

IoT ビッグデータのためのリアルタイム AI 技術



特集1
ハイテク推進
セミナー

大阪大学 産業科学研究所 産業科学AIセンター
センター長/教授 櫻井保志氏

1. はじめに

データ駆動型社会においてはIT、IoT などにより様々な端末やセンサから情報収集することになるが、Beyond5Gに関連して充実した通信環境、そして次世代IoT技術の進化により、データ量はこれまで以上に爆発的に増大する。ネットワークから大量に流れてくるビッグデータ、すなわちデータストリームには、刻々と変化する環境や突発的な外的要因の影響などにより、時間とともに変わり、時には急激に変化する時系列情報が含まれている。実社会において多種多様な機器やデバイスから発生するIoTデータストリームには、設備や機器、デバイスに共通する特徴を含む場合もあれば、個体差を示すこともあり、さらに様々な状況において発生する事象間の関係性や因果関係なども含んでいる。そのため、データストリーム解析においてはモデルを固定すること無く、時系列パターンの急激な変動や設備個別の特徴や傾向の変化を高速に検出し、モデル学習やモデル更新をリアルタイムに処理することが重要であり、このような新しい課題に対応することが、製造業をはじめとするIoT分野において重要となる。

本稿では、著者らがリアルタイムAI技術と呼んでいる、IoTビッグデータを高速に学習し、予測、要因分析、トラブル予知に関する情報をリアルタイムに提供する新たな技術基盤について述べる。その中でも特に3種類の要素技術、(1)リアルタイム予測と動的要因分析、(2)リアルタイム時系列テンソル解析、(3)リアルタイム特徴自動抽出、について概説する。また、製造業の生産ライン支援や自動車運行管理など、DXに関連する取り組みを実例とともに紹介する。

2. リアルタイム予測と動的要因分析

図1と図2はリアルタイム将来予測・動的要因

分析技術[1][2]の概要を示している。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から、リアルタイムに特徴や潜在的なトレンド(レジーム)を検出し、各レジーム間の動的な関係性を見つけることにより、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。

非線形動的システムに基づくリアルタイム学習：実世界における時系列データストリームは、様々な時系列パターンから構成され、外的要因等によって突発的に変化していく。本技術はデータストリームの最新時刻の潜在的トレンドや時系列パターンを動的に把握、モデル化し、将来値を継続的に予測し続ける。特に、様々な時系列ダイナミクスを非線形動的システムによって表現するとともに、時系列トレンドの急激な変化を変化点としてリアルタイムに検出し、モデルパラメータを瞬時に切り替え、柔軟に予測を行う。アルゴリズムとしては、まずデータストリームから学習した様々な非線形方程式のモデルをデータベース(モデルDB)に格納する。予測処理においては、現在の時系列パターンに合うモデル



講師 櫻井 保志 氏

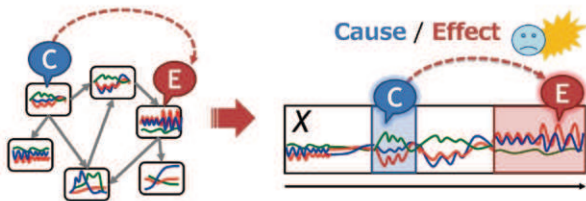


図1 リアルタイム将来予測・動的要因分析技術

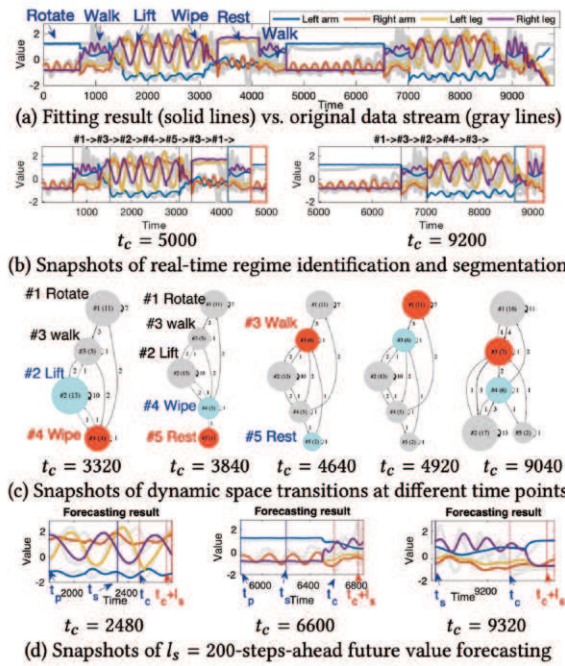


図2 予測と要因分析の結果

をDBから探索し、探索したモデルを用いて予測値を推定する。

動的要因分析：一般に、大規模な時系列データストリームは、自然現象や人々の社会活動、さらには様々な設備の動作状況等、様々な事象を表現している。本技術は、時系列ビッグ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、要因分析と動的空間遷移ネットワークを用いることにより、予測精度を向上させている。図2は本技術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、工場における作業員の両手足4箇所に加速度センサを設置し、100Hzで加速度を計測している。図2(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2(b)は、各時刻におけるリアルタイムパターン検出とモデル生成、図2(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、作業員の行動の間につながり（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2(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。本技術によってリアルタイム要因分析を実現することができ、例えば、スマート工場における装置故障、自動車走行における急ブレーキや急なハンドル操作など、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ことが可能となる。

3. リアルタイム時系列テンソル解析

リアルタイム時系列テンソル解析技術[3]は、リアルタイム予測技術[1][2]をベースに、テンソル解析技術として発展させることによって開発された基礎技術であり、図3は、複合データストリームの時系列テンソル解析、特にリアルタイム将来予測の様子を示している。本技術は、現時刻 t における動的パターン（図3青色箇所）を解析することにより、将来発生するイベント（図3赤色箇所）をリアルタイムに予測し続ける。ここでの例は、オンライン活動データを用いているが、本技術は、オンライン活動データにとどまらず、製造業における生産工程の設備データなど、様々な複合時系列ビッグデータストリームのリアルタイム解析・予測に適応することが可能である。

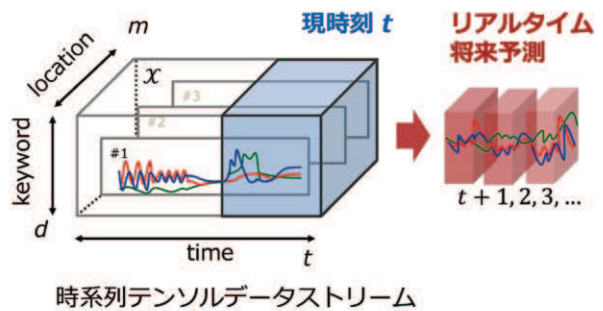


図3 リアルタイム時系列テンソル解析技術

図4は、Web上におけるキーワードアクセス件数推移データに対し本技術を適応した例を示してい

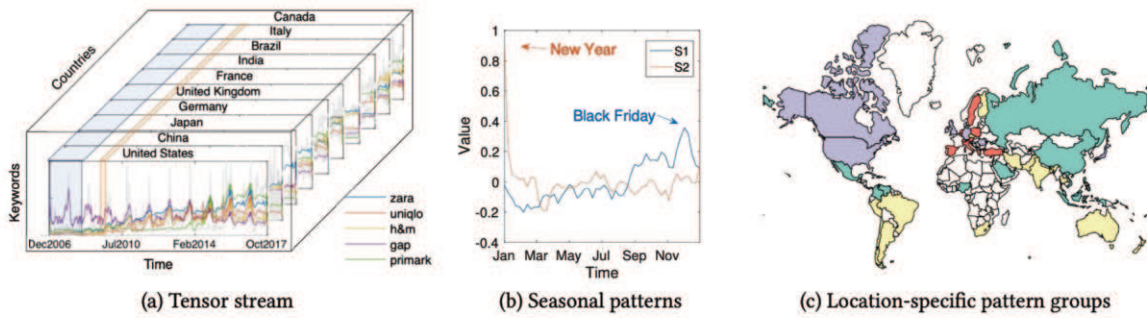


図4 テンソルデータストリームの解析結果

る。時間、地域、キーワードのように複数の属性を持つテンソルデータストリームが与えられたとき、最新の観測データ (図 4 (a) 青) を監視しながら潜在的なトレンドを発見し、適応的にモデルを変化させながら長期先のデータ (図 4 (a) 赤) を予測し続ける。このとき、図 4 (b) のように各地域で共通する周期 (季節) パターンを抽出し、それらに基づき、図 4 (c) に示すように類似パターンを有する地域のグループ化を行う。このように本技術は、非線形微分方程式に基づき、非線形性を有する時系列ダイナミクス、長期トレンド、周期性を同一のモデル空間で表現し、そして時系列テンソルの内部において類似した潜在トレンドを持つ属性データのグループ化を自動的かつ効率的に行う。

4. リアルタイム特徴自動抽出

リアルタイム特徴自動抽出技術 [4][5] は、大規模時系列データストリームに含まれる典型的なパターンを発見するための技術であり、図 5 は技術の概要を示している。大規模データストリームの解析には高度なマイニングアルゴリズムが求められるが、複雑な演算による計算コストに加えて高度なパラメータチューニング等による時間的・人的コストが高

くなり、実用化の際にはそれらが大きなボトルネックとなる。本技術では符号化理論に基づくモデル評価基準を応用し、解析データに関する事前知識を必要とせず、データの要約情報 (時系列パターンの種類・変化点) を自動的に抽出する。そして、与えられたデータストリームをリアルタイムに解析し、時系列パターンの種類と変化点を発見し、それらの特徴をモデルパラメータとして表現する。

図 5 は本技術を用いたリアルタイム解析の出力例である。図 5 上段はヒトの両手足 4 箇所に加速度センサを設置し、歩行と 3 種類 (両腕、右腕、左腕) のストレッチで構成される合計 4 種類の動作を捉えたデータストリームを示す。図 5 下段の番号は、本技術が検出したパターンの ID を示し、長方形の両端は各パターンの開始点・終了点を示す。本技術はセンサデータをリアルタイムに監視し、歩行、両腕のストレッチ、と次々にパターンの変化点を検出する。このとき、モデル評価基準を用いて新たなモデル生成の必要性を自律的に判断することにより、パターンの種類 (モデルの数) を自動的に決定する。図 5 に示すように、本技術は 8 つのパターン変化点を検出し、それらを 4 つの動作へと分類することに成功している。

また本技術は階層構造を有する確率モデルを用い

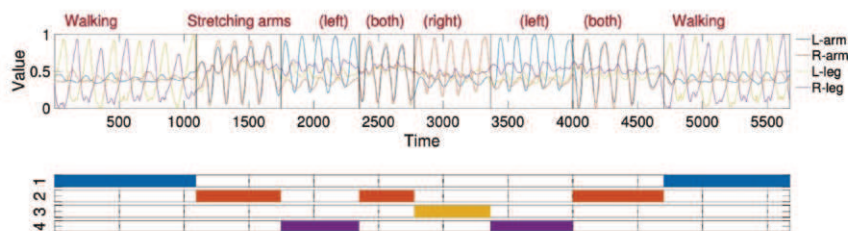


図5 モーションキャプチャデータからのリアルタイム特徴自動抽出

て過去に検出したパターンの特徴を表現し、新たなデータが観測されるたびに類似パターンの検索とモデルパラメータの推定を逐次的に行う。このとき、最後に観測したパターンのデータと過去に推定したモデルパラメータのみを保持するため、一度に大量のデータを処理する必要がなく効率的にデータストリームを処理することができる。

5. 社会実装と産業への展開

5.1 多角的テンソル特徴抽出技術の車両走行データ解析への応用

著者らの研究グループではトヨタ自動車と連携し、多角的テンソル特徴抽出技術 [6] を共同開発した。本技術はデータに含まれる重要なパターンの時間遷移やユーザー特性を自動的に抽出する。センサ、ユーザ、時間など複数ドメインで構成される時系列テンソルデータから各ドメインにまたがる複雑な特徴を時系列パターンとして検出し、各パターンをモデルパラメータとして要約する。

多角の特徴抽出：実社会において収集されるビッグデータは、センサ、デバイス、ユーザ、時間など複数のドメインを持つ時系列テンソルデータとして表現でき、本技術は、時系列テンソルに含まれる複雑な特徴を多角的に解析し、パターンの時間遷移とドメイン間の個体差を同時に抽出する。図6は車両走行データの解析結果である。図6(a)は(センサ、

ドライバー、時間) で構成される時系列テンソルからのパターン抽出結果であり、同色のセグメントが類似パターンのグループを表している。図6(b)は、出力結果を実際に走ったコース上にプロットしたものであり、図6(c)では詳細な出力結果を示している。本技術は、直進、右左折、徐行など車両走行における運転行動の時間遷移のみならず、ドライバごとの特性も同時に解析し、車両走行の様々な共通パターンを抽出すると同時に、車両走行のグループ化と、モデル化を完全自動で行う。

5.2 事業化に向けての取り組み

著者らの研究グループでは、様々な企業と連携し、社会実装に向けて製造業 DX 技術の実用化に取り組んでいる。その中でも本稿ではソニーセミコンダクタマニュファクチャリングとの取り組みを紹介する。

半導体製造工程における設備故障予測：上述した要素技術を統合、発展させ、複合時系列データからイベント予測を行うための AI ソフトウェアを開発した。センサデータの潜在的な動的パターンを時系列モデルとして要約し、特徴量として抽出することにより、イベントの要因分析を行いながら長期先のイベントの発生確率を予測することを可能とする。開発した AI ソフトウェアを活用して、CMOS イメージセンサーの半導体製造工程における DRY 装置のターボ分子ポンプの故障予測に関する評価実験を行っている。DRY 装置のターボ分子ポンプの突発故障は製造ラインへの影響が大きく、また高額パーツのライフ適正化をすすめることも可能となるため、故障予測のニーズは極めて高い。開発した AI 技術によって、DRY 装置のターボ分子ポンプ故障を事前に予測し、計画保全を実現する。

図7は、コントロールユニットにおける管理データ(軸の正常位置からのブレ、電流値、回転数など) およびターボ分子ポンプの振動計測データを解析し、装置状態を推定した結果である。図7の上段はオリジナルの入力情報となる22次元の設備管理および稼働データであり、最右端の時系列データの切れ目において故障が発生している。開発した AI ソフトウェアでは、センサデータに潜在する特徴的パターンやその変化点を時系列に捉え、装置状態を推定しており、図7下段はその解析結果である。装置状態を正常(青)、注意(オレンジ)、水色(故

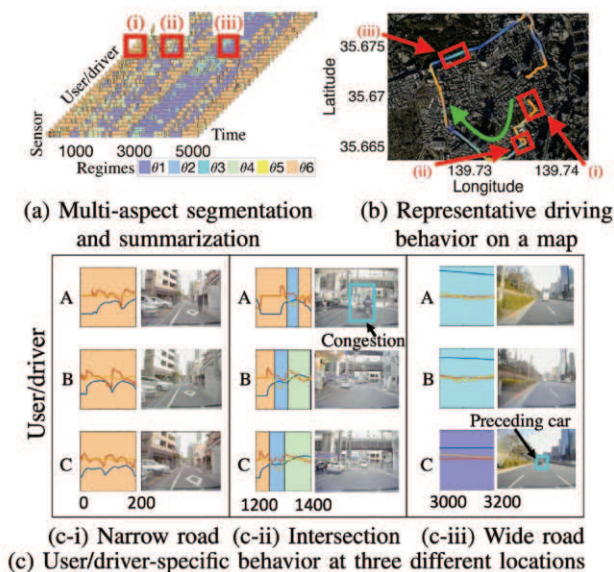


図6 車両走行データからの特徴自動抽出の結果
(トヨタ自動車との共同研究成果)

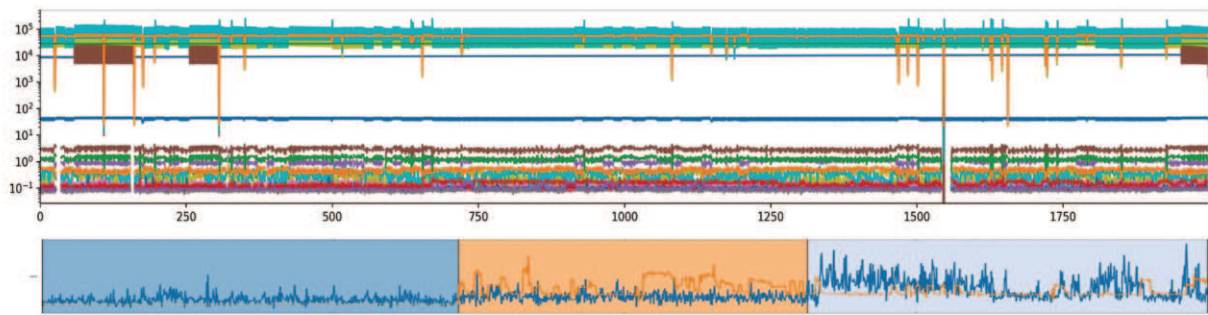


図7 設備稼働データの解析と故障予測

障直前) に分類しており、水色のセグメント幅は約13日である。すなわち、開発したAIソフトウェアによって、約13日前に故障の兆候を検出している。評価実験では、故障データは5件しかないものの、5件全ての故障の兆候を完璧に捉え、最短でも8日前に、最長で15日前に設備故障の予測に成功している。

6. おわりに

本稿では、著者らの研究グループにおいて開発したリアルタイムAI技術とその応用について述べた。IoT分野における様々なプロジェクトや事業に貢献するため、性能、機能、計算速度を向上させてAIソフトウェアとして開発しており、またその開発技術をもとにした社会実装や事業化をすすめている。

参考文献

- [1] Y. Matsubara, Y. Sakurai: "Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences", ACM SIGKDD Conference (KDD), pp. 1045-1054, August 2016.
- [2] Y. Matsubara, Y. Sakurai: "Dynamic Modeling and Forecasting of Time-evolving Data Streams", ACM SIGKDD Conference (KDD), pp. 458-468, August 2019.
- [3] K. Kawabata, Y. Matsubara, T. Honda, Y. Sakurai: "Non-Linear Mining of Social Activities in Tensor Streams", ACM SIGKDD Conference (KDD), pp. 2093-2102, August 2020.
- [4] Y. Matsubara, Y. Sakurai, C. Faloutsos: "AutoPlait: Automatic Mining of Co-evolving Time Sequences", ACM SIGMOD Conference, pp. 193-204, June 2014.
- [5] K. Kawabata, Y. Matsubara, Y. Sakurai: "Automatic Sequential Pattern Mining in Data Streams", ACM Int. Conf. Information and Knowledge Management (CIKM), pp. 1733-1742, November 2019.
- [6] T. Honda, Y. Matsubara, R. Neyama, M. Abe, Y. Sakurai: "Multi-Aspect Mining of Complex Sensor Sequences", IEEE Int. Conf. on Data Mining (ICDM), pp. 299-308, November 2019.