

# 異常検知の機械学習モデルを構築 トライアル分析で高精度な結果

ひとたび発生すると各所へ大きな影響損害を及ぼすことになる機器・設備の異常。そのため異常が起きたときに発生した個所を早くに特定することや、異常が起きる前に予兆を検知して対策を講じることが重要です。こうしたことからNTTデータ数理システムは機械学習などが可能なデータ分析プラットフォーム**Alkano**に、異常検知の分析に有効な時系列データからの特徴量抽出やそれらを用いた分析機能を追加しました。これらの分析機能の有効性を検証するため実施したモニター企業でのトライアル事例について、担当した技術者に聞きました。

## Interview

### Alkanoに追加した時系列データ分析機能は異常検知に有効

#### Alkanoに、異常検知に有効な時系列処理機能を新たに追加しました。

**山口** 異常検知とは普段と異なる挙動を見つけることです。その検知技術は、設備保守、製造ライン、社会インフラとさまざまな分野で活用されています。背景にはセンシング技術の向上により各種データが取得できるようになっていることや、データサイエンスや機械学習技術の発達があります。

製造業ひとつとってみても異常検知はさまざまな場面で役立ちます。例えば工場の冷却装置なら、センサーデータと故障履歴を用いて正常／異常な状態のパターンを分析することで早期の異常発見が可能になります。また気圧、湿度、気温など環境センサーのデータから配電盤の故障を予知できれば、事前に機器交換することで関係設備の稼働停止を回避できるでしょう。ほかにも、製造ラインのセンサーデータと不良品検査の結果を分析・モデル化することで、検査を行わずにセンサーの情報だけで自動的に不良品を発見する、といったことの実現につながります。

これらはいずれも機械学習を中心とした技術で解決できる可能性があります。当社の技術を組み合わせることで、異常の早期発見や予兆検知を必要とする企業のお役に立てるのではないかと考えて異常検知の分析に役立つ機能を開発し、新機能として**Alkano**に追加しました。

#### 提供開始に先立ちトライアル分析を行いました。その実施内容について教えてください。

**中道** モニター企業を募集し、異常検知分析のトライアルを実施しました。その中からA社の「空調設備の異常検知」をご紹介します。A社の化学品製造工場では吸収式冷温水発生機と呼ばれる空調設備をお使いです。空調設備が故障すると最悪の場合、製造中止など甚大な損害が出るのが想定されます。そのため異常の予兆が出た段階で対処する「予知保全」によって故障を防ぎたいと、モニター企業として手を挙げていただきました。

**尾崎** この空調設備には多数のセンサーが備わっていて、出入口温度、瞬時熱量、瞬時流量、外気温、外気湿度、燃料である都市ガス消費量などが随時記録されています。A社からは、それら約40個のセンサーについて、過去3年間の計測データと故障記録をご提供いただきました。トライアルではこの大量のデータをもとに、直近3日間の計測データを使って次の日に故障（≒異常）が起きるかどうかを予測する異常検知モデルをつくることを目標に設定。最終的にRandom ForestやSupport Vector Machine (SVM) と呼ばれる機械学習手法を用いることで、異常の予兆を高精度に検知できるモデルを構築することができました。



データマイニング部 山口 裕



データマイニング部 中道 祐希



データマイニング部 尾崎 博子

## PROFILE

### 山口 裕

データマイニング部マネージャー。顧客分析や製造現場の不良要因分析など、幅広いビジネス領域でデータサイエンスに取り組む。

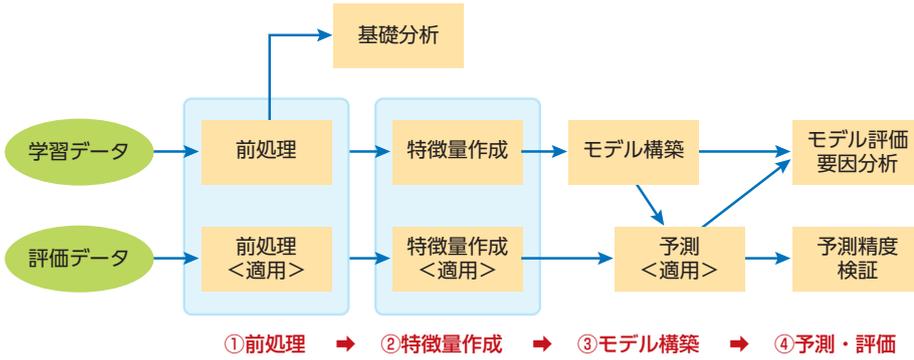
### 中道 祐希

データ分析のスペシャリスト。20年にわたり、異常検知を含む機械学習全般に関する受託分析に従事。近年では知識を活かして分析ツールの開発にも携わる。

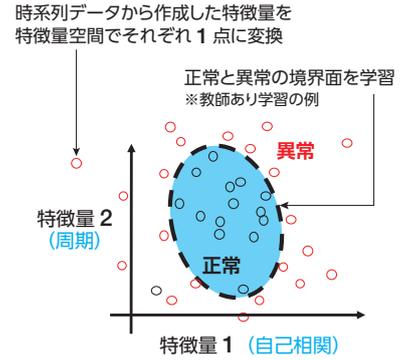
### 尾崎 博子

機械学習やテキストマイニングツールの開発に従事。豊富な経験に裏付けられた技術コンサルティングや、データサイエンス教育の講義には定評がある。

異常検知モデルの作成フロー



モデル構築イメージ



Alkanoでの分析フロー「時系列データの異常検知」の例



どのような技術を用いて、どのように異常検知モデルを作ったのですか。

**尾崎** 異常検知のための機械学習モデルの作り方・手順は、ある程度、定型化できます。一般的な手順は次のとおりです。まずモデルを構築するための学習データと、できあがったモデルを評価するための評価用データを用意します。次にデータの前処理と基礎分析を行い、前処理を施したデータを使って時系列データの特徴量を作成します。そこから機械学習モデルを構築し、モデルができあがったら評価用データに適用してみて、モデルの評価や予測精度の検証を行う、という流れです。つまり検討しなければいけない項目が、①前処理 ②特徴量作成 ③モデル構築 ④予測・評価と、大きく4工程あると言えます。これらの各工程で使える手法・技術は数多くあるので、状況に応じてもっとも有効なものを選びながら作業を行います。A社の空調設備もこのステップで進め、モデル構築後はパラメータチューニングでブラッシュアップ。それによって高精度なモデルとなりました。

「特徴量」という言葉が出てきましたが、これが異常検知モデルの構築に大きく関係するんですね。

**中道** そうです。例えば時系列データには、「時間的に隣り合う2点は同じように高い値を示す」といった「自己相関」

や、「12カ月単位で同じような変動を繰り返しやすい」といった「周期性」が見られることがあります。こうした特徴を定量的に表した数値が特徴量です。機械学習は、これらの部分的な時系列データから作成した複数の特徴量から正常と異常の境界を学習します。ですから特徴量の取り方はとても重要で、結果の精度に大きく関わってくるのです。

トライアル分析に対するお客様からの評価はいかがでしたか。

**尾崎** トライアルでは、①データの整形・前処理 ②基礎分析 ③特徴量作成 ④機械学習モデリング ⑤パラメータチューニング ⑥予測精度評価・要因分析 ⑦レポート作成を行いました。A社を含めトライアルのお客様からは「これまで気づかなかった傾向が見いだされ、新たな発見になった」「データの絞り込み方や時間単位のみとめ方、分析方法などが参考になった。これまでのやり方の見直しに役立てたい」「検討項目の洗い出し方がとても役立った。また一連の分析から、分析ポイントを学ぶこともできた」などのご感想をいただいています。

今後のAlkanoの異常検知分析についてお聞かせください。

**山口** 当社は異常検知を含む機械学習全般で、これまでたくさんの方の製造業のお客様から受託分析のお仕事をいただけてきました。これらを通して蓄積してきた知見・技術を結集したのがAlkanoの時系列分析機能です。「センサーデータをもっと活用したい」「設備の故障や不具合の兆候をとらえ、効率的な部品交換・修理をしたい」とお考えの方々に、当社の分析技術をテンプレートとし、利用される方の敷居を下げたこの異常検知分析ツールをご活用いただきたいと思います。私たちもこの先、分析やデータ処理の機能をさらに増やし、また新しいアルゴリズムの検討も行うなどして、一層、製造業の皆様にお役立ていただけるようにしてまいります。