

令和 4 年 5 月 20 日現在

機関番号：12602

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20221

研究課題名（和文）ヒト体内細菌叢時系列データを用いた細菌叢と疾患・健康状態の関係予測の研究

研究課題名（英文）Prediction of microbiome-disease association using time series human microbiome data

研究代表者

長谷川 嵩矩（Hasegawa, Takanori）

東京医科歯科大学・M&Dデータ科学センター・准教授

研究者番号：80753756

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：近年、ゲノム・シーケンス技術の発達に伴い、個人が持つ全ゲノム情報に加えて、腸内や口腔内に存在する細菌が疾患や健康状態に対して与える役割を明確化し、臨床応用を試みる新たな研究が始まっている。本研究では、弘前大学COIにおいて複数年に渡り継続的に毎年約1000人分取得しているヒト腸内・口腔内細菌叢の時系列メタゲノムのデータと、同時に採取したヒトゲノムデータ、並びに血液検査値などのバイタルデータから、細菌叢が疾患や健康状態に与える影響を予測するための統計的解析手法の開発を行った。

研究成果の学術的意義や社会的意義

細菌叢と疾患、もしくは健康状態の解析に関しては昨今活発に研究が開始されている。主なものは、単独の細菌やクラスター化された細菌群と疾患の関連性を統計的検定によって明らかにする研究であり、この先進的な解析例としてはベイズ統計モデルを用いた解析手法や、Dirichlet回帰を用いて健康状態から細菌叢のコンポジションを予測しようとする研究などである。一方で、より先進的なショットガンメタゲノムデータに対して複雑な統計モデルを考慮し、疾患や健康状態との関連性を解析した例は多くない。本研究では、このような次世代技術を用いて採取されたデータと健康の関連性を予測するものであり、学術的意義の大きい研究である。

研究成果の概要（英文）：With the recent development of genome sequencing technology, new research has begun to clarify the role of bacteria in the intestines and oral cavity on diseases and health conditions in addition to the whole genome information possessed by individuals, and to attempt clinical applications. In this study, time-series metagenomic data on the intestinal and oral microbiota of approximately 1,000 individuals collected annually at the Hirosaki University COI for multiple years, human genome data collected simultaneously, and vital data such as blood test values were used in an integrated manner. Then, we developed a statistical analysis method for predicting the impact of bacterial flora on diseases and health conditions.

研究分野：統計科学

キーワード：統計的時系列解析 状態空間モデル メタゲノム解析 ゲノム解析 生命情報学

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年、ゲノム・シーケンス技術の発達に伴い、個人が持つ全ゲノム情報を読み解くコストは著しく低減している。結果として、個人の持つ遺伝子多型の情報のみならず、コピー数変異やエピゲノム修飾などを含めた網羅的遺伝情報が大量に得られるようになり、疾患や血液検査データなどを含むバイタルデータとの関連性が積極的に研究されている。特に昨今では、腸内や口腔内に存在する細菌叢が、関節炎や食物アレルギー、癌、炎症性腸疾患、神経疾患、肥満、糖尿病など様々な疾患において重要な役割を果たすということが明らかになってきている (Hall et al. 2017, Kostic et al. 2015, Wlodarska et al. 2015)。これらの研究成果から、ヒト体内細菌叢の役割をより明確化しようという試みが活発化しており、臨床応用までを視野に入れた新たな研究が始まっている。

細菌叢の解析においては、従来から 16S rRNA と呼ばれる遺伝子領域の情報を用いて同定された細菌の存在量情報(以下、16S rRNA データとする)が用いられて来た。この方法は安価だが、細菌の存在量を求めることしか出来ず、細菌それぞれが有している遺伝子多型や遺伝的欠損、もしくは機能遺伝子などの情報を得ることは出来ない。一方で近年、ショットガン法と呼ばれる方法を用いて細菌のゲノムデータ(以下、ショットガンデータとする)を詳細かつ網羅的に得られるようになった。この技術を用いることで、例えば上図のように、ある特定の機能遺伝子を持つ細菌がどの程度存在しているのか、といった細菌叢の詳細な情報を用いた解析が可能になるため注目を浴びている。

2. 研究の目的

本研究の目的は上述のデータを用いて、「どのような細菌叢」が、「どのような疾患もしくは健康状態」に対して、「どのような制御関係を有するのか」を明らかにするための、統計的解析手法の開発を行うものである。ショットガンデータを用いた細菌叢解析は世界的に見ても未だ例が多いものではなく、網羅的解析に関しての試みは現在進行形で行われている課題である。特に本研究で取得されている時系列データは類を見ないものであり、その統計的解析手法の開発とそこから得られる知見の学術的独自性は非常に高い。一方で、16S rRNA データを用いた細菌叢解析に関しては、細菌叢と疾患との関連性が明確に認識されたのは近年ではあるものの、データ自体が比較的安価に得られることから、解析手法はショットガンデータに対するものより先んじている。加えて、細菌叢のデータ解析においては、16S rRNA とショットガンデータ双方において細菌叢のデータはコンポジションと言われる細菌の割合データとして扱われることが多い。このことから、本研究の初期段階においては、既存の 16S rDNA データやショットガンデータから得られる細菌のコンポジションを対象とし、既存研究を踏襲した解析並びにその理論的拡張を実施することになる。

3. 研究の方法

細菌叢解析の具体的なハードルとして、細菌叢はサンプルによってバリエーションが豊かで観測値の次元が高くなる一方で、サンプルによっては得られない細菌などが多く存在するため、データが疎であるという特徴がある。一般的な解析においては、多くのサンプルで存在量が低い細菌を解析から省く、知られている細菌のみに絞って解析を行うなどの手順が取られることが多いが、そういった細菌の与える影響を無視することになってしまう。他にも、細菌の存在量の推定値を得るために複数の技術が組み合わさっているため、得られる観測ノイズのモデルが比較的複雑であることや、生活習慣(外部環境)によって変化しやすい細菌群が存在すること、存在量に強い相関性が見られる細菌群の存在などが挙げられる。

本研究では既存研究をベースに、(i)各種バイタルデータを網羅的に扱えるように既存モデルを拡張することから始め、(ii)これを統計的時系列解析のモデルとして拡張し、(iii)ショットガンデータから得られるより詳細なゲノム情報を統合可能なモデルと方法論を確立する、という3段階のステップで統計的解析手法の開発を実施した。具体的には、申請者の専門とする統計的時系列解析手法の一つである、データ同化と呼ばれる、地学分野や細胞内システム解析で応用の進む統計的解析手法を適用した。本研究で取り扱う時系列データは疎であり、また欠損値を多く有しており、通常の時系列解析手法を適用することが困難である。データ同化手法を用いることで、自由度の高いモデル設計を行った上で、変数間の制御関係を推定することが可能となる。具体的には、総和が1となるような0から1までの実数値で構成されるコンポジションデータを取り扱えるようにモデルを拡張し、パラメータ推定とモデル選択を効率的かつより高精度に計算可能な統計学的手法を開発する必要がある。また、複数の細菌群の混合作用を考慮出来るようなモデルを設計することも重要である。必然的に、非線形モデルや、観測ノイズも非ガウス型であるモデルを設計することになるため、より広範なノイズを扱える統計モデルへの理論的拡張や、Monte Carlo 法を用いた解析手法等を用いることになる。この問題に対する解決策として、申請者の専門である線形、並びに非線形状態空間モデルを血液検査値や細菌叢データに対して、適切に隠れ変数とパラメータの推定が行えるように拡張したモデルを開発した。

4. 研究成果

(1) 単一疾患だけではなく各種バイタルデータを網羅的に扱えるような統計モデルの適用

近年、ヒト腸内細菌叢のコンポジションデータと疾患の関連性を、latent Dirichlet allocation というモデルを用いて解析する研究が成果を挙げている。本研究では、ヒト体内細菌叢メタゲノムのデータに留まらず、血液データなど各種バイタルデータも同時に採取しているため、まずはこれら複数のデータを統合的かつ網羅的に扱えるような非時系列のベイズ統計モデルを実装し、その予測精度を検証した。

(2) 提案モデルの時系列解析の枠組みへの拡張とショットガンデータ特有の情報を扱えるモデルへの拡張

上述した通り、時系列データへの拡張として、状態空間モデルを用いた統計的時系列解析手法の開発に着手した。ここでは、状態空間モデルの観測関数に非ガウス型の観測ノイズを組み込んだ state-space model with skew-t measurement noise という拡張された線形状態空間モデルを開発し、そのパラメータを最尤推定に基づく方法で推定する統計的解析手法までを開発した。これにより、ガウス型を持たない観測ノイズを持つデータに関しても、適切なパラメータの推定と将来値の予測を高速に実施することが可能になった。また同時に、より複雑な時系列データを適用可能にするために、既存の非線形状態空間モデルである ensemble Kalman filter を拡張することにも着手した。既存の研究に加えて、条件付き確率分布の計算をより高速に行える ensemble Kalman smoother という統計的解析手法を開発した。これに対して、16S rRNA データや、ショットガンデータから得られたデータを事前処理することで得られる細菌叢のコンポジションデータを適用することで、ヒト体内細菌叢と健康状態の関連性予測を実施した。前者の統計的解析手法に関しては既に論文化実施済みであり、後者の実データ解析に関しては、現在結果をまとめており、論文化を行っているところである。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 T. Hasegawa, R. Yamaguchi, M. Kakuta, M. Ando, J. Songee, I. Tokuda, K. Murashita and S. Imoto	4. 巻 100
2. 論文標題 Application of state-space model with skew-t measurement noise to blood test value prediction	5. 発行年 2021年
3. 雑誌名 Applied Mathematical Modelling	6. 最初と最後の頁 365-378
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.apm.2021.08.007	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 T. Hasegawa, R. Yamaguchi, A. Niida, S. Miyano, S. Imoto.	4. 巻 357
2. 論文標題 Ensemble smoothers for inference of hidden states and parameters in combinatorial regulatory model. Journal of the Franklin Institute	5. 発行年 2020年
3. 雑誌名 Journal of the Franklin Institute	6. 最初と最後の頁 2916-2933
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.jfranklin.2019.10.015	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 T. Hasegawa, R. Yamaguchi, M. Kakuta, K. Sawada, K. Kawatani, K. Murashita, S. Nakaji and Seiya Imoto	4. 巻 15
2. 論文標題 Prediction of blood test values under different lifestyle scenarios using time-series electronic health record	5. 発行年 2020年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0230172
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0230172	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計0件

〔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
---------	---------