TOSHIBA

大学共同利用機関法人 情報・システム研究機構
統計数理研究所

２０１９－８－２
株式会社　東芝
大学共同利用機関法人　情報・システム研究機構　統計数理研究所

大量の欠損を含むデータからでも不具合の要因を特定する機械学習アルゴリズム(AI)を開発
～品質低下や歩留悪化の要因を高速・高精度に特定し、製造現場の信頼性と生産性の向上に貢献～

株式会社東芝（本社：東京都港区、代表執行役社長：綱川智、以下「東芝」）と、大学共同利用機関法人 情報・システム研究機構 統計数理研究所（所在地：東京都立川市、所長：椿 広計、以下「統数研」）は、収集した製造データに多くの欠損値が含まれている場合でも、品質低下や歩留悪化などの要因を高速・高精度に特定する機械学習アルゴリズム「HMLasso（Least absolute shrinkage and selection operator with High Missing rate)」を開発し、最先端のアルゴリズム「CoCoLasso[注1]」と比べ推定誤差を約41%削減することに成功しました。本技術により、これまで活用の難しかった欠損値を多く含むデータでも高速・高精度な要因解析が可能となり、工場・プラントなど製造現場の生産性・歩留・信頼性の向上が期待できます。

東芝と統数研は本技術の詳細を、8月10日から16日に中国・マカオで開催される、AI分野で権威ある国際会議「The 28th International Joint Conference on Artificial Intelligence (IJCAI-19)」で発表するとともに[注2]、簡易プログラムをオープンソースソフトウェアとして8月2日より公開予定です[注3]。

工場・プラントなどの製造現場では、製造物の品質値や加工条件、設備の温度や圧力などの製造プロセスや設備稼働に関するデータが日々大量に収集・蓄積されています。これらのデータを活用し品質のばらつきを説明する回帰モデル[注4]を構築することができれば、品質や歩留が悪化する要因の特定と改善に大きく寄与することが可能となります。

しかし、実際に収集されるデータには測定ミスや通信エラーによる欠損が発生するだけではなく、抜き取り検査によって品質を確認することが多いため、1割程度しかデータを収集できない場合もあります。このような場合、予め欠損値を計算・補完してから解析するのが一般的ですが、欠損値が多いと膨大な計算が必要となり、要因解析の高速化・高精度化は困難でした。

そこで東芝と統数研は、欠損値の多いデータからでも高精度な回帰モデルを構築可能な新しい機械学習アルゴリズム「HMLasso」を共同開発しました。

本技術の特徴は以下の3つです。

（1） 欠損率が高い場合でも高精度に回帰モデルを構築

「CoCoLasso」は欠損率の高低を考慮しない設計のため、欠損率が高い項目に引きずられて全体の精度が下がってしまいます。一方、「HMLasso」は欠損率の高低に応じて柔軟に計算する設計のため、欠損率が高い項目があっても全体の計算精度が低下せず、高精度な回帰モデルの構築が可能です。

（２） 欠損値の補完プロセスを省略
欠損値を含むデータから直接、回帰モデルを構築することを可能とし、全体の計算時間を短縮します。

（３） 重要項目の自動絞り込み
データ項目が多い場合でも分析を実現するスパースモデリング技術[注5]の応用により、多くのデータ項目から品質や歩留への影響度の高い重要な項目だけを絞り込みます。

本技術の有効性は、理論と実験の両面から検証が完了しています。理論解析では、欠損率を活用することで誤差限界が最適になり、従来のアルゴリズムよりも優れていることを検証しました。数値実験では、平均欠損率 50%でデータ項目によっては欠損率が 90%以上となる人工データでベンチマークし、最先端のアルゴリズム「CoCoLasso」と比べて推定誤差を約 41%削減することに成功しました。

本技術を用いることで、大量の欠損を含むデータであっても、高い精度で要因解析を行うことが可能となります（図１）。今後、東芝と統数研は、本技術の汎用化・高速化に取り組むとともに、工場・プラントを含む様々な分野の実課題への適用を検証し、生産性・歩留・信頼性の向上に貢献してまいります。

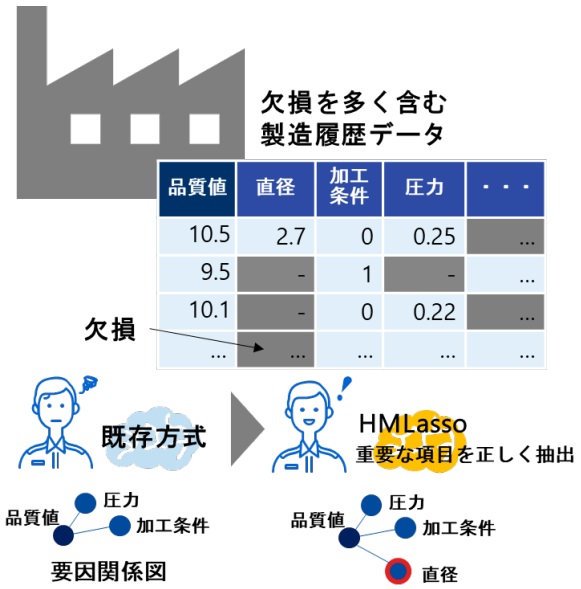

図１：HMLasso の活用イメージ

注１ Convex Conditioned Lasso（CoCoLasso）
出典：Datta, A., & Zou, H. (2017). CoCoLasso for high-dimensional error-in-variables regression. The Annals of Statistics, 45(6), 2400-2426.
注２ https://www.ijcai19.org/
注３ https://CRAN.R-project.org/package=hmlasso（公開時変更の可能性あり）
注４ 特定のデータ項目の値を他のデータ項目から説明するモデル
注５ スパースモデリング：変数選択とモデル化を同時に行う方法論

以上

TOSHIBA

大学共同利用機関法人 情報・システム研究機構
統計数理研究所

【報道機関からのお問い合わせ先】
（株）東芝　コーポレートコミュニケーション部　広報・ＩＲ室　03-3457-2100

大学共同利用機関法人　情報・システム研究機構
統計数理研究所　運営企画本部企画室 URA ステーション　050-5533-8580