

科学研究費助成事業 研究成果報告書

平成 29 年 5 月 6 日現在

機関番号：32305

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26350868

研究課題名(和文)クラウドコンピューティングによる健康ビッグデータ処理に関する研究

研究課題名(英文)Study on Machine Learning of Health Big Data Based on Cloud Computing

研究代表者

竹内 裕之(Takeuchi, Hiroshi)

高崎健康福祉大学・健康福祉学部・教授

研究者番号：00348129

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)： 毎日の粒度で長期に亘って時系列的に蓄積された健康ビッグデータを、これまでの研究で明らかになったデータの周期的な季節変動をフィードバックさせて処理し、個人の健康管理に役立つ知識を抽出する手法を開発した。1例として、試行中のクラウド型個人健康管理システムの1人のボランティアユーザが8年7か月に亘り蓄積した毎粒度の家庭血圧データとエネルギー消費/摂取に関わる生活習慣データとの関係を遅延相関分析法をコアとした健康データマイニング技術により分析した。その結果、家庭血圧の周期的な季節変動を変動幅に基づき補正することにより時系列データのばらつきは正規分布に近づき、両者の間に有用な相関ルールが得られた。

研究成果の概要(英文)： Long-term time-series data of an individual's home blood pressure measurements were analyzed to clarify the relation with lifestyle components such as energy expenditure and supply. We found that the distribution of daily blood pressure data approached normal when the seasonal variation observed in the time-series data was adjusted. This demonstrates that adjustment is desirable for producing appropriate rules between blood pressure and lifestyle components in the healthcare data mining process. The results of healthcare data mining using the adjusted data are summarized as follows:(1)The adjustment of blood pressure data affected the rule production process and different rules were extracted with the adjustment because the blood pressure data were categorized into three classes (higher, moderate, or lower) when being used for an output (target) variable.(2)Rules extracted with the adjustment were more useful for personal healthcare than rule extracted without adjustment.

研究分野：健康情報学

キーワード：健康データマイニング 個人健康管理システム 時系列データ解析 クラウドコンピューティング 機械学習 遅延相関分析

1. 研究開始当初の背景

(1) インターネットを活用した健康医療分野のユビキタス化が進展しており、最近の関連国際学会の潮流として、m(mobile)Health や p(personalized)Health といった概念が浸透している(例えば、Proc. 14th World Congress on Medical and Health Informatics (Medinfo 2013)(2013))。特に体重、体脂肪率、血圧など個人の健康に関するデータが家庭や職場などでも容易に取得できるようになり、携帯電話やスマートフォンといった携帯端末を通してインターネット上(クラウド)に蓄積できる技術が開発されている。

(2) 我々はいち早くこの潮流を捉え、基盤研究(C)「携帯電話と Web テクノロジーを活用した個人健康管理システムに関する研究」(平成 16-18 年度)および基盤研究(B)「特定健診・保健指導を支援するシステムの開発に関する研究」(平成 20-22 年度)により、クラウドで処理を行う自動健康データマイニングをコア技術とした「個人健康管理システム」および「特定健診・保健指導を支援するシステム」を開発してきた。これらのシステムは、時系列的にクラウドに蓄積された個人の生活習慣と日常の健康に関するデータを解析して相関ルールを抽出(健康データマイニング)し、結果を個人が参照できる、正にクラウドコンピューティングをいち早く具現化したものである。続く基盤研究(C)「健康データマイニングによる個人の特質クラスタ解析に関する研究」(平成 23-25 年度)では、クラウド上に蓄積された個人のデータを解析することにより、生活習慣と健康状態の相関に関わる特質の個人差を明らかにしてきている。また、ビッグデータとも呼べる、日毎の粒度の長期にわたる個人のデータを解析することにより、季節変動などのパターンにも興味深い特徴が見いだされている。

2. 研究の目的

(1) 本研究では、これらの成果をさらに発展させて、すでに開発した自動健康データマイニングのアルゴリズムによって生活習慣と健康状態に関わる相関ルールを抽出するのみでなく、日毎の粒度で時系列的に蓄積されたビッグデータを多角的に処理して、個人の健康管理に役立つ知識を抽出する手法を開発する。具体的には、

個人の体脂肪率や血圧等に季節変動が生じていることを反映(フィードバック)させて、長期間の生活習慣と健康状態の相関を求める処理方法、

季節をバーチャルアイテムとして捉えて相関ルールを抽出する処理方法、などを開発する。これらにより、個人差を含めたより本質的な相関関係を追及する。

(2) 個人の長期に亘る日毎の粒度の時系列データから季節変動など別の要因による変化のパターンを抽出し、そのパターンをバイアスとして処理することにより、生活習慣と健康状態の相関ルールに季節に依らない一貫性が出るかを検証する。これまでは、季節変動のルール生成への影響を抑えるために3カ月間(1季節)のデータに基づき相関ルールマイニングを行っていた。またこれまで行っている基盤研究で、生活習慣と健康双方に関して同じ目的・項目でデータが蓄積されているシステムユーザ例が出てきたので、意識的に同じデータ項目についてデータ蓄積するボランティアユーザを募り、生活習慣と健康状態の相関に個人差が生じている原因、即ち他の生活習慣の違いによるのか遺伝子レベルでの差異(遺伝子多型)によるのか等を、探っていく。

3. 研究の方法

(1) 我々が開発している健康データマイニングでは、「生活習慣の蓄積が健康状態に変化をもたらし、その影響は時間遅れをもって現れることがある」という極めてシンプルなモデルをベースとしている。すなわち、ある健康状態(例えば体脂肪率)の変化を目的変数とし、時間遅れを考慮したある期間の生活習慣(例えば運動による消費エネルギー)の蓄積を入力変数として相関ルールを抽出する。相関ルールの抽出には、あらかじめ時系列データを基にして、目的変数である健康状態の変化に影響を及ぼす生活習慣の蓄積をスクリーニングする。スクリーニングには、健康状態の変化(ある期間の差分)と遅延期間を考慮した生活習慣の蓄積(ある期間の加算)の時系列データ間のピアソンの積率相関係数を評価する。ここで、差分をとる期間、加算する期間、遅延期間の3つをパラメータとして相関係数を評価し、相関係数の絶対値が最大となる条件を見つける(遅延相関分析法)。そして、その最大値がある閾値を超えた場合に、その生活習慣の蓄積を健康データマイニングの入力変数として採用する。

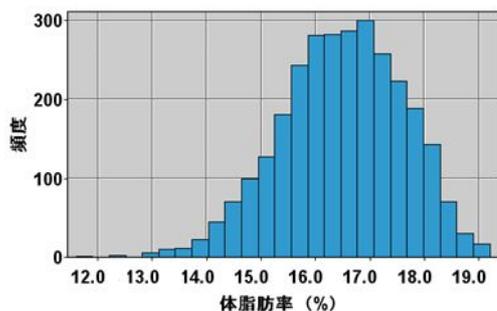
(2) 本研究では、健康状態の変化に季節変動など他の要因があることをアприオリとして捉え、長期にわたる健康ビッグデータを処理する手法を開発する。即ち、季節変動などのパターンを考慮して時系列相関を評価し、ルールマイニングを行う処理手法を開発する。また、これまでの研究で、蓄積するデータ項目(生活習慣、健康状態)が同じシステムユーザ例が出てきたので、意識的にこのようなボランティアを募り、抽出される相関ルールの個人差についてその原因を追究して、個人毎のテーラーメイド健康管理実現への足がかりを創る。

4. 研究成果

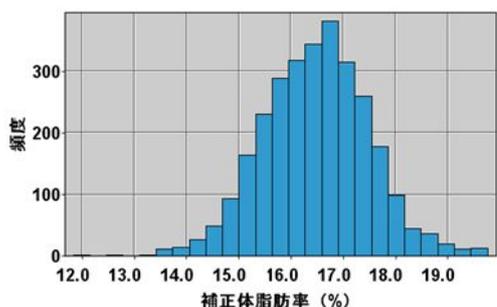
(1) 試行中のクラウド型個人健康管理シス

テムの1人のボランティアユーザが8年7か月に亘り蓄積した日毎粒度の体脂肪率データとエネルギーの消費と摂取に関わる生活習慣データに基づき、開発中の健康データマイニング技術を評価した。主な結果を以下に纏めた。

体脂肪率データは夏に低く冬に高い明らかな周期的季節変動を示し、その変動幅を基にした補正を施すことにより、時系列データのばらつきは正規分布に近づいた(図1)。



(a) 季節変動補正前



(b) 季節変動補正後

図1 体脂肪率時系列データの分布

季節を跨る時系列データを対象に健康データマイニングを実行する場合には、季節変動補正後のデータを用いることにより、ユーザの健康管理により役立つ相関ルールが得られた。

(2) 試行中のクラウド型個人健康管理システムの同じボランティアユーザがやはり8年7か月に亘り蓄積した日毎粒度の家庭血圧データとエネルギーの消費と摂取に関わる生活習慣データに基づき、開発中の健康データマイニング技術を評価した。主な結果を以下に纏めた。

家庭血圧のデータも夏に低く冬に高い明らかな周期的季節変動を示し、その変動幅を基にした補正を施すことにより、時系列データのばらつきは正規分布に近づいた。

家庭血圧についても季節変動補正後のデータを用いて健康データマイニングを実行することにより、ユーザの健康管理により役立つ相関ルールが得られた。

(3) これらの結果により、個人の長期に亘る

日毎粒度の体脂肪率や家庭血圧の時系列データから季節変動など別の要因による変化のパターンを抽出し、そのパターンをバイアスとして処理することにより、生成される相関ルールの有用性向上を実現できることが判った。

(4) 生活習慣と健康状態に関わる相関ルールの個人差に注目し、新たに健康データとして家庭での末梢血管血流速度計測を採りあげ、6人の被験者について、睡眠時間および有酸素運動との間の時系列データ解析を試みた。主な結果を以下に纏めた。

6被験者中4被験者において、気温の高い夏季期間に血流速度の顕著な増加が観測され、生体に備わる深部体温調節機能の現れと考えられた。他の2被験者については、気温が下がる秋口から血流速度が下がる傾向のみ観測された。

6被験者すべてにおいて、血流速度と睡眠時間の間に有意な正の相関がみられ、睡眠による血液循環改善の現れと考えられた。

血流速度と歩数の間に、3被験者で有意な正の、2被験者で負の相関がみられ、1被験者では有意な相関がみられず、未だ全くの仮説であるが、原因として運動による血液循環改善と大腿筋への血流配分の個人差が考えられた。

(5) 生活習慣と健康状態に関わる相関ルールの個人差に注目し、さらに新たな健康データとして家庭で取得する末梢血管中ヘモグロビン量を採りあげた。4人の被験者について、無侵襲の近赤外分光画像計測法を用いてヘモグロビン推定量を取得し、睡眠時間および運動習慣との間の遅延相関分析を行った。主な結果を以下に纏めた。

継続的に鉄分を多く含む食品やサプリメントを摂取することによりヘモグロビン推定量は増加傾向がみられた。

適度な継続的有酸素運動はヘモグロビン推定量を増加させる傾向がみられるが、アスリートが行うような激しいトレーニングは逆に減少させた。

睡眠時間が長いと個人差はあるがヘモグロビン推定量は増加する傾向がみられた。

(6) このように蓄積するデータ項目(生活習慣、健康状態)が同じになるボランティアを募り、抽出される相関ルールの個人差についてその原因を追究することにより、個人毎のテーラーメイド健康管理実現への足がかりをつかむことができた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計6件)

竹内裕之、児玉直樹、近赤外分光画像計

測法による末梢血液中ヘモグロビン濃度の時系列データ解析—個人の生活習慣との関係において—、高崎健康福祉大学紀要、査読有、16巻、2017、1-8

Hiroshi Takeuchi and Naoki Kodama, Time-Series Data Analysis of Home Blood Flow Velocity Measurements from Radial Artery in Relation to Lifestyle Factors、Proc. 4th IEEE International Conference on Biomedical and Health Informatics、査読有、2017、425-428

DOI : 10.1109/BHI.2017.7897296

竹内裕之、児玉直樹、個人健康管理の手段としての超音波ドップラー法による末梢血管血流速度計測、高崎健康福祉大学紀要、査読有、15巻、2016、1-9

Hiroshi Takeuchi, Naoki Kodama, and Shingo Takahashi, Time-Series Data Analysis of Long-Term Home Blood Pressure Measurements in Relation to Lifestyle、Proc. 15th World Congress on Health and Biomedical Informatics、査読有、2015、751-755

DOI : 10.3233/978-1-61499-564-7-751

竹内裕之、児玉直樹、高橋真悟、健康データマイニングの評価() - 長期間の時系列データに基づく事例、高崎健康福祉大学紀要、査読有、14巻、2015、1-12

Hiroshi Takeuchi and Naoki Kodama, Validity of Association Rules Extracted by Healthcare-Data-Mining、Proc. 36th Annual International Conference of the IEEE EMBS、査読有、2014、4960-4963
DOI : 10.1109/EMBC.2014.6944737

[学会発表](計6件)

竹内裕之、児玉直樹、個人の末梢血液中ヘモグロビン濃度の時系列データ解析、第9回データ工学と情報マネジメントに関するフォーラム(deim 2017)、2017.3.7、高山グリーンホテル(岐阜県)

Hiroshi Takeuchi and Naoki Kodama, Time-Series Data Analysis of Home Blood Flow Velocity Measurements from Radial Artery in Relation to Lifestyle Factors、4th International Conference on Biomedical and Health Informatics (BHI 2017)、2017.2.18、Orlando (USA)

竹内裕之、児玉直樹、超音波ドップラー法を用いて計測した個人の末梢血管血流速度の時系列データ解析、第8回データ工学と情報マネジメントに関するフォーラム(deim 2016)、2016.3.2、ヒルトン福岡シーホーク(福岡県)

Hiroshi Takeuchi, Naoki Kodama, and Shingo Takahashi, Time-Series Data Analysis of Long-Term Home Blood Pressure Measurements in Relation to

Lifestyle、15th World Congress on Health and Biomedical Informatics、2015.8.21、Sao Paulo (Brazil)

竹内裕之、児玉直樹、高橋真悟、個人の体脂肪率と生活習慣との相関ルール生成にデータの季節変動が及ぼす影響、第7回データ工学と情報マネジメントに関するフォーラム(deim 2015)、2015.3.3、磐梯熱海ホテル華の湯(福島県)

Hiroshi Takeuchi and Naoki Kodama, Validity of Association Rules Extracted by Healthcare-Data-Mining、36th Annual International Conference of the IEEE EMBS、2014.8.28、Chicago (USA)

6. 研究組織

(1) 研究代表者

竹内 裕之 (TAKEUCHI, Hiroshi)
高崎健康福祉大学・健康福祉学部・教授
研究者番号 : 0 0 3 4 8 1 2 9

(2) 研究分担者

倉林 正彦 (KURABAYASHI, Masahiko)
群馬大学・医学(系)研究科(研究院)・教授
研究者番号 : 0 0 2 1 5 0 4 7

(3) 研究分担者

児玉 直樹 (KODAMA, Naoki)
高崎健康福祉大学・健康福祉学部・准教授
研究者番号 : 5 0 3 8 3 1 4 6