

令和元年6月24日現在

機関番号：17401

研究種目：挑戦的萌芽研究

研究期間：2016～2018

課題番号：16K12430

研究課題名(和文) オンラインアクティビティの高速解析技術の開発

研究課題名(英文) Realtime Data Mining for Online Activities

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(工)・教授

研究者番号：30466411

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：本研究では、シーケンス間の複雑な関係性やダイナミクスを捉える新たなモデルとオンラインアクティビティのための高速解析手法を開発することを目的とする。多種多様なオンラインアクティビティのデータを解析するため、本プロジェクトでは3つの技術、非線形テンソル解析、データストリーム予測、リアルタイム要因分析を開発した。研究成果として、非線形テンソル解析技術はWeb分野におけるトップ国際会議であるWWW2016に採択され、またデータストリーム予測とリアルタイム要因分析技術はデータマイニング分野におけるトップ国際会議であるKDD2016とKDD2019に採択され、いずれの技術も国際的に高い評価を受けている。

研究成果の学術的意義や社会的意義

近年のIT技術の急速な発展により、Web空間ではデータ量が飛躍的に増大し、そして現在FacebookやTwitterなどの巨大なソーシャルネットワーク上では、日々大量の情報が高速に流通している。そして、これらの情報の流れは刻々と変化している。このようなWeb情報の時間発展の解析は、特定のビジネスのみならず、社会経済の活性化、環境、防災やエネルギーなど、重要な社会問題を解決するための効果的なアプローチとして期待されている。このような背景の中で本プロジェクトでは、オンラインアクティビティのための高速解析技術を開発した。

研究成果の概要(英文)：The goal of this research project is the development of mining and forecasting technologies for big data of Web information. Specifically, our algorithms (namely, non-linear tensor analysis, data stream forecasting, and realtime causality mining) handle a wide variety of big data, and capture the latent interaction between social events, and forecast future social activities.

研究分野：マルチメディア・データベース

キーワード：オンラインアクティビティ Webマイニング 将来予測

1. 研究開始当初の背景

近年の IT 技術の急速な発展により、Web 空間ではデータ量が飛躍的に増大し、そして現在 Facebook や Twitter などの巨大なソーシャルネットワーク上では、日々大量の情報が高速に流通している。そして、これらの情報の流れは刻々と変化している。このような Web 情報の時間発展の解析は、特定のビジネスのみならず、社会経済の活性化、環境、防災やエネルギーなど、重要な社会問題を解決するための効果的なアプローチとして期待されている。このような背景の中で本研究課題では、オンラインアクティビティのための高速解析技術を開発する。

2. 研究の目的

従来から Web におけるオンラインアクティビティの時系列解析は盛んに行なわれてきた。しかし、その多くは特定の事象のみを対象として単一の時系列シーケンスを個別に解析するか、もしくは複数シーケンスをまとめて解析する場合でも時間変化の類似性や相関のみに着目して解析することが多く、シーケンス間の複雑な関係性やダイナミクスを捉えることは難しかった。これらの問題を解決するため本研究では、シーケンス間の複雑な関係性やダイナミクスを捉える新たなモデルとオンラインアクティビティのための高速解析手法を開発することを目的とする。

3. 研究の方法

多種多様なオンラインアクティビティのデータに対して高度な処理を行なうため、本研究では従来研究とは異なる、新たなコンセプトの技術を開発した[T1, T2, T3]。

(1) 非線形テンソル解析

櫻井らは大規模テンソル解析、非線形時系列解析、特徴自動抽出という3種類の時系列解析技術をこれまで開発していたが、本研究では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。開発技術である CompCube については、研究成果が Web 分野におけるトップ国際会議である WWW2016 に採択され、国際的に高い評価を受けている[C5]。

(2) データストリーム予測

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 RegimeCast を開発した。研究成果はデータマイニング分野におけるトップ国際会議である KDD2016 に採択され、国際的に高い評価を受けている[C4]。

(3) リアルタイム要因分析

OrbitMap は、IoT/センサデータストリームをはじめとする大規模な時系列データに対し、リアルタイムに重要な特徴や潜在的なトレンド(レジーム)を発見し、各レジーム間の動的な関係性を抽出することで、長期的かつ継続的に時系列イベントストリーム内の重要な動的要因を監視し、将来のイベント予測を行う。研究成果は KDD2019 において採録が決定している[C1]。

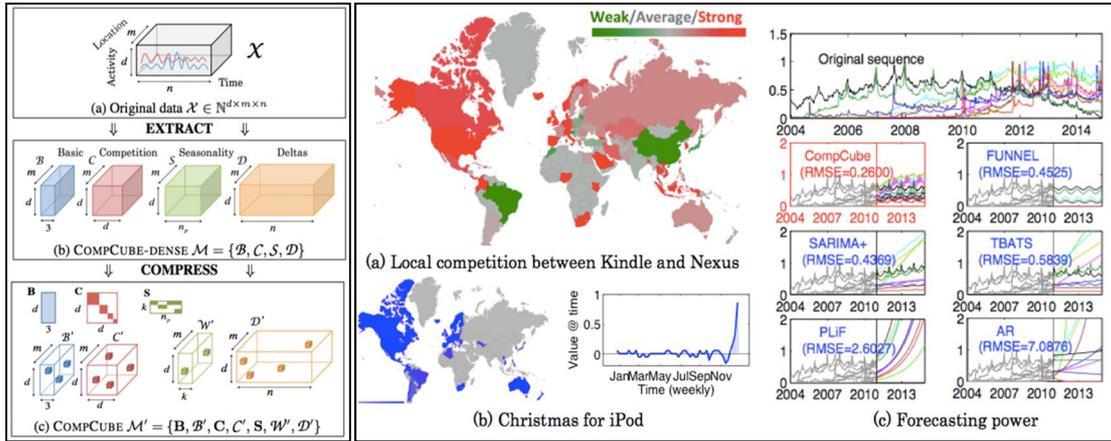
4. 研究成果

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 [C5]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

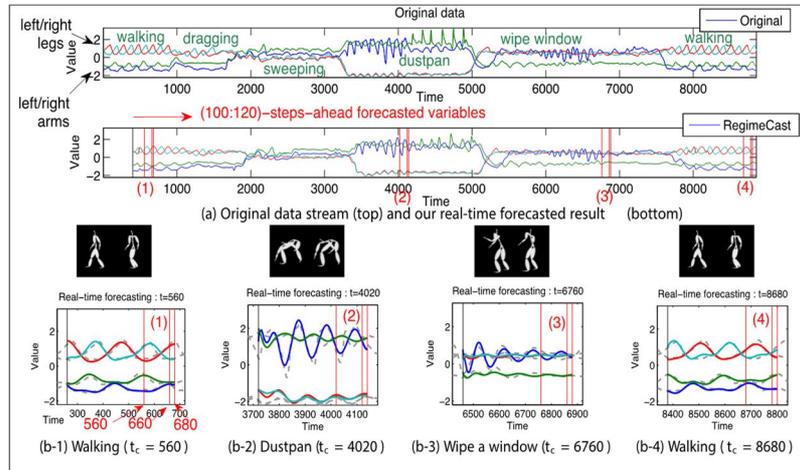
域における今後のユーザ行動の予測を行うことができる。



4.2 データストリーム予測

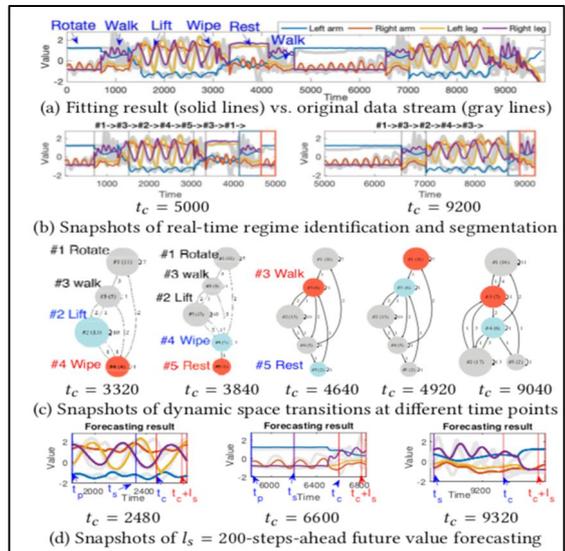
実社会の実データは、様々な環境変化や人々の状況や行動の変化を示しており、その変化は局的、突発的であることも多い。そこで、局所的もしくは突発的な変化に対して即座に対応することができる適応力のある予測技術 RegimeCast を開発した。これは、Web 情報をはじめとする時系列データストリームに対してリアルタイムに将来のパターンを予測し続ける技術であり、セグメンテーションとモデル推定、モデル選択、将来値の生成、全てをデータストリーム上で高速に処理する。図はセンサデータを用いた実験結果を示しており、RegimeCast は Web 情報でも同様に予測することができる。

右図の上段はオリジナルセンサストリー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スナップショットを示す。上述のように論文は KDD2016 に採択され、国際的に高い評価を受けている[C4]。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 リアルタイム要因分析

本研究では、時系列イベントデータから時系列モデル間の前後関係（要因 - 結果関係）を捉え、それらの事象の連鎖を動的空間遷移ネットワークとしてモデル化する。さらに、本技術は、要因分析と動的空間遷移ネットワークを用いることで、予測精度のさらなる向上も達成している。OrbitMap は Web 情報の解析に活用できるが、さらに IoT データにも適用可能である。図2は OrbitMap を用いた加速度センサデータの解析結果である。具体的には、工場における作業者の両手足4箇所加速度センサを設置し、100Hz で加速度を計測している。図(a)は、オリジナルのデータストリームの学習結果を示し、図(b)は、各時刻におけるリアルタイムレジーム検出例、図(c)は各時刻におけるネットワークの成長の様子を示す。図(d)は、学習した動的モデルとネットワークを用いたリアル



タイム予測の様子を示している。例えば、自動車走行における急ブレーキや急なハンドル操作、スマート工場における装置故障など、様々な事故やトラブルの兆候(サイン)をビッグデータから高速かつ自動的に抽出するためのリアルタイム要因分析を実現できる。上述のように論文はKDD2019に採択されている[C1]。

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、富士通研究所、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、三菱重工工作機械、日立製作所などと共同研究を実施している。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 (スペースの都合上、研究代表者である櫻井の業績のみ掲載する)

[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. Takuya Maekawa, Naomi Yamashita, Yasushi Sakurai: ``How Well Can a User's Location Privacy Preferences be Determined Without Using GPS Location Data?'', IEEE Transactions on Emerging Topics in Computing, Vol. 5, Number 4, pp. 526-539, December 2017.
- J3. 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.
- J4. 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.
- J5. 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.
- 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: ``大規模疫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 37 件)

(国際会議論文 6 件)

- C1. Yasuko Matsubara, Yasushi Sakurai: ``Dynamic Modeling and Forecasting of Time-evolving Data Streams'', ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining (KDD), Anchorage, Alaska, August 2019 (Acceptance Rate: 170/1200, 14.2%).

- C2. Kouki Kawabata, Yasuko Matsubara, Yasushi Sakurai: StreamScope: Automatic Pattern Discovery over Data Streams. ACM SIGMOD Workshop on Exploiting Artificial Intelligence Techniques for Data Management (aiDM), pp. 5:1-5:8, June 2018.
- C3. 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.
- C4. 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%).
- C5. 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%).
- C6. 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.

(チュートリアル講演 2 件)

- 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.

(特別講演、基調講演、招待講演 20 件)

- S1. Yasushi Sakurai: "Mining and Forecasting of Big Time-series Data", International Workshop on Behavior analysis and Recognition for knowledge Discovery (PerCom Workshop, BiRD 2019), Kyoto, March 11, 2019 (Keynote Speech).
- S2. Yasushi Sakurai: "Smart Analytics for IoT Big Data", ACM SIGAPP International Conference on Management of Emergent Digital EcoSystems (MEDES 2018), Tokyo, September 26, 2018 (Keynote Speech).
- S3. Yasushi Sakurai: "Real-time Forecasting of IoT Big Data: Foundations and Challenges", the 11th Korea-Japan Database Workshop (KJDB), Daegu, South Korea, November 19, 2016 (Keynote Speech).
- S4. Yasushi Sakurai: "Mining Big Time-series Data on the Web", POSTEC Database Workshop, POSTEC (Pohang University of Science and Technology), Korea, May 30, 2016 (Invited Talk).

他 16 件

(国内ワークショップ、研究会 9 件)

- W1. 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).

W2. 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).

W3. 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).

W4. 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).

他 5 件

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 4 件)

(審査待ち特許)

P1. 松原 靖子, 櫻井 保志: ``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", 特願 2016-138075, 2016 年 7 月 12 日.

P2. 櫻井保志, 松原靖子, 入船恭彰, 山室 冨, 川畑光希, 浦 晃, 河東 孝, 梅田裕平: ``付与方法、付与プログラム及び付与装置", 特願 2019-036210, 2019 年 2 月 28 日 (熊本大学, 富士通研究所, 共同特許).

(PCT 国際出願)

P3. 櫻井保志, 松原靖子: ``予測装置、パラメータ集合生産方法及びプログラム", 国際出願 PCT/JP2017/025236, 2017 年 07 月 11 日, 国際公開 WO2018/012487A1, 2018 年 1 月 18 日.

(国外出願・アメリカ)

P4. Yasuko Matsubara, Yasushi Sakurai: ``FORECASTING DEVICE, PARAMETER SET PRODUCTION METHOD AND PROGRAM", 16/316946, July 11, 2017 (US patent).

〔その他〕

ホームページ

<http://www.dm.sanken.osaka-u.ac.jp/yasushi/index-j.html>

<https://www.dm.sanken.osaka-u.ac.jp/>

6 . 研究組織

(1)研究分担者

研究分担者氏名：田島敬史

ローマ字氏名：Keishi Tajima

所属研究機関名：京都大学

部局名：大学院情報学研究科

職名：教授

研究者番号 (8 桁)：60283876

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。