

「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進
のための次世代アプリケーション技術の創出・高度化」

2018 年度
実績報告書

2015 年度採択研究代表者

大浪 修一

理化学研究所生命機能科学研究センター
チームリーダー

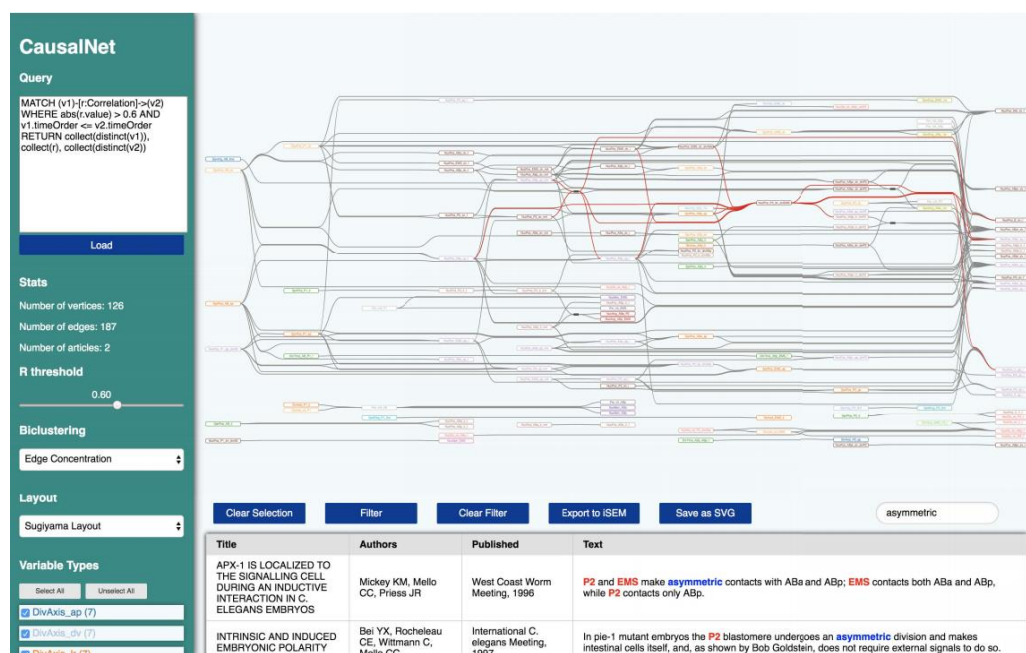
データ駆動型解析による多細胞生物の発生メカニズムの解明

§ 1. 研究成果の概要

本研究は、因果推論を用いたビッグデータの解析(川上グループ)とビッグデータ解析における可視化技術開発(小山田グループ)にそれぞれ高い実績を持つ2つの研究グループと連携し、大浪グループがこれまで進めてきた、生命動態の時空間定量計測データを中核データとして使用して多細胞生物の発生のメカニズムを解明するデータ駆動型の研究手法の開発を強力に推進することを目的とする。第4年次である平成30年度は、前年に引き続き、3グループが共同して、計算表現型解析、クラスタ解析、および表現型特徴の因果推論の手法の改良と、これらの解析結果からの知識発見を促進する可視化手法の開発を行った。加えて、大浪グループは発生のデータ駆動解析に適用する新たな大規模データの取得を、小山田グループは論文情報から因果関係を抽出するプログラムの開発を前年に引き続き行った。また、大浪グループは先端バイオイメージデータの世界規模での共有化に向けて、欧州の IDR との連携を開始した。

具体的には、大浪グループは、GO enrichment 解析の方法を改良し、前年度に遺伝子ノックアウト線虫胚(1142胚)にクラスタ解析を適用して分類した23クラスタ全てに対し、各クラスタに含まれる遺伝子の機能を推定した。また、ゲノム編集を用いた蛍光タンパク質標識株の作成を昨年度より継続して行い、計91遺伝子に対する蛍光標識株を樹立した。また、昨年度取得した線虫胚の1細胞期から90細胞期までの細胞核および細胞膜の動態の計測データ(50個体分)を対象に細胞動態の定量解析を行い、線虫胚の初期の50分裂の76%(38/50)が非対称分裂であることを発見した。更に、生命動態データとバイオイメージデータの統合データベースSSBDを更新し、103セットの生命動態データと453セットの画像データを新たに公開した。世界最大規模のバイオイメージデータベースである欧州のIDRと、先端バイオイメージデータの世界規模での共有化に向けて連携を開始した。川上グループは、これまでに抽出された特徴量間の因果推論手法の問題点を検討し

LiNGAMをはじめとした統計的因果推論手法についての検討と実装を行った。大浪グループから線虫胚の細胞核の4次元分裂動態の計測データの提供を受け、LiNGAMによる統計因果推論およびバリデーションを行った。また、多次元の特徴量データを、新規並べ替えアルゴリズムにより、低次元で表現する手法を開発し異なる次元のデータの連結可能性を検討した。小山田グループは大規模ネットワークに対して対話的にバイクラスタ分析を適用することで抽出される潜在変数の候補を、論文情報データベースや生命科学者の知見に基づいて効果的に絞り込み、統計的解析を支援する視覚的分析環境の構築し、使用事例を通じて有用性を示した。また、線虫に関する論文情報データベースから取得した約2万件のアブストラクトを対象に、因果関係記述候補の抽出を行った。その際に大量のテキストの中から効率よく候補を絞り込むため、線虫の細胞名に合致する記述かを判定する機械学習モデルの改良も行った。また、卵母細胞内のカルシウム濃度に関する時系列データに対して、特異的現象の発見を促す視覚的分析環境の開発と事例研究を行い、その成果を国際会議および学術論文にて公表した。



バイクラスタ分析結果と論文データ情報を統合した視覚的分析システムによる線虫の表現型特徴ネットワーク探索

【代表的な原著論文】

Yosuke Onoue, Koji Kyoda, Miki Kioka, Kazutaka Baba, Shuichi Onami, and Koji Koyamada, "Development of an integrated visualization system for phenotype character networks", 2018 IEEE Pacific Visualization Symposium (PacificVis), Kobe, pp.21-25, 2018

Momoko Imakubo, Jun Takayama and Shuichi Onami, "Improvement and evaluation of a

mathematical model for fertilization calcium waves in *Caenorhabditis elegans*”, IPSJ Transactions on Bioinformatics, vol. 11, pp.24-30, 2018

Kozen Umezawa, Hiroaki Natsukawa, Yosuke Onoue and Koji Koyamada, “A Visual Analytics System to Support the Formation of a Hypothesis from Calcium Waves Data”, Visual Informatics, vol.2, No. 1, pp.2-13, 2018

§ 2. 研究実施体制

(1) 研究代表者グループ(研究機関別)

① 研究代表者: 大浪 修一 (理化学研究所生命機能科学研究センター チームリーダー)

② 研究項目

- ・ 生命動態データとオミクス計測データを統合して多細胞生物の発生メカニズムを解明するデータ駆動型の研究手法の開発と実証
 - 開発済みの手法の改良
 - 最新データ及び改良手法を用いた発生メカニズムの推定
 - 推定した発生メカニズムの検証
 - 発生メカニズムを解明するデータ駆動型の新規手法の開発

(2) 共同研究グループ(1)

① 主たる共同研究者: 川上 浩司 (京都大学大学院医学研究科 教授)

② 研究項目

- ・ 計算表現型解析における新規統計手法の開発と実装
 - 表現型間の因果推論のための統計手法の開発
 - 因果推論により推定された発生モデルの適合性指標の開発
 - 発生メカニズムを解明する新規統計手法の開発
 - 新規統計手法により推定されたモデルの適合性指標の開発

(2) 共同研究グループ(2)

① 主たる共同研究者: 小山田 耕二 (京都大学学術情報メディアセンター 教授)

② 研究項目

- ・ 基礎生命科学の発見を促進するビッグデータ可視化技術の開発
 - 生命科学者のニーズ分析
 - 大規模グラフの粗視化技術の開発
 - 潜在変数探索支援技術の開発
 - 時系列複数密度データからの因果発見支援技術の開発