

XAI を応用した時系列要因分析技術の提案 Time-Series Cause Analysis Method Utilizing eXplainable AI

横井 直明[†]
Naoaki Yokoi

恵木 正史[†]
Masashi Egi

1. はじめに

近年、製造業において様々なデータの収集・活用が活発化している。この収集した膨大なデータに機械学習などの技術を適用することで、製品品質の安定化や設備の異常検知など、業務上の課題を解決する流れも加速している。工場の高度化に向けたデータ利活用および分析容易化は、生産性改善やコスト削減の観点から重要性が高い。しかし、実問題への機械学習の応用では、単に高い予測精度を持つモデルの構築だけでなく、その判断根拠をユーザに提示することで品質低下や異常発生の原因特定までできることが要求される。これに対する技術として、モデルの判断根拠を分析する説明可能 AI (以下 XAI) が注目されている[1]。

そこで、筆者らは工場内で収集される各センサデータなどから製造品質の悪化を引き起こす要因の分析に XAI を応用することを検討している。本稿では、時系列データを対象に(A)品質悪化プロセスの分析手法と、(B)時系列変数間の因果関係分析手法を提案する。(A)の手法では「製造工程のいつのどの要因が、品質に対してどれくらい寄与するか」といった品質の悪化に至るプロセスの分析を実現する。これにより、「工程終盤の C 温度の低下を抑えよう」といった具体的な対策の立案を支援する。また、(B)では「直接的な原因である C 温度は、その 5 分前の B 圧力に応じて変化する」といった時間的な遅れも考慮した根本原因の特定を実現とする。これにより、(A)の分析結果として人が直接制御できない要因が抽出された場合でも、制御可能な根本要因まで特定できるようにし、例えば「工程終盤の 5 分前に B 圧力の設定を 1. はじめに下げよう」などといった対策立案を支援する。

2. 時系列データの要因分析への XAI 活用

時系列データを対象にした要因分析に XAI を適用する場合、まず、製品の不良度といった品質値とセンサデータとの関係性をモデルに学習させる。その際、センサデータなどの時系列データを説明変数とする一方、目的変数となる品質値は製造ロットごとの非時系列データであることが多い。一般的なモデルの学習時には、この時系列な説明変数と非時系列な目的変数間の粒度を統一する。例えば、「A 温度」や「B 圧力」といった時系列変数から、製造ロットごとに「A 温度の最大値」「B 圧力の平均値」などの統計量などに集約することで目的変数同様ロット単位の特徴量に変換して学習データを作成する。次に、その学習データから構築したモデルと XAI を用いて、分析対象の製造データに対して、各特徴量の不良度への寄与度を算出する。そして、不良度への寄与度が高い特徴量を抽出することで品質悪化の原因を特定するといった分析作業が通常行われる。しかし、今後の品質改善に向けて、このような要因分析の結果を具体的な対策に繋げるには以下の問題がある。

- **集約された説明変数の解釈の問題**：例えば「C 温度の平均値が高いこと」が品質の悪化原因として抽出できても、その対策として「常に C 温度を低く保つ」や「ある時点での C 温度の上昇を抑止する」などいくつも対策が考えられる。そして、どれが最適な対策なのか判断できない。
- **非可制御な要因抽出による問題**：例えば「C 温度の平均値」が原因として抽出されても、そもそも「C 温度」が人が直接制御できない要因だった場合、「C 温度」に影響を及ぼし、かつ人が制御できる根本要因を特定できないと具体的な対策が立てられない。

3. 提案手法

本稿では、時系列変数を統計量などに集約せず、いつ・どの要因がどれだけ品質低下に寄与したか、といった品質悪化プロセスを分析する手法を提案する。また、「C 温度」に対して t 時刻前のどの要因が影響するなど、時系列を考慮した因果抽出し、より根本的な要因の特定を支援する時系列変数間の因果関係分析手法についても提案する。

3.1 品質悪化プロセスの分析手法

図 1 を用いて品質悪化プロセスの分析手法を説明する。まず、時系列データに対して、図 2 に示す処理を適用し、時系列変数を時間変数化する。時間変数とは、各時刻・各時系列変数の値をロットごとにフラット化した変数である。こうすることで、時系列データを統計量に集約せずに時間×時系列変数の数だけ生成した時間変数をモデルの説明変数とする。しかし、生成される時間変数データの次元数は時系列変数の数×1 ロット当たりのデータ数と膨大となる。一般に、学習データの次元数が多いとモデルの過学習のり

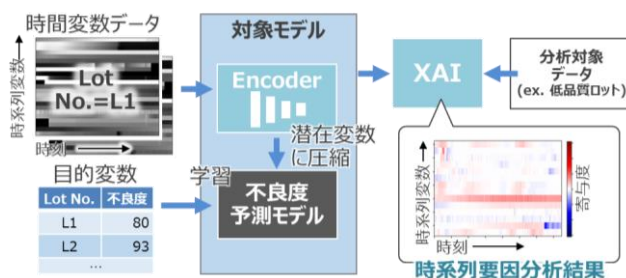


図 1 品質悪化プロセスの分析手法

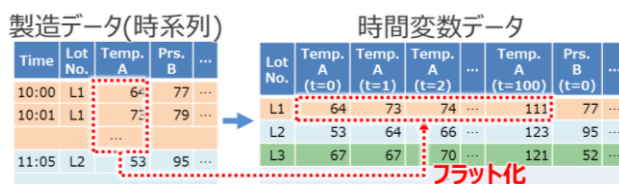


図 2 時間変数化の処理

[†]株式会社日立製作所 研究開発グループ Hitachi, Ltd.,
Research & Development Group

スクが高まる。そのため、事前に AutoEncoder で学習した Encoder を介することで、元の時間変数データから、その特徴を残しながらより低次元の潜在変数に圧縮することで、学習データを生成する。そして、潜在変数から製品の不良度を予測するモデルを構築後、分析対象の製造データなどに対して、各時間変数による品質への寄与度を XAI で算出する。その際、XAI への入力とするモデルは、不良度予測モデル単体ではなく、図 2 に示した対象モデル全体を一つのモデルとして入力する。これは、不良度予測モデルを入力とした場合、XAI は各潜在変数による不良度への寄与度を算出してしまい、人による解釈が困難となるためである。一方、Encoder と不良度予測モデルを含んだ対象モデルを入力とすることで時間変数に対する寄与度、つまり、各変数の各時刻が不良度にどれだけ寄与するかを算出できる。

3.2 時系列変数間の因果分析手法

次に、図 3 を用いて時系列変数間の因果分析手法を説明する。本手法では、ある時刻 t の各時系列変数の値(図 3 の ■ 部分)をその直前 τ 時刻間における全時系列変数(図 3 の全 ■ 部分)から予測するモデルをそれぞれ学習し、変数間の因果関係の抽出に用いる。ここで、 τ は直後の値を予測する際にどれくらいの期間のタイムウィンドウ幅を説明変数とするかを表し、図 3 ではタイムウィンドウ幅 $\tau = 4$ とした場合の例である。このように時系列変数ごとに学習したモデルに XAI を適用することで、例えば時刻 t の C 温度にはその 2 時刻前の Y 圧力による寄与が高いといった関係性を定量化でき、それらの結果をもとに Y 圧力($t-2$)→C 温度(t)などの因果関係をデータドリブンで抽出する。

4. 評価

提案手法の評価では、TEP から生成された時系列データセットを用いて、変数間の因果関係分析や品質悪化プロセスの分析を行った。TEP とは、米 Eastman Chemical 社が開発した化学プラントのプロセスシミュレータである[2]。TEP データには時系列変数として、試料の投入量や流量、反応槽の温度や圧力などがあり、一部の 14 変数についてはプラント構造に基づく理論的な因果関係が先行研究にて示されており[3]、これらの時系列変数を対象とした。

4.1 品質悪化プロセスの分析手法の評価

TEP データに基づき、特定の時系列変数を要因として算出した擬似的な不良度を目的変数とした。つまり、本手法によって、この意図的に組み込んだ疑似不良度の要因を正しく抽出できるかで有効性を評価した。疑似不良度は、時

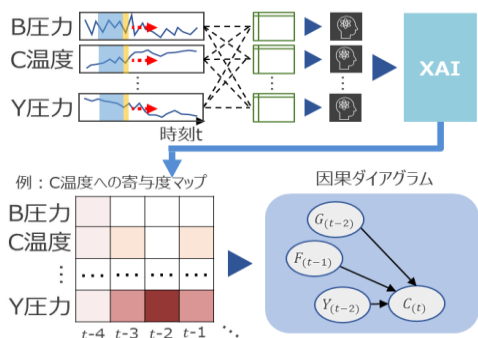


図 3 時系列変数間の因果分析手法

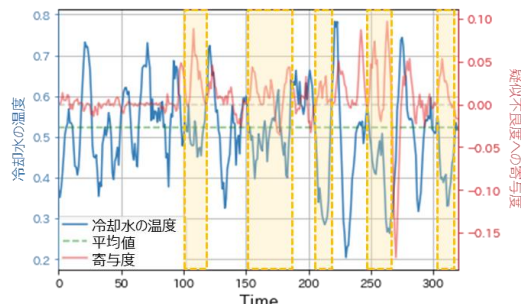


図 4 疑似不良度への寄与プロセス

系列変数から「冷却水の温度」などを要因とし、各時刻の値とそれら平均の差分値から算出した値をロットごとに合計することで各ロットについて算出した。図 4 に「冷却水の温度」が疑似不良度の増減にどう寄与したかを分析した結果を示す。「冷却水の温度」については、その平均値(図 4 の緑点線)よりも高いと疑似不良度を負の方向に、低いと正の方向に寄与するよう疑似不良度を算出した。そのため、図 4 内の青線と赤線が逆位相であることが期待され、その関係が大局的に成り立つ領域を黄色のハッチングで示した。ハッチング部分は全体の約 37.5%であり、完全一致の結果さえ得られなかったものの、本手法で期待される不良度への寄与傾向を一部抽出できることを確認した。

4.2 時系列変数間の因果分析手法の評価

TEP データに対し、先行研究で示されている理論的な因果関係を本手法で正しく抽出できるかで有効性を評価した。提案手法で各時系列変数について構築したモデルに対して XAI を適用し、寄与度が高い順に時系列変数のペアの上位 38 件を因果ありと判定した結果、理論的な因果関係の内 50%を抽出できるという結果が得られた。これは、従来の因果抽出手法である VAR-LiNGAM[4]の結果(上位 54 件で 50%抽出)と比べると因果抽出効率で 30%改善されるという結果であった。今回の TEP データは変数間の非線形な関係が強く、線形性を仮定する従来手法に対して、本手法では線形性を仮定しない機械学習モデルを用いたためより優れた結果が得られたと考えられる。

5. おわりに

製造業における時系列データを対象とした要因分析に XAI 技術を応用する技術について初期検討を行った。特に、品質悪化プロセスの分析手法および時系列変数間の因果分析手法の確立を目的に、提案手法の有用性を TEP データで評価した。その結果から、今後、両手法を組み合わせることで、時系列のままの原因分析および因果関係を辿ったより根本的な原因の特定といった、従来の XAI を活用した原因分析の問題点を解決できる見込みを得た。

参考文献

- [1] F. K. Došilović, M. Brčić, N. Hlupić, "Explainable artificial intelligence: A survey", MIPRO2018, pp.0210-0215 (2018).
- [2] J. J. Downs, E. F. Vogel, "A plant-wide industrial process control problem," Computers & Chemical Engineering, Vol.17, No.3 (1993).
- [3] X. Chen, J. Wang, and S. X. Ding, "Complex system monitoring based on distributed least squares method," IEEE Transactions on Automation Science and Engineering, Vol.18, No.4 (2021).
- [4] A. Hyvärinen, K. Zhang, S. Shimizu, P. O. Hoyer, "Estimation of a Structural Vector Autoregression Model Using Non-Gaussianity," Journal of Machine Learning Research, Vol.11, No.56 (2010).