

仕 様 書

1. 件名:bulk RNA-Seq データからの遺伝子制御関係推定技術の実装

2. 研究の概要

国立研究開発法人産業技術総合研究所（以下、「産総研」という。）生命工学領域生物プロセス研究部門では、国立研究開発法人新エネルギー・産業技術総合開発機構委託事業「カーボンリサイクル実現を加速するバイオ由来製品生産技術の開発」に関する研究を行っている。従来から、スモールスケールの発酵槽培養で取得したデータに基づいて開発されたスマートセルが、ラージスケールの発酵槽培養で想定していたパフォーマンスを発揮しない場合のあることが課題となっていた。その原因の一つとして、培養の過程で複数の細胞状態が混在することがあり、スケールが大きくなるほど不均一度が増すため、スモールスケールのデータに基づいて開発されたスマートセルが、ラージスケールにおいて想定した細胞状態になっていない（あるいはマジョリティを占めていない）ことがあると考えられる。高いパフォーマンスを示す細胞状態を誘導するためには、まず、細胞状態を推定する技術が必要であり、産総研ではそのための解析技術を開発してきており、解析パイプラインとして実装済みである。その解析技術では細胞状態にかかわる複数の遺伝子が同定される。本実装では、抽出された遺伝子から遺伝子制御関係を推定する技術に関する調査と、選定されたアルゴリズムの実装と既存の解析パイプラインへの統合を行う。

3. 作業の概要

本実装では、細胞状態が混在している系から取得した RNA-Seq データを bulk RNA-Seq データと呼ぶことにする。これまでの開発で、次元削減・クラスタリング・統計検定により細胞状態に影響を及ぼす遺伝子を同定する技術を開発してきた。それらの遺伝子間で直接的・間接的な制御関係を推定することができれば、それらの制御カスケード上での位置に関する情報が得られることになる。本件では、上述のような制御関係を因果関係と見なし、因果探索手法の適用を試みる。因果探索の様々な手法を実装した既存のオープンソースである LiNGAM パッケージ (<https://github.com/cdt15/lingam>) を調査し、パッケージ中で適用可能な探索手法・評価手法をこれまでに開発してきた手法（既存の解析パイプライン）に統合する。

4. 作業項目

- 4-1: LiNGAM に関する調査
- 4-2: LiNGAM による遺伝子制御関係推定コードの実装
- 4-3: 既存の解析パイプラインへの統合

5. 作業項目別仕様

5-1: LiNGAM に関する調査

5-1-1: パッケージについての調査

LiNGAM (<https://github.com/cdt15/lingam>) にどのような機能が含まれているかを調査し、それぞれの概要を簡潔に整理する。

5-1-2: 要素間の関係性についての調査

要素間の関係性には線形で表現できる場合、非線形でなければ表現できない場合があると考えられる。それぞれの場合に、どのような定式化がされているかを調査しとりまとめる。

5-1-3: 時系列データの取り扱いについての調査

LiNGAM に実装されている時系列データに対しての因果構造推定アルゴリズムを調査しとりまとめる。

5-1-4: 因果探索結果に対する評価手法についての調査

LiNGAM に実装されている評価指標やブートストラップ法について調査しとりまとめる。

5-2: LiNGAM による遺伝子制御関係推定コードの実装

5-2-1: 産総研が貸与する「TPM 正規化した遺伝子発現データおよび生産性等の数値データ」に含まれる 10~50 個程度の遺伝子を対象に、遺伝子発現データから遺伝子制御関係を推定する機能を、LiNGAM を用いて実装する。5-1-2~5-1-4 の機能を組み込むこととする。

5-3: 既存の解析パイプラインへの統合

5-3-1: 産総研にてこれまで開発してきた細胞状態推定技術（次元削減、クラスタリング、統計検定が実装されたパイプライン）に 5-2-1 の機能を統合する。

6. 特記事項

6-1: (1) 週 1 回程度、調達請求者と作業の進捗報告を行うこと。

(2) Linux 系 OS 環境の上で解析作業を行う。(本環境は請負者が用意する。) 使用言語は Python、R とする。(それ以外の言語を使用する場合は調達請求者に了承を得ることとする。) 解析作業の詳細については、調達請求者と協議すること。

(3) サプライチェーン・リスクに対応するため、「IT 調達に係る国等の物品等又は役務の調達方針及び調達手続きに関する申合せ」(平成 30 年 12 月 10 日関係省庁申合せ) に基づき対応を求めることがあるので応じること。

7. 請負者に必要な要件等

7-1: 本作業の従事者は、次の各項に関する知識や業務実績を有していること。

(1) 遺伝子制御関係、推定等に関して十分な知識を有し、因果推定の

実装が可能であること。

(2) バイオインフォマティクスの専門性を有し、バイオインフォマティクス系研究者と共同で、関連するアルゴリズムの開発およびバイオデータ解析の実績を有すること。

8. 貸与品

8-1: TPM 正規化した遺伝子発現データおよび生産性等の数値データ

貸与品は産総研から box 経由で貸与する。本作業終了後、貸与品は直ちに削除し、文面で削除した旨を報告すること。

9. 納入物品

9-1: 以下に示す作業内容報告書や実装プログラム（パイプライン）を個別のファイルとして電子媒体に記録し納入すること。なお、(2) の実装プログラム（パイプライン）は、一連の作業内容を1つにまとめて納入しても構わないものとする。

(1) 作業内容報告書

5. に示す作業項目で実施した作業内容を記述した報告書

(2) 実装プログラム（パイプライン）

5. に示す作業項目で作成したプログラム（ソースコード）

※USB メモリ等の外部電磁的記録媒体は用いないこと。

10. 納入場所

10-1: 国立研究開発法人産業技術総合研究所

生命工学領域生物プロセス研究部門 北海道センターD1 棟 325 号室
〒062-8517 北海道札幌市豊平区月寒東 2 条 17-2-1

11. 納入の完了

11-1: 作業完了の後、「9. 納入物品」に記載されている納入物品が過不足なく納入されていること、また、その内容・品質を確認し、本仕様書を満たしていることの確認を行い、納入の完了とする。品質に関しては、作業内容報告書に、具体的な解析プロトコル・パラメータと、解析結果の評価が定量的に示されていることで確認することとする。

12. 納入期限

12-1: 2025年2月27日

13. 付帯事項

13-1: 請負者は、成果に係る著作物の著作権を産総研に代金支払の対価として譲渡するものとし、著作者人格権を行使しないものとする。

13-2: 請負者は、契約条項に定める検査に合格後、直ちに別紙様式による著

作者財産権譲渡証書及び著作者人格権不行使証書を産総研に提出しなければならない。

13-3: 請負者は、産総研に対し、納品した成果品が第三者の知的財産権を侵害しないことを保証するものとする。なお、納品した成果品について、第三者の権利侵害の問題が生じ、その結果、産総研又は第三者に費用や損害が生じた場合は、請負者は、その責任と負担においてこれを処理するものとする。

13-4: 納入された成果に係る著作物及び納入後の電子媒体の動作不良を含む発注側の責めによらない納入の完了後1年以内の不具合等については、その交換・補修・訂正等を責任をもって無償で速やかに行うこと。

13-5: 本仕様書の技術的内容及び知り得た情報については、守秘義務を負うものとする。

13-6: 本仕様書の技術的内容に関する質問等については、調達請求者と協議すること。また、本仕様書に定めのない事項及び疑義が生じた場合は、調達担当者と協議のうえ決定する。

別紙様式

年 月 日

著 作 者 財 産 権 譲 渡 証 書

国立研究開発法人産業技術総合研究所 殿

請 負 者
住 所
会 社 名
代表者氏名

印

作業請負契約 (年 月 日 契約)
件 名

上記契約により作成したプログラムの所有権及び著作権（著作権法第 27 条及び第 28 条に規定する権利を含む）は、国立研究開発法人産業技術総合研究所に譲渡したことに相違ありません。ただし、自己所有していた権利は除くものとします。

別紙様式

年 月 日

著作者人格権不行使証書

国立研究開発法人産業技術総合研究所 殿

請 負 者
住 所
会 社 名
代表者氏名

印

作業請負契約 (年 月 日 契約)
件 名

上記契約により作成したプログラムの著作権（著作権法第 27 条及び第 28 条に規定する権利を含む）に係わる著作者人格権を行使しないことを約束します。

なお、著作者人格権を行使しようとする場合は、国立研究開発法人産業技術総合研究所の承認を得るものとします。