

機械学習による製品品質影響因子の特定

AI品質ナビゲーションシステムへの適用

Identification of factors affecting product quality with machine learning: Application to AI quality navigation system

渡邊 拓朗

Takuro Watanabe

平 昌和

Masakazu Taira

キーワード

機械学習, 品質影響因子, 特徴量生成, 特徴量選択

デジタル化の進展に伴い、製造業においても様々なデータ分析の取組みが行われている。特に、操業データを分析し、品質に影響を与えている因子を特定することは、顧客の要求の高度化・多様化を背景に、より高い品質レベルやイノベーションが求められているため、重要性が増している。しかし、これを行うには製造プロセスに関する知見と大量の操業データを扱うためのスキルの両方が求められる。特に多品種を生産しているプロセスにおいては品種ごとに分析する必要があり、膨大な時間と労力を要する。そのため、それらの低減が課題となる。この課題に対応するため、操業データから運転の状態を表す特徴を抽出し、その中から品質影響因子を自動的に特定する技術を開発した。本技術を用いて、特に品質影響因子の自動特定が困難なバッチプロセスを対象に、品質影響因子の探索機能と傾向監視、および品質異常時の自動調査機能を備えた Deep Anchor™を開発した。

With the advancement of digitalization, data analysis has become widely utilized in the manufacturing industry. In particular, identifying quality-influencing factors that influence quality has gained importance due to the demand for high-quality products that meet the sophisticated and diverse requirements of customers. However, this requires expertise in both manufacturing processes and data analysis. In particular, in manufacturing processes producing a wide variety of products, analyzing data for each product individually demands considerable time and effort. A key challenge is to reduce this burden. To address this, we have developed a technology that extracts features representing operating conditions from operational data and automatically identifies quality-influencing factors. Using this technology, we have developed Deep Anchor™, which is equipped with functions for exploring quality-influencing factors, monitoring process condition, and automatic investigation of the causes of quality anomalies, specifically targeting batch processes where the identification of quality-influencing factors is challenging.

1. はじめに

今日の製造業には、高度化・多様化する顧客要求に対応するための高い品質レベルやイノベーション、競争力を維持するための製造コストの低減、カーボンニュートラルをはじめとした社会的要求への対応など、様々な要件を満たすことが求められている。そのため、適切に製造条件を変更していくことが必要であり、また同一条件下での製造においても、設備特性の経年変化などにより製造条件と品質の関係性も変化していくため、製造条件の見直しが必要となる。しかし、各条件が品質にどのように影響を与えているかは必ずしも明確ではなく、条件変更による影響の大きさを定量的に評価することは困難である。

これに対する一般的なアプローチとして、温度や圧力といったセンサデータや原料組成などの運転条件の操業データの解析があるが、多品種を生産している製造現場においては、品種ごとに分析を行う必要があり、特に分析の負荷が高い。加えて、操業データの解析には、製造知見に加えて、統計解析の知見が必要であり、人材の育成も課題となる。

この課題を解決するために、操業者が短時間で操業データに含まれる品質に影響を与える因子を特定し、その影響の大きさを定量的に評価する技術を開発した。本稿では上記技術と、この技術を用いて、品質影響因子の探索と傾向監視、および品質異常時の自動調査を行うDeep Anchorについて報告する。

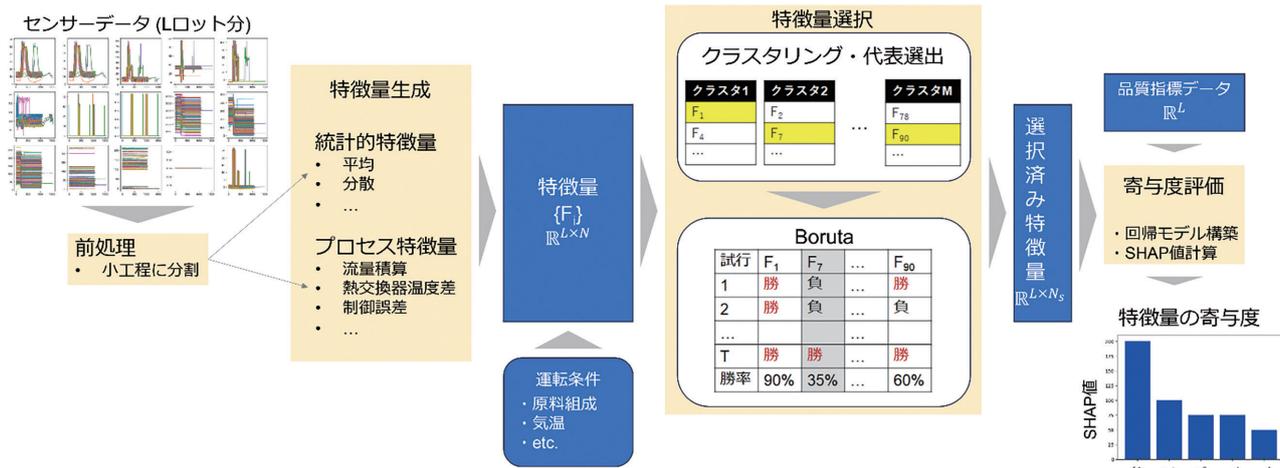


図1 提案技術の大まかな流れ

2. 提案技術

本技術の目標は、バッチプロセス^{注1}において操業データと製品品質のデータとの関係性を自動的に分析し、品質管理の改善につながる情報を抽出することである。操業データは主に多数の物理量の測定値から成り、これらの変数の品質指標への影響を評価する。変数同士の関係性を分析する方法として、回帰分析に基づく方法が広く用いられている。操業データを入力、品質指標データを出力とするモデルを構築し、入力変数の推定への影響の大きさを評価することで、品質に大きく寄与する入力変数を特定できる。線形回帰⁽²⁾、非線形回帰⁽²⁾、多変量回帰⁽²⁾等多様なモデルから適切に選んで活用すれば、様々な関係性に対応できる点が長所である。センサーデータ等時系列データを用いる場合、長時間の時系列データを入力とするのは難しいため、センサーデータの特徴を表す特徴量を入力とする場合が多い。近年では、Auto Encoder⁽³⁾やTransformer⁽⁴⁾などの深層学習を用いて入力の特徴を学習し、精度の良い回帰モデルを構築する研究がなされている。ただし、バッチプロセスの品質推定に回帰分析を適用するには、下記2点の課題がある。

1点目は、限られたデータから、品質に関連する情報を適切に抽出することである。バッチプロセスの品質指標データは生産単位(ロット)ごとに1度しか取られないことがほとんどである。この限られた品質データのみが得られている状態で、品質と関係が強い特徴量を適切に抽出することは困難が伴う。

2点目は、解釈性の向上である。回帰分析では、品質推定への影響が大きい入力変数を特定する。この情報を製造プロセスの改善に生かすには、その入力変数のプロセス上での意味を解釈する必要がある。入力変数の意味が不明瞭では、品質に影響することが分かっても具体的な対策の立案は難しくなる。

特に前述した深層学習を用いた方法は学習に大量のデータを必要とし、出力される特徴量の解釈も難しいため、この

目的での使用には問題が多い。

そこで、本技術では製造プロセスの知見と機械学習手法を組み合わせ、上記2点の課題に対応しつつ分析の自動化を目指す。まず、上記2点の課題を解決するため、操業データから、プロセスの知見に基づいた特徴量を生成する。人間が監視する際に注目するようなポイントに基づいた意味を持つ値を特徴量とすることで、効率よく品質に関わる情報を抽出できると期待される。この特徴量を回帰モデルの入力に用いると、少ないデータからでも精度よく品質との関連性を学習できると考えられる。また製造プロセス上、明確な意味を持つため、特徴量の解釈性も向上する。この分析を自動的に行うため、本技術ではまずプロセス知見に基づく典型的な特徴量を網羅的に生成し、その後、機械学習手法を用いて、品質に関係する特徴量のみを抽出する。

図1に、本技術の大まかな流れを示す。Lロット分のデータが取得されているとする。Lロット分それぞれに対し、一連の運転における各種センサの時系列データと原料組成などの運転条件、品質指標データが存在する。まずセンサーデータに前処理を施し、その後は特徴量生成、特徴量選択、寄与度評価の3つの手順を実施する。次節以降ではこれら3つの手順について説明する。まず、2.1節ではセンサーデータから特徴量を生成する方法について述べる。次に、2.2節にて生成した特徴量の品質への寄与の有無を判定し、寄与があるもののみを選択する手法を解説する。最後に、2.3節で選択された特徴量を用いて機械学習モデルを構築し、推定値への寄与を定量的に評価する方法について述べる。

2.1 特徴量生成

本節では、ロットごとにセンサーデータの特徴量を生成する手順を述べる。まず、前処理として時系列データの分割を行う。一般に製造プロセスは原料仕込みや反応、乾燥といった複数の工程に分かれており、それぞれの工程での処理が製品品質に影響する。そのため品質に影響する因子を探索する際は、工程ごとに分けて分析することが望ましいといえる。そこで、各ロットのセンサーデータを工程ごとに分割する。そして分割した各工程のデータに対して数理的な処理を施して、特徴量を生成する。数理的処理は大きく2種類に分けられる。

注1 製品の生産単位で製造プロセスを運転する生産方式。製造条件の調整をしやすいため、高機能製品の製造に広く用いられる。

1つ目は、平均や分散、最大値といった一般的な統計量である。これらの特徴量は、データの一般的な性質を抽出する目的で用いる。

2つ目は、製造プロセスの知見に基づいて設定した本技術独自の計算処理である。製造プロセスで測定されるデータは温度、流量、圧力などの物理量ごと、計測対象の装置ごとに、それぞれ監視すべき観点が存在する。本技術では、これら物理量や装置の性質を考慮して特徴量を設計する。例えば、流量の積算値はその装置に流入、流出した流体の総量を示す。温度の傾きや極値は、化学反応における温度制御の精度に関連していると考えられる。また、複数のセンサーデータを複合した特徴量も考えられる。例えば熱交換器の入口と出口の温度差は、冷却、昇温過程で出入りした熱量を示し、熱交換系の状態を表しているといえる。また変数の実測値と制御における目標値の偏差および偏差が十分小さくなるまでに要した時間を特徴量として、変数の制御性能を考慮している。

このようなプロセス操作の上で意味を持つと考えられる値は、様々な工業プロセスで共通して定義できる。本技術ではこれらを求める計算式をマスクとして保存しておき、実際に観測されたデータに適用して特徴量を生成する。その他、原料組成などバッチごとに決まる運転の特性値も特徴量として用いる。これらの特徴量のうちの特徴量が品質に関連するかはプロセス次第であるが、この段階においては考えられる特徴量を網羅的に生成する。品質に関連のない特徴量が生成されても、後の処理で自動的に除去されるためである。

2.2 品質に寄与する特徴量の選出

前節で述べた処理では、品質に関連する可能性がある特徴量を網羅的に生成した。ただし、一般に品質へ明確な影響を持つ特徴量は一部のみである。入力変数が過剰に存在すると精度の良い回帰モデルは構築し難いため、まずは品質に関係がある特徴量のみを選出する。この特徴量選択の処理は大きく2つの手順に分けられる。以降、それぞれの手順について解説する。

2.2.1 クラスタリングによる冗長特徴量の除去

前述した特徴量生成処理では、プロセス上意味を持つ可能性がある特徴量を網羅的に生成する。そのため、結果として同じような挙動を示す特徴量が生成される可能性もある。それらの冗長な特徴量はモデル精度を悪化させることがあるため、一つに絞ってから次の手順に進むことで、モデルの精度向上を図る。本技術では階層型クラスタリング⁽²⁾を用いて類似した特徴量を分類し、各クラスから一つずつ代表を選出することで冗長な特徴量を削除している。各クラスの代表は、品質指標との相関係数が最も高い特徴量とする。

2.2.2 Borutaによるスクリーニング

本項で解説する処理の目的は、品質に関連を持つ特徴量の抽出である。そのために、品質指標を出力とする回帰モデルにおいて、推定の精度向上に寄与できる特徴量を選択す

る。本技術では、寄与の有無を一貫した基準で判定できる特徴量選択手法として、Boruta⁽³⁾を採用した。Borutaは決定木⁽²⁾を利用した回帰モデルに基づく特徴量選択手法である。まず、決定木を利用したモデルについて、推定の原理と説明変数の寄与の評価法を述べる。決定木を用いたモデルの一種である、Random Forestのモデル構造を図2に示す。

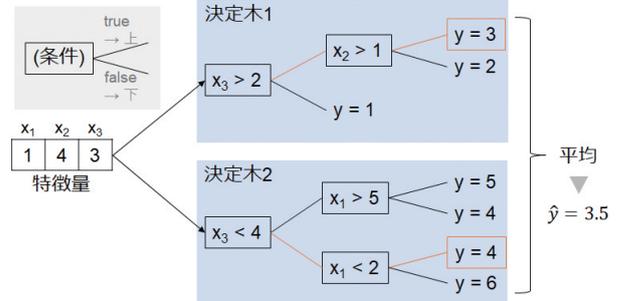


図2 Random Forestモデルの例

決定木は、入力された入力変数のうち一つの値を用いた条件分岐を繰り返し、分岐結果のパターンごとに推定値を割り当てるモデルである。Random Forestは乱数により使用するサンプルや変数を変えながら多数の決定木モデルを構築し、それらの推定値を統合して最終的な推定値を出力する。図2の例では2つの決定木モデルの推定値の平均を最終的な推定値としている。

決定木を用いたモデルには分岐への寄与を表すFeature Importance (FI)⁽²⁾という指標が提唱されており、各入力変数の寄与が数値化される点が特徴である。ただし、FIの値から寄与の有無を判断する定量的なしきい値は確立されていない。

Borutaでは上記問題に対応するため、寄与がない偽の特徴量との比較を行うことで、寄与の有無を一貫した基準で判断する。Borutaのアルゴリズムを図3に示す。

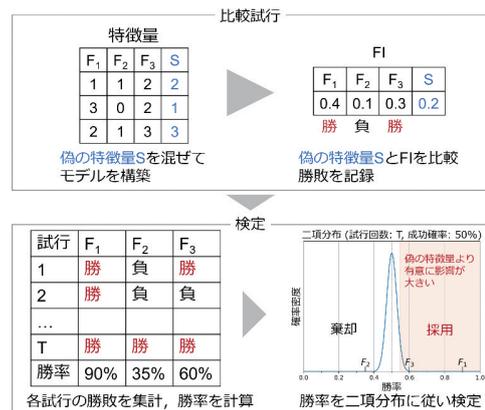


図3 Boruta概念図

まず、本質的には推定に寄与しない、偽の特徴量をランダムに生成する。その偽特徴量を本来の特徴量と混ぜてモデルを構築し、各特徴量の寄与の大きさを比較する。この処理を繰り返し試行し、偽特徴量と寄与の優劣を比較し、勝敗の結果(図3, 優: 勝, 劣: 負)を記録する。そして、「勝率(特徴量が偽の特徴量より寄与が大きい確率)は50%という

帰無仮説の下で二項検定を行う。この帰無仮説は、特徴量と偽特徴量の間で寄与に優劣がない、つまり、特徴量に寄与がないことを示す。この仮説が棄却されるなら、特徴量は偽特徴量より有意に大きな寄与を持つといえる。この仮説を棄却できるか否かで特徴量を取捨選択することで、しきい値の設定等を行うことなく、寄与の有無を一貫した基準で判定できる。

2.3 寄与の定量的評価

本節では、前節で選出した特徴量を用いて品質を推定する回帰モデルを構築する手法、そのモデルを解析して品質推定における各特徴量の寄与の大きさを評価する方法について述べる。寄与の大きさの評価手法として、本技術ではTreeExplainer⁽⁶⁾を用いた。TreeExplainerは回帰木を用いたモデルに対してSHAP (SHapley Additive exPlanation) 値⁽⁷⁾を定式化した評価指標である。SHAP値は回帰モデルの推定値への各入力変数の寄与を定量化する手法であり、各サンプル(入力変数、出力変数の組。本技術では1ロット分の特徴量と品質指標の組)において、推定値の期待値からの変動を各入力変数の寄与として割り当てる。

まず、SHAP値の一般的な性質と本技術での解釈について解説する。SHAP値の概念を図4に示す。

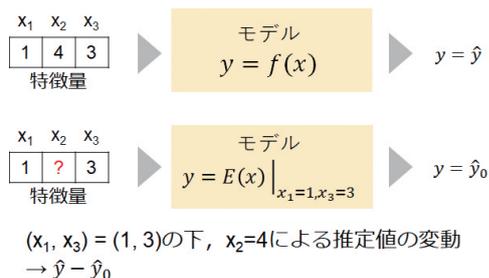
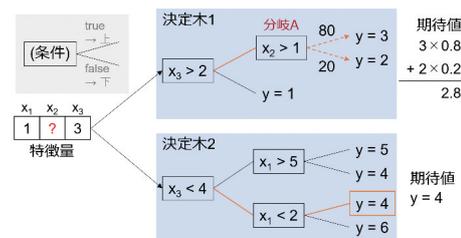


図4 SHAP値概念図

図4において、 x は入力となる特徴量、 y は出力となる品質指標、 f は回帰モデル、 E は f の期待値を表す。SHAP値はモデルの特徴量の一つに注目し、その変数の値が未知の場合の推定値の期待値を基準として、その変数による推定値の変動量を示す値である。図4の例では x_2 に注目し、 x_1 と x_3 のみが与えられた場合の推定の期待値と x_1, x_2, x_3 全てが所与の場合の推定値との差分を計算している。固定する変数の組み合わせを変えながらこの差分を求め、それを平均したものがSHAP値となる。

次に、TreeExplainerの詳細について述べる。回帰木は図2に示した通り、特徴量の値に応じた分岐を重ねた木構造を構築し、分岐の結果ごとに一つの推定値を割り当てる回帰モデルである。TreeExplainerでは注目している特徴量を用いた分岐全てについて、それぞれの分岐先へ進む割合を仮定して期待値を求める。その割合は、学習に用いたデータが実際にそれぞれの分岐先に進んだ割合とする。TreeExplainerにおける推定の期待値の考え方を図5に示す。図5の例では x_2 に注目し、 x_2 を用いた分岐(図中分岐A)を、各分岐先に進んだ学習データ数に基づいて重みづけする。



分岐Aにおいて、学習データの判定結果がtrue:false=80:20とする
 決定木1の期待値: $3 \times 80/100 + 2 \times 20/100 = 2.8$
 決定木2の期待値: 4

$\rightarrow x_2$ が未知の場合の期待値 $E(\hat{y}) = (2.8 + 4)/2 = 3.4$

図5 TreeExplainerにおける期待値の考え方

いま、学習データが100サンプル存在するとして、学習時に分岐Aの上下分岐先に進んだサンプル数をそれぞれ80、20とする。この時、分岐Aにおける分岐結果を上記重みづけ分岐で置き換え、期待値を求める。図5の場合、推定の期待値は3.4となる。つまり、 x_2 の値が未知の場合の推定の期待値は3.4となる。一方図2に示した事例では、 $x_2=4$ としたときに推定値が3.5となっている。これらより、このサンプルにおいては「 x_2 の値が4である場合に、期待値より推定値が0.1増加している」といえる。この変化の大きさが、このサンプルにおける x_2 のSHAP値となる。

このように、SHAP値は各サンプルの特徴量に対して、それが推定値に及ぼした変化の大きさを評価する指標である。この値は特徴量の値そのものだけでなく、品質指標等にも依存して変動する。これらの値に対するSHAP値の傾向を分析することで、全体的な品質への影響の大きさだけでなく、ある条件下での局所的な影響も分析できると期待される。例えば「異常なロットにおいてSHAP値が大きい特徴量は、品質異常の原因に関わるため検査をする」、「ロット全体にわたってSHAP値が大きい特徴量は、品質に大きく影響する主要因子であるため特に注視する」というように、SHAP値の傾向に応じて適切なアクションを立案できる。

3. 品質影響因子発見システムDeep Anchor

前章で述べた技術を用いて、品質影響因子発見システムDeep Anchorを開発した。Deep Anchorは品質影響因子の探索と傾向監視、および品質異常時の自動調査を行うアプリケーションソフトウェア製品である。本技術により分析は自動化されているため、データ分析のための専門的なスキルは必要なく、製造プロセスの知見を有する製造部門や品質管理部門のユーザーが独力でモデル構築を行うことができる。

3.1 システム構成

Deep Anchorは、コンフィギュレータ、サーバ、ビューアの3種類のノードで構成される(図6)。コンフィギュレータでは品質推定回帰モデルの構築を行う。ユーザーは、モデルの設計・実装から評価までの一連の作業を、コンフィギュレータ上にてワンストップで実行することができる。コンフィギュレータで構築されたモデルを使用し、サーバとビューアがオンラインで傾向監視と原因調査を行う。

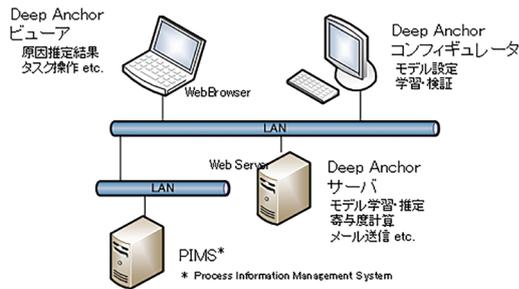


図6 Deep Anchor システム構成

3.2 主要機能と出力例

Deep Anchorは3つの機能を有する(図7)。以降、高機能樹脂製造プロセスにおける製品粘度の因子について出力した結果を用いて、各機能を解説する。

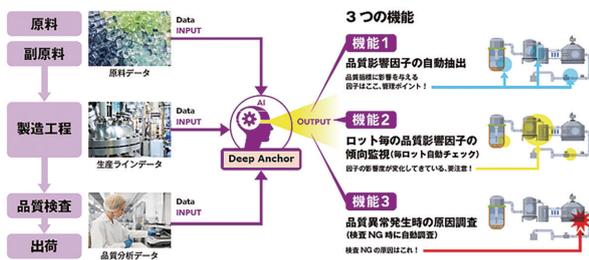


図7 Deep Anchor の機能

3.2.1 機能1(品質影響因子の自動抽出)

機能1は、品質影響因子の自動抽出である。過去の操業データおよび品質指標データから、品質に影響を与える特徴量を因子として示す。また各因子と品質指標の関係性を可視化することで、各因子の適切な範囲である制御推奨範囲を示す。これらの情報により、生産条件の確立までの時間を短縮し、新製品の早期市場投入に貢献する。

図8に、出力の一例を示す。図8左の棒グラフは抽出された因子の一覧である。棒グラフの長さは各因子の全ロットにおける平均SHAP値であり、各因子が品質指標に与えた全体的な影響を示している。この例では、粘度に最も影響を与えやすい因子が反応工程における反応槽温度であることが

わかる。図8右の散布図は粘度と上記因子のSHAP値を、全ロットに対してプロットしている。この例では、反応槽の温度が70℃を下回ると大きく粘度が落ちることを示している。

3.2.2 機能2(品質影響因子の傾向監視)

機能2は、品質影響因子の傾向監視である。機能1で抽出された因子の平均SHAP値や制御推奨範囲を、新たに運転されるロットにて監視することで、品質異常が起こる前に設備特性の経年変化等による変化を検知する。結果は画面上的アラーム、メールでユーザーに通知され、品質トラブルを未然防止、ロス(原材料、エネルギー、 manpower)と廃棄品の低減に貢献する。

図9は実際に傾向監視のアラームが出力された例である。図内の棒グラフは、因子の一つである原料Aの投入にかかった時間(仕込み時間)の平均SHAP値が増したことを示している。右グラフの散布図は機能1と同様に、品質指標と選択した特徴量の平均SHAP値のプロットである。今回のロット(オレンジ)と前ロット(青)において、過去の製造実績より原料Aの仕込み時間が長くなっており、粘度も下がっている。このことから、仕込み時間の増大が粘度低下の原因となっていると推察できる。つまり粘度低下の異常を防ぐには、原料Aの仕込み時間増大の原因を調査し、対応をするべきと示している。

3.2.3 機能3(品質異常発生時の自動原因調査)

機能3は品質異常発生時の自動原因調査である。この機能は、品質異常ロットを含めたデータに対して、図1にて示した一連の工程を実施し、異常ロットに対する各特徴量のSHAP値を評価することで実現している。一般に、異常発生時の原因調査にはスキルと時間を要するため、担当者の負担も高く、調査完了までには時間がかかる場合が多いが、Deep Anchorは品質異常を検知すると15~20分程度で原因を推定してユーザーにメールで通知する。推定した異常の原因は2.1節にて述べたプロセスの知見を考慮した特徴量であり、さらにその意味が文章にて出力されるため、人から見て納得性が高い。

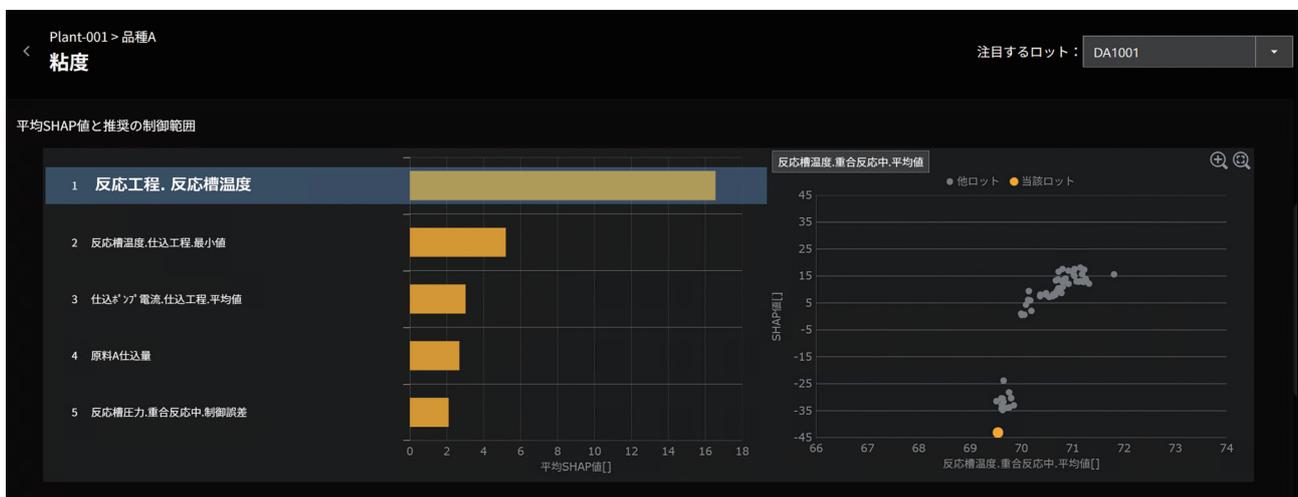


図8 品質影響因子と制御推奨範囲



図9 品質影響因子の傾向監視

図10は品質異常時の出力結果である。棒グラフは各特徴量のSHAP値を表しており、棒が赤い場合は正、青い場合は負の影響を持つ。乾燥工程における温度制御の整定時間という特徴量が、粘度に対して大きな負のSHAP値を持つことを示している。この結果を文章で示すとともに、それがどの程度粘度を下げたのかを棒グラフで示している。右の青い5つのエリアは選択式になっており、Deep Anchorが示した異常原因に関する時系列データ、特定した因子の管理図、過去の類似トラブル情報など、品質異常を調査する際に必要になると考えられる情報をダイレクトに呼び出せるようになっている。

3.3 自律型・品質管理システムへ

当社はDX（デジタルトランスフォーメーション）の基盤として、自律化（Autonomy）の技術を確立し、現在の生産工場において人が管理して行われている各マネジメント領域（生産管理、品質管理、設備管理など）に自律化システムとして適用することを目指している。自律化システムは、ミッションを与えられたAIが、人の介入を最小限に抑えながら、人よりも迅速かつ正確に遂行する仕組みである。図11は自律化システムを品質管理領域に適用した自律型・品質管理システムの機能で、Deep Anchorは、この中の機能1, 2, 3を担う製品である。



図10 品質異常の原因調査

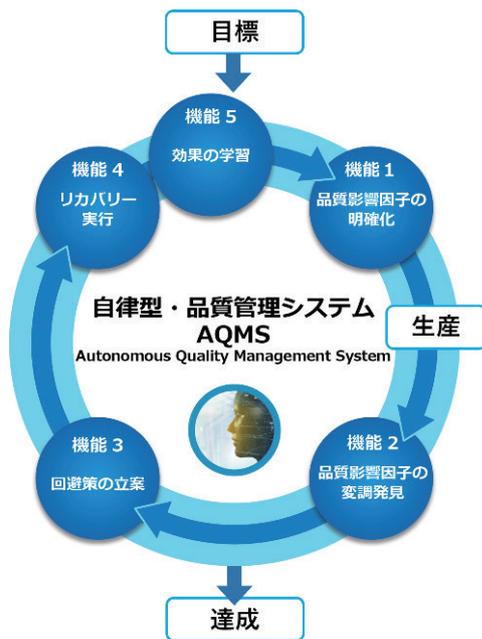


図11 自律型・品質管理システムの機能

(7) Lundberg Scott M., Lee Su-In: A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems 30, 2017年, pp.4765-4774, Curran Associates, Inc.

<商標>

Deep Anchorはアズビル株式会社の商標です。

<著者所属>

渡邊 拓朗 アズビル株式会社

AIソリューション推進部

平 昌和 アズビル株式会社

アドバンスオートメーションカンパニー

戦略事業開発3部

4. おわりに

本稿では、操業データから品質影響因子を特定する技術と、その技術を適用した製品Deep Anchorについて述べた。Deep Anchorの持つ3つの機能は、新製品の早期市場投入による社会イノベーションの加速、廃棄品やロスを減らすことによる環境負荷低減、品質管理担当者の時間的および心的負担の低減に貢献する。今後、本技術・本製品の拡張や新技術を組み合わせることで、自律型品質管理システムを完成させ、飛躍的な生産性の向上と持続可能な社会の実現に貢献していく。

<参考文献>

- (1) プラントにおける先進的AI事例集,
<https://www.kobe-kiankyo.jp/common/uploads/2021/06/20210611-news-021117-2.pdf>
(アクセス日 2024.10.25)
- (2) Trevor H, Robert T, et al.: 統計的学習の基礎 データマイニング・推論 予測, 2014年, 共立出版
- (3) Hinton, G. E., & Salakhutdinov, R. R.: Reducing the dimensionality of data with neural networks. Science, 2006年, Vol. 313 No. 5786, pp. 504-507.
- (4) Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems, 2017年.
- (5) Kursu, M. B., & Rudnicki, W. R.: Feature Selection with the Boruta Package. Journal of Statistical Software, 2010年, Vol. 36 No. 11, pp. 1-13.
- (6) Lundberg Scott M., Erion Gabriel, et al.: From local explanations to global understanding with explainable AI for trees, Nature Machine Intelligence, 2020年, Vol.2, No.1, pp.56-67, Nature Publishing Group

