

平成 30 年 6 月 25 日現在

機関番号：17401

研究種目：基盤研究(B) (一般)

研究期間：2015～2017

課題番号：15H02705

研究課題名(和文)大規模医療データのための統合的情報解析技術の開発

研究課題名(英文)Development of Advanced Analytical Technology for Large-scale Medical Data

研究代表者

櫻井 保志 (Sakurai, Yasushi)

熊本大学・大学院先端科学研究部(工)・教授

研究者番号：30466411

交付決定額(研究期間全体)：(直接経費) 13,300,000円

研究成果の概要(和文)：近年、医療機関では電子カルテからの診察記録や症例登録など、様々な医療データが大量に蓄積されている。医療データは検査や診断、投薬などの医療サービスの内容が患者毎に、そして時間情報とともに蓄積される。そして、ビッグデータ解析は医療サービスの質の向上および効率化を図り、医療の様々な問題を解決できる重要なアプローチとして期待されている。本研究では、このような大規模かつ多種多様な複合データをモデル化し、統合的に解析するための高度なデータマイニング技術を開発する。

研究成果の概要(英文)：With the aim of improving health care and medical services, we develop a unifying advanced analytical technology that will make it possible to extract latent relationships in complex time-stamped medical data (e.g., electronic medical records). Medical data analytical technology is considered important throughout the world. Our developed method can identify important dynamical patterns and forecast long-range future events, incrementally and automatically.

研究分野：データマイニング

キーワード：医療情報 時系列解析 予測

1. 研究開始当初の背景

近年、医療機関では電子カルテからの診察記録や症例登録など、様々な医療データが大量に蓄積されている。医療データは検査や診断、投薬などの医療サービスの内容が患者毎に、そして時間情報とともに蓄積される。そして、ビッグデータ解析は医療サービスの質の向上および効率化を図り、医療の様々な問題を解決できる重要なアプローチとして期待されている。本研究では、このような大規模かつ多種多様な複合データをモデル化し、統合的に解析するための高度なデータマイニング技術を開発する。医療データは多項目であり、またレコード数は非常に多い。開発した技術は、このような大規模な複合データを効果的かつ効率的に要約、モデル化し、そして項目間の関係性を抽出する。

2. 研究の目的

本研究では、電子カルテのような複数項目からなる医療データを複合医療データと呼ぶ。本研究では複合医療データを効率的に解析するための基盤技術、特に複合データの時系列予測技術を開発することを目的とする。

3. 研究の方法

医療に関する複合データは大規模かつ多様化しており、そのような複雑なビッグデータの時系列解析を行い、将来予測のような高度な処理を行なうため、本研究では従来の研究とは異なる、新たな研究テーマを提案し、研究開発を行なった [T1, T2, T3]。

(1) 非線形テンソル解析

本研究ではこれまで、(a)大規模テンソル解析、(b)非線形時系列解析、(c)特徴自動抽出という3種類の時系列解析技術を開発した。そして、これらの要素技術全てを融合させ、非線形テンソル解析技術を世界に先駆けて開発した。非線形テンソル解析技術により、時系列ビッグデータを自動的かつ統合的にモデル化し、全ての情報の関係性を抽出することが可能となり、情報予測においては大幅な精度向上を達成した。

(2) リアルタイム予測

人々の行動、自然現象や社会現象の発生における様々な事象の間の相互作用をモデル化し、リアルタイムに情報を提供するための予測技術を開発した。

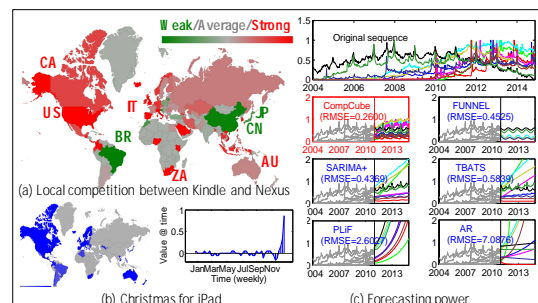
4. 研究成果

4.1 非線形テンソル解析

大規模テンソル解析、非線形時系列解析、特徴自動抽出の3つの要素技術を融合、発展

させ、多数の時系列データ間の複雑な関係性を表現するための非線形テンソル解析技術 CompCube [C4]を開発した。テンソルとして表現された複合データから、基本特徴、相互関係、季節性、外れ値の4つの情報を Global/Local 両視点から抽出する統合モデルである。非線形テンソル解析は計算機科学において世界で初めての取り組みであり、各時系列データに対し個々の独立したモデルを当てはめるのではなく、時系列ビッグデータを多角的に、かつ非線形性を考慮しながら分析することにより、各シーケンス間の潜在的な関係性をとらえることが可能となる。モデルの表現力を高めることにより精度を大幅に向上させ、現存する予測手法と比較し、世界最高の精度を示している。

開発した CompCube は、複数の属性から構成される複合データをテンソルとして表現し、さらにテンソルの時間発展を非線形方程式によってモデル化し、解析する。例えば、Web 上のユーザの地域別の活動データ (time, activity, location) に本手法を適用すれば、データから基本情報、競合関係、季節性、外れ値の4つの情報を、Global/Local 両視点から抽出することができる。各国における Google 検索件数のデータが与えられたとき、図のように提案手法は、競合関係 (Kindle vs. Nexus 等)、地域別季節性 (Christmas, Chinese New Year) 等の情報を抽出し、各地域における今後のユーザ行動の予測を行うことができる。



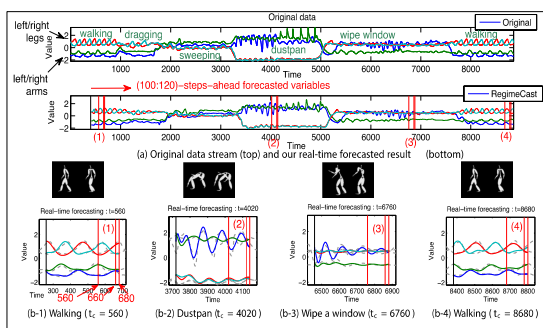
本手法を Web 情報に適用させた論文は WWW2015 および WWW2016 に採択されており、国際的に高い評価を受けている [C4, C6]。

4.2 リアルタイム予測

IoT データ、Web 情報、医療データなど、実社会の実データは、様々な環境変化や人々の状況や行動の変化を示しており、その変化は局所的、突発的であることも多い。そこで、局所的もしくは突発的な変化に対して即座に対応することができる適応力のある予測技術 RegimeCast を開発した。これは、時系列データストリームに対してリアルタイムに将来のパターンを予測し続ける技術であり、セグメンテーションとモデル推定、モデル選択、将来値の生成、全てをデータストリーム上で高速に処理する。

下図の上段はオリジナルセンサストリー

ム、中段は 100 ステップ先の予測結果、下段は各時刻におけるリアルタイム予測のナップショットを示す。



論文は KDD2016 に採択され、国際的に高い評価を受けている [C3]。リアルタイム性と適用力という、従来手法にはないコンセプトの本技術は、同会議において企業関係者から最も注目を浴びた技術であり、現在数多くの共同研究につながっている。

4.3 開発技術の医療情報解析への応用

本研究において開発した非線形テンソル解析技術は感染症のデータ解析に応用すれば、感染症患者数を高い精度で予測することができ、本研究ではそれを評価実験により確認した。また、項目数の多い複合データには、深層学習と組み合わせることにより、高い予測結果を実現することができる(例えば、[W1, W5]など)。

また、活動量や心拍数など、生体データの解析にはリアルタイム予測技術が有効であり、これを実証実験により確認した。介護データを解析することにより、要介護度を改善させるための知見を得ることが可能となる。今後、介護データ解析に関してさらに実験をすすめ、成果をまとめていく予定である。

4.4 その他の成果

上記を含めた時系列ビッグデータ解析の一連の取り組みは ACM Transaction などのトップ国際雑誌、ACM SIG などのトップ国際会議に数多く、継続的に採録され、国際的に高く評価されている。また、本研究では SIGMOD 2015、WWW 2016、KDD 2017 において 3 時間のチュートリアル講演をしている。これらはデータベース、Web、データマイニングの分野での最難関の国際会議であり、各々のチュートリアル講演はいずれも日本人としては初めての採択である。また、開発した時系列ビッグデータ解析技術は様々な分野の有力企業から注目されており、トヨタ自動車、トヨタ IT 開発センター、富士通研究所、SONY、ソニーセミコンダクタマニュファクチャリング、三菱重工エンジン&ターボチャージャ株式会社、日立製作所などと共同研究を実施している。

特にトヨタ IT 開発センターとは、車両走

行センサーデータの解析技術の共同開発を行っている。非線形テンソル解析技術により、大規模な車両走行センサーデータからの車両走行パターンの自動検出に成功している。ハンドル操作、加速や減速、停止など、車両走行の様々な共通パターンを抽出すると同時に、慎重な走行、スムーズで安定した走行、渋滞時の走行など車両走行のグループ化を完全自動で行なうことができる [J9]。

共同研究契約を結ぶ企業や技術開発で連携する法人が急激に増加しており、様々な分野の有力企業や法人と連携し、それぞれの応用ドメインでの技術の実証実験を行い、技術の実用化に取り組んでいる。

5. 主な発表論文等

(スペースの都合により研究代表者分のみ)

[雑誌論文](計 9 件)

J1. 川畑光希, 松原靖子, 櫻井保志: ``自動パターン検出のためのストリームアルゴリズム'', 情報処理学会論文誌:データベース, Vol. 11, No. 1, pp. 1-10, 2018 年 4 月.

J2. 松原靖子, 櫻井保志, Christos Faloutsos: ``大規模オンライン活動データの特徴自動抽出'', 情報処理学会論文誌:データベース, Vol.10, No.3, pp. 1-5, 2017 年 10 月.

J3. Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, Christos Faloutsos: ``Non-linear Dynamics of Information Diffusion in Social Networks'', ACM Transactions on the Web (TWEB), Vol. 11, No. 2, pp. 11:1-11:40, May 2017.

J4. Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos: ``Ecosystem on the Web: non-linear mining and forecasting of co-evolving online activities''. World Wide Web, Springer, Vol. 20, Issue 3, pp. 439-465, May 2017.

J5. 松原靖子, 櫻井保志, Christos Faloutsos: ``生態系モデルに基づくオンライン活動データの非線形解析'', 電子情報通信学会論文誌 D, Vol.J100-D, No. 4, pp.457-471, 2017 年 4 月 (2017 年度論文賞, 研究会推薦論文).

J6. 松原靖子, 櫻井保志: ``大規模データストリームのリアルタイム予測'', 情報処理学会論文誌:データベース, Vol.9 No.4, pp. 32-45, 2016 年 12 月.

J7. 松原靖子, 櫻井保志, Willem G. van Panhuis, Christos Faloutsos: ``大規模モ

病データのための非線形モデル解析", 情報処理学会論文誌:データベース, Vol.9, No.4, pp. 17-31, 2016年12月.

J8. Thinh Minh Do, Yasuko Matsubara, Yasushi Sakurai: "Non-linear Time-series Analysis of Social Influence", Journal of Information Processing, Vol. 24, No. 6, pp. 937-945, November 2016.

J9. 本田 崇人, 松原 靖子, 根山 亮, 櫻井保志: "車両走行センサデータからの自動パターン検出", 情報処理学会論文誌:データベース, Vol. 9, No. 3, pp. 1-13, 2016年9月.

[学会発表](計 32件)

査読付き国際会議論文

C1. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Smart Analytics for Big Time-series Data", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Tutorial, Halifax, Nova Scotia, Canada, August 2017.

C2. Thinh Minh Do, Yasuko Matsubara, Yasushi Sakurai: "Automatic and Effective Mining of Coevolving Online Activities", Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 233-246, Jeju, South Korea, May 2017.

C3. Yasuko Matsubara, Yasushi Sakurai: "Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 1045-1054, San Francisco, California, August 2016 (Acceptance Rate: 70/784, 8.9%).

C4. Yasuko Matsubara, Yasushi Sakurai, Christos Faloutsos: "Non-Linear Mining of Competing Local Activities", International World Wide Web Conference (WWW), pp. 737-747, Melbourne, Canada, April 11-15, 2016 (Acceptance Rate: 115/727, 15.8%).

C5. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Mining Big Time-series Data on the Web", International World Wide Web Conference (WWW), Tutorial, pp. 1029-1032, Montreal, Canada, April 2016.

C6. Yasuko Matsubara, Yasushi Sakurai,

Christos Faloutsos: "The Web as a Jungle: Non-Linear Dynamical Systems for Co-evolving Online Activities", International World Wide Web Conference (WWW), pp. 721-731, Florence, Italy, May 18-20, 2015 (Acceptance Rate: 131/929, 14.1%, Citations: 25).

C7. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Mining and Forecasting of Big Time-series data", ACM SIGMOD Conference, Tutorial, pp. 919-922, Melbourne, VIC, Australia, May/June 2015 (Citations: 27).

チュートリアル講演

T1. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Smart Analytics for Big Time-series Data", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Halifax, Nova Scotia, Canada, August 13, 2017.

T2. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Mining Big Time-series Data on the Web", International World Wide Web Conference (WWW), Montreal, Canada, April 12, 2016.

T3. Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos: "Mining and Forecasting of Big Time-series data", ACM SIGMOD Conference, Melbourne, VIC, Australia, June 3, 2015.

招待講演

I1. 櫻井保志: 日本コンピュータ化学会 2017 秋季年会 特別講演, "時系列ビッグデータのリアルタイム解析: 新技術と挑戦", 主催と場所: 日本コンピュータ化学会、日本化学会、くまもと県民交流館パレア 日時: 2017年10月21日

I2. 櫻井保志: くまもと技術革新・融合研究会 (RIST) シンポジウム, "時系列ビッグデータのリアルタイム解析: 新技術と挑戦", 主催と場所: くまもと技術革新・融合研究会 (RIST) 公益財団法人くまもと産業支援財団、ニューオータニホテルズ ザ・ニューホテル熊本 日時: 2017年10月19日

I3. 櫻井保志: 電子情報通信学会 PRMU/IE/MI/SIP 合同研究会 特別講演, "時系列ビッグデータのリアルタイム解析: 新技術と挑戦", 主催と場所: 電子情報通信学会パターン認識・メディア理解研究会 (PRMU), 画像工学

研究会(IE) , 医用画像研究会(MI) , 信号処理研究会(SIP), 名古屋工業大学
日時: 2017年5月25日

14. Yasushi Sakurai: Keynote Speech: “Real-time Forecasting of IoT Big Data: Foundations and Challenges”, the 11th Korea-Japan Database Workshop (KJDB), Daegu, South Korea, November 19, 2016.

15. 櫻井保志: 第19回情報論的学習理論ワークショップ (IBIS2016), “実社会データへの機械学習応用”
主催と場所: 電子情報通信学会情報論的学習理論と機械学習(IBISML)研究会, 京都大学
日時: 2016年11月17日

16. 櫻井保志: 電子情報通信学会 IE 研究会 映像情報メディア学会 ME-AIT 研究会 合同研究会 (感性とメディアおよび高精細度画像処理, 一般) “時系列ビッグデータ解析の新たな展開”
主催: 電子情報通信学会 画像工学研究会 (IE), 映像情報メディア学会メディア工学研究会 (ME), 映像表現&コンピュータグラフィックス研究会 (AIT), 福岡大学
日時: 2016年10月6日

17. 櫻井保志: 第5回生命医薬情報学連合大会 (IIBMP 2016), “ビッグデータのための非線形時系列解析とその応用”
主催と場所: 日本バイオインフォマティクス学会 (JSBi) 日本オミックス医療学会, 情報計算化学生物学会 (CBI 学会), 東京国際交流館プラザ平成
日時: 2016年9月28日

18. 櫻井保志: ERATO 感謝祭 Season3, JST ERATO 河原林巨大グラフプロジェクト 招待講演, “Mining and Forecasting of Big Time-series Data”
主催と場所: JST ERATO 河原林巨大グラフプロジェクト, 学術総合センター
日時: 2016年8月9日

19. Yasushi Sakurai: POSTEC Database Workshop 招待講演, “Mining Big Time-series Data on the Web”
主催と場所: POSTEC (Pohang University of Science and Technology), Department of Computer Engineering, 韓国 POSTEC
日時: 2016年5月30日

110. 櫻井保志: 日本基礎心理学会 2016年度第1回フォーラム 招待講演, “時系列ビッグデータ解析とその応用”
主催と場所: 日本基礎心理学会, 慶應義塾大学
日時: 2016年5月21日

111. Yasushi Sakurai: Invited Talk: “Mining and Forecasting of Big Time-series Data”, NII Shonan Meeting Seminar 082, Shonan Village Center, March 28, 2016.

112. 櫻井保志: ACM SIGMOD 日本支部第59回支部大会(第22回先端的データベースとWeb技術動向講演会)招待講演, “時系列ビッグデータ解析の新たな方向性”
主催と場所: ACM SIGMOD 日本支部, リコーITソリューションズ株式会社本社事業所
日時: 2015年6月13日

口頭発表(研究会およびワークショップ)

W1. 衛藤 亮太, 松原 靖子, 山下 和人, 國澤 進, 今中 雄一, 櫻井 保志: “電子診療報酬データの多角的解析”, 第10回データ工学と情報マネジメントに関するフォーラム (DEIM2018) 論文集, K6-3, 2018年3月4日-3月6日 (学生プレゼンテーション賞).

W2. 山室 冴, 松原 靖子, 川畑 光希, 井手 優介, 梅田 裕平, 古川 忠延, 丸橋 弘治, 稲越 宏弥, 大川 佳寛, 櫻井 保志: “深層学習を用いた時系列データの要約と分類”, 第10回データ工学と情報マネジメントに関するフォーラム (DEIM2018) 論文集, C3-3, 2018年3月4日-3月6日 (学生プレゼンテーション賞).

W3. 松原 靖子, 櫻井 保志: “大規模データストリームの将来予測アルゴリズム”, 第10回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 2017年9月18日-9月20日 (最優秀論文賞).

W4. 松原 靖子, 櫻井 保志, Willem G. van Panhuis, Christos Faloutsos: “大規模疫病データのための非線形モデル解析”, 第10回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 2017年9月18日-9月20日 (優秀論文賞).

W5. 衛藤 亮太, 松原 靖子, 山下 和人, 國澤 進, 今中 雄一, 櫻井 保志: “深層学習を用いた電子カルテ医療情報の多角的解析”, 第10回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 2017年9月18日-9月20日 (学生奨励賞).

W6. 川畑 光希, 松原 靖子, 櫻井 保志, “自動パターン検出のためのストリームアルゴリズム”, 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 (DEIM2016) 論文集, G1-6, 2016年2月29日-3月2日 (最優秀論文賞, 学生プレゼンテーション賞).

W7. 松原 靖子, 櫻井 保志, Christos Faloutsos, ``生態系モデルに基づくオンライン活動データの非線形解析'', 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 (DEIM2016) 論文集, G4-4, 2016 年 2 月 29 日-3 月 2 日 (優秀論文賞).

W9. 本田 崇人, 松原 靖子, 根山 亮, 櫻井 保志: ``車両走行センサデータからの自動パターン検出'', 第 8 回 Web とデータベースに関するフォーラム (WebDB Forum 2015) 論文集, pp. 134-141, 2015 年 11 月 24-25 日 (最優秀論文賞, 学生奨励賞).

W10. Minh Thinh Do, Yasuko Matsubara, Yasushi Sakurai: ``Non-linear Time-series Analysis of Social Influence'', 第 8 回 Web とデータベースに関するフォーラム (WebDB Forum 2015) 論文集, pp. 16-23, 2015 年 11 月 24-25 日 (企業賞(楽天)).

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 2 件)

名称: ``時系列データストリームのための情報予測装置、方法、及びプログラム”

発明者: 松原 靖子, 櫻井 保志

権利者: 熊本大学

種類: 特許

番号: 特願 2016-138075

出願年月日: 2016 年 7 月 12 日

国内外の別: 国内

名称: ``パラメータ集合生産方法及びプログラム”

発明者: 松原 靖子, 櫻井 保志

権利者: 熊本大学

種類: 特許

番号: PCT/JP2017/025236

出願年月日: 2017 年 07 月 11 日

国内外の別: 国外

取得状況 (計 1 件)

名称: ``電池型センサノード、電子機器識別装置、方法、及びプログラム”

発明者: 前川 卓也, 岸野 泰恵, 柳沢 豊, 櫻井 保志

権利者: 日本電信電話株式会社 (NTT)

種類: 特許

番号: 5746072

取得年月日: 2015 年 5 月 15 日

国内外の別: 国内

〔その他〕

受賞

A1. 2017 年度電子情報通信学会 論文賞 (2018 年 6 月 7 日)

A2. 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017) 最優秀論文賞 (2017 年 9 月 19 日)

A3. 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017) 優秀論文賞 (2017 年 9 月 19 日)

A4. 2016 年度電子情報通信学会, 画像工学研究会, IE 特別賞 (2017 年 5 月 25 日)

A5. 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 (DEIM2016) 最優秀論文賞 (2016 年 6 月 18 日)

A6. 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 (DEIM2016) 優秀論文賞 (2016 年 6 月 18 日)

A7. 第 8 回 Web とデータベースに関するフォーラム (WebDB Forum 2015) 最優秀論文賞 (2015 年 11 月 24 日)

ホームページ

<http://www.cs.kumamoto-u.ac.jp/~yasushi/index-j.html>

http://www.dm.cs.kumamoto-u.ac.jp/sakurai_lab/ja/

6. 研究組織

(1) 研究代表者

櫻井 保志 (Yasushi, Sakurai)

熊本大学・大学院先端科学研究部・教授
研究者番号: 30466411

(2) 研究分担者

有次 正義 (Masayoshi, Aritsugi)

熊本大学・大学院先端科学研究部・教授
研究者番号: 40282412

(3) 研究分担者

吉川 正俊 (Masatoshi, Yoshikawa)

京都大学・大学院情報学研究科・教授
研究者番号: 30182736

(4) 研究分担者

今中 雄一 (Yuichi, Imanaka)

京都大学・大学院医学研究科・教授
研究者番号: 10256919

(5) 研究協力者

Christos Faloutsos

カーネギーメロン大学・Dept. of Computer Science・教授