

令和 2 年 6 月 30 日現在

機関番号：23903

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K12779

研究課題名(和文)フローサイトメトリーデータの測定条件に対して頑健な計算論的分析法の開発

研究課題名(英文) Development of robust clustering method for cytometry data obtained under different measurement conditions

研究代表者

大佐賀 智(Osaga, Satoshi)

名古屋市立大学・医薬学総合研究院(医学)・助教

研究者番号：60790772

交付決定額(研究期間全体)：(直接経費) 1,000,000円

研究成果の概要(和文)：フローサイトメトリーデータの機械学習に基づく細胞集団の自動分類手法として、測定条件の違いによるデータの分布の大幅な変動をアフィン変換として記述し、自動分類時にアフィンパラメータを推定しそれを打ち消すように補正することで、測定条件に対して頑健な自動分類法を提案した。これを人工データと実際のフローサイトメトリーデータに適用し、2次元の低次元なデータであれば提案手法が有効であることが確かめられた。しかしながら、通常フローサイトメトリーで得られる10次元程度のデータについては、提案手法によるパラメータ推定が収束には至らず、適切な分類ができないことが明らかとなった。

研究成果の学術的意義や社会的意義

フローサイトメトリーにおいて測定条件によるデータの変動は大きな問題であり、実臨床や多施設共同研究での利用の障害になっている。本提案手法は高次元データでは十分なパフォーマンスを発揮できず実用に耐えるものにはなっていないが、測定条件に対して頑健な自動分析法を開発する上で本手法のアプローチは重要な示唆を与えるものと考えている。また、それは将来的にはフローサイトメトリーのより広い応用にもつながると期待している。

研究成果の概要(英文)：I suggested an automatic classification method for cell populations in flow cytometry dataset with consideration of samplewise variability caused by differences of measuring conditions. My method was applied to both artificial and actual flow cytometry data and confirmed to be effective when data were two-dimensional. However, for data of about 10 dimensions, usually obtained by flow cytometry, a parameter estimation of the proposed model did not converge, and proper classification was not possible.

研究分野：バイオインフォマティクス

キーワード：サイトメトリー 機械学習 クラスタリング

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

フローサイトメトリーは光散乱や細胞表面抗原の免疫蛍光染色によって細胞の鑑別や分類を行う手法であり、従来はデータの散布図から人の目で細胞集団を分類する方法が取られてきたが、技術改良によってより多数の抗原を同時に測定することが可能になるにつれて目視での分類が困難になってきているため、近年では機械学習に基づくさまざまな自動分類法が提案されている。しかしながら、フローサイトメトリーデータは測定に用いる試薬や反応条件、検査機器といった測定条件の違いに起因する変動が非常に大きいことが知られており、一方で自動分類法は各細胞の測定値の類似度に基づいたクラスタリングにより細胞分類を行うため、異なる条件で測定されたデータが混在するデータセットには適用が困難であった。この問題は、一般に測定条件のばらつきが大きい実臨床で得られるデータセットや多施設共同研究で得られるデータセットに自動分類法を適用するにあたって、特に大きな障害となっていた。

2. 研究の目的

本研究では、測定条件の違いに対して頑健なフローサイトメトリーデータの自動分類法の確立するため、(1) 既存手法である HDPGMM (文献) の拡張として、測定条件の違いによって生じるばらつきを明示的に導入した統計モデルの構築と、そのパラメータ推定アルゴリズムの定式化を試みた。また、(2) 提案手法による自動分類の性能を実データを用いて評価し、実臨床への応用可能性について検討することを目指した。

3. 研究の方法

(1) 各検体のデータにそれぞれ HDPGMM を適用することでそれらの分布を混合正規分布として推定し、検体間で分布が似通ったものになるようにアフィンパラメータを推定した。当初はアフィンパラメータの生成モデルを HDPGMM のモデルに追加して同時に推定することを想定していたがその定式化が困難だったため、各検体の分布を個別に HDPGMM で推定した後に検体間の分布の距離 (提案手法ではカルバック・ライブラー情報量を使用) が最小になるようにアフィンパラメータを推定する形で定式化を行った。なお、HDPGMM による混合正規分布の推定とアフィンパラメータの推定は交互に行い、アフィンパラメータが概ね収束するまで学習を継続するものとした。

(2) 初めにアフィン変換により完全に重なる条件で 2 次元混合正規分布にしたがう人工データを作成し、それに提案手法を適用することで正しく自動分類ができるかを検証した。次に同一の検体を異なる 2 条件で測定して得られたフローサイトメトリーデータについて提案手法を適用し、同様に正しく自動分類できるかを検証した。

4. 研究成果

(i) 人工データによる検証

3 つの 2 次元正規分布から構成される混合正規分布に従う人工データ 1 を生成し、それとアフィン変換により完全に重なる条件の混合正規分布から人工データ 2, 3 をそれぞれ生成した。このデータに提案手法を適用して自動分類したところ、人工データ同士の対応する正規分布から生成された点が正しく分類されることが確かめられた (図 1)。また、人工データでは真のアフィンパラメータは既知であるが、それが正しく推定されていることも確認した。

(ii) 実データによる検証

同一検体を異なる 2 条件で測定した 10 次元のフローサイトメトリーデータについて、データ間で分布の異なる 2 次元のみを取り出し提案手法を適用して自動分類したところ、対応すると考えられる集団が概ね正しく分類されることが確かめられた (図 2)。しかしながら、元の 10 次元データに提案手法を適用したところ、HDPGMM の推定とアフィンパラメータの推定が繰り返されアフィンパラメータがいつまでも収束しない状況が頻発した。これはアフィンパラメータの推定において HDPGMM の対数尤度とは異なる最適化基準が存在しているため、高次元のデータになるほどいずれかの最適化がもう一方の最適化を崩して結果的に堂々巡りに陥りやすいものと推測された。これを解決するためには HDPGMM のモデルにアフィンパラメータの推定を組み込み 1 つの対数尤度の最適化に落とし込むことが考えられるが、研究期間内にその具体的な方法の考案には至らなかった。

なお、本研究では複数の検体のデータが含まれた実臨床のデータセットへの提案手法の適用を予定していたため、一部のリファレンスデータに対しても提案手法を適用したが、これについても高次元データでは上記と同様の問題が発生した。複数の検体のデータを含むデータセットは計算量が非常に大きく結果の検討には非効率であることから、まずは同一検体から得られたデータでの最適化における問題の解決を優先したため、実臨床データについてそれ以上の検証はできなかった。

< 引用文献 >

Cron A et al. Hierarchical Modeling for Rare Event Detection and Cell Subset Alignment across Flow Cytometry Samples. PLOS Comput Biol. 2013;9:e1003130.

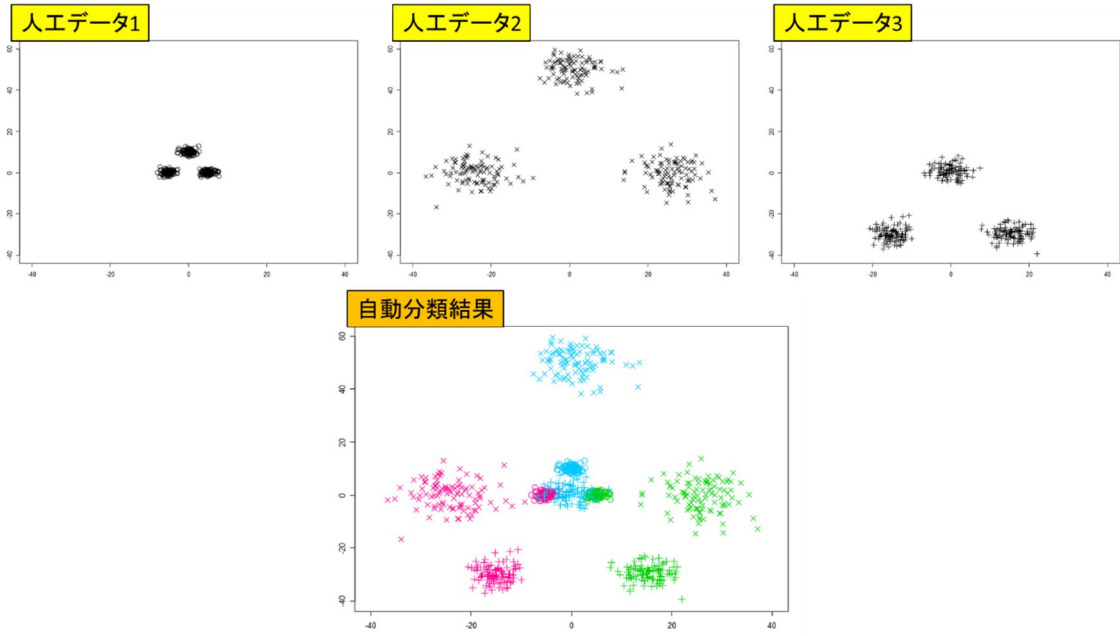


図 1. 人工データによる自動分類結果 (点の色が分類結果を表す)

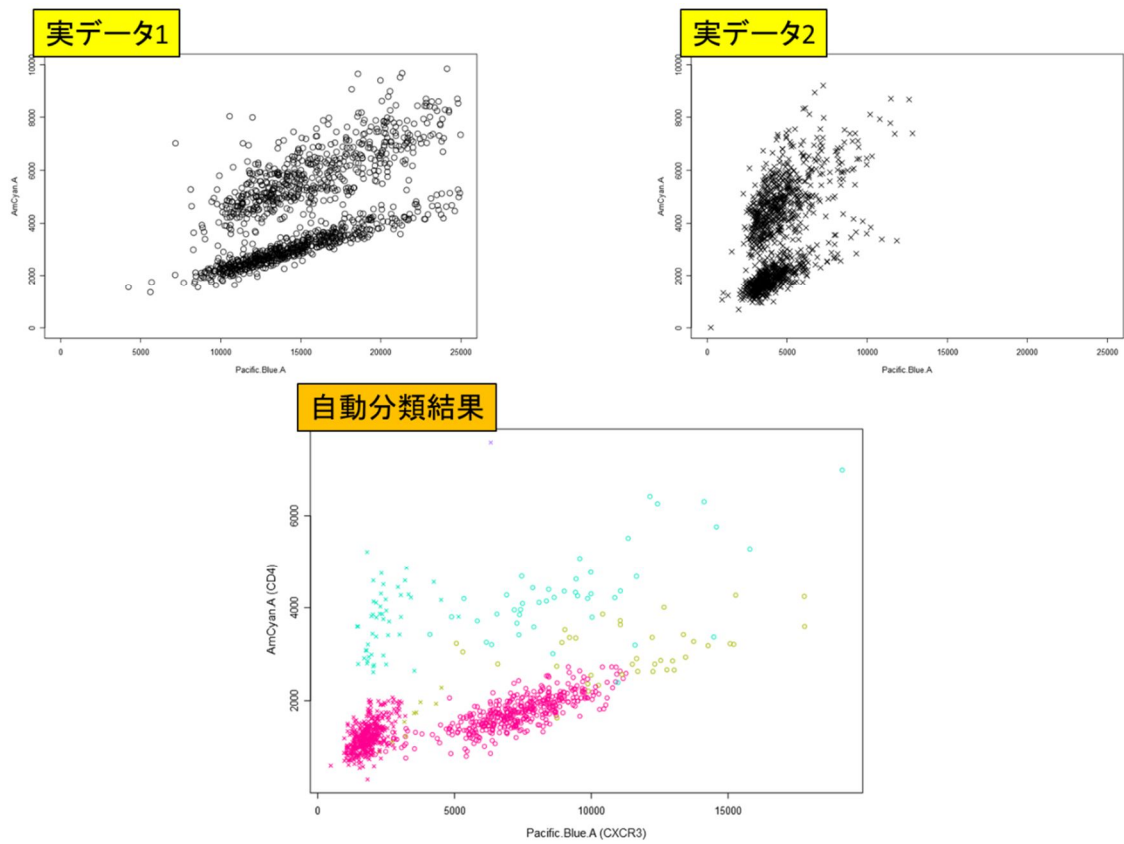


図 2. 実データによる自動分類結果 (点の色が分類結果を表す)

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計2件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Satoshi Osaga, Shinji Maeda, Takeshi Kamiya
2. 発表標題 Nonparametric Bayesian method for clustering of cytometry data obtained under different measurement conditions for each sample
3. 学会等名 CYTO 2018 (国際学会)
4. 発表年 2018年

1. 発表者名 Satoshi Osaga, Shinji Maeda, Takeshi Kamiya
2. 発表標題 Computational clustering method for cytometry data with sample-wise instrumental variability
3. 学会等名 ICCS 2017 (国際学会)
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	前田 伸治 (Maeda Shinji)	名古屋市立大学・呼吸器・免疫アレルギー内科学分野・助教 (23903)	