

卒業論文

統合報告書におけるテキストマイニングを用いた ESG 関連語の抽出

公立はこだて未来大学
システム情報科学部 複雑系知能学科
複雑系コース 1019044

貴船 怜士

指導教員 新美 礼彦

提出日 2023 年 1 月 26 日

BA Thesis

ESG (Environmental, Social, and Governance)-Related Words Extraction from Integrated Reports Using Text Mining

by

Reiji Kifune

Complex Systems Course, Department of Complex and Intelligent Systems
School of Systems Information Science, Future University Hakodate

Supervisor: Ayahiko Niimi

Submitted on January 26th, 2023

Abstract

In recent years, environmental, social, and governance (ESG) investment, which considers companies' ESG factors, has been expanding. Therefore, knowing a company's ESG efforts (ESG information) is crucial. The integrated report is one way to obtain such information, but due to the large number of descriptions, it takes a lot of time and effort to obtain ESG information. Therefore, in this study, we attempt to solve the aforementioned problem by extracting ESG-related words from integrated reports using text mining and marking them on the text. The ESG-related words in this study are words similar to "environment," "society," and "governance," extracted using Word2Vec. The experimental results show that the marking of ESG-related words is effective in some cases and not in others. One possible reason for the lack of effectiveness is that the marking of ESG-related words may be confusing to the judges. In conclusion, the above issues could not be solved by the proposed method in this study.

Keywords: ESG, Related Words, Text Mining

概要：

近年、企業の環境、社会、ガバナンス要素を考慮した ESG 投資が拡大している。そのため、企業の ESG についての取り組み（ESG 情報）を知ることが重要となる。これを知る手段として統合報告書があるが、記述量の多さから、ESG 情報を獲得するには、多くの時間と労力を要してしまう。そこで、本研究では、統合報告書からテキストマイニングを用いて ESG 関連語を取得し、文章上でマークすることで上記の問題を解決できるか試みる。ここでの ESG 関連語は、Word2Vec を用いて抽出した「環境」「社会」「ガバナンス」に類似する語とする。実験の結果、ESG 関連語へのマークは、有効である場合とそうでない場合があることがわかった。有効性を示さない場合の要因としては、ESG 関連語へのマークが判断を悩ませていることが考えられる。結論として、本研究の提案手法をもって上記の課題を解決することはできなかった。

キーワード： ESG, 関連語, テキストマイニング

目次

第 1 章	序論	1
1.1	背景	1
1.2	課題	2
1.3	目的	2
1.4	複雑系コースにおける卒業研究の位置付け	2
1.5	論文構成	2
第 2 章	関連手法	3
2.1	形態素解析	3
2.2	Word2Vec	3
2.2.1	CBOW	3
2.2.2	Skip-gram	4
2.3	コサイン類似度	5
第 3 章	関連研究	6
3.1	ESG 分野で Word2Vec を扱った研究	6
3.2	文字強調による読解支援に関する研究	7
第 4 章	提案手法	8
4.1	提案手法概要	8
4.1.1	データの収集	9
4.1.2	データ形式の変換とテキストの整形	9
4.1.3	形態素解析	9
4.1.4	Word2Vec モデルを用いた ESG 関連語の取得	10
4.1.5	ESG 関連語にマーク	18
第 5 章	実験と評価および考察	19
5.1	予備実験	19
5.1.1	課題	19

5.1.2	実験目的	19
5.1.3	参加者	19
5.1.4	日時	19
5.1.5	場所	20
5.1.6	実験条件	20
5.1.7	実験材料	20
5.1.8	実験手順	22
5.1.9	実験結果	23
5.1.10	実験結果のまとめ	24
5.2	本実験	26
5.2.1	課題	26
5.2.2	実験目的	26
5.2.3	参加者	26
5.2.4	日時	26
5.2.5	場所	26
5.2.6	実験条件	26
5.2.7	実験手順	26
5.2.8	実験結果	27
5.2.9	実験結果のまとめ	30
第 6 章	結論	33
6.1	まとめ	33
6.2	今後の展望	34
	参考文献	36

第 1 章

序論

本章では, 研究の背景と課題, 目的, 所属コースにおける卒業研究の位置付けについて述べる.

1.1 背景

近年, 従来の財務情報だけでなく, 企業の環境 (E: Environment), 社会 (S: Social), ガバナンス (G: Governance) の非財務情報への対応も考慮した ESG 投資が近年急速に拡大しており [1], 企業の長期的な成長と経営には, ESG の 3 つの観点が必要だという考え方が世界中で広まっている. 「ESG」という言葉がここまで浸透するようになったのは, 2006 年当時国連事務総長であった, コフィ・アナンが機関投資家に対して, PRI (Principles for Responsible Investment: 国連責任投資原則 以下, PRI) の 6 つの責任投資原則を実践するよう呼びかけたことがきっかけだとされている. 日本においては, この拡大が広まった要因として, 2015 年 9 月の年金積立金管理運用独立行政法人 (GPIF) による国連の責任投資原則 (Principles for Responsible Investment (PRI)) への署名が ESG 投資拡大の契機になったと言われている. 主に企業関係者や投資家に対して ESG は注目されている. その企業関係者や投資家たちにとっては, ESG 情報を獲得することは重要である. ESG 情報を取得する手段はいくつかあるが, 1 つの手段として, 各企業がそれぞれ発行する統合報告書がある. 統合報告書は, 株主や投資家, 取引先, 金融機関, 従業員などに向けて経営実態や持続的な成長への取り組み, 中長期的な価値創造などを発信する媒体であり, 企業の財務情報に加え, 非財務情報をまとめた報告書である. KPMG ジャパンの調査 [2] によると, 統合報告書の発行企業数は年々増加しており, 今後, ますます注目され有用される機会が増えることが予想される. また, 非財務情報への重要度が増す中で, 海外では 2000 年以降, テキストマイニングや機械学習などを用いた企業の情報開示の研究が蓄積され, 日本も同様に 2010 年前後から財務報告に関するテキストの分析が進められている.

1.2 課題

そして、財務情報に加え、非財務情報についても記述されている統合報告書の中には、100ページを超えるもあり、KPMG ジャパンの調査によると、日本企業 712 社の 2021 年度統合報告書の平均ページ数は約 72 ページだという。この中から、読者が ESG 情報を獲得するには、多くの時間と労力を要するという問題を抱えている。特に ESG に関しての専門知識を有さない人にとっては、着目すべき単語がわからず、ESG 情報の獲得は困難である。

1.3 目的

そのため、文章上で ESG 関連語にマーカーを付与し強調することで、統合報告書が抱える「ESG 情報を獲得するには、多くの時間と労力を要する」という問題を解決することが可能になると考えた。そこで、まず ESG 関連語を抽出し、文章上でマークする。そして被験者実験を行い、提案手法が実際に上記の課題に対して有効かどうか確かめることを目的とする。目的の対象は ESG に関しての専門知識を有さない人とし、被験者実験では ESG についての知識を有していない大学生を対象とした。

1.4 複雑系コースにおける卒業研究の位置付け

本学の複雑系コースのカリキュラム・ポリシーは複雑系の観点から、法則性の解明、数理モデル化、実世界への適用に関する具体的な課題に取り組み、その結果の評価を通じて、新しい方法論や、学問領域を切り拓く能力を育むことである。本研究では、企業の統合報告書からテキストマイニングによって、ESG 関連語を抽出し、これを利用することで、「統合報告書から ESG 情報を獲得するには、多くの時間と労力を要する」という課題を解決することを試みている。本研究で以上の課題を解決できた場合、本研究がとった手法やアプローチが実社会でも適用でき、役立てることが期待できる。これを複雑系コースのカリキュラム・ポリシーに基づいた、本研究の位置づけとする。

1.5 論文構成

本節では、本論文の論文構成について述べる。第 1 章では、本研究の背景と課題、目的を述べた。第 2 章では本研究と関連する手法について述べる。第 3 章では本研究と関連する研究を述べる。第 4 章では本研究で提案する手法について述べる。第 5 章では本研究の実験と評価および考察を述べる。第 6 章では本研究のまとめと今後の展望について述べる。

第 2 章

関連手法

本章では、本研究の根幹となるテキスト処理に関する技術や手法について説明する。

2.1 形態素解析

文章を形態素区切りで分割し、文章を構成している単語の情報を得る処理が形態素解析である。本研究での形態素解析は、形態素解析ソフト MeCab[3] を用いて、日本語テキストの分かち書き、品詞の分類と名詞の抽出を行った。

2.2 Word2Vec

Word2Vec[4] は、Tomas Mikolov が米グーグルの研究者だったときに開発した自然言語処理の手法の一つであり、単語を数値ベクトルとして表現する手法として多くの応用で利用されている。Word2Vec には、意味の近い単語から生成されたベクトルは類似したベクトルとなる特徴があるという前提をもつ。手法の特徴として単語を高次元空間のベクトルとして表現する定量化手法であり、単語の類似度やベクトルの加算、減算することで別な単語を求めることができる。また、Word2Vec には、周辺単語群から該当単語を推定する Continuous Bag-of-Words Model (以下、CBOW とする) と、該当単語から周辺単語群を推定する Continuous Skip-gram Model (以下、Skip-gram とする) の 2 種類のモデルが存在する。本研究では、Skip-gram モデルを作成し使用している。

2.2.1 CBOW

本節では CBOW について解説する。このモデルは図 2.1 を見て分かるように、入力層、中間層、出力層の 3 層からなるが、入出力が Skip-gram モデルの逆となっている。出力は中心の単語 $w(t)$ であり、入力をその前後の単語 $w(t - c), \dots, w(t - 1), w(t + 1), \dots, w(t + c)$ とするニューラルネットワークである。つまり、Skip-gram モデルとは反対に、周辺の単語が

ら中心にある単語を予測する問題をニューラルネットワークに学習させるモデルである。

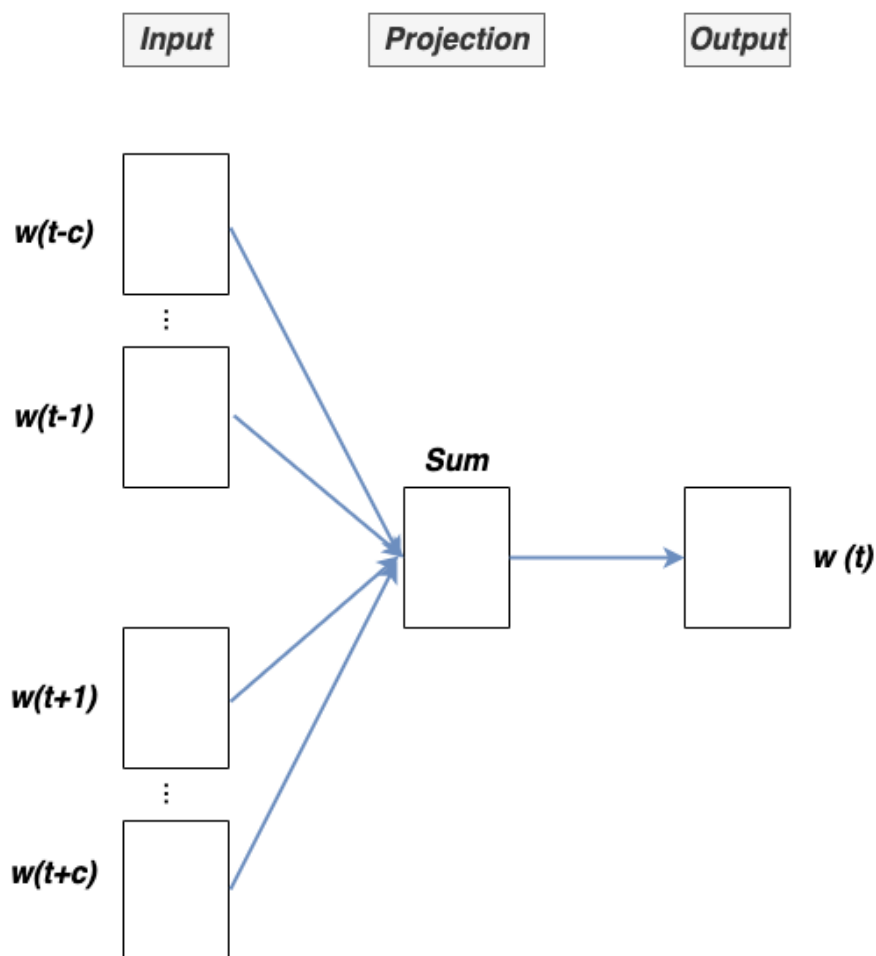


図 2.1 CBOW

2.2.2 Skip-gram

本節では Skip-gram について解説する. Skip-gram は CBOW を逆転させたモデルであり, 図 2.2 から分かるように, Skip-Gram は入力層, 中間層, 出力層の 3 層からなり文章中のある単語 $w(t)$ を入力, その前後の単語 $w(t - c), \dots, w(t - 1), w(t + 1), \dots, w(t + c)$ を出力とするニューラルネットワークである. ある単語が入力に与えられたとき, ボキャブラリー中の単語がその「周辺」である確率を学習している. Skip-gram の最終目的はモデルの構築ではなく, 中間層の重みを獲得することにある. 本研究は, Window = 10 の範囲内を「周辺」と定義しモデルを構築している.

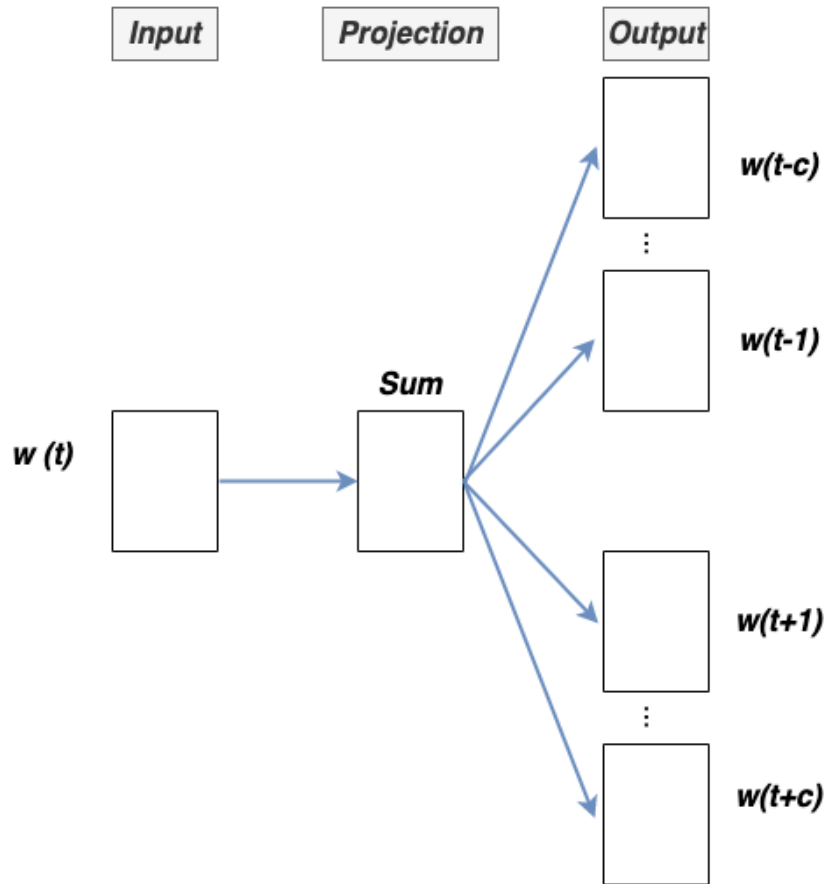


図 2.2 Skip-gram

2.3 コサイン類似度

ベクトル空間モデルにおいて、文書ベクトル同士を比較して文書の類似度を計算する手法がコサイン類似度である。コサイン類似度の計算では2つの文書ベクトルの内積を計算してベクトル同士の成す角度の近さを表現するので、コサイン類似度が1に近いほど2つの文書が類似していることになる。逆に、コサイン類似度が0に近いほど、類似度は低く関係性はみられない。本研究では、単語ベクトル同士を対象にコサイン類似度の計算をし、「環境」、「社会」、「ガバナンス」に類似している上位100単語を選出している。第4章の表 4.4, 4.5, 4.6 で示している。以下、コサイン類似度算出の数式を記す。

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\vec{q}}{|\vec{q}|} \cdot \frac{\vec{d}}{|\vec{d}|} = \frac{\sum_{i=1}^{|V|} q_i d_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \cdot \sqrt{\sum_{i=1}^{|V|} d_i^2}}$$

第 3 章

関連研究

本研究では Word2Vec を用いた ESG 関連語の取得と, 取得した ESG 関連語にマーカーを付与し強調することで, 統合報告書が抱える「ESG 情報を獲得するには, 多くの時間と労力を要する」という問題を解決することを試みる. また, 被験者実験を行い, 以上の問題に対して, 本研究の手法は有用であるかを確認する. そこで ESG 分野で Word2Vec を扱った研究と, 文字強調に関する研究について説明する.

3.1 ESG 分野で Word2Vec を扱った研究

まず, 河村らの研究に, 機械学習を用いて統合報告書から ESG 関連ページを推定する研究がある [5]. この研究では, 統合報告書より, E, S, G それぞれに対する特徴語を抽出し, 特徴語を用いたルールベースによりラベルを付与し, 学習データを自動生成. そして, 学習したモデルを用いて, 統合報告書から ESG 関連ページを推定している. 特徴語の抽出には, 統合報告書から単語 “ESG” を含むページを取得し, 取得したページの文集合から Word2Vec により 100 次元の分散表現を学習. 統合報告書に出現する名詞 n の「環境」「社会」「ガバナンス」に対する正規化自己相互情報量, 名詞 n が出現する確率に基づくエントロピーが共に上位である単語を特徴語候補とし, 特徴語候補を k-means 法でクラスタリングすることで ESG 特徴語としている. 本研究でも同様に, 統合報告書を対象データに Word2Vec も用いて ESG に関連する単語の抽出を行っている. しかし, Word2Vec の学習方法が本研究と異なっており, 本研究では統合報告書内に出現するすべてのテキストを対象に Word2Vec に分散表現を学習させている. 次に, 秋山らの研究に, ESG トピックワードを抽出し, 各企業の ESG スコアを評価, ESG ファクターの分析を行った研究がある [6]. この研究では, ESG というキーワードの曖昧さにより, 各企業の開示文書の内容から ESG への取り組みを評価するのは難しいとし, Word2Vec を用いて各単語をベクトル化し, 「環境」「社会」「ガバナンス」に類似する ESG トピックワードを抽出している. 加えて, 各企業に関連するニューステキスト内にどの程度 ESG トピックワードに類似する単語を含むかを算出し, 客観的に各企業の ESG スコア

を評価している。本研究とは、Word2Vec を用いて各単語をベクトル化し、「環境」「社会」「ガバナンス」に類似する単語の抽出を行っている点で類似している。また河村らの研究同様、本研究とは Word2Vec の学習方法が異なっている。まず対象のデータがニューステキストであり、ネガティブ単語を除去した、ポジティブなニューステキストのみを Word2Vec の学習材料としている。

3.2 文字強調による読解支援に関する研究

テキスト情報の内容把握の認知促進を手助けするために、太字・下線や斜体文字などで文字強調を行い提示する方法がある。まず本多らの研究 [7] では、文字強調が文章内容の認知に与える影響を調査している。まず、太字や下線などの文字の強調は、読み手への情報伝達にどのような効果を持つのかテキスト情報の文章を読解させ設問に解答させるとともに、事後アンケートとして強調文字の感想と、読解時にどのように利用したのかなどを調べている。筆者らは、被験者をランダムに3つの群「A 群：適切な強調文字」、「B 群：的外れな強調文字」、「C 群：強調文字なし」に分けて、実験を行なった結果、設問への正答率は A 群 > B 群 > C 群に高くなり、事後のアンケートでは強調文字に対して「読む速度を落として読んだ」などの対応をしていたとの回答を得ている。しかし、読解時間を計測したところ、実際は、強調文字ありの方が有意に短いという結果となった。本研究では、本多らの研究と異なり、色付きのマーカーを付与させることで文字を強調させる。一方で、本研究は本多らの文章を読解させ設問に解答させる方法、事後のアンケートで強調文字に対する感想を聞く一連の流れを参考に実験を行う。次に中尾らの研究 [8] では、品詞情報、係り受け構文情報、意味論的、語用論的情報である文型情報を総合して文中の重要語句をマークすることで強調表示することで、主題に関連する重要箇所が、内容上の意図と共に正確に特定できるようになったことを述べている。本研究も中尾らと同様にマーカーによる文字強調している。しかし、中尾らは文構造に着目した重要語のマークをしているのに対し、本研究は「環境」「社会」「ガバナンス」に類似する語にマークしている。

第4章

提案手法

本章では, 本研究で提案する手法について説明を行う.

4.1 提案手法概要

本研究では, 統合報告書からの ESG 情報獲得には多くの時間と労力を要するという問題に対し, ESG 関連語を文章上でマークする手法を提案し, 問題解決を試みる. 主に, 以下の5つの手順を踏む. また, 提案手法の中で用いる, ツール・ライブラリとそのバージョンについて表 4.1 にて記載する.

1. 統合報告書データ形式を PDF から TXT へ変換し, 以降のテキスト処理が可能な形にする
2. テキストデータに対して, 形態素解析を行いテキストを整形する
3. 整形したテキストデータから Word2Vec モデルを作成する
4. Word2Vec モデルを活用して, ESG 関連語を取得する
5. 文章上に出現する ESG 関連語にマークする

表 4.1 利用した各種ツール・ライブラリとそのバージョン

利用したツール・ライブラリ	バージョン
Python	3.9.13
scikit-learn	1.0.2
Pandas	1.4.4
Numpy	1.21.5
mecab-python3	1.0.5
pdfminer.six	20221105
gensim	4.1.2

4.1.1 データの収集

まず、本研究でデータ分析の対象とする、統合報告書テキストデータに関する既存データセットの調査を行った。しかし、データセットとしてまとめられているものはなかったため、自ら統合報告書の収集を行った。

- 対象データ：統合報告書
- データ形式：PDF ファイル
- 入手先：各企業のホームページ
- 入手数：150 社分
- 入手範囲：CSR 企業白書 2022 年度版 [9] に記載されている ESG ランキングより

4.1.2 データ形式の変換とテキストの整形

まずデータを扱えるよう以下の工程でデータの処理を行った。

1. データ形式を変換
2. ルールベースにより、テキストを整形

1. では、どの統合報告書でも入手時点では PDF 形式であるため、TXT ファイルへの変換を行う。TXT ファイルへの変換には、Python のライブラリの 1 つである PDFMiner を用いた。他にも、PyPDF2, ReportLab といった PDF ファイルから TXT ファイルに変換に用いられるライブラリがある。しかし、日本語で記述されている統合報告書に対応できるライブラリではないため、日本語対応が可能である PDFMiner を用いた。ここで、2. の作業を行う。本来、統合報告書には画像やグラフが埋め込まれているため、変換後には、テキスト間に大きな行間が開いたり、改行が過剰に入るなどしている。そこで、空白、改行、句読点、記号、数字などは削除し、150 ある TXT ファイルを統合し、1 つの TXT ファイルとした。中身のテキストは 1 つの文章となるようにした。

4.1.3 形態素解析

以下の手順で形態素解析を行った。

1. テキストの分かち書き（1 回目）
2. 複合名詞の扱いの検討
3. テキストの分かち書き（2 回目）
4. 名詞の抽出

手順 1. では、分かち書きを行った。分かち書きとは、英語のように単語間を切り離す書き方である。しかし、日本語で記述されたテキストは分かち書きをされていないため、自然言語処理ができないという問題点がある。そこで、MeCab というツールの中の「-Owakati」オプションを用いて分かち書きを行った。このとき「地球温暖化」として抽出したい単語が、「地球」と「温暖化」のように分かれて分かち書きされてしまう問題が起きた。MeCab の標準のシステム辞書では正しく分割できない複合名詞や固有表現などがあることがわかった。そこで、手順 2. ではこのような複合名詞を 1 つの単語として抽出するために「mecab-ipadic-NEologd」という辞書を用いた。mecab-ipadic-NEologd は、多数の Web 上の言語資源から得た新語を追加することでカスタマイズした MeCab 用のシステム辞書であり、週 2 回以上更新され、新語・固有表現に強く、Web 上の言語資源を活用しているため語彙数も多い。加えて、オープンソース・ソフトウェアであるという特徴を持っている。この辞書を用いることで、地球温暖化も 1 つの単語として認識され、複合名詞が分割されるという問題は解決した。手順 3 では再び分かち書きをし、各単語の品詞分類を行なった。手順 4. では分類した品詞の中から名詞のみを抽出した。本研究では、ESG 関連語の抽出に重要な意味を持つとして名詞を抽出し、その他の品詞については今回は対象外とした。

4.1.4 Word2Vec モデルを用いた ESG 関連語の取得

以下の手順で Word2Vec を用いた ESG 関連語の取得を行う

1. 既存の Word2Vec モデルを入手
2. 1. を用いた ESG 関連語の取得
3. 2 と統合報告書出現単語の比較
4. Word2Vec モデルの作成
5. ESG 関連語の取得

Word2Vec モデルを適用し、文章中の単語をベクトル化することで関連する単語の抽出や単語同士の類似度を調べたり単語の線形計算ができるようになる。本研究ではコサイン類似度を計算し、「環境」、「社会」、「ガバナンス」に対してコサイン類似度の高い上位 100 単語を取得し、これらを ESG 関連語とする。このように、「環境」、「社会」、「ガバナンス」に類似する単語を取得し、文章上でこれらの単語がまとまっている箇所は ESG についての記述がなされている箇所であると見なすことができると考えている。手順 1. では、まず Word2Vec モデルを用いた場合に、どのような ESG 関連語が入手できるのかを確認するために、既存の Word2Vec モデル [10] を用い、ESG 関連語を取得することとした。このモデルは、日本語版 wikipedia 本文から作成されている。以下、「環境」、「社会」、「ガバナンス」に対してコサイン類似度の高い上位 100 単語を表 4.2, 4.3, 4.4 に記す。

表 4.2 「環境」に対してコサイン類似度の高い上位 100 単語

環境	多様性	環境政策学
地球環境	地球環境問題	保全
生態系	保全	文化
社会環境	安全	土地利用
自然環境	土壌	秩序
生物多様性	風土	汚染
生活環境	エコロジー	医療
景観	気候	生態系サービス
自然	リスクマネジメント	経済
資源	防災	植生
健康	イノベーション	開発途上国
環境問題	生物圏	食の安全
住環境	森林	社会福祉
生態系	大気汚染	化学物質
社会	生物	産業構造
生物多様性	科学技術	ケア
環境政策	メンタルヘルス	安全性
持続可能な開発	エコロジー	土壌
気候変動	生命	テクノロジー
公衆衛生	産業	生活空間
海洋	環境保護	土地利用
人間活動	持続可能性	遺伝的多様性
気候変動	生物群集	保健
環境保全	消費者	衛生
環境保全	経済発展	微生物
福祉	人間社会	状況
水質	職場環境	創造産業
地域社会	国際協力	適応
持続可能性	地球温暖化	公共財
生活	環境保護	効率性
環境教育	共生	社会的
人間関係	野生動物	自然保護
環境問題	安全保障	バイオテクノロジー

表 4.3 「社会」に対してコサイン類似度の高い上位 100 単語

社会	市場経済	道徳
社会的	現代社会	市民社会
道徳	個人主義	人間社会
人間関係	意識	認識論
経済	自己	一般社会
文化	価値観	習慣
価値観	人間	実践
社会構造	国民	概念
精神	知識	階級闘争
民主主義	行動	社会関係
政治	社会制度	良心
倫理	家族	官僚制
共同体	文明	社会生活
社会集団	観念	理性
生活	国家	貧困
秩序	思考	現代社会
イデオロギー	自我	社会集団
組織	学問	科学
規範	経験	存在論
集団	平等主義	社会階級
市民社会	正義	国際関係
思想	感情	社会一般
地域社会	ジェンダー	民族
個人	観念	自然法
資本主義	日常生活	知性
自由	無意識	価値
身体	階層	方法論
認識	宗教	社会秩序
常識	国際社会	健康
理性	権力	生産様式
大衆	心理	知覚
制度	社会階層	地縁
人格	環境	儀礼

表 4.4 「ガバナンス」に対してコサイン類似度の高い上位 100 単語

逆機能	サンテニャン＝グランリユー	オリバー・ウィリアムソン
全体主義体制	三菱 UFJ リサーチコンサルティング	社会構成主義
私的年金	イスラーム共和制	バランス・オブ・パワー
要素従属性	応用倫理学	学校法人産業能率大学
原因において自由な行為	社会的公正	パンデクテン法学
ウォーラステイン	教育行政学	bureaucracy
金融ビッグバン	公共選択論	社会システム論
プルーデンス政策	自由貨幣	間接侵略
村社会	アルファ・グループ	消費者保護
ミハウ・カレツキ	ミネゼンガー	IFRSs
日本興亜損保	キャリア段位	エリ・ヘクシャー
サステナビリティ	本木北町	皇国史観
代数的整数	シカゴ学派_(政治学)	#矢野主税
社員権	社会保障法	論理主義
摩擦的失業	社会的企業	華夷思想
地域社会学	マクロ経済政策	市場構造
家産制	形式社会学	マス・コラボレーション
生産性分析	教育方法学	ゾーンクウェー
プルデンシャル生命保険	全的墮落説	非合理主義
公的扶助	統計的仮説検定	単独介入
非攻	レギュレーション理論	連結貸借対照表
非公務員化	平等権	使者_(法律用語)
文化多元主義	科学リテラシー	社団国家
租税法	新カント学派	NEO_(証券市場)
第三の位置	アメリカ合衆国の医療	知識産業
仁愛	地方開発事業団	三徳_(儒学)
未開社会	A/B テスト	安全保障貿易管理
オークシー	ファストファッション	地域力
サピア＝ウォーフの仮説	家産国家	公的領域
ヒューマンホールディングス	機能的クレーム	度線分析
国際金融のトリレンマ	大阪健康福祉短期大学	光生館
腐食疲労	教育方法	山口県立萩商業高等学校
構造言語学	メディア効果論	地方公務員法

手順 3. では, 既存の Word2Vec モデルで用いた ESG 関連語の統合報告書内での出現頻度を調査した. 結果を以下に示す.

- E: 環境に関する語 43/100 語
- S: 社会に関する語 26/100 語
- G: ガバナンスに関する語 3/100 語

既存モデルから抽出された ESG 関連語は統合報告書内では, 使用されていない語が多く存在した. 以上の結果より, Word2Vec モデルは作成元となるドキュメントに依存するため, 統合報告書よりモデルを作成する必要があることがわかった. 手順 4.gensim という自然言語処理のライブラリを使用して word2vec のモデルを作成する. 入力には, 4.1.2 で形態素解析を行なった統合報告書 150 社分のテキストを用いる. 今回モデルを学習させるにあたって以下のような設定を施した.

```
model = word2vec.Word2Vec(sentences,  
                           sg=1,  
                           vector_size=100,  
                           min_count=1,  
                           window=10,  
                           hs=1,  
                           negative=0)
```

sg=1 は, skip-gram を使用したことを指す. vector size=100 は, 生成されるベクトルの次元数が 100 であることを指す. min count=1 は, 登場回数が 1 回未満の語彙は無視することを指す. window=10 は, current word の前 10 個, 後ろ 10 個が学習に用いられることを指す. hs =1 は, ソフトマックスを使用したことを指す. negative=0 は, 0 より大きければ関連しない語彙を使ってネガティブサンプリングをすることを指す. このようにして, 統合報告書のテキストより word2Vec モデルを作成し, 前述したように, 「環境」, 「社会」, 「ガバナンス」に対してコサイン類似度の高い単語上位 100 語を取得した. また, それぞれ 100 語に含まれていたひらがな 1 字であったり文章中では大きな意味をなさないとしたものは, ESG 関連語から外し対象外とした. 以下 ESG 関連語の取得結果を表 4.5, 4.6, 4.7 で示す.

表 4.5 「環境」に対してコサイン類似度の高い上位 100 単語

価値観	特徴	風水害
方	課題	お客さま
衣料	安心	医療従事者
浮き彫り	質	注力
衛生	身近	一変
変化	環境改善	ライフスタイル
環境負荷	急速	強靱
境	産業構造	モビリティ
ニーズ	くらし	要請
変容	プリンク	トータル
グループビジョン	出し	結束
入手	特定	市場構造
環境問題	Life	デジタル化
街	豊か	提供
Material	柔軟	方向性
HealthCare	アクセラレーション	ストック
安全	多様化	甚大
良質	激化	誠実
酷暑	高齢化	対面
経営環境	喪失	生
不連続	人びと	困難
タイムリー	能動	世界の自動車
融合	ニューノーマル	レジリエンス
下市場	生まれ	深刻
共生	世界同時不況	エントリー
不測	Home	MAS
察知	クリーン	時代
都市	今後	前
快適	長寿	発想
健康	思	部品
便利	健	共通
世の中	環境技術	関心
レジリエント	Energy	ミーティング

表 4.6 「社会」に対してコサイン類似度の高い上位 100 単語

貢献	モビリティ	リティ
いのち	共通認識	炭素
安心	汚染	マテリアリティ
クリーン	先手	技術力
人びと	サステナビリティ	さまざま
グループビジョン	理念	エンス
豊か	突拍子	鍵
循環型社会	環境負荷	健
サイクル	健康	昨日
快適	大地	リード
環境改善	特長	試み
実現	浮き彫り	前
都市	年造	経営資源
医療従事者	未来へ	分野
患	持続可能性	シフト
長寿	レジリエント	ストック
共生	国際情勢	人類
Solution	深刻	課題
酷暑	リサイ	ニューノーマル
対立	省資源	使命
セット	喫緊	きた
薄肉	融合	際
日々	サービ	サーキュラーエコノミー
重点的	付加価値	Material
解決	提供	急ピッチ
近く	プリン	衣料
環境技術	CareforPeople	循環
便利	ダイナミック	康
くらし	防止	Outcome
良質	デカッ	グリーンケミ
身近	安全	Tomorrow
世界	可能	異
サークル	構築	

表 4.7 「ガバナンス」に対してコサイン類似度の高い上位 100 単語

ホルダー	改修工事	会計
組み	洗い出し	絶対
高分子化学	気候変動	室
経験	構成員	コーポ
担保	持続可能性	緩和
要素	リティ	物的
判断基準	イノベティブ	検討
双方	風水害	洗
経済合理性	実質	投入
毎年	説明	総合力
バックグラウンド	社会貢献	ティ
運営	広範	配慮
立案	反映	循
グローバル企業	ラル	総務部
とおり	一元的	革新的
浸透	事業部門	分析
執行責任	使用量	した
開発プロセス	ボン	サークル
ステークホルダー	発想	視点
分布	サス	会
業務提携	結果	審議
環	答申	出
会議	考	理念
意義	甚大	レートガバナンス
ポテンシャル	補佐	ゼロ
留意	委員	事項
外	分科	重要
適用範囲	ヒアリング	予防措置
骨子	監査役会	深刻
サステナビ	対象	追加
実効	valueofthecircularEconomy	サイバーセキュリク
地球環境	コンセプト	多面
サステナビリティ	最適	

4.1.5 ESG 関連語にマーク

上記の過程で, 取得した ESG 関連語, E (環境) 要素の関連語には緑色のマーク, S (社会) 要素の関連語にはオレンジ色のマーク, G (ガバナンス) 要素の関連語には紫色のマークを付与した. 次に, この提案手法をもって, 1.2 で前述した課題を解決できるか被験者実験を行い確認する. 被験者実験の詳細については, 第 5 章で述べる.

第 5 章

実験と評価および考察

ここでは被験者実験について述べる。まず予備実験を行い、続いて本実験を行った。

5.1 予備実験

実験手順や実験時間、対象者にかかる負荷の確認、本実験のリハーサルを兼ねて予備実験を行った。

5.1.1 課題

近年の ESG 投資の拡大と共に、企業関係者や投資家たちにとって、ESG 情報を獲得することは重要となっている。しかし、記述量の多さより、統合報告書から ESG 情報を獲得するには、多くの時間と労力を要してしまう。

5.1.2 実験目的

統合報告書の文章上で、取得した ESG 関連語にマークすることで ESG 情報の記載箇所の把握が容易となり、以上の点が解決するかどうか確認する。

5.1.3 参加者

参加者は、大学生 3 名である。

5.1.4 日時

予備実験は、2023 年 1 月 12 日 (金) から 1 月 14 日にかけて行った。

5.1.5 場所

オンライン上で行った. 実験説明やタスクの記入には Google Forms を用いた.

5.1.6 実験条件

参加者は ESG についての専門知識を有さないこと. 統合報告書を閲覧する際に PC やタブレットなどの大画面で文書を閲覧できる環境や, 印刷して目を通せる環境を有すること.

5.1.7 実験材料

まず 3 社分の統合報告書を用意する. ここではそれぞれ文書 A, 文書 B, 文書 C とする. 次に, それぞれの文書に対して, 何も付与されていないそのままの文書 (1 とする). 取得した ESG 関連語にマーカーを付与した文書 (2 とする). 不作為にマーカーを付与した文書 (3 とする) を作成する. 以上の作業により A-1, A-2, A-3, B-1, B-2, B-3, C-1, C-2, C-3 の計 9 つの文書を用意できた. 以下, 文書 A としているオムロン株式会社の統合報告書 [11] を例に, マーカーなしの文書 A-1, ESG 関連語にマーカーを付与した文書 A-2, 不作為にマーカーを付与した文書 A-3 のイメージを図 5.1, 5.2, 5.3 で掲示する.



図 5.1 マーカーなしの統合報告書 (A-1)

5.1.8 実験手順

参加者1人にたいして、3つの文書を読み3つのタスクに解答してもらおう。まず、実験材料として3種類の統合報告書から抜粋した文書を用意する。手順1. 実験の概要について説明を行う。手順2. ESGに関する基本事項について説明を行う。手順3. 統合報告書から ESGに関する記述がされている段落を問うタスクを与える。この際には、タスクを与えられてから、解答を終えるまでの時間の計測を行う。手順4. その他2種類の統合報告書で手順3. と同じ作業を行う。また、被験者が解答する文書の順序が結果に影響する可能性を考慮し、実験では以下のように、解答する文書の順序を指定した。この際、文書の内容を覚えてしまうことを防ぐため、被験者1人には、違う内容物の3つの文書 A, B, C を使用する。

1人目には文書 A-1, B-3, C-2 という順番。2人目には C-1, A-2, B-3 という順番。3人目には B-3, A-1, C-2 という順番で一つの文書に与えられる1つタスクをそれぞれ解答してもらった。以下、与えたタスクを記述する。

文書 A のタスク

ESGのうち、E（環境）について具体的な目標が記されているのはどの段落ですか？

文書 B のタスク

気候変動に対しての取り組みが具体的に記述されているのは、どの段落ですか？

文書 C のタスク

二酸化炭素に関する記述が最も多いのはどの段落ですか？

また、タスクに加え、本実験の参考になるように、ESG 関連語にマーカーを付与した文書を読んだ感想、その他実験に関する感想を自由に記述してもらった。

5.1.9 実験結果

評価指標（タスクを終えるのにかかった時間、タスクの正誤）

1人目：大学生 A さん

使用文書	解答時間	タスクへの解答	正誤判定
A - 2	5分40秒	13	正解
B - 1	2分38秒	19	正解
C - 3	1分59秒	9	正解

感想
<ul style="list-style-type: none"> ・マーカーがない文書に比べて、マーカーがある文書の方が読みやすい印象 ・マーカーが引かれている内容はそれほど印象に残らず、ランダムな文書と ESG に関する部分にマーカーが引かれた文書の間にはそれほど違いは感じなかった

2人目：大学生 B さん

使用文書	解答時間	タスクへの解答	正誤判定
C - 1	4分56秒	9	正解
B - 2	7分43秒	19	正解
A - 3	6分56秒	13	正解

感想
<ul style="list-style-type: none"> ・どの段落に何が書いてあるかはパッと全体を見た時に理解できると感じたが、これも ESG に関する単語なのかと思いながら読んだ。 ・マークがない方が集中して読めたかもしれない。 ・文章 A の場合、私自身が ESG に関する理解度が低かったために何が E に該当するのかを探り探りになったのかと感じる。 ・また文章の量と文章の難易度にばらつきがあったように感じた。 ・文章 C は問題が分かりやすく、かつ短かったので1番読みやすく感じた。 ・読みやすさや要した時間がマーカーの有無に左右されているのかはなんとも言えないが、マーカーの付与の基準が前もって分かっているのもっと読みやすかったかもしれない。

3人目：大学生 C さん

使用文書	解答時間	タスクへの解答	正誤判定
B - 3	3分 11秒	19	正解
A - 1	0分 32秒	13	正解
C - 2	0分 37秒	9	正解

感想
<ul style="list-style-type: none"> ・ 漢字 1 文字だけ, または単語の途中でマーカーされており, 見辛い ・ 線が引かれているもののうち, 本当は ESG に関連してないものや別の色でマーカーすべき単語がありそう

5.1.10 実験結果のまとめ

文書	解答時間	正誤
A-1	0分 32秒	正解
A-2	5分 40秒	正解
A-3	6分 56秒	正解
B-1	2分 38秒	正解
B-2	7分 43秒	正解
B-3	3分 11秒	正解
C-1	4分 56秒	正解
C-2	0分 37秒	正解
C-3	1分 59秒	正解

タスクへの解答は 3 人全員が全ての文書で正解であった。解答時間に関しては, A, B はマーカーなしの文書への解答時間が短く, C は ESG 関連語にマークした文書への解答時間が一番短いという結果だった。しかし, 1 人目の平均解答時間は 3 分 23 秒, 2 人目の平均解答時間は 6 分 30 秒, 3 人目の平均解答時間は 1 分 26 秒と想定以上に解答時間に個人差が大きかったためマーカーの有無による解答時間の変化は正しく計測できなかった。以上の結果より ESG 関連語を用いて「統合報告書から ESG 情報を獲得するには, 多くの時間と労力を要する」という課題を解決するという目的は達成できなかった。実験がうまくいかなかった要因として, 被験者の少なさ, 実験説明の不十分さが考えられる。実験期日を踏まえ, 即座に修正可能である実験説明を改善し, 本実験に臨むことにした。以下 3 点に関する説明を修正し本

実験を行う。

1. 文書の種類に関して
2. ESG 関連語に関して
3. 実験目的に関して

主に、文書の種類、ESG 関連語、実験目的についての事前説明が不適切だったり不十分であったりした。まず、1. 予備実験の事前説明では、マーカーを付与した文書が「ESG 関連語にマーカーを付与した文書」、「不作為にマーカーを付与した文書」の 2 種類があることを事前に示していた。しかし、本研究の実験は、マークによる先入観を図る実験なので「関連語にマークした二つの文書」とする必要があった。マークのあるなし以外にランダムにマークされたものかどうかを被験者が当てるタスクが混ざってしまっていた。次に、2. ESG 関連語についての説明である。実験後の感想より、ESG 関連語を一般的なキーワードのことだと誤解されていたように、本研究で ESG 関連語としているものには、一見 ESG とは関係のない単語まで取り上げているため、被験者を混乱させてしまう要因を作ってしまった。そして、3. 実験目的に関しての事前説明も改善すべき箇所がある。二人目の被験者は、1 人目の被験者よりタスクを終えるのに、どの文書に対しても時間がかかっていた。後の質問で、実験中は内容をしっかり把握しようと、時間をかけ過ぎてしまったと感じていることがわかった。本研究の目的は取得した ESG 関連語を用いて ESG 情報が記載されている場所の把握を容易にすることにある。そのため内容把握は、本研究の目的とは少しずれが生じている。この問題をなくすために、事前に実験目的に関して詳細な説明を行う必要があった。

5.2 本実験

予備実験の内容を踏まえ、本実験を行う。

5.2.1 課題

近年の ESG 投資の拡大と共に、企業関係者や投資家たちにとって、ESG 情報を獲得することは重要となっている。しかし、記述量の多さより、統合報告書から ESG 情報を獲得するには、多くの時間と労力を要してしまう。

5.2.2 実験目的

統合報告書の文章上で、取得した ESG 関連語にマークすることで ESG 情報の記載箇所の把握が容易となり、以上の点が解決するかどうか確認する。

5.2.3 参加者

参加者は、大学生 7 名である。

5.2.4 日時

本実験は、2023 年 1 月 16 日 (月) から 1 月 25 日 (水) にかけて行った。

5.2.5 場所

オンライン上で行った。実験説明やタスクの記入には Google Forms を用いた。

5.2.6 実験条件

参加者は ESG についての専門知識を有さないこと。統合報告書を閲覧する際に PC やタブレットなどの大画面で文書を閲覧できる環境や、印刷して目を通せる環境を有すること。

5.2.7 実験手順

参加者 1 人にたいして、3 つの文書を読み 3 つのタスクに解答してもらおう。まず、実験材料として 3 種類の統合報告書から抜粋した文書を用意する。手順 1. 実験の概要について説明を行う。手順 2. ESG に関する基本事項について説明を行う。手順 3. 統合報告書から ESG に関する記述がされている段落を問うタスクを与える。この際には、タスクを与えられてから、解

答を終えるまでの時間の計測を行う。手順 4. その他 2 種類の統合報告書で手順 3. と同じ作業を行う。また、被験者が解答する文書の順序が結果に影響する可能性を考慮し、実験では以下のように、解答する文書の順序を指定した。この際、文書の内容を覚えてしまうことを防ぐため、被験者 1 人には、違う内容物の 3 つの文書 A,B,C を使用する。6 人には使用文書の上から順に文書に与えられる 1 つタスクをそれぞれ解答してもらった。以下、与えたタスクを記述する。

文書 A のタスク

温室効果ガス排出量に関する目標を記述しているのはどの段落ですか？

文書 B のタスク

現在、企業組織が進めているサステナビリティ経営について記述されているのはどの段落ですか？

文書 C のタスク

現在、技術的課題があることを示し、E（環境）への取り組みについて記述されているのはどの段落ですか？

5.2.8 実験結果

1 人目：大学生 D さん

使用文書	解答時間	タスクへの解答	正誤判定
B - 1	1 分 59 秒	8	不正解
C - 3	1 分 58 秒	18	正解
A - 2	0 分 58 秒	30	不正解

感想
・色と ESG 各々のイメージの対応関係があまりないため、どれがどの色か認識するのに最初は時間がかかった。

2 人目：大学生 E さん

使用文書	解答時間	タスクへの解答	正誤判定
C - 2	2 分 46 秒	18	正解
A - 1	1 分 51 秒	13	正解
B - 3	4 分 06 秒	18	正解

感想
<ul style="list-style-type: none"> ・ マーカーがある文書はない文書に比べて、その段落で何について書かれているかが分かりやすい。 ・ 全体的に馴染みのない言葉が多かったため、色分けされている方が読みやすかった。

3人目：大学生 F さん

使用文書	解答時間	タスクへの解答	正誤判定
A - 3	4分41秒	13	正解
B - 2	5分40秒	17	不正解
C - 1	5分56秒	18	正解

感想
<ul style="list-style-type: none"> ・ マーカーありは色によってどの段落でどのような内容が述べられているのかが分かりやすかった。 ・ ダミーが散りばめられていたり色が沢山散りばめられているため気が散った場合もあった。 ・ マーカーなしは余計な情報が無くシンプルで、まっすぐ文章を読む分には読み易かったが、どこに何が書いてあるか一目では判断しにくく、目的の段落付近にたどり着くまで時間がかかった。

4人目：大学生 G さん

使用文書	解答時間	タスクへの解答	正誤判定
B - 1	2分58秒	18	正解
C - 3	5分06秒	18	正解
A - 2	1分25秒	13	正解

感想
<ul style="list-style-type: none"> ・ 一番最初に読んだ b の文章は、読みやすかったが熟読する必要があった。 ・ 2 番目に読んだ C の文章はマーカーで気が散って読みづらく感じ、問いの中に具体的な名詞（温室効果ガス、サステナビリティなど）がなく解答を見つけづらい。 ・ 最後の A は問いが平易だったのと、マーカーの分布で解答しやすい。

5人目：大学生 H さん

使用文書	解答時間	タスクへの解答	正誤判定
A - 3	2分20秒	13	正解
B - 2	3分10秒	18	正解
C - 1	4分50秒	18	正解

感想
<ul style="list-style-type: none">・マークがない文章の方が、少し読むのが煩わしく感じる・しかしマークがあるとその部分に着目してしまい、内容が入ってきにくい

6人目：大学生 I さん

使用文書	解答時間	タスクへの解答	正誤判定
A - 3	4分00秒	13	正解
B - 2	3分00秒	18	正解
C - 1	5分30秒	18	正解

感想
<ul style="list-style-type: none">・マーカのない文書は要点を見つけることが難しい・ある文章では情報の整理がしやすい

7人目：大学生 J さん

使用文書	解答時間	タスクへの解答	正誤判定
B - 3	6分42秒	18	正解
A - 1	2分04秒	13	正解
C - 2	5分38秒	18	正解

感想
<ul style="list-style-type: none">・特に気になることはなかった

5.2.9 実験結果のまとめ

文書	参加者	解答時間	正誤
A-1	2人目：Eさん	1分51秒	正解
A-1	7人目：Jさん	2分04秒	正解
A-2	1人目：Dさん	0分58秒	不正解
A-2	4人目：Gさん	1分25秒	正解
A-3	3人目：Fさん	4分41秒	正解
A-3	5人目：Hさん	2分20秒	正解
A-3	6人目：Iさん	4分00秒	正解
B-1	1人目：Dさん	1分59秒	不正解
B-1	4人目：Gさん	2分58秒	正解
B-2	3人目：Fさん	5分40秒	不正解
B-2	5人目：Hさん	3分10秒	正解
B-2	6人目：Iさん	3分00秒	正解
B-3	2人目：Eさん	4分06秒	正解
B-3	7人目：Jさん	6分42秒	正解
C-1	3人目：Fさん	5分56秒	正解
C-1	5人目：Hさん	4分50秒	正解
C-1	6人目：Iさん	5分30秒	正解
C-2	2人目：Eさん	2分46秒	正解
C-2	7人目：Jさん	5分38秒	正解
C-3	1人目：Dさん	1分58秒	正解
C-3	4人目：Gさん	5分06秒	正解

以上の結果より、不正解箇所の確認、解答時間、実験終了後の感想についての確認を行った。まず、9つの文書中3つの不正解があり A-2 は、比較的早い時間で解答されていたが、異なる段落を選択していた。A の問いは、「温室効果ガス排出量に関する目標を記述しているのはどの段落ですか？」というものである。誤って選択した段落は温室効果ガスという単語が含まれており、この単語の「室」が緑色でマークされて、一際目立っていた。しかし、排出量に関する目標は記述されていない。次に B-1 についてである。B の問いは「現在、企業組織が進めているサステナビリティ経営について記述されているのはどの段落ですか？」というものである。被験者の選択した段落 8 では、中長期経営方針について述べられており、持続可能という単語も含まれていた。そのため、段落 8 がサステナビリティ経営について述べられていると判断したと考えられる。次に不正解箇所 B-2 についてである。選択段落は 17 で正解段落は 18 であり、後の被験者への確認で「サステナビリティ」というキーワードが多くマークされていることから場所はすぐに絞り込めたが、17 と 18 で非常に迷い、解答時間もかかってしまったことがわかった。17 はサステナビリティ経営について初めて具体的に述べられている段落ではあるが、現在のサステナビリティ経営について記述されている段落 18 である。いずれもタスクの解答としては不正解だが、ESG についての記載がある段落を選択していた。

次に解答時間に関して、1 人目の平均解答時間は 1 分 38 秒、2 人目の平均解答時間は 2 分 54 秒、3 人目の平均解答時間は 5 分 25 秒、4 人目の平均解答時間は 3 分 9 秒、5 人目の平均解答時間は 3 分 23 秒、6 人目の平均解答時間は 4 分 10 秒、7 人目の平均解答時間は 4 分 38 秒であり、予備実験と比較して、解答時間の個人差は少なくできていることが確認できた。また、文書ごとの平均解答時間をみたところ、A は A-2 < A-1 < A-3、B は B-1 < B-2 < B-3、C は C-3 < C-2 < C-1 という結果になった。A に関しては ESG 関連語へマークされた文書への解答時間が一番短く、また感想からも ESG 関連語へのマークの有効性が確認できた。B、C に関しては、ESG 関連語へのマークが及ぼす影響を確認できなかった。しかし、実験語の感想からは文書 A 同様にマークされている文書は読みやすく感じている被験者が多いことがわかった。実験後に記述してもらった感想には、マーカーのある文書に関してポジティブな意見とネガティブな意見があった。まず、ポジティブな意見として、マーカーがある文書はその段落が何について記述されているのか分かり易い。全体的に馴染みがない単語が多いため、色分けされていて読みやすい。マーカーの分布で解答しやすかった。というものであった。続いてネガティブな意見には、色と ESG 各々のイメージの対応関係があまりないので、どれがどの色か認識するのに時間がかかる。色が沢山散りばめられているため気が散った。という意見があった。マーカーのない文書に対しては、余計な情報が無くシンプルで、まっすぐ文章を読む分には読み易い。しかし、どこに何が書いてあるか一目では判断しにくく、目的の段落付近にたどり着くまで時間がかかる。読みやすいが熟読する必要がある。というものがあつた。

以上の結果から、提案手法は統合報告書から ESG 情報を獲得するには、多くの時間と労力を要するという課題に対して、有効な場合と、そうでない場合が確認できた。ESG 関連語へのマークが有効でなかった理由としては、前述した文書 B の問題のように、ESG 関連語への

マークが段落選択の判断を悩ませていることが考えられる。また、実験後の感想からは、ESG 関連語にマークされている文書は、段落に何が記述されているのか分かり易い。情報整理がしやすいなどと、マークされていない文書に比較して良い印象をもっていることがわかった。これらの理由として、文章上で ESG 関連語にマークが集中している箇所が一眼でわかり、そこは ESG 情報が記述されている箇所である可能性が高く、読者は優先的にその箇所を読んでいるからだと考えられる。本研究の提案手法は、課題解決のための ESG 情報取得の観点では成功し、マークをつける単語の選び方は適切だったと考えている。しかし、実験場での問題点があった。今回の実験では、被験者の数が十分でないこと、被験者それぞれ文書の読書速度に差があること、マークされた ESG 関連語が段落選択の判断を悩ませていることを事前に把握できなかったことが本実験がうまく運ばなかった要因であると考えている。本研究での目的対象は ESG に関する専門知識を有さない人と設定しているため、被験者実験では ESG について事前知識を有していない大学生を対象とした。仮に、被験者実験の対象が ESG についての知識を有する人物であるとする、マークされている単語への妥当性の点で不安を覚えるかもしれないが、マークされている ESG 関連語をヒントに、惑わされることなく情報の取捨選択をし、容易に ESG 情報へ辿り着き、タスクの正解率は被験者が大学生の場合と比較して高くなり、解答にかかる時間は短くなると予想する。

第6章

結論

6.1 まとめ

本研究では、統合報告書から ESG 関連語を取得し、取得した ESG 関連語を用いることで、文章の中から ESG 情報が記載されている場所の把握が困難であるという点を解決できるのか確認するために被験者実験を行った。第 1 章では、本研究の背景を説明し、統合報告書において、ESG 情報が記載されている場所の把握が困難であることを指摘した。この問題の解決のために、ESG 関連語を取得し、被験者実験を行い確認することを論じた。第 2 章では、本研究を進めるにあたって関連手法について論じた。主に、自然言語処理技術に関して記述した。第 3 章では、本研究と関連する、ESG 情報を扱った研究、読解支援に関する研究について論じた。それぞれの関連研究と本研究との類似点、相違点について述べている。第 4 章では、ESG 関連語取得に向けた、提案手法について論じた。ESG 関連語の取得には主に、データとなる統合報告書の収集、テキストデータの抽出、形態素解析、Word2Vec モデルを作成、コサイン類似度の計算という手順を踏んだ。「環境」、「社会」、「ガバナンス」に対してのコサイン類似度の高い単語上位 100 語をここでの ESG 関連語とし、抽出した。第 5 章では、取得した ESG 関連語が上記の問題点の解決の要因になるか確かめるための実験について述べた。実験は予備実験、本実験と 2 段階にわけて行った。予備実験には 3 人に協力してもらい、実験手順や実験時間、対象者にかかる負荷の確認をした。ここで、事前の実験説明の仕方に問題があることが被験者へのアンケートからわかった。続いて、予備実験の内容を踏まえ、本実験を行った。実験の結果、文書ごとの平均解答時間をみたところ、A は $A-2 < A-1 < A-3$ 、B は $B-1 < B-2 < B-3$ 、C は $C-3 < C-2 < C-1$ という結果になった。A に関しては ESG 関連語へマークされた文書への解答時間が一番短く、B、C に関しては 2 番目の長さであった。以上の結果から、提案手法は統合報告書から ESG 情報を獲得するには、多くの時間と労力を要するという課題に対して、有効である場合とそうでない場合があることがわかった。ESG 関連語へのマークが有効でなかった理由としては、ESG 関連語へのマークが段落選択の判断を悩ませていることが考えられる。実験後の感想から、ESG 関連語へマークされた文書は、その段

落では何が記述されているのか分かり易いという様に、読者に対して、読みやすいという印象を多分に与えていることがわかった。本研究の提案手法は、課題解決のための ESG 情報取得の観点では成功し、マークをつける単語の選び方は適切だったと考えている。しかし、実験上での問題点があった。今回の実験では、被験者の数が十分でないこと、被験者それぞれ文書の読書速度に差があること、同様の ESG 関連語が複数の段落へ出現していることで、段落選択の判断を悩ませていることを事前に把握できなかったことが本実験がうまく運ばなかった要因であると考えられる。本研究の結論としては、提案手法をもって、課題を解決することはできなかった。

6.2 今後の展望

被験者実験の改善を行う。今回本実験がうまく運ばなかった要因として、被験者の数が十分でないこと、被験者それぞれ文書の読書速度に差があること、マークされた同様の ESG 関連語が複数の段落へ出現していることで、段落選択の判断を悩ませていることを事前に把握できなかったことが考えられる。以上の点を踏まえ、被験者実験を再度行い、統合報告書から ESG 情報を獲得するには、多くの時間と労力を要するという課題を解決できているかどうか、定量的データをもって示したい。

謝辞

本研究を進めるにあたり, 丁寧にご指導下さった新美礼彦教授に深く感謝いたします。また, 新美研究室の皆様, 被験者実験に協力して下さった皆様, その他研究に関するアドバイスをいただいた方々に深く感謝いたします。

The authors would like to thank Enago (www.enago.jp) for the English language review.

参考文献

- [1] 経済産業省. ESG 投資. https://www.meti.go.jp/policy/energy_environment/global_warming/esg_investment.html 参照 2022-10-28.
- [2] KPMG. Kpmg サステナブルバリューサービス・ジャパン. <https://assets.kpmg/content/dam/kpmg/jp/pdf/2022/jp-corporate-reporting-20220406.pdf> 参照 2022-12-29.
- [3] Mecab: Yet another part-of-speech and morphological analyzer. <http://taku910.github.io/mecab/> 参照 2022-12-29.
- [4] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.
- [5] 河村康平, 高野海斗, 酒井浩之, 永並健吾, 中川慧. 機械学習を用いた統合報告書の ESG 関連ページの推定. 人工知能学会金融情報学研究会, Vol. 2021, No. FIN-027, p. 11, 2021.
- [6] 秋山祥伍, 江口潤一, 鈴木智也. ニューステキストを用いた ESG ファクター運用. 人工知能学会金融情報学研究会, Vol. 2020, No. FIN-024, p. 183, 2020.
- [7] 本多ふく代, 本多薫. 文字強調が文章内容の認知に与える影響. 日本人間工学会第 63 回大会, Vol. 58, No. Supplement, pp. 1G4-03-1G4-03, 2022.
- [8] 中尾 桂子・森下淳也. 表示切替式読解支援における教育用「文型」に基づいた構造化. 人文科学とコンピュータシンポジウム じんもんこん 2011, 第 2011 巻, pp. 75-80, dec 2011.
- [9] 東洋経済新報社. ESG 企業ランキング 2022 年版. 東洋経済新報社, 2022.
- [10] 鈴木正敏. 日本語 wikipedia エンティティベクトル. http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/ 参照 2022-10-23.
- [11] オムロン株式会社. Omron 統合レポート 2021. https://www.omron.com/jp/ja/ir/irlib/irlib_ar21.html 参照 2022-11-10.