年、相互作用や相関関係によって幹細胞からの細胞分化のメカニズムに影響することが知られているからである。しかし人間の遺伝子は約 20,000 遺伝子あると言われており、それらの全て網羅的に相関関係を解析することは困難である。

2. 研究の目的

本研究の目的は、大規模遺伝子制御ネットワーク推定アプリケーションの支援にある。大規模な遺伝子制御ネットワークの解析は、大規模な並列計算機環境を用いても困難である。そのため、計算量の軽減を図る工夫が必要となる。そこで本研究では以下の項目を研究開発する。

- A. スケジューラフレームワークの開発
- B. ワークスティーリングによる遺伝子制御ネットワーク推定の開発

また本研究では当初の予定になかった遺伝子ネットワーク推定における可視化も取り組むことができた。これは遺伝子ネットワーク推定の複雑な解析結果を高解像度で可視化することで、生命科学への理解を深める上で非常に重要と言える。

3. 研究の方法

まずワークスティーリングを実現するためのジョブスケジューラフレームワークの開発に取り組んだ。ジョブスケジューラフレームワークとは、スケジューリングアルゴリズムに対するインターフェースを定義することで、ポータビリティを向上させることに他ならない。つまりスケジューリング機構とスケジューリングアルゴリズムを分離することでフレームワークを実現できる。ジョブスケジューラには並列分散環境つまりクラスタ計算機にて広く適用されている Sun Grid Engine をベースに開発に取り組んだ。アルゴリズム部分を分離することで、ワークスティーリング手法におけるアプリケーションからのフィードバックを受け付けることができる。ワークスティーリング手法の主たる目的は、ネットワーク推定における相関が得られない遺伝子間の計算を打ち切り、ネットワーク全体の推定のスループットを向上させることにある。

フレームワークの次に、ワークスティーリング手法を取り入れた遺伝子制御ネットワーク推定アプリケーションに改良に取り組んだ。遺伝子制御ネットワーク推定のアプリケーションは SiGN-BN[1]をベースとした。

SiGN-BN は京コンピュータで動作させるために開発されたアプリケーションであるが、汎用的に開発されており、一般的なインテル CPU でのクラスタ計算機上でも動作する。SiGN-BN での遺伝子制御ネットワーク推定アルゴリズムは、Bootstrap 法を用いることで統計的な蓋然性を算出している。Bootstrap 法では、遺伝子発現データにランダムサンプリングを行い、一部のデータを取得し、次に一部のデータを対象とした遺伝子制御ネットワーク推定による部分グラフの生成を行っている。これらの手順を繰り返すことで部分グラフを大量に作成することで、最終的に部分グラフを結合し、大規模な遺伝子制御ネットワークの推定結果を出力する。しかし再標本による遺伝子の重複などがあり、これらを排除することで計算スループットの向上を目指した。

4. 研究成果

本研究では、当初の目的どおり、ジョブスケジューラフレームワークの開発を行い、またワークスティーリング手法を用いた遺伝子制御ネットワーク推定アプリケーションの改良を行った。図1は、ワークスティーリング手法を用いた遺伝子制御ネットワーク推定アプリケーションの実行時の挙動である。マスターは並列分散環境におけるマスターノードであり、スレーブ1、スレーブ2はそれぞれワーカーノード上の挙動である。「loop」で囲われた箇所は主に SiGN-BN の挙動であり、そこ計算結果の書き込みや、検索処理を NFS サーバ上の共有ディスクに行うことで、既に行った推定計算を打ち切る機構を組み込むことができる。

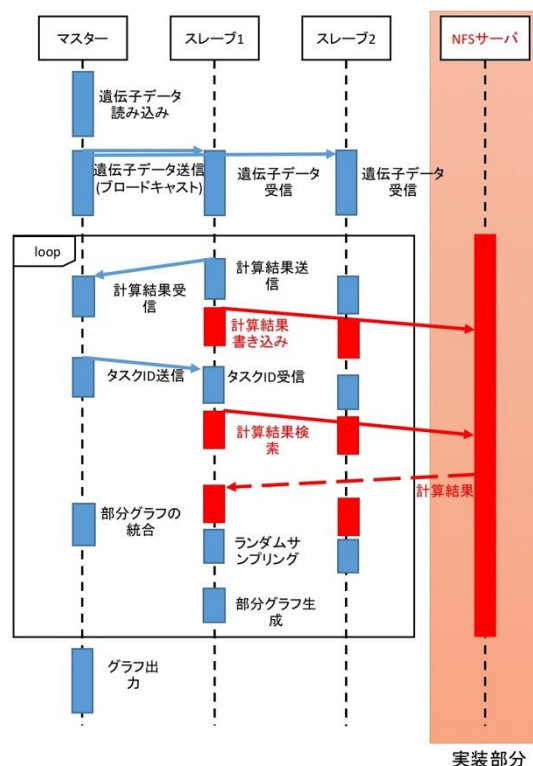


図1. シーケンス図

また当初の予定にはなかったネットワーク推定結果を可視化するためのアプリケーションの改良も行った。図2は可視化アプリケーションを高解像度可視化装置で表示した例である。大規模遺伝子制御ネットワーク推定の結果は、約 20,000 ノードの遺伝子が互いの相関をグラフにおけるエッジで表示することとなり、通常のデスクトップディスプレイでは網羅的に表示することが困難である。もしくは表示できてそれらを統合的に理解することは困難である。そのため複数のディスプレイを組み合わせることで仮想的な1つの高解像度ディスプレイを構成する大規模なタイルドディスプレイが必要となる。本研究においてネットワーク推定可視化アプリケーションである Gephi[2]をタイルドディスプレイへの表示対応することができた。



図2 . 可視化装置への表示

#### < 引用文献 >

- [1]. Y. Tamada, T. Shimamura, R. Yamaguchi, S. Imoto, M. Nagasaki, S. Miyano, "SiGN: Large-scale gene network estimation environment for high performance computing," *Genome Informatics*, Vol.25, no.1, pp.40-52, 2011.
- [2]. M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media (ICWSM09)*, pp.361-362, 2009.

#### 5 . 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

#### [ 雑誌論文 ] (計 3 件)

Ryoichi Jingai, Yoshiyuki Kido, Susumu Date, Shinji Shimojo, A High Resolution Graph Viewer for Multi-monitor Visualization Environment, *The Review of Socionetwork Strategies*, Volume 9,

Issue 1, pp.19-27, 2015 年 06 月 (査読有)

DOI: 10.1007/s12626-015-0052-5

Yasuhiro Watashiba, Susumu Date, Hirotake Abe, Yoshiyuki Kido, Kohei Ichikawa, Hiroaki Yamanaka, Eiji Kawai, Shinji Shimojo, Haruo Takemura, Performance Characteristics of an SDN-enhanced Job Management System for Cluster Systems with Fat-tree Interconnect, *Proceedings of 6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom2014)*, pp.781-786, 2014 年 12 月 (査読有)

DOI: 10.1109/CloudCom.2014.82

渡場康弘, 木戸善之, 伊達進, 阿部洋丈, 市川昊平, 山中広明, 河合栄治, 竹村治雄, 計算資源とネットワーク資源を考慮した割当ポリシーを配備可能とするジョブ管理フレームワーク, *電子情報通信学会論文誌*, Vol.J97-D, No.6, pp.1082-1093, 2014 年 06 月 (査読有)

[http://search.ieice.org/bin/summary.php?id=j97-d\\_6\\_1082](http://search.ieice.org/bin/summary.php?id=j97-d_6_1082)

#### [ 学会発表 ] (計 7 件)

Arata Endo, Yoshiyuki Kido, Susumu Date, Yasuhiro Watashiba, Kiyoshi Kiyokawa, Haruo Takemura, Shinji Shimojo, Improvement of Scalability in Sharing Visualization Contents for Heterogeneous Display Environments, *International Symposium on Grids & Clouds 2016 (ISGC2016)*, Academia Sinica, Taipei, Taiwan, 2016 年 03 月 17 日

Yoshiyuki Asai, Yoshiyuki Kido, Kyota Kamiyoshi, Taishin Nomura, Hiroaki Wagatsuma, An Ontology-Based Model Search Extended to Deal with Accessibility Issues of Neuroinformatics Metadata with Multidirectional Attributes, *INCF Japan Node International Workshop: Advances in Neuroinformatics 2015*, ENEOS Hall, Univ. of Tokyo, Meguro-ku, Tokyo, 2015 年 11 月 26-27 日

木戸善之, 生命科学シミュレーションと大規模可視化, 日本化学会情報化学部会 第二回若手の会, 化学会館, 東京都千代田区, 2014 年 11 月 29 日

Yasuhiro Watashiba, Yoshiyuki Kido, Susumu Date, Hirotake Abe, Kohei

Ichikawa, Hiroaki Yamanaka, Eiji Kawai, Haruo Takemura, Prototyping and Evaluation of a Network-aware Job Management System on a Cluster System Leveraging OpenFlow, 19th IEEE International Conference on Networks (ICON), pp.1-6, Orchard Hotel Singapore, Singapore, 2013年12月12日

Tomoyoshi Nakayama, Yoshiyuki Kido, Hiromi Daiyasu, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, Estimation of Dynamic Gene Regulatory Networks for Cell Differentiation by Splitting Time Course Data, Joint Conference on Informatics in Biology, Medicine and Pharmacology, Tower Hall Funabori, Edogawa-ku, Tokyo, 2012年10月15-17日

木戸善之, 「京」を利用したインシリコ創薬基盤を目指して, 第35回情報化学討論会, 2D1s-2D1s, 広島大学, 広島県東広島市, 2012年10月5日

木戸善之, 生命科学における「京」の必要性 - 「京」を用いた大規模遺伝子制御ネットワークの推定 -, 第66回バイオメクフォーラム21研究会, 大阪大学, 大阪府豊中市, 2012年09月29日

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

なし

## 6. 研究組織

### (1) 研究代表者

木戸 善之 (KIDO, Yoshiyuki)

国立研究開発法人理化学研究所・生命システム研究センター・客員研究員

研究者番号: 70506310