

審査の結果の要旨

氏名 趙世涛

本論文は、トランスクリプトーム発現解析分野において、サンプル間の全体的な類似度を任意の群間で客観的に示すための指標として、平均シルエットスコア (AS) の利用を提案したものである。RNA-seq とマイクロアレイデータを用いて、発現変動遺伝子の割合 (PDEG) と AS 値との間に正の相関があることを示し、その具体的な活用法と今後の展望を述べている。本論文は、4つの章から構成されている。

第一章では、本研究の背景や目的、そして意義を述べている。トランスクリプトーム解析の歴史とともに、AS の計算や発現変動解析の入力である遺伝子発現行列を得るまでの一般的な手順を概説している。また、本研究の着想に至った発現変動解析の結果と階層的サンプル間クラスタリング (HSC) の結果の関連性に関する先行研究を述べるとともに、最適なクラスター数に関する情報を得る目的で利用されてきたシルエットスコアについて述べている。

第二章では、データ解析環境や解析データ、そして方法について述べている。解析はフリーソフトウェア R 上で行われ、HSC、発現変動解析、RNA-seq シミュレーションカウントデータの生成には TCC 発現変動解析パッケージを用いている。AS を従来のクラスター間ではなく、任意の群間で計算する手順について述べている。また、用いたリアルデータ (RNA-seq とマイクロアレイ) についての詳細を述べている。

第三章では、結果と考察を述べている。3.1 節は、RNA-seq カウントデータの 2 群間比較結果について述べている。まず、先行研究で用いられた RNA-seq の発現変動解析用リアルカウントデータに対して、様々な群間での AS 値と PDEG 値を計算している。先行研究で示された根拠は発現変動パターンの分類結果と HSC 結果の主観的な評価に基づくものであったが、AS が比較する群間の全体的な類似性を客観的なスコアとして利用できることを示している。次に、2 群間比較用シミュレーションデータの解析を行い、反復数 (同一群内のサンプル数) の影響を調べている。結果として、発現変動解析結果として得られた PDEG の推定値は反復数 (Nrep) 依存性があり、反復数が多くなるほど真の値に近づく傾向にあることが述べられている。一方 AS 値は反復数非依存であり、反復数が少ない場合 (例: Nrep = 3) でも多少のゆらぎはあるものの、真の値に近い結果が得られることが示されて

いる。

次に、同様の解析を酵母のリアルカウントデータに対して行った結果についても述べている。これは、42 サンプル vs. 44 サンプルの 2 群間比較用データであり、発現変動解析結果として PDEG = 78.1%を、そして AS = 0.7289 という結果を得ている。各群から 3 サンプルずつランダム抽出して得られた反復数の少ない 2 群間比較の場合には、発現変動解析結果として得られた PDEG の推定値は 40%程度以下であるが、AS 値は 0.73 程度と安定した結果が得られている。また、AS が 0 に近い値となるデータの具体例として、ヒトのリアルカウントデータの 2 群間比較 (女 vs. 男) も行っている。AS が 0 付近のデータの場合には、反復数に関係なく PDEG も 0%付近となることが示されている。これは、発現変動解析結果として有意な発現変動遺伝子がほとんど得られなかった場合の理由付けとして、「サンプル間の全体的な類似度が比較する群間で非常に高いためである」ということを AS 値という客観的なスコアが使用できることを示唆するものである。

3.2 節は、マイクロアレイデータの 2 群間比較結果について述べている。AS 値と PDEG 値の相関があることを示しつつ、データの前処理法の違いによって数値分布の違いが生じうることも示している。3.3 節は、AS の多群間比較への拡張について、3 群間比較を例に述べている。AS 自体は、本来最適なクラスター数を得る目的で様々なクラスター数の候補ごとに算出するものであるため、3 群間比較の場合でも計算式をそのまま適用可能ではある。しかしながら多群間比較の場合は、サンプルごとのシルエットスコアを算出する際に「そのサンプルが属する群以外の最も近い群との平均距離」しか考慮されない。この問題を回避するため、AS の計算式に改良を加え、全ての群間の全体的な類似度を評価可能な指標 AAS を提案しその有用性を示している。

第四章では、総合討論として、結果のまとめや今後の展望を述べている。

以上、本論文はこれまで主観的に行われてきた遺伝子発現データの任意定義群における全体的な類似度を客観的に示す指標として、平均シルエットスコアの利用を提案したものである。発現変動解析結果の理解や解釈を行う上でも有用であることが示されており、学術上応用上寄与するところが少なくない。よって、審査委員一同は、本論文が博士 (農学) の学位論文として価値あるものと認めた。