Vol.2019-CE-148 No.13
2019/2/17

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

尾崎敏司[†1]

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の人材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．自己学習の起点と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初学者が把握するのは難しい．そこで，本研究では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 による特徴ベクトルを用いてガイドラインの文書内容の体系的な内容を提示する手法の提案を行い，質的コーディングにより実施した結果と比較することでその評価を行い，実用可能性を確認した．

Proposal to visualize a content of information security guideline based on pre-provided security framework

SATOSHI OZAKI[†1]

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の人材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」の必要性を指摘しており，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伺える．

前述の「情報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の人材不足のうち，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．

2. 関連技術と本研究の目的

セキュリティ人材育成に関する課題を解決するために，学習を促す手法について様々な提案がなされている．

例えば，2018 年には中矢誠らにより，複数人でプレイする Web ゲームサイトを題材とし，攻防型ハッキング競技としての体験的な演習が提案されている[6]．また，これに限らず，実際にセキュリティに関連した技術的な問題に挑戦することで，セキュリティに関連した技術を身に着ける CTF (Capture the Flag) によるアプローチに関した研究が多くみられる[7][8]．CTF 形式の学習では，コミュニティの育成を兼ねているためか複数人での学習が前提となっている

†1 筑波大学
University of Tsukuba

ことが多い．

演習環境に注目したものでは，ネットワークセキュリティ教育に重点を置き，仮想環境で演習環境を構築し実際の攻撃シナリオを演習することのできる環境の提案が行われている[9].

これらの学習手法について技術的な側面での学習として有用であると考えられるが，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，学習者自身の包括的な自己学習を促すアプローチが必要になると考えられる．

また，中小企業においては，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求められており，業務への適用を前提とした自己学習が重要な位置を占めていると考えられる．

実業務に基づいた自己学習の起点となる対策ガイドラインは多く公開されており，経産省で整理されているものに限っても 150 を超える[10]．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，一見してその項目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．

分野全体の全体像を把握できる情報が提示されていることは，学習者が自己の学習方策を立てる上で重要になると考えられ，また，利用しているガイドラインと全体像の差分を把握することは次のセキュリティ対策の方策を考える上で重要な情報になると考えられる．

そこで，本研究では，セキュリティ関連のガイドラインについてセキュリティ分野の全体像を把握できるような形で内容を提示する手法について検討し，その手法の評価を行う．

3. 提案手法

本研究では，全体像を意識した体系的な内容の提示を行うため，米国立標準技術研究所（NIST）から発行された Cybersecurity Framework 1.1[11] 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ことを検討した．既存のモデルに基づいて内容の提示を行うことにより，文書毎に個別に内容分析を行う場合に比べて，セキュリティ対策活動の全体像の把握や，文書間の内容の比較が容易になると考えられる．

また，今回，分類精度の量的な評価を可能にするために，事前に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た．これらについてもこの章で記載をする．

この章では，まず内容提示の枠組みとなる Cybersecurity Framework と解析対象である「中小企業の情報セキュリティ対策ガイドライン」について概要の説明を行い，次に，質的コーディングによる評価用データの作成方法について述べる．最後に，提案手法について記載する．

3.1 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，現在ある多様なサイバーセキュリティアプローチを体系化・構造化し，企業に示している（重要インフラのサイバーセキュリティを向上させるためのフレームワーク 1.0[12]　p2 より引用）”.

今回の対象とする文書は日本語であるため，IPA が発行している本文書を翻訳した「重要インフラのサイバーセキュリティを向上させるためのフレームワーク 1.0」を利用して解析を行っている．この文書は Cybersecurity Framework 1.0 を基に作成されたものであり，今回利用した Cybersecurity Framework 1.1 との差分部分については，英語版からの翻訳を行って解析に利用している．

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，参考情報の四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
特定 (ID)	資産管理 (ID.AM)：組織が事業目的を達成することを可能にするデータ、職員、デバイス、システム、施設を特定し、事業目標と自組織のリスク戦略との相対的重要性に応じて管理している。	ID.AM-1: 企業内の物理デバイスとシステムの一覧を作成している。	• CCS 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
		ID.AM-2: 企業内のソフトウェアプラットフォームとアプリケーションの一覧を作成している。	• CCS CSC 2 • COBIT 5 BAI09.01, BAI09.02, BAI09.05 • 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
		ID.AM-3: 企業内の通信とデータの流れの図を用意している。	• CCS CSC 1 • COBIT 5 DSS05.02 • ISA 62443-2-1:2009 4.2.3.4 • ISO/IEC 27001:2013 A.13.2.1 • NIST SP 800-53 Rev. 4 AC-4, CA-3, CA-9, PL-8

図 1 重要インフラのサイバーセキュリティを向上させるためのフレームワーク 1.0「表 2 フレームワークコア」より部分的に引用

機能の一意の識別子	機能	カテゴリーの一意の識別子	カテゴリー
ID	特定	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	リスク管理戦略
PR	防御	PR.AC	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	伝達
		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	伝達

図 2 重要インフラのサイバーセキュリティを向上させるためのフレームワーク 1.0「表 1 機能の一意の識別子とカテゴリーの一意の識別子」を引用

注意点として，図 2 は，Cybersecurity Framework 1.0 を基に作成されたものであり，Cybersecurity Framework 1.1 では「特定」の機能の下に新たに，「サプライチェーンマネジメント」のカテゴリが作成されている．この「サプライチェーンマネジメント」に関連した部分については，自ら Cybersecurity Framework 1.1 の該当部分を翻訳したものを解析に利用している．

3.2 中小企業の情報セキュリティ対策ガイドライン

このガイドラインは，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 年に法改正等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．特に問題を抱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ふれるドキュメントであろうと考えられるため，今回の解析・評価の対象とした．

3.3 質的コーディングによる評価用のデータの作成

提案手法の定量的な評価を実施する評価用のデータを得るために，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，「中小企業の情報セキュリティ対策ガイドライン」に対してテンプレートコーディングを実施した．

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一つの意味を成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．

こととした．

実際のコーディングの結果については，付録として表 3 に記載をした．

コード（サブカテゴリ―）が割り当てらてられた数について，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．

3.4 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方法 tf-idf （Term Frequency-Inverse Document Frequency） 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 $\mathbf{c_i}$ 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 $\mathbf{l_j}$ のコサイン類似度で記載することができる．従って，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文章中にあるかを表すスコア $S_i(C_i)$ は，この総和となるので，式 1 で評価することができると考えられる．

$$S_i(C_i) = \sum_j \frac{\vec{l_j} \cdot \vec{c_i}}{|\vec{l_j}||\vec{c_i}|} \quad \cdots\cdots \text{式 1}$$

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．

1. Cybersecurity Framework 1.1 内で各カテゴリ C_i について記述されている部分を確認し，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抽出し，カテゴリ C_i の文書の一部として取り扱った．
2. 抽出した全文書集合に対して，分かち書きを実施して，名詞句だけの集合に変換した．
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 毎に，tf-idf による特徴語ベクトル $\mathbf{c_i}$ を作成した．

4. 適用対象の文章から 1 行ごと文字列を抜き出し特徴語ベクトル $\mathbf{l_j}$ を計算し，カテゴリの特徴語ベクトル $\mathbf{c_i}$ との間のコサイン類似度を計算した．
5. カテゴリ毎に算出したコサイン類似度の総和を取り，提案手法のスコア $S_i(C_i)$ を計算した．

作成された特徴語ベクトルの次元は 360 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，分かち書きには，Mecab[13] を，tf-idf とコサイン類似度の計算は scikit-learn[14] のライブラリを利用している．

表 2 カテゴリ毎に抽出された特徴語上位 10 個

分類コード	カテゴリー	特徴語　上位10個　　()はtf-idfの計算結果
IDENTIFY（特定）	資産管理	管理(0.353)、リスク(0.287)、ビジネス(0.270)、事業(0.241)、特定(0.227)、戦略(0.215)、こと(0.215)、組織(0.186)、資産(0.177)、サイバーセキュリティリスク(0.171)
	ビジネス環境	ビジネス(0.378)、リスク(0.317)、管理(0.298)、上(0.227)、特定(0.211)、理解(0.191)、こと(0.181)、戦略(0.181)、順位(0.173)、優先(0.173)
	ガバナンス	管理(0.359)、リスク(0.334)、ビジネス(0.329)、サイバーセキュリティリスク(0.233)、理解(0.207)、特定(0.195)、上(0.192)、戦略(0.184)、こと(0.184)、ガバナンス(0.178)
	リスクアセスメント	ビジネス(0.336)、リスク(0.292)、サイバーセキュリティリスク(0.260)、管理(0.224)、こと(0.219)、特定(0.202)、企業(0.195)、組織(0.180)、理解(0.173)、戦略(0.164)
	リスク管理戦略	リスク(0.464)、ビジネス(0.321)、管理(0.320)、戦略(0.225)、特定(0.190)、こと(0.180)、順位(0.160)、サイバーセキュリティリスク(0.160)、優先(0.160)、組織(0.148)
	サプライチェーンリスク	リスク(0.398)、管理(0.297)、ビジネス(0.258)、特定(0.242)、サプライチェーンリスク(0.224)、評価(0.202)、確立(0.197)、組織(0.178)、順位(0.172)、優先(0.172)
Protection（防御）	アクセス制御	アクセス(0.402)、保護(0.362)、承認(0.314)、防御(0.234)、制御(0.234)、トランザクション(0.209)、ユーザ(0.188)、デバイス(0.188)、限定(0.183)、施設(0.162)
	意識向上およびトレーニング	保護(0.358)、トレーニング(0.320)、向上(0.305)、意識(0.305)、防御(0.305)、セキュリティ(0.197)、手順(0.174)、情報(0.156)、ため(0.155)、技術(0.153)
	データセキュリティ	保護(0.487)、防御(0.325)、データ(0.245)、性(0.230)、情報(0.202)、セキュリティ(0.185)、完全(0.168)、ため(0.165)、技術(0.163)、トレーニング(0.163)
	情報を保護するためのプロセスおよび手順	保護(0.524)、防御(0.311)、手順(0.236)、情報(0.199)、プロセス(0.187)、セキュリティ(0.177)、技術(0.168)、制御(0.162)、アクセス(0.162)、ため(0.161)
	保守	保守(0.399)、保護(0.376)、防御(0.320)、制御(0.240)、アクセス(0.200)、修理(0.187)、手順(0.182)、意識(0.160)、向上(0.160)、トレーニング(0.160)
	保護技術	保護(0.479)、防御(0.318)、技術(0.318)、制御(0.191)、手順(0.181)、セキュリティ(0.181)、アクセス(0.175)、ため(0.170)、トレーニング(0.159)、意識(0.159)
Detection（検知）	異常とイベント	検知(0.570)、異常(0.463)、イベント(0.418)、タイムリー(0.208)、発見(0.154)、サイバーセキュリティイベント(0.143)、継続(0.142)、把握(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	検知(0.602)、モニタリング(0.425)、継続(0.275)、サイバーセキュリティイベント(0.207)、発見(0.200)、異常(0.200)、セキュリティ(0.182)、的(0.158)、イベント(0.150)、機能(0.143)
	検知プロセス	検知(0.744)、異常(0.286)、イベント(0.229)、プロセス(0.193)、継続(0.193)、タイムリー(0.192)、発見(0.191)、モニタリング(0.175)、機能(0.137)、サイバーセキュリティイベント(0.132)
Response（対応）	分析	復旧(0.697)、伝達(0.182)、状態(0.180)、軽減(0.180)、計画(0.172)、通常(0.165)、機能(0.161)、ため(0.146)、運用(0.135)、者(0.134)
	伝達	復旧(0.741)、計画(0.257)、改善(0.229)、状態(0.179)、軽減(0.179)、通常(0.165)、機能(0.161)、教訓(0.148)、ため(0.146)、運用(0.135)
	改善	復旧(0.763)、計画(0.288)、維持(0.170)、サイバーセキュリティイベント(0.156)、タイムリー(0.152)、実施(0.150)、ため(0.137)、機能(0.121)、後(0.113)、中(0.113)
	低減	対応(0.619)、分析(0.393)、低減(0.213)、支援(0.188)、機能(0.178)、計画(0.175)、サイバーセキュリティイベント(0.172)、作成(0.159)、改善(0.159)、の(0.151)
	対応計画	対応(0.670)、検知(0.253)、サイバーセキュリティイベント(0.246)、伝達(0.216)、実施(0.173)、計画(0.159)、対処(0.152)、低減(0.152)、維持(0.134)、分析(0.134)
Recoveriy（復旧）	改善	対応(0.684)、改善(0.289)、低減(0.193)、教訓(0.187)、計画(0.180)、分析(0.170)、機能(0.162)、サイバーセキュリティイベント(0.157)、活動(0.144)、作成(0.144)
	伝達	対応(0.520)、低減(0.404)、分析(0.194)、影響(0.186)、機能(0.185)、インシデント(0.184)、サイバーセキュリティイベント(0.178)、改善(0.164)、作成(0.164)、計画(0.164)
	復旧計画	対応(0.708)、計画(0.288)、発生(0.218)、検知(0.214)、サイバーセキュリティイベント(0.208)、実施(0.200)、中(0.151)、後(0.151)、低減(0.129)、対処(0.129)

表 1 フレームワークコアに基づいた提案手法によるスコアと質的コーディングによるスコア

	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	提案手法のスコア	正規化後のスコア	カテゴリ	正規化後のスコアの差	質的コーディングによるスコア	正規化後のスコア	機能ごとの平均値
IDENTIFY（特定）	85.10625	83.7107	0.755397	資産管理	0.142494	38	0.61290323	28.33333
		82.49747	0.738171	ビジネス環境	0.609138	8	0.12903226	
		84.70715	0.769546	ガバナンス	0.069164	52	0.83870968	
		90.35975	0.849806	リスクアセスメント	0.188516	41	0.66129032	
		86.25294	0.791494	リスク管理戦略	0.710849	5	0.08064516	
		83.1095	0.746861	サプライチェーンリスク	0.327506	26	0.41935484	
Protection（防御）	87.11211	68.38192	0.537746	アクセス制御	0.408714	8	0.12903226	20.16667
		88.64493	0.825457	意識向上およびトレーニング	0.196425	39	0.62903226	
		100.6844	0.996404	データセキュリティ	0.835113	10	0.16129032	
		100.9377	1	情報を保護するためのプロセスおよび手順	0	62	1	
		80.47797	0.709496	保守	0.709496	0	0	
		83.54582	0.753056	保護技術	0.720798	2	0.03225806	
Detection（検知）	37.21109	31.82964	0.018746	異常とイベント	0.018746	0	0	2.333333
		49.29422	0.266723	セキュリティの継続的なモニタリング	0.15382	7	0.11290323	
		30.5094	0	検知プロセス	0	0	0	
Response（対応）	59.17382	56.82235	0.373614	分析	0.341356	2	0.03225806	6.8
		68.60442	0.540905	伝達	0.266712	17	0.27419355	
		53.80571	0.330781	改善	0.16949	10	0.16129032	
		60.31133	0.423153	低減	0.390895	2	0.03225806	
		56.3253	0.366556	対応計画	0.318169	3	0.0483871	
Recoveriy（復旧）	40.48532	43.0634	0.178252	改善	0.033091	9	0.14516129	3.666667
		37.3109	0.096573	伝達	0.096573	0	0	
		41.08167	0.150114	復旧計画	0.117856	2	0.03225806	

4. 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⇔赤：高）とともに表 2 の「提案手法による解析」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記載している．

機能ごとのスコアの平均値を見ると，「特定」「防御」についての記述の値が 80 程度で高く，次いで「対応」が 60 程度のスコアを示している．そのためこの文書は，特に「特定」「防御」について記載されているものだと予想される．

カテゴリのスコアを見ると，「防御」の機能の中でも特に「データセキュリティ」と「情報を保護するためのプロセス及び手順」の値が 100 程度で特に高くなっていることがわかる．

「検知」「対応」「復旧」の項目は全体的に低いスコアを示しており，特に，「検知」の「異常とイベント」，「検知プロセス」の項目が低い値であった．

以上より，このガイドラインはセキュリティ対策活動の内特に「特定」と「防御」の機能について重点的に述べており，セキュリティポリシーや保護を維持するためのプロセスや手順，社内コミュニケーションについて厚く記載があり，保護手法としてはデータセキュリティの側面からの記述が多くアクセス制御の側面からの記述が少なくなっていると考えられる．

5. 評価

本研究では，提案手法の評価を行うために事前にコーテ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．

質的コーディングによるスコア $SQ_i(C_i)$と機能ごとの平均値については，表 2 の質的コーディングの列に記載した．

$S_i(C_i)$ ，$SQ_i(C_i)$について，式 2 で，正規化を行った．正規化後のスコアとして表に記載をした．

$$N(X) = \frac{X - x_{min}}{|x_{max} - x_{min}|} \quad \cdots\cdots\text{式 2}$$

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．

提案手法のスコアを正規化した値 $N(S_i)$ を要素として持つベクトルを **M** とし，質的コーディングによるスコアを正規化した値 $N(SQ_i)$ を要素として持つベクトルを **Q** とする（式 3）．

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

このベクトル **M**, **Q** についてコサイン類似度を計算したところ，0.791 であった．

また，機能ごとの平均値についても同様の操作を行い，コサイン類似度を計算したところ 0.966 であった．

この結果より，機能による文書内容表示については，実用的に使える程度の精度をもち，カテゴリによる分類についても参考情報としては十分な精度にあると考えている．

しかしながら，質的コーディングのスコアと提案手法のスコアが著しく異なるカテゴリが確認できた．表 2 上で，正規後のスコアの差分の絶対値$|N(S_i) - N(SQ_i)|$が 0.5 を超えるものについて赤くマークを付けた．これらについて，次の議論と制限の項目で検討を行う．

6. 議論と制限

正規化したスコアの差分が 0.5 を超えたのは「ビジネス環境」,「リスク管理戦略」,「データセキュリティ」,「保守」,「保護技術」の 5 つのカテゴリであった．この著しい違いは全て提案手法のスコア $N(S_i)$ 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．例えば，Identify（特定）の機能に属するカテゴリは全て,「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同士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．

一方で，数値的には表れていないが，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，特徴語と思われるのにも関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．

例えば，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が，この文章を Cybersecurity Framework1.1 の特徴語でベクトル化した場合，「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あらわれていたが,「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「重要インフラのサイバーセキュリティを向上させるためのフレームワーク 1.0」の文書中に「委託」という単語自体が表れていな

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．

また，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．しかし，コーディングは提案手法の実験前に行い，コーディングの結果についてもレビューを実施した．
2. 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り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. 結論

セキュリティ人材の自己学習の補助のために，ガイドラインの内容を体系的に学習者に事前提示することを目的として，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た．

実際に，「中小企業の情報セキュリティ対策ガイドライン」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 0.966，カテゴリ（23 項目）レベルでも 0.791 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．

今後の展望として，手法の改善の観点では，fastText[15], doc2vec[16] 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．また，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，今回情報セキュリティの分野を対象に内容提示を試みたが，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内容提示を行うことができる可能性があると考えている．

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”. https://www.ipa.go.jp/security/fy23/reports/jinzai/, (参照 2019-01-24).
[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）”. https://www.ipa.go.jp/files/000040646.pdf, (参照 2019-01-24).
[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”. https://www.nisc.go.jp/conference/cs/pdf/jinzai-sesaku2018set.pdf, (参照 2019-01-24).
[4] “情報セキュリティ事故に関わるアンケート調査”. http://lab.iisec.ac.jp/~hiromatsu_lab/files/jiko-questionnaire_result.pdf , (参照 2019-01-24).
[5] “法人組織におけるセキュリティ実態調査 2017 年版”. https://appweb.trendmicro.com/doc_dl/select.asp?type=1&cid=236, (参照 2019-01-24)
[6] 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. 2018,vol 12, 研究報告コンピュータと教育（CE), p1-8
[7] 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. 2017, 第 79 回全国大会講演論文集 p 737 -73
[8] 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, 2017 第 79 回全国大会講演論文集 p 739 - 740
[9] 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, 2018, インターネットと運用技術シンポジウム論文集, 92 - 99
[10] “運用者向けセキュリティ関連コンテンツ一覧”. http://www.meti.go.jp/policy/netsecurity/secdoc/ope_contents.html, (参照 2019-01-24)
[11] ”CYBERSECURITY FRAMEWORK”. https://www.nist.gov/cyberframework, (参照 2019-01-24)
[12]“重要インフラのサイバーセキュリティを向上させるためのフレームワーク”. https://www.ipa.go.jp/files/000038957.pdf, (参照 2019-01-24)
[13] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237
[14] “scikit-learn”. https://scikit-learn.org/stable/, (参照 2019-01-24)
[15] “fastText”. https://fasttext.cc/, (参照 2019-01-24)
[16] Le，Q．and Mikolov，T.，“ Distributed Representations of Sentences and Documents ”，2014, CoRR，abs/1405．4053，p1-9

付録

付録 A.1 表 3 分類コード一覧 1

付録 A.1 表 4 分類コード一覧 2

事前に実施したテンプレートコーディングの分類コード群と実施結果

表 3 分類コード一覧 1

カテゴリ	サブカテゴリ（コード）	コードが割り振られた文の数
IDENTIFY¥資産管理	組織内の物理デバイスとシステムが目録になっている	3
IDENTIFY¥資産管理	企業内のソフトウェアプラットフォームとアプリケーションの一覧を作成している。	5
IDENTIFY¥資産管理	企業内の通信とデータの流れの図を用意している。	3
IDENTIFY¥資産管理	外部情報システムの一覧を作成している。	1
IDENTIFY¥資産管理	リソース（例：ハードウェア、デバイス、データ、ソフトウェア）を、分類、重要度、ビジネス上の価値に基づいて優先順位付けしている。	16
IDENTIFY¥資産管理	すべての従業員と第三者である利害関係者（例：供給業者、顧客、パートナー）に対して、サイバーセキュリティ上の役割と責任を定めている。	10
IDENTIFY¥ビジネス環境	サプライチェーンにおける企業の役割を特定し、伝達している	5
IDENTIFY¥ビジネス環境	重要インフラとその産業分野における企業の位置付けを特定し、伝達している。	1
IDENTIFY¥ビジネス環境	企業のミッション、目標、活動に関して優先順位を定め、伝達している。	1
IDENTIFY¥ビジネス環境	重要サービスを提供する上での依存関係と重要な機能を把握している。	1
IDENTIFY¥ビジネス環境	重要サービスの提供を支援する、レジリエンスに関する要求事項を定めている。	0
IDENTIFY¥ガバナンス	自組織の情報セキュリティポリシーを定めている。	18
IDENTIFY¥ガバナンス	情報セキュリティ上の役割と責任について、内部と外部パートナーとで調整・連携している。	13
IDENTIFY¥ガバナンス	プライバシーや市民の自由に関する義務を含む、サイバーセキュリティに関する法規制上の要求事項を理解し、管理している。	20
IDENTIFY¥ガバナンス	ガバナンスとリスク管理プロセスがサイバーセキュリティリスクに対応している。	1
IDENTIFY¥リスクアセスメント	資産の脆弱性を特定し、文書化している。	3
IDENTIFY¥リスクアセスメント	情報共有フォーラム／ソースより、脅威と脆弱性に関する情報を入手している。	8
IDENTIFY¥リスクアセスメント	内外からの脅威を特定し、文書化している。	0
IDENTIFY¥リスクアセスメント	ビジネスに対する潜在的な影響と、その可能性を特定している。	14
IDENTIFY¥リスクアセスメント	リスクを判断する際に、脅威、脆弱性、可能性、影響を考慮している。	6
IDENTIFY¥リスクアセスメント	リスクに対する対応を定め、優先順位付けしている。	10
IDENTIFY¥リスク管理戦略	リスク管理プロセスが自組織の利害関係者によって確立、管理され、承認されている。	0
IDENTIFY¥リスク管理戦略	自組織のリスク許容度を決定し、明確にしている。	3
IDENTIFY¥リスク管理戦略	企業によるリスク許容度の決定が、重要インフラにおける自組織の役割と、その分野に特化したリスク分析の結果に基づいて行われている。	2
IDENTIFY¥サプライチェーンリスク	"サイバーサプライチェーンのリスクマネジメントプロセスが、組織の利害関係者によって、特定、確立、評価、管理され、合意が取れてい	18
IDENTIFY¥サプライチェーンリスク	サプライチェーンリスク¥"サイバーサプライチェーンのリスクアセスメントプロセスを使用して、情報システム、コンポーネント、およびサービスのサプライヤと第三者パートナーを特定、優先順位付け、評価している	1
IDENTIFY¥サプライチェーンリスク	組織のサイバーセキュリティプログラムとサイバーサプライチェーンリスクマネジメント計画の目的に合うように設計された適切な措置を実施するために、サプライヤーと第三者パートナーとの契約が使用されている	3
IDENTIFY¥サプライチェーンリスク	サプライヤとサードパーティのパートナーは、契約上の義務を果たしていることを確認するために、監査、テスト結果、または他の形式の評価を使用しており、定期的にその評価している。	4
Protection¥アクセス制御	承認されたデバイスとユーザの識別情報と認証情報を管理している。	4
Protection¥アクセス制御	資産に対する物理アクセスを管理し、保護している。	2
Protection¥アクセス制御	リモートアクセスを管理している。	0
Protection¥アクセス制御	最小権限および職務の分離の原則を取り入れて、アクセス権限を管理している。	2
Protection¥アクセス制御	適宜、ネットワークの分離を行って、ネットワークの完全性を保護している。	0
Protection¥アクセス制御	同一性が証明され、認証情報に結びついており、相互に断定することができる	0
Protection¥アクセス制御	ユーザ、デバイス、その他の資産はトランザクションのリスク（例えば個人のセキュリティ/プライバシーリスクと、組織的なもの）に見合った認証（例えば、一要素、二要素）認証が実施されている。	0
Protection¥意識向上およびトレーニング	すべてのユーザに情報を周知し、トレーニングを実施している。	15
Protection¥意識向上およびトレーニング	権限を持つユーザが役割と責任を理解している。	6
Protection¥意識向上およびトレーニング	第三者である利害関係者（例：供給業者、顧客、パートナー）が役割と責任を理解している。	2
Protection¥意識向上およびトレーニング	上級役員が役割と責任を理解している。	13
Protection¥意識向上およびトレーニング	物理セキュリティおよび情報セキュリティの担当者が役割と責任を理解している。	3
Protection¥データセキュリティ	保存されているデータを保護している。	1
Protection¥データセキュリティ	伝送中のデータを保護している。	0
Protection¥データセキュリティ	資産について撤去、譲渡、廃棄プロセスを正式に管理している。	2
Protection¥データセキュリティ	可用性を確保するのに十分な容量を保持している	0
Protection¥データセキュリティ	データ漏えいに対する保護対策を実施している。	7
Protection¥データセキュリティ	ソフトウェア、ファームウェア、および情報の完全性の検証に、完全性チェックメカニズムを使用している。	0
Protection¥データセキュリティ	開発・テスト環境を実稼働環境から分離している。	0
Protection¥データセキュリティ	"ハードウェアの完全性の検証に、完全性チェックメカニズムを使用している	0

表 4 分類コード一覧 2

カテゴリ	サブカテゴリ（コード）	コードが割り振られた文の数
Protection¥情報を保護するためのプロセスおよび手順	情報技術／産業用制御システムのベースラインとなる設定を定め、維持している。	0
Protection¥情報を保護するためのプロセスおよび手順	システムを管理するためのシステム開発ライフサイクルを導入している。	0
Protection¥情報を保護するためのプロセスおよび手順	設定変更管理プロセスを導入している。	0
Protection¥情報を保護するためのプロセスおよび手順	情報のバックアップを定期的に実施、保持し、テストしている。	2
Protection¥情報を保護するためのプロセスおよび手順	自組織の資産の物理的な運用環境に関するポリシーと規制を満たしている。	1
Protection¥情報を保護するためのプロセスおよび手順	ポリシーに従ってデータを破壊している。	2
Protection¥情報を保護するためのプロセスおよび手順	保護プロセスを継続的に改善している。	23
Protection¥情報を保護するためのプロセスおよび手順	保護技術の有効性について、適切なパートナーとの間で情報を共有している。	1
Protection¥情報を保護するためのプロセスおよび手順	対応計画（インシデント対応および事業継続）と復旧計画（インシデントからの復旧および災害復旧）を実施し、管理している。	29
Protection¥情報を保護するためのプロセスおよび手順	対応計画と復旧計画をテストしている。	1
Protection¥情報を保護するためのプロセスおよび手順	人事に関わる対策にサイバーセキュリティ（例：アクセス権限の無効化、従業員に対する審査）を含めている。	0
Protection¥情報を保護するためのプロセスおよび手順	脆弱性管理計画を作成し、実施している。	3
Protection¥保守	自組織の資産の保守と修理 は、承認・管理されたツールを用いて、タイムリーに実施し、ログを記録している。	0
Protection¥保守	自組織の資産に対する遠隔保守は、承認を得て、ログを記録し、不正アクセスを防げる形で実施している。	0
Protection¥保護技術	ポリシーに従って監査記録／ログ記録の対象を決定、文書化し、そうした記録を実施して、レビューしている。	1
Protection¥保護技術	ポリシーに従って取り外し可能な外部記録媒体を保護し、そうした媒体の使用を制限している。	0
Protection¥保護技術	最小機能の原則を取り入れて、システムと資産に対するアクセスを制御している。	0
Protection¥保護技術	通信ネットワークと制御ネットワークを保護している。	1
Protection¥保護技術	通常条件下と悪条件下での回復力の要件を満たすための機構(例えば、フェールセーフ、ロードバランス、ホットスワップ)が実装されている。	0
Detection¥異常とイベント	ネットワーク運用のベースラインと、ユーザとシステム間の予測されるデータの流れを特定し、管理している。	0
Detection¥異常とイベント	攻撃の標的と手法を理解するために、検知したイベントを分析している。	0
Detection¥異常とイベント	イベントデータを複数の情報源やセンサーから収集し、相互に関連付けている。	0
Detection¥異常とイベント	イベントがもたらす影響を特定している。	0
Detection¥異常とイベント	インシデント警告の閾値を定めている。	0
Detection¥セキュリティの継続的なモニタリング	発生する可能性のあるサイバーセキュリティイベントを検知できるよう、ネットワークをモニタリングしている。	0
Detection¥セキュリティの継続的なモニタリング	発生する可能性のあるサイバーセキュリティイベントを検知できるよう、物理環境をモニタリングしている。	1
Detection¥セキュリティの継続的なモニタリング	発生する可能性のあるサイバーセキュリティイベントを検知できるよう、個人の活動をモニタリングしている。	0
Detection¥セキュリティの継続的なモニタリング	悪質なコードを検出できる。	6
Detection¥セキュリティの継続的なモニタリング	悪質なモバイルコードを検出できる。	0
Detection¥セキュリティの継続的なモニタリング	発生する可能性のあるサイバーセキュリティイベントを検知できるよう、外部サービスプロバイダの活動をモニタリングしている。	0
Detection¥セキュリティの継続的なモニタリング	権限のない従業員、接続、デバイス、ソフトウェアのモニタリングを実施している。	0
Detection¥セキュリティの継続的なモニタリング	脆弱性スキャンを実施している。	0
Detection¥検知プロセス	説明責任を果たせるよう、検知に関する役割と責任を明確に定義している	0
Detection¥検知プロセス	検知活動は必要なすべての要求事項を満たしている。	0
Detection¥検知プロセス	検知プロセスをテストしている。	0
Detection¥検知プロセス	イベント検知情報を適切な関係者に伝達している。	0
Detection¥検知プロセス	検知プロセスを継続的に改善している。	0
Response¥対応計画	イベントの発生中または発生後に対応計画を実施している。	3
Response¥伝達	対応が必要になった時の自身の役割と行動の順番を従業員は認識している。	4
Response¥伝達	対応計画に従って情報を共有している。	2
Response¥伝達	対応計画に従って、利害関係者との間で調整を行っている。	3
Response¥伝達	サイバーセキュリティに関する状況認識を深めるために、外部利害関係者との間で任意の情報共有を行っている	5
Response¥伝達	サイバーセキュリティの状況認知をより広するために、外部の利害関係者と自発的な情報共有が実施される	3
Response¥分析	検知システムからの通知を調査している。	0
Response¥分析	インシデントがもたらす影響を把握している。	1
Response¥分析	フォレンジクスを実施している。	0
Response¥分析	対応計画に従ってインシデントを分類している。	0
Response¥分析	部および外部の情報源から組織に開示された脆弱性を受信、分析、および対応するためのプロセスが確立されている（例：内部テスト、セキュリティ情報、またはセキュリティ研究者）	1
Response¥低減	インシデントを封じ込めている。	1
Response¥低減	インシデントを低減している。	1
Response¥低減	新たに特定された脆弱性に関して、許容できるリスクである場合にはその旨を文書化し、そうでない場合には低減している。	0
Response¥改善	学んだ教訓を対応計画に取り入れている。	4
Response¥改善	対応戦略を更新している。	6
Recoveriy¥復旧計画	イベントの発生中または発生後に復旧計画を実施している。	2
Recoveriy¥改善	学んだ教訓を復旧計画に取り入れている。	4
Recoveriy¥改善	復旧戦略を更新している。	5
Recoveriy¥伝達	広報活動を管理している。	0
Recoveriy¥伝達	イベント発生後に評判を回復している。	0
Recoveriy¥伝達	復旧活動について内部利害関係者と役員、そして経営陣に伝達している。	0