

平成 31 年 4 月 22 日現在

機関番号：12601
研究種目：若手研究(B)
研究期間：2017～2018
課題番号：17K12647
研究課題名(和文) 時系列検診データとゲノム・メタゲノム情報を統合した生活習慣病の重篤化予測の研究

研究課題名(英文) Predicting severity of lifestyle-related diseases by integrating time-series health records with human genome and metagenome data

研究代表者
長谷川 嵩矩 (Hasegawa, Takanori)

東京大学・医科学研究所・助教

研究者番号：80753756
交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：近年、日本を含む多くの先進諸国において、総人口に占める高齢者比率の増加とそれに伴う生活習慣病患者の増加が世界的な問題になっている。本研究では、これまで西欧諸国を中心に研究が進んできた心疾患のリスク予測手法を日本人に適用可能な形に修正・拡張し、近年研究の進むゲノム・腸内細菌叢メタゲノム情報を統合した生活習慣病の新たなリスク予測手法を開発した。特に、血液などのバイタル値が持つ非線形な影響を考慮した予測精度の高い統計的時系列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
世界的に大きな注目を浴びる生活習慣病の早期リスク予測手法の開発に対して、データ同化手法を適用することで重篤化のシミュレートを行い、状態推定とリスク予測を行う点に特色がある。また本研究は臨床応用へのインパクトが高いと思われる。アジア人に対する慢性疾患のリスク予測手法の開発は遅れており、指標の提案は高齢化先進国と呼ばれる日本のみならず、今後高齢化が加速するアジア圏内において1つの重要な指標として取り入れられる可能性がある。

研究成果の概要(英文)：In recent years, in many developed countries including Japan, an increase in the ratio of elderly people to the total population and an increase in patients with lifestyle-related diseases have become a global issue. Through the improvement and extension of existing risk prediction methods for heart disease whose research has been advanced mainly in western countries, we newly develop a risk prediction method for Japanese population that integrates genome and bacterial flora metagenomic data. In particular, we developed a statistical time series analysis method with high prediction accuracy, taking into account non-linear effects of vital values for lifestyle related diseases.

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 リスク予測 ゲノムデータ 腸内細菌叢ゲノム 健康診断データ

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

近年、日本を含む多くの先進諸国において、総人口に占める高齢者人口の割合が増加しており、それに伴う生活習慣病患者の増大が世界的な問題となっている。医学の発展に伴い、感染症の占める死亡者数の割合は減少の一途を辿っている一方で、世界的に死因の大きな割合を生活習慣病が占めるようになっており、中高年・高齢者の生活習慣病の重篤化が医療・福祉費の大きな負担となっている。中でも心疾患・脳疾患 (Cardio/Cerebral Vascular Disease, CVD) は生活習慣病による死因の 50%以上の割合を占めるため、その兆候の発見と予測・予防は世界的に重要なテーマとなっている。

2. 研究の目的

心疾患のリスク予測手法の開発は 2000 年代から活発に行われており、有名なものとしては、Framingham General CVD Risk Indicator, Reynolds CVD risk score, ACC/AHA CVD risk calculator, JBS3 risk score, MESA risk score など様々な方法が提唱されている。しかしながら、これらの予測手法は日本社会への適用に際して以下に述べる 3 つの問題が指摘できる。(i) 提唱されている手法は欧米を中心に行われたコホートスタディのデータを用いて作成されており、アジア人に対するリスクの識別性能が低い。(ii) 近年では一般的に採取されている血液データの一部や、ゲノム・ヒト体内細菌叢メタゲノム情報などが用いられていない、(iii) 変数の影響を直感的に理解出来るよう、簡易な一般化線形モデルでリスク計算がなされており、またバイタル値など説明変数の時系列遷移も考慮していない。

本研究の目的は、これまで西欧諸国を中心に研究が進んできた心疾患のリスク予測手法を日本人に適用可能な形に修正・拡張することからスタートし、近年研究の進むゲノム・腸内細菌叢メタゲノム情報を統合した生活習慣病の新たなリスク予測手法を開発することである。特に、血液などのバイタル値が持つ非線形な影響を考慮した予測精度の高い統計的時系列解析手法の開発までを実施することである。

3. 研究の方法

本研究計画は、(i) 既存の CVD リスク予測手法の拡張、(ii) ゲノム・腸内細菌叢メタゲノム情報を統合した生活習慣病のリスク予測手法の開発、(iii) 提案手法の非線形状態空間モデルへの拡張の 3 段階に分けることが出来る。初年度は (i) と (ii) に着手した。終了年度では、(iii) に着手するが、申請者はデータ同化に関する論文を多数出版しており、これらをベースに統計的解析手法の開発を行った。具体的には、Kalman filter の拡張として近年提案されている closed-skew t-distributed Kalman filter を用いることで、生活習慣病の重篤化シミュレーションを実施する。

4. 研究成果

まず、既存の CVD リスク予測手法の日本人への適用とモデルの拡張、知見の蓄積、加えてゲノム・腸内細菌叢メタゲノムデータを用いた生活習慣病のリスク予測手法の開発を行った。心疾患のリスク評価手法は幾つも提案されており、まずはこれらの手法に日本のコホートから得られる検診データを直接適用した。同時に、既存の指標では用いられていないバイタル情報や生活習慣データを用いた場合の、心疾患や重度慢性疾患 (高血圧症・糖尿病・慢性腎不全・肝硬変・慢性膵炎) の発症・重篤化に対する予測精度の評価と予測精度の高いモデルの選択を行うことで、今後の解析に用いるべき変数の評価と選定を行った。ここにおいては、Genome Wide Association Study の技法を用いて、個々人が持つ遺伝情報 (SNPs) のうちバイタルデータに影響を持つ因子を抽出することや、腸内細菌叢メタゲノムデータからコンポジションのデータを抽出し、解析の変数として用いる準備を行った。

次に、取得されている時系列の健診データを扱うため方法として、状態空間モデルを用いた解析手法の開発に着手した。具体的には、線形の状態空間モデルを検診データに対して適切に適用できるようにモデルの提案を行い、上述した変数を導入し (遺伝的データは取得年度に限られているため、データ数担保の観点から本解析には用いなかった)、隠れ変数と外部環境変数の依存関係を推定した。これにより、相関性の高い観測変数を同時に扱い、それらの制御関係を推定することが可能になった他、生活習慣などの外部環境変数がバイタルデータや将来の疾患リスクに与える影響を推定することが可能になった。

終了年度においては、前年度に開発されたモデルを非線形・非ガウス型の状態空間モデルに拡張し、これを実データに対して適用した。まず前年度の解析に基づいて、ある健康状態と生活習慣、遺伝子多型情報を持つ個人の重度慢性疾患に関わるバイタルデータ値の将来的な値の予測や生活習慣病の発症を予測する新たなモデルの開発に着手した。ここで状態空間モデルの隠れ変数間の因果関係を推定し、適したモデル選択するためには、観測データが得られたときの隠れ変数の条件付き確率分布を正確に求める必要がある。前年度の研究から明らかになったこととして、健康診断データのノイズの形状はガウス型ではなく、例えば食事によって中性脂肪の値が大きく上がるように、片側に裾の重い分布になっていることが挙げられる。このような

特徴的な現象に対処する方法として、近年 Kalman filter の理論的拡張の1つとして提案されている、closed-skew t-distributed Kalman filter を適用した。同モデルを用いることで、前述したようなノイズを適切に表現することが可能になり、このモデルをベースにした提案モデル内の必要な構造・パラメータの推定を可能とする手法の開発も同様に行った。実際に提案手法を時系列の健康診断データに対して適用することで、健康診断時の観測ノイズの影響を吸収しつつ、環境からのバイタル値の影響を明らかにすることが可能であることが明らかになった。本研究結果は現在学術論文として執筆中である。

5. 主な発表論文等

〔雑誌論文〕(計1件)

T. Hasegawa, K. Kojima, Y. Kawai, and M. Nagasaki. Time-Series Filtering for Replicated Observations via a Kernel Approximate Bayesian Computation, IEEE Transactions on Signal Processing, vol. 66(23), pp6148-6161, 2018.

〔学会発表〕(計1件)

長谷川嵩矩, 新井田厚司, 山口類, 井元清哉. 状態空間モデルを用いた時系列ヘルスケアデータの解析と応用, 統計関連学会連合(招待講演), 2018.

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
出願年：
国内外の別：

取得状況(計 件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕

ホームページ等

6. 研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号(8桁)：

(2)研究協力者

研究協力者氏名：

ローマ字氏名：

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。