

時系列データ モデルフリー分析技術

吉永 直生 外川 遼介 網代 育大

要旨

時系列データ モデルフリー分析は、ディープラーニングを活用して時系列データの特徴を取り出し、照合できるようにすることで、高速、高精度な状態判別を実現する技術です。既存のデータ分析技術とは異なり、固定的な数理モデルを前提としないため、モデルチューニングの手間を大幅に削減できることが特徴です。本技術をシステムの監視業務に適用することで、現在のシステム状態がいつもと同じで問題ない、あるいは、いつもと違う、過去のあのときと似ている、といった監視員による判断を自動化することができます。本稿では、技術の特徴を説明するとともに、運用監視を含めたユースケースを紹介します。



時系列データ/データ分析/ディープラーニング/機械学習/異常検知/早期検知/予測/予知/最適化

1. はじめに

プラントや車両、データセンター、通信システムなどに設置されたセンサーから日々大量のデータが生み出されています。データには、数値やテキスト、音声・動画など、さまざまな種類がありますが、時間的に変化する時系列データの分析には、データの特徴を数式で表現した数理モデルがよく用いられています。センサーなどのデータから機械学習によって数理モデルを構築し、構築したモデルを使って予測や異常検知、分類などが行われます。

しかし、このような数理モデルを用いた方式には、データの前処理やパラメータのチューニングに手間がかかり、モデルの構築や検証、評価に多くの時間が費やされるという課題がありました。NECは、このような課題に対し、前処理やチューニングの手間を大幅に低減する時系列データ モデルフリー分析技術を開発しました。本技術は、固定的な数理モデルを前提とせず、時系列データから機械的に抽出した特徴同士を比較することで、システムの状態を高精度に判別するものです。本稿では、この時系列データ モデルフリー分析技術の中身とユースケースを紹介します。

2. 時系列データ モデルフリー分析技術

2.1 概要

時系列データ モデルフリー分析技術とは、センサーなどから収集された時系列データを分析し、検索可能にすることでシステムの状態判別を実現するAI技術です。本技術は、例えば、プラントや道路・橋梁、鉄道・自動車などの社会インフラシステムに設置されたセンサーから収集されたデータを用いて、正常や異常、異常の前兆などの状態を高速かつ高精度に判別することができます。システムの状態を正しく判別することで、運用監視における異常の検知や診断、予測などを実現します。

本技術は、対象から得られるデータをモデル化するのではなく、データの特徴を比較することで対象の状態を判別します。システムの運用監視において、監視員が経験的に行っているセンサーデータの特徴点の発見と同様のことをAIで実現しています。

数理モデルを用いる方法は、モデルに合致するデータの選別、加工や、システムの動作モードの変化に起因するモデルのチューニングや切り替えを必要とします。一方、本技術はデータの特徴をそのまま取り出して比較するというシンプルなアプローチのため、前述したような手間をかけ

ることなく導入できます。また、データが少量しかない初期の段階から運用を開始し、運用しながら精度を向上させることが可能です。

2.2 技術の特長

本技術の特長は、次の2点です。

1つ目は、データの特徴を抽出可能な特徴抽出エンジンを、データからの学習によって自動生成する点です(図1)。

検索に適した特徴抽出のために、本技術は、データの(1)時間的な変化と、(2)センサー間の関係性の2つに着目しました。エンジン内部でそれぞれの特徴を抽出し、抽出後にこれらの特徴を合成します。合成した特徴は、最後にバイナリデータに変換されます。

2つ目の特長は、ディープラーニングとNEC独自の学習指標で、特徴抽出エンジンを効率的に生成する点にあります(図2)。はじめに、学習対象データを蓄積されたデータ(セグメント)からランダムに1つを選択します(図2の9/9)。次に、検証用データとして数個のデータを選択します(図2の9/1、2、3、4、5)。次に、それぞれのデータを特徴抽出エンジンで特徴量(バイナリデータ)へと変換します。ここで、学習対象の9/9のデータと一番似ているのが9/5のデータだとすると、本来9/9の特徴量と一番近い特徴量は9/5であって欲しいのですが、今は9/4

の特徴量が一番近くなっています。この結果をもとに、9/5の順位が上がるように特徴抽出エンジンのパラメータが調整されます。

既存手法では特徴量の類似度を学習指標としていましたが、類似度の改善が必ずしも順位の改善につながらないという課題がありました。本技術では、学習指標として順位を直接扱うことで、前述の課題を解決しています。ただし、ディープラーニングは「順位」のような離散的な値を扱うのが苦手なため、次の近似を行うことで効率的な学習を実現しています。

ある学習対象 y_q (図2の9/9)に対し、上位に来て欲しいデータを y_t (図2の9/5)、その他を y_j (図2の9/1、2、3、4)とします。このとき、 y_t の順位は次のように表現できます。

$$1 + \sum_j I(\|H(y_q) - H(y_t)\| \geq \|H(y_q) - H(y_j)\|)$$

ここで、 $H(z)$ は z の特徴量、 $I(a \geq b)$ は $a \geq b$ の場合に1、その他は0となる関数です。前述の式では、 y_t よりも y_q に特徴量が似ている y_j の数をカウントし、 y_t の順位を求めています。ただ、 y_t は離散値であるため、次の近似によって学習を可能にしています。

$$1 + \sum_j \sigma(\|H(y_q) - H(y_t)\| - \|H(y_q) - H(y_j)\|)$$

ここで、 $\sigma(z) = \frac{1}{1 + \exp(-z)}$ はシグモイド関数です。

2.3 本技術の具体的な動作

システムの運用監視を例に、具体的な動作の流れを説明します。その他のユースケースは第3章で説明します。

図3に示したように本技術を活用した運用監視は、次の3フェーズからなります。

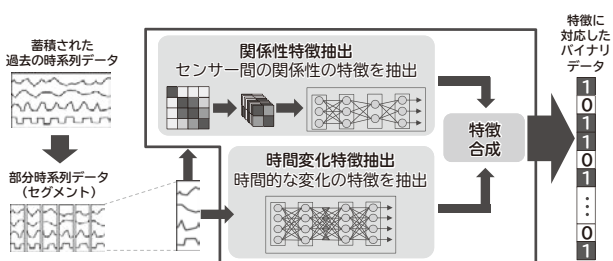


図1 特徴抽出エンジン

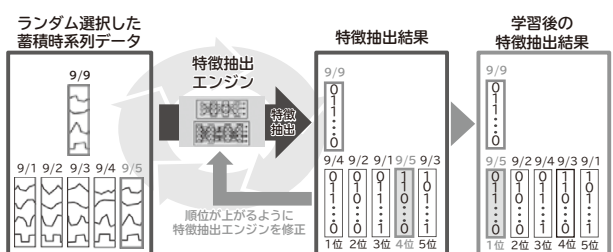


図2 学習方法

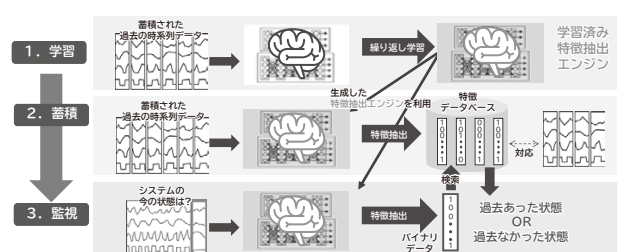


図3 本技術による運用監視の流れ

最初に、学習フェーズで特徴抽出エンジンを生成します。運用対象に設置された多種多様なセンサーから収集・蓄積された時系列データを、一定時間ごとの部分時系列（セグメント）に分割します。この各セグメントを繰り返し学習することで、時系列データをバイナリデータへ変換するディープニューラルネットワーク（図3中の学習済み特徴抽出エンジン）を得ます。

次の蓄積フェーズでは、生成した特徴抽出エンジンを用いて、各セグメントの特徴を抽出し、これをバイナリ形式に変換して特徴データベースに格納します。これにより、データ量を圧縮し、高速な検索を実現します。

最後の監視フェーズで、特徴データベースを使って運用対象の状態を把握します。まず、特徴抽出エンジンを用いて現在の時系列データセグメントをバイナリデータに変換し、データベースに格納された過去のバイナリデータと照合します。検索結果上位の過去データを参照することで、現在の状態が過去のいつと似ているかを判別することができます。検索の結果、類似の過去データが見つからなかった場合は、過去に発生したことのない未知の状態、つまり異常な状態である可能性が高いと分かります。このように、正常・異常を問わず、検索結果から運用対象の状態を高速かつ高精度に判別することができます。

2.4 性能評価

人体に付けた加速度センサーや心拍数など52センサーの時系列データに、「立っている」「走っている」などの13の行動状態タグが付与されたデータを用いて精度を

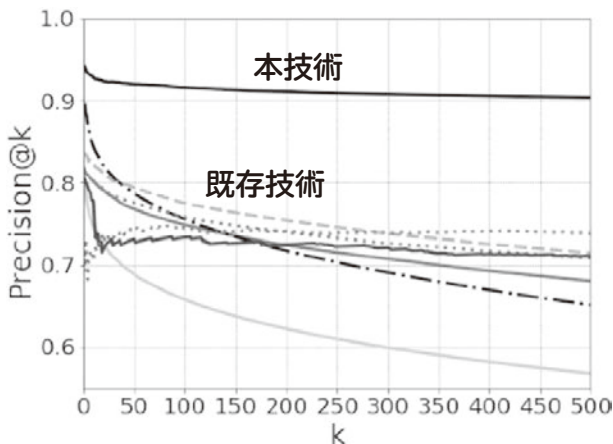


図4 生体データによる精度検証結果

検証しました¹⁾。ある行動を取っていた時刻の時系列データを検索キーにして過去データを検索したときに、検索結果にどれだけ同じ状態のデータが含まれるかどうかを評価しました。具体的には、Precision@kという、検索結果上位k個のうち、現在の状態と同じ状態が含まれる割合を評価指標としました。この値が高いほど、正しく検索できていることになります。

評価の結果、図4に示すとおり本技術が既存技術より10%以上高精度であることが分かりました。既存技術は検索精度が最高でも90%で、検索個数の増加に従って精度が落ちていきますが、本技術は500個検索した状態でも90%以上の精度を維持しており、大量のデータのなかから状態を正しく判別することに優れていると言えます。

3. 時系列データ モデルフリー分析のユースケース

ここでは、時系列データ モデルフリー分析技術のユースケースを3つ紹介します。1つ目は、本技術の高い判別性能を利用して、システム運用監視における障害対応を迅速化する例です。2つ目と3つ目のユースケースとして、運用監視以外の例を紹介します。

3.1 過去の対応記録に基づく初動対応の迅速化

システム運用時には、いつ発生するか分からない障害への迅速な対応が課題です。現在のデータを過去のデータと照合するだけでなく、過去の障害に紐付いた対応記録と連携させることで、現在の異常と近い特徴を持つ過去の障害対応記録から、そのときの対処内容を提示することができます。

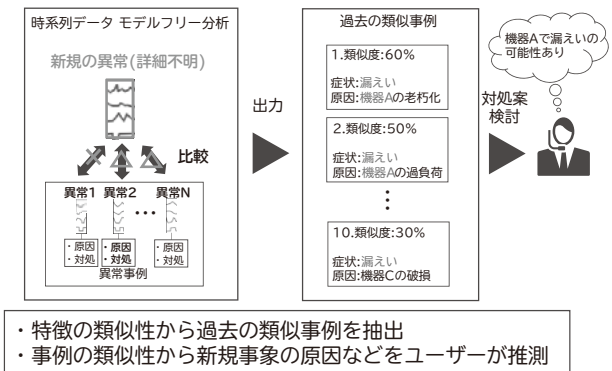


図5 対処案提示

また、本技術は、現在の異常がどの種類の障害に近いかを提示(分類)するだけでなく、個々の事例との類似度を提示します。そのため、これまで発生したことがない未知の事象であっても、類似する過去事例を参考に調査を進めていくことができます。類似度が100%に近い過去事例がない場合でも、上位の事例の症状や原因、発生個所を提示することで、保守作業による異常事象の理解や原因調査を支援できます(図5)。これにより、障害発生時の対応が早まるとともに、必要な作業が短縮されます。

3.2 前兆の検知によるイベントの発生予測

時系列データと事例が蓄積されることで、同種の事例に存在する共通の特徴を取り出せるようになります。例えば、運用監視では、装置故障の前に共通的に現れる前兆の特徴をとらえることで、障害を未然に防止できるようになります。運用監視以外でも、例えば、売上データやSNSなどのトレンドデータの分析から、需要の急増に伴う商品の在庫切れや、逆に急減による在庫過多のような好ましくない事象を事前に検知できる可能性があります(図6)。このような需要やトレンドの変化の特徴をとらえることで、事前に増産や減産などの対策を打つことができます。将来的には、SNSでのキーワード出現数推移から、アーティストのゲリラライブの前兆をとらえ、事前に警察官を配備するといった活用も考えられます。

3.3 過去の実績に基づく将来予想

新商品の発売直後でまだデータが少なく、回帰分析などによる売上の予測が難しい場合でも、本技術を使って、

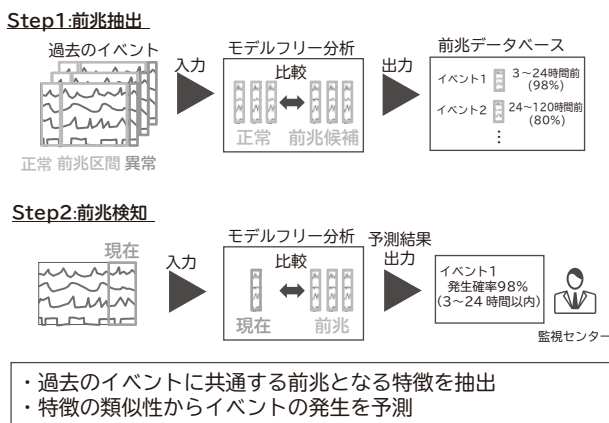


図6 イベントの予測

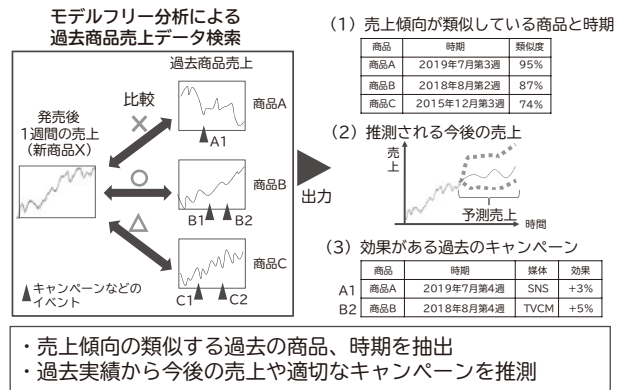


図7 売上状況分析

過去に発売した商品のなかから類似の売上傾向を持つ商品を見つけることで、売上を予測できる可能性があります。

具体的には、まず新商品の発売直後の売上データを、過去のさまざまな製品の売上データと照合し、特徴が類似する既存の商品を見つけます(図7(1))。次に、見つかった類似商品の売上実績パターンをもとに、新商品の売上が今後どのような経緯をたどるかを予測します(図7(2))。更に、類似商品に対して行ったキャンペーンの実績データを別途参照することで、売上向上に寄与する広告媒体やキャンペーン内容に関する示唆が得られます(図7(3))。このようにして、過去の実績に基づく将来予測や打つべき手立ての判断が可能となります。

4. おわりに

本稿では、ディープラーニングを活用して時系列データの特徴を取り出し、照合できるようにすることで、高速、高精度な状態判別を実現する時系列データ モデルフリー分析技術を紹介しました。本技術は、大規模複雑化したシステムの運用監視を始め、売上推移や生体データなど、さまざまな時系列データに適用することで、新たなソリューションの開発や既存のソリューションの強化に貢献します。

今後は、数値だけでなく、テキストや音声、動画といったマルチモーダルなデータを分析できるように技術を強化する予定です。長年にわたり取り組んできた時系列データ分析技術の研究開発を継続・発展させることで、安全・安心な社会の実現を目指します。

参考文献

- 1) D. Songほか: Deep r-th root of rank supervised joint binary embedding for multivariate time series retrieval, Proceedings of the 24th ACM SIGKDD, pp.2229-2238, 2018.

執筆者プロフィール

吉永 直生

セキュリティ研究所
主任

外川 遼介

セキュリティ研究所
主任

網代 育大

セキュリティ研究所
プリンシパルクリエイター

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報 (日本語)

NEC Technical Journal (英語)

Vol.72 No.1 新たな社会価値を生み出すAI特集

新たな社会価値を生み出すAI特集によせて
AIとデータ活用によるデジタライゼーションの拡大

◇ 特集論文

AIの社会実装に向けた取り組み

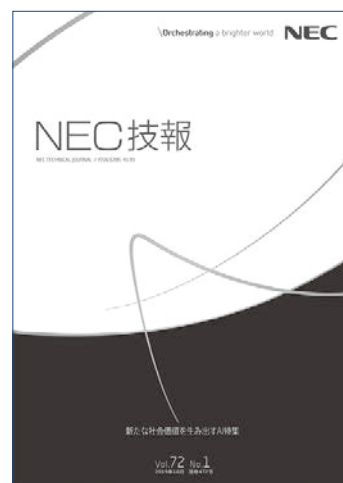
「NECグループ AIと人権に関するポリシー」とその実践に向けた取り組み
AI時代の人材育成

デジタルトランスフォーメーションを加速するAI活用サービス・ソリューション

「みんなで創るAI」を支えるNEC Advanced Analytics Platform (AAPF)
物体指紋認証技術による個人識別機能の活用
画像処理コントローラへのディープラーニング活用による外観検査ソリューション
通信予測制御技術を活用した車両の遠隔監視ソリューション
働き方改革や健康経営を支える「NEC感情分析ソリューション」
オフィスのセキュリティと利便性を向上する「顔認証ソリューション for オフィス」
業務自動化・省力化を実現する自動応答ソリューション (AIチャットボット) の概要
ビジネス創造へのワークシフトを加速するソリューション (AI for Work Shift Support) の概要と実証事例
AIを有効活用する「NEC Energy Resource Aggregation クラウドサービス」
容体変化予兆検知技術による早期退院支援の取り組み
予防・健康領域に対するデータ活用による効果的なアプローチ
AIを活用したインサイトマーケティング事業の共創
時代のムードを味わえる「あの頃はCHOCOLATE」の開発

人とともに未来を創る最新のAI技術

あらゆる小売商品を認識可能にする多種物体認識技術
ネットワークインフラを活用して実世界を見える化する光ファイバセンシング技術
熟練者の意思決定を模倣する意図学習技術
グラフベース関係性学習 (GraphAI)
時系列データ モデルフリー分析技術
社会インフラの最適運用を支援する論理思考AI
少量データ向け深層学習技術
AIを支えるコンピューティングプラットフォーム



Vol.72 No.1
(2019年10月)

特集TOP