

センサーから読み解く 製造工程の異常検知に活用

働き方改革や人口減少に伴う人手不足に対応するため、AIの活用を検討する企業が増えています。一方でAIといえども、使う人間に確かな知見や意思がないと狙った結果は得られません。また、多くの企業がツールの導入ではなく、活用で挫折することも報告されています。NTTデータ数理システムでは、AI活用への取り組みに不安がある方向けにサポートサービスを行っています。業務知識や自社のデータに対する知見はあるが分析経験がなく、しかし成果を求められている担当者に対して、その目的やデータに合わせ、幅広い経験や知識を持つ技術者が、親身に相談にのります。その技術者の1人、データマイニング部シニアアナリストの中道祐希は、センサーデータを利用した異常判別を実現した際に、Visual R Platform (VRP) を活用。その分析の進め方について聞きました。



NTTデータ数理システム
データマイニング部
中道 祐希

Interview

分析前のデータ整理も集計も。VRPで現状把握

相談者がどのようなことに困っていて、何を解決しようとしていたのかを教えてください。

中道 ご相談者は大手メーカーのICT部門に所属される方で、データを活用することで、さらなる品質向上につながる「気づき」が得られないか、という課題をお持ちでした。製造現場で記録されたデータの一部を持参されたので、お持ちになったデータについてヒアリングを行いました。データは時系列に記録された100種類以上のセンサーデータで、一定の時間間隔で測定されており、1レコード1000時点前後あります。1レコード毎に、このようなテーブル形式で保存されていました。それぞれのレコードには、別途費用をかけて行ったテスト結果（合格、不合格）が紐づいていました。分析の目的は「不合格となる原因が何か知りたい」「新たなセンサーデータから、テストをしなくても合否が判定できるようにしたい」ということでした。ご相談者はすでに、ご自身で目星をつけたセンサーのグラフ化を行っていましたが、不合格に結びつくセンサーや規則性は発見できていませんでした。

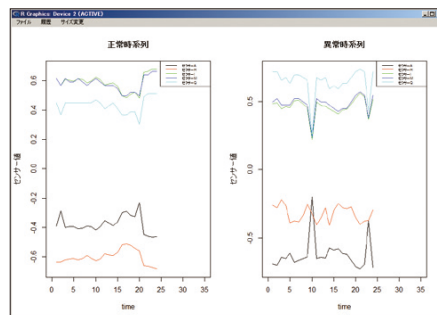
問題の難しさを教えてください。

中道 通常の判別問題は1行に原因となり得るデータと、結果のデータが保存されている状態で分析を行い、結果に対する関連性を特定し、予測を行います。そのための手法はさまざま開発され、代表的なものにはDecision Tree（決定木）やニューラルネットワークがあります。

分析データのイメージ

| | | | |
|------|----|----|--------|
| Time | V1 | V2 | } Good |
| 0001 | } | } | |
| 0100 | } | } | |
| Time | V1 | V2 | } Bad |
| 0001 | } | } | |
| 90 | } | } | |

長さの異なる時系列データを結合し、教師の値をつける



時系列のパターンをグラフで把握

PROFILE

中道 祐希

2004年入社。異常検知を含む機械学習全般に関する受託分析に従事。近年は異常検知ツールの開発、コンサルティング、データサイエンティスト教育プログラムなどの自社ソフトを用いた顧客の支援を中心に取り組む。

しかし、今回は1つの結果データに対して、1000行を超えるセンサーデータが紐づいています。これまでの判別手法では分析ができません。そこでさまざまな工夫を行う必要があったのです。

相談を受けて、最初に取り組んだことを教えてください。

中道 現状把握です。まずはVRPを用いて、各センサーデータの基本的な統計量を計算し、欠損、抜け、異常値などを、集計結果やグラフを用いてピックアップ、データを分析しやすい形に整備しなおしました。VRPは、NTTデータ数

理システムが開発したRをエンジンとした統計解析ツールで、統計量の計算はもちろんですが、データ加工機能も充実しているため、データ整備が手軽に行えます。このときも、すぐにその後の分析に使えるようなデータに整備することができました。そして、集計結果を確認することで、各センサーデータの意味と動き、範囲を把握しました。集計は現状把握をするのに役立つ指標で、大量のデータを理解し、データの特徴を把握するために欠かせないものです。AIが活用される時代でもそれは変わりません。

VRPの豊富な機能で分析。結果をもとに異常検知モデルを構築

センサーの数が100種類以上あるとは、人間が一度に把握できる量を超えていたのですね。その問題にはどのように対応したのでしょうか。

中道 現状把握をすることで、可動範囲の大きいセンサー、まったく値が変わらないセンサーなど、いくつかのパターンが想定できました。そこで、センサーをグループに分けることを考え、センサーの関連性とその構造を、相関係数とクラスター分析の手法を用い、分析しました。その結果、同じような動きをするセンサーをまとめることができました。この作業で、特徴的な動きをしているセンサー、より全体に影響を及ぼしそうなセンサーを特定しました。

みが立ちました。現在は人間がテストデータをチェックするための参考指標として用いていますが、将来的には自動判定を目指しています。

最後に、今回の異常検知分析に用いられているVRPについて、開発で工夫をされた点、ユーザーに喜ばれている点を教えてください。

中道 Rは世界中の研究者が利用していることもあり、最新の統計や機械学習の手法が搭載されています。しかし、データを読み込み、集計するためには、プログラミングを覚え、コードも正しく記述する必要があります。VRPは、忙しいビジネスマンがスピーディに強力なRによる分析を行えるよう、ノンプログラミングでデータ処理、可視化、分析ができるアイコンを備えています。分析結果も分かりやすく、実務に役立てやすい形式で出力されるよう、工夫されています。

重要なセンサーを特定したわけですね。でも、1つの可否判定に対して1000行のデータが結びついているので、このままだと予測モデルの作成ができない問題は解決されませんが…。

中道 そうなのです。そこで、レコード毎の特徴量の作成を行いました。特徴量の作成には統計量計算やその他、さまざまな方法があり、複数を候補とします。この候補を用いて学習データの「合格」「不合格」の2グループの判別ができる特徴量をVRPのElastic Netなどを用いて探索します。その後、モデル学習を行い、異常検知モデルを構築しました。

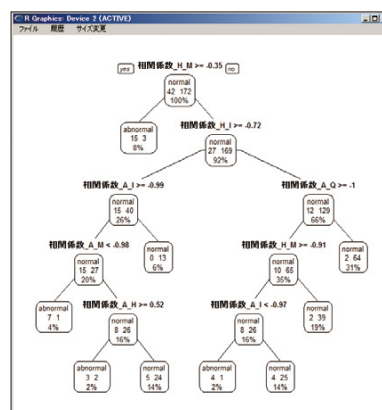
今回は分析をすることで、重要なセンサーを抽出し、レコード毎の特徴量を算出し、モデル学習を行い、不合格の原因となる複数のセンサーの条件、規則性を見つけ出すことに成功しました。

VRPを使った方が、プログラミングなしに高度な分析ができることと驚き、時間の短縮になると喜んでくださるのを見るのはとても嬉しいですね。また、VRPはR関数にユーザーが好みのインターフェースをつけて実行ができる、独自アイコン構築機能を持っています。この機能により、Rに搭載された最先端の解析手法をすぐに使いやすい形でとり入れることができます。世界中で高く評価されているRが手軽に使える環境を実感していただきたいと思います。

結果をどのように活用されたのでしょうか？

中道 機械学習はブラックボックスだと言われる一方で、人間による原因把握に役立つ手法もあります。このような手法を試すことで、不合格となる条件を見つけ、現場の気づきにつながりました。今回は製品の試作実験段階でのデータ分析でしたが、本格製造時の不具合の早期発見に役立ちますし、品質向上のための知見を得ることができたので、結果は製品設計にも活かせます。

また、今後さらに精度のよいモデルを作れば、一部のセンサーの値だけで、合格・不合格の判定が可能となる見込



人間にも理解が進みやすい学習手法Decision Tree