

機械学習を用いた有価証券報告書のサステナビリティ情報の可視化

構造的トピックモデル (Structural Topic Model) による記述情報把握

経済調査部 研究員 中田 理恵
政策調査部 主席研究員 鈴木 裕

[要約]

- 2023 年 3 月末決算から、有価証券報告書には新たに「サステナビリティに関する考え方及び取組」の欄が追加された。本稿では機械学習の手法（構造的トピックモデル）を用いて、700 社の有価証券報告書を対象に記述内容のトピック（テーマ）の分布を可視化した。
- サステナビリティ開示初年度の各企業の対応は、期待通り、人的資本と環境関連の開示に重心が置かれたものであった。また、一部業種では自社の業態や企業価値への影響を踏まえてより重要性の高い項目に比重を置いた開示が確認された。総じて、新設された開示制度の狙いに応えた企業側の対応があったと言えそうだ。
- 今後は、この開示情報を誰がどのように利用したのかを探ることが課題となる。ESG 投資家が利用し、世界の ESG 投資を日本に呼び込むうえで役立つ開示であったのかを明らかにすることが期待される。開示政策により上場企業に課せられた新たな開示事務負担を上回るベネフィットが創出されているか否かを検証する必要がある。
- 開示によって、世界の ESG 資金が日本に向かうことで、資金調達の条件改善や、株価形成への好影響が観察されれば、企業側にとっても一層の良質な情報開示に向けて取り組みを加速する動機となるはずだ。

はじめに：新たに開示されたサステナビリティ情報、記述型式で把握に課題

企業経営における環境・社会課題への関心が急速に高まっているように見える。環境負荷や人権の観点からサプライチェーンを見直す必要が生じるなど、直接的に企業活動に影響を及ぼしているだけでなく、そうした企業に投資をする機関投資家の判断要素としても注目されるようになった。この背景には、いわゆる ESG 投資の存在感が高まっていることを指摘できるだろう。もし環境・社会・ガバナンスの諸課題に企業がどのような取り組みを見せているかを投資判断において重要な要素と考える ESG 投資が規模を拡大しているとするれば、企業側は資金調達や株

価値形成の観点からも ESG 投資家の目を引き付ける必要がある。投資家の関心が ESG 課題であるとなれば、証券市場行政においても、ESG 関連の情報開示を拡充する政策が打ち出されることになる。

実際、岸田文雄内閣の看板政策である「新しい資本主義」のロードマップを示した「新しい資本主義のグランドデザイン及び実行計画」（2022年6月7日閣議決定）では、「国による大規模かつ中期・戦略的な財政出動等と呼び水として、世界の ESG 資金を呼び込む」と記すなど（p. 21）、ESG 投資にかける政権の期待は大きい。

もともと、こうした期待がどれほど現実味のあるものかは、疑いの余地もないではない。よく知られている通り、ESG 投資を謳う金融商品の中には、単に宣伝上 ESG の用語をちりばめているだけで、実体は他の金融商品と大した違いはないものもある。こうした欺瞞的な ESG 対応は「グリーンウォッシュ」や「ESG ウォッシュ」と呼ばれている。また、米国を中心に ESG 投資が資産運用機関の受託者責任と合致するか否かについて疑問が持たれている。環境や社会問題への対応を重視するあまり、有利な投資機会を見逃しているとなれば、ESG 投資は資産運用機関の顧客の損失につながりかねないとの批判もある。

このように ESG 投資の規模や意義については、疑いの余地があるものの、わが国において ESG 投資の一層の拡大に向けて、上場企業の情報開示制度の整備が進められてきた。2023年3月末決算から、有価証券報告書の「経営方針、経営環境及び対処すべき課題等」と「事業等のリスク」の間に、新たに「サステナビリティに関する考え方及び取組」の欄が追加された¹。ESG 投資における投資判断に資するものとして開示が求められたものだ。サステナビリティ情報の開示にあたっては、各企業において自社の業態や経営環境、企業価値への影響等を踏まえ、サステナビリティ情報の重要性を判断することが求められている。

「サステナビリティ」の範囲に関しては、例えば、環境、社会、従業員、人権の尊重、腐敗防止、贈収賄防止、ガバナンス、サイバーセキュリティ、データセキュリティなどに関する事項がサステナビリティ情報に含まれ得ると考えられている。このうち、人的資本に関しては開示が義務付けられている。他の項目は各企業の適切な判断に委ねられるものの、この開示制度が設けられた経緯から、気候変動に関連する取り組みについても、法令上の義務ではないが、かなり強く開示が期待されている。つまり、人的資本と気候変動が、新たなサステナビリティ開示制度の中核にある。

ここまでの説明で明らかなように、サステナビリティ開示において、人的資本以外に開示されるサステナビリティ課題は各企業の判断によることになる。つまり、各企業でまちまちの内容が開示される可能性があるということだ。また開示形式は、文章中心の記述的开示であるため従来のような数的処理による全体把握が難しく、個々の企業の記述内容を読み込むために多大な時間を要することとなる。このような記述形式の情報開示におけるデメリットは、特に投資のユニバース全体の ESG 関連の取り組みや経営戦略について把握する際などに分析の障害とな

¹ 詳しくは、下記の大和総研レポートを参照されたい。

藤野大輝「[開示府令の改正が公布・施行](#)」大和総研レポート、2023年2月7日

り得よう。

そこで本稿では機械学習の手法を用いて、多数の企業の ESG 情報に関する記述の概要を効率的かつ視覚的に把握することを試みる。具体的には、機械学習による自然言語処理の一手法である構造的トピックモデル (Structural Topic Model、詳細は後述) を用いて有価証券報告書のサステナビリティに関する考え方及び取組欄の記述内容をトピック (テーマ) 別に分類し、業種別にどのようなトピックを重点的に記載しているか、またそのトピック内にてどのような単語が多用されているかを可視化する。

分析手法：構造的トピックモデルはより高精度のトピック分類・分析が可能

対象データ

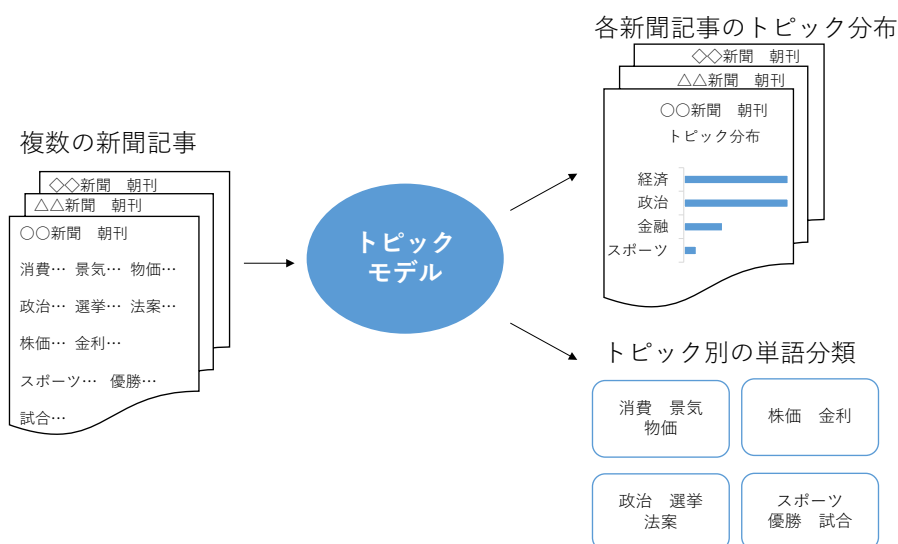
対象とする有価証券報告書は 2023 年 6 月末時点で EDINET (金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム) から取得可能な TOPIX1000 に属する企業の 2023 年 3 月期の有価証券報告書、700 社分とした。有価証券報告書中の「サステナビリティに関する考え方及び取組」欄全文及び、その内訳項目である戦略、指標及び目標欄に記載された内容を分析対象とする² (図表 1 左)。なお、その他の内訳項目であるガバナンス、リスク管理は記述量が少なく、かつトピックが限定されているため分析対象から除外した。業種別のサンプル数は図表 1 右の通りである。分析対象 700 社のうち、いくつかの企業はテキストデータが登録されていない等の理由で分析対象から除外している。また、1 社当たり各項目の平均字数は図表 2 の通りである。

² 企業の入力方式により EDINET に登録されている「サステナビリティに関する考え方及び取組」欄と内訳項目 (戦略、指標及び目標) の記載内容が一致しているものとそうでないものがあることに留意されたい。分析対象のデータでは、後者の内訳項目は概ね、前者の「サステナビリティに関する考え方及び取組」の該当箇所の抜粋となり記述内容は重複関係にあるが、うち 6～7 社 (項目により異なる) は「サステナビリティに関する考え方及び取組」欄と重複しない形で登録されている。

構造的トピックモデル (Structural Topic Model、STM) とは

トピックモデルとは複数の文書のテキストを読み込み、各文書に含まれる単語の出現頻度を元に、どのようなトピック (テーマ、話題) が潜在的に存在するかを推定する手法の一つである。例えば複数のトピックを持つ新聞記事のデータをモデルに読ませることで、各新聞記事におけるトピックの分布の推定やトピック別に単語を分類することができる (図表3)。

図表3 トピックモデル概念図



(出所) 大和総研作成

トピックモデルの代表的手法である「潜在的ディリクレ配分法 (LDA)」では、各文書に含まれる単語の出現頻度及び事前に指定したトピック数を元に、各文書を構成するトピックの分布及び各トピックに属する単語の分布を推定する。

本分析では LDA を改良した「構造的トピックモデル」を用いた。LDA はトピックの推定においてテキスト内の単語出現頻度のみを参照するが、構造的トピックモデルはテキストに加えて、その他の特徴量を元に推定する。このため、より詳細なトピック分類・分析が可能となる。例えば新聞記事のテキストから複数のトピックを推定する際に、その記事を掲載した新聞社はどこであるかといった情報を加味することで、新聞社ごとによく取り上げるトピックの違いに配慮した推定や、同一のトピックでも新聞社により異なる単語を使う傾向にある場合でも一つのトピックとして認識するといったこと等が可能となる。有価証券報告書のサステナビリティに関する情報では業種ごとに重点分野が異なることが推測されるため、業種情報を特徴量として取り込むことができる構造的トピックモデルの使用によって、推定結果がより精緻となることが期待できる。

モデルの適用

モデルを適用するにあたり、テキストを単語（形態素）ごとに分割し、出現頻度が 20 文書以下と少ない単語は取り除いた。また、トピックモデルではトピック分割にあたってトピック数をあらかじめ指定する必要がある。5~25 のトピック数を検討し、複数の評価指標の値を参考に 20 に決定した。推定に使用するデータとしてはテキストのほか、17 業種分類に基づく業種情報とテキストの記載欄（全体、戦略、指標及び目標）を特徴量に加えた。

トピック推定結果：企業は人的資本と環境対応の開示に重心、一部業種ではより重要性の高い項目に比重を置いた開示も

トピック一覧：持続的成長、環境関連、人的資本関連などのトピックを抽出

トピックの推定結果は**図表 4**の通りである。**図表 4**ではトピックと当該トピックに属する特徴的な単語を並べている。トピックごとの特徴的単語はモデルが算出した FREX という値に基づき上位 15 単語を抽出した。FREX はトピック内における単語の出現頻度の高さ与其他のトピックと比較した排他性（特定のトピックによく出現し、他のトピックでは出現しにくい）の両者を加味した値である。トピックの名称はトピックモデルが分類した単語に沿って筆者が決定している。

トピック一覧を見ると、持続的成長、環境関連、人的資本関連、IT、生産関連、コンプライアンス、ガバナンス、リスク管理体制等が抽出されている。そのほか、金融や建設関連等特定業種と関連の強いトピックもみられる。

各トピックに属する単語グループの内容を見ると、人による直感的な分類と近い形で分類できていると言えよう。なお、モデルの仕様上、単語が各トピックに出現する確率に基づいて割り振られているため、表中の一部の単語は複数のトピックに属している。例えば人材という単語は「人材戦略」、「人材多様性」のどちらにも属している。

図表4 トピック及びトピックに属する単語一覧

トピック	トピックに属する単語 (FREX 上位 15 単語)
持続的成長	持続 発展 社会 解決 地球環境 貢献 可能 経営理念 持続的 社会的 豊か 地球 企業価値 創造 理念
気候変動	TCFD 気候変動 気候 提言 情報開示 シナリオ分析 認識 タスクフォース 関連 財務 リスク 開示 影響 機会 賛同
気候リスク評価	気温 需要 コスト シナリオ 炭素税 上昇 C 価格 激甚 物理 政策 小 増加 減少 IEA
温室効果ガス	SCOPE TCO 排出 算定 間接 CO2 温室効果ガス 量 排出削減 カーボンニュートラル 電気 電力 熱 スコープ GHG
人材戦略	人材戦略 リーダー サーベイ 自律 社員 人事 層 DX 人材 キャリア 自ら 配置 幹部 異動 ポジション
人材育成	人材 育成方針 育成 自律 挑戦 キャリア 醸成 変革 スキル 財 一人ひとり 活躍 心理的 風土 自律的
人材多様性	人材 人材育成 国籍 性別 能力 職場環境 整備 社内 発揮 個性 階層 確保 十分 最大限 多様
女性活躍推進	男性 割合 管理職 育児休業 男女 職 女性 賃金 中途採用 女性労働 労働者 差異 登用 外国人 比率
勤務制度、WLB	勤務 育児 優良 法人 休暇 介護 健康経営 短時間 休職 面談 研修 定年 ストレスチェック 受診 在宅
IT	データセンター IT 創 ソリューション ネットワーク デジタル パーパス 領域 投資 テクノロジー 経済 サービス 自然資本 通信 モデル
生産	原料 生産 プラスチック 原材料 リサイクル 食 工場 製造 調達 素材 資源 製品 包装 生産拠点 食品
安全・品質管理	事故 品質 電動 水素 安全 化学物質 技術 重大 船舶 発生件数 製品 クリーン 電池 災害 自動車
コンプライアンス	人権 情報セキュリティ 規範 人権尊重 法令 遵守 防止 デューデリジェンス 指針 取引 腐敗 セキュリティ ハラスメント 尊重 compliance
金融	当行 投融資 セクター 与信 ファイナンス 融資 信用リスク 金融 銀行 額 コンサルティング 態勢 毀損 お客様 職員
建設	期 施工 物件 工事 建設 住宅 建築 不動産 取扱 営業 受注 建物 メンテナンス 性能 作業
ガバナンス	委員長 審議 監督 取締役会 委員会 会議 代表取締役社長 事項 ガバナンス 取締役 執行役員 委員 決議 議長 執行
リスク管理体制	リスク管理 リスクマネジメント 識別 リスク 規程 全社 委員会 報告 統括 部署 管理 策 内部統制 評価 部門
値	件 単体 値 率 比 実績 KPI 累計 株 指標 目標 受講 件数 スコア 毎年
WEB サイト関連	www https coj マテリアリティ jp 詳細 sta abi 参照 web サイト in com lit ご覧 su
その他	執行役 株 医療 医薬品 DE&I 患者 エクイティ DEI CEO 各社 手当 報酬 世界 コミットメント ミッション

(注) トピックの名称はトピックモデルが分類した単語に沿って筆者が決定した。

(出所) 各社有価証券報告書をもとに大和総研作成

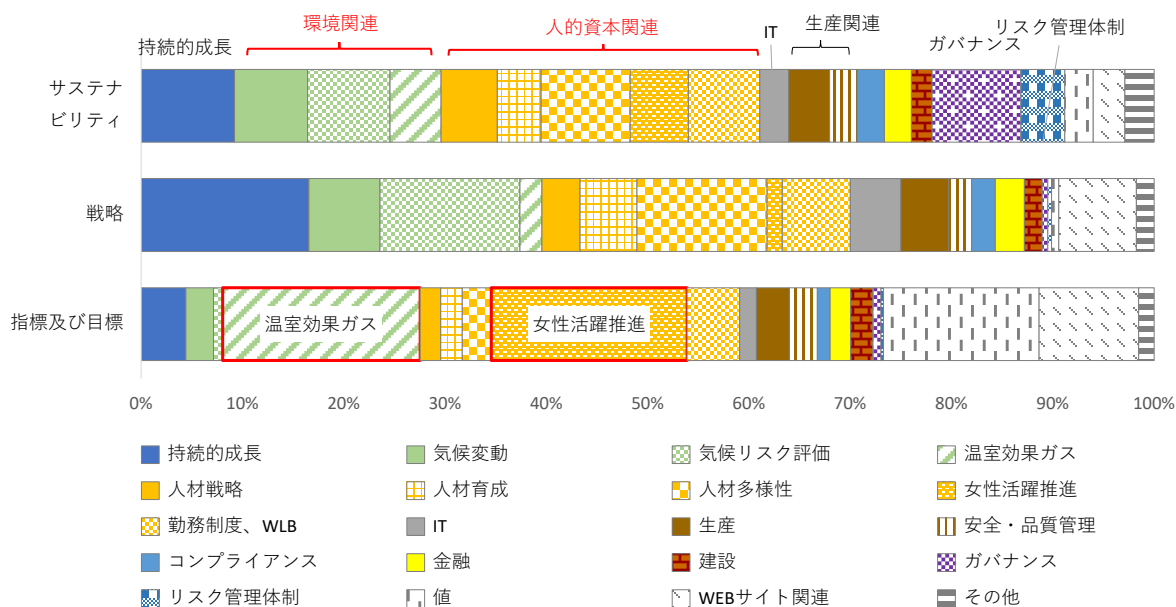
項目別トピック分布：人的資本関連と環境関連が過半を占める

トピックモデルでは文書ごとのトピックの分布も算出される。**図表 5**は、「サステナビリティに関する考え方及び取組」欄の全体と内訳項目（戦略、指標及び目標）のトピック分布（トピックごとの記述量の割合）を示している。図表を見やすくするため筆者の判断で20トピックのうち一部を環境関連、人的資本関連、生産関連として同系色でまとめた。概ね、実際の記述内容を読んだ印象と近い形でのトピック分布となっている。ただし、先に述べた通りモデルによる推定値であるため、実際の記述の分布と正確に一致しているとは限らない点には留意されたい。

まず、全体を見てみると人的資本関連と環境関連で過半を占めている。人的資本については開示が義務付けられており、また、環境関連の取り組みも金融審議会ディスクロージャーワーキング・グループにて記述を推奨されていたトピックである。企業側は概ね期待に応える形で両トピックに重心を置いて開示したとみられる。細かい内訳を見ると、環境関連では、気候変動、気候リスク評価、温室効果ガス関連のトピックが概ね同程度の割合で記述されており、人的資本関連も同様に人材配置、人材育成、人材多様性、女性活躍推進、勤務制度・WLB等が同程度の割合を占めている。

内訳項目の一つである戦略のトピック分布は、「サステナビリティに関する考え方及び取組」全体の記述とほとんど同様である。より細かく見ると、環境関連では気候リスク評価に関する議論、人的資本関連では人材多様性に関する議論がより重点的に述べられている傾向にある。

もう一つの内訳項目である指標及び目標欄では、環境関連では「温室効果ガス」、人的資本関連では「女性活躍」が重点的に言及されている。戦略と指標及び目標において、環境関連・人的資本関連の内訳項目の言及割合に差があるが、この点は既に指標として開示が求められているトピックについて記述が偏った可能性がある。東京証券取引所のコーポレートガバナンス・コードでは、プライム市場上場会社を中心にTCFD（気候関連財務情報開示タスクフォース）に基づいた開示を求めている。このため既に自社の温室効果ガスの排出量や目標値を把握・設定している企業が今回の分析対象に多いことが、サステナビリティ情報の指標・目標欄における「温室効果ガス」トピックへの高い言及割合につながった可能性がある。また、「女性活躍」については、有価証券報告書において女性活躍関連の定量情報（女性管理職比率、男女間賃金格差等）の公開が求められていることが影響したと推測される。

図表5 「サステナビリティに関する考え方及び取組」欄におけるトピック分布推定結果


(注) 各項目のサンプル数はサステナビリティ：693社、戦略：601社、指標及び目標：601社。

(出所) 各社有価証券報告書をもとに大和総研作成

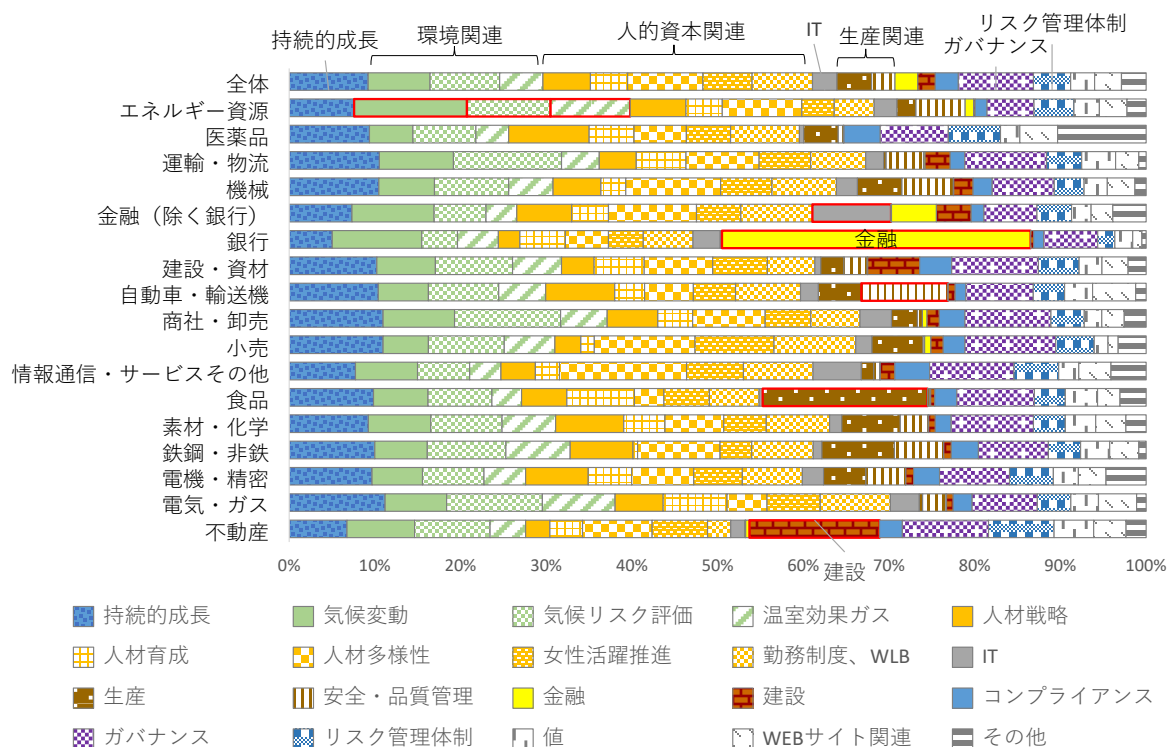
業種別トピック分布：一部業種は自社ビジネスと関連性の強いトピックに比重を置いて開示

図表6は「サステナビリティに関する考え方及び取組」欄のトピック分布を業種別で示している。ほとんどの業種が全体平均の分布と近い分布になっているが、いくつかの業種では、自社のビジネスに関連性の高いトピックにより比重を置いた開示をしていることが分かる。

例えば、エネルギー資源業では環境関連への言及割合が高い。業務の性質上、自然環境に影響を与えやすく同時に受けやすいため、特に重点的に述べられていると考えられる。同様に、銀行業は「金融」、自動車・輸送機業は「安全・品質関連」、食品業は「生産」、不動産業は「建設」のトピックへの言及割合が高くなっている。また、金融（除く銀行）では「IT」トピックへの言及割合が高い。実際の記述を見ると、経営における重要課題（マテリアリティ）としてデジタルトランスフォーメーションや新たなビジネスモデルの創出を挙げているケースや、人的資本に関する開示でIT人材の確保・育成方針を記述しているケースが散見され、金融業の持続的成長においてITの活用が重要視されていることが分かる。

サステナビリティ情報の開示規定では、各社が自社ビジネスにとって重要性・関連性の高い項目に重点を置いて開示ができるように敢えて開示項目を各企業の判断に委ねたが、業種別に見た結果では、ある程度施策の意図に沿った開示がなされているようである。

図表6 「サステナビリティに関する考え方及び取組」欄の業種別トピック分布推定結果



(注) サンプル数 693 社。

(出所) 各社有価証券報告書より大和総研作成

おわりに：開示制度の狙いに応えた企業側対応、今後の焦点は利用の実態へ

本稿では、構造的トピックモデルを用いて有価証券報告書のサステナビリティに関する記述情報のトピック分布を可視化した。分析の結果、サステナビリティ開示初年度の各企業の対応は、期待通り、人的資本と環境関連の開示に重心が置かれたものであった。また、一部業種では自社の業態や企業価値への影響を踏まえて、より重要性の高い項目に比重を置いた開示が確認された。総じて、新設された開示制度の狙いに応えた企業側の対応があったと言えそうだ。また同時に、記述的开示の分析における機械学習の有用性を示せたと言えよう。

今後は、この開示情報を誰がどのように利用したのかを探ることが課題となる。ESG 投資家が利用し、世界の ESG 投資を日本に呼び込むうえで役立つ開示であったのかを明らかにし、開示政策のベネフィットを測定することが期待される。

そもそも ESG 投資はその規模が不明であることに加えて、受託者責任上の摩擦の懸念も残る投資手法を採用する投資資金を日本に呼び込もうとして、そのための開示制度を新設し、上場企業に新たな開示事務負担を負わせたのであるから、それを上回るベネフィットが果たして開示制度にあったのかは、今後、検証の必要があるだろう。もし開示によって、世界の ESG 資金が日本に向かうことで、資金調達条件が改善したり、株価形成に好影響が観察されたりすれば、企業側にとっても一層の良質な情報開示に向けて取り組みを加速する動機となるはずだ。