OSIsoft is now part of AVEVA

機械学習とデータ準備の基本

始める前に知っておきたいこと

企業が機械学習（ML：machine learning）を使って膨大なオペレーションデータを分析し、プロセスの改善を図る機会が増えています。ところが、プロセス改善のために既に「ビッグデータ」を活用している企業が、初めて機械学習を取り入れる際に、多くの不明点や取りこぼしを抱えることは珍しくありません。

この紙面では機械学習とは何か、そして機械学習を用いた分析に必要なデータ準備についての基本を説明します。

機械学習はデータ分析で「インテリジェンス」を実現する

機械学習とは、ソフトウェア内蔵モデルを用いて、データを解析しながら自動的に知識を蓄積していく仕組みです。RやPythonなどの言語で書かれたアルゴリズムは、モデルの目的に基づいて、モデルがどのように「学習データ」と相互作用するかを定義付けします。

機械学習は、予測精度を向上させたり、データの中に潜む未知のソリューションを特定したりなど、データ分析の限界をさらに押し広げます。産業界での応用例としては、データサイエンティストやエンジニアが、アセットの効率化、予知保全の改善、より正確性の高い使用法、需要予測、事象の根本原因分析（RCA）の迅速化などの課題に対応するために機械学習を活用しています。

機械学習では、適切なデータを適切なアルゴリズムに提供することにより、産業分野における多くの一般的な問題を解決します。ここで、有益な成果を得るためにデータが重要な要素となります。機械学習プロジェクトでは、学習モデルに供給する種類に富んだ膨大な高品質データが必須です。

機械学習はいつ使用するべきか

機械学習が問題解決に役立つ5つの場面

1. 分類

AかBか、を判断したい場合。たとえば、「このポンプは90日後には故障するだろうか」のような「はい」「いいえ」で答えられる質問は、機械学習に向いています。

2. 異常検知

正常範囲外かどうかを知りたい時。たとえば、圧力値の異常を知らせるアラート受信をします。

プロセスのばらつきを減らす

世界的な製薬企業のベーリンガーインゲルハイム社は、動物用ワクチンの製造過程で機械学習を利用し、プロセスのばらつきに迅速に対応しています。具体的には、多変量解析と統計モデリングを用いて、プログラマブルロジックコントローラ（PLC）内の対応ステップを「EventFrames」でトリガーするシステムを構築しました。これにより、プロセスのばらつきが最も問題になりそうなステージに集中してオペレーターが監視できるようになりました。
ケーススタディを読む。

3. 回帰

どれくらいの数量になるか予測する場面。たとえば、「この機械を２シフト分追加で稼働させると、エネルギー消費はどのくらい増えるか」など。

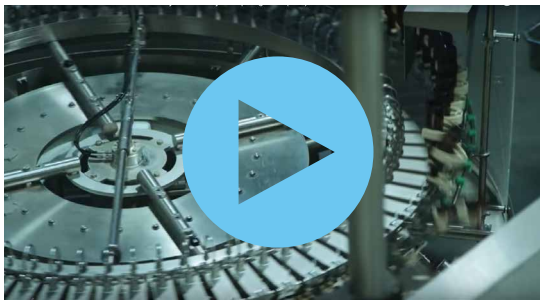
4. クラスタリング

データをどのように整理すればよいか?この問いには正解も不正解もありませんので、類似性のある指標に基づいてグループ化を行います。たとえば、「どの種類のポンプが似たような経過で故障するか」など。

5. 強化学習

今後の施策展開方針を助けます。レインフォースメントにより、モデルが過去のアクションや経験から得られたフィードバックを用いて、インタラクティブな環境でトライ＆エラーを繰り返しながら学習します。自動化システム（温度制御システムなど）のパフォーマンス向上や、生産能力予測の精度を上げるのに適しています。

品質管理のモデリング



オレゴン州に本社を置くデシューツ・ブリュワリー社は、バージニア州に新しい醸造所を建設する際、既存の醸造所のオペレーションデータとMicrosoft Cortana（機械学習機能を搭載したデジタルアシスタント）を利用して、新拠点のオペレーションをモデル化できました。これには、Azureベースのシステムのデータセキュリティも含まれています。
動画を見る。

機械学習には
高品質なデータとデータ準備が不可欠

機械学習には、正しく分類された、豊富なデータが必要です。機械学習は実際の状況を完全に網羅したデータ、つまり様々な過去のデータセットが揃って初めて機能します。ユーザーは、関連するすべてのシステムからデータを確実に収集しなくてはなりません。適切なデータセットが1つでも欠けていると、機械学習の成果の質に影響します。

また、精度の高いコンテキストを得るには、データの多様性も重要です。たとえば、外部の研究機関が発表する大気データを取り入れると、再生可能エネルギー事業者は、太陽光と風力の予測を改善でき、予想発電量を算出するのに役立ちます。

機械学習の利用を検討しているオペレーションチームは、データ準備の過程で問題に突き当たるかもしれません。最も一般的な課題の1つに、オペレーションデータがサイロ化されたシステムや古い機器に保存されているという点があります。また、データが異なる拠点や地域に保存されている場合もあります。このような状況下で機械学習の利用実現のために最優先すべきは、将来的に関連すると想定される、あらゆるオペレーションデータを**単一のデータ管理システムに集める**ことです。

データの準備が完了したら、機械学習プロジェクトを開始できますが、まだ注意が必要です。アルゴリズムに投入する前に、データを合成し、コンテキストを与えることが重要です。

そこでデータ分析に調査的フェーズを設け、問題となった項目があればデータサイエンティストやアナリストが技術の専門家と話し合えるようにしましょう。これには2つの理由があります。1つ目は、現場の専門家から関連データの早期解析に役立ちそうなコンテキストに沿った洞察を得られ、プロジェクトの効率化やタイムラインの短縮化につながります。2つ目は、そのコンテキストに沿った洞察を活用すると、これまで考慮されてこなかった必要なデータパラメータを特定できます。機械学習の学習データに関連パラメータが欠けていたら、そのモデルは意義のある実用的な結果を提供できません。

たとえば、望ましいパラメータの範囲外で動作している機器があり、データサイエンスチームが原因を特定しようとしているとします。その場合、当然、過去の機器使用データや、プラント環境に関するデータ（中央部センサーから得られる気温や湿度データなど）を利用するでしょう。ところが、現場のオペレーターであれば、その機器が冷気にさらされる場所、あるいは他の機器よりも少し傾斜のある場所に設置されているなど、固有の重要な情報を把握しているかもしれません。

結論

機械学習に必要なデータ準備を実現するシステム

機械学習は、さまざまな産業オペレーションを全く新たに生まれ変わらせる可能性を秘めています。OSIsoftは、機械学習プロジェクトで使うデータの準備に必要なあらゆるステップをサポートしています。PI Systemを導入すると、チームはオペレーション全体をまたいだ複数のソースから大量かつ高精度の時系列データを収集、分析、可視化、共有することができます。

CLMLEN-091819