様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 １３ 日現在

機関番号：１２６０８
研究種目：挑戦的研究（萌芽）
研究期間：2018～2021
課題番号：１８Ｋ１８５６６

研究課題名（和文）テキスト分析による有価証券報告書の文章情報の情報価値の分析

研究課題名（英文）Analysis of the value of textual information in annual reports through natural language processing

研究代表者
井上　光太郎（Inoue, Kotaro）

東京工業大学・工学院・教授

研究者番号：９０３８１９０４

交付決定額（研究期間全体）：（直接経費）　　4,800,000円

研究成果の概要（和文）：本研究では、有価証券報告書のテキスト情報を自然言語処理により分析し、定性データが定量データの持たない情報価値を持つかを検証した。分析結果として有価証券報告書のテキスト情報が定量データのみの予測より将来のM&Aの発生を高い確率で予測可能であること、有価証券報告書の「事業等のリスク」の記載内容がその企業の翌期の株価に対し説明力を持つこと、内閣府令による有価証券報告書の記述内容の充実要請に基づく有価証券報告書の記載変化が、企業と投資家の間の情報非対称性を緩和効果を持ち、株式流動性の改善に貢献することを示した。上記の一連の研究は、２本の査読付論文、2本の招待論文として学術誌に掲載（掲載決定含む）した。

研究成果の学術的意義や社会的意義
本研究の成果は、従来の会計数値などの定量的データの充実に加え、テキストデータの充実が、企業と投資家の間の情報の非対称性の緩和に貢献し、株式市場の流動性改善につながることを示した。自然言語処理による企業開示情報の分析が市場効率性や事業戦略の予測に貢献することを示した点は学術的な意義がある。また、企業の開示情報における定性的テキスト情報の充実が、会計数値などでは観測できない情報を投資家に提供することを示し、制度的な対応の意義があることを示した点で今後の制度設計に有用な情報を提供しており、社会的意義がある。

研究成果の概要（英文）：In this study, textual information in corporate annual reports was analyzed using natural language processing to test whether textual data has valuable information that quantitative data does not. As a result of the analysis, we found that textual information in annual reports can predict the occurrence of future M&A with higher probability than quantitative data alone. The "Business and Other Risks" section of the annual reports have explanatory power with respect to the company's stock price in the following fiscal year. The changes in the content of annual reports based on the Cabinet Office Ordinance on Enhancing the Contents of Annual Reports have the effect of mitigating the information asymmetry between firms and investors, and contribute to improving stock liquidity.
The above series of studies were published as two peer-reviewed papers and two invited papers in academic journals.

研究分野：ファイナンス

キーワード：テキストマイニング　有価証券報告書　ファイナンス　開示制度　市場流動性　M&A　リスクプレミアム

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

株式市場における個別企業の株価形成や、企業経営者の投資意思決定は、過去の企業の財務情報や定量的な市場の予測情報のみならず、経営者自身による外部環境認識、企業の強みや課題、または研究開発の状況等にも影響を受けることを考える。後者の情報は、従来のファイナンス研究で用いられてきた財務情報や株価情報などでは計測困難な情報である。一方で、そうした情報は、企業や経営者が自社の将来の見通しや直面する課題を記述した文章情報から読み取れる可能性がある。筆者たちは、近年、発展の著しい自然言語処理手法を活用し、企業が発表する自社の事業に関するテキスト情報を分析することで各企業の事業の詳細な内容や経営者のリスク認識などを変数化し、これを用いて既存の仮説を検証することで新たな知見を得られるのではないかとの見通しをもった。

２．研究の目的

本研究は、従来は株価、財務数値、株主構成データなど専ら数値データを用いた研究が中心だった経営財務研究分野に、自然言語処理を用いたテキスト分析手法を導入し、テキストで提供される定性情報の有用性を明らかにすることを目的として開始した。具体的には、全ての上場企業が発行する有価証券報告書の定性記述データが、従来はほとんど分析対象になっていないことに注目し、これを対象にテキスト分析を行うことで、既存のコーポレートファイナンス研究の複数の重要課題についての新たな知見が得られるかを探索した。
特に有価証券報告書のテキスト情報から（１）経営者が認識する自社の直面するリスクを変数化でいるか、（２）事業内容、事業戦略、または研究開発戦略に関する情報から企業の投資戦略を説明できるか、（３）テキスト情報の充実が、経営者と投資家の間の情報の非対称性を緩和し、株式市場の効率性の改善に資するかなどの論点を明らかにすることを目的とした。

３．研究の方法
本研究では、日本の全上場企業をカバーする企業自身が作成する有価証券報告書の文書情報に対してテキスト分析をいち早く適用し、この新たな研究アプローチの可能性を検証しようという挑戦的研究である。有価証券報告書は、全上場企業が定期的かつほぼ統一のフォーマットで発行しており、全上場企業を一律に分析対象とすることが可能な点、また内容が会計監査対象になっており正確な点でテキスト分析の対象として価値があるとの見通しを持っていた。
本研究の開始時点において、有価証券報告書を対象とするテキスト分析の先行研究は国内ではほとんどなかった。一方で、海外では少数だが研究成果が報告されており、研究アプローチについてはある程度の見通しがついた状態で開始した。研究においては、ファイナンス研究者である研究代表者の井上と研究分担者の池田に加え、自然言語処理および機械学習を専門とする研究分担者の中田という異分野の研究者がチームを組んで研究を開始した。
近年、統計的潜在意味解析を行うトピックモデルが提案され、テキストから「意味」を抽出することが可能となった。これは、複数の単語の共起性を潜在的意味（トピック）とみなした上で、テキストが複数の潜在的意味から構成されているという確率的生成モデルであり、本研究でも文章情報をトピックごとに分類するために採用している。さらにニューラルネットワークの研究をテキスト分析に取り入れたニューラル言語モデルである Word2Vec、単語の埋め込みを、文章や文の埋め込みに拡張した Doc2Vec なども採用し、最新の様々な手法によるテキストのベクトル化を試みた。これらの手法を用いて有価証券報告書のテキスト情報を変数化し、定量的検証モデルに反映させた。

４．研究成果
第１に、有価証券報告書の「事業等のリスク」の記載内容が、その企業の１期先の株式市場におけるリスクプレミアムを予測するかについて、機械学習手法である Word Embedding を用いて株式リスクに関する新たな独自の辞書を作成し、有価証券報告書に記載されたリスク情報が翌期の株価に反映される当該企業のリスクに対し説明力を持つことを示した。一方で有価証券報告書が企業のリスクに関する記述の毎年の変化が乏しく、株式市場に織り込まれた以上の新規情報を提供していないことを示した。この結果は、査読付き論文として『証券アナリストジャーナル』に掲載した。
第２に有価証券報告書に記載された事業内容、研究開発動向の企業間の類似性が将来の M&A の発生を予測するかを検証した。分析結果として有価証券報告書に記載された事業内容、研究開発動向の企業間の類似性が将来の M&A の発生確率に強い正の効果を持つこと、業種の異なる 2 社間の多角化 M&A において類似性が正の効果を持ち、範囲の経済性への期待が M&A の背景にあることを示した。この研究の成果は、査読付き論文として『金融経済研究』に掲載した。
第３に、内閣府令による有価証券報告書の記述内容の充実の要請による 2020 年 3 月期からの

有価証券報告者の記載内容の変化が、企業と投資家の間の情報非対称性を緩和したかを Difference in Difference 分析手法を用いて検証した。この検証結果は、2020 年 3 月期にそれ以前の時期に比較して有意に大きな記述内容の変化が観測できたこと、特に 4 大会計事務所の監査先企業において記述内容の著しい充実化が見られ、それらの企業で株式の流動性の大きな改善が見られたことを示した。この結果は、有価証券報告書のテキスト情報が経営者と投資家の間の情報の非対称性を緩和し、効率的な株価形成に資することを示す。この研究成果については、『証券アナリストジャーナル』への招待論文としての掲載が決定している。

第 4 に有価証券報告書の「事業等のリスク」から文章ベクトルを生成し、それを特徴量とする新しいファクターモデルを構築し、抽出したファクターを評価するために、アセットプライシングモデルの妥当性を評価する GRS 検定を行った。その結果、有価証券報告書から抽出したファクターを既存のファクターに追加することで、p-value が改善することを確認した。特徴量生成に文脈を考慮可能な BERT を用いた場合は、word2vec や LDA を用いた場合に比べて改善幅が大きいことを示し、自然言語処理の手法が、有価証券報告書からの情報抽出に重要な役割を果たしていることを示した。これらの成果は、査読付き国際会議 2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA 2021)で発表している。

第 5 に有価証券報告書の中から ESG 関連文を抽出する機械学習モデルの構築を行った。具体的には、有価証券報告書の経営方針項目及び事業等のリスク項目の文に対しアノテーションを行うことで ESG 関連文データセットを作成し、その ESG 関連文データセットを利用して BERT のファインチューニングを行い、ESG に関連する部分を抽出した。年度・企業が異なる検証データを使って検証を行った結果、BERT を ESG 関連文抽出タスクでファインチューニングすることで、年度・企業が異なるデータに対しても精度良く汎化し、ESG 関連文を抽出できることが分かった。さらに個別企業レベルで ESG 関連文を抽出すると共に、複数企業の ESG 関連文に関して集計することで、ESG 情報開示の動向の可視化に成功した。これらの研究成果については、第 26 回人工知能学会・金融情報学研究会で報告している。

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 佐藤隆清、池田直史、井上光太郎	４．巻 59 (1)
２．論文標題 有価証券報告書のテキストマイニングによる株式のリスクファクター分析	５．発行年 2021年
３．雑誌名 証券アナリストジャーナル	６．最初と最後の頁 99-111
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 有
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 小室幸人、池田直史、 井上光太郎	４．巻 45
２．論文標題 企業間の類似性とM&A：テキスト分析アプローチ	５．発行年 2022年
３．雑誌名 金融経済研究	６．最初と最後の頁 1-21
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 有
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 田中研人、木村遥介、中田和秀、井上光太郎	４．巻 forthcoming
２．論文標題 企業の情報開示と株式の市場流動性：記述定性情報のケース	５．発行年 2022年
３．雑誌名 証券アナリストジャーナル	６．最初と最後の頁 forthcoming
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 井上光太郎、中田和秀	４．巻 74
２．論文標題 自然言語処理の発展と有用性	５．発行年 2022年
３．雑誌名 企業会計	６．最初と最後の頁 16-26
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 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 土橋諒太，中田和秀
２．発表標題 BERTを用いた有価証券報告書からのESG関連文抽出
３．学会等名 第26回 人工知能学会 金融情報学研究会
４．発表年 2021年

１．発表者名 小室幸人，池田直史，井上光太郎
２．発表標題 企業間の類似性とM&A：テキスト分析アプローチ
３．学会等名 International Workshop: Digital Innovation in Finance（国際学会）
４．発表年 2018年

１．発表者名 佐藤隆清，池田直史，井上光太郎
２．発表標題 有価証券報告書のテキストマイニングによる株式のリスクファクター分析
３．学会等名 International Workshop: Digital Innovation in Finance（国際学会）
４．発表年 2018年

１．発表者名 小室幸人，池田直史，井上光太郎
２．発表標題 企業間の類似性とM&A：テキスト分析アプローチ
３．学会等名 日本経営財務研究学会大会
４．発表年 2018年

１．発表者名
佐藤隆清，池田直史，井上光太郎

２．発表標題
有価証券報告書のテキストマイニングによる株式のリスクファクター分析

３．学会等名
日本経営財務研究学会大会

４．発表年
2018年

１．発表者名
田中研人、木村遥介、中田和秀、井上光太郎

２．発表標題
企業の情報開示と株式の市場流動性：記述定性情報のケース

３．学会等名
日本ファイナンス学会第3回秋季研究大会

４．発表年
2021年

１．発表者名
Kota Ishizuka, Kazuhide Nakata

２．発表標題
Text Mining for Factor Modeling of Japanese Stock Performance

３．学会等名
the 2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA 2021)（国際学会）

４．発表年
2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
「有価証券報告書のテキストマイニングによる株式のリスクファクター分析」データ公開
http://www.me.titech.ac.jp/~inouelab/Webzisyo.pdf

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
研究分担者	中田　和秀 (Nakata Kazuhide) (00312984)	東京工業大学・工学院・教授 (12608)	
研究分担者	池田　直史 (Ikeda Naoshi) (90725243)	日本大学・法学部・准教授 (32665)	

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