会計不正検知におけるテキスト情報の役割

矢野 直 [1,2]　小川 泰弘 [1]
[1] 名古屋市立大学　[2] 矢野公認会計士事務所
tadashi.yano@yano-accounting.com, ogawa@ds.nagoya-cu.ac.jp

概要

会計不正は発覚後の損失が大きく，早期検知が重要である。近年は，財務開示文書に含まれるテキスト情報を対象とした研究が進む一方で，数値情報とテキスト情報の役割を明示的に区別し，自然言語処理手法を複数組み合わせて体系的に比較した検証は十分でない。本研究では，数値情報のみ，テキスト情報のみ，および両者を組み合わせた条件の下で，TF-IDF，Word2Vec，BERT を用いた手法，ならびに大規模言語モデルによる推論を対象として，会計不正検知性能を比較する。

1　はじめに

会計不正検知は，観測可能な公開情報から不正の有無を推定するという点で，一般に不均衡データ下の分類問題として位置づけられる。実務的観点からは，会計不正は企業活動および資本市場に深刻な影響を及ぼし，発覚後には財務的損失に加えて，調査・訂正・再発防止対応に多大なコストを要する。そのため，公開情報に基づく早期の不正検知は，学術的にも実務的にも重要な研究課題である。

従来の会計不正検知研究は，財務比率や会計数値に基づく統計的手法を中心に発展してきたが，近年では自然言語処理技術の進展に伴い，企業が公開する財務開示文書に含まれるテキスト情報を活用した研究も増加している。

日本において有価証券報告書は，投資家に対して企業の財務状況や経営の実態を包括的に伝えるための主要な財務開示文書であり，財務諸表に基づく数値情報に加えて，事業内容，リスク要因，経営者による分析など，企業活動を理解するための多様なテキストを含んでいる。

例えば Yazawa ら [1] は，有価証券報告書テキストと財務数値を併用することで会計不正検知性能が向上することを示した。また Sugiura ら [2] は，日本の財務開示文書を対象としたベンチマークデータセット EDINET-Bench を公開し，大規模言語モデル（LLM）を含む多様な手法を評価している。しかし，既存研究では，財務数値に基づく数値情報とテキスト情報の役割が必ずしも明確に整理されておらず，また伝統的手法から BERT，LLM までを同一条件下で体系的に比較した検証は十分ではない。

本研究では，有価証券報告書を対象とし，数値情報とテキスト情報を明示的に区別した上で，会計不正検知におけるそれぞれの有用性を検証することを目的とする。具体的には EDINET-Bench を用い，数値情報のみを用いる構成，テキスト情報のみを用いる構成，および両者を結合した構成の三条件を定義し，複数の手法における性能を比較する。

本研究の貢献は，(1) 数値情報とテキスト情報を明確に分離した評価枠組みを提示する点，(2) TF-IDF や Word2Vec といった伝統的手法から，BERT のファインチューニングおよび LLM までを統一条件下で比較する点にある。これにより，会計不正検知においてどの情報がどの程度有効であるかを整理し，今後のモデル設計に対する基礎的知見を提供する。

2　関連研究

会計不正検知に関する研究は，主として財務数値に基づく統計的手法を中心に発展してきた。一方で，法定財務開示文書には財務数値のみならず自然言語テキストが含まれることから，テキスト情報を用いた不正会計検知が近年注目されている。

テキスト情報を用いた会計不正検知の初期的なアプローチとして，単語頻度や TF-IDF，辞書ベース指標などの Bag-of-Words 系特徴量を用い，財務開示文書テキストを定量化する手法が用いられてきた。宮後ら [5] は，米国における法定財務開示文書を対象とした先行研究を整理し，財務数値に加えて，単語頻度や辞書ベース指標等のテキスト情報を用いた機械学習が不正検知に活用されてきたことを体系的にまとめている。

テキストの構造的側面に着目した研究として，

Mayew ら [3] は，米国における法定財務開示文書に対してトピックモデルを適用し，特定の話題構造が，不正会計と統計的に関連することを示した。この研究は，法定財務開示文書テキストに潜在する意味構造が不正検知に有用な情報を含み得ることを示した代表的な先行研究である。

近年では，文脈依存表現を用いる深層学習手法の導入が進んでいる。Ketelaar ら [4] は，米国における法定の年次財務開示文書の一部のセクションを対象に，BERT をファインチューニングした文脈依存的な文書表現を用いて分析をし，従来の語彙ベース手法と比較して高い不正検知性能を報告している。

日本を対象とした研究として，Yazawa ら [1] は，有価証券報告書のうち MD&A に加え，経営方針，リスク情報，コーポレート・ガバナンス等の記述を対象に，テキスト情報と財務数値を併用することで不正検知性能が向上することを示した。

また，Sugiura ら [2] は，日本の法定財務開示文書を対象としたベンチマークデータセット EDINET-Bench を公開し，不正検知を含む複数のタスクについて，LLM を含む多様な手法を評価している。

以上の既存研究は，財務開示文書に含まれるテキスト情報の有用性を示している一方で，分析対象や用いる情報，評価設定は研究ごとに異なっている。本研究はこの点に着目し，数値情報とテキスト情報の役割を明示的に区別した上で，特徴量抽出型手法から BERT，さらに LLM に至るまでを同一条件下で体系的に比較・検証する。これにより，会計不正検知におけるテキスト情報の役割を，数値情報との比較の下で明らかにすることを目的とする。

3 実験設定

本節では，本研究における実験設定について述べる。本実験の目的は，有価証券報告書に含まれる数値情報とテキスト情報を明確に区別した上で，会計不正検知におけるそれぞれの有用性を同一条件下で比較することである。

この目的のため，数値情報のみ（S-only），テキスト情報のみ（T-only），および両者を組み合わせた構成（S+T）を定義し，複数の代表的手法に対して統一的に評価する。以下では，使用したデータセット，特徴量の構成，適用した手法，および評価方法について述べる。

3.1 データセット

本研究では，日本の法定財務開示文書を対象としたベンチマークデータセット EDINET-Bench を用いる。会計不正検知タスクでは，各サンプルが 1 件の有価証券報告書に対応し，不正会計の有無を二値ラベルとして付与している。

EDINET-Bench では，訓練データ 865 件，テストデータ 224 件の単一分割が提供されている。この分割に基づく会計不正検知タスクは，有価証券報告書（企業 × 年度）単位で定義されており，不正（陽性）サンプルは全体の 49% を占める，均衡なデータセットである。しかし，有価証券報告書は内容や記載形式に多様性があるため，単一分割に基づく評価では推定結果の安定性を十分に担保できない。

そこで本実験では，評価の安定性を確保するため，訓練データとテストデータを統合した上で，5-fold クロスバリデーションによる評価を行う。この際，同一企業に関する複数年度の有価証券報告書が学習・評価の双方に同時に含まれることを防ぐため，企業あたり 1 件の有価証券報告書のみを残すよう重複を除去し，企業単位のデータセットを構成した。その結果，合計 754 件の企業データを分析対象とした。

3.2 特徴量の構成方法

本研究では，以下の二種類の入力を用いる。

- 数値特徴ベクトル：有価証券報告書の数値情報をベクトル化したもの
- テキスト特徴ベクトル：有価証券報告書のテキストをベクトル化したもの

数値特徴ベクトルは本節で述べ，テキスト特徴ベクトルは手法に応じて次節以降で述べる。

本研究で用いる数値特徴ベクトルは，有価証券報告書冒頭に掲載されている主要な経営指標等の推移に基づくものである。EDINET-Bench では，当該数値特徴ベクトルが会計基準の違いを考慮して整理されており，本研究では，その整理結果として提供される 22 項目の主要な経営指標を用いた。各指標について過去 5 年分の値を用い，1 件の有価証券報告書あたり 22 項目 × 5 年の 110 次元からなる数値特徴ベクトルを構成した。

数値特徴ベクトルの値については，学習時に平均 0，分散 1 となるよう標準化した。

EDINET-Bench には，より詳細な財務数値も含まれているが，本研究では，数値特徴ベクトルとテキスト特徴ベクトルの寄与を明確に比較するため，主要な経営指標等の推移に限定した。

S+T 構成では，数値特徴ベクトルとテキスト特徴ベクトルを連結し，分類器に入力した。

3.3 手法

本研究では，数値特徴ベクトルとテキスト特徴ベクトルの組合せ（S-only / T-only / S+T）を統一条件として設定し，複数の手法を比較した。比較対象には，伝統的な Bag-of-Words 系手法から，文脈依存表現を用いる深層学習手法，さらにゼロショットで推論する事前学習済み LLM を含め，表現能力および統合方法の違いが性能に与える影響を検証する。

3.3.1 TF-IDF / Word2Vec

TF-IDF および Word2Vec については，数値特徴ベクトルとテキスト特徴ベクトルを入力とする二値分類問題として定式化し，ロジスティック回帰分類器を用いて評価した。

圧縮前の TF-IDF 表現は約 1.7 万次元となった。これに対して LSI を適用し，100 次元に削減したテキスト特徴ベクトルを用いた。

Word2Vec では，埋め込み次元数を 100 とし，文書中の各単語に対応する単語埋め込みの平均により，テキスト特徴ベクトルを構成した。なお，汎用的に事前学習された Word2Vec モデルは用いず，各クロスバリデーションの訓練データに基づいて新たに単語埋め込みを学習した。

いずれも次節の BERT-PCA と次元数を揃えることにより，特徴ベクトルの次元数の違いによる影響を排除した。

3.3.2 BERT

事前学習済み日本語 BERT として，東北大学が公開している `cl-tohoku/bert-base-japanese-v2` を用いた。有価証券報告書は入力長が長いため，テキストを一定長ごとに分割し，各チャンクの `[CLS]` トークン出力を平均することで，テキスト特徴ベクトルを構成した。

BERT-PCA では，得られた高次元表現（768 次元）に PCA を適用し，100 次元に削減した。この構成は，BERT 本体のパラメータを更新せずに文脈依存表現を特徴量として利用することで，特徴量抽出型手法との比較を可能にすることを目的としている。

BERT-FT では，T-only および S+T 構成において，テキスト特徴ベクトルに基づき BERT 本体を含む全パラメータを更新するファインチューニングを行った。モデル構成としては，BERT 本体に分類ヘッドを付加したものを用い，非線形モデルによる統合効果を検証した。S-only 構成では，テキスト特徴ベクトルを用いず数値特徴ベクトルのみを分類ヘッドに入力し，BERT 本体は使用・更新せず，分類ヘッドのみを学習した。

3.3.3 大規模言語モデル（LLM）

LLM では学習を行わず，プロンプトに基づくゼロショット推論を実施し，会計不正の可能性を 0 から 1 の連続値として出力させた。

他の手法との比較可能性を確保するため，LLM に対しても S-only，T-only，S+T の三つの入力構成を定義した。LLM は特徴ベクトルを入力とする教師あり学習モデルではないため，本研究では数値情報およびテキスト情報をテキストとして提示する。

S-only 構成では，数値特徴ベクトルと同一の情報源である主要な経営指標等の推移（22 項目）を LLM が解釈可能な形式に整形した表現（TOON）[1] として入力し，T-only 構成では有価証券報告書のテキストを，S+T 構成では両者を入力した。

3.4 評価方法

LLM 以外の，訓練データを用いてパラメータ学習を行う手法（以下，教師あり学習モデル）については，5-fold クロスバリデーションを用いて評価を行い，各評価指標の平均値および標準偏差を報告した。各 fold においては，検証データ上で F1 スコアが最大となる閾値を選択し，当該閾値に基づいて評価指標を算出した。

一方，LLM については，分類器としての学習や最適化を行わないゼロショット推論として用いるため，EDINET-Bench において公開されているテスト分割を用いた単一評価を行った。この際，テストデータに基づく閾値の調整による情報漏洩を避ける目的から，閾値は 0.5 に固定し，他の手法と同一の評価指標に基づいて比較した。

評価指標としては，Precision，Recall，F1 スコア，Average Precision（AP），Matthews Correlation Coefficient

1） TOON（Textualized Object-Oriented Numbers）は，数値情報を項目名と値の組からなるテキスト表現に変換する形式である（`https://github.com/toon-format/toon`）。

表 1　会計不正検知における各手法の性能比較

Method	Feature	Precision	Recall	F1	AP	MCC	ROC-AUC
S（共通）		0.382 ± 0.068	0.715 ± 0.195	0.484 ± 0.056	0.390 ± 0.040	0.249 ± 0.095	0.633 ± 0.077
TF-IDF (LSI)	T	0.428 ± 0.063	0.770 ± 0.135	**0.541** ± 0.034	**0.481** ± 0.081	**0.345** ± 0.057	**0.715** ± 0.052
	S+T	0.367 ± 0.056	0.775 ± 0.122	0.493 ± 0.050	0.424 ± 0.056	0.252 ± 0.086	0.667 ± 0.060
Word2Vec	T	0.399 ± 0.088	0.760 ± 0.080	0.514 ± 0.057	0.433 ± 0.076	0.283 ± 0.108	0.674 ± 0.069
	S+T	0.376 ± 0.066	0.790 ± 0.172	0.500 ± 0.059	0.413 ± 0.049	0.270 ± 0.109	0.654 ± 0.072
BERT-PCA	T	0.382 ± 0.050	0.845 ± 0.094	0.521 ± 0.042	0.424 ± 0.061	0.307 ± 0.068	0.681 ± 0.062
	S+T	0.396 ± 0.049	0.700 ± 0.061	0.502 ± 0.033	0.412 ± 0.058	0.269 ± 0.067	0.671 ± 0.053
BERT-FT	S	0.286 ± 0.023	**0.980** ± 0.033	0.442 ± 0.023	0.332 ± 0.075	0.137 ± 0.060	0.532 ± 0.065
	T	**0.433** ± 0.065	0.665 ± 0.104	0.522 ± 0.071	0.450 ± 0.107	0.313 ± 0.106	0.688 ± 0.083
	S+T	0.411 ± 0.091	0.805 ± 0.089	0.535 ± 0.060	0.441 ± 0.089	0.325 ± 0.103	0.701 ± 0.068
LLM	S	0.619	0.533	0.573	0.618	0.140	0.615
	T	0.688	0.631	0.658	0.671	0.287	0.654
	S+T	0.621	0.672	0.646	0.640	0.184	0.621

教師あり学習モデルは 5-fold CV（平均値 ± 標準偏差）を示す。太字は教師あり学習モデル内での各評価指標の最高値であり，LLM は評価方法が異なるため太字による比較対象には含めない。なお，S（数値特徴ベクトルのみ）は TF-IDF / Word2Vec / BERT-PCA で同一入力・同一分類器のため同一結果となる。

（MCC），および ROC-AUC を用いた。

4　実験結果と考察

実験結果を表 1 に示す。表 1 より，TF-IDF, Word2Vec，BERT-PCA といった特徴量抽出型手法では，いずれの手法においても T-only 構成が S-only 構成を一貫して上回る傾向が確認された。特に F1, AP，ROC-AUC の結果は，テキスト特徴ベクトルが数値特徴ベクトルよりも会計不正検知に有用な情報を含む可能性を示している。

一方で，多くの手法において Recall の標準偏差が大きい結果となった。これは，有価証券報告書および会計不正の内容が多様であり，限られた訓練データではその多様性を十分に網羅できないことを示唆している。

S+T 構成では，特徴量抽出型手法では T-only 構成を上回る結果は得られなかった一方，BERT-FT では S+T 構成において最も高い F1 が得られた。このことから，数値特徴ベクトルとテキスト特徴ベクトルの統合効果は，モデルの表現能力や統合方法に強く依存することが示唆される。

また，LLM は T-only 構成において相対的に高い性能を示し，テキスト情報に基づく判断の有効性が確認された一方で，S-only および S+T 構成では性能向上は限定的であり，特に Recall の観点では改善の余地が示唆された。

5　まとめ

本研究では，有価証券報告書を対象として，数値情報およびテキスト情報を用いた会計不正検知手法を比較検証した。TF-IDF，Word2Vec，BERT に基づく特徴量抽出手法，BERT のファインチューニング，および LLM による推論を対象とし，数値情報のみを用いる構成，テキスト情報のみを用いる構成，および両者を組み合わせた構成の下で評価した。

実験結果から，複数の手法においてテキスト情報のみを用いる構成が数値情報のみを用いる構成を上回り，有価証券報告書に含まれるテキスト情報が会計不正検知に有用であることが示された。一方で，数値情報とテキスト情報を組み合わせた構成の有効性はモデルの特性に依存し，非線形に統合する手法では性能向上が確認された。

また，LLM はテキスト情報に基づく判断において一定の有効性を示したが，Recall の観点からは，実運用に向けて追加的な工夫が必要であることが示唆された。

今後の課題としては，標本数の拡大に加え，解釈性の高い手法による特徴分析や，数値情報とテキスト情報の関係性をより柔軟に捉える統合モデルの検討が挙げられる。本研究で得られた知見は，会計不正検知における情報設計およびモデル選択に関する基礎的な指針を提供するものと考えられる。

参考文献

[1]Kenichi Yazawa, Kazuo Araragi, Yoshinao Itakura, Teppei Usuki, Daichi Hattori, and Satoshi Mizuno. Detecting Financial Misconduct Using NLP and Machine Learning: Evidence from Japan. **SSRN Electronic Journal**, 2024.

[2]Issa Sugiura, Takashi Ishida, Taro Makino, Chieko Tazuke, Takanori Nakagawa, Kosuke Nakago, and David Ha. EDINET-Bench: Evaluating LLMs on Complex Financial Tasks using Japanese Financial Statements. **arXiv preprint** arXiv:2506.08762, 2025.

[3]William J. Mayew and Mohan Venkatachalam. What Are You Saying? Using Topic to Detect Financial Misreporting. **Journal of Accounting Research**, 2019.

[4]Florian Ketelaar and Ana Mićković. Artificial Intelligence in Fraud Detection: Textual Analysis of 10-K Filings. **Maandblad voor Accountancy en Bedrijfseconomie**, 99(2):61–71, 2025. doi:10.5117/mab.99.132881.

[5]宮後圭佑，佐藤夏輝，小村亜唯子，平井裕久．Form 10-K のテキストを使用した不正会計検知モデルの研究．**研究報告インテリジェント・インフォマティクスと運用技術**, 21:107–125, 2024. doi:10.24677/riim.21.0_107.

謝辞

本研究は JSPS 科研費 23K25155 の助成を受けたものである．

A　実験設定の詳細

A.1　主要な経営指標等の推移

EDINET-Bench の summary が提供している主要な経営指標等の推移（22 項目）を表 2 に示す。これらは有価証券報告書に基づく制度上定義された数値情報であり，企業の収益性，財務状態，キャッシュ・フロー，および株主価値に関する側面を表す。

表 2　主要な経営指標等の推移（22 項目）

指標名
売上高
経常利益
親会社株主に帰属する当期純利益
包括利益
純資産額
総資産額
１株当たり純資産額
１株当たり当期純利益又は当期純損失
潜在株式調整後１株当たり当期純利益
自己資本比率
自己資本利益率、経営指標等
株価収益率
営業活動によるキャッシュ・フロー
投資活動によるキャッシュ・フロー
財務活動によるキャッシュ・フロー
現金及び現金同等物の残高
当期純利益又は当期純損失
税引前利益 (IFRS)
親会社株主に帰属する当期純利益 (IFRS)
当期包括利益: 親会社の所有者に帰属（IFRS）
従業員数
平均臨時雇用人員

A.2　特徴量抽出型手法のハイパーパラメータ

TF-IDF，Word2Vec，および BERT-PCA に基づく特徴量抽出型手法では，全ての入力構成（S-only, T-only, S+T）において，ロジスティック回帰分類器（L2 正則化，class_weight=balanced，solver=liblinear，max_iter=4000）を共通して用いた。

TF-IDF では，TfidfVectorizer により unigram 特徴量を構成し，min_df=5 により低頻度語を除去した上で，TruncatedSVD を用いて 100 次元に削減した。

Word2Vec では，skip-gram （sg=1） を用い，vector_size=100, window=5, min_count=2, epochs=20 の設定で単語埋め込みを学習した。テキスト特徴ベクトルは，文書内の単語埋め込みベクトルの平均により構成した。

A.3　前処理およびトークナイザ設定

TF-IDF および Word2Vec に基づく手法では，日本語形態素解析器 MeCab を用い，UniDic-light（unidic-light）に基づいて語彙を構成した。

BERT モデルでは形態素解析は行わず，事前学習済みモデルに対応する WordPiece ベースのトークナイザを用いた。

A.4　BERT モデルの設定

BERT-PCA および BERT-FT では，有価証券報告書が最大入力長を超えることに対処するため，テキストを最大 510 トークンごとに分割し，各チャンクの [CLS] トークン出力を用いた。

BERT-PCA では，複数チャンクから得られた [CLS] ベクトルを各チャンクの有効トークン長に基づいて重み付き平均し，その後 PCA を適用して 100 次元に削減した。

BERT-FT では，BERT 本体に分類ヘッドを付加し，クロスエントロピー損失に基づくファインチューニングを行った。分類ヘッドは Dropout 層と全結合層 1 層から構成され，最終出力には Sigmoid 関数を用いて二値分類確率を算出した。S+T 構成では，分類ヘッド入力段階でテキスト特徴ベクトルと数値特徴ベクトルを連結した。

A.5　LLM 推論設定

LLM による推論には GPT-5.2 を使用し，同一入力に対する出力の安定性を高めるため，temperature=0.0 に設定した。出力は JSON 形式のみを許可し，不正会計リスクを表す連続値スコア（0.0–1.0）と，日本語による簡潔な判断理由の 2 項目を出力させた。

入力テキストが長文となる場合には，文字数 6,000 を上限として非重複に分割し，各チャンクに対して独立に推論を実行した。文書単位のスコアは，チャンクごとの出力スコアの最大値により集約した。

System prompt

あなたは，日本企業の有価証券報告書や決算短信などの開示文書を分析し，不正会計リスクを評価する専門家です。入力として与えられたテキストのみに基づき，不正会計の可能性を 0.0 から 1.0 の範囲で評価してください。出力は必ず JSON 形式のみとし，それ以外の説明文は出力しないでください。

User prompt

mode={S-only / T-only / S+T}
以下の入力テキストのみに基づき，不正会計リスクを 0.0–1.0 の数値で評価してください。
{ ‘‘fraud_score’’: 数値,
‘‘reason_ja’’: ‘‘判断理由（日本語，簡潔に）’’ }
上記以外の形式では出力しないでください。