

# BERT を用いた 気候変動リスク文の抽出と可視化

Short Review  
2023年2月

投資工学研究所  
川崎 正勝

## 1. はじめに

気候変動問題への世界的な関心の高まりから、ESG 投資やサステナブル・ファイナンスの取組が重要となっている。投資家や金融機関にとって、企業の気候変動問題への取組は重要な判断要素であり、企業の取組に関する情報開示の枠組みや法令は、近年急速に整備されている。

2017年6月にTCFD<sup>1</sup>は、企業など<sup>2</sup>に対し気候変動の情報開示の推奨と開示の枠組みを提言した(TCFD [2017])。開示のフレームワークとして、「ガバナンス」、「戦略」、「リスク管理」、「指標と目標」の4項目が開示推奨項目とされ、また、重要な要素として気候変動のリスクと機会が11項目(リスク6項目、機会5項目)に分類された。

日本においては、2021年6月に改訂された「コーポレートガバナンス・コード」(東京証券取引所 [2021])により、東京証券取引所のプライム市場上場企業(プライム企業)は「TCFD またはそれと同等の枠組みに基づく開示の質と量の充実」が求められ、コーポレート・ガバナンス報告書において対応することとなった<sup>3</sup>。

また、2022年11月に「企業内容等の開示に関する内閣府令」(金融庁 [2022])が公表され、有価証券報告書の記載事項について改正がなされた。2023年3月期以降の有価証券報告書では、「サステナビリティに関する考え方及び取組」の記載欄が新設され、「ガバナンス」、「リスク管理」については必須記載事項、「戦略」、「指標と目標」については重要性に応じて記載が求められる。

一方、テキスト分析の分野では企業の開示文書から情報を抽出する研究が進んでいる。例えば、土橋ら[土橋・中田 2022]は、有価証券報告書から機械学習モデルを用いて ESG に関連する文を抽出し、ESG 情報の開示は、時価総額が大きい企業ほど積極的であることを示した。また、若月ら[若月他 2022]はテキスト分析を用いて、統合報告書の評価スコアの算出を行っている。

本稿では、気候変動に関する情報開示の要請が急速に強まっている中での企業の対応状況について、TCFD 提言における気候変動リスク6項目の記載に着目する。機械学習の手法を用いて開示情報から気候変動リスク6項目に該当する気候変動リスク文を抽出し、可視化することで、企業の気候変動リスクの情報開示の変化を定量的に分析する。対象とする開示文書は、企業の法的な開示書類であり、統一さ

<sup>1</sup> TCFD とは、「Task Force on Climate-related Financial Disclosures」の略で金融安定理事会によって気候関連の情報開示及び金融機関の対応をどのように行うかを検討するために設立された気候関連財務情報開示タスクフォースのことである。

<sup>2</sup> 債券や株式を発行する全ての組織 ((TCFD [2017]))

<sup>3</sup> 2022年4月4日以降に開催される定時株主総会後に提出されるコーポレート・ガバナンス報告書が対象。

れたフォーマットで記載される有価証券報告書とする。

本稿の構成は、以下の通りである。2章で気候変動リスク6項目に該当する文を抽出するモデル（抽出モデル）の作成に関して説明する。2.1では抽出対象とするテキストデータについて、2.2では抽出モデルの学習に用いるデータのラベリングについて、2.3では抽出モデルについて説明する。3章で抽出モデルを使って抽出した結果を用いて、気候変動リスク6項目に関する記載量の変化を時系列で分析する。

## 2. 抽出モデルの作成

### 2.1. 対象テキストデータ

本稿では、2017年～2022年の3月が本決算の企業で、有価証券報告書が取得できた企業を対象企業とする。対象企業の有価証券報告書の「事業等のリスク」に記載されている文を、句点や改行などを考慮したルールベースにより一文ずつに分割し、一文が15文字以上の文を、抽出対象の文（以下、対象テキストデータ）とした。年度ごとの企業数と文数を図表1に示す。

図表1 対象テキストデータの基本情報

年度	2017	2018	2019	2020	2021	2022
企業数	2,422	2,419	2,399	2,387	2,368	2,345
文数	87,849	90,334	92,290	120,736	126,121	132,544

（出所）日興リサーチセンター作成

### 2.2. データラベリングについて

本稿ではTCFD提言に基づき気候変動リスクを6項目とする。各リスクとその具体例を図表2に示す。気候変動リスクは「移行リスク」と「物理的リスク」の2つのカテゴリーに分類される。「移行リスク」は、脱炭素社会への移行に伴う幅広い分野（規制、技術、市場、評判）におけるリスクであり、「物理的リスク」は、台風や洪水などの異常気象の激甚化（急性）と、より長期的な気象の変化や影響（慢性）に分類される。

また、単に該当する単語があるかではなく、重要な文を抽出するために3つの条件に当てはまる文を抽出対象とする。条件は以下の通り。

- ① リスクの原因が気候変動に関連している
- ② リスクの内容が具体的である
- ③ リスクが企業に影響を与えている

図表 2 気候変動リスク 6 項目とその具体例

リスク カテゴリー	リスク 項目	具体例
移行 リスク	規制	炭素税、GHG 排出規制、製品への規制強化、運営コストの増加、開示義務
	技術	低炭素製品への置き換え、新技術や設備への投資の先行投資や失敗
	市場	エネルギーミックスの変化、消費者の嗜好の変化、原材料コストの上昇
	評判	顧客・投資家の評判の変化、ステークホルダーの懸念の増大
物理的 リスク	急性	異常気象の激甚化
	慢性	平均気温の上昇、降水・気象パターンの変化、水不足

(出所) 日興リサーチセンター作成

2.1 の対象テキストデータから、気候変動に関連する文をキーワード（「二酸化炭素」や「気候変動」などの一般的な語）によるスクリーニングで抽出した。抽出したテキストデータに対し、気候変動リスク各 6 項目の正例データがそれぞれ 200 を超えるようにラベリングを行い、学習用データを作成した。ラベリングは 4 人で行い、正例・負例<sup>4</sup>を多数決で判定し、多数決で判定できない場合や信頼性に疑義が生じた場合は再度相談して合議により判定した。

### 2.3. 抽出モデル

機械学習モデルには、代表的な言語モデルである BERT (Bidirectional Encoder Representations from Transformers) を用いる。BERT は、大規模教師なしデータを用いる双方向型の Transformer による事前学習モデルであり、少量の教師ありデータで Finetuning することにより、汎用性が高く、高性能な予測モデルを作成することができると言われている。

本稿では、気候変動リスク 6 項目について抽出モデルをそれぞれ作成する。2.2 で作成した学習用データを用いて Finetuning を行い、訓練データ、検証データ、評価データそれぞれで正例データと負例データが 1:1 となるように分割した。実装においては、BERT は東北大学乾・鈴木研究室が公開している日本語 Wikipedia の学習モデル<sup>5</sup>を用いた。

気候変動リスクのリスク項目ごとの評価結果は図表 3 に示す。F 値は適合率と再現率の調和平均であり、二項分類の統計において一般的な指標である。また、適合率は正例と予測したもののうち、本当に正例だったものの割合で、再現率は本当の正例のうち、正例として予測した割合であり、適合率と再現率はトレードオフの関係にある。「評判」以外の項目で F 値は 0.8 を超えた。また、全体的に適合率より再現率の方が高く、正例判定を出しやすいモデルであると解釈できる。

<sup>4</sup> 正例…気候変動リスク各 6 項目に該当する抽出対象の文  
負例…気候変動リスク各 6 項目に該当しない抽出対象外の文

<sup>5</sup> 下記 HP にて、Pretrained Japanese BERT models released / 日本語 BERT モデルは公開されている。  
(<https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>)

図表 3 気候変動リスクのリスク項目ごとの評価結果

リスク項目	F 値	適合率	再現率
規制	0.87	0.82	0.92
技術	0.93	0.92	0.94
市場	0.87	0.81	0.94
評判	0.77	0.67	0.90
急性	0.86	0.80	0.94
慢性	0.82	0.82	0.82

(出所) 日興リサーチセンター作成

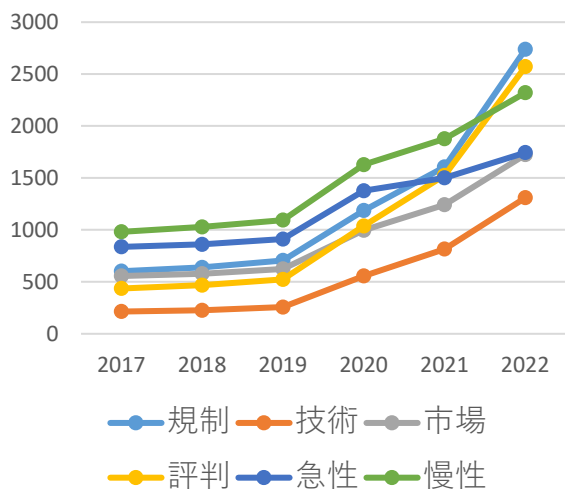
### 3. 時系列での気候変動リスクの抽出結果の可視化

2章で作成した抽出モデル（リスク項目ごとの6種類）による気候変動リスクの抽出結果を、文・企業ごとに、記載数、記載割合、増加率の時系列（2017年度～2022年度）で集計した（図表4・5）。

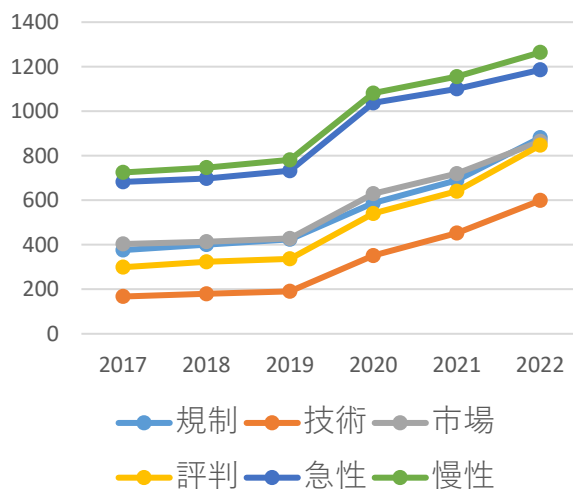
図表4-1の記載文数、図表5-1の記載企業数、図表4-2の記載文割合、図表5-2の記載企業割合を確認すると、すべての項目、すべての年度において増加している。これは、近年の気候関連情報の開示の質と量の充実という動きと合致している。また、図表4-3の記載文増加率、図表5-3の記載企業増加率をみると、特に2020年度からすべての項目で急激に増加している。これは、2019年1月に行われた有価証券報告書の開示の充実を求める内閣府令（金融庁[2019]）の影響だと考えられる。内閣府令は、「財務情報及び記述情報の充実」として「事業等のリスクについて、顕在化する可能性の程度や時期、リスクの事業へ与える影響の内容、リスクへの対応策の説明を求める」としている。また、2022年度は2021年度よりも記載文数、記載企業数の増加率が大きく、2021年の「コーポレートガバナンス・コード」（東京証券取引所[2021]）の改定による、プライム企業の気候変動リスクの開示義務化が影響していると思われる。

図表4-1、4-2と図表5-1、5-2をみると、記載文数、記載文割合では2020年度以前の年度で、記載企業数、記載企業割合では全ての年度で「物理的リスク」（急性、慢性）が多い。一方で、図表4-3、5-3の増加率を確認すると、傾向として、特に2020年度以降では、「移行リスク」（規制、技術、市場、評判）の増加率が大きく、個別にみると、技術と評判の増加率が大きい。気候変動リスクの整理と記載内容の充実に伴い、気候変動の直接的なリスクである災害の激甚化や気温の上昇などの「物理的リスク」だけでなく、それまで注目度の低かった「移行リスク」の記載が増えたと考えられる。

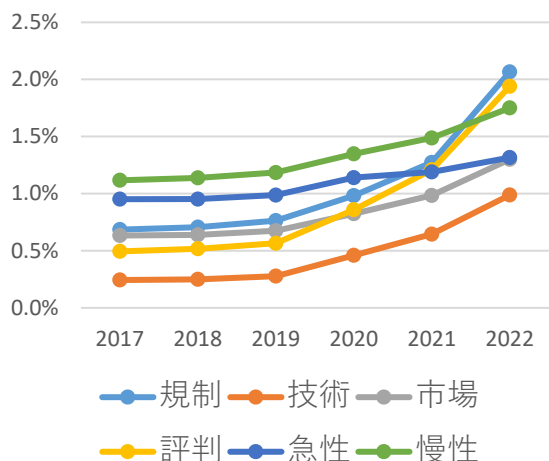
図表 4-1 記載文数



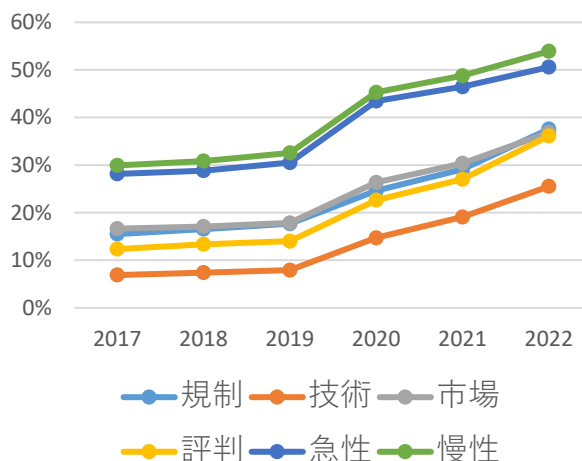
図表 5-1 記載企業数



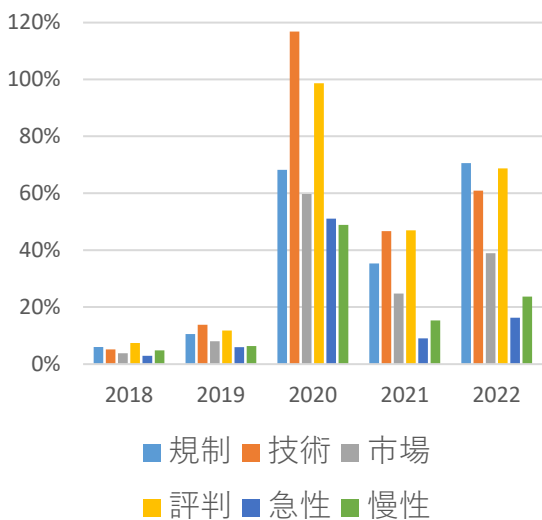
図表 4-2 記載文割合



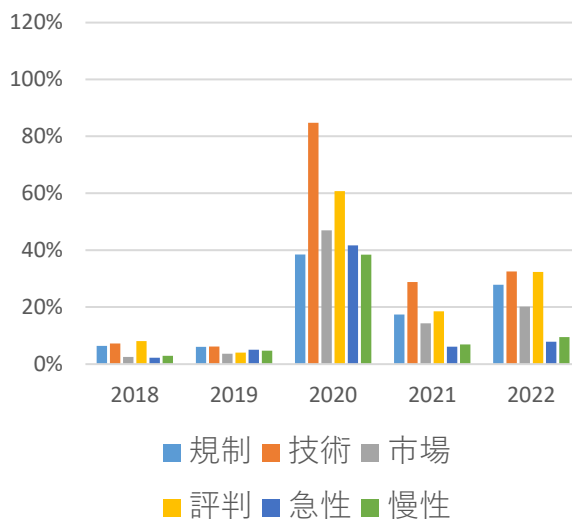
図表 5-2 記載企業割合



図表 4-3 記載文増加率



図表 5-3 記載企業増加率

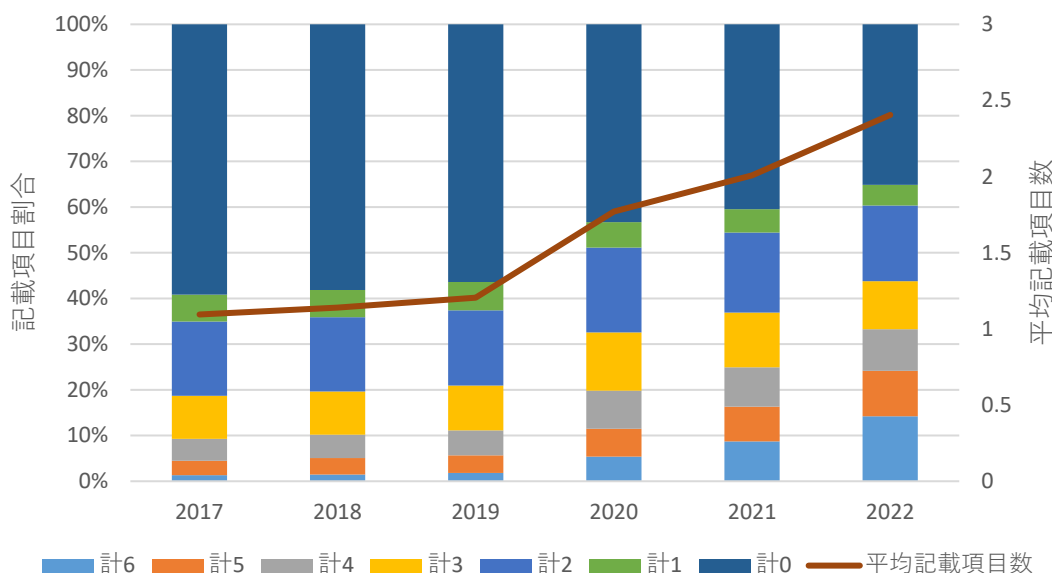


(出所) 各社有価証券報告書より日興リサーチセンター作成

次に、企業が気候変動リスクをいくつ記載しているかを図表 6 に示す。

図表 6 からは、企業の気候変動リスクの記載数が毎年増えていることが確認できる。平均記載リスク項目数は 2017 年度の 1.1 から 2022 年度の 2.4 になり、2 倍以上に増えている。また、2017 年度には気候変動リスクを一つも記載していない企業が 60%近くあったが、2020 年度に大きく減少し、2022 年では 35%となっている。特に 4 項目以上記載している企業が大きく増加しており、気候変動リスクの記載項目数が多ければリスクを丁寧に記載していると解釈すると、年々記載内容は充実してきていると言える。

図表 6 気候変動リスクの記載項目割合と平均項目数の推移



(出所) 各社有価証券報告書より日興リサーチセンター作成

#### 4. おわりに

本稿では、開示情報から TCFD 提言による気候変動 6 項目のリスクの記載量に着目して、機械学習の手法を用いて気候変動リスク文を抽出し、可視化することで、企業の気候変動リスクの情報開示の変化を定量的に分析した。

抽出した気候変動リスク文の記載量を確認すると、特に 2020 年度から急激に記載量が増えていることが分かった。リスク項目別にみると、「移行リスク」に関する記載の増加が大きく、企業の記載が「物理的リスク」だけでなく、「移行リスク」にも拡大している。また、企業の気候変動リスクの記載数は年々増加しており、気候変動リスクの記載のない企業は 2017 年度の 60%から 2022 年度には 35%まで減少し、平均記載リスク項目数は 1.1 から 2.4 まで倍増しており、年々記載が充実してきていること

が確認された。

今後の課題として、抽出モデルの拡張と、抽出対象の拡大が挙げられる。抽出モデルの拡張としては、本稿では気候変動の6項目のリスクに着目したが、更なる抽出対象として、TCFD提言における、気候変動の機会の5項目、さらに、「ガバナンス」、「戦略」、「リスク管理」、「指標と目標」の4項目などがある。今後、TCFD提言に沿った情報開示が期待されるが、企業の記載がどの項目に当てはまるかを確認することは、開示状況を確認する上で重要だと考える。また、抽出対象の拡大では、有価証券報告書だけでなく、統合報告書やCSRレポートなども対象となる。統合報告書やCSRレポートは有価証券報告書と異なり、記載の内容・形式は定まっておらず、自由記述となっている。分析には困難が伴うが、自由記述のため、企業固有の情報を捉えることができる可能性がある。

2023年度からは、有価証券報告書で、「サステナビリティに関する考え方及び取組」の記載欄が新設され、開示情報の更なる充実が予想される。今後も気候変動に関する情報に注目して、有益な分析情報の発信に努めたい。

## 参考文献

TCFD [2017], 「気候関連財務情報開示タスクフォースによる提言 (最終版)」

[https://assets.bbhub.io/company/sites/60/2020/10/TCFD\\_Final\\_Report\\_Japanese.pdf](https://assets.bbhub.io/company/sites/60/2020/10/TCFD_Final_Report_Japanese.pdf)

東京証券取引所 [2021], 「コーポレートガバナンス・コード」

<https://www.jpx.co.jp/news/1020/nlsgeu000005ln9r-att/nlsgeu000005lne9.pdf>

東京証券取引所 [2022], 「コーポレートガバナンス・コードへの対応状況について」

<https://www.jpx.co.jp/news/1020/nlsgeu000006jro6-att/nlsgeu000006jrqr.pdf>

金融庁 [2019], 「企業内容等の開示に関する内閣府令等の一部を改正する内閣府令」

<https://www.fsa.go.jp/news/30/sonota/20190131/191018.pdf>

金融庁 [2022], 「企業内容等の開示に関する内閣府令等の一部を改正する内閣府令」

<https://www.fsa.go.jp/news/r4/sonota/20230131/03.pdf>

土橋、中田 [2022], 「BERT を用いた有価証券報告書からの ESG 関連分抽出」、第 28 回 人工知能学会 金融情報学研究会

<https://sigfin.org/?028-25>

若月、南、河合、松原、花城 [2022], 「テキストマイニングを活用した統合報告書評価の試み」、第 26 回 人工知能学会 金融情報学研究会

<https://sigfin.org/?026-02>

(END)