

令和 5 年 6 月 26 日現在

機関番号：14301

研究種目：研究活動スタート支援

研究期間：2021～2022

課題番号：21K21316

研究課題名（和文）幾何学と関数データ解析による医学・生命科学の時空間データマイニング

研究課題名（英文）Geometric and functional data analysis for spatiotemporal data mining in medicine and life sciences

研究代表者

岡田 大瑚 (Okada, Daigo)

京都大学・医学研究科・助教

研究者番号：10911852

交付決定額（研究期間全体）：（直接経費） 2,400,000円

研究成果の概要（和文）：本研究では、幾何学や関数データ解析を用いた、時間構造や空間構造を持った医学・生命科学領域のデータ駆動的な解析を行うための新規のデータ解析手法を開発した。特に、加齢ゲノミクス領域において、非線形な加齢変化を網羅的に検出し、その代表的なパターンを分類する関数データ解析に基づく新規のDICNAP法を開発した。この手法を公共DNAメチル化データセットに適用することで、DNAメチル化マーカーの非線形な加齢変化の全体像をあきらかにすることができた。

研究成果の学術的意義や社会的意義

加齢に過程で生じる変化には、線形なものだけでなく非線形な変化も数多くあることが知られている。本研究の学術的意義は、オミックスデータからデータ駆動的に加齢変化パターンの全体像を取得する関数データ解析に基づく手法を開発したことである。また、関数データ解析を用いたオミックス解析の有意義な実施例を示したことも意義の1つである。

研究成果の概要（英文）：In this study, we developed a novel data analysis method based on geometry and functional data analysis for data-driven analysis in the medical and life sciences domain with temporal and spatial structures. In particular, we developed a novel DICNAP method based on functional data analysis to comprehensively detect nonlinear aging changes and classify their representative patterns in the aging genomics field. By applying this method to a public DNA methylation dataset, we were able to reveal the overall picture of nonlinear aging changes in DNA methylation markers.

研究分野：計算生物学

キーワード：幾何学 関数データ解析 老化 オミックス

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年の医学・生命科学研究では、時間構造や空間構造の両方を持ったデータからの情報抽出問題は様々な状況で現れる。本研究では、臨床医学および基礎生命科学における代表的かつ重要性の高い時空間データ解析課題として、蛍光造影画像検査データ解析と加齢エピゲノム解析に取り組む。これら2つの課題は一見大きく異なるものの、類似のデータ構造を持っており統一的なアプローチが可能であると考えられる。

2. 研究の目的

本研究の目的は、(1) 蛍光造影検査が示す病巣のデータ駆動的な評価および自動検出プログラムの開発、(2) エピゲノムデータの変化を定量的に評価するためのプログラムの開発という2種類の課題に取り組むことを通じて、医学・生命科学分野における時空間データから情報抽出を行うための、幾何学や関数データ解析を応用した手法を確立することである。

3. 研究の方法

(1) 幾何学と関数データ解析を用いた蛍光造影画像の解析

蛍光造影検査は、眼科医療において造影剤が眼底の血管を流れる様子を時系列で撮影することで眼の病気を診断する検査手法であり、臨床医学における典型的な時空間データからの情報抽出問題である。眼科医は、病巣部においては"造影剤が毛細血管からじわじわとしみだしてくる"という特有の状態を示すことから病巣部を同定する。眼底蛍光造影画像の病巣部位を情報科学を用いて定量的に評価することや病巣を自動で同定・検出することが求められている。

本研究では、以下のような方法でこの課題にアプローチした。画像データを時系列に沿って縦に結合すると、このような幅×高さ×時間の3次元アレイとして表現できる。ここで、"毛細血管からのじわじわとした染み出し"はこのアレイ中の錐体として定義でき、シグナル値の時系列変化は図形上での関数として表現できる。このように、時空間データ中の情報は時空間データ中に埋め込まれた図形と関数のペアとして網羅的に列挙する。図形の分類および特徴づけには幾何学的な特徴量が利用でき、関数の分類及び特徴づけには関数データ解析と呼ばれる統計手法群を用いて行う。

(2) 加齢エピゲノムデータの解析

老化は、がんや認知症等の加齢性疾患の原因であり、医学研究の重要な課題である。加齢研究において、身体の老化の程度を評価する生体分子老化バイオマーカーの開発が中心的なテーマである。例えば、テロメア長や p16 遺伝子の発現量は、年齢との相関が報告されており、老化の代表的なバイオマーカーとされている。近年、網羅的な生体分子発現データ(オミックスデータ)を利用した老化バイオマーカー探索研究が盛んに行われている。

老化バイオマーカーの探索では、通常、年齢との直線的な関係が前提とされるが、加齢に伴う非線形な変化も多く存在することが考えられている。例えば、加齢性疾患のリスクは年齢に比例して増加するわけではなく、中年以降にリスクが加速度的に増大することが示されている。また、疾患ごとに発症リスクの変化パターンが異なることが知られている。生体における各種形質や生体分子発現量の加齢に伴う変化は、背後にある生物学的メカニズムの変化を反映しているとされる。しかし、加齢に伴う非線形変化の全体像はまだ明らかになっていない。

この問題に対処するため、潜在的な加齢変化パターンをデータ駆動的に同定する関数データ解析を用いた新たなデータ解析手法、DICNAP(Data-driven Identification and Classification of Nonlinear Aging Patterns)法を開発した。この方法は、非線形関数解析と関数データ解析を組み合わせた複数段階のプロセスから構成される。まず、年齢と関連するDNAメチル化サイトを識別するためMICと呼ばれる統計指標を用いて非線形相関分析を行った。MICとメチル化強度の間の非線形な相関を計算した。その結果を用いて、ゲノムワイドなバイオマーカーを無関連、線形増加、線形減少、非線形の4つのパターンに分類する。次に、これらの年齢関連サイトを線形および非線形パターンに分類した。MICとピアソンの相関係数を用いて非線形度を指数化し、非線形指数が一定以下のサイトを線形な加齢関連サイトとした。さらに、これら線形サイトはピアソンの相関係数に基づいて増加または減少とに分類した。三番目のステップでは、非線形な年齢関連メチル化サイトのノンパラメトリックな関数推定を行った。スプライン平滑化によるノンパラメトリック回帰を用いて、年齢関連の変化の関数を推定した。次に、関数主成分分析(FPCA)を用いて、非線形な年齢関連メチル化サイトを低次元空間に埋め込んだ。FPCAは関数データを入力として受け取り、各関数の主成分(PC)スコアを出力する。このステップでは、先のステップで得られた加齢変化パターンの関数にFPCAを適用し、サイトを主成分座標空間に埋め込んだ。

最終ステップでは、非線形サイトの PC 座標を使用して多変量解析を行った。K-means クラスタリングを用いて非線形なサイトを分類し、各クラスタの PC 座標値の中央値を中心とし、各クラスタの典型的なパターンを再現した。さらに、非線形なマーカーを関数の形状に基づいて分類し、各グループの代表的な加齢変化パターンを同定した。このようなワークフローを用いて生体分子情報と加齢との関連を網羅的に解析することで、非線形な加齢変化の全体像が観察できるようになる。

この手法を、公共データベースから取得されたメチロームデータセットに適用し、加齢と関連する DNA メチル化マーカーの中には、どのような加齢変化パターンが存在しているかの全体像を調べた。

加えて、加齢によって生体分子発現の制御ができなくなることを考えると、バイオマーカー発現の変動性は加齢によって変化するはずである。DNA メチル化強度の変動性が加齢によって変化する場合、ノンパラメトリック回帰を適用したときの残差はこの変化を反映すると予想される。これまでの研究で、DNA メチル化の変動性が加齢に関連して増加することが示されている。しかしながら、この変動性が加齢によってどのように変化するかは明らかではない。そこで、提案手法をこの残差の関数に対しても適用し、メチル化強度の変動性が加齢に伴ってどのように変化していくのかについても、その加齢変化パターンを網羅的に調べた。

4. 研究成果

(1) 幾何学と関数データ解析を用いた蛍光造影画像の解析

Python を用いて開発したプログラムを、京都大学病院で取得された蛍光造影検査の実画像データに適用した。しかしながら、蛍光造影検査画像は、画像ごとの違いが大きく、病巣領域を抽出する汎用的なアルゴリズムの開発には至らなかった。今後も引き続き取り組んでいくことを予定している。

(2) 加齢エピゲノムデータの解析

提案手法を公開データベースから取得したヒトの大規模な DNA メチロームデータに適用し、メチル化強度の加齢変化の全体像とその潜在的な関数パターンを調査した。DNA メチル化はエピゲノムレベルでの典型的な生体分子情報であり、老化と関連する DNA メチル化マーカーが多数同定されている。

本研究では、DICNAP と名付けた新しいデータ駆動型計算アプローチを開発し、2つの異なる研究から得られた4つの大規模メチロームデータセットに適用した。この手法では、線形変化に加えて非線形変化にも注目し、潜在的な非線形パターンを調べた。データ駆動型解析の利点の一つは、過去の生物学的知識を用いることなく、大規模データセットから新たな生物学的知識を発見できることである。データ駆動型の計算機的アプローチもまた、ゲノム研究において様々なデータセットやタスクを解析するために採用されており、新しい生物学的洞察をもたらしている。

本研究により、DNA メチル化の加齢パターンには線形だけでなく、非線形な変化も多く存在することが判明した。代表的な非線形なパターンとして、単調に増加または減少するいくつかの関数パターンが抽出された。

本研究では、加齢に関連する DNA メチル化強度は、非線形の単調な脱メチル化パターンによって特徴付けられ、60歳前後を境にその傾向が緩やかになることを見出した。これは安定的に観察されるパターンであり、データセット間で再現された。DNA 複製中のシトシンメチル化のエラーは、加齢に伴う DNA 脱メチル化の受動的な原因である。加えて、ヒトの細胞分裂率は加齢とともに減少する。このパターンの原因は、加齢に伴う細胞分裂の頻度の低下により、DNA の脱メチル化が遅くなることかもしれない。

また、変動関数の非線形加齢変化は、メチル化強度が加齢とともに変化した部位の約 10% で同定された。同定された安定な加齢変化パターンは、変動が単調に増加し、中年期以降に加速する関数である。これらのパターンを示す部位では、単調に進行する生命システムの制御の破綻が中年期に加速することを示唆している。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Okada Daigo, Cheng Jian Hao, Zheng Cheng, Yamada Ryo	4. 巻 67
2. 論文標題 Data-driven comparison of multiple high-dimensional single-cell expression profiles	5. 発行年 2021年
3. 雑誌名 Journal of Human Genetics	6. 最初と最後の頁 215 ~ 221
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s10038-021-00989-9	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Okada Daigo, Zheng Cheng, Cheng Jian Hao, Yamada Ryo	4. 巻 44
2. 論文標題 Cell population based framework of genetic epidemiology in the single cell omics era	5. 発行年 2021年
3. 雑誌名 BioEssays	6. 最初と最後の頁 2100118 ~ 2100118
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/bies.202100118	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 Daigo Okada, Jian hao Cheng, Cheng Zheng, Ryo Yamada
2. 発表標題 非線形な加齢変化に着目した老化に対するDNAメチル化バイオマーカーの探索と分類
3. 学会等名 日本人類遺伝学会第66回大会
4. 発表年 2021年

1. 発表者名 Information geometric approach to embed multiple single cell expression profiles into low dimensional coordinate space
2. 発表標題 Daigo Okada, Jian Hao Cheng, Ryo Yamada
3. 学会等名 Pacifichem 2021
4. 発表年 2021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
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------