報道発表

宇都宮大学
UTSUNOMIYA UNIVERSITY

令和8年9月14日

会計仕訳の異常について「本来のあるべき仕訳は何だったのか」まで推測する説明可能AI（XAI）を開発

～異常と判定した会計仕訳について、原因とみられる項目と修正候補まで示す～

【発表のポイント】

・これまでのAIによる会計仕訳の異常検知の多くは、仕訳を「異常」と判定するだけであったため、財務諸表監査における仕訳テストなどの監査手続を実施する際や会計事務所が巡回監査をする際に、「なぜ異常なのか」も、「本来の正常な仕訳は何だったのか」も、監査人が一から調べ直すしかなく、大きな手間がかかっていた。

・異常と判定した会計仕訳について、判定に効いた項目と、その項目をどの値に変えればAIが正常と判定するのかという修正候補を、自動で示す手法を開発した。AIの判定を覆す最小の変更を示す「反実仮想説明」を会計仕訳の異常検知に持ち込んだのは、本研究が初。

・大企業のERP(統合基幹業務システム)で使われる形式の仕訳データ約53万件で検証。異常の正解ラベルを使わずに検知のしきい値を決め、評価用に人工的に作成して注入した120件の異常を、全件検知した。正常な仕訳約8万件のうち、誤って異常とした件数は31件。

・代表的な既存手法3種（Wachter法、DiCE、FACE）との比較では、変更後にAIが全項目を正常と判定する割合を100%に保ったまま、人工異常で改変した項目を正解としたときの完全一致率が最も高く、変更する項目数も最も少なかった（平均1.18項目）。

■研究概要

宇都宮大学 データサイエンスセンターに設置している「宇都宮大学・TKC AIソリューション共同研究室」所属の川面洋平准教授(宇都宮大学 データサイエンス経営学部／データサイエンスセンター副センター長)と齊藤大輔准教授（株式会社TKC／宇都宮大学 データサイエンスセンター）は、会計監査で使う仕訳データの異常検知について、AIが異常と判定した理由を監査人に分かりやすく示し、さらに本来の正常な仕訳を推測する手法を開発しました。

これまでのAIによる仕訳異常検知の多くは、正常な取引だけを学習したオートエンコーダ（注1）が、うまく再現できなかった仕訳を「異常」として拾い上げる方式でした。ただし出てくるのは再構成誤差（注2）という数値ひとつです。監査手続の実施および監査上の判断を行う際に、この数値だけでは足りません。

開発した手法は、検知、寄与項目のランキング、修正候補の探索という3段階からなります（図1)。異常と判定した仕訳について、まず既存の説明手法RESHAPE（注3）で再構成誤差を項目ごとに分解し、異常

判定への寄与が大きい項目から順に並べます。次に、その順位を手がかりに反実仮想説明（注 4）、つまり「この項目をこの値に変えれば AI が正常と判定する」という最小限の修正候補を探します。

反実仮想説明はこれまで、画像認識のように「A か B か」を判定する分類 AI を対象に研究されてきました。会計仕訳の異常検知に持ち込んだのは、本研究が初めてです。第 2 段階の RESHAPE は先行研究（Mueller ら、2022 年）の手法であり、その寄与度を道しるべに、会計仕訳向けの修正候補探索（第 3 段階）を組み立てたところが本研究の中核にあたります。

本研究成果は、9 月 9 日、学術誌「IEEE Access」オンライン版に掲載されました。

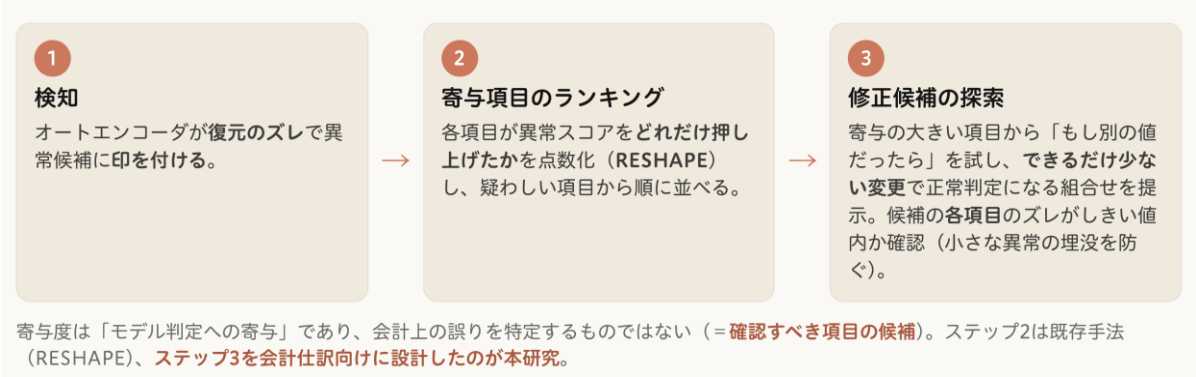

図 1　提案手法の 3 ステップ。異常を見つけるところで止めず、寄与の大きい項目の順位付けと修正候補の提示までを一続きで行う。

■研究の背景

不正や誤りのある仕訳を見つけ、経営者による内部統制の無効化リスクの判断に資するべく、有効な仕訳テストなどの監査手続を実施することは、会計監査の長年の課題です。従来は「この勘定科目でこの金額を超えたら確認する」といった条件を人が書き出すルールベースの方式が主流でしたが、書かれていない手口は素通りしてしまいます。そこで近年は、オートエンコーダ（注 1）をはじめとする AI を使う手法の研究が進んできました。正常な取引だけを学習させておけば、異常の正解ラベルを用意しなくても、学習した正常パターンから外れた仕訳を拾えます。あらかじめ想定できない逸脱も拾いやすい方式です。

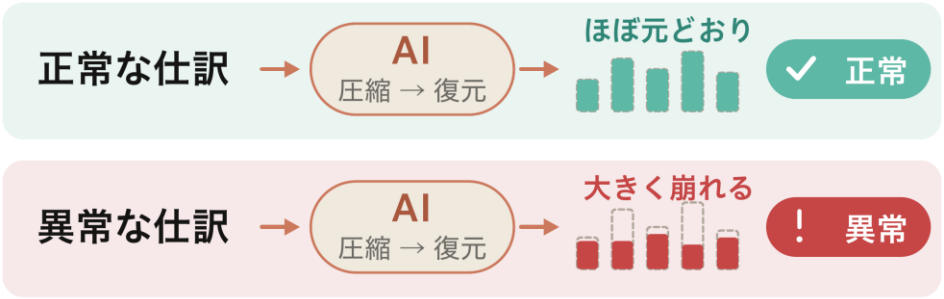

図 2　オートエンコーダによる異常検知。正常な仕訳は圧縮しても元どおりに復元できるが、学習していない異常な仕訳は復元が大きく崩れる。このズレの大きさが異常の指標になる。

ただし監査の現場では、「異常か否か」の二値判定だけでは足りません。監査手続の実施および監査上の判断を行うためには、「なぜこの取引が異常なのか」、「修正するにはどこをどう変えれば正常になるのか」がわからなければなりません。多くの従来手法が出す再構成誤差（注 2）は仕訳ひとつにつき数値ひとつで、項目ごとの説明と修正候補を一体で示すものではありませんでした。

「どの項目をどの値に変えれば AI の判定が変わるか」を示す説明手法は、反実仮想説明（注 4）と呼ばれ、研究が進んでいます。ただしその大半は、画像を「A か B か」に振り分ける分類 AI を想定したものです。オートエンコーダはこれと事情が違います。修正候補が正常かどうかを項目ごとの再構成誤差のしきい値で確かめる必要があり、勘定コードのような離散値と金額のような連続値が混ざった空間を探索しなければなりません。オートエンコーダ特有のこうした事情をまとめて扱う手法は、これまで考案されていませんでした。

そこで、寄与の大きい項目の特定と修正候補の提示をひとつのパイプラインにまとめ、監査人がそのまま確認できる形で出力する手法を設計しました。

反実仮想説明により、修正するにはどこをどう変えれば正常になるのかが分かるようになることで、これまで監査対象企業やその業界への知識や経験が無いと難しかった、証憑書類等に疑義を持つことも可能となり監査の質の向上につながります。

■研究方法

大企業の ERP として広く使われる SAP の形式で先行研究が公開している会計仕訳データセットから、正常な仕訳明細行 532,789 行を抽出しました。評価には「どの項目を改変したか」という正解が要るため、改変内容を記録しながら異常を人工的に注入しています。注入した異常は 4 種類、計 120 件であり、学習データにないコードへの置換、単独では正常でも組合せが未出現になるもの 2 種類、カテゴリ項目と金額を同時に改変したものです。

データの項目は、総勘定元帳キー、利益センタ、転記キー、勘定コード、会社コード、通貨キー、国内通貨金額、伝票通貨金額の 8 つです。

図 3　RESHAPE による寄与項目のランキング。仕訳ひとつにつき数値ひとつだった異常スコアを、勘定コードや金額といった項目ごとの寄与度に分解し、疑わしい順に並べる。

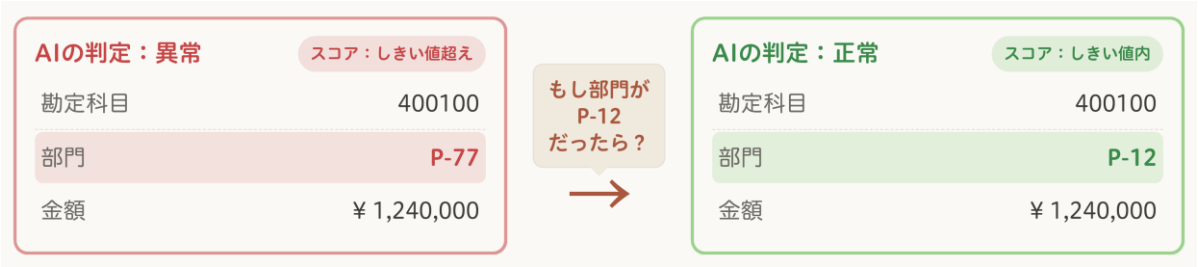

図 4　反実仮想説明による修正候補（模式例）。部門 P-77 はそれ自体はよくある値だが、この勘定科目との組合せが学習データにない。P-12 に変えると AI の判定が正常に変わる。この最小の変更がそのまま修正候補になる。※AI が正常と判定する候補であり、会計上の正しさは最終的に人が判断する。

■研究成果

検知については、ラベルを見ずに選んだしきい値で、注入した 120 件の異常を全件見つけました。誤検知は正常なテストデータ 79,919 件のうち 31 件（偽陽性率（注 5） 3.9×10^{-4}）です。4 種類の異常はいずれも取りこぼしがありませんでした。誤検知 31 件のうち 26 件では、総勘定元帳キーの項目別再構成誤差が最も大きくなっていました。仕訳全体の異常スコアが大きい順に並べれば、監査人が確認の優先順位をつけられます。

修正候補については、120 件中 30 件（25.0%）で改変した項目の集合と完全に一致し、99 件（82.5%）で少なくとも 1 つの正解項目を含みました。変更を提案する項目数は平均 1.18 項目で、確認する候補はおおむね 1〜2 項目に収まります。

同じデータ・同じ制約で既存の代表的手法と比べました。改変した項目の集合との完全一致率は本手法 25.0%に対し、Wachter 法 18.3%、DiCE 10.8%、FACE 6.7%。変更後に AI が全項目を正常と判定した割合は本手法・DiCE・FACE が 100%、Wachter 法は 33.3%でした。この 100%を満たした 3 手法のなかで、変更項目数が最も少なかったのは本手法の平均 1.18 項目でした（DiCE 1.89 項目、FACE 2.98 項目）。修正候補の探索時間は 1 件 0.06 秒で、Wachter 法の 9.92 秒、DiCE の 8.02 秒を大きく下回りました。FACE は 0.004 秒とさらに速いものの、正常な学習データの 1 行をまるごと写す方式のため、完全一致率は 6.7%、変更項目数は 2.98 項目です。

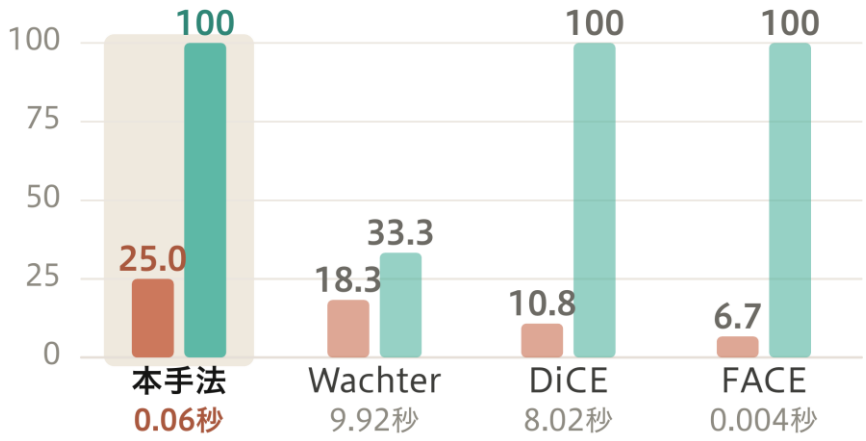

図 5　既存手法との比較。左（茶）＝改変された項目の組との完全一致率（%）、右（緑）＝修正候補が AI に正常と判定された割合（%）。手法名の下は 1 件あたりの修正候補の探索時間（RESHAPE の寄与度計算は含まない）。

■今後の展望（研究のインパクトや波及効果など）

これまで多くはアラートを出すだけだった AI の異常検知が、有効な監査手続の実施および監査上の判断の根拠と、具体的な修正候補まで出せるようになります。AI が示す変更候補は平均 1.18 項目に絞られます。監査作業時間の短縮による「効率性」や監査手続の利害関係者や検査当局への説明責任に係る「有効性」については、実データの範囲や分析対象量の拡大と経験豊富な監査人による検証が必要であり、この点がこれからの課題となります。また、同じ枠組みは、金融取引の不正検知や製造データの品質管理など、異常検知に説明が求められる他分野にも広げられる可能性を持っています。

一方で、実務で使うにはまだ越えるべき点があります。今回の検証は人工的に注入した 4 種類の異常が対象で、実際の不正はもっと巧妙です。今後は実際の企業データを用いて発展をさせ、実務での運用に近づけます。

■論文情報

論 文 名 ： Explainable Anomaly Detection in Accounting Journal Entries Using Autoencoders and Counterfactual Explanations

雑誌名：IEEE Access

著者：Yohei Kawazura（川面洋平、宇都宮大学 データサイエンス経営学部／データサイエンスセンター）、Daisuke Saitou（齊藤大輔、株式会社 TKC／宇都宮大学 データサイエンスセンター）

DOI：10.1109/ACCESS.2026.3727143

URL：https://doi.org/10.1109/ACCESS.2026.3727143

■用語説明

注 1　オートエンコーダ（Autoencoder）：入力データをいったん低次元に圧縮し、そこから元のデータを復元するように学習するニューラルネットワーク。正常なデータだけで学習させると、正常なパターンは復元できるが、学習したことのない異常なデータは復元に失敗する。この失敗の度合いを異常の指標として使う。

注 2　再構成誤差：オートエンコーダが復元したデータと、もとの入力データとのずれの大きさ。大きいほど、学習した正常パターンから外れている。

注 3　SHAP／RESHAPE：SHAP は、ゲーム理論のシャープレイ値に基づき、AI の出力に各入力項目がどれだけ寄与したかを算定する手法。RESHAPE はこれを会計監査向けに特化させた先行研究（Mueller ら、2022 年）の手法で、オートエンコーダの再構成誤差を勘定コードや金額といった項目単位の寄与度に分解する。本研究はこの寄与度を、修正候補を探す順序として利用している。

注 4　反実仮想説明（Counterfactual Explanation）：「もし○○がこの値だったら、AI の判定は変わっていた」という形で、判定を覆す最小限の変更を示す説明手法。本研究では「この項目をこの値に変えれば AI が正常な仕訳と判定する」という、監査人が確認すべき修正候補として使う。

注 5　偽陽性率（False Positive Rate）：正常なデータのうち、誤って異常と判定した割合。本研究では 3.9×10^{-4}で、正常な仕訳およそ 2,600 件に 1 件にあたる。

■英文概要

In financial auditing, an autoencoder (AE) can flag a journal entry as anomalous, but the auditor still needs to know which attributes triggered the flag and how to correct the entry. We address both questions with a pipeline that uses attribute-level SHAP (SHapley Additive exPlanations) attribution (RESHAPE) for root-cause identification, and RESHAPE-guided beam search with gradient-based numeric optimization for counterfactual explanation (CF) generation. We use two thresholds: global binary cross-entropy (BCE) loss for detection, and per-attribute thresholds for CF validity. The detection threshold is selected label-free, from a target false-positive rate measured on a held-out clean validation set, so that no anomaly labels enter threshold selection. On 532,789 normal SAP-format accounting journal entries, split into training, validation, and test partitions with 120 injected anomalies held out for evaluation, the detector flagged all 120 anomalies (100%

recall) while raising only 31 false positives on the 79,919-entry clean test set (false-positive rate 3.9×10^{-4}). On those 120 anomalies, the CF pipeline exactly matched the ground-truth modified attributes in 30 cases (25.0%) and hit at least one ground-truth attribute in 99 cases (82.5%), averaging 1.18 modifications per CF, and outperformed the Wachter, DiCE, and FACE counterfactual baselines, achieving the highest exact-match rate at full validity with the sparsest corrections.

本件に関する問い合わせ

（研究内容について）

国立大学法人 宇都宮大学　データサイエンス経営学部　准教授　川面 洋平

TEL：028-649-5312　E-mail：kawazura@a.utsunomiya-u.ac.jp

（報道対応）

国立大学法人 宇都宮大学 企画総務部企画総務課 広報・渉外係

TEL：028-649-5201　FAX：028-649-5027　E-mail：kkouhou@a.utsunomiya-u.ac.jp