

〔研究ノート〕

事業等のリスクの分析

—記載内容のトピックにもとづくアプローチ—

土屋 和之

1. はじめに

有価証券報告書における記述情報の開示は、開示府令が改正されて大幅な開示の拡充が図られており、その一部はすでに適用が進んでいるところである。同時に、このような記述情報の開示拡充の効果を検証するため、開示の実態など、記述情報の分析は、今後、記述情報の開示のあり方を議論する上で、必要不可欠となるであろう。

一方で、EDINETでの法定開示書類のXBRL形式の開示によって、こうした記述情報が大量のテキスト情報として入手可能となったことから、記述情報について、これまでとは異なる分析が可能となっている⁽¹⁾。

記述情報のなかでも、これまで取り上げられることの多かった事業等のリスクについて、従来の研究をふまえて、単語の出現頻度をもとに文書の類似度を比較して、業種によって事業等のリスクが似ている業種と、そうでない業種があることを明らかにした⁽²⁾。そこで今回は、自然言語処理の分野で利用されているトピックモデルによって、事業等のリスクの記載内容を分析し、業種による類似度の違いの中身を明らかにすることにしたい。

本研究は次のような構成となっている。すでに前回の研究で事業等のリスク開示の導入の背景から、開示府令の規定などを整理し、開示を対象にした研究を再検討した。その後、開示府令が改正され、事業等のリスクについても記載内容等について改正があったことから、どのような経緯で開示府令の改正が行われたのか、特に事業等のリスクについてどのような改正が行われたのかを概観しておく。

次に、分析の方法と対象について検討する。本研究では事業等のリスクの記載内容をトピックモデルを使って推定する。先行研究をふまえて、事業等のリスクをパラグラフに分割して、トピックモデルによる記載内容の推定を行っている。

次に、分析の結果を検討する。まず、トピックモデルによって事業等のリスクの記載内容がどのように推定されたのかを示す。次に、それにもとづいて、各社の事業等のリスクの記載内容をどう決定するかを明らかにする。その結果、分析の対象となった会社全体の事業等のリスクの記載内容を概観した上で、業種ごとにどのような内容が事業等のリスクに記載されているのかを見ていくことにする。

最後に分析結果をまとめるとともに、今後の課題について検討することにしたい。

(1) テキストに表れている感情を分析した感情分析や、その株価への影響に関する研究はすでに数多く行われている(首藤 [2019], 奥村 [2018])。

(2) 土屋 [2018]。

2. 事業等のリスクの開示をめぐる動向

事業等のリスクの開示については、2019年1月31日に施行された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（平成31年内閣府令第3号）により、開示内容の拡充が図られることとなった⁽³⁾。

この改正は、2018年6月28日に公表された金融審議会ディスクロージャーワーキング・グループ報告を受けたものである。この報告書では、企業価値向上のため企業と投資家の対話を促進するため「財務情報」および「記述情報」の充実など4項目について提言がなされている。このうち、記述情報を、財務情報を理解するために必要な情報であるとし、具体的に投資家が経営者の視点から企業を理解するための情報を提供し、財務情報全体を分析するための文脈を提供するとともに、企業収益やキャッシュ・フローの性質やそれらを生み出す基盤についての情報提供を通じ将来の業績の確度を判断する上で重要としている。その上で、経営戦略・ビジネスモデル、MD&A、リスク情報、その他について検討している。

事業等のリスクについては、全体としてみると、一般的なリスクの羅列となっている記載が多く、外部環境の変化にかかわらず数年間記載に変化がない開示例も多いほか、経営戦略やMD&Aとリスクの関係が明確でなく、投資判断に影響を与えるリスクが読み取りにくいと指摘を行っている。また、投資判断に当たっては、企業固有のリスク、リスクが顕在化した際の影響度、リスクへの対応策等の開示が重要とも指摘している。そこで、経営者視点からみたリスクの重要度の順に、発生可能性や時期・事業に与える影響・リスクへの対応策等を含め、企業固有の事情に応じたより実効的なリスク情報の開示を促していく必要があると指摘している⁽⁴⁾。

ディスクロージャーワーキング・グループ報告を受けて、改正された内閣府令では、事業等のリスクについて次のように改正が行われている。有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、経営者が連結会社の経営成績等の状況に重要な影響を与える可能性があるとして認識している主要なリスクについて、当該リスクが顕在化する可能性の程度や時期、当該リスクが顕在化した場合に連結会社の経営成績等の状況に与える影響の内容、当該リスクへの対応策を記載するなど、具体的に記載することとし、記載にあたっては、リスクの重要性や経営方針・経営戦略等との関連性の程度を考慮して、わかりやすく記載するように求めている。事業等のリスクに関する規定は、2020年3月31日に終了する事業年度に係る有価証券報告書から適用されることとされているが、2019年3月31日から早期適用することも妨げないとされている⁽⁵⁾。

また、開示府令の改正に合わせ、報告書で求められていたプリンシパルベースのガイダンスとして、2019年3月19日に「記述情報の開示に関する原則」と、投資家とアナリストと企業が開示の好事例について意見交換する場を設置し、そこで収集された好事例をと

(3) 有価証券報告書での事業等のリスクの開示の導入から、今回の改正までの経緯については、土屋 [2018] で整理している。

(4) 金融庁 [2018, p. 6]。

(5) 開示項目によって適用時期が異なっている。適用時期の概要は八木原他 [2019, p. 98] にまとめられている。

りまとめた「記述情報の開示に関する好事例集」が公表されている⁽⁶⁾。

原則は、総論と各論から構成され、総論では、記述情報の開示全般に共通する原則、考え方やおよび望ましい開示に向けた取り組みを示し、各論では、事業等のリスクなど、法令上記載が求められている事項に関して、開示の考え方や望ましい開示に向けた取り組みを提示している。

事業等のリスクについては、一般的なリスクの羅列ではなく、投資家の判断に重要な影響を及ぼす可能性のある事項を具体的に記載すること、その際に、取締役会等でそのリスクの影響の程度や発生蓋然性に応じて、リスクの重要性をどう判断しているかについて説明すること、リスクの記載の順序については、取締役会等での重要性の判断を反映することなどが挙げられている。

好事例集では、各開示例として着目したポイントを原則に対応する形でコメントしており、有価証券報告書だけでなく、統合報告書など任意の開示書類における開示例のうち有価証券報告書での開示に参考になるものも含めている。事業等のリスクについては、事例として挙げられた各社について、適切な分類が行われていること、具体的であること、わかりやすいことが指摘されている。

3. 分析の方法と対象

トピックモデルとは、文書が生成される過程を、確率を用いてモデル化した確率モデルであり⁽⁷⁾、文書の内容を表すテーマであるトピックを推定する方法である。文書のトピックを推定することで、文書を分類したり、検索語に近い文書を探すことが可能になる。

ここではトピックモデルの代表的なモデルである LDA (Latent Dirichlet Allocation : 潜在ディリクレ配分法) を利用する。LDA では、次のように文書の生成過程をモデル化する。文書ごとにトピック分布が存在し、そこからトピックが生成され、トピックごとの単語の出現分布から単語が文書に出現する。LDA では、実際に観察される文書中の単語から、観察されないトピック分布と単語の出現分布を計算することでトピックを推定するのである⁽⁸⁾。

前回の研究では、k 平均法というクラスタリングの手法によって事業等のリスクを分類した。k 平均法では 1 つの文書が 1 つのクラスタに属すると考え、分類が行われる。一方、LDA では、1 つの文書が複数のトピックを持つと考えるので、トピックにもとづくクラスタリングとみなせば 1 つの文書が複数のクラスタに属すると考え、分類が行われる。

本研究では、このトピックモデルによって事業等のリスクのトピックを推定し、事業等のリスクの記載内容を分析することにした。分析の対象となるのは、2018 年 4 月 1 日から 2019 年 3 月 31 日までの間に決算日を迎える 4093 社のうち、2019 年 3 月末時点で東京証券取引所の上場銘柄一覧として業種が与えられている 3618 社の有価証券報告書であ

(6) 原則と好事例集については藤岡他 [2019] に詳しい。

(7) 岩田 [2015, p. iv]

(8) 計算のアルゴリズムにはいくつかあるが、アルゴリズムの詳細は本研究の範囲を超えるので、ここでは取り上げない。詳しくは、佐藤 [2015] の第 4 章を参照されたい。また、実際の分析については後述する。

る。トピックを推定する事業等のリスクは有価証券報告書のXBRLインスタンスの該当箇所を抽出している。このうち、2019年11月30日までに訂正有価証券報告書が提出されたものについては訂正有価証券報告書のXBRLインスタンスを使用している⁽⁹⁾。また、事業等のリスクについては、2019年3月31日の有価証券報告書から改正開示府令の早期適用が認められているが、今回はこれを考慮していない。

LDAに必要なデータは、単語と出現頻度をベクトルとして表現したbag-of-wordsである。トピックモデルでは文書ごとにbag-of-wordsを用意する必要がある。そこで1社ごとの事業等のリスクを1つの文書とすることが考えられるが、本研究では次のようにして事業等のリスクを分割することにした。

Bao and Datta [2014] は米国の年次報告書Form 10-Kに記載されるリスク情報であるItem 1A. Risk Factorについて、トピックモデルを用いた研究である。Bao and Datta [2014] は、Risk Factorを読むと、1つのセンテンスに1つのトピックが記述されていると指摘している。そこでBao and Datta[2014] は、トピックモデルを用いる場合に、Risk Factor全体を1つの文書として、そこに含まれるトピックを推定するのではなく、1つのセンテンスを1つの文書として、センテンスのトピックを推定する方法をとっている⁽¹⁰⁾。

一方、有価証券報告書の事業等のリスクについても、図表1のように、適当な見出しのもと、項目ごとに多くの場合1つの文章ではなく、複数の文章で、リスクが説明されている。したがって、事業等のリスクについては、1パラグラフ1トピックという記載が取られているとみることができる。

そこで、本研究では、事業等のリスクを、適当な区切り文字でパラグラフに分割することとした。適当な区切り文字で分割が困難な場合は、分割されずに複数のリスク項目が混じってしまうため、分析の精度向上の点から、今回は分析の対象から外すこととする。これによって分析の対象は2791社となっている。もともと複数のリスク項目の記載がない事業等のリスクは分析の対象からははずれることになる。したがって、今回の分析の対象となる事業等のリスクでは、記載されているリスク項目は2つ以上あることになる。

また、事業等のリスクでは、リスクの記載に先立って、リスクが有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資者の判断に重要な影響を及ぼす可能性のある事項であること、将来に関する事項は、当連結会計年度末現在に会社が判断したものであることなどが記載されているが、この記載は、リスクの内容に関する記載ではないことから削除する。したがって、図表1の事業等のリスクはパラグラフに分割し、記載内容の説明に関する部分を削除することによって、図表2のように3つのパラグラフになる。

以上から、分析の対象となるのは、2719社の事業等のリスクに記載された分割された24974のパラグラフとなる。24974のパラグラフを文書とみなし、パラグラフを分かち書

(9) ファイルの入手には2019年3月から公開されているEDINET APIを利用している。EDINET APIについては、金融庁企画市場局企業開示課 [2019] を参照のこと。

(10) Bao and Datta [2014] は、local-LDAと呼んでいる。もともとは製品やサービスに関するオンラインレビューの分析のためにBrody and Elhadad [2010] が採用したものである。

2【事業等のリスク】

有価証券報告書に記載した事業の状況、経理の状況等に関する事項のうち、投資家の判断に重要な影響を及ぼす可能性のある事項には以下のようなものが考えられます。なお、将来に関する事項については、当連結会計年度末（2019年3月31日）現在において判断したものであり、特に経営成績に重要な影響を与える恐れがあるものを中心として記載しております。

なお、以下のリスクが顕在化する可能性は、現時点において極めて低いと考えております。

(1) 製品及び製品開発におけるリスク

当社の主力製品である統合業務ソフトウェア「OBIC7シリーズ」は現在、総売上高の半分以上を占める主力製品であります。このコアシステムと周辺サブモジュール群はそれを支えるプラットフォームや開発言語の進化・変遷により開発のタイミングや製品の開発手法に大きな影響を及ぼすものであり、製品開発における発売時期の遅延やコストの上昇により業績が影響されるリスクがあるものと考えております。

(2) モチベーションの高い人材の流失とノウハウの喪失に関するリスク

当「情報サービス業界」は他の装置産業等に比べ労働集約的な側面があり「人材のモチベーション」が、よりダイレクトに業績に影響する可能性のある業界であります。当社は優秀な人材の確保と育成に毎年多くの時間とコストをかけ将来性豊かな社員の育成に努めてまいりました。仮想ではありますが、敵対的な買収者による奇襲攻撃的な企業買収行為が起きた場合には、人心の混乱を招き、結果としてモチベーションの高い人材の流失やノウハウの喪失を招くとすれば、安定的・継続的に成長・発展させていく企業活動にとって致命的な損失であり、当社の経営成績に重大な影響を与える可能性があると考えております。

(3) 顧客から預かるテストデータに関する情報管理におけるリスク

当社は、事業の性格上、システム導入や運用サポート時に、顧客企業のデータや情報を取り扱う場合があります。取り扱いに際しては、個人情報保護法に準拠して、情報管理規定の整備、研修を通じた社員への周知徹底、インフラのセキュリティ強化などにより、管理の強化・徹底と漏洩の防止に努めております。しかしながら、情報の授受、運搬時における紛失や盗難などにより、顧客企業の個人情報が漏洩した場合には、当該顧客からの損害賠償請求による費用発生や、社会的信用の低下などにより、当社の経営成績に影響を与える可能性があると考えております。

図表1：事業等のリスクの例

製品及び製品開発におけるリスク当社の主力製品である統合業務ソフトウェア「OBIC7シリーズ」は現在、総売上高の半分以上を占める主力製品であります。このコアシステムと周辺サブモジュール群はそれを支えるプラットフォームや開発言語の進化・変遷により開発のタイミングや製品の開発手法に大きな影響を及ぼすものであり、製品開発における発売時期の遅延やコストの上昇により業績が影響されるリスクがあるものと考えております。

モチベーションの高い人材の流失とノウハウの喪失に関するリスク当「情報サービス業界」は他の装置産業等に比べ労働集約的な側面があり「人材のモチベーション」が、よりダイレクトに業績に影響する可能性のある業界であります。当社は優秀な人材の確保と育成に毎年多くの時間とコストをかけ将来性豊かな社員の育成に努めてまいりました。仮想ではありますが、敵対的な買収者による奇襲攻撃的な企業買収行為が起きた場合には、人心の混乱を招き、結果としてモチベーションの高い人材の流失やノウハウの喪失を招くとすれば、安定的・継続的に成長・発展させていく企業活動にとって致命的な損失であり、当社の経営成績に重大な影響を与える可能性があると考えております。

顧客から預かるテストデータに関する情報管理におけるリスク当社は、事業の性格上、システム導入や運用サポート時に、顧客企業のデータや情報を取り扱う場合があります。取り扱いに際しては、個人情報保護法に準拠して、情報管理規定の整備、研修を通じた社員への周知徹底、インフラのセキュリティ強化などにより、管理の強化・徹底と漏洩の防止に努めております。しかしながら、情報の授受、運搬時における紛失や盗難などにより、顧客企業の個人情報が漏洩した場合には、当該顧客からの損害賠償請求による費用発生や、社会的信用の低下などにより、当社の経営成績に影響を与える可能性があると考えております。

図表2：分割された事業等のリスク

きした上で、ストップワードを削除し、名詞のみからなる bag-of-words を作成し⁽¹¹⁾、トピックモデルによる分析を行う。

4. トピックモデルによる事業等のリスクの分析

トピックモデルによるトピックの推定では、あらかじめトピック数を付与する必要がある。今回はトピック数を30に設定している。これまでの事業等のリスクの開示を調査した研究では、リスクをさまざまな基準で分類している。財務会計基準機構 [2005] が開示されたリスクを17に分類したのをはじめ、中野 [2010] は7つの分類を設けて、具体例として10のリスクを挙げている。また、野田 [2016] は12のリスクカテゴリーを設けている。このうち、張替 [2008] で、もっとも多い27のリスクに分類されていること、Bao and Datta [2014] でも先行研究をふまえて27に設定していること、事業等のリスクをパラグラフに分割した際に、パラグラフのもっとも多い会社でパラグラフが29であることなどから30として⁽¹²⁾トピックの推定を行った⁽¹³⁾。

分かち書きされた24974パラグラフをもとに、トピック数を30として計算した結果、推定されたトピックとそのトピックを代表する単語は図表3のとおりである。トピックモデルはいわゆる教師なし機械学習に当たるので、トピックモデルによってトピックのラベルは付与されない。また、トピックの番号は順不同である。

以上のようにトピックが推定されたので、各パラグラフのトピックを次のように決定する。各パラグラフについてトピックと出現確率が与えられるので、このうちもっとも出現確率が高いトピックをそのパラグラフのトピックに決定する。たとえば図表2のように3つの分割されたパラグラフは、図表4のようにトピックとその出現確率が計算される。例えば1つ目のパラグラフは、トピック4の確率が0.033、トピック5が0.373、トピック9が0.093、トピック12が0.034、トピック16が0.040、トピック18が0.103、トピック23が0.196、トピック24が0.036、トピック25が0.065であることを示している。このうちもっとも確率が高いのはトピック5なので、このパラグラフのトピックは5に決定される。したがって、この事業等のリスクのトピックは、トピック5、トピック28、トピック18であると推定される。

以上のようにして、24974のパラグラフのトピックを決定し、各トピックの数を集計したものが図表5である。また、図表6はトピック14に分類されたパラグラフの中から3

(11) 前回と同様、日本語の分かち書きを MeCab (<http://taku910.github.io/mecab/>) によって行い、MeCab が使う辞書として、mecab-ipadic-NEologd (<https://github.com/neologd/mecab-ipadic-neologd/>) を使用している。

(12) 生成されたモデルの汎化能力 (性能) の尺度となる perplexity をトピック数の決定の基準とすることもできる (佐藤 [2015, pp. 135-136]) が、金 [2018, p. 89] では、トピック数の決定にはさまざまな方法が提案されているが、どの方法も決定的な方法ではないことが指摘されている。こうしたことから今回はトピック数を先行研究などから決定している。

(13) 今回は、Python と、Python のトピックモデルのライブラリである gensim (<https://radimrehurek.com/gensim/>) の Idamodel を使用している。オプション等は指定していない。なお、gensim は、LDA のアルゴリズムとして変分ベイズ法を採用している。詳しくは以下に説明がある。<https://radimrehurek.com/gensim/models/ldamodel.html>

土屋和之：事業等のリスクの分析

トピック	トピックを代表する単語
0	訴訟, 重要, 収入, 提起, 結果, 関連, 内容, 依存, 事業活動, 経営
1	顧客, 販売, 機器, 技術, 製品, 装置, 製造, 業界, 医療, 向け
2	契約, 取引, 財務, 連結会計, 損失, 締結, 重要, 残高, 制限, 米ドル
3	取引先, 債権, 管理, 回収, 与信, 信用, 取引, 信用リスク, 悪化, お客さま
4	賃貸, 不動産, 物件, 期間, 加盟, 受注, 収支, 案件, 官公庁, 予算
5	競合, 競争, 商品, 開発, 他社, 市場, 激化, 製品, 販売, 新製品
6	工事, 建築, ガス, 受注, 仕入れ, 遅延, 環境規制, 電気, 得意, 請負
7	業務, 体制, 火災, 管理体制, プロジェクト, 整備, 内部, 組織, 強化, 管理
8	取引, 変更, 売上高, 販売, 特定, 依存, 要因, 顧客, テロ, 割合
9	売上高, 減少, 売上, 傾向, 商品, 企業グループ, 四半期, 季節, 構成, 天候
10	契約, 提携, ブランド, 権利, ライセンス, 手数料, 提供, 展開, 承認, 獲得
11	新規, 変化, 投資, 計画, 拡大, 環境, 対応, 開発, 確保, 成長
12	給付, 債務, 退職, 制度, 費用, 変更, 資産, 運用, 計算, 結果
13	動向, 市場, 需要, 業界, 住宅, 販売, 製品, 経済状況, 景気, 株式
14	店舗, 自然災害, 災害, 出店, 事故, 地震, 大規模, 設備, 被害, 施設
15	金利, 環境, 変化, 経済, 動向, 情勢, 設備投資, 資金調達, 調達, 資金
16	リース, 価格競争, シェア, 同業他社, 統合, 顧客, 介護, 市場, 技術革新, 携帯電話
17	固定資産, 減損, 資産, 減損損失, 適用, 計上, 減損会計, 低下, 処理, 保有
18	システム, 情報, 管理, 業務, 顧客, 情報システム, お客様, 信用, 情報セキュリティ, 食材
19	侵入, 排除, 多様, 印刷, 甚大, タイ, ダウン, コンピュータウイルス, 発電, 不慮
20	個人情報, 管理, 信用, 情報, 漏洩, 外部, 委託, 社会的, 流出, 個人情報保護
21	投資, M&A, 企業, 技術, 調査, 成果, 買収, 出資, 検討, 財務
22	保有, 株式, 下落, 有価証券, 不動産, 投資, 時価, 価値, 資産, 評価損
23	価格, 原材料, 調達, 製品, 仕入, 供給, 上昇, 商品, 生産, 高騰
24	自動車, 配送, 業者, 仕入, パートナー, 価格, 固定金利, 調達, 商品, 相場
25	製品, 品質, 欠陥, 品質管理, 保険, 問題, 製造, 商品, 製造物責任, 保証
26	知的財産権, 第三者, 侵害, 保護, 製品, 知的財産, 訴訟, 技術, 取得, 他社
27	海外, 為替, 為替レート, 取引, 為替相場, 換算, 中国, 地域, 通貨, 現地
28	確保, 人材, 育成, 採用, 優秀, 重要, 教育, 社員, 制度, 成長
29	規制, 法令, 法的, 許可, 遵守, 関連, 変更, 許認可, コンプライアンス, 登録

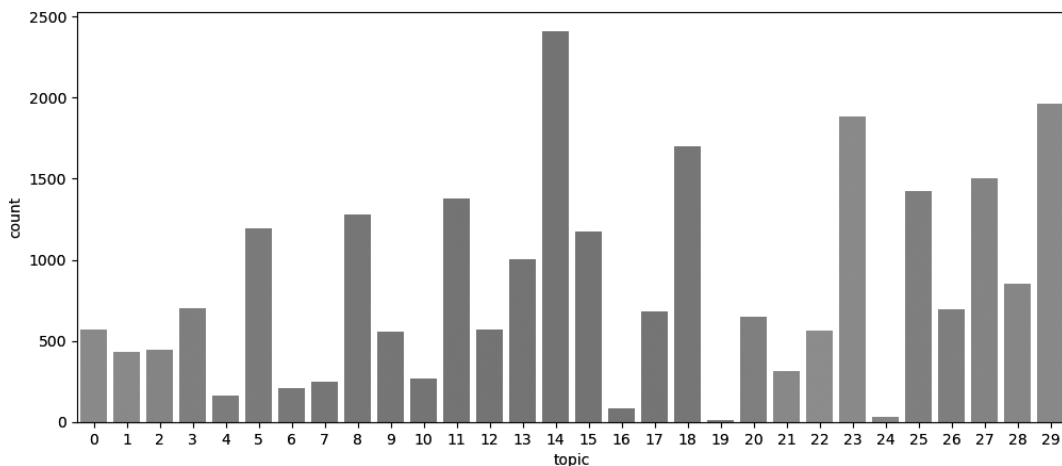
図表 3：トピックとトピックを代表する単語

つを、図表 7 はトピック 28 に分類されたパラグラフの中から 3 つを挙げている。

前回の研究では、先行研究をふまえ、業種ごとに事業等のリスクの開示の類似度に違いがあることを示した。そこで、どのようなトピックがどのくらい開示されているかを業種

(4, 0.033), (5, 0.373), (9, 0.093), (12, 0.034), (16, 0.040), (18, 0.103), (23, 0.196), (24, 0.036), (25, 0.065)
(3, 0.119), (8, 0.037), (10, 0.058), (16, 0.087), (18, 0.076), (21, 0.127), (28, 0.474)
(1, 0.111), (7, 0.030), (12, 0.037), (18, 0.524), (20, 0.224), (25, 0.026), (28, 0.029)

図表4：図表2の各パラグラフに与えられたトピックと確率



図表5：24973パラグラフのトピックとその数

ごとに見てみよう。なお、東証33業種のうちパラグラフの分割ができずに含まれなかった4業種を除く29業種が対象となっているが、7業種は会社数が少ないため、ここでは除外する。

図表8は、22業種について業種ごとの30のトピックの記載割合を%で表示したものである。業種の中でもっとも記載の多いトピックの割合を網掛けにしている。これを見ると製造業中心にトピック23の記載が多くなっている。また、他のトピックの記載割合が多い業種でも、いくつかの特定のトピックの記載が多い。一見全く違う業種であっても、認識しているリスクが類似していると考えられる。

前回の研究で行ったクラスタリングによると、銀行業、情報・通信業、建設業、不動産業、小売業で類似度が高いことが明らかになった。このうち不動産業は唯一トピック29の記載割合が多くなっている。また、銀行業と情報・通信業ではトピック18が、小売業と建設業ではトピック14がそれぞれ多くなっている。一方で、クラスタリングでは把握できなかった、トピック5の記載割合が高い医薬品と電気機器、トピック8の記載割合の高い輸送用機器、トピック14の記載割合が高い繊維製品、卸売業、トピック18の記載割合が高いサービス業、トピック27の記載割合の高いゴム製品と精密機器が特徴的であることが理解できる。

事業等のリスクに関する議論では、同じようなリスクが記載されていることが指摘されている。これは、今回推定されたトピックのうち、同じトピックの組み合わせが記載されていること、と考えることができるだろう。

重複、順番を無視して、全く同じトピックが記載されているケースを調べると、6社が同じトピックを記載しているケースが1、5社が同じケースが1、4社が同じケースが8、

自然災害、天候、伝染病等冷夏暖冬等の異常気象、台風や地震等の自然災害が発生した場合や伝染病が日本国内で流行した場合、当社グループの業績と財政状態に影響を及ぼす可能性があります。これらは予期できぬことですが、収益性の低下を招き、業績に悪影響を及ぼす可能性があります。

自然災害、事故、テロ、戦争等に関するリスク当社グループが事業活動を行っている地域では、地震、台風等の自然災害の影響を受ける可能性があります。同様に火災等の事故災害、テロ、戦争等が発生した場合、当社グループの拠点の設備等に大きな被害を受け、その全部又は一部の操業が中断し、生産及び出荷が遅延する可能性があります。また、損害を被った設備等の修復のために多額の費用が発生し、結果として、当社グループの業績及び財政状態に影響を及ぼす可能性があります。

災害等の発生について当社は、地震、台風、洪水等の自然災害や火災等の事故、及び新型インフルエンザ等の感染症の流行の発生を想定し、必要とされる安全対策や安否確認体制の構築等、事業継続計画を策定して影響の回避に努めております。しかしながら、大規模な災害、事故等が発生した場合は、全ての影響を回避することは困難であり、財政状況及び業績等に影響を及ぼす可能性があります。

図表 6：トピック 14 に分類されたパラグラフ

人材確保・維持について当社グループの成長を支えている最大の資産は人材であり、優秀な人材の採用と維持は当社にとって重要な課題であると認識しております。当社グループでは、優秀な人材の確保と育成については最大限の努力を払っておりますが、優秀な人材を確保・育成できない場合、また事業変革に伴うニーズにマッチした人材の補充ができない場合、当社グループの経営成績や成長に大きな影響を及ぼす可能性があります。

人材の確保について当社グループの中核事業である建設事業は、優秀な資格者と高度な技術による施工実績の良好な評価が、事業を継続拡大するためのベースとなっており、それゆえに優秀な人材を獲得し維持する必要があります。当社グループの人事部門は、優秀な人材を確保するため注力していますが、当社グループが必要とする人材が計画どおり確保できなかった場合は、当社グループの業績及び財務状況に悪影響を及ぼす可能性があります。

人材の確保及び育成について当社においては人材が重要な経営資源であり、当社の中長期的な成長のためには、採用計画に沿った人材の確保が不可欠な要素となっております。また様々な教育・研修制度や各種インセンティブ制度を用意し、それらの人材の入社後においても、人材の動機付けと、徹底的な教育・育成を行っております。しかしながら、今後の経営環境の急激な変化等により、人材の確保や育成が計画通りに進まない場合や、予測の範囲を超える多数の退職者が同時期に発生した場合等は、当社の業績に影響を与える可能性があります。

図表 7：トピック 28 に分類されたパラグラフ

3社が同じケースが21、2社が同じケースが164となっている。したがって、残りの2357社は同じトピックの組み合わせの記載の会社はいないことになる。

5. おわりに

本研究では、文書のトピックを推定するトピックモデルを使って、事業等のリスクの記載内容を分析した。事業等のリスク全体を1つの文書として扱うのではなく、事業等のリスクをリスク項目ごとに分割し、リスク項目が記載されたパラグラフを1つの文書として扱ってトピックモデルを使用している。

本研究で明らかになったことは次の2点である。1つ目は、パラグラフを1つの文書としたことによって、おおむねトピックが推定できている点である。事業等のリスクには多くのリスクが記載されることから、全体を1つの文書として扱っていると、分割したときに比べ、

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
食料品	1.51	0.58	0.93	2.09	0.58	3.59	0	0.12	4.87	2.67	0.46	4.06	1.97	3.24	10.66	2.32	0.23	4.17	6.72	0.12	1.51	0.46	2.2	15.99	0.23	11.59	1.27	5.79	1.62	8.46
建設業	1.69	0.62	1.62	7.24	2.31	1.31	8.86	1.31	1.54	2.77	0.46	2.54	3.16	2.16	9.78	8.86	1.08	0.92	3.24	0	1.62	0.69	5.47	7.09	0.15	7.16	0.08	4.08	2.47	9.71
金属製品	1.35	1.68	0.84	3.37	1.52	5.56	0.84	0.51	6.9	1.85	0.34	2.02	3.03	5.22	11.62	5.72	0	2.02	2.53	0	0.67	0.84	2.02	14.48	0	7.91	2.36	7.58	0.84	6.4
ガラス・土石製品	1.4	1.68	1.96	3.36	1.12	5.88	0.84	0.56	3.92	0.84	0	4.2	3.08	8.12	9.52	5.04	1.12	2.24	1.96	0	0.56	0.56	2.24	14.85	0	6.16	2.8	8.68	1.12	6.16
繊維製品	1.61	0.54	0.54	4.29	0.54	3.75	0	0.54	7.51	4.02	2.95	4.56	4.83	5.09	9.92	4.56	0.27	0.8	4.83	0	1.88	0.54	2.95	8.85	0.27	6.7	3.75	9.65	1.07	3.22
パルプ・紙	3.12	0	3.12	2.6	0	3.65	0	0	4.17	0.52	1.04	1.56	4.17	6.25	11.98	7.81	0	3.12	2.6	0	0.52	0.52	1.56	17.71	0	3.12	2.08	10.42	1.56	6.77
化学	2.83	1.11	0.68	1.6	0.31	6.27	0.25	0.18	6.21	1.04	0.43	3.93	2.46	4.55	10.69	2.95	0.12	2.89	3.01	0	0.61	1.23	1.78	12.6	0	7.68	5.04	9.4	1.66	8.48
医薬品	8	6.59	0.47	0	0	14.59	0	0	4.94	0.47	6.59	4.94	0.94	1.41	4.94	1.18	0	0.94	5.65	0	1.18	0.47	2.59	7.06	0	6.35	7.76	4.24	0.47	8.24
ゴム製品	0.63	1.26	1.89	0	0	1.89	0	0	7.55	1.89	0	0	5.03	3.14	10.06	4.4	0	4.4	3.77	0	0.63	3.14	3.14	11.95	0	11.32	5.03	12.58	1.26	5.03
輸送用機器	1.76	0.44	1.61	1.61	0.15	5.71	0.73	0.15	13.91	0.88	0.29	2.2	3.22	4.83	10.25	2.64	0.29	1.32	3.51	0	0.44	0.59	1.02	10.69	0.15	9.37	3.95	10.54	2.05	5.71
鉄鋼	0.64	0	1.29	1.29	0.96	2.89	1.29	0.64	6.43	1.29	0.64	2.57	3.54	5.47	14.15	4.5	0	2.89	2.25	0	0.64	0.96	4.5	19.29	0	4.5	1.29	7.4	1.93	6.75
非鉄金属	1.9	0.38	3.04	1.14	0	2.28	0	0	7.22	0	0.38	3.42	1.52	3.8	11.03	7.98	0	3.04	3.04	0	1.14	0.76	3.04	13.31	0	8.37	5.32	8.37	1.52	7.98
機械	21	2.15	1.87	1.98	0.17	6.24	0.68	0.23	7.48	1.47	0.34	3.57	2.32	4.71	9.81	4.59	0.34	2.1	3.63	0	0.57	0.85	2.38	9.92	0	7.26	3.97	9.64	3	6.63
電気機器	1.52	3.1	1.98	1.17	0.2	9.19	0.15	0.36	7.82	1.17	0.36	4.82	2.34	5.58	8.63	3.05	0.15	1.93	3.5	0.05	1.07	1.12	1.83	8.43	0	8.12	4.87	9.04	3.05	5.43
精密機器	1.26	4.53	2.52	1.01	0	7.56	0.25	0.5	8.56	1.01	0.76	6.3	0.76	5.54	6.05	3.02	0	2.27	2.27	0	1.51	1.51	1.51	8.31	0	7.81	6.55	10.08	2.27	6.3
その他製品	1.34	1.88	0.94	2.42	1.07	6.44	0.13	0.54	5.77	2.68	0.54	3.36	1.88	4.97	10.74	3.89	0.4	1.48	5.91	0	2.68	0.81	1.88	12.21	0	7.11	2.82	7.52	2.55	6.04
情報・通信業	3.04	2.26	1.6	1.13	0.7	4.6	0.23	3.27	5.06	2.26	2.18	12.89	1.44	3.58	4.09	1.52	0.43	1.36	14.8	0.08	2.49	2.92	1.36	1.13	0.04	3.54	5.49	2.57	8.14	5.77
サービス業	3.6	2.1	2.14	2.06	1.46	3	0.6	1.8	3.64	2.93	2.1	9.94	1.65	3.23	8.4	3.45	0.56	3	10.73	0.04	4.95	1.43	1.54	1.73	0.19	3.19	1.99	2.7	6.23	9.64
卸売業	1.69	2.51	1.73	6.45	0.17	3.94	0.17	0.56	5.84	2.73	0.82	3.76	2.38	4.93	8.91	5.06	0.39	2.73	5.15	0.13	2.34	1.47	3.16	8.31	0.17	5.71	1.08	8.48	2.12	7.1
小売業	1.99	1.39	2.83	2.39	0.24	4.26	0.04	0.72	2.35	3.87	1.2	6.42	0.8	3.63	16.74	3.83	0.24	5.42	7.05	0.04	6.3	1.04	0.76	5.1	0.48	3.91	1.24	3.63	3.59	8.53
銀行業	0.88	0	2.95	10.62	0	2.65	0	3.83	1.47	1.18	0	2.06	10.62	0.88	4.72	11.5	0	7.67	16.52	0	2.95	0.88	5.31	0	0.29	0.29	0	2.06	0.88	9.73
不動産業	4.23	0.9	1.54	2.3	2.3	1.92	2.69	1.41	1.54	4.35	0.13	3.07	1.15	4.74	7.3	14.98	0.9	3.46	2.56	0	5.89	0.9	2.82	3.33	0	4.99	0.38	1.15	4.1	14.98

図表8：トピックとトピックを代表する単語

トピックの推定が困難になるとみられる。これは、有価証券報告書の記述情報だけでなく、近年拡大しているいわゆる ESG 情報など任意開示の記述情報の分析についても、あらかじめ文書を分割しておけば、トピックモデルによるトピックの推定が可能になることを示している。

2つ目は、対象会社全部のトピックの推定ができたことで、先の研究で明らかになった、業種ごとの記載内容の類似度の違いが説明できることである。東証 33 業種のうち、銀行業、情報・通信業、建設業、不動産業、小売業は事業等のリスクの記載内容は類似していると見られる。トピックの推定によって、これが製造業を中心に記載されているトピックと比べて、これらの 5つの業種では異なるトピックの記載が多くなっていることが理由と考えられる。

一方で、本研究に関連して、さらに検討を要する点として、次の 2 点を挙げるができる。まず 1 点目は、事業等のリスクのパラグラフへの分割に関わる問題である。事業等のリスクのリスク項目ごとの記載は開示府令等で規制があるわけではないため、会社によって数字や記号をさまざまに使用している。今回は試行錯誤で分割しているが、うまく分割されていないケースも見られ、トピック推定に影響を与えている可能性がある。また約 3600 社のうち分割が困難な約 900 社を対象から除いているため、東証上場会社全体を把握できたと結論するのは難しいだろう。2 点目は先の研究でも指摘したとおり、分析に使用する単語の問題である。汎用の辞書を用いて形態素解析を行っているので、有価証券報告書で使用される単語が適切に分かち書きされていない可能性がある。また、分析にあたって除去すべき単語を決定するという問題もある。事業等のリスクを分析するために、必要な単語、あるいは必要のない単語を決定するのは非常に困難である。こうした問題は今後、記述情報を分析するためにはぜひとも検討しておかなければならないだろう。

有価証券報告書の記述情報については、第 2 節で取り上げたとおり、改正された開示府令に従った開示が行われることになる。事業等のリスクについては、リスクの記載順を重要性の判断を反映したものにすることが求められている。本研究で示したように事業等のリスクをパラグラフごとに区分すれば、トピックの推定と同時に、リスクの重要度をどのように判断しているかを分析する手掛かりとなるだろう。このように開示府令の改正によって、事業等のリスクの開示をこれまでとは異なる角度から分析することが可能になると考えられる。こうした分析についても今後取り組むことにしたい。

[参考文献]

Bao, Yang and Anindya Datta, Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures, *Management Science*, Vol.60 No.6, June 2014, pp. 1371-1391.

Brody, Samuel and Noemie Elhadad, *An Unsupervised Aspect-Sentiment Model for Online Reviews*, Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics20, 10, pp. 804-812.

岩田具治, 『トピックモデル』, 講談社, 2015 年 12 月。

奥村雅史, 「解説が進むテキスト情報と会計情報への示唆」, 『企業会計』, 第 70 巻第 7 号,

2018年7月, pp. 4-5。

金明哲, 『テキストアナリティクス』, 共立出版, 2018年8月。

金融庁企画市場局企業開示課, 『EDINET API仕様書 version 1.0』, 2019年3月。

金融庁金融審議会, 「金融審議会ディスクロージャーワーキング・グループ報告—資本市場における好循環の実現に向けて—」, 2018年6月28日, https://www.fsa.go.jp/singi/singi_kinyu/tosin/20180628/01.pdf

金融庁, 『記述情報の開示に関する原則』, 2019年3月19日,

<https://www.fsa.go.jp/news/30/singi/20190319/01.pdf>

金融庁, 『記述情報の開示に関する好事例集』, 2019年3月19日,

<https://www.fsa.go.jp/news/30/singi/20190319/02.pdf>

黒沼悦郎, 「ディスクロージャーワーキング・グループ報告について: 有価証券報告書の情報拡充を中心に」, 『資料版商事法務』, 第413号, 2018年8月, pp. 6-11。

小林一男, 「有価証券報告書における「事業等のリスク」等の開示実態調査(調査報告)について」, 『JICPA ジャーナル』, 第594号, 2005年1月, pp. 26-30。

財務会計基準機構編, 『有価証券報告書における『事業等のリスク』等の開示実態調査』, 財務会計基準機構, 2005年2月。

佐藤一誠, 『トピックモデルによる統計的潜在意味解析』, コロナ社, 2015年8月。

首藤昭信, 「会計学研究の展開と非財務情報の重要性」, 『週刊経営財務』, 第3392号, 2019年1月, pp. 10-19。

高村大也, 『言語処理のための機械学習入門』, コロナ社, 2010年8月。

土屋和之, 「事業等のリスクの分析—記載内容の類似度にもとづくアプローチ」, 『千葉商大論叢』, 第55巻第2号, 2018年3月, pp. 113-133。

中野貴之, 「財務諸表外情報の開示実態」, 山崎秀彦編『財務諸表外情報の開示と保証—ナラティブ・リポーティングの保証—』, 2010年10月, pp. 133-150。

野田健太郎, 『有価証券報告書における定性情報の分析と活用—リスク多様化にともなう望ましい対話のあり方—』, 経済経営研究, 第37巻第1号, 日本政策投資銀行設備投資研究所, 2016年5月。

張替一彰, 「有価証券報告書事業リスク情報を活用したリスク IR の定量評価」, 『証券アナリストジャーナル』, 第46巻第4号, 2008年4月, pp. 32-44。

藤岡由佳子, 前田和哉, 國分優子, 「「記述情報の開示に関する原則」および「記述情報の開示の好事例集」について」, 『企業会計』, 第71巻第7号, 2019年7月, pp. 81-95。

八木原栄二, 岡村健史, 堀内隼, 片岡素香, 「企業内容等の開示に関する内閣府令の改正」, 『企業会計』, 第71巻第5号, 2019年5月, pp. 89-100。

(2020.1.19 受稿, 2020.2.26 受理)

〔抄 録〕

本研究では有価証券報告書の事業等のリスクを、トピックモデルによってトピックを推定することで記載内容を分析した。特に事業等のリスク全体を1つの文書として分析するのではなく、事業等のリスクをリスク項目ごとにパラグラフに分割し、そのパラグラフを1つの文書として分析することで、おおむねトピックの推定が可能になったと考えられる。推定されたトピックからパラグラフのトピックを決定し、事業等のリスクに、どのようなトピックが記載されているかを明らかにすることができた。さらに、これによって、どのような業種で、どのようなリスクの記載割合が高いかを明らかにすることができた。