

# IoT ビッグデータのための非線形解析システムの研究開発 (162110003)

## Non-linear Mining and Forecasting of IoT Big Data

### 研究代表者

松原 靖子 大阪大学 産業科学研究所  
Yasuko Matsubara ISIR, Osaka University

研究期間 平成 28 年度～平成 30 年度

### 概要

IoT ビッグデータは、交通システムや環境測定等における各種センサネットワーク等、様々な分野で大量に発生し続けている。本研究では、IoT ビッグデータを対象とし、様々な現象、活動の時間的な推移を分析・モデル化することにより、重要なパターンの発見、将来のイベントの予測を効果的、効率的に行なうシステムを開発した。具体的には、本研究のフェーズ I では、時系列ビッグデータのためのリアルタイム予測技術を開発し、センサデータのみならず、環境データや Web 情報等、多様なデータに対して適用し、時系列モデルやアルゴリズムの有効性を検証した。また、本研究のフェーズ II においては、IoT ビッグデータの統合的なモデル化とモデルパラメータ推定手法の研究開発を行ない、さらに、IoT 非線形予測技術の開発、リアルタイム要因分析技術と IoT スマートサービスの開発を行なった。開発したシステムは、大規模実データを用いた評価実験を行ない、技術の有用性を確認した。

### 1. まえがき

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。そ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、大規模時系列データの解析技術である。時系列データは、Google や Twitter に代表される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、交通システムや環境測定等における各種センサネットワーク等、様々な分野で大量に発生し続けている。本研究は、これらの時系列ビッグデータを対象とし、様々な現象、活動の時間的な推移を分析・モデル化することにより、重要なパターンの発見、将来のイベントの予測を効果的、効率的に行なうシステムの開発を目的とする。本研究のフェーズ I では、時系列ビッグデータのためのリアルタイム予測技術を開発し、センサデータのみならず、環境データや Web 情報等、多様なデータに対して適用し、時系列モデルやアルゴリズムの有効性を検証する。続いて、本研究のフェーズ II では IoT ビッグデータの統合的なモデル化とモデルパラメータ推定手法の研究開発をする。より具体的には、IoT 非線形予測技術の開発、リアルタイム要因分析技術と IoT スマートサービスの開発を行う。本研究課題ではさらに、IoT スマートサービスの大規模実証実験および産業界への技術移転を目的とし、複数の国内企業と連携し、センサデータストリーム等、実社会における IoT ビッグデータの解析に取り組み、技術の実用化に取り組む。

### 2. 研究開発内容及び成果

本研究では、平成 28 年度において、リアルタイム予測技術である RegimeCast を開発した。また、平成 28 年度において得られた知見や基盤技術を元に、平成 29 年度において、RegimeCast を改良および高度化し、IoT ビッグデータを統合的にモデル化、推定し、動的な要因分析を行うための基盤技術を実装した。RegimeCast は大規模なデータストリームの中から重要な特徴や潜在的なトレンドを発見し、長期的かつ継続的に将来の非線形的なイベント予測を行うことを可能とする高度な解析技術である。

一般的に、時系列データの中には複数の時系列パターンが含まれており、それらのパターンは不規則的に出現および推移していく。そこで RegimeCast は、(a) 時系列ストリームデータの中から事前知識なしに現時刻の状況に応じて多様なパターンをモデル学習し、さらに、(b) ストリーム内の突発的な変化点を検出し、現時刻の時系列パターン

を瞬時に探索、認識することで、将来のイベントを柔軟に予測する。本研究における開発技術は、その有用性と実用性により、国内外の学術コミュニティにおいて非常に高い評価を受けている (KDD2016、Tutorial@KDD2017、WebDB Forum 2017 最優秀論文賞受賞、KDD2019)。さらに、本技術の国内企業との産学連携と社会への技術移転に向け、国内外において特許化を行なった (国内：特許出願済、海外：PCT、U.S. Patent 出願済)。

図 1 は RegimeCast に基づいて開発したリアルタイム予測システムの出力結果である。実験データとしてモーションセンサデータを用いており、図 1(a)の上段はオリジナルのデータストリーム、図 1(a)下段は 100 から 120 時刻先をリアルタイムかつ継続的に予測した結果である。事前学習なしに様々なパターンを高精度に予測していることが分かる。図 1(b-1)から(b-4)まではリアルタイムに時系列データをパターン分けし、オンライン学習した結果である。4 つのパターンは非線形解析に基づき少ない誤差でモデル化し、これら 4 つのモデルパターンを使い分けながらリアルタイムに予測値を推定する。

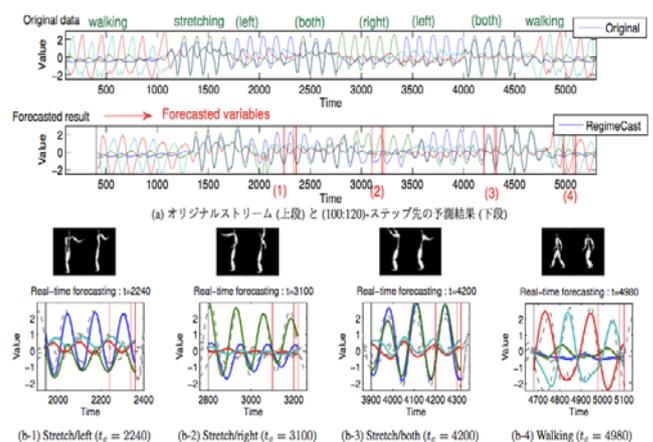


図 1. リアルタイム予測システムの実験結果例。

また、平成 30 年度の取り組みとして、これまでに開発をおこなった IoT 非線形解析技術をさらに発展させ、技術の産業化、社会への技術移転を行うための取り組みを行なった。より高度かつ複雑なデータに対し解析を行うための技術を検討していくため、既存の時系列データのみならず、新たな実データとして、IoT、モーションセンサ、視線デ

ータ、スマート工場関連等のセンサデータストリーム等、様々なビッグデータを生成、収集し、大規模な評価実験を行なった。さらに、共同研究を行なっている企業と連携しながら、現場における大規模実証実験のための検討、データ収集を行なった。本研究では、多項目からなる基本項目データに加え、IoT デバイスから計測値を取得し、それらを統合させてデータ管理を行う。例えば、車両走行データの場合、車種やユーザに関する情報は基本項目データであり、速度や前後加速度、ヨーレートなどは IoT デバイスから計測値である。しかし、製造部門で扱う IoT データは単なる多次元シーケンスではなく、基本項目データも含む複雑なデータである。そこでこれまで研究代表者が開発した技術である (a) 非線形時系列解析、(b) 大規模テンソル分析、(c) 特徴自動抽出、全てを融合させ、IoT データを統合的にモデル化し、全ての情報の関係性を自動的に抽出するとともに、急激な特徴の変化にも対応できる予測技術を開発した (図 2)。様々な分野の有力企業と連携し、それぞれの応用ドメインでの技術の実証実験を行い、技術を実用化するための共同研究を実施した。また、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。本研究で開発した技術は、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、現時点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。

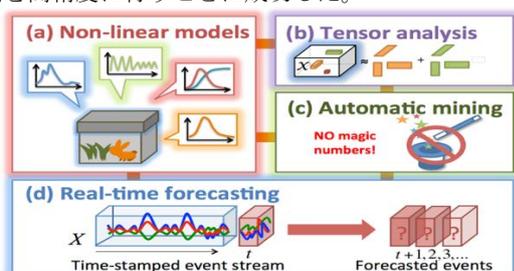


図 2. これまでの研究成果の融合、深化、発展。

### 3. 今後の研究開発成果の展開及び波及効果創出への取組

IoT ビッグデータを活用し、今後の動向予測や要因分析、リアルタイム情報提供を行うことにより、社会を豊かにするサービスを提供することができる。交通ネットワークにおける車両走行センサデータを解析・予測すれば、渋滞緩和や交通事故防止が期待でき、環境データを解析することにより災害時の被災者支援が可能となる。また、デジタルヘルスケア分野において IoT デバイスからのデータ (体温、血圧、心電など) を取得、解析することにより、病気予防に役立てることができ、社会的意義は大きい。

### 4. むすび

我が国において、自動車などの組込・制御機器分野で強みがある点に鑑みれば、その製品の付加価値を高めるために IoT ビッグデータ解析技術が極めて重要になると考えられる。アメリカの後追いではなく、日本独自の研究を推進するべきであり、研究開発者らは産学連携を重視し、製造業の変革に向けた取り組みを行っている。例えば、本研究で開発した動的要因分析・予測技術をベースに、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ている。本開発技術は汎用性が非常に高く、様々な場面において重要なパターンを自動的に抽出、解析、予測することができる。現在も複数企業との共同研究を通して本技術の実用化に向けて継続的に開発を進めている。様々な分野の企業との共同研究を進展させた技術開発、その大規模な連携こそが未来を大きく変える力とな

る。今後も有力な企業と協調しながら、スマート工場や運転支援サービスなど、IoT ビッグデータ解析に関する世界のデファクトスタンダードとも呼ぶべき技術を開発することを念頭に、先駆的な技術開発に取り組んでいく。

#### 【査読付き論文リスト】

- [1] 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、2017, Vol.10 No.3, pp.1-5, 2017, 10月5日。
- [2] 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, 10.1145/3057741, May 2017.
- [3] 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, 12月。

#### 【査読付き口頭発表論文リスト】

- [1] 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 4-8, 2019.
- [2] Yasushi Sakurai, Yasuko Matsubara, Christos Faloutsos, Smart Analytics for Big Time-series Data, ACM SIGKDD 2017 Tutorial, August 13, 2017.
- [3] 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 (KDD2016), San Francisco, California, August 13-17, 2016.

#### 【招待講演リスト】

- [1] 松原靖子、時系列ビッグデータ解析の新たな展開、2018年度人工知能学会全国大会(第32回)6月6日、2018。
- [2] 松原靖子、時系列ビッグデータの特徴自動抽出とリアルタイム将来予測、第9回ステアラボ人工知能セミナー、2017年4月21日、4月21日、2017。
- [3] 松原靖子、時系列ビッグデータのリアルタイム将来予測と知的社会サービスへの展開、電子情報通信学会総合大会、スマート無線(SR)研究会、スマート無線における機械学習応用、2017年3月22日。

#### 【申請特許リスト】

- [1] Yasuko Matsubara, Yasushi Sakurai, 予測装置、パラメータ集合生産方法及びプログラム、PCT 国際特許出願、US patent (JST 知財活用支援事業)、2018。
- [2] 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、出願日:2019年2月28日。
- [3] 松原靖子、櫻井保志、予測装置、パラメータ集合生産方法及びプログラム、出願番号: 特願 2016-138075、出願日: 2016年7月12日。

#### 【受賞リスト】

- [1] Yasuko Matsubara, Research on Analysis of Time-series Data and Nonlinear Dynamic Systems IPSJ/ACM Award for Early Career Contributions to Global Research (2018年6月6日)。
- [2] 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、第10回 Web とデータベースに関するフォーラム、最優秀論文賞 (2017年9月19日)。
- [3] 松原靖子、2016年度 情報処理学会 山下記念研究賞 (2017年3月16日)

#### 【本研究開発課題を掲載したウェブページ】

- [1] <http://www.dm.sanken.osaka-u.ac.jp/~yasuko/>