情報処理学会論文誌 Vol.60 No.12 2196–2210 (Dec. 2019)

情報セキュリティに関連するガイドラインの Cybersecurity Framework に基づいた文書内容の可視化手法の提案とその評価

尾崎 敏司[1,a)]

受付日 2019年3月12日, 採録日 2019年9月11日

概要：2012 年に独立法人情報処理推進機構により提示された「情報セキュリティ人材の育成に関する基礎調査」と 2014 年のその追加調査によると，約 8.1 万人の情報セキュリティ人材の不足が指摘されており，現在もその育成は課題となり続けている．自己学習や実業務の補助になると考えられるガイドラインは多く公開されているが，これらのガイドラインがセキュリティ業務のどの部分に該当するのかを学習者が把握することは難しい．そこで本研究では，学習者の体制化方略を補助する文書評価の枠組みを検討することを目的として，米国国立標準技術研究所の公開している Cybersecurity Framework をもとに，tf-idf による特徴語のベクトルを用いてガイドラインの内容の可視化を行う手法の提案を行い，提案手法の妥当性と有利性の観点で評価を行った．情報セキュリティに関する 4 つのガイドラインに対して提案手法を適用して得た結果と，質的データ分析のテンプレートコーディングを実施して得た結果をコサイン類似度とピアソンの積率相関係数で比較したところ，フレームワークコアの機能で見た場合コサイン類似度の平均 0.907，カテゴリで見た場合平均 0.761 となり，相関係数も機能では強い正の相関を示し，カテゴリでも正の相関を示した．これにより提案手法の分析結果の妥当性について確認できた．また，4 つのガイドラインに対して tf-idf と k 平均法によるクラスタリングとトピック分析を行った結果と Cybersecurity Framework のテキストマイニングの結果を比較することで，体制化方略に求められる要求を満たすという観点で，枠組みを用いることに有利性があることを確認した．

キーワード：セキュリティ，自己調整学習，体制化，テキストマイニング，可視化，tf-idf，Cybersecurity Framework

Proposal to Visualize a Content of Information Security Guideline Based on Cybersecurity Framework

Satoshi Ozaki[1,a)]

Received: March 12, 2019, Accepted: September 11, 2019

Abstract: According to the “Basic Survey on the Training of Information Security Human Resources” presented by the Information-technology Promotion Agency in 2012 and its additional survey in 2014, the shortage of information security human resource is estimated about 81,000 people and training for information security human resource have been a problem. Many guidelines are published and helpful for self-study and actual work, but it is difficult for the learner to grasp which part of the security activity can be learned by them. In order to create the document evaluation procedure to enforce learner’s organizational strategy, we proposed the method to visualize contents based on the Cybersecurity Framework published by the National Institute of Standards and Technology with the vector of features extracted by tf-idf. We evaluated the accuracy of the visualization result and considered the advantage of applying Cybersecurity Framework in text mining. With cosine similarity and Pearson product-moment correlation coefficient, we compare the results obtained by applying the proposed method to the four guidelines on information security and the results obtained by the template coding of qualitative data analysis. The average of cosine similarity was 0.907 at “functions” of the framework core and was 0.761 at the “categories”. The correlation coefficient of “functions” are positive-correlated strongly and that of “category” are positive-correlated. This reveals the validity of the proposed method. We also compare the results of two text mining procedure (k-means with tf-idf and topic analysis with LDA) to 4 guidelines and the results of text mining to Cybersecurity Framework and found the advantage that using Cybersecurity Framework in text analysis provides systematic background logic and framework to analysis results, which is required for organizational strategy.

Keywords: security, self-regulated learning, organizational strategy, text mining, visualization, tf-idf, cybersecurity framework, k-means clustering, topic analysis

1. はじめに

2012 年に独立法人情報処理推進機構（IPA）により提示された「情報セキュリティ人材の育成に関する基礎調査」[1] と 2014 年に行われた追加分析 [2] によると，約 8.1 万人の情報セキュリティ人材の不足が指摘されており，現在もその育成は課題となり続けている．また，内閣サイバーセキュリティセンターのサイバーセキュリティ人材の育成に関する施策連携ワーキンググループが結成されており，2018 年に「サイバーセキュリティ人材の育成に関する施策間連携ワーキンググループ報告書」[3] が作成されている．この報告書では，セキュリティの専門家であるスペシャリストと，一般的な社内の IT オペレーションを実施しているゼネラリストの間に，エキスパートと呼ばれる「自社事業とセキュリティ活動をよく知り，現場と経営をつなぐ人材」の必要性が指摘されており，引き続き企業における人材育成の必要性が求められていることがうかがえる．

前述の「情報セキュリティ人材の育成に関する基礎調査」の追加分析によると，約 8.1 万人の情報セキュリティ人材の不足のうち，現在セキュリティ人材を保持していない企業において新たに必要とされる人数は 6.1 万人と推計されている．同時期に情報セキュリティ大学院大学により行われた「情報セキュリティ事故対応に関わるアンケート調査」[4] の結果においても，無回答層を含めた場合，中小企業における約 75%がセキュリティ担当者を保持していない可能性が示唆されており，担当者をおいている場合でも約 41%が兼任の担当者 1 名のみの状態であった．トレンドマイクロ株式会社が 2018 年 9 月に発行した「法人組織におけるセキュリティ実態調査 2017 年版」[5] においては，従業員規模とセキュリティ対策の包括度に相関関係があることが指摘されており，特に，中小企業において引き続き限られた人材・資源の中でセキュリティ対策を実施していくことが必要になると考えられる．

実業務に基づいた自律的な学習の教材となりうる運用者向けのガイドラインは多く公開されており，経済産業省で整理されているものに限っても 150 を超える [6]．これらのガイドラインは 30 種程度に分類はされているものの，その項目は体系立てられたものになっておらず，そのガイドラインがセキュリティ対策活動のどの部分に該当するかを把握することは難しい．これにより学習者は，適切なガイドラインを選択できず，また，学習をどのように進めるのかの計画（学習方略）を立てにくくなる．

この問題を軽減するため，本研究では，エキスパート人材を目指す学習者の学習の補助を行うための文書評価の枠組みについて検討することを目的とした．本研究では，Cybersecurity Framework 1.1 [7] を文書評価の枠組みとして用い，実際に Cybersecurity Framework に基づいたテキストマイニングを行うことでガイドラインの文書内容を可視化する手法の提案を行い，下記の 2 つの観点で検証を行った．

1) 提案手法による結果と質的コーディングの結果の比較を行うことで，Cybersecurity Framework に基づいたテキストマイニングの結果の妥当性を検証した．
2) 提案手法による結果と，枠組みを前提としないクラスタリングやトピック分析の結果との比較を行うことで Cybersecurity Framework を文書評価の枠組みとして用いる有利性について確認した．

枠組みに基づいた文書評価を参照することで，学習者は，自身がセキュリティ対策のどの部分に該当する分野を学習しているかを把握し，また，文書どうしでの比較を行い関連付けて理解することができるため，学習効率が高まることが期待できる．

本提案手法は，第 148 回「コンピュータと教育研究発表会」の研究論文セッションで行った「情報セキュリティに関連するガイドラインの内容提示の手法の提案とその評価」の研究報告 [8] と同一である．ただし，本論文では，下記の点を更新している．

1) 2019 年 1 月に発行された「重要インフラのサイバーセキュリティを改善するためのフレームワーク 1.1 版」[9] に基づき再解析・再評価を行った．
2) 解析・評価の対象とする文書数を増やした．
3) プログラムのバグの修正を行った．
4) 有利性の確認のため，tf-idf と k 平均法を用いたクラスタリングと，LDA（Latent Dirichlet Allocation）[10] によるトピック分析を実施した．

2. 関連研究

この章では，セキュリティ人材育成に関する学習方法の近年の研究について確認し，その後，自己調整学習についての先行研究について述べる．最後に，テキストマイニングとしての本研究の立ち位置について述べる．

2.1 セキュリティ人材の育成に関する先行研究

セキュリティ人材の教育手法の研究としては，近年では，CTF（Capture the Flag）によるアプローチに関した研究が多く見られる [11], [12], [13]．また，攻撃手法の学習上課題となることが多い演習環境に注目したものも多い [14], [15]．

これらの教育手法は，技術的な側面での学習補助として有用であると考えられる．しかし，エキスパートつまり「自社事業とセキュリティ活動をよく知り，現場と経営をつなぐ人材」という観点では，技術に限らない広い視点での学習活動が求められている．これを実現するためには，

[1] 筑波大学
University of Tsukuba, Tsukuba, Ibaraki 305–8577, Japan
[a)] s1730150@s.tsukuba.ac.jp

学習者の主体的な学習を補助するアプローチが必要になると考えられる．

エキスパートの育成を目的とした研究では，孫らによる大学・大学院のカリキュラムに対する研究があげられる [16]．孫らは，アメリカ国立標準技術研究所（NIST）の下に設置されている NICE（The National Initiative For Cybersecurity Education）が定義した Cybersecurity Workforce Framework [17] に基づいて大学と大学院におけるセキュリティ教育課程カリキュラムの分析を行った．この研究では，Cybersecurity Workforce Framework の 783 個の技術能力項目を 62 種類の項目に集約している．この 62 の項目について，a) Cybersecurity Workforce Framework 内の単語出現数でつけた 62 項目の順位と，b) 大学・大学院カリキュラムと 62 項目の対応付けを行い科目の数でつけた順位の 2 つの順位を作成し a) と b) の間の Spearman 順位相関係数を用いて，大学・大学院カリキュラムの妥当性を検証している．

この研究は，大学・大学院教育を対象に，カリキュラム開発の要求分析を目的として，Cybersecurity Workforce Framework に基づいた項目を単語の出現数や科目数で順位づけて比較を行い，教育課程全体の妥当性を評価している．しかしながら，本研究では，主に社会人の学習者に利用されるガイドラインを対象に，学習者の主体的な学習の補助や実務実施の補助を目的として，NIST の発行している Cybersecurity Framework 1.1 に基づいてテキストマイニングの一種である tf-idf（Term Frequency-Inverse Document Frequency）を用いて分析行い，文書ごとの可視化を行った．

2.2 自己調整学習に関する研究

自己調整学習とは，1990 年代からアメリカの教育心理者 Zimmerman らが中心となって提案している教育心理学の理論体系で，学習者の主体的な学習方略を重要視している．学習方略とは，学習に取り組む戦略のことで，自己調整学習では，学習方略を大きく，認知的方略（例：関連付けて覚える），メタ認知的方略（例：勉強時間と学習範囲を記録する），動機付け方略（例：学習の目的を書き出す）に分けている [18]．

このうち認知的方略に含まれる体制化方略と図示化方略は，国内では，松沼により英語の現在完了形の学習において，実験的に学習効果の向上が確認されている [19]．体制化方略とは，「何らかの理論や枠組みによって学習要素を相互に関連付けて整理する方法」であり，図示化方略とは，文字どおり「図示により整理」を行う方法である．

本研究では，他分野ではあるが先行研究で効果が確認されている体制化方略につながる形での情報提示を検討した．

2.3 テキストマイニングに関する研究

セキュリティ関連の文書の内容について包括的な検討を行っているものでは，暗号化 API のユーザビリティについて，API の解説文書の内容の不備を指摘した Acar らの研究が有名である [20]．しかし，これは暗号化 API という限られた分野で人手による評価を行っており，複数の分野にまたがる 150 を超えるガイドラインについて同様の評価方法をとるのは難しいと考えられる．

そこで本研究ではテキストマイニングによる手法を提案する．Gupta らによると，テキストマイニングの利用目的は大きく，分類，クラスタリング，概念リンク，可視化，要約，情報抽出，質疑応答，トピック追跡などに分けられるといわれており，本研究で行う取り組みは，文書の可視化や要約に該当すると考えられる [21]．

可視化や要約の取り組みでは，文書の内容・構造などを仮定せずに対象の文書を目的に合致する形式で再構成し可視化を行う方針と，文書の内容・構造を事前に仮定して，それに従って解析・再構成を行う方針が存在すると考えられる．前者の方針は多く研究がなされており，近年でも，Park らによる対話的に語彙を入力させて文書の内容を可視化する手法が研究されている [22]．一方，後者の方針による研究事例は少なく，たとえば，2006 年の赤石による物語構造に基づき動的に連想される情報を提示するフレームワークの研究 [23] をあげることができる．しかしながら，特に情報セキュリティ関連の文書について，フレームワークに基づいて可視化を試みたものは存在していない．

情報セキュリティ教育の分野で，事前に内容や構造を仮定して解析する利点として，下記の 3 つをあげることができる．

1) この分野では，すでにいくつかのフレームワーク（モデル，規準）が存在している．
2) これらのフレームワークは，分野についての体制化された情報として考えることができ，この情報に基づいて文書の可視化を行うことで体制化方略につながる情報提示を容易に行えると考えられる．
3) 既往のフレームワークを用いることで，文書上に記述がない分野がある場合に，その分野について記載がないことを表現することができると考えられる．

3. 本研究の目的

そこで本研究では，セキュリティ関連のガイドラインに関して学習者の体制化方略を補助することを目指し，文書評価に Cybersecurity Framework の枠組みを使うことの有利性と評価結果の妥当性を検討することを目的として，実際に Cybersecurity Framework を用いた可視化を試み，下記の 2 つの観点からその評価を行う．

1) 質的コーディングの結果との比較を行い，Cybersecurity Framework に基づいたテキストマイニングの結果の妥当性

を検証する．

2) 枠組みを前提としないクラスタリングやトピック分析の分析結果と Cybersecurity Framework に対する比較を行う．これにより，枠組みを前提としないテキストマイニングの方法では，体制化方略を促すための最低限の要求を満たすことができない可能性があることを確認し，枠組みに従った可視化を行う提案手法の有利性を確認する．

体制化方略は，理論や枠組みに基づいて学習要素を相互に関連付けて整理する学習方略である．本手法ではこれを補助することを目的とするため，可視化結果は，体系だった理論や枠組みに基づいて学習要素が相互に関連付けられ整理された形で提示されることが要求される．

Cybersecurity Framework を文書評価に用いることが，実際の体制化方略を通じた学習効果に，どれほど影響を及ぼすかについての検証は本研究では扱わず今後の課題とする．

4. 提案手法

本研究では，Cybersecurity Framework 1.1 の「フレームワークコア」と呼ばれるモデルを基に，文書内容の提示を行う．フレームワークコアは，Cybersecurity Framework の一部で，サイバーセキュリティ対策と期待される効果について体系的にまとめたモデルである．このモデルに基づいて内容の可視化を行うことにより，学習者は，自分がセキュリティ対策活動の全体像の中のどの部分について学習したのか把握し文書どうしの内容の比較を行えるため，学習内容の体制化を行いやすくなると考えられる．

まず，Cybersecurity Framework についての説明を行い，次に提案する可視化手法について説明する．

4.1 Cybersecurity Framework

このフレームワークは，重要インフラストラクチャにおけるセキュリティ対策向けに作成されており，現在，産業界で効力を発揮している標準，ガイドライン，およびベストプラクティスを集約することで，現在ある多様なセキュリティ対策を体系化・構造化し，企業に示している．

このフレームワークで提示されているフレームワークコアは，機能，カテゴリ，サブカテゴリ，参考情報の 4 つで構成されている（**図 1**）．機能は，基本的なサイバーセキュリティ対策の最も上位の構成要素として「特定」,「防御」,「検知」,「対応」,「復旧」の 5 つが定義されている（**図 2**）．カテゴリは，機能をさらにセキュリティの効果によって分類したものであり，サブカテゴリは，さらに具体的な対策に分類したものである．参考情報は，各サブカテゴリについて期待される成果を達成するための，既存の標準・ガイドライン・ベストプラクティスについてまとめたものである．ただし，参考情報はあくまで例であり包括的なものではない．

機能	カテゴリー	サブカテゴリー	参考情報
識別（ID）	**資産管理（ID.AM）**: 自組織が事業目的を達成することを可能にするデータ、人員、デバイス、システム、施設が、識別され、組織の目的と自組織のリスク戦略における相対的な重要性に応じて管理されている。	ID.AM-1: 自組織内の物理デバイスとシステムが、目録作成されている。	CIS CSC 1 COBIT 5 BAI09.01, BAI09.02 ISA 62443-2-1:2009 4.2.3.4 ISA 62443-3-3:2013 SR 7.8 ISO/IEC 27001:2013 A.8.1.1, A.8.1.2 NIST SP 800-53 Rev. 4 CM-8, PM-5
		ID.AM-2: 自組織内のソフトウェアプラット	CIS CSC 2

図 1 重要インフラのサイバーセキュリティを改善するためのフレームワーク 1.1 版の「表 2 フレームワークコア」より部分的に引用

Fig. 1 Adapted from “Table 2 Cybersecurity Framework Core” in “Framework for Improving Critical Infrastructure Cybersecurity”.

機能の識別子	機能	カテゴリーの識別子	カテゴリー
ID	識別	ID.AM	資産管理
		ID.BE	ビジネス環境
		ID.GV	ガバナンス
		ID.RA	リスクアセスメント
		ID.RM	リスクマネジメント戦略
		ID.SC	サプライチェーンリスクマネジメント
PR	防御	PR.AC	アイデンティティ管理とアクセス制御
		PR.AT	意識向上およびトレーニング
		PR.DS	データセキュリティ
		PR.IP	情報を保護するためのプロセスおよび手順
		PR.MA	保守
		PR.PT	保護技術
DE	検知	DE.AE	異常とイベント
		DE.CM	セキュリティの継続的なモニタリング
		DE.DP	検知プロセス
RS	対応	RS.RP	対応計画の作成
		RS.CO	コミュニケーション
		RS.AN	分析
		RS.MI	低減
		RS.IM	改善
RC	復旧	RC.RP	復旧計画の作成
		RC.IM	改善
		RC.CO	コミュニケーション

図 2 重要インフラのサイバーセキュリティを改善するためのフレームワーク 1.1 版の「表 1 機能とカテゴリーの識別子」を引用

Fig. 2 Adapted from “Table 1 Function and Category Unique Identifiers” in “Framework for Improving Critical Infrastructure Cybersecurity”.

先行研究 [16] の Cybersecurity Workforce Framework はセキュリティ対策業務の役割と必要なスキルセットや能力について整理しているのに対して，Cybersecurity Framework は，セキュリティ対策で用いられる手法を整理している．本研究で，Cybersecurity Framework を用いた理由としては，

1) ガイドラインでは手法・手順が多く記載されることが予想されたため，役割ではなく，手法の観点で整理を行ったほうが適切な体制化が行えると考えた．
2) IPA より Cybersecurity Framework 1.1 の翻訳版である「重要インフラのサイバーセキュリティを改善するためのフレームワーク」が発行されていた．

ことがあげられる．

本研究では日本語文書を対象とするため，この「重要インフラのサイバーセキュリティを改善するためのフレーム

ワーク」を基に解析を行っている．

4.2　フレームワークコアに基づく可視化手法（提案手法）

Cybersecurity Framework 1.1 のフレームワークコアのカテゴリに基づいて，文書中の単語の重要度を評価する手法である tf-idf により特徴語ベクトルを作成し，解析対象の文章の各センテンスとのコサイン類似度で類似度を測定することで，フレームワークコアに基づいた内容の推定を行った．

解析対象の文書中のある行 L_j が，フレームワークコアのあるカテゴリ C_i にどの程度関連しているか（つまり内容が類似しているか）は，Cybersecurity Framework 1.1 の記述を基に作成したカテゴリ C_i の特徴語ベクトル $\boldsymbol{c_i}$ と，ある行 L_j に対して Cybersecurity Framework 1.1 の統計情報を基に作成した特徴語ベクトル $\boldsymbol{l_j}$ のコサイン類似度で記載することができる．文章全体の中にカテゴリ C_i に関連する記述がどの程度あるかを表すスコア $S_i(C_i)$ は，$\boldsymbol{l_j}$ と $\boldsymbol{c_i}$ のコサイン類似度の総和となるので，式 (1) で評価することができると考えられる．

$$S_i(C_i) = \sum_j \frac{\vec{l_j} \cdot \vec{c_i}}{|\vec{l_j}| \cdot |\vec{c_i}|} \tag{1}$$

具体的には，下記の手順でスコア $S_i(C_i)$ の計算を行った．手順を図示したものを**図 3** として示す．

1. 「重要インフラのサイバーセキュリティを改善するためのフレームワーク」内で各カテゴリ C_i について記述されている部分を実際に読んで確認し，カテゴリごとに抽出した．また，カテゴリ C_i が属している機能に関する記述も同様に実際に読み抽出し，カテゴリ C_i の文書の一部として取り扱った（図 3 ① を参照）．具体的には，「重要インフラのサイバーセキュリティを改善するためのフレームワーク」の本文 2.1 節や付録 A などの各機能，カテゴリについての説明を用いている．
2. 抽出した各カテゴリと機能の文書集合に対して，Mecab [24] を用いて標準のシステム辞書で分かち書きと形態素解析を実施して名詞を取り出し，名詞だけの集合に変換した（図 3 ① を参照）．
3. 名詞句による文書集合に対して，それぞれのカテゴリ C_i ごとに，tf-idf による特徴語ベクトル $\boldsymbol{c_i}$ を作成した．また，同時に「重要インフラのサイバーセキュリティを改善するためのフレームワーク」の語彙と文書頻度を得た（図 3 ① を参照）．
4. 適用対象の文章から 1 行ごと文字列を抜き出し「重要インフラのサイバーセキュリティを改善するためのフ

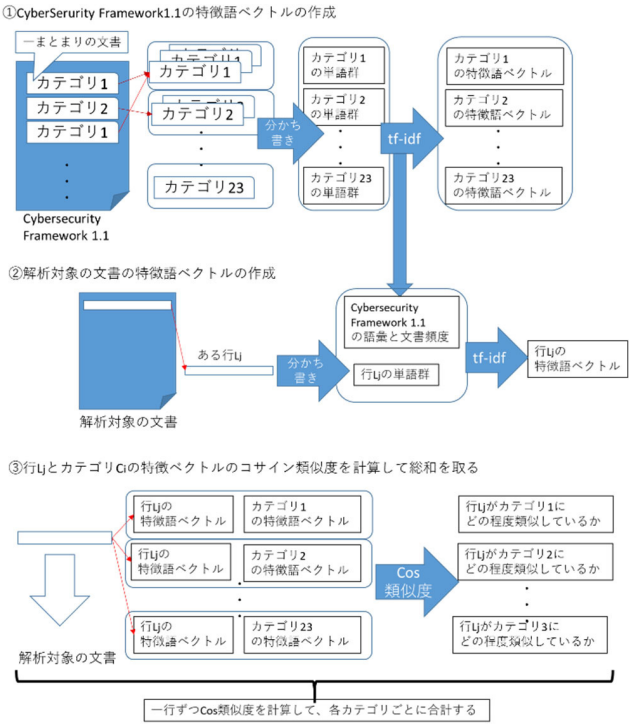


図 3　提案手法の手順

Fig. 3　Schematic diagram of the proposal.

表 1　カテゴリごとに抽出された特徴語上位 10 個

Table 1　Top 10 words extracted for each category by tf-idf.

機能	カテゴリ	特徴語上位10個　()はtf-idfの計算結果
IDENTIFY（特定）	資産管理	管理(0.367)、リスク(0.298)、ビジネス(0.281)、事業(0.250)、特定(0.236)、戦略(0.223)、組織(0.193)、資産(0.183)、サイバーセキュリティリスク(0.177)、理解(0.157)
	ビジネス環境	ビジネス(0.351)、管理(0.302)、リスク(0.295)、上(0.264)、理解(0.233)、順位(0.193)、優先(0.193)、サイバーセキュリティリスク(0.175)、付け(0.175)、組織(0.157)
	ガバナンス	管理(0.398)、リスク(0.317)、サイバーセキュリティリスク(0.289)、ビジネス(0.267)、理解(0.256)、上(0.209)、ガバナンス(0.200)、組織(0.199)、特定(0.158)、環境(0.149)
	リスクアセスメント	リスク(0.351)、ビジネス(0.305)、サイバーセキュリティリスク(0.289)、管理(0.266)、特定(0.214)、資産(0.210)、組織(0.199)、企業(0.193)、アセスメント(0.181)、機能(0.173)
	リスク管理戦略	リスク(0.523)、管理(0.276)、ビジネス(0.267)、戦略(0.224)、順位(0.178)、サイバーセキュリティリスク(0.178)、優先(0.178)、組織(0.169)、許容(0.167)、度(0.167)
	サプライチェーンリスク	リスク(0.368)、管理(0.302)、サプライチェーンリスク(0.263)、ビジネス(0.253)、評価(0.237)、組織(0.198)、順位(0.185)、優先(0.185)、サイバーセキュリティリスク(0.169)、特定(0.150)
Protection（防御）	ID管理とアクセス制御	アクセス(0.532)、認可(0.306)、保護(0.211)、認証(0.197)、防御(0.196)、制御(0.196)、トランザクション(0.175)、管理(0.174)、ユーザ(0.158)、デバイス(0.158)
	意識向上およびトレーニング	トレーニング(0.344)、向上(0.322)、意識(0.322)、保護(0.303)、防御(0.215)、セキュリティ(0.212)、責任(0.193)、関連(0.181)、手順(0.171)、教育(0.168)
	データセキュリティ	保護(0.494)、データ(0.280)、性(0.265)、完全(0.249)、防御(0.240)、情報(0.226)、可用性(0.211)、機密(0.188)、セキュリティ(0.182)、記録(0.150)
	情報を保護するためのプロセスおよび手順	保護(0.524)、手順(0.267)、防御(0.242)、情報(0.218)、プロセス(0.213)、セキュリティ(0.176)、範囲(0.174)、コミットメント(0.174)、経営(0.154)、目的(0.139)
	保守	保守(0.449)、保護(0.317)、修理(0.263)、制御(0.225)、防御(0.225)、実施(0.194)、手順(0.179)、アクセス(0.168)、用(0.155)、コンポーネント(0.155)
	保護技術	保護(0.501)、技術(0.363)、防御(0.242)、手順(0.193)、セキュリティソリューション(0.189)、セキュリティ(0.184)、ポリシー(0.181)、制御(0.170)、確保(0.167)、レジリエンス(0.155)
Detection（検知）	異常とイベント	検知(0.569)、異常(0.462)、イベント(0.417)、タイムリー(0.232)、把握(0.154)、発見(0.154)、サイバーセキュリティイベント(0.142)、継続(0.142)、モニタリング(0.142)、可能(0.125)
	セキュリティの継続的なモニタリング	モニタリング(0.523)、検知(0.501)、継続(0.289)、サイバーセキュリティイベント(0.253)、セキュリティ(0.190)、有効(0.173)、検証(0.173)、異常(0.157)、発見(0.157)、識別(0.144)
	検知プロセス	検知(0.776)、異常(0.273)、イベント(0.226)、プロセス(0.224)、タイムリー(0.205)、継続(0.151)、テスト(0.140)、発見(0.137)、サイバーセキュリティイベント(0.126)、モニタリング(0.126)
Response（対応）	分析	対応(0.571)、分析(0.439)、実施(0.196)、支援(0.186)、適切(0.178)、サイバーセキュリティイベント(0.178)、低減(0.165)、対処(0.165)、検知(0.165)、計画(0.148)
	コミュニケーション	対応(0.535)、機関(0.216)、執行(0.216)、法(0.216)、コミュニケーション(0.216)、調整(0.191)、間(0.191)、利害(0.182)、内外(0.173)、関係(0.172)
	改善	対応(0.655)、改善(0.311)、教訓(0.269)、過去(0.203)、現在(0.203)、活動(0.197)、意思(0.163)、計画(0.155)、サイバーセキュリティイベント(0.150)、低減(0.139)
	低減	緩和(0.393)、対応(0.381)、低減(0.323)、インシデント(0.269)、根絶(0.236)、拡大(0.236)、影響(0.182)、サイバーセキュリティイベント(0.174)、実施(0.174)、対処(0.162)
	対応計画	対応(0.726)、計画(0.295)、発生(0.223)、検知(0.219)、実施(0.214)、サイバーセキュリティイベント(0.214)、対処(0.132)、低減(0.132)、維持(0.123)、タイムリー(0.116)
Recoveriy（復旧）	コミュニケーション	復旧(0.628)、者(0.187)、計画(0.168)、ベンダ(0.165)、被害(0.165)、コミュニケーションコーディネーティングセンター(0.165)、オーナー(0.165)、インターネットサービスプロバイダ(0.165)、csirt(0.165)、機能(0.14
	改善	復旧(0.712)、計画(0.302)、改善(0.259)、教訓(0.224)、将来(0.169)、機能(0.145)、軽減(0.135)、実現(0.135)、状態(0.135)、阻害(0.135)
	復旧計画	復旧(0.779)、計画(0.294)、維持(0.184)、タイムリー(0.173)、サイバーセキュリティイベント(0.159)、実施(0.159)、機能(0.124)、阻害(0.115)、状態(0.115)、軽減(0.115)

レームワーク」の語彙と文書頻度を用いて，特徴語ベクトル $\boldsymbol{l_j}$ を計算し，カテゴリの特徴語ベクトル $\boldsymbol{c_i}$ との間のコサイン類似度を計算した（図 3 ② を参照）．

5. カテゴリごとに算出したコサイン類似度の総和をとり，提案手法のスコア $S_i(C_i)$ を計算した（図 3 ③ を参照）．

作成された特徴語ベクトルの次元は 361 次元で，各カテゴリの上位 10 の単語については，**表 1** として記載した．

プログラミング言語は python を用い，tf-idf とコサイン類似度の計算は scikit-learn [25] を利用している．

5. 実験

この章では，まず提案手法の効果確認のための実験に用いた文書と，評価用のデータの作成方法について説明し，その後提案手法と評価用データの結果について述べる．評価用のデータの作成には，解析対象のガイドラインに対して質的コーディングを行っている．このコーディングの詳細については，5.2 節で説明を行う．

また，Cybersecurity Framework の枠組みを用いないテキストマイニング手法との比較を行うための，tf-idf と k 平均法を用いてクラスタリングを行う実験と，LDA によるトピック分析の実験の手順について述べる．

5.1 解析対象の文書

経済産業省の運用者向けガイドライン [6] のうち，タイトル内に，「中小企業」の単語が含まれているものを用いた（**表 2**）．これは，特に問題をかかえていると思われる中小企業のセキュリティ担当者が，最初にこれらのガイドラインにふれると考えられるためである．各ガイドラインの概要を以下に示す．

「中小企業の情報セキュリティ対策ガイドライン」[26] は，中小企業の IT 利用の活用が進む中で中小企業がセキュリティ対策に取り組むための指針として 2009 年に作成され，2017 年に法改正など最新の情報を基に改定されたものである．このガイドラインには，チェックリストなどが同梱されており，学習目的のみだけでなく実際にガイドラインに基づいた運用を行えるよう工夫がされている．

「中小企業向けはじめてのマイナンバーガイドライン」[27] は，中小企業の担当者が「マイナンバーガイドライン」[28] を読む前に概要を学習することを想定して作成された文書である．今回検証の対象とした文書の中で最もページ数が少ない．

「中小企業 BCP（事業継続計画）ガイド」[29] は，2008 年に中小企業庁により発行された，中小企業が行うべき事業継続計画について記載した文書である．Cybersecurity Framework 1.1 においても事業継続と災害復旧に対する言及は存在するため，今回解析・評価対象に含めた．

「出社してから退社するまで中小企業の情報セキュリティ対策実践手引き」[30] は，中小企業で一般的な業務に潜む情報セキュリティ上のリスクを洗い出し，評価，対策することを支援するための文書である．評価に用いた文書の中で最もページ数が多いが，第 2 部（p.20 以降）については，1 ページにつき 1 項目分の説明で統一されている．

表 2 解析・評価対象とした文書一覧
Table 2 Documents for analysis and test.

文書名	文書の概要	ページ数
中小企業の情報セキュリティ対策ガイドライン	リスクアセスメントやポリシー策定などセキュリティ対策の導入についての文書	54
はじめてのマイナンバーガイドライン	マイナンバーに関する法令や運用についての文書	8
中小企業ＢＣＰ（事業継続計画）ガイド	災害時の事業継続計画についての文書	43
出社してから退社するまで中小企業の情報セキュリティ対策実践手引き	一般的な業務に潜む情報セキュリティ上のリスクを洗い出し、評価、対策することを支援する文書	110

5.2 質的コーディングによる評価用のデータの作成

提案手法の精度を測定するためには，「解析対象の文書の内容を人がどのように理解しているか」を定量的に表すことが必要とされる．そこで本研究では，質的コーディングの方法を用いて文書の内容を人の手で分析・定量化することで，提案手法の評価に用いるデータとした．

質的コーディングとは，質的データ解析の方法の 1 つで，文書に表れる表現に対してコード（符号）を割り当てることで，文書の内容について整理を行う手法である．事前にコーディングに用いる語群（コード群）を定義するテンプレートコーディングと，繰り返し文書を読みながら都度コードを作成していくオープンコーディングに大別される．

提案手法の評価を行うには，提案手法と比較可能な形式で文書の質的解析を行う必要がある．そこで，Cybersecurity Framework 1.1 のフレームワークコアのサブカテゴリをコード群として，解析対象にした 4 つの文書に対してテンプレートコーディングを実施した．この結果を提案手法の結果と比較することで，提案手法の評価が可能になる．

コーディングを実施する際には，

a) 原則，1 センテンスごとに評価を行う．「用語の説明 + 用語を用いた文」，「説明 + 補足事項」などの 2 つ以上のセンテンスで 1 つの意味をなしていると考えられた部分には，そのまとまりでの評価を実施している．
b) 複数のサブカテゴリに該当すると考えられた場合には，複数のコードを割り振る．
c) 図表など，画像として添付されている項目はコーディングの対象に含めない．
d) コード群に適切なコードが存在しないと思われる場合には，その文に対してコードの割り振りは実施しない．

こととした．

たとえば，「中小企業の情報セキュリティ対策ガイドライン」の中にある「パソコンにはウイルス対策ソフトを入れてウイルス定義ファイルを自動更新するなどのように，パソコンをウイルスから守るための対策を行っていますか？」という文には「悪質なコードを検出できる」というコード

表 3　各文書に対する提案手法と質的コーディングの実施結果
Table 3　Results of the proposed procedure and template coding for each document with color scale.

a)　中小企業の情報セキュリティ対策ガイドライン

	提案手法による解析					質的コーディング		
機能	機能ごとの平均値	スコア	正規化スコア	カテゴリ	正規化後のスコアの差	スコア	正規化スコア	機能ごとの平均値
Identify（特定）	63.4631	60.58109	0.6017	資産管理	0.0112	38	0.6129	28.3333
		62.66334	0.6294	ビジネス環境	**0.5003**	8	0.1290	
		64.10374	0.6485	ガバナンス	0.1902	52	0.8387	
		67.09194	0.6882	リスクアセスメント	0.0269	41	0.6613	
		63.7634	0.6440	リスク管理戦略	**0.5633**	5	0.0806	
		62.57506	0.6282	サプライチェーンリスク	0.2088	26	0.4194	
Protection（防御）	70.2238	43.9369	0.3806	アクセス制御	0.2516	8	0.1290	20.1667
		73.78872	0.7771	意識向上およびトレーニング	0.1481	39	0.6290	
		90.56535	1.0000	データセキュリティ	**0.8387**	10	0.1613	
		89.1779	0.9816	情報を保護するためのプロセスおよび手順	0.0184	62	1.0000	
		59.64659	0.5893	保守	**0.5893**	0	0.0000	
		64.22752	0.6501	保護技術	**0.6179**	2	0.0323	
Detection（検知）	33.1768	26.96574	0.1552	異常とイベント	0.1552	0	0.0000	2.3333
		49.67934	0.4569	セキュリティの継続的なモニタリング	0.3440	7	0.1129	
		22.88544	0.1010	検知プロセス	0.1010	0	0.0000	
Response（対応）	36.2112	40.64815	0.3369	分析	0.3047	2	0.0323	6.8000
		44.90455	0.3935	コミュニケーション	0.1193	17	0.2742	
		26.56475	0.1499	改善	0.0114	10	0.1613	
		40.69605	0.3376	低減	0.3053	2	0.0323	
		28.24251	0.1721	対応計画	0.1238	3	0.0484	
Recovery（復旧）	20.9207	15.28354	0.0000	改善	0.1452	9	0.1452	3.6667
		29.58491	0.1900	コミュニケーション	0.1900	0	0.0000	
		17.89355	0.0347	復旧計画	0.0024	2	0.0323	

b)　中小企業向けはじめてのマイナンバーガイドライン

	提案手法による解析					質的コーディング		
機能	機能ごとの平均値	スコア	正規化スコア	カテゴリ	正規化後のスコアの差	スコア	正規化スコア	機能ごとの平均値
Identify（特定）	13.1319	14.50888	0.9173	資産管理	**0.6096**	4	0.3077	3.5000
		13.59565	0.8527	ビジネス環境	**0.8527**	0	0.0000	
		13.20291	0.8248	ガバナンス	0.1752	13	1.0000	
		15.67625	1.0000	リスクアセスメント	**0.9231**	1	0.0769	
		10.78312	0.6535	リスク管理戦略	**0.6535**	0	0.0000	
		11.02435	0.6706	サプライチェーンリスク	0.4398	3	0.2308	
Protection（防御）	7.7915	7.611298	0.4288	アクセス制御	0.0327	6	0.4615	4.5000
		8.446021	0.4880	意識向上およびトレーニング	0.1274	8	0.6154	
		8.609577	0.4995	データセキュリティ	0.1158	8	0.6154	
		8.5153	0.4929	情報を保護するためのプロセスおよび手順	0.1083	5	0.3846	
		6.209951	0.3296	保守	0.3296	0	0.0000	
		7.357107	0.4108	保護技術	0.4108	0	0.0000	
Detection（検知）	2.8879	2.521282	0.0684	異常とイベント	0.0684	0	0.0000	0.6667
		3.968102	0.1708	セキュリティの継続的なモニタリング	0.0170	2	0.1538	
		2.174301	0.0438	検知プロセス	0.0438	0	0.0000	
Response（対応）	5.1365	4.617821	0.2169	分析	0.2169	0	0.0000	0.0000
		9.342579	0.5515	コミュニケーション	**0.5515**	0	0.0000	
		3.164115	0.1139	改善	0.1139	0	0.0000	
		5.386191	0.2713	低減	0.2713	0	0.0000	
		3.171816	0.1144	対応計画	0.1144	0	0.0000	
Recovery（復旧）	3.2189	1.555793	0.0000	改善	0.0000	0	0.0000	0.0000
		6.522643	0.3517	コミュニケーション	0.3517	0	0.0000	
		1.578266	0.0016	復旧計画	0.0016	0	0.0000	

c)　中小企業 BCP（事業継続計画）ガイド

	提案手法による解析					質的コーディング		
機能	機能ごとの平均値	スコア	正規化スコア	カテゴリ	正規化後のスコアの差	スコア	正規化スコア	機能ごとの平均値
Identify（特定）	35.8055	49.7153	1.0000	資産管理	0.0000	18	1.0000	13
		35.8844	0.6664	ビジネス環境	0.2781	17	0.9444	
		29.1137	0.5030	ガバナンス	0.0030	9	0.5000	
		36.8698	0.6901	リスクアセスメント	0.2543	17	0.9444	
		32.1633	0.5766	リスク管理戦略	0.2433	6	0.3333	
		31.0864	0.5506	サプライチェーンリスク	0.0605	11	0.6111	
Protection（防御）	19.8708	12.3003	0.0974	アクセス制御	0.0419	1	0.0556	5.5
		17.769	0.2294	意識向上およびトレーニング	**0.5484**	14	0.7778	
		20.7976	0.3024	データセキュリティ	0.1357	3	0.1667	
		28.8285	0.4961	情報を保護するためのプロセスおよび手順	0.3372	15	0.8333	
		17.3627	0.2196	保守	0.2196	0	0.0000	
		22.1665	0.3354	保護技術	0.3354	0	0.0000	
Detection（検知）	12.3075	12.6695	0.1063	異常とイベント	0.1063	0	0.0000	0
		15.9913	0.1865	セキュリティの継続的なモニタリング	0.1865	0	0.0000	
		8.26154	0.0000	検知プロセス	0.0000	0	0.0000	
Response（対応）	23.5848	26.3673	0.4368	分析	0.4368	0	0.0000	1.8
		35.2885	0.6520	コミュニケーション	0.1520	9	0.5000	
		15.8842	0.1839	改善	0.1839	0	0.0000	
		20.3853	0.2925	低減	0.2925	0	0.0000	
		19.9985	0.2831	対応計画	0.2831	0	0.0000	
Recovery（復旧）	47.5215	45.2986	0.8935	改善	0.3935	9	0.5000	9.666667
		48.3188	0.9663	コミュニケーション	**0.5774**	7	0.3889	
		48.9473	0.9815	復旧計画	0.2593	13	0.7222	

d)　出社してから退社するまで中小企業の情報セキュリティ対策実践手引き

	提案手法による解析					質的コーディング		
機能	機能ごとの平均値	スコア	正規化スコア	カテゴリ	正規化後のスコアの差	スコア	正規化スコア	機能ごとの平均値
Identify（特定）	208.2439	232.312	0.8955	資産管理	0.2876	31	0.6078	15.0000
		211.0702	0.7960	ビジネス環境	**0.7764**	1	0.0196	
		231.5448	0.8919	ガバナンス	**0.6370**	13	0.2549	
		201.8335	0.7527	リスクアセスメント	0.4193	17	0.3333	
		188.1098	0.6884	リスク管理戦略	0.4531	12	0.2353	
		184.593	0.6719	サプライチェーンリスク	0.3582	16	0.3137	
Protection（防御）	194.1280	203.8884	0.7623	アクセス制御	0.2377	51	1.0000	27.3333
		157.5106	0.5450	意識向上およびトレーニング	0.3489	10	0.1961	
		254.6157	1.0000	データセキュリティ	0.0784	47	0.9216	
		209.3266	0.7878	情報を保護するためのプロセスおよび手順	0.0161	41	0.8039	
		146.5039	0.4934	保守	0.4934	0	0.0000	
		192.9226	0.7109	保護技術	0.4168	15	0.2941	
Detection（検知）	76.9775	72.72001	0.1477	異常とイベント	0.1281	1	0.0196	17.0000
		114.1258	0.3417	セキュリティの継続的なモニタリング	**0.5406**	45	0.8824	
		44.08675	0.0135	検知プロセス	0.0845	5	0.0980	
Response（対応）	76.7417	91.13232	0.2340	分析	0.1555	4	0.0784	2.4000
		102.5866	0.2876	コミュニケーション	0.2288	3	0.0588	
		50.26428	0.0425	改善	0.0425	0	0.0000	
		82.27769	0.1925	低減	0.1925	0	0.0000	
		57.44748	0.0761	対応計画	0.0219	5	0.0980	
Recovery（復旧）	64.1140	41.19798	0.0000	改善	0.0588	3	0.0588	4.0000
		97.00023	0.2615	コミュニケーション	0.1634	5	0.0980	
		54.14372	0.0607	復旧計画	0.0214	2	0.0392	

（符号）を割り振っている．この例では，文中に「悪質なコード」などの単語は表れていないが，「ウイルス」が「悪質なコード」を指していると読み取れ，その検出技術の導入を促しているため，このコードが適切であると判断した．

また，本研究での提案手法の解析結果を見ることでコーディング結果に影響が出ないように，各解析対象の文書に対して提案手法を適用する前に，質的コーディングを実施し，レビューを行った．

最終的なコーディングの結果については，付録として表 A·1 に記載をした．

定量評価を実施するため，コードが割り当てられた文の数について，カテゴリごとに和をとり，質的コーディングによる記述数を表すスコア $SQ_i(C_i)$ とした．

5.3　提案手法と質的コーディングの結果

提案手法のスコア $S_i(C_i)$ は，カラーコード表示（緑：低

表 4 各文書での tf-idf と k 平均法によるクラスタリングの結果
Table 4 Results of the tf-idf and k-means clustering for each document.

文書	ID	各クラスタに属する文書の割合	各クラスタの特徴語上位10個 （）内はクラスタの中心の数値
中小企業の情報セキュリティ対策ガイドライン	0	0.69885	情報(0.02856), 場合(0.01983), 個人(0.01821), 評価(0.01783), 利用(0.01698), 事故(0.01675), 管理(0.01552), 発生(0.01536), 経営(0.01527), 業務(0.01375)
	1	0.17449	対策(0.18686), セキュリティ(0.16468), 情報(0.14280), 実施(0.04883), ポリシー(0.04310), 必要(0.03864), 組織(0.02491), 実行(0.02390), 経営(0.02285), 検討(0.02228)
	2	0.05314	重要(0.32306), 度(0.20306), 取組(0.12208), 情報(0.08597), 管理(0.04146), 資産(0.03906), 算定(0.03883), 値(0.03826), 判断(0.03647), 記入(0.03273)
	3	0.03100	ガイドライン(0.3846), 中小(0.37181), 企業(0.35302), セキュリティ(0.26426), 情報(0.23549), 対策(0.22681), 不利益(0.04112), ポリシー(0.03777), 重要(0.02922), 策定(0.02353)
	4	0.04252	リスク(0.43500), 値(0.15342), 大(0.08709), 算定(0.06775), 実施(0.05376), 対策(0.05018), 分析(0.04246), 小(0.03734), もの(0.03488), 資産(0.03139)
中小企業向けはじめてのマイナンバーガイドライン	0	0.56809	等(0.06108), 社(0.04444), 会社(0.03704), ルール(0.02963), 保管(0.02653), 提供(0.02368), ページ(0.02114), 取得(0.02106), 策定(0.02014), 漏えい(0.02007)
	1	0.21401	委託(0.6672), 先(0.1712), 再々(0.1253), 必要(0.07223), 監督(0.06806), 許諾(0.06631), 者(0.05391), 場合(0.04593), 最初(0.03182), 同様(0.02976)
	2	0.08560	個人(0.2094), 番号(0.1587), 情報(0.1108), 必要(0.1107), 特定(0.1003), 等(0.09197), 廃棄(0.06397), 適切(0.06290), 場合(0.05693), 監督(0.05212)
	3	0.07393	措置(0.37213), 安全(0.3684), 管理(0.3613), 物理(0.1227), 人的(0.08918), 技術(0.08678), 組織(0.08295), 適正(0.07380), 取扱い(0.07067), 者(0.05513)
	4	0.05837	者(0.28143), 事業(0.2021), 取扱(0.15215), 事務(0.14225), 担当(0.13453), 規模(0.13401), 中小(0.10849), 等(0.06162), 情報(0.06053), 特定(0.05778)
中小企業BCP（事業継続計画）ガイド	0	0.76885	員(0.19636), 従業(0.17962), 企業(0.14986), 中小(0.10533), 等(0.03995), 安否(0.038226), ｂｃｐ(0.03527), 確認(0.03290), 家族(0.03027), 時(0.02902)
	1	0.05068	事業(0.24649), 復旧(0.15476), 中核(0.15051), 目標(0.09285), 継続(0.09193), 時間(0.09128), 計画(0.03599), 時(0.03017), 経営(0.02677), 特定(0.02575)
	2	0.01545	担当(0.60588), 社員(0.40559), 調達(0.09212), 輸送(0.07909), 搬送(0.07853), 経理(0.07722), 調整(0.07675), 加工(0.07088), 避難(0.02913), 用(0.02585)
	3	0.07169	緊急(0.01327), 時(0.01282), 等(0.01163), 社(0.01146), 工場(0.01139), ｂｃｐ(0.01122), 連絡(0.01102), 確保(0.01023), 代替(0.00961), 資金(0.00939)
	4	0.09333	bcp(0.39292), 企業(0.12469), 導入(0.10977), 策定(0.06965), 運用(0.05710), 済み(0.05058), サイクル(0.04360), 中小(0.04050), 発動(0.03278), 会社(0.03135)
出社してから退社するまで中小企業の情報セキュリティ対策実践手引き	0	0.71998	情報(0.04748), 影響(0.02871), 性(0.02829), 実体(0.02775), 備考(0.02570), 運用(0.02518), ポイント(0.02488), 処(0.02321), サービス(0.02247), 保存(0.02198)
	1	0.03205	管理(0.18360), 記憶(0.17526), 媒体(0.17464), 策(0.16805), メモリ(0.15172), usb(0.15063), 関連(0.14567), アプリケーション(0.13664), ネットワーク(0.13356), 項目(0.07300)
	2	0.11500	対策(0.42138), セキュリティ(0.24812), 人的(0.19670), 技術(0.19415), 目的(0.16577), 現状(0.15350), レベル(0.14550), 情報(0.00560), 記述(0.00461), シート(0.00405)
	3	0.08805	責任(0.70439), 実施(0.69439), 参考(0.00934), 本人(0.00000), 管理(0.00000), 員(0.00000), 従業(0.00000), システム(0.00000), 外(0.00000), 要因(0.00000)
	4	0.04492	者(0.40052), 本人(0.31734), 管理(0.26461), 員(0.25278), 従業(0.25248), システム(0.22107), 外(0.17806), 要因(0.17748), 偶発(0.12668), 訪問(0.12284)

表 5 各文書でのトピック分析の結果
Table 5 Results of the topic analysis for each document.

文書	ID	トピック分布	各トピックの特徴語上位10個（）はそのトピックでの単語の出現頻度
中小企業の情報セキュリティ対策ガイドライン	0	0.06667	必要(0.03955), 管理(0.02585), 組織(0.02514), 攻撃(0.02100), 委託(0.01817), 確認(0.01786), リスク(0.01638), 実施(0.01587), 重要(0.01480), 脅威(0.01473)
	1	0.06675	利用(0.03893), 重要(0.03505), 管理(0.03068), 脅威(0.02035), ポリシー(0.02032), 脆弱(0.02032), 必要(0.01894), 資産(0.01735), 自社(0.01605), 責任(0.01422)
	2	0.06778	実施(0.03551), 企業(0.03373), 場合(0.03120), 可能(0.02179), 先(0.01930), ガイドライン(0.01745), ウイルス(0.01654), 中小(0.01630), 実行(0.01516), 事業(0.01432)
	3	0.73175	企業(0.05300), ガイドライン(0.04213), 中小(0.03897), リスク(0.02880), 値(0.02081), 場合(0.01717), 発生(0.01711), 重要(0.01607), 度(0.01406), 漏えい(0.01405)
	4	0.06706	個人(0.03154), 経営(0.02648), 自社(0.02088), 事故(0.01993), 従業(0.01659), 員(0.01600), 責任(0.01545), 利用(0.01483), 管理(0.01449), 資産(0.01429)
中小企業向けはじめてのマイナンバーガイドライン	0	0.05007	委託(0.06956), 場合(0.06889), 監督(0.06856), 番号(0.05833), 先(0.05220), 提供(0.05139), 廃棄(0.04385), 保管(0.04346), 必要(0.04140), 担当(0.03498)
	1	0.05000	事務 (0.08497), 取扱 (0.08290), 委託 (0.05833), 適切 (0.05212), 担当 (0.048119), 取得 (0.04766), 必要 (0.04037), 保険 (0.03253), 監督 (0.03251), 利用 (0.02945)
	2	0.79787	措置(0.10067), 管理(0.10059), 安全(0.10047), 事業(0.04735), 中小(0.03960), 規模(0.03957), 物理(0.03956), 対応(0.03947), 方法(0.03204), 漏えい(0.03172)
	3	0.05137	特定(0.15690), 番号(0.0845), 事業(0.0653), 事務(0.0502), 取扱い(0.04), 管理(0.0370), 適正(0.0337), 機器(0.0295), 安全(0.0278), 廃棄(0.0256)
	4	0.05069	委託(0.09274), 必要(0.06774), 場合(0.05529), 事業(0.05428), 事務(0.03986), 員(0.03558), 従業(0.03521), 廃棄(0.03385), 先(0.03144), 書類(0.03068)
中小企業BCP（事業継続計画）ガイド	0	0.10089	時(0.06236), 緊急(0.05439), 復旧(0.04648), 員(0.03791), 従業(0.03744), 事業(0.03609), 災害(0.02606), 確保(0.02197), 目標(0.01958), 時間(0.01899)
	1	0.10001	情報(0.03419), 継続(0.02883), 資源(0.02657), 事業(0.02526), 代替(0.02395), 訓練(0.02310), 場所(0.02305), データ(0.01991), 企業(0.01996), バックアップ(0.01971)
	2	0.59841	例(0.05541), 記入(0.02587), 行(0.02416), 様式(0.02287), 連絡(0.02228), 工場(0.02161), 方法(0.02036), 企業(0.02009), 時(0.01942), 機械(0.01868)
	3	0.10013	BCP(0.04445), 確保(0.02999), 対策(0.02871), 企業(0.02646), 緊急(0.02187), 協力(0.02100), 家族(0.01749), 要員(0.01705), 代替(0.01506), 取引(0.01421)
	4	0.10056	企業(0.03914), 事業(0.03366), 電話(0.03129), 連絡(0.02546), 影響(0.02354), 経営(0.02178), ＢＣＰ(0.02119), 中小(0.01906), 代替(0.01860), 先(0.01848)
出社してから退社するまで中小企業の情報セキュリティ対策実践手引き	0	0.20000	セキュリティ(0.07954), ネットワーク(0.07023), アプリケーション(0.05325), 媒体(0.04930), 記憶(0.04798), 利用(0.03683), レベル(0.03652), メモリ(0.03496), USB(0.03278), 現状(0.03144)
	1	0.20000	管理(0.10552), システム(0.06640), 本人(0.06300), 従業(0.04803), 員(0.04802), 対策(0.03959), 業務(0.03949), 外(0.03151), 要因(0.03144), 技術(0.02334)
	2	0.20000	サービス(0.06010), セキュリティ(0.04434), サーバー(0.04275), PC(0.03802), 対策(0.037121), ファイル(0.02906), 影響(0.02758), リスク(0.02727), 机上(0.02623), 目的(0.02515)
	3	0.20000	情報(0.11040), 機器(0.04279), 機密(0.03590), スマート(0.03535), 完全(0.03448), 可用性(0.03306), 電子(0.03220), デバイス(0.03169), 適法(0.02880), コピー(0.02776)
	4	0.20000	資料(0.07814), 参考(0.07308), 付き(0.07241), 番号(0.06903), 保存(0.05582), ポイント(0.04547), 運用(0.04463), 理(0.04406), 備考(0.04055), 監査(0.01896)

⇔ 赤：高）とともに**表 3**「各文書に対する提案手法と質的コーディングの実施結果」の a) から d) の「提案手法による解析」の「スコア」に記載した．また，フレームワークコアの機能ごとにスコアの平均値を計算し「機能ごとの平均値」として記載している．質的コーディングによるスコア $SQ_i(C_i)$ についても，表 3 の「質的コーディング」の「スコア」に記載をし，「機能ごとの平均値」についても同様に計算して記載した．

「提案手法による解析」の「スコア」や「機能ごとの平均値」に，それぞれの文書の特徴が見て取れる．

たとえば，表 3 a) の「中小企業の情報セキュリティ対策ガイドライン」は，Identify（特定）と Protection（防御）の機能に関連したカテゴリのスコアが高く，リスクアセスメントやポリシの作成などの対策の準備段階に重点が置かれていることが予測される．実際，この文書では，リスクアセスメントとセキュリティポリシ作成を促している．

一方で，表 3 c) の「中小企業 BCP（事業継続計画）ガイド」では，Identify（特定）の「資産管理」のカテゴリや Recovery（復旧）の機能に含まれるカテゴリのスコアが高く，資産の洗い出しや，復旧計画などについて記載されていることが予測される．実際，この文書には名前のとおり，災害復旧計画を含む BCP（事業継続計画）の要素が多く記述されている．

定量的な評価については，次の評価の章で検討を行う．

5.4 枠組みを用いないテキストマイニング手法の実験

Cybersecurity Framework 1.1 を評価の枠組みとして用いている有利性を確認するために，解析対象の文書それぞれについて，tf-idf と k 平均法によるクラスタリングと，トピック分析を用いて文書内容の分析を行った．それぞれの実験結果については，**表 4**，**表 5** に記載する．

5.4.1 tf-idf と k 平均法によるクラスタリング

今回の提案手法では tf-idf を用いているため，比較対象として tf-idf を用いたクラスタリングによる実験を実施した．各解析対象の文書を対象に tf-idf を用いて特徴語ベクトルを作成し，k 平均法を用いてクラスタリングを行った．k 平均法は非階層型のクラスタリングのアルゴリズムで，対象の集合を指定した個数のクラスタに分割することができる．今回の実験では，クラスタの数は Cybersecurity Framework 1.1 の機能の数に合わせて 5 と設定した．

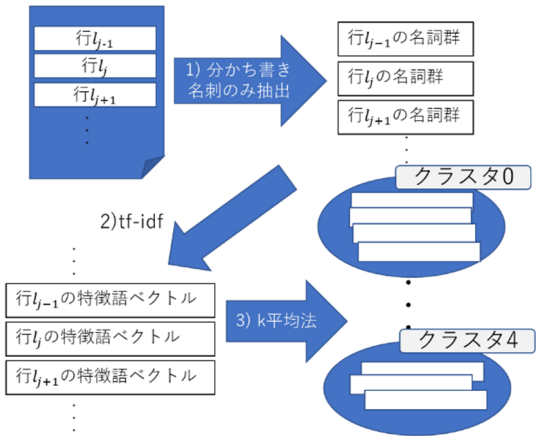


図 4 tf-idf と k 平均法によるクラスタリングとの手順

Fig. 4 Schematic diagram of the experimental procedure with tf-idf and k-means clustering.

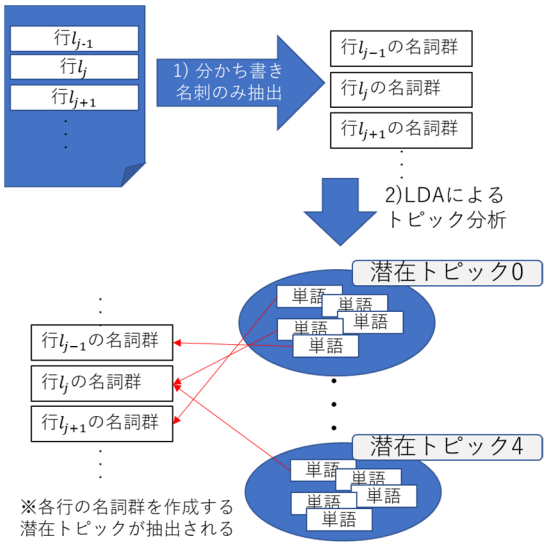


図 5 トピック分析の手順

Fig. 5 Schematic diagram of the experimental procedure with topic analysis.

具体的には，解析対象の文書それぞれに対して，以下の手順で解析を実施した．手順の図示を**図 4** に記載する．

1) 文書中の 1 行を 1 つの文章とし，Mecab を用いてその標準辞書で分かち書きと形態素解析を行い名詞の文書集合を作成した．
2) 作成した文書集合を用いて tf-idf を計算し，各文書の特徴語ベクトルを作成した．
3) scikit-learn を用いて，クラスタの数を 5 と設定して，k 平均法によるクラスタリングを実施した．

各クラスタの特徴について把握するために，各クラスタの中心における特徴語ベクトルのうち，上位 10 個を抽出した．この結果は，表 4 の「各クラスタの特徴語上位 10 個」に記載する．

5.4.2 トピック分析

今回の提案手法において Cybersecurity Framework 1.1 の枠組みを仮定していることの比較として，潜在的なトピックの存在を仮定するトピック分析の手法を用いて実験を行った．トピック分析とは，文書が複数の潜在的なトピックから確率的に生成されると仮定するトピックモデルに基づいたテキストマイニング手法である．

各解析対象の文書に対して gensim [31] を用いて LDA によるトピック分析を実施した．LDA によるトピック分析では，ある個数のトピックが文書内に潜在していると仮定して解析を行う．今回の実験では，Cybersecurity Framework 1.1 の機能の数に合わせて 5 つとした．

具体的には，解析対象の文書それぞれに対して，以下の手順で解析を実施した．手順の図示を**図 5** に記載する．

1) 文書中の 1 行を 1 つの文章とし，Mecab を用いてその標準辞書で分かち書きと形態素解析を行い名詞の文書集合を作成した．
2) 作成した文書集合に対して，gensim を用いて，LDA によるトピックモデルを作成した．この際トピックの数は 5 としている．

各トピックの特徴について把握するため，トピックの出現確率が高い上位 10 個の特徴語を抽出した．結果については表 5 の「各トピックの特徴語上位 10 個」に記載する．

6. 評価

まず，質的コーディングによる結果と提案手法との定量的な比較を行い提案手法の結果の妥当性について検討する．

その後，枠組みを仮定しないテキストマイニング手法による分類結果と提案手法で用いた Cybersecurity Framework のフレームワークコアの機能による分類の間の比較を行い，提案手法の有利性について検討する．

6.1 コーディング結果との比較による評価

本研究では，提案手法の提示内容の妥当性の評価を行うために事前にコーディングを行い各カテゴリの記述数を表す $SQ_i(C_i)$ を計算している．これを用いて提示内容が適切かどうかの評価を行う．

$S_i(C_i)$，$SQ_i(C_i)$ について，式 (2) で，正規化を行った．正規化後のスコアは表 3 の「正規化スコア」に記載をした．

$$N(X) = \frac{X - x_{min}}{|x_{max} - x_{min}|} \tag{2}$$

ここで X はデータセット全体を表し，各要素 x の正規化後の値を $N(x)$ と表すこととする．表 3 上で，正規後のスコアの差分の絶対値 $|N(S_i) - N(SQ_i)|$ を正規化後のスコアの差として表記し，これが 0.5 を超えるものについては，太字と二重下線で強調した．これらについて，次の議論と制限の 7.2 節で検討を行う．

提案手法のスコアを要素として持つベクトルを $\boldsymbol{M}$ とし，質的コーディングによるスコアを要素として持つベクトルを $\boldsymbol{Q}$ とする（式 (3)）．

表 6 提案手法の結果と質的コーディングの結果の間のコサイン類似度とピアソン積率相関係数

Table 6 Cosine Similarity and Pearson product-moment correlation coefficient between results of the proposed procedure and template coding.

文書名	コサイン類似度		ピアソン積率相関係数	
	カテゴリ(23)	機能(5)	カテゴリ(23)	機能(5)
中小企業の情報セキュリティ対策ガイドライン	0.768	0.932	0.605	0.902
中小企業向けはじめてのマイナンバーガイドライン	0.637	0.866	0.424	0.761
中小企業BCP（事業計画）ガイド	0.846	0.918	0.704	0.810
出社してから退社するまで中小企業の情報セキュリティ対策の実践手引き	0.792	0.912	0.607	0.699

$$\vec{M} = (S_1(C_1), \cdots, S_{23}(C_{23}))$$
$$\vec{Q} = (SQ_1(C_1), \cdots, SQ_{23}(C_{23})) \quad (3)$$

このベクトル $\boldsymbol{M}$，$\boldsymbol{Q}$ は計測手法が異なるため，直接の比較は難しい．そこで類似性にのみ注目し，コサイン類似度とピアソンの積率相関係数を各文書で計算した．その結果を**表 6** の「カテゴリ」として示す．また，機能ごとの平均値についても同様の計算したものを表 6 の「機能」として示す．機能で見た場合コサイン類似度のへ平均は 0.907，カテゴリで見た場合の平均は 0.761 とであった．また，相関係数も機能では強い正の相関を示す数値であり，カテゴリでも正の相関を示す数値であった．したがって，提案手法による結果と質的コーディングによる結果には類似性が見られるため，提案手法の結果は人の感覚にそった妥当性のあるものであると考えられる．

6.2 枠組みを用いないテキストマイニング手法との比較よる評価

本研究では，有利性の確認のため，解析対象の文書に対して枠組みを用いない方法でもテキストマイニングを実施した．この分類結果と Cybersecurity Framework 1.1 の枠組みとの比較を行うことで，枠組みを用いる利点を確認する．

6.2.1 Cybersecurity Framework の機能の特徴語の抽出

枠組みを用いないテキストマイニングの結果が，Cybersecurity Framework 1.1 のフレームワークコアのどの機能に類似しているかを比較するために，tf-idf を用いて，機能ごとの特徴語を計算した．計算では，4.2 節で「重要インフラのサイバーセキュリティを改善するためのフレームワーク」から抽出したカテゴリ C_i についての記述をそれぞれの機能ごとにまとめて，機能についての文書とし，機能の文書の間で tf-idf の計算を行った．特徴語ベクトルの次元は 376 次元で，各カテゴリの上位 10 の単語については，**表 7** として記載した．

表 7 機能ごとの特徴語上位 10 個

Table 7 Top 10 words extracted for each Function by tf-idf.

機能	機能毎の特徴語 上位10個 ()はtf-idfの計算結果
Identify（特定）	ビジネス(0.35377), リスク(0.34304), 管理(0.32180), サイバーセキュリティリスク(0.26164), 組織(0.22458), 順位(0.17688), 優先(0.17688), 理解(0.17275), 機能(0.17033), 特定(0.17029)
Protection（防御）	保護(0.48204), 防御(0.31241), アクセス(0.28507), 情報(0.20460), 制御(0.19916), 技術(0.19916), トレーニング(0.19525), 向上(0.19135), 意識(0.18744), 保守(0.17182)
Detection（検知）	検知(0.70770), 異常(0.28660), モニタリング(0.27327), サイバーセキュリティイベント(0.23486), イベント(0.21521), 継続(0.18919), タイムリー(0.16286), 発見(0.15633), 機能(0.14898), セキュリティ(0.12722)
Response（対応）	対応(0.62111), サイバーセキュリティイベント(0.23948), 分析(0.23292), 低減(0.23186), 対処(0.20609), 機能(0.19641), 支援(0.18868), 検知(0.18706), 実施(0.16572), 計画(0.15965)
Recovery（復旧）	復旧(0.64329), 計画(0.26673), 機能(0.23887), サイバーセキュリティイベント(0.20397), 阻害(0.16710), 軽減(0.16710), 実現(0.16710), 状態(0.16710), 改善(0.14121), 策定(0.13481)

表 8 各クラスタの中心の特徴語ベクトルとワークコアの機能の特徴語ベクトルのコサイン類似度

Table 8 Cosine Similarity between feature words vector of cluster center and one of function of framework core.

文書名	a) 中小企業の情報セキュリティ対策ガイドライン				
クラスタID	0	1	2	3	4
ラベル	個人情報の管理	セキュリティ対策ポリシー	資産管理	企業向けのガイドラインによる対策	リスク管理
Identify(特定)	0.2325	0.1159	0.2029	0.1129	0.3405
Protection(防御)	0.2718	0.2911	0.2025	0.1514	0.0945
Detection(検知)	0.1103	0.1921	0.0384	0.0917	0.0408
Response(対応)	0.1991	0.1607	0.0404	0.0488	0.0920
Rcovery(復旧)	0.1242	0.1021	0.0267	0.0339	0.0322

文書名	b) 中小企業向けはじめてのマイナンバーガイドライン				
クラスタID	0	1	2	3	4
ラベル	情報の管理	委託先の監督	個人番号	物理的な安全管理	事業者の取り扱い
Identify(特定)	0.0625	0.0251	0.1267	0.2287	0.1346
Protection(防御)	0.1506	0.0185	0.1234	0.1165	0.0580
Detection(検知)	0.0381	0.0071	0.0672	0.0173	0.0202
Response(対応)	0.0598	0.0190	0.0825	0.0237	0.1077
Rcovery(復旧)	0.0628	0.0116	0.0382	0.0240	0.0581

文書名	c) 中小企業BCP（事業継続計画）ガイド				
クラスタID	0	1	2	3	4
ラベル	事業継続計画での従業員の安否確認	中核事業の復旧計画	輸送調達搬送などの調整	工場における緊急時のBCP	企業でのBCP導入
Identify(特定)	0.0817	0.1552	0.0002	0.1115	0.0660
Protection(防御)	0.0262	0.0449	0.0093	0.1214	0.0208
Detection(検知)	0.0128	0.0607	0.0000	0.0571	0.0145
Response(対応)	0.0490	0.0547	0.0044	0.1450	0.0199
Rcovery(復旧)	0.0505	0.2993	0.0048	0.1372	0.0531

文書名	d) 出社してから退社するまで中小企業の情報セキュリティ対策実践手引き				
クラスタID	0	1	2	3	4
ラベル	情報運用の影響	IT環境の管理	セキュリティ対策	実施責任	システムの管理
Identify(特定)	0.1954	0.1525	0.0446	0.0118	0.1411
Protection(防御)	0.2811	0.0817	0.1817	0.0979	0.0548
Detection(検知)	0.0895	0.0095	0.1247	0.0635	0.0180
Response(対応)	0.1236	0.0076	0.0739	0.1163	0.0302
Rcovery(復旧)	0.0974	0.0069	0.0541	0.0745	0.0661

6.2.2 k 平均法によるクラスタリングの結果との比較評価

各クラスタリングの特徴語を基に各クラスタのラベルを考案した．この結果を**表 8** の「ラベル」に記載した．付けられたラベルを見る限りでは，どの文書においてもおおむねその文書の代表的な話題を表現しているように見える．その一方で，これらのラベルは，Cybersecurity Framework 1.1 のフレームワークコアの機能と比較すると下記のような印象をうけた．

1) セキュリティ対策に関係のない分類がある．
2)（1 文書内での）分類に偏りがある．
3) 体系立てられた分類にはなっていない．

そこで，各クラスタと各機能の間の関係性を把握するため各クラスタの中心での特徴語ベクトルと 6.2.1 項で計算した機能の特徴語ベクトルの間のコサイン類似度を計算し，それらの傾向の類似性について確認を行った．その結果を，カラーコード表示（緑：低 ⇔ 赤：高）とともに表 8 に記載する．

まず，「中小企業向けはじめてのマイナンバーガイドライン」のクラスタ ID1 の「委託先の監督」や，「中小企業 BCP（事業継続計画）ガイド」のクラスタ ID2 の「輸送・調達・搬送などの調整」のように，どの機能とも類似性の低い項目が確認された．これらは 1) のセキュリティ対策に直接関係のない分類であると考えられる．

また，たとえば「中小企業の情報セキュリティ対策ガイドライン」では，5 つすべてのクラスタで，Identify（特定）の機能との間で他の機能に比べて相対的に高い類似性を示しており，それにより 2) のような分類に対する偏りを感じたものと考えられる．

1 つの機能と強く類似しているクラスタは，3 つのみ（「中小企業の情報セキュリティ対策ガイドライン」のクラスタ ID4 の「リスク管理」，「中小企業向けはじめてのマイナンバーガイドライン」のクラスタ ID3 の「物理的な安全管理」，「中小企業 BCP（事業継続計画）ガイド」のクラスタ ID1 の「中核事業の復旧計画」）であり，残りのクラスタは，複数の機能と類似した部分を持っていると考えられる．Cybersecurity Framework は 1 つの体系化されたセキュリティ対策の枠組みであるため，これから外れていることで，3) の体系化されていないという印象を得たものと考えられる．

6.2.3 トピック分析の結果との比較評価

トピック分析の結果についても同様の方法で比較を実施した．特徴語から検討したトピックについては，**表 9** の「トピック」の項目に記載した．

トピック分析の結果においては，クラスタリングの場合とは異なり，1) セキュリティ対策に直接関係のないトピックが現れることはなかったが，2) の 1 つの機能に対する偏りがクラスタリングの場合と同じく，「中小企業の情報セキュリティ対策ガイドライン」に発生しているのが確認された．また，ほとんどのトピックが複数の機能と類似性を持っており，同様に少なくとも 1 つの体系には従っていない分類になっていることが確認された．

6.2.4 枠組みを用いないテキストマイニングとの比較評価まとめ

今回，tf-idf と k 平均法によるクラスタリングとトピック分析の 2 つの方法で分類を行い，その結果を確認したが，どちらにおいても，文書内容を表す分類結果を得ることはできるが，分類結果の偏りが発生し，（少なくとも 1 つの）体系に従っていない分類になっていることが確認された．

しかしながら，体制化方略は，理論や枠組みに基づいて学習要素を相互に関連付けて整理する学習方略であり，体系だった理論や枠組みに基づいて学習要素が相互に関連付けられ整理された形で提示されることが要求される．

提案手法においては Cybersecurity Framework 1.1 の枠組みに従ってテキストマイニングを行うため，体制化方略の要求である「体系だった理論や枠組み」による情報提示を行うことができる．この点において提案手法の可視化は，有利性があると考えられる．

表 9 トピックの出現頻度のベクトルとワークコアの機能の特徴語ベクトルのコサイン類似度

Table 9 Cosine similarity between appearance frequency vector of words of each topic and the feature words vector of function of framework core.

文書名	a) 中小企業の情報セキュリティ対策ガイドライン				
トピックID	0	1	2	3	4
トピック	包括的な管理の必要性	ポリシーに基づいた管理の重要性	中小企業におけるガイドライン実施	中小企業のリスク管理ガイドライン	経営者と従業員の責任
Identify(特定)	0.3413	0.2846	0.2226	0.2812	0.2417
Protection(防御)	0.1896	0.2469	0.1285	0.1580	0.2139
Detection(検知)	0.0558	0.0582	0.0899	0.0579	0.0551
Response(対応)	0.1126	0.1099	0.2160	0.1412	0.1301
Rcovery(復旧)	0.0740	0.1019	0.1112	0.0864	0.0856

文書名	b) 中小企業向けはじめてのマイナンバーガイドライン				
トピックID	0	1	2	3	4
トピック	委託先の番号管理の監督	事務における適切な情報の取り扱い	物理的な安全管理	番号と事業・事務	委託先の従業員
Identify(特定)	0.0593	0.0859	0.2144	0.2363	0.1263
Protection(防御)	0.0766	0.0529	0.0736	0.0678	0.0504
Detection(検知)	0.0294	0.0376	0.0075	0.0348	0.0211
Response(対応)	0.0506	0.0606	0.1312	0.0238	0.0565
Rcovery(復旧)	0.0279	0.0330	0.0286	0.0104	0.0197

文書名	c) 中小企業BCP（事業継続計画）ガイド				
トピックID	0	1	2	3	4
トピック	緊急時の目標復旧時間と従業員確保	事業継続のための情報管理と訓練	工場における事例，連絡先の管理	事業継続計画と緊急時の関係性	事業継続計画における連絡先
Identify(特定)	0.0794	0.1505	0.1011	0.0994	0.1340
Protection(防御)	0.0499	0.1340	0.0915	0.0805	0.0668
Detection(検知)	0.0354	0.0843	0.0291	0.0633	0.0348
Response(対応)	0.0630	0.0659	0.0527	0.1027	0.1068
Rcovery(復旧)	0.2396	0.1022	0.0378	0.1013	0.1009

文書名	d) 出社してから退社するまで中小企業の情報セキュリティ対策実践手引き				
トピックID	0	1	2	3	4
トピック	IT環境でのセキュリティ	従業員を含めたシステム管理	サーバ/サービスでのリスク対策	スマートデバイスに関するCIA	監査のための参考資料
Identify(特定)	0.0692	0.2134	0.0471	0.0273	0.0593
Protection(防御)	0.1933	0.1424	0.1182	0.1896	0.0322
Detection(検知)	0.0811	0.0543	0.2156	0.0408	0.0125
Response(対応)	0.0455	0.0448	0.0902	0.0237	0.0167
Rcovery(復旧)	0.0121	0.0301	0.1515	0.0324	0.0209

7. 議論と制限

評価結果をもとに本研究で行った提案の制限と改善方法について議論を行う．

7.1 記述内容の正確性と十分性についての制限

本研究の提案手法では，あるカテゴリもしくは機能についての記述があるかないか（多いか少ないか）について可視化することは可能だが，記述内容の正確性や十分性を可視化することはできない．

これは，本質的にはテキストマイニングを行ったことによる制限だが，Cybersecurity Framework のフレームワークコアでは，サブカテゴリの項目まで用いて詳細まで判定することで，問題の緩和を行える可能性がある．

ただし，本研究では，サブカテゴリでの特徴語ベクトルの作成の検討は行っていない．これは，特徴語ベクトル作

成には，文書量が不十分であると思われたためである．そのため，サブカテゴリでの解析を実施するためには，「フレームワークコア」の「参考情報」を用いるなどして，文書量を増やすなどの改善策を検討する必要があると考える．

7.2 誤差が発生したケースについての検討

この節では，正規化したスコアの差分が 0.5 を超えたカテゴリ，つまり，表 3 中で太字二重下線を用いてマークした項目について検討を行う．

7.2.1 同一機能内のカテゴリの特徴語が類似しているために誤差が発生したケース

太字二重下線でマークした項目のうち，「中小企業 BCP（事業継続計画）ガイド」の「意識向上およびトレーニング」のカテゴリと，「出社してから退社するまで中小企業の情報セキュリティ対策実践手引き」の「セキュリティの継続的なモニタリング」のカテゴリ以外のものは，提案手法のスコア $N(S_i)$ が，質的コーディングのスコア $N(SQ_i)$ を大きく超えていることで差が発生している．

この原因を検討するため，カテゴリの上位の特徴語について実際に確認した．表 1 によると各機能で見た場合には，上位の特徴語が類似していることが確認できる．たとえば，Identify（特定）の機能に属するカテゴリはすべて，「リスク」や「ビジネス」といった単語を上位の特徴として持っている．つまり，本研究の提案手法では，同一の機能に属するカテゴリどうしではスコアの差が発生しにくいという制限が存在すると考えられる．

これは，カテゴリ C_i について記述された部分を抽出する際に，C_i が属する機能についての記述も対象に含めた影響であると考えられる．そのため，カテゴリに関連した記述の抽出をする方法を変更することで改善される可能性がある．

7.2.2 特徴語の可能性がある単語が特徴語ベクトルに含まれていないために誤差が発生したケース

「中小企業 BCP（事業継続計画）ガイド」の「意識向上およびトレーニング」のカテゴリと，「出社してから退社するまで中小企業の情報セキュリティ対策実践手引き」の「セキュリティの継続的なモニタリング」のカテゴリでは，質的コーディングのスコア $N(SQ_i)$ が提案手法のスコア $N(S_i)$ を大きく超えていることで差が発生している．

この差の原因を推測するため，「出社してから退社するまで中小企業の情報セキュリティ対策実践手引き」の「セキュリティの継続的なモニタリング」のカテゴリ内のサブカテゴリの質的コーディングの結果を確認した．このカテゴリ内で最も多くコードとして割り振られたのは，「発生する可能性～（中略）～物理環境をモニタリングしている」のサブカテゴリと「悪質なコードを検出できる」の 2 つであった．どちらも割り振られた文の数は 10 個である．

このうち「悪質なコードを検出できる」がコードとして割り振られた文の 1 つに「コンピューター（PC、サーバー）をアンチウイルスソフトウェアで定期的（1 週間に 1 度程度）にスキャンする」という文がある．この文の特徴語ベクトル $\boldsymbol{L_j}$ を計算したところ，「スキャン」は上位の特徴語として現れたが，「アンチウイルス」ないし「ウイルス」は特徴語として出現しなかった．同様に残りの 9 個の文についても検証したところ，「ワーム」，「スパイウェア」，「トロイの木馬」，「ボッド」などの用語も特徴語として抽出されなかった．以上より，質的コーディングの際，コーディングの手がかりとなるが，提案手法の特徴語ベクトルに含まれていない単語があり，これがスコア差を生んでいると考えられる．

また，実験結果には明示的に表れていないが，他のカテゴリにおいても実際にコーディングを実施した結果と特徴語を比較してみると，特徴語と思われるにもかかわらず特徴語ベクトル中には出現していない単語がある事例が確認された．

たとえば，「中小企業の情報セキュリティ対策ガイドライン」のコーディングで「ビジネス環境」のカテゴリに属するコードが，「業務上の関係者（顧客，取引先，委託先，代理店，利用者，株主など）からの信頼を高めるには，…（中略）…，整理しておくことが重要です．」という文に割り振られている．この文に対して同様に特徴語ベクトル $\boldsymbol{L_j}$ を計算すると，「関係」，「業務」，「顧客」，「経営」，「利用者」などの単語については，上位の特徴語として現れていたが，「委託」の単語が特徴語ベクトルに存在していないことが確認できた．

これは，特徴語ベクトルの作成に利用した Cybersecurity Framework 1.1（重要インフラのサイバーセキュリティを改善するためのフレームワーク 1.1 版）の文書中に該当の単語自体が現れていなかったことが原因である．そのため，この問題については特徴語ベクトル作成時に利用する文書数を増やすか，適切な類義語や関連語を追加することで改善できると思われる．

7.3 質的コーディングの制限

今回質的コーディングの結果を評価データとして利用したが，注意点がある．

1. テンプレートコーディングでよく行われる複数人でのコーディングの実施と統計的なすり合わせ処理は，今回実施していない．この影響を低減させるため，提案手法の試行前に質的コーディングを実施し，コーディング結果についてもレビューを行った．
2. Cybersecurity Framework 1.1 の 106 項目のサブカテゴリをコード群とした．これは適切なコードの数より多いと考えられる．しかし，評価段階では，カテゴリに集約して利用しており，フレームワークコアにおいてサブカテゴリとカテゴリは包括的な関係にあると考

えられるので，同一カテゴリ内で発生するレベルでのコードの割当てミスや解釈違いについては，評価結果への影響はないと考えられる．

7.4 枠組みを用いないテキストマイニング手法の制限

今回，文書の内容・構造を仮定しないテキストマイニング手法として，tf-idf と k 平均法によるクラスタリングと LDA によるトピック分析を実施したが，注意点がある．

1. k 平均法と LDA どちらも確率的な要素を含む手法であるため，パラメータによっては異なるトピックやクラスタが現れる可能性がある．
2. 特徴語を基にトピックの内容やラベルを決める際に主観的な判断が含まれる．

8. 結論

本研究は，学習者の体制化方略を補助することを目指し，文書評価のための枠組みとして Cybersecurity Framework を用いること検討した．実際に Cybersecurity Framework 1.1 のフレームワークコアのカテゴリを基にして，tf-idf による特徴語ベクトルを作成し，解析対象の文書の内容をスコア化する手法を提案し，提案手法による解析結果について，質的コーディングの結果と比較することで結果の妥当性の評価を行った．また，枠組みによって文章の内容や構造を仮定しないテキストマイニング（tf-idf と k 平均法によるクラスタリングとトピック分析）の分類結果と，Cybersecurity Framework 1.1 のフレームワークコアの機能とを比較することで Cybersecurity Framework 1.1 を枠組みとして用いることの有利性の確認を行った．

コーディング結果による評価では，実際に，4 種類のガイドラインに対して本手法を適用しスコアを算出し，事前に質的コーディングをした結果をスコア化したものとコサイン類似度による類似性の確認を行ったところ，フレームワークコアの機能（5 項目）レベルでは平均 0.907，カテゴリ（23 項目）レベルでも平均 0.761 となり，提案手法の実用可能性を提示することができた．

一方で，本提案手法では，同一機能に属するカテゴリ間ではスコアの差が発生しにくく，カテゴリレベルでの正確性を欠く場合があるという制限や，特徴語ベクトル作成時の文書量の少なさが要因となり，解析対象の文書内に特徴語と思われる単語があっても評価対象にならないことがある制限が確認された．

tf-idf と k 平均法によるクラスタリングとトピック分析による解析では，Cybersecurity Framework 1.1 のような枠組みに従わない分類やトピックが発生することが確認された．Cybersecurity Framework 1.1 は 1 つの体系化されたセキュリティ対策の枠組みであり，体制化方略を促すためには，体系だった情報の提示が要求される．したがって，Cybersecurity Framework 1.1 の枠組みに基づいて文書内容を解析し可視化する手法は，セキュリティ分野において体制化方略を実現するための要求を満たすという観点で，枠組みを利用しないテキストマイニング手法と比較して有利性があると考えられる．

今後の課題として，実際に Cybersecurity Framework 1.1 に基づいた文書内容の提示を行うことで，学習効率がどの程度向上するのか，ユーザ実験の実施が必要である．

また，手法の改善の観点では，fastText [32]，doc2vec [33] などの方法を tf-idf の代わりに用いることも可能であると考えている．文書量の増加や類義語・関連語を利用した精度の向上についても検討することができる．

応用の観点では，セキュリティ分野でも異なるフレームワークを用いて，異なる観点での体制化を試みることが考えられる．たとえば，Cybersecurity Workforce Framework を用いることで役割の観点から再整理を行うことができると考えられる．また，異なる分野でも，1) 専門性が高く使用される用語が決まっており，2) その分野についての包括的な構造モデルが提示されている分野であれば，同様の手法を用いてそのモデルに基づいた可視化を行うことができる可能性があると考えている．

謝辞 コーディングと本論文のレビューにご協力いただいた皆様に，謹んで感謝の意を表する．

参考文献

[1] 「情報セキュリティ人材の育成に関する基礎調査」報告書について，入手先 ⟨https://www.ipa.go.jp/security/fy23/reports/jinzai/⟩（参照 2019-03-12）．
[2] 情報セキュリティ人材不足数等に関する追加分析について（概要），入手先 ⟨https://www.ipa.go.jp/files/000040646.pdf⟩（参照 2019-03-12）．
[3] サイバーセキュリティ人材の育成に関する施策間連携ワーキンググループ報告書，入手先 ⟨https://www.nisc.go.jp/conference/cs/pdf/jinzai-sesaku2018set.pdf⟩（参照 2019-03-12）．
[4] 情報セキュリティ事故対応に関わるアンケート調査，入手先 ⟨http://lab.iisec.ac.jp/˜hiromatsu_lab/files/jiko-questionnaire_result.pdf⟩（参照 2019-03-12）．
[5] 法人組織におけるセキュリティ実態調査 2017 年版，入手先 ⟨https://appweb.trendmicro.com/doc_dl/select.asp?type=1&cid=236⟩（参照 2019-03-12）．
[6] 運用者向けセキュリティ関連コンテンツ一覧，入手先 ⟨http://www.meti.go.jp/policy/netsecurity/secdoc/ope_contents.html⟩（参照 2019-03-12）．
[7] CYBERSECURITY FRAMEWORK，入手先 ⟨https://www.nist.gov/cyberframework⟩（参照 2019-03-12）．
[8] 尾崎敏司：情報セキュリティに関連するガイドラインの内容提示の手法の提案とその評価，CE-148，研究報告コンピュータと教育（CE），pp.1–8 (2019).
[9] 重要インフラのサイバーセキュリティを改善するためのフレームワーク 1.1 版，入手先 ⟨https://www.ipa.go.jp/files/000071204.pdf⟩（参照 2019-03-12）．
[10] Blei, D.M., Ng, A.Y. and Jordan, M.I.: Latent Dirichlet allocation, *Journal of Machine Learning Research*, Vol.3, pp.993–1022 (2003).
[11] 中矢 誠，富永浩之：Web ゲームサイトを題材とした攻

防型ハッキング競技の環境構築と運用実践—試行実践に基づいて改善を行った本番実践の結果と分析，研究報告コンピュータと教育（CE），Vol.12, pp.1–8 (2018).
[12] 阿部隆幸，中矢　誠，太田翔也，富永浩之：学校機関ごとの個別情報を組み込んだ情報セキュリティの導入教育のためのクイズ形式のアドベンチャーゲームの試作，第 79 回全国大会講演論文集，pp.737–773 (2017).
[13] 楠目　幹，阿部隆幸，中矢　誠，富永浩之：情報セキュリティの導入教育のための大会イベント BeeCon におけるハッキング競技 CTF の問題構築，第 79 回全国大会講演論文集，pp.739–740 (2017).
[14] 湯川誠人，井口信和：仮想マシンを用いた攻防戦型ネットワークセキュリティ学習支援システムにおけるネットワーク型 IDS を用いた不正侵入シナリオの実装，インターネットと運用技術シンポジウム論文集，pp.92–99 (2018).
[15] 福山和生, 谷口義明, 井口信和：仮想マシンを活用したネットワークセキュリティ学習支援システムの実装と評価，情報処理学会論文誌，Vol.57, No.3, pp.931–935 (2016).
[16] 孫　英敬，山口由紀，島田　創，高倉弘喜，谷口義明：技術能力に注目した情報セキュリティ教育課程開発のためのカリキュラム分析，情報処理学会論文誌，Vol.58, No.5, pp.1163–1174 (2017).
[17] NICE Cybersecurity Workforce Framework，入手先 ⟨https://www.nist.gov/itl/applied-cybersecurity/nice/resources/nice-cybersecurity-workforce-framework⟩（参照 2019-03-12）.
[18] 瀬尾美紀子：自律的・依存的援助要請における学習観とつまずき明確化方略の役割—多母集団同時分析による中学・高校生の発達差の検討，教育心理学研究，Vol.55, pp.170–183 (2007).
[19] 松沼光奉：学習内容の体制化と図作成方略が現在完了形の学習に及ぼす効果，教育心理研究．Vol.55, pp.414–425 (2007).
[20] Acar, Y., Backes, M., Fahl, S., Garfinkel, S., Kim, D., Mazurek, M.L. and Stransky, C.: Comparing the Usability of Cryptographic APIs, *IEEE Symposium on Security and Privacy* (2017).
[21] Gupta, V. and Lehal, G.S.: A Survey of Text Mining Techniques and Applications, *Journal of Emerging Technologies in Web Intelligence*, Vol.1, No.1, pp.60–76 (2009).
[22] Park, D., Kim, S., Lee, J., Choo, J., Diakopoulos, N. and Elmqvist, N.: ConceptVector: Text Visual Analytics via Interactive Lexicon Building using Word Embedding, *IEEE Trans. Visualization and Computer Graphics*, Vol.24, No.1, pp.361–370 (2018).
[23] 赤石美奈：文書群に対する物語構造の動的分解・再構成フレームワーク，人工知能学会論文誌，Vol.21, No.5A, pp.428–438 (2006).
[24] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. 2004 Conference on Empirical Methods in Natural Language Processing* (*EMNLP-2004*), pp.230–237 (2004).
[25] scikit-learn, 入手先 ⟨https://scikit-learn.org/stable/⟩（参照 2019-03-12）.
[26] 中小企業の情報セキュリティ対策ガイドライン，入手先 ⟨https://www.ipa.go.jp/security/keihatsu/sme/guideline/index.html⟩（参照 2019-03-12）.
[27] 中小企業向けはじめてのマイナンバーガイドライン，入手先 ⟨http://www.ppc.go.jp/files/pdf/160114_chusho.pdf⟩（参照 2019-03-12）.
[28] マイナンバーガイドライン，入手先 ⟨http://www.ppc.go.jp/legal/policy/⟩（参照 2019-03-12）.
[29] 中小企業 BCP（事業継続計画）ガイド，入手先 ⟨http://www.chusho.meti.go.jp/bcp/download/bcp_guide.pdf⟩（参照 2019-03-12）.
[30] 出社してから退社するまで中小企業の情報セキュリティ対策実践手引き，入手先 ⟨https://www.jnsa.org/result/2013/chusho_sec/data/chusho_security_tebiki_20140331.pdf⟩（参照 2019-03-12）.
[31] gensim, 入手先 ⟨https://radimrehurek.com/gensim/⟩（参照 2019-07-13）.
[32] fastText, 入手先 ⟨https://fasttext.cc/⟩（参照 2019-03-12）.
[33] Le, Q. and Mikolov, T.: Distributed Representations of Sentences and Documents, CoRR, abs/1405.4053, pp.1–9 (2014).

付　　録

A.1 コーディング結果

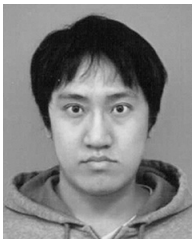
Cybersecurity Framework のサブカテゴリをコード群として，テンプレートコーディングを実施した結果を**表 A·1** として記載する．ただし，表中にはサブカテゴリの本文ではなく，フレームワークコアで定義されたサブカテゴリの識別子を用いて表記している．

表 A·1 テンプレートコーディングの実施結果

Table A·1 Results of template coding with sub-categories of cybersecurity framework.

機能	カテゴリ	サブカテゴリ識別子	中小企業の情報セキュリティ対策ガイドライン	中小企業向けはじめてのマイナンバーガイドライン	中小企業BCP～事業継続計画～ガイド	出社してから退社するまで中小企業の情報セキュリティ対策実践手引き
Identify	資産管理	ID.AM-1	3	0	4	8
		ID.AM-2	5	0	0	8
		ID.AM-3	3	0	0	1
		ID.AM-4	1	0	1	6
		ID.AM-5	16	0	10	2
		ID.AM-6	10	4	3	6
	ビジネス環境	ID.BE-1	5	0	4	1
		ID.BE-2	1	0	2	0
		ID.BE-3	1	0	3	0
		ID.BE-4	1	0	5	0
		ID.BE-5	0	0	3	0
	ガバナンス	ID.GV-1	18	5	4	6
		ID.GV-2	13	4	5	6
		ID.GV-3	20	4	0	1
		ID.GV-4	1	0	0	0
	リスクアセスメント	ID.RA-1	3	1	0	3
		ID.RA-2	8	0	0	2
		ID.RA-3	0	0	0	1
		ID.RA-4	14	0	10	3
		ID.RA-5	6	0	2	4
		ID.RA-6	10	0	5	4
	リスク管理戦略	ID.RM-1	0	0	0	3
		ID.RM-2	3	0	2	5
		ID.RM-3	2	0	4	4
	サプライチェーンリスク	ID.SC-1	18	0	8	9
		ID.SC-2	1	1	0	0
		ID.SC-3	3	0	0	7
		ID.SC-4	4	0	0	0
		ID.SC-5	0	2	3	0
Protection	アクセス制御	PR.AC-1	4	3	0	13
		PR.AC-2	2	2	1	12
		PR.AC-3	0	0	0	12
		PR.AC-4	2	1	0	5
		PR.AC-5	0	0	0	3
		PR.AC-6	0	0	0	2
		PR.AC-7	0	0	0	4
	意識向上およびトレーニング	PR.AT-1	15	1	4	9
		PR.AT-2	6	7	2	1
		PR.AT-3	2	0	3	0
		PR.AT-4	13	0	5	0
		PR.AT-5	3	0	0	0
	データセキュリティ	PR.DS-1	1	1	1	6
		PR.DS-2	0	1	0	4
		PR.DS-3	2	2	0	1
		PR.DS-4	0	0	2	5
		PR.DS-5	7	4	0	27
		PR.DS-6	0	0	0	2
		PR.DS-7	0	0	0	1
		PR.DS-8	0	0	0	1
	情報を保護するためのプロセスおよび手順	PR.IP-1	0	0	0	0
		PR.IP-2	0	0	0	0
		PR.IP-3	0	0	0	8
		PR.IP-4	2	0	6	12
		PR.IP-5	1	0	1	1
		PR.IP-6	2	3	0	4
		PR.IP-7	23	0	0	0
		PR.IP-8	1	0	0	0
		PR.IP-9	1	0	6	3
		PR.IP-10	0	0	2	1
		PR.IP-11	3	1	0	3
		PR.IP-12	29	1	0	9
	保守	PR.MA-1	0	0	0	0
		PR.MA-2	0	0	0	0
	保護技術	PR.PT-1	1	0	0	2
		PR.PT-2	0	0	0	4
		PR.PT-3	0	0	0	0
		PR.PT-4	1	0	0	6
		PR.PT-5	0	0	0	3

機能	カテゴリ	サブカテゴリ識別子	中小企業の情報セキュリティ対策ガイドライン	中小企業向けはじめてのマイナンバーガイドライン	中小企業BCP～事業継続計画～ガイド	出社してから退社するまで中小企業の情報セキュリティ対策実践手引き
Detection	異常とイベント	DE.AE-1	0	0	0	0
		DE.AE-2	0	0	0	0
		DE.AE-3	0	0	0	1
		DE.AE-4	0	0	0	0
		DE.AE-5	0	0	0	0
	セキュリティの継続的なモニタリング	DE.CM-1	0	0	0	8
		DE.CM-2	1	0	0	10
		DE.CM-3	0	1	0	9
		DE.CM-4	6	1	0	10
		DE.CM-5	0	0	0	3
		DE.CM-6	0	0	0	0
		DE.CM-7	0	0	0	4
		DE.CM-8	0	0	0	1
	検知プロセス	DE.DP-1	0	0	0	0
		DE.DP-2	0	0	0	1
		DE.DP-3	0	0	0	1
		DE.DP-4	0	0	0	1
		DE.DP-5	0	0	0	2
Response	対応計画	RS.RP-1	3	0	0	4
	伝達	RS.CO-1	4	0	7	2
		RS.CO-2	2	0	2	1
		RS.CO-3	3	0	0	0
		RS.CO-4	5	0	0	0
		RS.CO-5	3	0	0	0
	分析	RS.AN-1	0	0	0	0
		RS.AN-2	1	0	0	0
		RS.AN-3	0	0	0	0
		RS.AN-4	0	0	0	0
		RS.AN-5	1	0	0	0
	低減	RS.MI-1	1	0	0	0
		RS.MI-2	1	0	0	0
		RS.MI-3	0	0	0	0
	改善	RS.IM-1	4	0	0	1
		RS.IM-2	6	0	0	4
Recovery	復旧計画	RC.RP-1	2	0	13	3
	改善	RC.IM-1	4	0	2	1
		RC.IM-2	5	0	7	4
	伝達	RC.CO-1	0	0	3	0
		RC.CO-2	0	0	2	0
		RC.CO-3	0	0	2	2



尾崎 敏司 （学生会員）

2013 年トレンドマイクロ株式会社入社 Customer Technology Specialist. 2015 年産業技術大学院大学情報アーキテクチャ専攻修了．2017 年筑波大学大学院システム情報工学研究科リスク工学専攻博士課程入学.