個人情報保護方針のクラスター分析による企業の取り組み状況の評価

奥原 雅之[1,a)]　藤本 正代[2]

概要：企業における情報セキュリティの取り組みの姿勢や，その成熟度を外部から知ることは一般に難しい．各企業が公開している個人情報保護方針などの文書は，その企業の取り組みを知る手がかりとなるべき重要な文書であるが，この内容から企業の姿勢を正確に知ることは，一般の消費者には困難である．本稿では，日本の地方銀行が公開する個人情報保護方針などの文書に対して，テキストマイニングの手法であるクラスター分析を用いることにより，それらの文書のグルーピングを行い，それぞれのグループ形成状況の分析を行う．さらに，それらのグループと各企業の特性を示す変数との関連性について調査する．

キーワード：セキュリティガバナンス，個人情報保護方針，クラスター分析

Cluster Analysis of Privacy Policies to Evaluate Company Initiatives

Masayuki Okuhara [1,a)]　Masayo Fujimoto[2]

Abstract: It is generally difficult to know the attitude and maturity of the information security efforts of a company from the outside. Documents such as privacy policy published by each company are important documents that should provide clues about the company's efforts. However, it is difficult for the average consumer to accurately know a company's stance from the contents of these documents. In this paper, we analyze documents such as privacy policies published by Japanese regional banks. We use cluster analysis, a text mining technique, to group these documents and analyze the formation of each group. Furthermore, we investigate the relationship between these groups and variables that indicate the characteristics of each company.

Keywords: security governanse, privacy policy, cluster analisys

1. はじめに

情報技術があらゆる業務の基盤となる現代社会において，情報システムの安定稼働は社会的要請の一つとなっている．情報システムの稼働を阻害する要因の一つとして，情報セキュリティインシデントの発生がある．営利，非営利にかかわらず，組織において情報セキュリティインシデントが発生すれば，情報の漏えい，情報の喪失，情報システムの停止などの被害が発生する．その被害は，インシデントを発生させた組織だけではなく，その組織の機能に依存する第三者（一般的には取引先や顧客など）にも及び，大規模なインシデントではその影響は一つの国の社会全体に波及する規模となる．このような背景もあり，情報セキュリティの確保を，企業の社会的責任（Corporate Social Responsibility, CSR）を構成する要素の一つとして捉える動きが生まれた（例えば [1]）．すなわち，企業が適切に情報セキュリティ対策を実施することに責任を負うことが，社会に対する企業の責任であるという概念である．

視点を変えて，企業外部のステークホルダーから見れば，

[1] 東京都立産業技術大学院大学
Advaned Institute of Industorial Technology, Shinagawa, Tokyo 140–0011, Japan
[2] 情報セキュリティ大学院大学
Institute of Information Technology, Yokohama, Kanagawa 211–8533, Japan
[a)] okuhara-m@aiit.ac.jp

ある企業が情報セキュリティに対してどのような取り組みを行っているかは，重大な関心事となる．今日の日々増大するセキュリティ脅威に対して，どのようなセキュリティ投資を行っているか，そしてそのセキュリティ投資は有効に機能しているのか，これらの評価によって第三者からの企業の価値は大きく変わるだろう．これは言葉を変えれば，その企業の情報セキュリティガバナンスの達成度（あるいは成熟度）を外から知ることである．しかしながら，外部からある企業の情報セキュリティガバナンスを知る手段は限られている．そのような取り組みの一つとして，経済産業省が提唱する報告書モデルに基づく情報セキュリティ報告書 [2] の，企業による自主的な発行がある．このモデルは 2007 年に制定されており，このモデルに基づく情報セキュリティ報告書を発行している企業も少なからず存在しているが，一般企業においてはこのような企業は今なお少数派である．

あるいは，外部団体による認証の取得状況も，企業の情報セキュリティガバナンスの状況を知る手がかりとなる．情報セキュリティマネジメントシステム（ISMS）の整備状況を対象とする，わが国における情報セキュリティマネジメントシステム適合性評価制度（一般には ISMS 認証，あるいは JIS Q 27000 認証として知られる）の認証取得企業数は，2022 年 8 月現在で 7,000 社を超えている [3]．また，個人情報保護のためのマネジメントシステムの整備状況を対象とする，プライバシーマーク制度（P マーク認証とも呼ばれる）の認証取得企業数は，同じく 2022 年 8 月現在で 17,000 社以上となっている [4]．これらの認定取得の有無は当然，企業の情報セキュリティガバナンスの状況を知るための有力な情報源となるが，認定取得の基準がベースラインに基づくものであり，ガバナンスの成熟度の段階を評価できないこと，認定取得企業が情報処理関連企業や一般消費者向けの B to C 企業に偏っていることなどから，情報セキュリティガバナンスの達成度の指標とするには課題も多い．

そこで，本稿では，企業が公式に開示している，個人情報保護に関連する文書の表現から，その企業の情報セキュリティへの取り組みを評価することを試みる．現代の企業は，情報セキュリティ方針や，個人情報保護方針など，多くの情報セキュリティに関連する文書を公表している．これらの文書は，法令などの公的な要請事項に基づくものもあり，その内容はある程度の相同性が認められるが，例えば個人情報保護宣言のように，その企業の独自性が強く発揮される文書もある．これらの文書は，その企業の経営者の見解や理念を始め，その企業の文化そのものを反映している可能性があり，その企業の取り組み姿勢，例えば情報セキュリティガバナンスの整備と何らかの相関を持つことも考えられる．

本稿では，わが国の地方銀行が公開している，個人情報保護関連の文書を対象とし，これらの文書における一般言語としての言葉の出現頻度と，その組織の特性に相関があるかどうかを，テキストマイニングの技法を用いて評価することを試みる．

本稿の構成は次の通りである．2 章では，本稿に関連する先行研究について述べる．3 章では，本稿で分析対象とする文書の定義，および本稿で使用するテキストマイニングの手法について述べる．4 章では，個人情報保護関連文書全体を分析した結果について述べる．5 章では，個人情報の利用目的の部分に対象を絞った分析について述べる．6 章では，定量データとしての個人情報開示手数料をを対象とした分析について述べる．7 章では，分析の総括と，今後の課題などについて述べる．

2. 関連研究

個人情報保護方針（プライバシーポリシー）そのものの分析を対象とした研究は，これまで多数行われている．McDonald らは，一般的な米国のインターネット利用者が，日常的に接する個人情報保護方針をすべて読んだ場合にかかる時間とコストの総量を推定している [5]．これによれば，全米のインターネット利用者がインターネット利用中に出会う個人情報保護方針を熟読すると仮定した場合，それに必要な時間の合計を機会損失として計算すると，年間約 7810 億ドルに相当する．Zaeem らは，データマイニングの手法を用いて個人情報保護方針の内容を分析する手法を提案し，それに基づいて個人情報保護方針の内容を自動的に要約するツールを提案した [6]．中村らは，完全性を失わなずにリスクの内容を要約することを目的とし，日本語で書かれたプライバシーポリシーを日本語形態素解析により分析した上で，機械学習によりリスク評価に必要なラベルを推測し，その精度を評価した [7]．

公開情報に基づいて企業のセキュリティへの取り組みを評価した研究としては，ISMS 認証の取得や CSO の設置など，外部から観測できる情報と，その企業におけるサイバーインシデントの発生頻度の相関をロジスティック回帰により分析したもの [8] がある．

3. 手法

3.1 方針

本稿では，各企業が公開している個人情報保護方針関連文書を収集し，それを分析することで，その企業の情報セキュリティや個人情報保護に対する姿勢を評価することを試みる．まず考えられる方法として，各文書に登場する，情報セキュリティリスクに関連するキーワードの出現回数を数え，それに基づいて分析対象文書のグルーピングを行うことが考えられる．しかし，この方法では，関連するキーワードの選出の時点で分析者の主観が入ってしまうため，グルーピングの結果の客観性が担保できないという問

題がある．そこで，本稿では，あえて文書に出現する言葉の意味の分析を排除し，純粋に自然言語としての言葉の出現頻度だけを分析するために，テキストマイニングの手法を適用することとした．テキストマイニングの技法の一つであるクラスター分析を使用することで，自然言語で記述された多数の文書を，その文書に登場する言葉の出現頻度に基づき，それぞれの文書の類似度を判定することができる．この技術を使うことにより，複数の企業で，ほぼ同じ内容の文書を開示しているようなケースを検出することができる．また，類似度によっていくつかのグループに明確に別れることがあれば，何らかの理由による「流派」がこれらの文書を開示している企業間に存在していることを示唆している．これによって得られたグループは，純粋に数値計算による結果であり，分析者の一切の主観を排除することができる．

3.2 分析対象

本稿では，日本国内の企業の例として，特に個人情報の保護に積極的に取り組んでいると考えられる，銀行業界を対象とした．日本の銀行は，都市銀行，信託銀行，外国銀行支店，地方銀行，第二地方銀行，その他の銀行などから構成されている [9]．このうち，支店数などが多く消費者が日々接する機会が多いものは都市銀行，地方銀行，第二地方銀行であるが，今回はサンプル数（銀行数）が適切な数となることを考慮し，地方銀行（62 行）を対象とすることとした．分析対象の文書は，各行が Web サイトで公開している個人情報の保護方針に関連する文書（以降「個人情報保護方針文書」と呼ぶ）とし，具体的には以下の 3 種を分析対象とした．

(1) 個人情報保護方針文書全体（4 章)
(2) 個人情報の利用目的（5 章）
(3) 個人情報開示手数料（6 章）

3.3 クラスター分析

本節では，本稿で使用するテキストマイニングおよびクラスター分析の手法について説明する．

3.3.1 日本語形態素解析

テキストマイニングは，テキストから知見を引き出す技術である [10]．テキストの要約などの処理をコンピュータに委ねようとするものがテキストマイニングであり，コンピュータに委ねることで，要約などの処理の再現性が保証される．テキストマイニングで最も重要な処理は，テキストを単語に分解することである．日本語テキストの場合は，英語などの西欧言語と異なり，単語の間にスペースなどの区切り文字を入れる慣習がないため，まず長大なテキストの文字列を単語として認識できる単位に分解することが必要となる．この処理のためには，日本語の文法知識が必要となる．例えば動詞や形容詞といった，活用する単語については，出現した単語がどの活用形かという判断をしないと，単語の切り分けを正しく行うことができない．例えば，「私は本を読んだ」という文は，「私」「は」「本」「を」「読んだ」ではなく，「私（名詞)」「は（助詞)」「本（名詞)」「を（助詞)」「読む（動詞)」「だ（助動詞)」と分解しなければならない．このように，単語の活用形などを考慮し，正しく単語に分解する技術を一般に形態素解析と呼ぶ．形態素解析は，その対象言語の文法に強く依存するため，言語ごとに解析のための解析器が存在する．日本語を対象とした形態素解析器としては，MeCab[11]，JUMAN[12]，JANOME[13] などが知られている．本稿では，R 言語環境で動作する MeCab の実装である，RMeCab[14] を形態素解析に利用した．

3.3.2 頻度分析

テキストマイニングでよく使われる手法の一つに，頻度分析がある．これは，複数の文書を対象に，文書内に出現する単語の頻度を測定し，その関係を調べるものである．この目的のために，一般には単語文書行列（Term-Document Matrix）を作成する．単語文書行列は，行方向に単語の出現頻度，列方向に文書を配置した二次元の行列である．単語文書行列で扱う出現頻度は，実際に各文書に出現した回数でもよいのだが，特に各文書間でその大きさの差が大きい場合には，その文書の大きさが分析の結果に影響が出ることがある．これを回避するために，出現頻度の正規化を行う．テキストマイニングの分野でよく使用される正規化方法として，TF と IDF が挙げられる．TF は Term Frequency の略で，文書中に出現する頻度そのものである．IDF は Inverse Document Frequency の略で，文書全体のうち，特定の単語がどれだけの文書に出現しているかを反映した指標である．IDF は，どの文書にも均等に出現する単語の重みを下げ，特定の文書だけに頻出する単語の重みを上げる．本稿では，IDF として RMeCab が採用している，以下の定義を用いる．

$$IDF = \log \frac{N}{n_i} + 1 \tag{1}$$

ここで，N は文書の総数，n_i は単語 i が出現している文書の数である．一般に，テキストマイニングでは TF と IDF それぞれの結果を掛け合わせ，それをコサイン正規化した以下の式で表される TF*IDF と呼ばれる重みを採用することが多い．ここで，m は単語の種類の数である．

$$\sqrt{\sum_{i=1}^{m} (TF \times IDF)^2} \tag{2}$$

3.3.3 クラスタリング

単語文書行列が作成できれば，それに従ってクラスター分析を行うことができる．クラスター分析では，各文書間の類似度を距離とみなして，その距離に基づいて文書をグループ化していく．距離の算出方法はいくつかあるが，本

稿では最も単純なユークリッド距離を使用する．これは，二つの文書間で単語の出現頻度の差を取り，それを二乗し合計したものの平方根を取るものである．すなわち，出現した単語のリストを位置ベクトルと見たときの，多次元空間における二点間の距離そのものと言ってよい．すべての二文書間の組み合わせがわかれば，それに基づいてクラスタリングが実施できる．クラスタリングの手法もいくつか提唱されているが，本稿では Ward 法を使用する．Ward 法では，集合 P と Q があるとき，

$$d(P,Q) = E(P \cup Q) - E(P) - E(Q) \qquad (3)$$

で定義される $d(P,Q)$ を，P と Q の距離とする．ここで，$E(A)$ は，A のすべての点から A の質量中心までの距離の二乗の総和である．Ward 法は，クラスタの各値からその質量中心までの距離を最小化するため，分類感度が高いとされている．

4. 個人情報保護方針の分析

一般に「個人情報保護方針」あるいは「プライバシーポリシー」と呼ばれる文書群は，Web サイトで一般に広く公開されるのが通常の姿になっている．例えば，日経 225 の通称で呼ばれる日経平均株価の対象企業 225 社（以降「日経 225 構成企業」と呼ぶ）は，すべて自社の Web サイトで個人情報保護方針文書を公開している．

これらの文書の構成は企業によって異なるが，以下のような要素から構成されるのが一般的である．

(1) 個人情報保護宣言
(2) 個人情報保護法[*1]が求める公表事項
  (a) 個人情報の利用目的（第 21 条）
  (b) 個人情報の第三者提供（第 27 条）
  (c) 個人情報の開示，訂正，利用停止（第 33 条から第 35 条まで）
(3) クッキーポリシー（Web サイトのプライバシーポリシー）

このうち「個人情報保護宣言」は，通常個人情報保護方針文書の先頭に置かれるもので，その企業の個人情報に対する保護の基本的な方針を記述している．また「クッキーポリシー」は，その Web サイトで利用者がどのように追跡されるかなどの情報が記載されてている部分であり，Cookie の取扱いがその代表例である．

本稿においては，テキストマイニングの対象となる文書の範囲を厳密に統一しないと，結果の有意性が保証できない．しかし，実際の個人情報保護対象方針文書構成は，企業によって多種多様であり，その境界を厳密に定義することは難しい．例えば「個人情報保護方針」のような単独の Web ページがサイト上に用意されていれば，その切り出しは容易だが，コンテンツごとに細かくページが分かれていたり，個人情報以外の利用者への告知情報と混在しているケースでは，どこまでを分析対象とするか，どうしてもあいまいな領域が残ってしまう．そこで今回の分析では，以下に述べるルールで文書の切り出しを行った．個人情報保護方針に関する記述が 1 ページにまとまっている場合は，原則としてページ全体を対象の文書として切り出した．ただし，ヘッダーやフッター，サイドメニュー，パンくずなど，Web ページの構成部品となる部分は除外した．また，個人情報保護方針が複数のページに分割されている場合は，個人情報保護に関するトップページからのリンクに従い，個人情報保護関連と認められるページを巡回して対象文書としてキャプチャーした．

[*1] 平成十五年法律第五十七号 個人情報の保護に関する法律

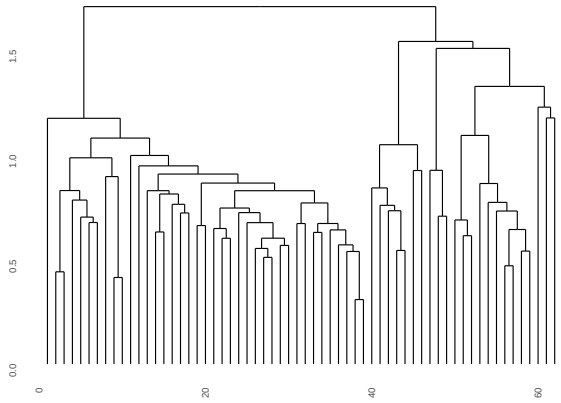

図 1 地方銀行の個人情報保護方針のデンドログラム
Fig. 1 Dendrogram for privacy policies of local banks

この方針に従い，地方銀行 62 行の個人情報保護方針を 2021 年 7 月 13 日から 8 月 1 日の期間に収集した．クラスタリングの結果のデンドログラムを**図 1** に示す．この結果から，以下のことがわかる．

- 構成がまったく同じという文書は存在しなかった．すなわち，いわゆる「コピー＆ペースト」で流用したようなコンテンツは，今回の銀行間では見られなかった．
- 各銀行の文書は，大きくは 2 種類のクラスターに別れているが，それよりも下位においてはそれほど明確なクラスターを形成していない．

ここで，全体を二つのクラスターに分割する．図 1 の右のクラスターを Cluster 1，左を Cluster 2 とする．Cluster 1 に属する銀行数は 23 行，Cluster 2 は 39 行である．地方銀行は地域ごとに所轄する財務局が定められている．この所轄財務局とクラスター分布の関係を**表 1** に示す．この表から，特に関東，中国・四国，沖縄で Cluster 2 の比率が高いことが伺える．

このように，ある程度の傾向のようなものは見て取ることができるが，明確な企業（銀行）のグループの形成や，顕著な特性などは観測することができなかった．この一つの理由として，切り出している個人情報保護関連文書の自由度が大きいため，切り出し方にどうしても主観が混入し

表 1 クラスター分析結果の地域分布
Table 1 Region distribution of cluster anaysis

cluster ID	1	2
北海道財務局	1	0
東北財務局	6	4
関東財務局	2	10
東海財務局	5	2
北陸財務局	1	3
近畿財務局	4	3
中国財務局	0	5
四国財務局	0	4
福岡財務局	2	4
九州財務局	2	2
沖縄総合事務局	0	2

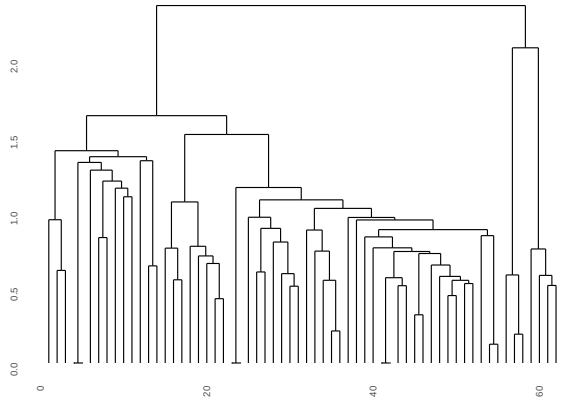

図 2 地方銀行の個人情報利用目的のデンドログラム
Fig. 2 Dendrogram for privacy data purpose of local banks

てしまい,「同じものを比較する」という点において十分な精度が得られていない可能性があげられる．この点から，次章ではより区切りが明確な，個人情報保護文書の一部分を分析対象とすることを試みることとした．

5. 個人情報利用目的の分析

個人情報保護の制度において，個人情報の利用目的は，重要な意味を持つ項目である．しかしながら，日本における個人情報の利用目的の記述は，一般に欧米に比べて具体性を欠く傾向があるとの指摘もある [15]．この点を鑑み，個人情報保護方針文書から，特に個人情報の利用目的を表記した部分のみを切り出し，前章と同様にクラスター分析を行った．この結果のデンドログラムを**図 2** に示す．

この図からは，以下のことが読み取れる．

- 前回の試行と異なり，まったく構成が等しいと思われる銀行のペアが 3 組あった．これらについて調べたところ，いずれも該当する銀行同士が同じ持ち株会社（フィナンシャルグループ）などに属しているケースであった．この点から，持ち株会社による企業グループ間では個人情報保護方針の内容を共通化する取り組みが進んでいることが推察される．
- 全体的になめらかなグラデーションのような形を示しており，明確なクラスターを形成していない．

表 2 利用目的のクラスターと ROE の関係
Table 2 Relationship between cluster of purpose and ROEs

cluster ID	N	mean	sd
1	3	3.20	0.36
2	33	2.80	1.66
3	12	2.90	0.81
4	8	2.84	1.13
5	3	2.70	1.35

このデンドログラムに基づき，5 個のクラスターとなるようにクラスター分割を行った．さらに，これらのクラスターと企業の業績との関連の有無を確認するため，各クラスター構成銀行の ROE（自己資本利益率）の平均値を求めた．ここで ROE を指標としたのは，一般的に企業業績を評価するために用いられる指標の一つであり，かつ企業規模によらない性質があるためである．なお，ROE の数値は東洋経済新聞社の会社四季報 2021 秋版に依った．各クラスターの構成企業数，ROE の平均と標準偏差を表**表 2** に示す．ROE の平均値については，多少のばらつきはあるものの，t 検定では有意性は見られなかった．

6. 個人情報開示手数料の分析

本章では，定量的な指標として，個人情報開示手数料について分析する．個人情報保護法第 33 条の規定により，個人は自身の情報の開示を事業者に求めることができるが，同 38 条の規定により，事業者は情報開示のための手数料を請求することが認められている．今回分析対象とした 62 行の地方銀行は，いずれも情報開示のための手続きを公開している．このうち，Web サイト上で情報開示手数料の情報を見つけることができなかった 2 行を除外し，残り 60 行について個人情報開示のための手数料の分布を調査した．ここで，手数料については消費税を含む金額で統一している．また，内容によって異なる手数料を規定している銀行については，一番基本となる情報開示のための手数料を対象とした．各行の手数料は，440 円から 3300 円（いずれも税込み）の範囲に分布した．一番頻度が高かった金額は 1100 円であった．分布のヒストグラムを**図 3** に示す．

手数料の金額と，その銀行の業績に関連があるかを調べるために，手数料と ROE の相関を調べた結果を**図 4** に示す．図から見てわかるように，手数料と ROE にはほぼ相関が見られない．相関係数の分析も，有意とはならなかった．つまり，業績がよい，あるいは悪いという状況は，手数料にはあまり影響していないという状況を見て取ることができる．

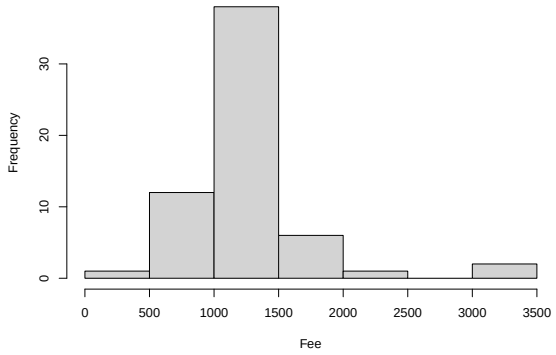

図 3 個人情報開示手数料の分布
Fig. 3 Histgram of diclosing fee

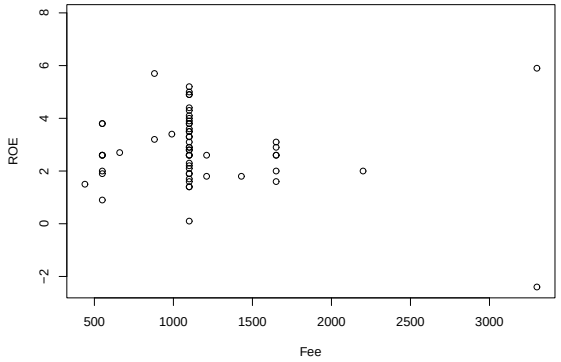

図 4 個人情報開示手数料と ROE の関係
Fig. 4 Relation between disclosing fee and ROE

7. おわりに

本稿では，わが国の地方銀行が公開している，個人情報保護関連文書を対象とし，これらの文書における一般言語としての言葉の出現頻度と，その組織の特性に相関があるかどうかを，テキストマイニングの技法を用いて評価することを試みた．得られた事実として，まず各企業の個人情報保護関連文書が，明確なグループを形成しないということである．分析を行う前の段階では，個人情報保護宣言は，企業グループや，業種などの属性に従い，ある程度の明確なグループを形成することを期待していたが，実際はそのようなことはなく，収集された個人情報保護宣言は，比較的連続した関連性を示した．このことは，各企業はそれぞれに独自性のある個人情報保護宣言文書を独自に作成していることを示しており，特定のサンプル文書をほぼそのまま流用するような，いわゆるコピーペーストな作成の仕方は行われていないことを示唆している．また，今回の分析では，文書によって分類された銀行のクラスターと，ROE などの指標との有意な相関は見られなかった．今後はデータの蓄積量が多い他の指標との関連を調べる，あるいは個人情報保護関連文書以外の企業が公開している文書を分析対象にするなどして，何らかの形での相関が得られる手法が確立できないかを調査したい．

参考文献

[1] 大木栄一郎：情報セキュリティは CSR である，ITMedia エンタープライズ（オンライン），入手先 ⟨https://www.itmedia.co.jp/im/articles/0603/10/news126.html⟩（参照 2021-11-15）．

[2] 経済産業省：情報セキュリティ報告書モデル（改訂版），METI（オンライン），入手先 ⟨https://cio.go.jp/node/2121⟩（参照 2021-11-15）．

[3] 情報マネジメントシステム認証センター：ISMS 認証取得組織検索，ISMS-AC（オンライン），入手先 ⟨https://isms.jp/lst/ind/⟩（参照 2022-8-19）．

[4] 一般財団法人日本情報経済社会推進協会：プライバシーマーク付与事業者検索，JIPDEC（オンライン），入手先 ⟨https://entity-search.jipdec.or.jp/pmark⟩（参照 2022-8-19）．

[5] McDonald, A. M. and Cranor, L. F.: The Cost of Reading Privacy Policy, *A Journal of Law and Policy for the Information Society*, Vol. 4, pp. 543–568 (2008).

[6] Zaeem, R. N., German, R. L. and Barber, K. S.: PrivacyCheck: Automatic Summarization of Privacy Policies Using Data Mining, *ACM Transactions on Internet Technology*, Vol. 18, No. 4, p. 53 (2018).

[7] 中村 徹，ウェルデルファエル B. テスファイ，バネッサブラカモンテ，清本晋作，鈴木信雄：日本語のプライバシポリシに対する完全性を考慮したリスク要約手法の評価，情報処理学会論文誌，Vol. 62, No. 1, pp. 332–345 (2021).

[8] 山田道洋，池上和輝，菊池浩明，乾 孝治：セキュリティマネジメントによるサイバーインシデントリスク削減の評価，情報処理学会論文誌，Vol. 61, No. 12, pp. 1781–1791 (2020).

[9] 金融庁：銀行免許一覧（都市銀行・信託銀行・その他），FSA（オンライン），入手先 ⟨https://www.fsa.go.jp/menkyo/menkyoj/ginkou.pdf⟩（参照 2022-8-19）．

[10] 石田基弘：R によるテキストマイニング入門，森北出版株式会社 (2017).

[11] 京都大学情報学研究科－日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクト：MeCab: Yet Another Part-of-Speech and Morphological Analyzer, github (online), available from ⟨http://taku910.github.io/mecab/⟩ (accessed 2022-03-28).

[12] 黒橋・楾・村脇研究室：日本語形態素解析システム JUMAN，京都大学（オンライン），入手先 ⟨https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN⟩（参照 2022-03-28）．

[13] 打田智子：Janome v0.4 documentation (ja), github (online), available from ⟨https://mocobeta.github.io/janome/⟩ (accessed 2022-03-28).

[14] 石田基弘：アールメカブ，rmecab.jp（オンライン），入手先 ⟨http://rmecab.jp/wiki/index.php?RMeCab⟩（参照 2022-08-21）．

[15] 高木浩光：高木浩光さんに訊く、個人データ保護の真髄 - いま解き明かされる半世紀の経緯と混乱，JILIS（オンライン），入手先 ⟨https://cafe.jilis.org/2022/03/18/160/⟩（参照 2022-8-20）．