

テキストマイニングによる大学の統合報告書分析

田澤彩紀[†] 古川巧[†] 三浦拓海[†] 平野雄大[†] 原田淳子[†] 橋本隆子[‡]

[†][‡]千葉商科大学 商経学部 〒272-8512 千葉県市川市国府台1-3-1

E-mail: [†] {b910332, b910200, b930139, b930200, b930063}@st.cuc.ac.jp, [‡] takako@cuc.ac.jp

あらまし 統合報告書とは、企業の知的資産（無形資産）や財務データなどの情報をもとに、今後の事業展開や経営ビジョン、経営理念、価値創造のための方針や戦略、それらのプロセスなどをまとめたものである。近年、日本の大学も統合報告書を発行するようになり、社会貢献や価値創造といった活動をアピールしようとしている。本稿では、大学の統合報告書をテキストマイニング手法により分析することで、大学の特徴評価を試みている。分析の結果、統合報告書は大学の特性をよく表しており、大学評価のための新たな指標として有用な資源となる可能性があることがわかった。

キーワード 統合報告書、大学、可視化、テキストマイニング、クラスタリング

1. はじめに

統合報告書とは、企業の財務情報に加え、企業統治や社会的責任(CSR)、知的財産(無形資産)などの非財務情報に基づいて、今後の事業展開、経営ビジョン、価値創造のための方針や戦略等をまとめたものである。2020年時点で約600社の企業が統合報告書を発行しているが、近年は日本の大学も統合報告書を発行するようになっている。2020年は国立大学を中心に18の大学が統合報告書を発行しており、統合報告書を通じて、社会貢献や地域連携など、従来の大学の評価基準に依存しない新たな価値をアピールしようとしている。その意味で、統合報告書は大学の特徴を表現していると言える。統合報告書を分析することで、その特徴を可視化し、理解することが可能となると考える。

そこで本論文では、19大学24本の統合報告書を対象として、テキストマイニングの手法を用いて分析を行い、統合報告書が大学の特徴を表しているかを評価する。テキストマイニングの手法としては、統合報告書内に出現する単語（名詞）の頻度情報に基づくワードクラウドを用いた可視化、LDA (Latent Dirichlet Allocation) [1]によるクラスタリングを行った。分析結果から統合報告書が大学の価値を表現するとともに、その類似性から大学をグループ化することが可能となったことがわかった。

2. 先行研究 – 統合報告書分析

統合報告書の分析については、河村ら[2]が、機械学習の手法を用いて企業の統合報告書のESG (Environment, Social and Governance) 関連ページを推定する手法を提案している。これは、数十ページに及ぶ長文の報告書からESG関連の情報を容易に見つけることを目的としている。村井ら[3]は、企業の環境報告書に対してテキストマイニングを行っている。報告

書から抽出されるキーワードと環境保全コストの金額との関係を分析し、コストの多寡に応じてキーワードが変化する傾向があることを示している。これらの研究は、統合報告書内のESG、環境といったトピックに注目しており、統合報告書全体を捉えた特徴分析はしていない。

中尾ら[4]は、統合報告書、環境報告書、サステナビリティ報告書内のトップマネジメントメッセージを対象とし、深層学習を用いて社会・環境ラベルを自動的に付与することで、企業のトップメッセージにおける環境・社会情報開示の傾向を分析している。テキストマイニングの分析対象を経営トップのメッセージに限定しており、統合報告書全体を評価していない。統合報告書には財務情報と非財務情報の両方が含まれており、本研究では報告書全体を評価することで、より幅広い情報の分析を目指す。

大坪ら[5]は、統合報告書を含む企業の非財務報告書を対象として、形態素解析によって抽出した単語とその出現頻度による報告書の類型化を試み、組織の規模が報告書の形態に影響を与えていることを示唆した。我々も同様のアプローチをとっているが、対象は大学の統合報告書であり、LDAによるクラスタリングを行うことで、統合報告書の特徴づけを目指している。

3. 大学による統合報告書

日本の大学における統合報告書の歴史は浅い。日本には795校の大学があり、そのうち国立が86校、公立が94校、私立が615校である。表1から分かるように、大学の統合報告書の発行は2018年度の東京大学から始まり、その後2019年度は11校、2020年度は18校と徐々に増えつつある。日経BPコンサルティングの調査によると、統合報告書を発行している大学は全て国立大学であり、2020年度まで、公立大学、私立大学は統合報告書を発行していない[6]。千葉商科大学は

2021 年の 8 月に、私立大学の先駆けとして統合報告書を発行している。

表 1 統合報告書を発行している大学一覧（2018-21 年度）（2021 年 12 月時点）＊太字は私立大学

2018 (1 校)	2019 (11 校)	2020 (18 校)	2021 (14 校)
東京大学	東京大学 宇都宮大学 筑波大学 千葉大学 一橋大学 新潟大学 福井大学 三重大学 神戸大学 岡山大学 東京海洋大学	東京大学 宇都宮大学 筑波大学 千葉大学 一橋大学 新潟大学 福井大学 三重大学 神戸大学 岡山大学 東京海洋大学 佐賀大学 信州大学 島根大学 滋賀大学 東京外語大学 北海道教育大学 滋賀医科大学	東京大学 宇都宮大学 新潟大学 神戸大学 筑波大学 一橋大学 岡山大学 三重大学 小樽商科大学 島根大学 東京工業大学 浜松医科大学 北海道教育大学 千葉商科大学

4. テキストマイニングによる分析

4.1. 対象とした統合報告書

今回評価対象としたのは北海道教育大学、東京大学、宇都宮大学、筑波大学、一橋大学、新潟大学、三重大学、神戸大学、岡山大学、島根大学、小樽商科大学、千葉商科大学、東京工業大学、浜松医科大学、東京海洋大学、千葉大学、福井大学、滋賀医科大学、佐賀大学の計 19 大学の 24 統合報告書である。千葉商科大学以外は全て国立大学であり、大学によって 2020 年度と 2021 年度発行のもの、2020 年または 2021 年単年度のみ発行のものがある (表 2)。各大学の統合報告書は最も短いもので 12 ページ(小樽商科大学 21 年度版)、長いもので 90 ページ(東京大学 21 年度版)に及ぶが、平均して 50 ページ程度で構成されている。それぞれ各大学の理念や目標・ビジョンや研究・教育・地域活動などが記載されている。

4.2. ワードクラウド

統合報告書の内容を可視化するために出現単語 (名詞) の頻度に基づき、ワードクラウドを作成した。ワードクラウドとは文章中で出現頻度が高い単語を選んで、その頻度に応じた大きさで図示する手法である。ワードクラウドは、形態素解析ライブラリ MeCab[30]を用い、「大学」や各大学名といった超頻出単語を削除して作成した。図 1 に調査対象である 24 報告書のワードクラウドの結果を示す。

全体の傾向として、「研究・教育・学生・社会」といった単語が多いことがわかる。以下に頻出単語の特徴を 4 つ挙げる。

表 2 対象とした統合報告書 (19 大学 24 報告書、太字は私立大学)

ID	大学名	2020 年度	2021 年度
1	北海道教育大学[7][8]	○	○
2	東京海洋大学[9]	○	
3	東京大学[10][11]	○	○
4	千葉大学[12]	○	
5	筑波大学[13][14]	○	○
6	宇都宮大学[15]	○	
7	一橋大学[16][17]	○	○
8	神戸大学[18][19]	○	○
9	滋賀医科大学[20]	○	
10	浜松医科大学[21]		○
11	佐賀大学[22]	○	
12	福井大学[23]	○	
13	島根大学[24]		○
14	三重大学[25]		○
15	東京工業大学[26]		○
16	岡山大学[27]		○
17	新潟大学[28]		○
18	小樽商科大学[29]		○
19	千葉商科大学[30]		○

特徴-1. 「研究・教育・学生・社会」

「研究・教育・学生・社会」といった単語は、教育機関・研究機関である大学らしい単語である。この単語を高頻出で持つ大学は、東京大学、千葉大学、東京工業大学、筑波大学など、比較的高偏差値の大学が挙げられる。

特徴-2. 「収益・収入・経営・財務」

これらは通常は、営利組織である企業に当てはまるような単語である。しかし、こうした単語は、国立大学の法人化や、少子化による大学経営への危機感を示していると考えられる。この単語を高頻出で持つ大学は、東京大学、神戸大学、福井大学、小樽商科大学などである。

特徴-3. 「医療・病院・医学・看護」

これらの単語は、主に医科大学 (滋賀医科大学、浜松医科大学) の報告書に多く出現する。医科大学の場合、大学の役割として、医学・医療的な側面を打ち出していると解釈できる。これらの単語は SDGs の 17 目標の 1 つである「すべての人に健康と福祉を」に当てはまる単語でもある。島根大学(21 年度)や神戸大学(20 年度)などの医学部を持つ大学も使用していると想定できる。

特徴-4. 「地域・環境・国際・グローバル・海外」

これらは特に国際化や環境を表現する単語である。地域に根ざした取り組みをすすめる大学やグローバル人材の育成を目指す大学がこのような単語を使用して

1.北海道教育大学(20 年度)



1.北海道教育大学 (21 年度)



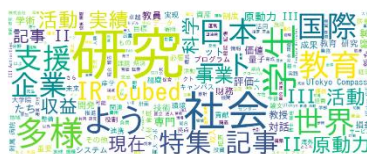
2.東京海洋大学(20年度)



3.東京大学(20 年度)



3.東京大学(21 年度)



4.千葉大学(20 年度)



5.筑波大学(20 年度)



5.筑波大学(21 年度)



6.宇都宮大学(20 年度)



7.一橋大学(20 年度)



7.一橋大学(21 年度)



8.神戸大学(20 年度)



8.神戸大学(21 年度)



9.滋賀医科大学(20 年度)



10.浜松医科大学(21 年度)



11.佐賀大学(20 年度)



12.福井大学(20 年度)



13.島根大学(21 年度)



14.三重大学(21 年度)



15.東京工業大学(21 年度)



16.岡山大学(21 年度)



17.新潟大学(21 年度)



18.小樽商科大学(21 年度)



19.千葉商科大学(21 年度)



図 1 対象報告書のワードクラウド

いと推測できる。この特徴を表す大学は、三重大学、筑波大学、神戸大学、佐賀大学、島根大学、新潟大学などである。

ワードクラウドの結果から、上記のような特徴が見られ、統合報告書によって大学の特徴を表現することが可能であることが確認できた。上記の分類は目視による判断であるが、次章では LDA による分析について記述する。

4.3. LDA による分析

次に、トピックモデル LDA を用いて、統合報告書のクラスタリングを試みる。LDA には Python Gensim [31] を使い、24 本の統合報告書のトピック分類を行った。形態素解析ライブラリである MeCab [30] を用いて、統合報告書の全文から単語を抽出し、それを LDA の入力とした。ワードクラウドと同様に、対象となる大学名、助詞などのストップワードをあらかじめ削除し、名詞、動詞、形容詞を入力単語とした。トピック数は Perplexity の値を考慮し、6 とした。

表 3 にトピック分類の結果と、そのトピックに最も所属確率の高い統合報告書名（大学名と該当年度）を示した。単語はそのトピックにおいて寄与の大きい順に並べている。図 2 は、LDA の結果を可視化している。円の大きさは、そのトピックに所属する大学数（多ければ大きい）を示している。また、前章の特徴との関連も記載した。

前章におけるワードクラウドの結果分析（目視）において、**特徴-1**「研究・教育・学生・社会」、**特徴-2**「収益・収入・経営・財務」、**特徴-3**「医療・病院・医学・看護」、**特徴-4**「地域・環境・国際・グローバル・海外」といった単語群が統合報告書の特徴づけると考察した。これらの特徴が、表 3 の LDA の結果にも現れている。

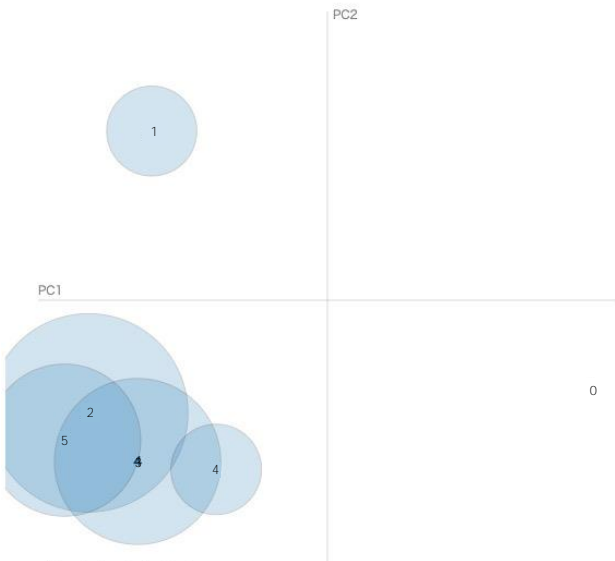


図 2 LDA のクラスタリング結果可視化

以下に前章の特徴と LDA の結果との関係を考察する。

特徴-1. 「研究・教育・学生・社会」

これは、図 3 の Topic 3 に対応していると考えられる。ここには、北海道教育大学(20-21 年度版)、東京大学(20-21 年度版)、東京海洋大学(20 年度版)、神戸大学(21 年度版)といった大学が所属しており、教育・研究と社会活動、支援をアピールしようとする意図が感じられる。図 3 の Topic 4 もこの特徴を有しているが、Topic 3 との大きな違いは、Topic 3 においては、「研究」の寄与が最も大きいのに対し、Topic 4 では、「研究」よりも「学生」の寄与が大きいところにある。

特徴-2. 「収益・収入・経営・財務」

これは、図 3 の Topic 2 と 3 に対応すると考えられる。単語の寄与度を考慮すると、Topic 3 よりも Topic 2 のほうがややこの特徴が強い。一橋大学、千葉大学、東京工業大学などを始めとする国立大学が、大学の収益、経費に対して配慮をしていることが伺える。

特徴-3. 「医療・病院・医学・看護」

これは、図 3 の Topic 5 に相当すると考えられる。医療系の大学、または医学部をもち、それをアピールしたい大学がこのトピックに主に所属している。

表 3 LDA によるトピック分類の結果

Topic ID	単語	統合報告書[大学名(年度)]
Topic 0	研究、学生、支援、社会、教育、地域、収益、連携、推進、実施	所属確率の高い大学はなし
Topic 1 ⇒ (特徴-4)	学生、研究、地域、教員、支援、設置、学部、4 月、教育、学長、推進、評価	宇都宮大学(20 年度)
Topic 2 ⇒ 特徴-2 特徴-4	研究、教育、学生、収益、経費、実施、支援、本学、推進、社会、国際、プログラム	一橋大学 (20-21 年度)、 三重大学 (21 年度)、 千葉大学 (20 年度)、 小樽商科大学 (21 年度)、 新潟大学(21 年度)、 東京工業大学 (21 年度)、 福井大学 (20 年度)、 筑波大学(20-21 年度)
Topic 3 ⇒ 特徴-1 (特徴-2)	研究、教育、学生、活動、社会、教員、支援、事業、本学、収益、経費、設置	北海道教育大学 (20-21 年度)、 東京大学 (20-21 年度)、 東京海洋大学 (20 年度) 神戸大学(21 年度)、
Topic 4 ⇒ (特徴-1)	学生、研究、教育、社会、地域、SDGs、支援、活動、実施、企業、取り組み、連携	千葉商科大学(21 年度)、 岡山大学 (21 年度)、
Topic 5 ⇒ 特徴-3 特徴-4	研究、地域、教育、医療、推進、支援、収益、病院、本学、連携、附属、設置	佐賀大学(20 年度)、 島根大学 (21 年度)、 浜松医科大学 (21 年度)、 滋賀医科大学 (20 年度) 神戸大学(20 年度)、

特徴-4. 「地域・環境・国際・グローバル・海外」

これは、図3のTopic5とTopic2に当たると考えられる。例えば筑波大学はスーパーグローバル大学創生支援プログラムなどを展開しており、国際的な活動に注力している。またTopic1にも地域の観点が含まれており、宇都宮大学が地域活動をアピールしていることが分かる。

この他に、Topic4はSDGsに注力している大学のグループであると想定できる。実際、千葉商科大学と岡山大学の統合報告書には、他大学に比べてSDGs活動に関する記述が多く、大学がSDGsを重視していることがよく表現されている。

また、神戸大学は、2020年度はTopic5のグループに所属していたが、2021年度はTopic3のグループに移動した。大学の方針や、社会ニーズの変化などによって、統合報告書の内容が変化していることを示唆しており、時系列的な評価も興味深い。

5. まとめ

本論文では、20大学24本の統合報告書を対象として、テキストマイニングの手法を用いて分析を行い、統合報告書が大学の特徴を表しているかを評価した。分析結果から統合報告書が大学の価値を表現するとともに、特徴によって大学をグループ化することが可能となった。

本来、偏差値、就職率、研究成果といった指標で評価されてきた大学の価値は、SDGsの登場や社会状況の変革、少子高齢化などによって、より社会的な観点を重視したものになりつつある。統合報告書は、従来の大学の価値に、そうした新たな価値が加わった社会的な存在価値を表現しようとしている。その意味で統合報告書の重要性はこれからますます増大すると想定できる。

今後も継続的に大学の統合報告書について分析を行い、社会状況の変化や大学の特徴に従って、統合報告書がどのように変遷していくかを評価し、大学の新たな価値を表現するための指標の明確化につなげていきたいと考えている。またこれらの評価結果を外部の報告書やデータを用いて分析することで、より多角的な視点から分析を行っていきたいと考える。

参 考 文 献

- [1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022, 2003.
- [2] 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”, 第27回人工知能学会 金融情報学研究会 (SIG-FIN) (2021).
- [3] 村井孝行, 中條良美, 朴恩芝, 前田利之, “テキストマイニングによる環境コスト支出の要因分析”, 経営情報学会 全国研究発表大会要旨集 2011 年秋季全国研究発表大会 (pp. 76-76) (2011).
- [4] 中尾悠利子, 石野亜耶, 岡田斎, “ニューラルネットワークによるサステナビリティ情報のテキスト分析——経営トップメッセージの環境・社会記述分析への適用”, 企業と社会フォーラム学会誌, 第8号, pp. 57-72 (2019).
- [5] 大坪史治, 黄海湘, “非財務報告書の類型化の試み”, 獨協経済, (100), 93-99 (2017).
- [6] 大学版統合報告書の現状と課題, 日経 BP コンサルティング ブランド戦略&マーケティング情報メディア, https://consult.nikkeibp.co.jp/ccl/atcl/20210712_1/ (2021)
- [7] 国立大学法人 北海道教育大学統合報告書 2020, <https://www.hokkyodai.ac.jp/files/00000500/00000525/2020tougouhoukokusyo.pdf>
- [8] 国立大学法人 北海道教育大学統合報告書 2021, <https://www.hokkyodai.ac.jp/files/00000500/00000525/2021tougouhoukokusyo.pdf>
- [9] 東京海洋大学 2020 統合報告書, <https://web-pamphlet.jp/kaiyodai/2021e2/book.pdf>
- [10] 東京大学 統合報告書 2020, <https://www.u-tokyo.ac.jp/content/400148351.pdf>
- [11] 東京大学 統合報告書 2021, <https://www.u-tokyo.ac.jp/content/400175829.pdf>
- [12] 千葉大学 統合報告書 2020, <https://www.chiba-u.jp/general/disclosure/announce/files/announce/integratedreport.pdf>
- [13] 筑波大学 統合報告書 2020, <https://www.tsukuba.ac.jp/about/disclosure-report/pdf/integrated-report2020.pdf>
- [14] 筑波大学 統合報告書 2021, <https://www.tsukuba.ac.jp/about/disclosure-report/pdf/integrated-report2021.pdf>
- [15] 宇都宮大学 アクションプラン&フィナンシャル 統合報告書 2021, https://www.utsunomiya-u.ac.jp/docs/action-financial2021_1203.pdf
- [16] 国立大学法人 一橋大学 統合報告書 2020, <https://www.hit-u.ac.jp/guide/information/pdf/R2/IntegratedReport2020.pdf>
- [17] 国立大学法人 一橋大学 統合報告書 2021, <https://www.hit-u.ac.jp/guide/information/pdf/R3/IntegratedReport2021.pdf>
- [18] 神戸大学 統合報告書 2020, <https://web-pamphlet.jp/kobe-u/2020e9/book.pdf>
- [19] 神戸大学 統合報告書 2021, <https://web-pamphlet.jp/kobe-u/2021e27/book.pdf>
- [20] 滋賀医科大学 統合報告書 2020, https://www.shiga-med.ac.jp/sites/default/files/2021-03/SUMS%20Integrated%20Report2020_web1_0.pdf
- [21] 国立大学法人 浜松医科大学 統合報告書 2021, https://www.hama-med.ac.jp/about-us/mt_files/integrated_report2021.pdf
- [22] 佐賀大学 統合報告書 2020, <https://www.saga-u.ac.jp/koukai/intergrated2020.pdf>
- [23] 福井大学 統合報告書 2020, <https://www.u-fukui.ac.jp/wp/wp-content/uploads/r1tougou.pdf>
- [24] 島根大学 統合報告書 2021, <https://www.shimane->

u.ac.jp/_files/00252270/tougouhoukoku2021-2.pdf

- [25] 三重大学 統合報告書 2021, <https://www.mie-u.ac.jp/profile/files/7a91181272dbcaae028b9f1c4f96bc87.pdf>
- [26] 東京工業大学 統合報告書 2021, <https://www.titech.ac.jp/public-relations/pdf/integrated-report-2021.pdf>
- [27] 岡山大学 統合報告書 2020, https://www.okayama-u.ac.jp/up_load_files/freetext/kaikaku-annual/file/OU_integratedreport2020_v.pdf
- [28] 新潟大学 統合報告書 2020, <https://www.niigata-u.ac.jp/wp-content/uploads/2020/11/ir2020.pdf>
- [29] 国立大学法人 小樽商科大学 統合報告書 2021, <https://www.otaru-uc.ac.jp/wp-content/uploads/2021/11/8010a1f62c357655d5ff7c346e8737d0.pdf>
- [30] 千葉商科大学 統合報告書 2021, <https://www.cuc.ac.jp/integratedreport/2021/>
- [31] Taku Kudo. “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”, <https://taku910.github.io/mecab/> .
- [32] Gensim, https://radimrehurek.com/gensim/auto_examples/core/run_core_concepts.html