

ホワイトペーパー

実例に学ぶ予知保全向けデータ活用

～MATLABを使った故障予測で一步先のメンテナンス～

はじめに

近年は製造業、インフラなど様々な現場においてデータ活用が進んでおり、「予知保全 (Predictive Maintenance)」の取り組みが注目を浴びている。予知保全とは、機器に取り付けられた各種センサーから得られる情報から機器の故障時期を予測し、適切なタイミングでのメンテナンス実施を目指すものである。メンテナンスコストの削減だけでなく、予期しない故障を減らすことによる稼働率や安全性の向上も期待されている。

稼働率ダウンが致命的になる発電設備や、安全性を追求する航空機業界では早い段階から取り組みが進んでいたが、ここ数年でその適応範囲は広がりを見せ、工作機械や生産設備なども含め、様々なところで予知保全が実現されつつある。多くのプラントで老朽化が進むほか、保守・安全管理の実務を担ってきたベテラン技術者が引退の時期を迎えつつあることから、今後予知保全のニーズは更に高まることが予想される。本稿では予知保全システムの開発事例と関連技術を紹介する。

大量のセンサーデータに潜む知見

大量のデータから得られる情報を活用することで新たなビジネスに繋げる取り組みは予知保全に限らない。データに基づいたシステムの診断と評価や、今までベテラン技術者の勘や経験に頼っていた部分の自動化など、統計や機械学習といった分析手法を活かした成功事例も出てきている。ただ、サーバなどにデータは蓄積されてはいるものの、そのデータを直接目で確認するような監視用途以外に有効活用されていない状況も多く見受けられる。ここで、故障や性能劣化を予測診断するようなより高度な診断システム開発に取り組む際の課題を考えてみる。

1つ目の課題は解析すべき時系列データの規模である。例えばある鉄鋼プラントでは1分周期で数百項目の時系列データがサーバに蓄積されるが、一定期間のデータをもとにシステムの故障予知を行うことを考えると、解析するデータは大規模なものになる。この点に関しては、近年大規模なデータを取り扱う環境(例えば大容量で安価なデータストレージ、高速で安価な演算処理装置、クラウドや分散処理用のインフラ)が整備されつつあり、ハードウェアの面ではアプローチしやすくなってきている。2つ目の課題はいわゆるデータサイエンティストの不足である。上記の環境が整えられたとしても、重要なのは性能劣化や予兆診断を行うアルゴリズム開発効率である。アルゴリズム開発には統計などの数理的な知識だけでなく、コンピュータサイエンスの知識・プログラミング技術、さらには監視対象の機器に対する経験・知識も重要になってくる。

以下ではこれらの課題を乗り越え、予知保全システム構築を実現した例を紹介する。

予知保全システムの開発事例

ここでは、Mondi Gronau(以下、Mondi)を例として予知保全システムの実例を紹介する。Mondiは梱包及び紙製品の世界的なリーディングカンパニーであり、そのプラスチック製品及び薄膜フィルムの生産量は年間1800万トンに及ぶ。約900名の作業員が約60台の射出成形機、印刷、接着及び巻き取り機械を24時間365日稼働させているが、機械の故障によるダウンタイムや原材料の破棄に伴い毎月数百万ユーロのコストが発生していた。これらのコストを最小化し、プラントの生産効率を最大化するために、Mondiは状態監視及び予知保全システムの開発に取り組んだ。

開発されたシステムは機械オペレーターがすぐに明確な形で機械の状態を確認できるアプリケーションとなっており(図1)、機械の潜在的な問題も特定でき、是正処置や深刻な問題の回避を行えるようになっている。予知保全システム開発の流れを下記の4つのステップに分けて紹介する。

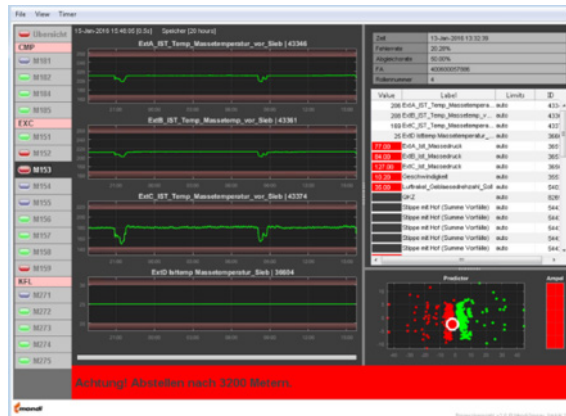


図1: 開発されたアプリケーションでオペレーターは機械の状態を逐次確認できる

1. センサーデータ収集

Mondi で使用される射出成形機やその他の機械は非常に大型で複雑な機構のものが多い。個々の機械は最大5台のプログラマブルロジックコントローラー (PLC) によって制御され、機械のセンサーから温度、圧力や速度などの性能パラメータを取得し記録していた。パラメータ数は1台の機械につき300-400個あり、取得されるデータ量は1日当たり7 GBにも及ぶ。Mondi では今回の取り組みにおいて、工場内のすべての機械からイーサネット経由でOracle® のデータベースに収集する仕組みを構築した。これらのデータを分析し、機械の潜在的な問題を特定するアルゴリズムの開発を行った。

2. データ前処理

様々なソースのデータから予測モデルを開発する際、取り出されたデータはそのまま使用されることは少なく、予めデータの前処理を行う必要があることが多い。例えば、欠損値や異常値がある場合や、データによって使用しているタイムスタンプの形式が異なる場合もあるからである。Mondi の場合、予測に寄与しないと判断される変動の無いデータの削除だけでなく、1分毎に記録される製造設備の性能パラメータにくわえて、約1時間毎に記録された品質状態をあわせて処理できるように前処理を行った。

3. 故障予測モデル開発

前処理を行い整形されたデータに対して、機械学習アルゴリズムを適用し予測モデルの学習を実施した(図2)。過去のデータからパターンや傾向を自動で学習するステップである。ただ、機械学習と一言で言っても線形回帰などシンプルなものから、深層学習など高度で複雑な手法まで様々なアルゴリズム(手法)が存在する。一般的に故障予測を実現できる決まった手法が存在するわけではなく、様々な手法から最適なものを検討する必要がある。Mondiではニューラルネットワーク、k最近傍法、バギングされた決定木、サポートベクターマシン (SVM) といったよく使われる機械学習アルゴリズムを評価した。一部原理と特徴を表1にまとめる。実際のセンサーデータを用いてそれぞれの手法による予測精度を検証した結果、バギングされた決定木が最も正確に予測を行うアルゴリズムであると結論付けられた。通常、学習用に使用した同じデータをアルゴリズムの評価に使用すると正しく性能が評価できないため、例えばデータの60%をモデル学習に使用し、残りの40%を評価に使用するといったことが行われる。

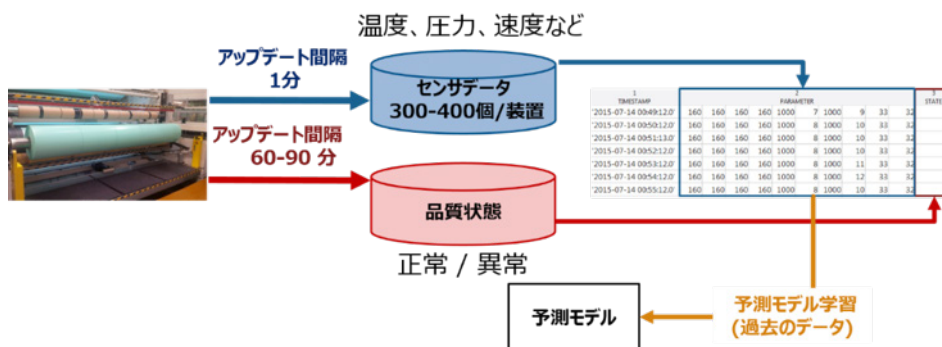


図2: 異なるサンプリング時間で記録された過去のデータから予測モデルを作成

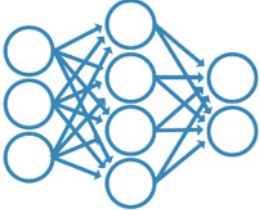


アルゴリズム名とイメージ図	原理	最適な用途
<p>ニューラルネットワーク</p> 	<p>人間の脳にヒントを得たニューラルネットワークは、互いに密接に結びついたニューロンのネットワークから成り、このネットワークを介することで、入力データを望ましい出力へと関連づける。ニューロン間の結びつきの強さを繰り返し修正することによってネットワークはトレーニングされ、与えられた入力データに対して正しい応答を出力するようになる。</p>	<ul style="list-style-type: none"> • 利用できるデータが増加していく場合で、モデルを絶えずアップデートしたい場合 • 入力データに予期せぬ変化が生じる可能性がある場合 • モデルの解釈のしやすさが重要な問題ではない場合
<p>k最近傍法(kNN)</p> 	<p>kNNは、データセット内の最近傍点のクラスに基づいて、対象物をカテゴリー化する。kNNによる予測では、互いに距離が近い対象物同士は類似していると仮定している。ユークリッド距離、市街地距離、コサイン距離、チェビシェフ距離などの距離測度を用いて最近傍点を見つける。</p>	<ul style="list-style-type: none"> • シンプルなアルゴリズムが必要な場合 • 学習対象のモデルのメモリ使用量・予測速度があまり問題にならない場合
<p>サポートベクターマシン(SVM)</p> 	<p>全てのデータポイントをあるクラスと別のクラスとに分ける線形の決定境界(超平面)を特定することによってデータを分類する。SVMでは、カーネル変換を用いて、線形分離できないデータをより高次元に変換することにより、線形の決定境界を見つけることもある。</p>	<ul style="list-style-type: none"> • 高次元で線形分離できないデータの場合 • シンプルで解釈がしやすく、正確な分類器が必要な場合

表1: 機械学習アルゴリズムの原理と特徴

4. 工場への展開

データベースから取得した機械のセンサーデータは上述の前処理を経た後、過去のデータを使って学習された故障予測モデルによって、リアルタイムで異常の有無が自動診断される(図3)。一連の処理を実行するプログラムは予知保全システムとしてアプリケーション化され、オペレーターが機械の状態を確認できるインターフェイスを備えたソフトウェアとして工場内で稼動しており、中断のない安定した運用が可能になっている。現在のところモニター対象の機械は8台だけだが、年間5万ユーロを超えるコスト削減が実現されている。Mondi は残りの機械のデータについても今後解析を進める計画である。

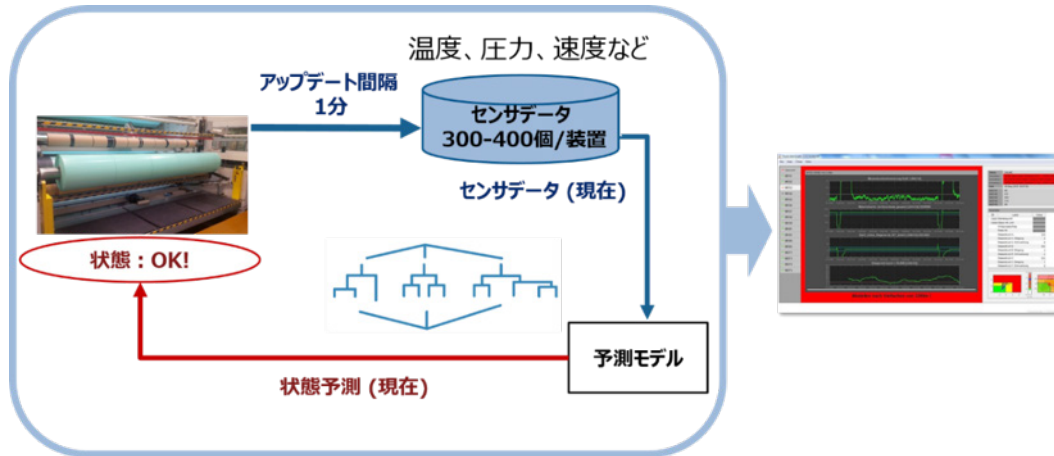


図3: 事前に学習した予測モデルでリアルタイムに機器の状態を予測する仕組みをアプリケーション化し工場に展開

MATLAB を使った予知保全ワークフロー

Mondi の実施した方法は一例であり、残念ながらすべての用途に対してそのまま適応できるわけでない。データの集め方、データの種類、サンプリング周波数などが異なれば、前処理に必要な作業も状況に応じて適切に変更する必要があるだろう。最も正確な予測を実現する機械学習アルゴリズムも異なることが想定される。更には予測を行うプログラムをアプリケーションではなく、C言語化し機器に直接組み込むことが求められるかもしれない。このように、用途によって最終的な方法は様々であることが想定されるが、予知保全を実現するために必要なステップを一般化してみると下記の4項目に落とし込むことができる(図4)。



図4: 予知保全システム構築に必要なワークフローをカバーするMATLAB プラットフォーム

それぞれMondiのケースに当てはめると:

- データへのアクセス: データが蓄積されるデータベースへアクセスしデータを吸い出す作業
- データの前処理: 異常値削除・変数選定・時刻の同期処理
- 予測モデルの構築: 複数の機械学習アルゴリズムの適用・評価
- システムへの統合: アプリケーション化して工場に展開

Mondiは予知保全システムを構築するツールとして、図4のすべてのステップにおいてMATLAB® を使用して開発を行った。当時Mondi では機械学習の専門的な知識を持つデータサイエンティストは存在しなかったが、わずか数ヶ月で予知保全システムの開発に成功している。MathWorks が提供するMATLAB は、データ解析、モデリングとシミュレーション、可視化、そしてプログラミング機能を提供する統合開発環境である。データ解析機能として信号処理や画像処理、統計解析、機械学習、最適化などの機能を備えており、さまざまな業種・業態の複雑なデータ解析をシンプルに実現することが可能だ。以下にそれぞれのステップについて説明する。

1. データへのアクセス

機械のログデータはデータベースに蓄積するだけでなく、CSVやMicrosoft® Excel® ファイルなどに書き出している場合も多く見かける。場合によっては画像データをもとに故障の兆候を捉えることも有効だ。MATLAB は1つの統合環境で次のような様々なデータソースを予知保全に活用することができる。

- データベース (ODBC/JDBC準拠)、データウェアハウス、分散ファイルシステム (Hadoop)
- OPCサーバ内の産業プラント関連のリアルタイムデータおよびヒストリカルデータ
- テキスト、スプレッドシート、XML、CDF/HDF、動画像、音声

例えば特定のデータロガーから生成されるバイナリファイルは、他のソフトウェアで解析を行う場合、専用ソフトウェアにより Excel ファイルに書き出す必要があることが多い。これらをMATLAB から直接読み取らせることで、ソフトウェアをまたぐ作業を省き工数を大幅に削減することも可能になる。

2. データの前処理

データの前処理は、この作業だけで数日を要してしまうことがあり、大規模な時系列データを扱うエンジニアから最も削減したい工数の一つであると言われることが多い。下記に代表的な前処理事業例を列記する。

- リサンプリング: 計器によりサンプリング周期が異なる場合、同じタイムスタンプにリサンプリングする。
- 欠損値/異常値処理: 計器の異常もしくはシステムの停止等の通常運転時以外のデータを処理する。
- ノイズ除去: プロセスデータは、例えば流量計では気泡の混入等で大きく測定値が振れる場合がある、これらのノイズをフィルタなどで除去する。

これらの作業はデータ規模が大きくなると手作業では非常に工数がかかり、処理の自動化が求められる。また、見逃されがちだが機械学習アルゴリズムはノイズと意味ある重要な情報とを自動で区別できるほど賢くない。異常値を無視してよいのか、あるいは、その異常値はモデルが考慮すべき現象(すなわち機械の故障や異常)を表しているのかを見極める必要もあるので、データの分析には必ず監視対象の機器に詳しい技術者の知見や経験が求められる。

3. 予測モデルの構築

機械学習アルゴリズムは何十種類もあり、特徴もそれぞれであり何にでも万能に使えるものはない。Mondi が行ったように、最適な手法を見つけるには個々の課題について様々な手法を試してみて評価する必要がある。更には、機械学習アルゴリズムを決める上では予測精度だけではなく、学習速度、メモリ使用量や予測結果の解釈のしやすさも重要な要素になってくる。例えば予測のアルゴリズムを機器に組み込むことを考えると、メモリを大量に必要とする計算負荷の高いアルゴリズムは適さない。新しく得られる機器からのデータを使って、予測モデルを都度アップデートすることを考えるならば、再学習にかかる時間も検討しなければならない。場合によってはブラックボックスでの予測ではなく、予測結果がある程度人間にも解釈可能であることという要求もある。結果、アルゴリズムの選択はそれぞれが持つ特徴の間でトレードオフを探る道程となる。

ただ、用いるべき機械学習アルゴリズムはデータの種類、予測したい内容によってある程度決まってくる部分もある。機械学習のアルゴリズムは、まず用意できるデータの種類によって大きく「教師あり学習」と「教師なし学習」に分けられる(図5)。Mondiのように過去のログデータに予測すべき結果(正常・異常)の情報が存在する場合は、教師あり学習。ただ、予知保全や故障予測を考えた場合、異常時のデータを用意できない場合も多い。その場合はデータをクラスターに分けるなどの教師なし学習のアプローチを選択する。教師あり学習はさらに「回帰」または「分類」のかたちをとる。

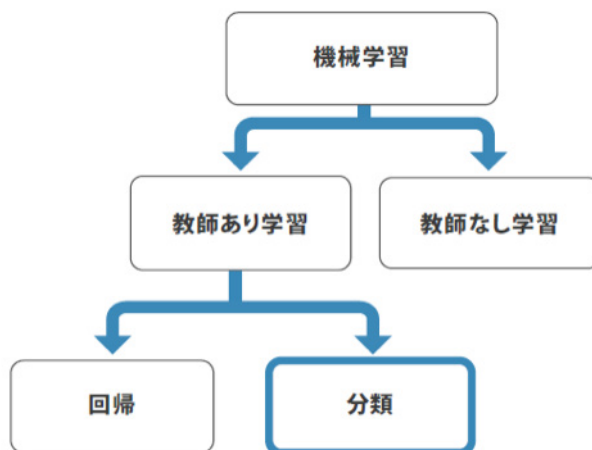


図5: 機械学習アルゴリズムの種類

分類では離散的な応答を予測する。例えば機械が正常に動いているか、異常を示しているかといった場合で、データを特定のグループやクラスに分類するための学習を行う。回帰では一定のレンジをもつ連続的な応答を予測する。例えば、温度変化や機器が故障するまでの時間、電力需要の変動などを予測する場合に使用される。

最後に多少込み入った話になるが、予測モデルを改善する方法の1つ「特徴選択」を紹介する。関連性が最も高く、予測力が最も高い特徴量(すなわち変数)を特定する作業を指し、これは特に数多くの変数を持つ多次元データを取扱う際に役立つ。特徴量の数を減らすことにより、予測精度を高めるだけでなく、必要なメモリ容量を減らしたり、処理時間を短縮したりすることができ、予測結果も解釈しやすくなる。一般的な特徴選択の手法として以下がある。

- ステップワイズ回帰: 予測精度に改善が見られなくなるまで、逐次的に特徴量の加減を行う。
- 逐次特徴選択: 特徴量の加減を反復的に行い、それぞれの変更がモデルの性能に及ぼす影響を評価する。
- L1正則化: 縮小推定を用い、予測に関係ない特徴量の重み(係数)をゼロにしモデルへの影響を取り除く。
- Neighborhood component analysis(NCA): 出力を予測する際に各特徴量の持つ重みを見つけ出し、重みの低い特徴量を除外する。

ここまでアルゴリズム選択の指針、そして予測モデルを改善する手法について紹介した。MATLABには上記の様なデータの前処理や特徴選択を行う関数ライブラリが多数用意され、欠損値処理などもプログラム1行で実現できる。また複数の機械学習アルゴリズムを検証することを目的にしたアプリ(図6)も用意されており、これらもMondiの開発の速さに寄与した。

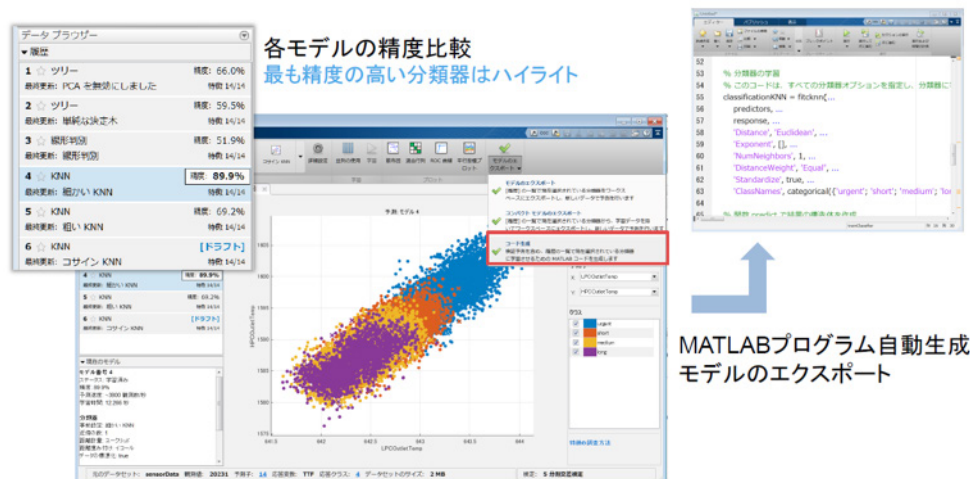


図6:分類学習器アプリは、データを分離するモデルの学習と検証を簡単に実行できるGUI

4. システムへの統合

前のステップまでに構築された予測モデルは、最終的に予知保全システムへと実装される。Mondi では予測モデルは状態監視システムの一部となり、アプリケーション(exe)化され、工場に展開された。実際の運用方法を考えると他にも様々な選択肢が見えてくる。予知保全システムを活用するのは整備スタッフか、機器のオペレーターなのか。またPC上で動くアプリが適切な場合もあれば、タブレットで持ち運べる実装がよい場合もある。確認するエンジニアが多い場合、サーバに実装し外部からWebアクセスして状態を確認する方が適切かもしれない。

MATLAB で開発した機能は、コードを書き直すことなく、Java、.NET、Excel、Python、C/C++ など、幅広いプログラムと連携できるコンポーネントとしてパッケージ化できるという点でも開発工数削減に寄与する。PC上で起動する単独のアプリケーション(exe)や共有ライブラリ (DLL、Javaクラス)として共有するだけでなく、Excel上ですべてを完結できるようExcel のアドオンを生成することも可能だ。また、Apache™やMicrosoft IISといったWebサーバやSOAP/RESTといったWebサービスを連携させ、Webアプリケーション、エンタープライズアプリケーションの一部として実行することも可能である。

まとめ

Mondi を例に、予知保全システム開発のワークフローについて紹介した。予知保全に限らず一般的にデータ解析を進める上では前処理や手法の選定・評価など多くの試行錯誤が求められ、このステップをいかに効率よく実行できるかが開発期間短縮のカギを握る。機械学習を適用する部分だけが注目を集めがちだが、データはどこから来るのか、そして最終的にアルゴリズムはどう実装するのか、それらの点も踏まえてシステム全体を考えた開発が重要になる。その際には必ず監視対象の機器に詳しい技術者の知見や経験が求められ、場合によっては部門間の協調も重要になる。MATLAB をアルゴリズム開発からエンタープライズシステムへの統合まで、プラットフォームとして組織的に利用することにより、よりいっそう資産共有及び再利用が加速され、開発効率向上及びメンテナンスコスト削減が期待できる。

関連リソース

[予知保全 解説ページ](#)

[Predictive Maintenance Toolbox™ 製品ページ](#)