

令和 5 年 6 月 19 日現在

機関番号：32675

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K01923

研究課題名（和文）機械学習を応用した財務諸表の不正発見

研究課題名（英文）Financial fraud detection using machine learning

研究代表者

坂上 学（Sakaue, Manabu）

法政大学・経営学部・教授

研究者番号：50264792

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は、機械学習やテキストマイニングの手法を用いて、会計不正の発見の可能性を高めることができるかを検証することにある。近年、有価証券報告書全体のデータがXBRL形式で入手できるようになり、財務情報とナラティブ情報について、全上場企業のデータが入手可能となった。これらの全データを取得するとともに、データの前処理をおこない、分析データの蓄積・整形を進めた。また東京商工リサーチによる「不適切な会計・経理の開示企業」のデータベースを入手し、不正をおこなった企業の特徴を把握するために、同業種で同規模の企業グループとの比較を行ったが、両企業群について、大きな特徴的差異というものは観察できなかった。

研究成果の学術的意義や社会的意義

これまで不正会計に関する研究は、財務分析を中心とした研究はそれなりの蓄積があったが、ナラティブ情報のような非財務情報に対する分析については十分な蓄積があるとはいえない。そこに近年発展の著しい機械学習やテキストマイニングの手法を応用して不正会計を行った企業のもつ特徴をとらえようとする追加的貢献をもたらすことに、本研究の学術的意義がある。本研究の結果はあまり芳しいものではなかったが、今後より進展すれば、不正会計を行っている企業を捉えることが可能となり、不正会計によってもたらされる損失を回避できる可能性が高まる点において、社会的意義があるといえる。

研究成果の概要（英文）： The purpose of this study is to examine whether machine learning or text mining methods can increase the likelihood of detection of accounting fraud over the conventional analysis method. The entire securities report data of all listed companies uploaded to the EDINET System is available in XBRL format from the fiscal years of 2013. As a result, we have access not only to financial information, but also to narrative information. The first step of this study was obtaining all of these data, pre-processing the data, and accumulating the data for analysis.

Then, using the database of the "Companies Disclosing Improper Accounting Survey (2008-2021)" by Tokyo Shoko Research, a comparison of data with a group of companies of similar size in the same industry was conducted to grasp the characteristics of the financial data of companies that have committed fraud. However, while no major differences of financial data and word and phrases characteristics of each company.

研究分野：会計学

キーワード：不正会計 機械学習 テキストマイニング XBRL

1. 研究開始当初の背景

近年、機械学習を応用して財務諸表の不正発見の試みをおこなった研究はいくつかなされてきたが、本研究課題を申請した時点において、畳み込みニューラルネットワークのようなディープラーニングを応用した深層学習を用いた研究は、存在していなかった。また機械学習の応用分野でもあるテキストマイニングの形態素解析をおこない、頻度分析を中心とする重要語の評価も組み合わせてどれだけ不正発見の精度を高めることができるのかについても、日本においては研究の蓄積はほとんどなされていない。

2. 研究の目的

上述したように、機械学習アルゴリズムを応用することによって、どれだけ不正発見の精度を高めることができるかを検証することが本研究の当初の目的であった。さまざまな文献をレビューした結果、ディープラーニングの代表的なアルゴリズムである畳み込みニューラルネットワークでは、株価や財務データのようなランダムウォークするデータを用いた場合には、ほとんど有意な成果が得られないことが判明したため、当初の目的をやや変え、テキストマイニングを中心とした研究に集中し、果たして不正発見の精度を高めることができるかについて探求することを目的とすることにした。

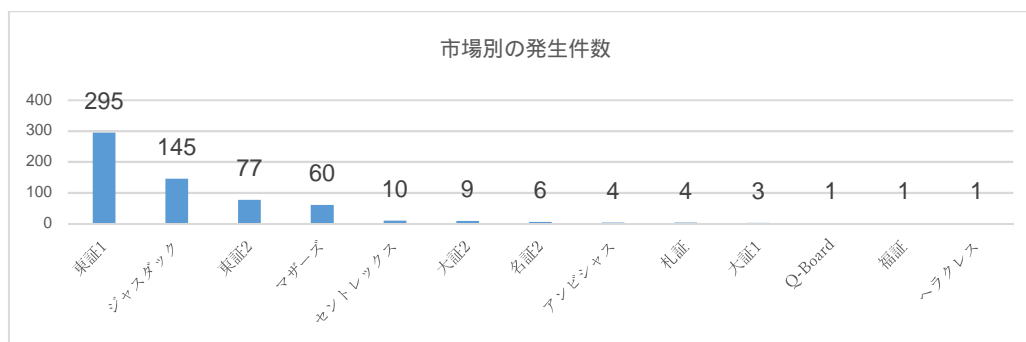
3. 研究の方法

まず、2014年度以降のすべての上場企業の有価証券報告書データを収集することから始めた。2014年度以降の有価証券報告書としたのは、財務データのみならず、すべてのナラティブな記述がXBRL化されて入手可能となったためである。次のステップとしては、収集したデータから不要なXBRLタグやHTMLタグを取り除き、またノイズとなりやすい英数字記号をすべて削除するという前処理をおこなった。この前処理に多くの時間を要し、研究計画の見直しを迫られる結果となったことは想定外であった。また不正会計をした企業を識別するため、東京商工リサーチ社による『全上場企業「不適切な会計・経理の開示企業」調査 2008～2021年度』を入手し、その全容について分析を行うとともに、不正会計を実施した企業についてのテキストマイニングを実施した。具体的な方法としては、形態素解析による不正企業のコーパス(全語彙)を明らかにし、頻度分析やTF・IDF分析を通じて、不正会計に固有な文言のスコアリングを実施し、不正のない企業との比較分析をおこなうことである。

4. 研究成果

不正会計企業のテキストデータについて、形態素解析はすませたものの、その解析に多くの時間がかかり、当初予定していた分析の多くを期間内に終わらせることはできなかった。とりわけTF・IDFのようにサンプルの母数を全上場企業とした場合、計算結果のアウトプットを取得できてはいるが、元のデータの5～10倍ほどに膨れ上がり、Excelのような普通のソフトウェアでまったく扱いきれないため、スクリプトを書いて処理をさせることになるが、多くの時間を要し、またエラーの発生に悩まされ、もう少し時間が必要となってしまった。それでも、以下のような傾向についての知見を得ることができた。

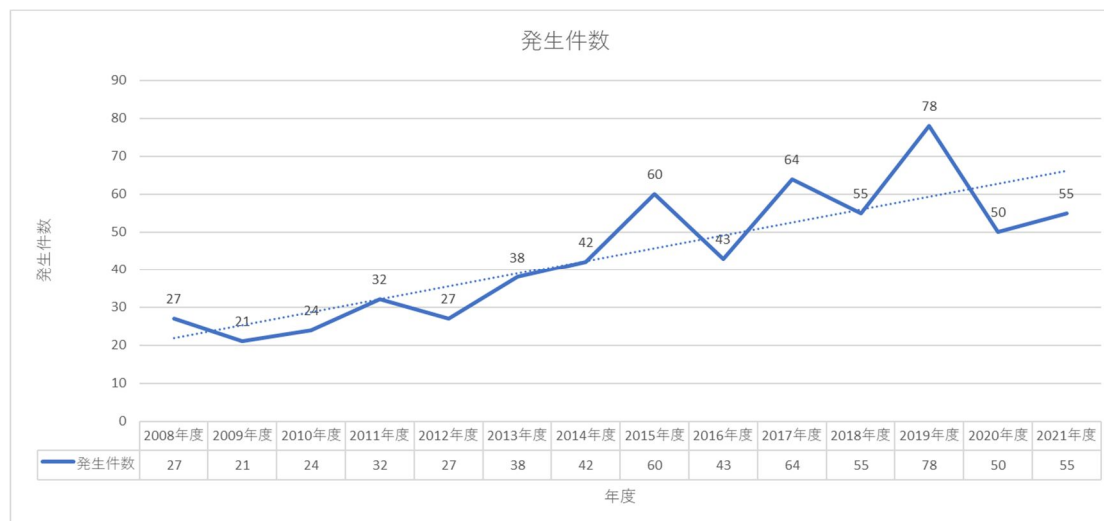
2008～2021年度において不正会計が報告された企業数は616件であった。まずは市場別の発生件数を見てもみると、以下のとおりであった。



市場規模の大きい東証1部が最も多くの会計不正が報告されているが、上場企業数に対する発生件数の割合を見てもみると、東証1部が13.5%、ジャスダックが20.9%、東証2部が16.2%、

マザーズが 14.2%と、ジャスダックが上場企業数の割には不正会計の報告件数が多いということが分かる。ジャスダックが中小型株を中心とする市場であることと何か関係があるのかもわからない。

続いて年度別に発生件数をみると、以下のような結果であった。多少のこぼこはあるものの、全体的には増加傾向にあることが見て取れる。



続いて 2008～2021 年度の不正会計の報告件数別に企業数を見てみると、以下のとおりであった。

不正報告数	1 回	2 回	3 回	4 回	5 回
企業数	463 社	55 社	10 社	2 社	1 社

最も不正会計があったと指摘された企業は、5 回の「東芝」である。実は、3 回の中に東芝の子会社（議決権を 52.7% 所有）である「東芝テック」も含まれているため、東芝グループとしては、合計で 8 回の不正会計が報告されていることになり、ダントツで 1 位ということになる。

以上の予備的な調査を終え、続いてテキストマイニングによる解析へと入るわけであるが、本報告書を執筆している時点で、まだデータ解析の結果が十分に得られていないことをお詫びしなければならない。データの前処理に膨大な時間がかかることと、解析結果は出力されているものの、その内容の吟味が大量のデータに阻まれてなかなか進まないのが原因である。この点については、当初の研究計画が甘かったことを反省しなければならない。

しかしながら、これまでに分かったことを示すと、会計不正の報告があった企業は、東芝のように業績悪化が引き金になっていると思われる企業もあるが、必ずしも業績が悪くないのにも関わらず会計不正を行う企業もあり、財務パフォーマンスだけを見ただけでは一概には言えないということ。今回の対象となった期間（2008 年～2021 年）には、リーマンショックに代表される金融危機の直後の期間と、新型コロナウイルス禍がちょうど始まった時期が含まれており、センチメント分析を行ったとしても、業種や企業固有の事情により、ネガティブな用語が必ずしも会計不正の発生とは結びつかない傾向がみられること（いくつかのサンプルを使った予備的な解析の結果という限定はあるが）、TF・IDF 分析をするにしても、母集団となる文書の範囲を、会計不正が報告された企業のみを対象とするのか、不正をおこなった企業と同規模同業種の企業とのマッチドペア解析でおこなうのか、それとも市場全体の企業を対象とするのかによっても、そのスコアが大きく異なるはずであり、一概には不正会計をおこなう企業の特徴を抽出することが困難であること、等々である。市場全体を母集団として TF・IDF 分析を実施することについては、現在研究室に導入している高性能ワークステーションでも、主として搭載メモリの関係でデータを分割して行わなければならないことが、解析の実施を困難なものにしている。本研究課題の研究期間は終了するが、ヒトゲノム解析が十数年を要したように、全上場企業を対象としたテキストマイニングについては、これからも諦めずに続けていく所存である。

なお本研究の期間が終わった後も引き続き分析を継続するが、当初予定していたコーパスの確定、頻度分析、TF・IDF 分析等の完了の目途がたった暁には、その成果を書籍として刊行する予定である。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 1件）

1. 著者名 坂上 学	4. 巻 201
2. 論文標題 会計領域における先端複合研究の動向	5. 発行年 2022年
3. 雑誌名 会計	6. 最初と最後の頁 167-175
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂上 学	4. 巻 73
2. 論文標題 会計DXの歩み	5. 発行年 2021年
3. 雑誌名 企業会計	6. 最初と最後の頁 16-30
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂上 学	4. 巻 32
2. 論文標題 AIを用いた会計研究と機械学習アルゴリズムの理解	5. 発行年 2020年
3. 雑誌名 会計・監査ジャーナル	6. 最初と最後の頁 647-692
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂上 学	4. 巻 73
2. 論文標題 研究方法のパラダイムシフトに向けて	5. 発行年 2021年
3. 雑誌名 企業会計	6. 最初と最後の頁 44-45
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂上 学	4. 巻 30
2. 論文標題 ITやフィンテックが税務会計研究に与える影響	5. 発行年 2019年
3. 雑誌名 税務会計研究	6. 最初と最後の頁 57-72
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 坂上 学、田口 聡志、上枝 正幸、廣瀬 喜貴	4. 巻 17
2. 論文標題 実験会計研究の未来	5. 発行年 2020年
3. 雑誌名 イノベーション・マネジメント	6. 最初と最後の頁 21~37
掲載論文のDOI (デジタルオブジェクト識別子) 10.24677/riim.17.0_21	査読の有無 無
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計5件(うち招待講演 0件/うち国際学会 0件)

1. 発表者名 坂上 学
2. 発表標題 先端複合研究の動向
3. 学会等名 日本会計研究学会
4. 発表年 2021年

1. 発表者名 坂上 学
2. 発表標題 統一論題解題
3. 学会等名 日本簿記学会第35回全国大会(8月24日、中央大学)
4. 発表年 2019年

1. 発表者名 坂上 学
2. 発表標題 ワークショップ解題
3. 学会等名 神戸大学RIAM第101回ワークショップ(9月29日、神戸大学)
4. 発表年 2019年

1. 発表者名 坂上 学
2. 発表標題 ITやフィンテック等の発展が税務会計研究に与える影響
3. 学会等名 税務会計研究学会第30回全国大会
4. 発表年 2018年

1. 発表者名 坂上 学
2. 発表標題 研究方法のパラダイムシフトをもたらすもの
3. 学会等名 日本経営分析学会・日本ディスクロージャー研究学会連合大会
4. 発表年 2018年

〔図書〕 計8件

1. 著者名 野口昌良・清水泰洋・中村恒彦・本間正人・北浦貴士編著	4. 発行年 2020年
2. 出版社 中央経済社	5. 総ページ数 187
3. 書名 会計のヒストリー-80	

1. 著者名 瀧博編著	4. 発行年 2020年
2. 出版社 同文館出版	5. 総ページ数 278
3. 書名 日本監査研究学会リサーチシリーズ テクノロジーの進化と監査 - AIとデジタル技術が拓く新たな監査の可能性 -	

1. 著者名 河崎照行編著	4. 発行年 2020年
2. 出版社 中央経済社	5. 総ページ数 496
3. 書名 会計研究の挑戦 - 理論と制度における「知」の融合	

1. 著者名 岩崎勇編著	4. 発行年 2021年
2. 出版社 税務経理協会	5. 総ページ数 168
3. 書名 AI時代に複式簿記は終焉するか	

1. 著者名 菊谷 正人	4. 発行年 2019年
2. 出版社 税務経理協会	5. 総ページ数 504
3. 書名 会計学と租税法の現状と課題	

1. 著者名 橋本尚	4. 発行年 2019年
2. 出版社 同文館出版(株)	5. 総ページ数 524
3. 書名 現代会計の基礎と展開	

1. 著者名 河崎照行編	4. 発行年 2019年
2. 出版社 中央経済社	5. 総ページ数 424
3. 書名 会計制度のパラダイムシフト 経済社会の変化が与える影響	

1. 著者名 原俊雄編	4. 発行年 2019年
2. 出版社 中央経済社	5. 総ページ数 216
3. 書名 簿記と帳簿組織の機能 歴史的・国際的視点から	

〔産業財産権〕

〔その他〕

<p>「坂上大学の業績一覧」『法政大学経営学部 坂上研究室』 http://www.sakaue-lab.org/research/works/ 「坂上大学の業績一覧」『法政大学経営学部 坂上研究室』 http://www.sakaue-lab.org/research/works/ 法政大学経営学部 坂上研究室 http://www.sakaue-lab.org/</p>
--

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------