

研究終了報告書

「複合時系列イベントストリームに基づくリアルタイム将来予測と社会行動支援サービスの構築」

研究期間：2016年12月～2021年2月

研究者：松原 靖子

1. 研究のねらい

現在の高度に発達した情報化社会において、我々が取り扱うことのできるデータの量は飛躍的に増大している。このような状況下において、現在特に重要視されているのが、大規模時系列イベントデータストリームの高速解析技術である。時系列イベントデータは、車両・交通システムや通信・モバイル機器等における各種センサネットワーク、Google や Amazon に代表されるオンライン・ソーシャルメディア上における各ユーザのアクティビティ等、様々なフィールドで大量に発生し続けている。また、これらのデータは、現実世界のニュース、季節、社会活動等の複合的な要因に影響を受けながら、リアルタイムかつ動的に推移している。

本研究では、大規模かつ多様な時系列イベントストリームを対象とし、重要な情報を自動的にかつ複合的に解析・統合・管理することにより、将来発生するイベントをリアルタイムかつ高精度に予測し続けるための基盤技術を開発する。時系列イベントストリームのリアルタイム予測は、今後の知的社会の構築や国内産業の発展において、非常に重要な基礎技術である。過去に蓄積された膨大な時系列データに基づき、今後の動向予測を高速に行うことで、様々な社会サービスをリアルタイムに提供することができる。具体的な活用事例としては、製造業における設備監視や故障・不良品発生の実時間予測が挙げられる。同様に、交通ネットワークにおける車両走行センサデータのリアルタイム解析により、渋滞緩和や交通事故防止が期待できる。デジタルヘルスケア分野においては、IoT デバイスからの生体データを解析することにより、生活習慣病に代表されるような慢性疾患の予防、あるいは工場の精密作業や長距離トラック等におけるストレス軽減と事故防止につなげることが可能となる。さらに Web 上の人々の動向を瞬時に予測することができれば、人々や企業の活動、その関係性を把握し、市場トレンドや行政ニーズの把握、災害時の被災者支援、さらには犯罪による緊急情報の提示等が実現可能となる。

本研究で開発するシステムは、大規模な実証実験において有効性と実用性を確認するとともに、様々な国内有力企業と連携し、時系列イベントストリームのリアルタイム予測技術の製品への導入やサービス事業化、技術移転の取り組みを実施する。

2. 研究成果

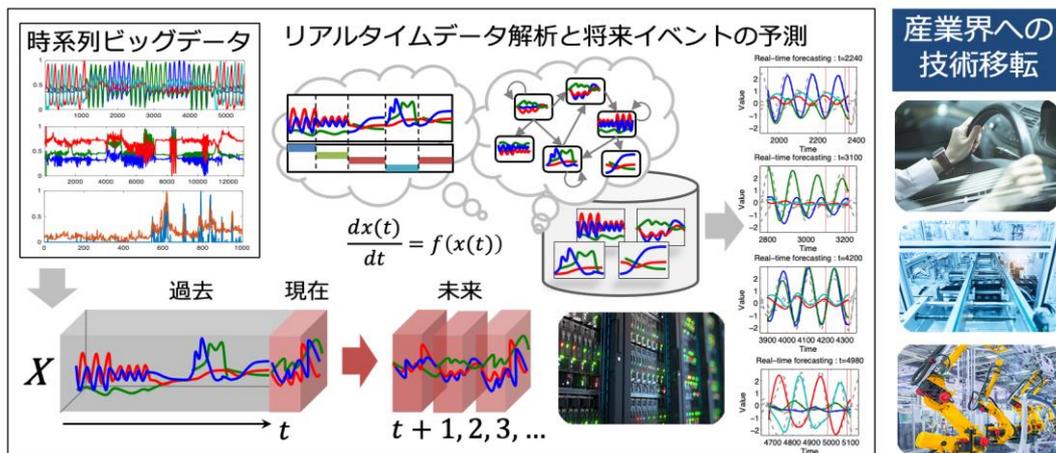
(1) 概要

本研究では、IoT/センサデバイス、Web 上のアクティビティ等、大規模かつ多様な時系列ビッグデータストリームを対象とし、重要な情報を自動的にかつ高速に解析・予測するための基盤技術を開発した。主な独自技術として、非線形モデリングに基づくリアルタイム要因分析・予測技術を考案し、国際的な場で極めて高い評価を得ている。具体的には、時系列ビッグデータストリームに基づく動的モデル学習と将来予測技術 (RegimeCast @ KDD2016) を高度に融

合することにより、複雑な時系列データの動的要因分析と将来予測をリアルタイムかつ高精度に行う新技術を開発した。2019 年に発表した最新の要因分析・予測技術 (OrbitMap @ KDD2019) は、深層学習を含む現存する最先端の予測手法の中で現時点において世界最高の予測精度と計算速度を示している。開発技術は、重要なパターンや動的な前後関係を非線形モデルに基づき高速学習することにより、複雑な時系列データのダイナミクスの要因分析を行い、リアルタイムかつ高精度に将来イベントの予測を行う。

学術貢献としては、データマイニング分野の最難関のトップ国際雑誌、国際会議において数多くの基盤技術を発表している (TWEB2017, KDD2019, CIKM2019, ICDM2019, KDD2020 等)。また KDD2017 においては、「Smart Analytics for Big Time-series Data」という題目で、本研究に関する 3 時間のチュートリアル講演を実施した。さらに研究代表者は 2018 年度において、情報処理学会と ACM により新設された国際的な表彰として、IPSJ/ACM Award for Early Career Contributions to Global Research を、受賞者第一号として受賞した。これは情報学の分野において優れた成果をあげ、今後も科学技術の発展への貢献と国際的な活躍が期待される若手研究者に贈られる賞である。本事業により、日本人の若手研究者として初めてチューリング賞授賞式 (ACM Awards Banquet) に招待された。また、電気通信普及財団第 36 回テレコムシステム技術賞 (入賞)、及び、2020 年度マイクロソフト情報学研究賞も受賞した。

本研究では、学術面で世界最高レベルの研究成果を出すとともに、国内有力企業と連携し、社会実装に向けた活動を積極的に実施した。トヨタ自動車との共同研究においては、特徴自動抽出技術 (SIGMOD2014) を発展させ、大規模な車両走行センサ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 (ICDM2019)。その他にも、開発技術は国内外の様々な企業から注目され、本研究をベースとした産学共同研究を複数件継続的に実施している。



(さきがけ研究) 時系列ビッグデータのリアルタイム解析と将来予測に基づく社会活動支援

図1 本研究の概念図

(2) 詳細

本研究では、時系列ビッグデータ解析と将来予測に関する基礎研究を行い、他に類を見ない独創的な技術開発を実施した。主な研究成果として、非線形動的モデリングに基づくリアル

タイム要因分析・予測技術を開発した。また、社会実装にむけた取り組みとして様々な企業との共同研究を実施した。以下において、開発技術の取り組みについて記載する。

時系列ビッグデータのリアルタイム要因分析・予測技術の開発

本研究では、研究代表者がこれまでに開発している時系列ビッグデータストリームに基づく動的モデル学習と将来予測技術 RegimeCast (KDD2016 にて発表) を発展・拡張し、より実用的かつ高性能な解析技術として、リアルタイム要因分析・予測技術 OrbitMap (KDD2019 にて発表、阪大プレスリリース発表¹⁾) を開発した。本技術は、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リアルタイムに行う。

世界におけるビッグデータ・AI 関連の技術については、画像や映像処理、自然言語処理等の分野を中心に、主に深層学習を用いたソフトウェアの開発が活発であり、急速に応用範囲が拡大している。その一方で、深層学習は非常に強力ではあるものの、ブラックボックスモデルの使用による学習結果の説明や理解が困難であること、モデル構造やパラメータの精密なチューニングを要すること、大量の教師データと多大な計算時間を必要とするため、リアルタイムに学習ができない、刻々と変化する環境の中で動作を柔軟に適應することができない、時系列データのような動的情報の処理に限界があり高い精度で予測することが難しい等、現時点では数多くの技術上、実用上の課題が存在する。

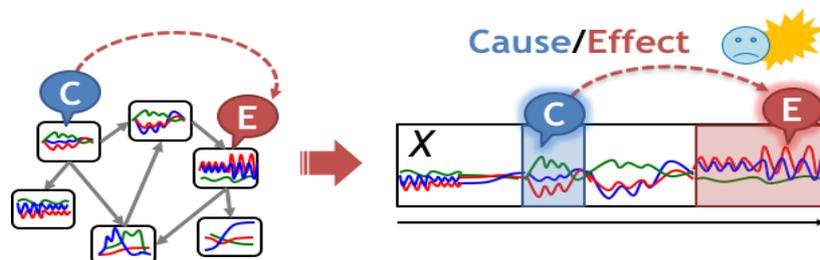


図 2 時系列データストリームにおける動的関係(要因-結果)の解析の様子

本研究で開発した時系列ビッグデータに基づくリアルタイム要因分析・予測技術 OrbitMap の概要を図 2 に示す。OrbitMap は、既存の深層学習に基づく技術では実現できなかった高い精度の予測や最適化、リアルタイム学習と情報出力、要因分析や結果説明を可能とする世界最高水準の革新的なデータ解析手法である。より具体的には、提案手法は主に IoT ビッグデータ等を対象とし、時系列モデル間の因果関係(要因-結果関係)を捉え、事象の連鎖をモデル化することにより、高精度かつ高速に要因分析・将来予測を行う。深層学習を含む現存する最先端の予測手法の中で現時点において世界最高の予測精度と計算速度を示しており、例えば製造業データに対し、最新の深層学習による予測手法と比較し最大で、約 670,000 倍の高速化、約 10 倍の高精度化(予測誤差 88%減)を達成している。

¹ https://resou.osaka-u.ac.jp/ja/research/2019/20191120_1

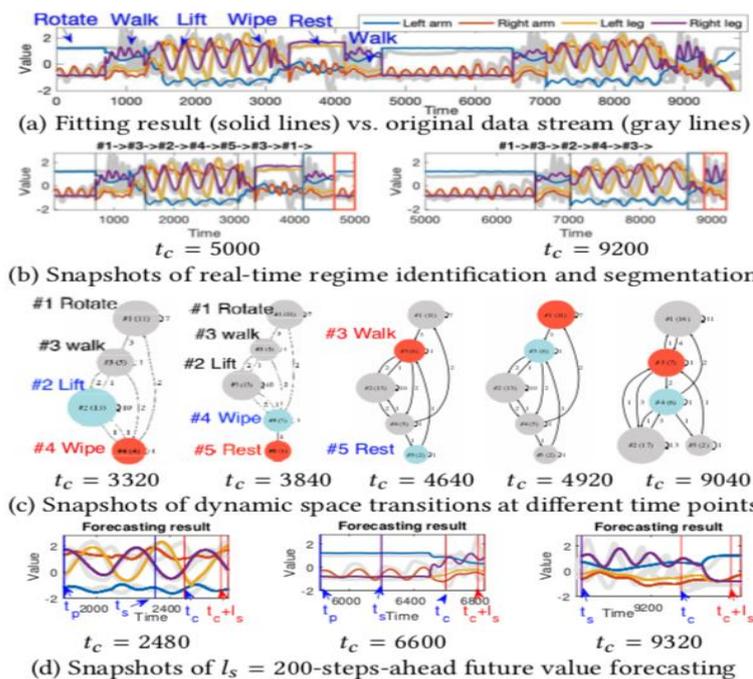


図3 モーションセンサデータを用いた要因分析技術の解析結果例

図3はOrbitMapを用いたリアルタイム要因分析の出力例である。ここでは、加速度センサデータを用いて解析している。より具体的には、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。より具体的には、作業者の行動の間のつながり(回転する→歩く→持ち上げるなど)をネットワークとして示している。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の様子を示している。ここでは、200単位時刻先(つまり、2秒先)の行動を予測している。現時刻 t_c において、時刻 t_c+l_s を予測している。ここで、 l_s は予測する長さを示す。本技術の特長として、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。

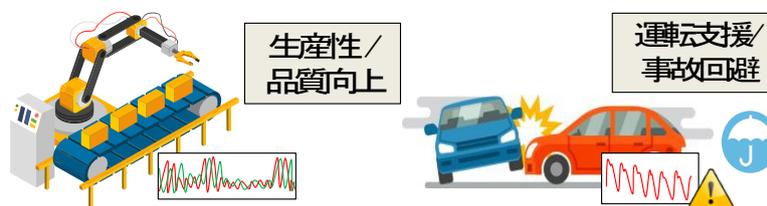


図4 モーションセンサデータを用いた要因分析技術の解析結果例

本研究による開発技術を発展・適用することで、様々な産業応用が期待できる。例えば、自動車走行における急なブレーキやハンドル操作、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をビッグデータから高速かつ自動的に抽出するための要因分

析をリアルタイムに行うことが可能となる(図 4)。

時系列ビッグデータ解析の社会実装に向けた取り組み

社会実装に関する取り組みについては、本研究期間中において、様々な企業と継続的に共同研究を実施し、社会・産業貢献として成果を挙げている。

例えば、トヨタ IT 開発センター(2019 年 4 月よりトヨタ自動車に合併)との共同研究においては、これまでの研究成果の一つである時系列センサデータ解析技術 AutoPlait (SIGMOD2014 にて発表)を発展させ、大規模な車両走行センサ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(IEEE ICDM 2019 にて発表)。図 5 は研究成果の例である。この車両走行のデータ集合には合計 31 の多次元シーケンスが含まれており、シーケンスの各次元は速度(青)、左右加速度(赤)、前後加速度(緑)を示している。図の上段は 31 のトリップシーケンスを運転状況に合わせて 3 つにグループ化し、その解析結果を地図上にプロットしたものである。図下段は提案手法が自動抽出した複数のセグメントパターンをグループ毎に示している。提案手法は、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う。

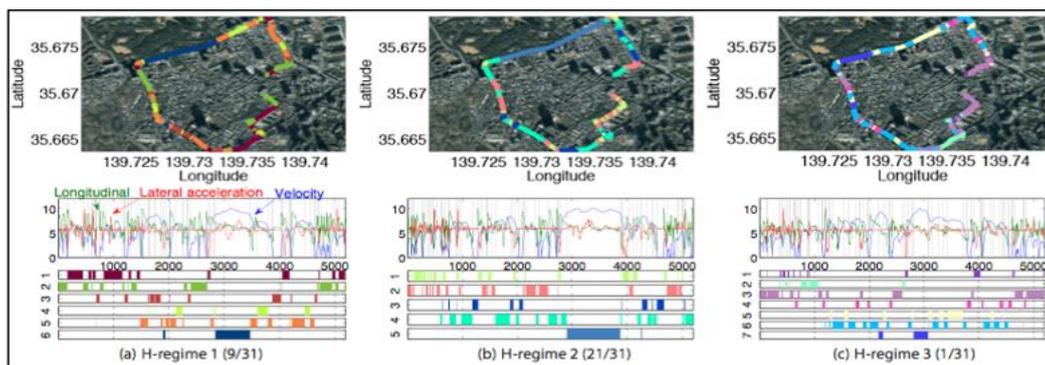


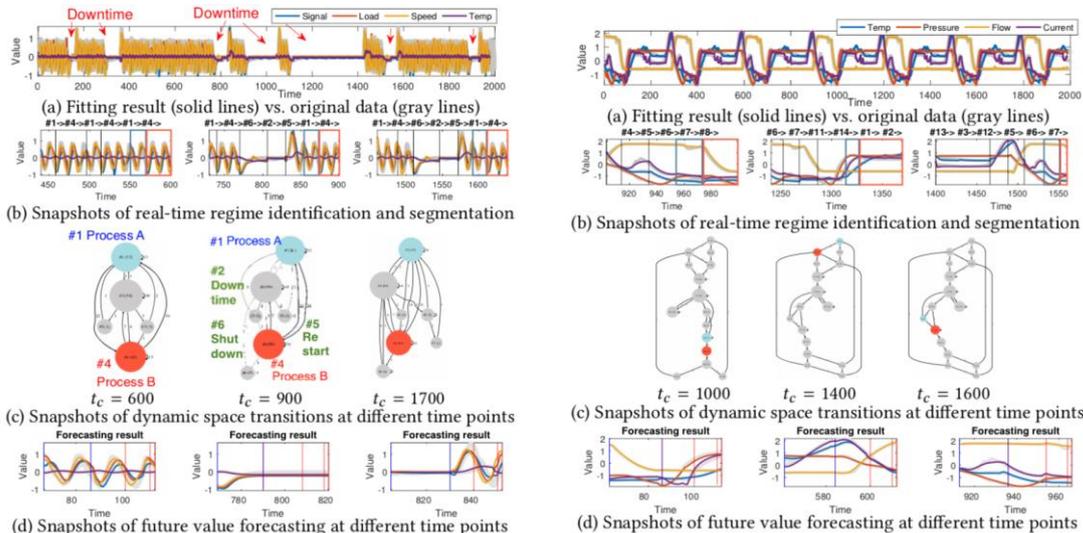
図 5 車両走行時系列ビッグデータにおける特徴自動抽出の例

また、本研究において開発した時系列ビッグデータ解析に関する独自技術は、その汎用性と有効性から、上記以外にも国内外の様々な企業から注目されている。例えば、スマート工場における設備監視の高度化を行うための研究開発等、研究代表者が開発した独自技術をベースとした継続的な共同研究を複数実施しており、開発しているリアルタイム予測技術が、産業界において非常に強い需要があることがわかる。これまでには、トヨタ自動車、富士通研究所、三菱重工エンジン&ターボチャージャ、三菱重工工作機械、ソニーセミコンダクタマニュファクチャリング、電通デジタルなど様々な企業の方々とスマート工場、車両走行データ解析、生体情報解析、デジタル広告などのテーマで実用化、社会実装に向けて共同研究を実施している。

3. 今後の展開

時系列ビッグデータ解析とリアルタイム将来予測技術の社会実装への取り組み

本研究では、時系列ビッグデータ解析と将来予測に関する基礎研究を行い、他に類を見ない独創的な要素技術を数多く考案しており、国際的な場できわめて高い評価を得ている。今後の展開として、これまでに開発した時系列ビッグデータ解析とリアルタイム将来予測に関する独自技術を発展させ、社会実装を実現するための取り組みを積極的に行っていく。本研究では既に、国内企業と連携しながら、現場における実証実験を開始している。例えば、スマート工場における設備監視の高度化を行うための研究開発等、研究代表者が開発した独自技術をベースとした共同研究を複数実施している。製造業をはじめとする産業界において、IoT を巡る環境が急速に変化し、生産性や品質の向上、スマート工場化に関する技術開発が急務となっている。例えば、スマート工場の実現には、設備・稼働状態などの様々なデータをリアルタイムに収集し、人の指示を介さず自律的に判断し、工程を最適化するような統合的な高度支援システムの構築が必要である。特に、実際の現場におけるスマート工場のデータには、各種センサからの連続値データのみならず、製品番号、設備稼働情報なども含まれるため、数多くの要素が含まれる複合ビッグデータをリアルタイムかつ動的に学習する必要がある。また、様々な事象の発生を事前に予測し、またその発生要因を自動検出することで、トラブル回避やコスト最小化等の行動最適化をするための技術が求められている。図は、本研究で開発した要因分析技術をベースに、実際の企業で運用されているスマート工場においてリアルタイムデータ解析を行なった例である。図に示す通り、開発した技術は、工場内で大量に生成され続けるセンサデータストリームに対し、現時刻における特徴や動的パターンを非線形モデルとして学習、表現し、将来予測を高精度に行うことに成功した。今後も、様々な分野の有力企業と連携することで、技術を実用化するための取り組みを進めていく。



(a) CNC 装置²におけるセンサ解析例 (b) CMP 装置³におけるセンサ解析例

図 6 スマート工場におけるリアルタイム要因分析と将来予測の様子 (KDD2019 にて発表)

² ソニーセミコンダクタマニュファクチャリングより提供

³ 三菱重工エンジン & ターボチャージャより提供

4. 自己評価

本研究では、IoT/センサデバイス、Web 上のアクティビティ等、大規模かつ多様な時系列ビッグデータストリームを対象とし、重要な情報を自動的かつリアルタイムに解析・予測するための基盤技術を開発した。学術貢献に関しては、データマイニング分野の最難関のトップ国際雑誌、国際会議において数多くの研究成果を発表し、これらの成果は国際的に高く評価されている (TWEB2017, KDD2019, CIKM2019, ICDM2019, KDD2020 等)。

特に、リアルタイム要因分析・予測技術 OrbitMap (KDD2019) は、世界最高の予測精度と計算速度を示しており、最新の深層学習と比較し最大で、約 670,000 倍の高速化、約 10 倍の高精度化 (予測誤差 88%減) を達成している。また KDD2017 においては、「Smart Analytics for Big Time-series Data」という題目で、本研究をベースとした、最新の時系列ビッグデータ解析に関する 3 時間のチュートリアル講演を実施した。

本研究では、上記以外にも、時系列ビッグデータに基づくリアルタイム解析に関する様々な基礎研究を実施した。例えば、非線形動的モデリングを用いた Web 情報の時間発展に関する研究については、研究代表者による非線形モデルに基づく時系列解析に関する研究 (KDD 2012 にて発表) をベースとし、リアルタイム処理を採り入れて研究を格段に発展させることで、トップ国際学術雑誌である ACM Transactions on the Web (TWEB) にて成果を発信している。ACM KDD 2012 および TWEB において発表した本研究成果はその後、情報拡散過程解析の流行を作り出しており、国際的にも高い評価を得ている。また、本研究成果に基づき、電気通信普及財団第 36 回テレコムシステム技術賞 (入賞) を受賞した。さらに、この取り組みは、現時点において既に 350 件以上の論文に引用されている (現時点での Google Scholar における、KDD2012 および TWEB2017 の引用件数の合計値)。

また、これまでの時系列ビッグデータ解析を基礎として発展させた数々の取り組みによって、研究代表者は 2018 年度において、情報処理学会と ACM により新設された国際的な表彰として、IPSJ/ACM Award for Early Career Contributions to Global Research を、受賞者第一号として受賞した。本事業により、日本人の若手研究者として初めてチューリング賞授賞式 (ACM Awards Banquet) に招待された。2020 年度においては、情報処理学会マイクロソフト情報学研究賞も受賞している。

さらに本研究では、学術的貢献のみならず、情報技術に基づいた社会変革の時代に対応し、新たな社会システムのデザインを可能にするための情報基盤技術の創出を実現するために、社会実装に向けた活動として、産学共同研究を積極的に実施した。本研究において開発した要素技術は、その汎用性と有効性から、国内外の様々な企業から注目され、開発技術をベースとした共同研究を複数開始している。特に、製造業においてはスマート工場関連プロジェクトにおいて、設備故障の予測、不良品発生 of 予測に関するソフトウェアを開発しており、現在、事業導入が検討されている。

最後に、本さがけ研究のプロジェクト開始当初より現在まで、研究総括の黒橋禎夫先生、そして、領域アドバイザーの先生方、JST の皆様から様々な確かつ重要な指導や細やかな支援をいただいたことにより、多くの研究成果を得ることができました。深く御礼申し上げます。

5. 主な研究成果リスト

- (1) 代表的な論文 (原著論文) 発表 (上限3件)

研究期間累積件数:26件

1. Yasuko Matsubara, Yasushi Sakurai, “Dynamic Modeling and Forecasting of Time-evolving Data Streams”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD2019**), Anchorage, Alaska, August 2019, Acceptance Rate:170/1200, 14.2%.

本論文では、IoT、 センサストリーム等の様々な時系列パターンから構成されるデータストリームが与えられたとき、その中に含まれる動的パターンを高速にモデル学習し、また、将来予測をし続けるための手法である OrbitMap について述べる。OrbitMap は、時系列データストリーム内の複雑な時間発展を非線形動的システムと動的空間遷移を用いて表現し、リアルタイムにモデル学習および将来予測を行う。提案手法は (a) 大規模データストリーム内の重要な動的パターンを柔軟に表現し、(b) 多種多様な実データに対しリアルタイム将来予測を行うことができる。また、提案手法は (c) データストリームの長さに依存せず、各時刻において、最適なモデル学習を高速に行う。実データを用いた実験では、OrbitMap が様々な時系列データストリームに対し、重要な動的パターンを発見し、長期的な予測を行うことを確認し、最新の既存手法と比較し大幅な精度、性能向上を達成していることを明らかにした。また、工場施設内において生成される IoT データストリームに対し本技術を適用し、提案手法の実用性と有効性を示した。

2. Takato Honda, Yasuko Matsubara, Ryo Neyama, Mutsumi Abe, Yasushi Sakurai, “Multi-Aspect Mining of Complex Sensor Sequences”, IEEE International Conference on Data Mining (**ICDM2019**), Beijing, China, Acceptance Rate: 95/1046, 9.08%.

本論文では、車両走行データのための自動パターン検出手法である CubeMarker について述べる。CubeMarker は、位置情報をもとに様々な車両走行センサデータが与えられたときに、おのこの道路や場所における車両走行の特徴を抽出し、それらの情報を統計的に要約、表現する。すなわち、走行データに基づく高度な道路地図情報を提供する。具体的に提案手法は、(a)車両走行データをテンソルとして表現した後、そこから複数の部分シーケンスに共通する主要な走行パターンを抽出する。(b)その際の計算量は入力データのサイズに対して線形である。さらに、最も重要な点として、(c)提案手法はパラメータに依存しない。すなわち、事前情報の付与またはパラメータのチューニングを行うことなく、大規模車両走行データの特徴抽出とパターン検出を自動で行うことができる。実データを用いた実験では CubeMarker が様々な車両走行データの中から主要パターンや外れ値シーケンスを効果的かつ効率的に検出することを確認した。

3. 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**), Volume 11 Issue 2, May 2017.

ブログやソーシャルネットワークサービスをはじめとするインターネットメディアの普及により、オンライン上での情報の伝播速度が増している。あるニュースは緩やかに拡散し、ゆっくりと減衰していく。一方で、別のニュースは急激に伝わると同時に、すぐに消えていく。本論文では、ソーシャルメディア上における情報拡散と減衰を表現するモデルである SpikeM について述べる。SpikeM は有限のノード(ユーザ)を仮定するとともに、情報の拡散過程においてパワー則に基づく減衰パターンを有し、また周期性を併せ持つ。情報拡散過程のパターンが与えられたとき、提案手法は (a) ユーザの数とネットワーク構造、(b) あるイベントの発生時刻と

そのイベントの質(もしくは注目度)、(c) イベント直後に影響を受けるノード(すなわち、ユーザ)の数をパラメータとして用い、それらのダイナミクスを柔軟に表現する単一の非線形モデルを推定する。モデルの利用により、拡散する情報の質やネットワーク規模のような有用な情報を推定することが可能となり、さらに外れ値検出や時系列予測等の実用的なタスクを処理することができる。

(2) 特許出願 (上限 3 件)

研究期間累積件数: 9 件 (特許公開前のもも含む)

1	発 明 者	松原靖子、櫻井 保志
	発 明 の 名 称	予測装置、予測方法およびプログラム
	出 願 人	大阪大学
	出 願 日	2019/8/1、2020/7/30
	出 願 番 号	特願 2019-142295、PCT/JP2020/029178
	概 要	本発明は、データストリーム中の一部のカレントウィンドウを用いて将来発生する可能性のあるイベント情報をリアルタイムで予測する技術である。本発明では、レジーム間の推移情報を導入することで、即答性、適応性、及び精度の高い予測処理を可能にする。
2	発 明 者	松原靖子、櫻井 保志
	発 明 の 名 称	予測装置、パラメータ集合生産方法及びプログラム
	出 願 人	熊本大学
	出 願 日	2017/7/11
	出 願 番 号	PCT/JP2017/025236
	概 要	本発明では、大規模な時系列データストリームを用いて、高精度に長期予測を実現する装置を提案する。本発明で用いる数理モデルは、非線形要素を含み、データストリームの非線形を表現する。また、レジームシフトを導入することにより、高精度な長期予測を実現する。
3	発 明 者	本田崇人、松原靖子、川畑光希、櫻井保志
	発 明 の 名 称	時系列解析に基づくイベント予測装置、イベント予測方法およびプログラム
	出 願 人	大阪大学
	出 願 日	2020/1/22
	出 願 番 号	特願 2020-8388
	概 要	本発明は、時系列センサデータに基づくイベント予測技術に関する技術である。本発明では、時系列ビッグデータに対し、多角的な動的パターンを自動抽出と特徴量の要約を行うことにより、将来のイベント発生を長期的かつ高精度に予測するための装置を提案する。

(3) その他の成果 (主要な学会発表、受賞、著作物、プレスリリース等) (上限 5 つまで記載)

- 1) 松原靖子, 2020 年度マイクロソフト情報学研究賞 (2021 年 3 月 18 日受賞)
- 2) Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, Christos Faloutsos, 電気通信普及財団 第 36 回テレコムシステム技術賞, “Nonlinear Dynamics of Information Diffusion in Social Networks” (2021 年 3 月 26 日受賞)

- 3) Koki Kawabata, Yasuko Matsubara, Takato Honda, Yasushi Sakurai: “Non-Linear Mining of Social Activities in Tensor Streams”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (**KDD2020**), pp. 2093–2102, Virtual Event, California, August 2020, Acceptance Rate: 216/1279, 16.8%.
- 4) Yasuko Matsubara, IPSJ/ACM Award for Early Career Contributions to Global Research, “Analysis of Time-series Data and Nonlinear Dynamic Systems” (2018年6月6日受賞)
- 5) Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos, “Smart Analytics for Big Time-series Data”, ACM SIGKDD 2017 Tutorial, (**KDD2017**).