# プライバシーポリシー分類による法律遵守の分析

森 啓華[1,a)]　永井 達也[1]　高田 雄太[1]　神薗 雅紀[1]

**概要**：企業が個人情報保護方針を公表する手段の一つにプライバシーポリシーがある．プライバシーポリシーは，個人情報保護法および業界別に定められたガイドラインでその規程が推奨されており，一部の内容に関しては公表が義務付けられている．しかし，定期的な法改正や企業が提供するサービスの変化から，法律の要求事項に対してプライバシーポリシーの改訂が追いついていない現状がある．そこで本研究では，畳み込みニューラルネットワークを用いてプライバシーポリシーの記載内容を分類し，その分類結果と法律の要求事項を比較分析することで，プライバシーポリシーの内容の過不足を明らかにする．幅広い業界のプライバシーポリシーを分析した結果，8 割以上のポリシーにおいてファーストパーティにおける個人情報の取り扱いやデータセキュリティに関して十分な記載が確認されたが，9 割以上のポリシーにおいて第三者提供に関する記載が不足していた．これら記載の過不足は，ガイドラインの有無やビジネスの特性によって業界間で差が生じていることがわかった．

**キーワード**：プライバシーポリシー，個人情報保護法，畳み込みニューラルネットワーク

# Compliance Analysis by Privacy Policy Classification

Keika Mori[1,a)]　Tatsuya Nagai[1]　Yuta Takata[1]　Masaki Kamizono[1]

***Abstract:*** Companies inform users how they collect personal information on a privacy policy. A privacy policy is recommended by Personal Information Protection Law and guidelines and some policy's contents are required to be disclosed. However, companies are struggling with updating policies against the legal requirements due to the periodic law revisions and business changes. This study clarifies sufficient or unsufficient privacy policies by classifying the content using convolutional neural network and comparing the classification results with the legal requirements. As a result, over 80% of policies sufficiently described the handling of personal information by first parties and security measures, and over 90% unsufficiently described the provision of data to third-parties. We found that these differences in the amount of description depended on the industry guidelines and business characteristics.

***Keywords:*** Privacy Policy, Personal Information Protection Law, Convolutional Neural Network

## 1. はじめに

EU の GDPR や日本の個人情報保護法は，企業組織に個人情報の取り扱いに関する情報の公表を義務付けている．ウェブサイトにおけるプライバシーポリシーは，その公表手段として一般的に使用されているとともに，企業が顧客やユーザ，株主に対して個人情報保護方針を示すために使用されている．しかしながら，EU におけるコーポレートサイトのプライバシーポリシーの調査によると，50%以上のプライバシーポリシーが GDPR における収集データのカテゴリに関する公表事項を満たしていないことが報告されている [1].

GDPR の要求事項とは異なるものの，国内企業も日本の個人情報保護法の要求事項に従ってプライバシーポリシーによる法遵守が求められるが，定期的な法改正やビジネスの変化にプライバシーポリシーの更改が追いついていない現状がある．これらの背景を踏まえ，本研究の Research

[1] デロイト トーマツ サイバー合同会社
Deloitte Tohmatsu Cyber LLC
[a)] keika.mori@tohmatsu.co.jp

**表 1** 業界ごとの個人情報の取り扱いに関するガイドライン概要（一部抜粋）

| 名称 | 概要 |
|---|---|
| 電気通信事業における個人情報保護に関するガイドライン [2] | 電気通信事業者に対して，アプリケーションソフトウェアを提供する場合に，アプリケーションによる情報の取得等について明確かつ適切に定めたプライバシーポリシーの公表を推奨している．（第 14 条） |
| 金融分野における個人情報保護に関するガイドライン [3] | 金融分野における個人情報取扱事業者に対して，関係法令等を遵守する旨，個人情報を目的外に利用しない旨，苦情処理に適切に取り組む旨，個人情報の利用目的の通知・公表等の手続や開示等の手続等，個人情報の取扱いに関する質問及び苦情処理の窓口，本人から求めに応じて利用停止をする旨，委託処理，個人情報の取得元又はその取得方法等をプライバシーポリシーの公表事項として推奨している．（第 18 条） |
| 医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス [4] | 医療・介護関係事業者に対して，関係法令及び本ガイダンス等を遵守する旨，個人情報に係る安全管理措置の概要，開示等の手続，第三者提供の取扱い，苦情への対応，利用目的等をプライバシーポリシーの公表事項として推奨している． |
| 健康保険組合等における個人情報の適切な取扱いのためのガイダンス [5] | 健保組合等に対して，関係法令及び本ガイダンス等を遵守する旨，個人情報に係る安全管理措置の概要，開示等の手続，第三者提供の取扱い，苦情への対応等，利用目的等をプライバシーポリシーの公表事項として推奨している． |
| 信書便事業分野における個人情報保護に関するガイドライン [6] | 信書便事業者に対して，プライバシーポリシーの公表を推奨している．（第 12 条） |

Questions として以下の三つを定めた．

(1) 日本のプライバシーポリシーはどういった内容がどの程度過不足しているか？

(2) 日本のプライバシーポリシーは国内の法規制をどの程度遵守できているか？

(3) 業界ごとのビジネス特性はプライバシーポリシーの記載内容や法遵守率に影響を与えるか？

これらを明らかにするため，本研究ではプライバシーポリシーの記載内容の過不足を評価する．具体的には，まず畳み込みニューラルネットワーク（CNN）を用いて，プライバシーポリシーの記載内容のカテゴリに分類する．次に，カテゴリ分類結果を法律の要求事項と比較することで，記載内容の過不足を評価する．

本研究の貢献は以下の通りである．

- 日本のプライバシーポリシー 2,545 件を分析し，必須要求事項の遵守率が 60.5%であることを示す．
- 法遵守率は業界ごとに異なり，金融および公共業界の遵守率が高い傾向にあることを示す．
- 業界ごとのガイドラインや内部規則におけるプライバシーポリシーに公表すべき内容の具体的な記載が，高い法遵守率につながることを示す．

## 2. 研究背景

### 2.1 個人情報保護法

日本の個人情報保護法（個人情報の保護に関する法律）は，個人情報を取り扱う事業者が遵守すべき義務を定めている．現行法では個人情報や要配慮個人情報，匿名加工情報の取り扱いについて，実施すべきセキュリティ対策や第三者提供時の内部規則などが定められているが，社会や経済の情勢の変化を踏まえて三年おきに見直されている．これら規則の一部では，個人情報取り扱いに関する情報の公表が要求されている．例えば，18 条では個人情報を取得する目的の公表，36 条では作成した匿名加工情報に含まれる情報項目の公表が求められている．

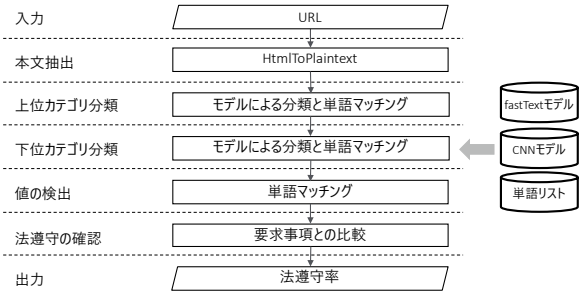


**図 1** 提案手法の分析フロー

### 2.2 個人情報保護に関するガイドライン

個人情報保護委員会は，個人情報の保護に関する法律についてのガイドラインを公表し，具体的な指針を定めている．加えて，一部の業界では**表 1** に記載するような業界個別のガイドラインが定められている．各業界のガイドラインは個人情報保護法を基礎とし，各業界に属する組織が適切に個人情報を取り扱うための具体的な指針を定めている．これらのガイドラインではプライバシーポリシーの公表も推奨されており，特に金融 [3] や医療 [4]，公共 [5] 業界のガイドラインでは，プライバシーポリシーで公表すべき事項について具体的に記載されている．

### 2.3 プライバシーポリシー

プライバシーポリシーには，企業組織の個人情報保護に関する考え方や個人情報の取り扱い方が記載されている．その他，前述のプライバシー法規制で求められる情報の公表を果たすためにプライバシーポリシーが活用されている．しかしながら，多くのプライバシーポリシーが GDPR の要求事項を十分に満たしていないという指摘もあり，プライバシーポリシーによる法遵守は容易ではない現状がある [1], [7]．日本においても，定期的な法改正やビジネスにおける個人情報の様々な利活用に即して，プライバイシーポリシーを改訂できていない同様の状況であると考えられる．

## 3. 提案手法

本研究ではプライバシーポリシーの記載内容の過不足および法遵守を評価する手法を提案する．提案手法は**図 1** に示す分析フローに従い，記載内容のカテゴリ分類と，分類結果と個人情報保護法における要求事項の比較分析を行う．

### 3.1 プライバシーポリシーの本文抽出

プライバシーポリシーを掲載している多くのウェブサイトには，ヘッダやフッタメニュー等のプライバシーポリシーには関係のないノイズが含まれている．入力の URL から取得した HTML ファイルに含まれるこれらノイズを除外し，プライバシーポリシー本文を抽出するために，`HtmlToPlaintext` [8] を利用した．`HtmlToPlaintext` は，HTML をパースし自然言語処理やヒューリスティクスにより本文を抽出できるが，英語にしか対応していないため，提案手法では日本語向けに一部更新して使用した．

### 3.2 文章のベクトル変換

後続の CNN モデルの入力データを準備するため，抽出したプライバシーポリシー本文をベクトルへ変換する．本文を，`MeCab` [9] を用いて単語ごとに区切り，`fastText` [10] を用いて各単語に対応するベクトルへ変換する．`fastText` は，自然言語処理により単語の類似を学習することにより，類似する単語をその距離が近くなるベクトルへ変換できる．本研究では，あらかじめ一万件以上のコーポレートサイトにおけるプライバシーポリシーを学習させた `fastText` モデルを使用した．

### 3.3 プライバシーポリシーのカテゴリ分類

#### 3.3.1 モデルによる分類

CNN モデルを用いて，前節で変換したベクトルを**図 2** に示す上位カテゴリおよび下位カテゴリに分類する．図 2 のカテゴリは，Wilson ら専門家が作成した OPP-115 データセット [11] における分類体系を基にしており，個人情報の取り扱いからプライバシーポリシーの管理に渡る幅広いカテゴリが網羅されている．CNN のモデルは Harkous らの手法 [12] を参考に，**図 3** に示す構成で構築する．なお，下位カテゴリは上位カテゴリごとに存在するため，上位カテゴリごとに下位カテゴリ向けの分類モデルを構築する．これら CNN モデルの入力はプライバシーポリシーの各行に含まれる単語のベクトルとし，その出力は上位カテゴリおよび下位カテゴリに属する確率とする．なお本研究では，出力された確率が 50%以上であるカテゴリを分類結果として採用する．

#### 3.3.2 上位および下位カテゴリの更新

提案手法では，図 2 に示したとおり，既存の分類体系 [11] の上位カテゴリに “共同利用・委託先” を追加し，“Do Not Track” を削除した 10 種類を使用する．これは，日本の法律において “特定の者との間で共同して利用される個人データが当該特定の者に提供される場合” は，第三者提供とは異なる要件が定められていることと，日本の法律では DNT の言及がなく，一部ブラウザで機能提供が停止されていることが理由である [13], [14]．下位カテゴリも上位カテゴリと同様に日本の法律を踏まえ，25 件のカテゴリを追加し，1 件のカテゴリを削除した．

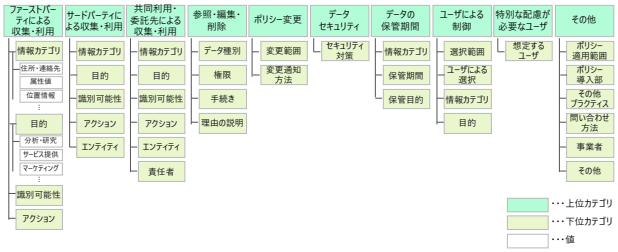


**図 2** 分類カテゴリ

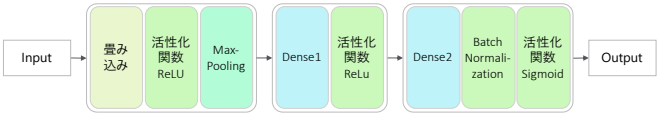


**図 3** 上位および下位カテゴリ分類モデルの構成

#### 3.3.3 単語マッチング

図 2 のカテゴリは，海外の法規制を想定したカテゴリが含まれているため，日本のプライバシーポリシーにあまり該当しないカテゴリが存在すると考えられる．この場合，学習データ不足により CNN を用いた分類は精度が落ちるため，提案手法では単語マッチングによる分類も適用する．単語マッチングに使用する単語リストは，法律の条文に使用されている単語や関連するプライバシーポリシーの記述に頻出する単語を目視で確認し，作成する．

### 3.4 値の検出

法規制の中には，カテゴリの粒度ではなく特定のキーワードや値の記載有無が問われる法案がある．例えば，個人情報保護法では “ファーストパーティ” の “収集目的” として，“第三者への提供” を記載することが要求されている法案や，“データセキュリティ” の “セキュリティ対策” として，“委託先の監督” や “従業員の教育” を記載することが推奨されている法案がある．これらの要求は文章ではなく単語粒度の情報であるため，前節と同様に単語マッチングを通じてキーワードや値の記載有無を検出する．ただし，単語マッチングに用いる単語リストは，学習データが十分あると考えられるため TF-IDF を用いて作成する．この時，プライバシーポリシー全体に値の検出を適用すると多くの値を検出してしまうため，下位カテゴリの分類結果に対して適用する．

**表 2** 法遵守の確認に使用した論理式の例

| 法案 | 条件事項 | 公表事項 |
|---|---|---|
| 第 18 条 1 項 | ファーストパーティ $\in L$ | ファーストパーティ_目的 $\in L$ |
| 第 23 条 2 項 | ( サードパーティ $\in L$) $\vee$ ( サードパーティ_アクション_ファーストパーティから受領 $\in L$) | ( ファーストパーティ_目的_第三者への提供 $\in L$)$\wedge$ ( サードパーティ_情報カテゴリ $\in L$)$\wedge$ ( サードパーティ_アクション $\in L$)$\wedge$ ( 参照・編集・削除_ユーザの権限_第三者提供の停止 $\in L$)$\wedge$ ( 参照・編集・削除_手続き_第三者提供の停止 $\in L$) |
| 第 36 条 3 項 | ファーストパーティ_識別可能性_統計・匿名 $\in L$ | ファーストパーティ_情報カテゴリ $\in L$ |
| 第 36 条 4 項 | (( サードパーティ $\in L$) $\vee$ ( サードパーティ_アクション_ファーストパーティから受領 $\in L$))$\wedge$ ( ファーストパーティ_識別可能性_統計・匿名 $\in L$)$\wedge$ ( サードパーティ_識別可能性_統計・匿名 $\in L$) | ( サードパーティ_アクション $\in L$) $\wedge$ ( サードパーティ_情報カテゴリ $\in L$) |
| 第 37 条 | サードパーティ_識別可能性_統計・匿名 $\in L$ | ( サードパーティ_アクション $\in L$) $\wedge$ ( サードパーティ_情報カテゴリ $\in L$) |

### 3.5 法遵守の確認

カテゴリ分類結果と法律の要求事項を照らし合わせ，法に遵守しているか確認する．具体的には，法律の法案ごとにおける要求事項をカテゴリや値の粒度で表現し，プライバシーポリシーの各行にラベル付けされた上位カテゴリ，下位カテゴリ，値の情報と比較する．個人情報保護法では，一定の条件を満たす場合に対応すべき事項が定められている．例えば，第 18 条 1 項では “個人情報取扱事業者は，個人情報を取得した場合は，あらかじめその利用目的を公表している場合を除き，速やかに，その利用目的を，本人に通知し，又は公表しなければならない．” と定められている．したがって，提案手法では “A のラベルを含む場合に，B のラベルを含むかどうか” といった形式の論理式を法案ごとに作成し，カテゴリ分類結果がその論理式を満たすか評価する．評価に使用した論理式の一部を**表 2** に示す．プライバシーポリシー内に含まれるラベルの集合を $L = \{l_i\}$ とし，ラベル $l_i$ には上位カテゴリ，上位カテゴリ_下位カテゴリ，上位カテゴリ_下位カテゴリ_値の三種類が含まれる．例えば，法案第 18 条 1 項では，“ファーストパーティ” の上位カテゴリが含まれる場合に，“ファーストパーティで収集した情報の利用目的” の下位カテゴリを含むかどうかを評価する．

## 4. 評価結果

### 4.1 データセット

#### 4.1.1 学習データセット

日本のコーポレートサイトに公開されているプライバシーポリシーを使用する．対象企業は，日経 DIGITAL 掲載企業 [15] の内，プライバシーポリシーの記載内容が充実している企業 64 社を金融や小売，情報通信，エナジーなど幅広い業界から選定した．これら企業のプライバシーポリシーのデータを用いて，3 名の研究者が行単位で上位カテゴリおよび下位カテゴリのラベル付けを行い，単語単位で値のラベル付けを行った．上位カテゴリのラベル付け結果を**表 3** に示す．全 5,164 ラベルのうち，2 名以上が同一のラベルをつけた 3,097 ラベルを学習データセットとして採用した．下位カテゴリについても，上位カテゴリと同様に 2 名以上が同一のラベルをつけた場合を採用し，計 2,775 ラベル採用した．値のラベルは単語単位で付与されており，人によって範囲が微細に異なったため，1 名以上が付けたラベルを採用した．

カテゴリによって学習データ数に偏りがあり，“サードパーティ_アクション” など一部の下位カテゴリにおいて学習データが存在しなかった．これら学習データに存在しないカテゴリには，CNN を用いたカテゴリ分類を適用できないため，本評価では法遵守の確認の対象外とした．

#### 4.1.2 単語リスト

学習データセットを作成した結果，上位カテゴリ “データの保管期間”，“ユーザによる制御”，および “特別な配慮が必要なユーザ” に属するデータはそれぞれ 7，4，5 件と少なかったため，これらのカテゴリは CNN モデルではなく，単語マッチングにより検出する．これらのカテゴリが少なかった要因として，日本の法律では個人データの消去義務がなく，保管期間も定められていないこと，ユーザによる制御方法に関する具体的な規則がないこと，子どもの個人情報の取り扱いに関する規則が定められていないことが考えられる．上位カテゴリの単語マッチングに使用した単語例として，“データの保管期間” カテゴリでは “保管”，“保存”，“期間” など，“ユーザによる制御” カテゴリでは “拒否” および “無効”，“特別な配慮が必要なユーザ” カテゴリでは “海外”，“在住”，“子ども” などがある．下位カテゴリも同様に，“共同利用・委託先” の下位カテゴリである “共同利用の責任者”，および “参照・編集・削除” の下位カテゴリである “ユーザからの請求に応じない理由の説明” に単語マッチングを適用した．

加えて，値の検出では，学習データセットの値ラベルの付いた単語を対象に TF-IDF を算出し，TF-IDF が 0.2 以上かつ同一下位カテゴリ内の 1 つの値ラベルでのみ 0.2 を超えている単語を単語リストとして採用した．

**表 3**　学習データとしてラベル付けした上位カテゴリのデータ数

| 上位カテゴリ | 3 人一致 | 2 人一致 | 1 人一致 |
|---|---|---|---|
| ファーストパーティによる収集・利用 | 524 | 428 | 416 |
| サードパーティによる収集・利用 | 128 | 142 | 161 |
| 共同利用・委託先による収集・利用 | 322 | 320 | 292 |
| 参照・編集・削除 | 184 | 127 | 446 |
| ポリシー変更 | 22 | 25 | 116 |
| データセキュリティ | 230 | 82 | 82 |
| データの保管期間 | 5 | 2 | 5 |
| ユーザによる制御 | 0 | 4 | 21 |
| 特別な配慮が必要なユーザ | 4 | 1 | 15 |
| その他 | 227 | 320 | 513 |
| 合計 | 1,646 | 1,451 | 2,067 |

**表 4**　業界別プライバシーポリシーのデータ数と単語数

| 業界 | URL 数 | ポリシー数 | 単語数 | | |
|---|---|---|---|---|---|
| | | | 最小 | 最大 | 平均 |
| 小売 | 303 | 294 | 196 | 7,378 | 1,147.3 |
| 金融 | 125 | 125 | 171 | 8,892 | 1,951.3 |
| 商社 | 246 | 226 | 150 | 5,559 | 1,053.8 |
| 情報通信 | 390 | 388 | 179 | 14,429 | 1,645.5 |
| 公共 | 35 | 33 | 157 | 4,460 | 1,299.8 |
| 建設・不動産 | 217 | 212 | 171 | 6,330 | 1,194.0 |
| 航空運輸 | 130 | 118 | 172 | 5,647 | 943.5 |
| 機械・製造 | 609 | 517 | 160 | 6,370 | 1,022.8 |
| 医療・健康 | 156 | 132 | 173 | 3,981 | 1,003.9 |
| エナジー | 19 | 18 | 201 | 5,056 | 2,210.0 |
| その他 | 497 | 482 | 177 | 13,863 | 1,331.5 |
| 合計 | 2,727 | 2,545 | 150 | 14,429 | 1,260.5 |

#### 4.1.3　テストデータセット

Hoovers D&B [16] に登録されている東証上場企業 3,273 社のコーポレートサイトのトップ URL (重複を除き 3,267 件) にアクセスし，リンク先の URL やリンクテキスト，取得したコンテンツタイトルに “privacy policy”，“プライバシーポリシー”，“個人情報保護” 等のキーワードが含まれるコンテンツを収集する．その結果，合計 2,674 件のプライバシーポリシーを収集できた．

### 4.2　本文抽出結果

テストデータに含まれるコンテンツからプライバシーポリシー本文を抽出する．リンクのみの短い本文が存在したため，単語数が 150 語に満たない 129 サイトはノイズとして除外した結果，2,545 件の本文を抽出できた．抽出結果の詳細を**表 4** に示す．抽出できたプライバシーポリシー本文の長さは企業ごとに異なる傾向を示した．本文に含まれる平均単語数は，航空運輸業界が最も少なく 943.5 語であり，エナジー業界が最も多く 2210.0 語であった．

### 4.3　分類モデルの精度評価結果

図 3 の畳み込み層において，ハイパーパラメータである kernel サイズおよび batch 数の最適値を探索する．まず，上位カテゴリの分類モデルを用いて，batch 数を 20 に固定し，kernel サイズごとの精度を比較した．具体的には，

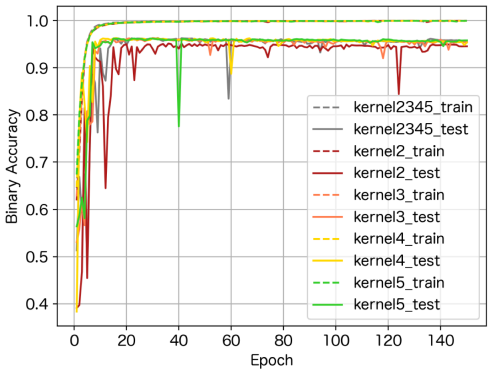


**図 4**　上位カテゴリ分類モデルの kernel 探索

kernel サイズ 2，3，4，5 のフィルター 1 つ使用した場合と，1 つずつ計 4 つ使用した場合の合計 5 通りのモデルを作成した．それぞれのモデルで，epoch 数 50，100，150 の時の精度を比較した結果を**図 4** に示す．高い分類精度と短い学習時間を両立する kernel サイズ 4 を採用した．なお，kernel サイズ 4 および epoch100 の時，F 値による分類精度は 0.77 であった．次に，kernel サイズを 4 に固定し，batch 数 10，20，40，80 と epoch 数 50，100，150 の時の精度をそれぞれ比較した．kernel 探索と同様，分類精度と学習時間を考慮し，batch 数 40 および epoch100 を採用した．なお，batch 数 40 および epoch100 の時，F 値による分類精度は 0.78 であった．

下位カテゴリの分類モデルも同様に最適なハイパーパラメータを探索する．分類対象であるプライバシーポリシー本文は上位カテゴリと同じであるため，上位カテゴリ分類と同様に kernel サイズは 4 を採用した．batch 数 と epoch 数は，下位カテゴリの数がそれぞれ異なるため，モデルごとに最適な値を探索した．上位カテゴリの分類モデルにおけるハイパーパラメータ探索の時と同様に実施した．その結果，batch 数は各下位カテゴリのラベル数と同じ値，epoch 数は 150 が最適であった．なお，下位カテゴリの分類モデル 7 件の F 値による分類精度は，平均 0.83，最大値 0.97，最小値 0.67 であった．

### 4.4　カテゴリ分類結果

#### 4.4.1　上位カテゴリの分類結果

業界ごとに，各上位カテゴリのラベルが存在したプライバシーポリシーの割合を**表 5** に示す．9 割以上のプライバシーポリシーにおいて “ファーストパーティによる情報の取り扱い” や “データセキュリティ” についての記述されていることがわかる．一方で，“ポリシー変更” に関する記述は，情報通信業界では 7 割以上，公共や金融，小売業界では 5 割以上のプライバシーポリシーで記述されていたが，エナジー業界や航空運輸業界では 4 割に満たなかった．表 1 の業界別ガイドラインが定められていると，ガイドラインへの準拠を確認する際にプライバシーポリシーの内容

**表 5**　各上位カテゴリを含むプライバシーポリシーの割合 (90%以上を緑字，50%以下を赤字)

| | 小売 | 金融 | 商社 | 情報通信 | 公共 | 建設・不動産 | 航空運輸 | 機械・製造 | 医療・健康 | エナジー | その他 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ファーストパーティによる収集・利用 | 99.7 | 99.2 | 99.1 | 98.4 | 100 | 99.1 | 97.5 | 98.8 | 98.5 | 100 | 99.4 |
| サードパーティによる収集・利用 | 85.4 | 79.0 | 84.4 | 78.8 | 78.8 | 88.2 | 82.2 | 87.2 | 83.3 | 83.3 | 85.4 |
| 共同利用・委託先による収集・利用 | 89.5 | 95.2 | 90.7 | 96.1 | 97.0 | 92.0 | 87.3 | 87.4 | 84.8 | 88.9 | 92.7 |
| 参照・編集・削除 | 86.4 | 89.5 | 77.3 | 84.5 | 87.9 | 87.3 | 75.4 | 84.3 | 85.6 | 77.8 | 84.3 |
| ポリシー変更 | 52.4 | 53.2 | 46.2 | 71.5 | 54.5 | 47.2 | 39.8 | 46.5 | 47.0 | 38.9 | 60.3 |
| データセキュリティ | 94.2 | 96.0 | 96.0 | 96.1 | 90.9 | 93.9 | 95.8 | 93.6 | 92.4 | 100 | 94.6 |
| データの保管期間 | 5.78 | 3.23 | 6.22 | 12.4 | 6.06 | 4.72 | 5.08 | 7.17 | 3.03 | 5.56 | 8.37 |
| ユーザによる制御 | 20.4 | 12.1 | 19.6 | 31.9 | 33.3 | 15.1 | 16.1 | 19.2 | 10.6 | 55.6 | 20.9 |
| 特別な配慮が必要なユーザ | 4.76 | 2.42 | 3.11 | 5.70 | 12.1 | 0.47 | 2.54 | 7.75 | 4.55 | 0.00 | 4.39 |
| その他 | 100 | 99.2 | 99.6 | 99.2 | 100 | 100 | 98.3 | 99.8 | 99.2 | 100 | 99.6 |

**表 6**　各下位カテゴリを含むプライバシーポリシーの割合（80%以上を太字）

| 上位カテゴリ | 下位カテゴリ | 小売 | 金融 | 商社 | 情報通信 | 公共 | 建設・不動産 | 航空運輸 | 機械・製造 | 医療・健康 | エナジー | その他 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ファーストパーティ | 情報カテゴリ | 71.4 | 75.8 | 68.9 | 73.6 | 63.6 | 73.1 | 67.8 | 74.0 | 73.5 | **88.9** | 74.7 |
| ファーストパーティ | 目的 | **86.1** | **90.3** | **83.1** | **87.8** | **90.9** | **88.2** | 79.7 | **83.9** | **82.6** | **88.9** | **86.8** |
| サードパーティ | 目的 | 38.4 | 40.3 | 33.3 | 48.7 | 51.5 | 37.7 | 32.2 | 40.5 | 37.1 | 55.6 | 37.9 |
| 共同利用・委託先 | 責任者 | 26.5 | 48.4 | 28.0 | 40.9 | 45.5 | 22.6 | 18.6 | 28.5 | 30.3 | 61.1 | 41.6 |
| 共同利用・委託先 | 情報カテゴリ | 48.0 | 65.3 | 51.1 | 76.7 | 57.6 | 50.5 | 41.5 | 42.4 | 41.7 | 77.8 | 58.6 |
| 共同利用・委託先 | 目的 | 36.7 | 57.3 | 39.6 | 50.3 | 48.5 | 41.5 | 29.7 | 32.9 | 30.3 | 66.7 | 41.6 |
| 共同利用・委託先 | エンティティ | 69.4 | **85.5** | 77.3 | **86.5** | **90.9** | 76.4 | 74.6 | 72.1 | 69.7 | **83.3** | **80.3** |
| 参照・編集・削除 | 手続き | 38.1 | 59.7 | 37.8 | 50.0 | 45.5 | 38.7 | 33.9 | 32.6 | 37.9 | 72.2 | 42.3 |
| ポリシー変更 | 通知方法 | 30.6 | 27.4 | 26.7 | 27.7 | 12.1 | 24.1 | 23.7 | 32.4 | 31.1 | 11.1 | 33.5 |
| データセキュリティ | セキュリティ対策 | **88.4** | **83.9** | **90.2** | **90.7** | **84.8** | **84.9** | **94.1** | **89.9** | **90.2** | **94.4** | **89.7** |
| その他 | 問い合わせ先 | 66.3 | **84.7** | 75.6 | **84.5** | **84.8** | 72.2 | 66.1 | 68.8 | 74.2 | 77.8 | 78.9 |
| その他 | その他のプラクティス | 24.5 | 33.1 | 27.1 | 27.7 | 27.3 | 23.6 | 36.4 | 24.8 | 24.2 | 50.0 | 27.6 |

も同時に見直されると考えられ，他業界よりもポリシー変更の頻度や記述が多いと推測できる．また，“データ保管期間” に関する内容は，情報通信業界で 10%以上で記述されていたが，他業界では 1 割に満たなかった．情報通信業界では，ガイドラインにてデータの消去が義務付けられているため，他業界よりデータ保管期間に関する規則が定められている企業が多く，プライバシーポリシー内でデータ保管期間について言及する傾向があると考えられる．

#### 4.4.2　下位カテゴリの分類結果

業界ごとに各下位カテゴリの記載があったプライバシーポリシーの割合を**表 6** に示す．なお，下位カテゴリは数が多いため，少なくとも 1 つの業界において 80%以上の記載があった下位カテゴリ，または業界間の記載割合が 20%以上差のある下位カテゴリのみを抜粋する．

前節で記述が多かった “ファーストパーティによる情報の取り扱い” および “データセキュリティ” の上位カテゴリの中で，特に記載されることが多かった下位カテゴリは，“ファーストパーティによる情報の利用目的” と “セキュリティ対策” であった．また，金融，情報通信，公共業界では問い合わせ先の記載があるプライバシーポリシーが 80%を超えた．これらの結果も，表 1 のガイドラインにおける推奨事項を反映していると考えられる．

エナジー業界のプライバシーポリシーは，他業界と比べて幅広いカテゴリを網羅していることが分かる．特に，他業界と比べて共同利用に関する記載が充実していた．エナジー業界では，サービス提供のために顧客情報の共同利用が必須であり，共同利用の目的となる業務内容や共同利用先のエンティティがガス事業法 [17] で定められていることが要因であると考えられる．

#### 4.4.3　値の検出結果

プライバシーポリシーの記載内容の具体性および充実度を調査するため，各下位カテゴリごとにプライバシーポリシーあたりの値の検出数を算出する．各下位カテゴリの属する値の検出総数を，当該下位カテゴリを含むプライバシーポリシー数で割ることにより，値の平均検出数を求めた．その結果を**表 7** に示す．なお，下位カテゴリは数が多いため，いずれかの業界で 平均 3 件以上の値が検出された下位カテゴリのみ抜粋する．全業界においてセキュリティ対策の値は 8 件以上検出されており，“データへのアクセス制御” や “匿名加工情報の識別の禁止” が多く記述されていた．中でも金融，情報通信，公共業界では 10 件以上の値が検出され，記載内容が充実していることがわかる．これらプライバシーポリシーの具体性の高さは，表 1 のガイドラインにおける説明の充実度が影響していると考えられる．また，前節の結果と同様，エナジー業界はそのビジネス特性より，共同利用する具体的な目的や情報カテゴリに関する値が多く検出されており，記載内容の充実度も高いことがわかる．

**表 7** 各下位カテゴリを含むプライバシーポリシーにおける値の平均検出数 (値の検出数を種類数で割った値が 0.7 以上を太字)

| 上位カテゴリ | 下位カテゴリ | 値の種類数 | 小売 | 金融 | 商社 | 情報通信 | 公共 | 建設・不動産 | 航空運輸 | 機械・製造 | 医療・健康 | エナジー | その他 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ファーストパーティ | アクション | 9 | - | 2.0 | 0.0 | 1.6 | - | 2.3 | - | 3.0 | - | - | 0.4 |
| ファーストパーティ | 目的 | 10 | 3.8 | 4.8 | 3.9 | 5.0 | 3.1 | 4.7 | 4.0 | 3.7 | 3.7 | 5.5 | 3.8 |
| 共同利用・委託先 | 情報カテゴリ | 15 | 1.6 | 2.3 | 0.8 | 1.1 | 0.7 | 0.8 | 1.6 | 1.2 | 1.1 | 7.2 | 1.6 |
| 共同利用・委託先 | 目的 | 9 | 1.6 | 2.3 | 1.3 | 2.1 | 1.3 | 2.3 | 1.8 | 1.3 | 1.8 | 3.4 | 1.7 |
| 共同利用・委託先 | エンティティ | 4 | 2.7 | **3.6** | 2.4 | **2.9** | 2.5 | **3.3** | 2.7 | 2.6 | 2.4 | **3.3** | **4.2** |
| 参照・編集・削除 | ユーザの権限 | 6 | 3.5 | **4.9** | 4.1 | **5.1** | **5.0** | 3.9 | 3.7 | 3.6 | 3.9 | **4.9** | **4.6** |
| データセキュリティ | セキュリティ対策 | 12 | **9.7** | **10.3** | **8.5** | **10.0** | **10.1** | **9.4** | **9.3** | **9.3** | **9.6** | **9.6** | **10.0** |
| データの保管期間 | 保管期間 | 4 | 2.6 | **3.0** | 1.3 | 1.8 | 1.0 | 2.0 | 1.8 | 1.8 | 2.0 | 2.0 | 2.1 |

### 4.5 法遵守の確認結果

プライバシーポリシーのカテゴリ分類結果を表 2 に記載した条件事項および公表事項の論理式に適用し，法律への遵守率を算出する．法案ごとの遵守率を算出し，要件の種類ごとの平均を求めた結果を**図 5** に示す．なお，法律で公表を義務付けられている要件を “必須”，推奨されている要件を “推奨”，内部処理は規定されているが公表に関しては言及されていない要件を “参考” とした．

全業界の平均遵守率は必須項目では 60.5%だったが，推奨項目や参考項目では 50%を下回った．どの企業も法律の必須項目を意識してプライバシーポリシーを作成していることが考えられる．業界別の必須項目における平均遵守率は，金融および公共業界が 70%を超え，次いで医療業界が 68%と高かった．一方で，情報通信や建設・不動産，航空運輸，エナジー業界は 50%前後と低かった．前節と同様に具体的なガイドラインの存在が企業の対応を促進させ，法遵守率向上に繋がったと考えられる．

必須項目のうち，半数以上のプライバシーポリシーが条件事項を満たした法案は法案 18 条 1 項と法案 23 条 2 項の 2 件であった．法案 18 条 1 項ではファーストパーティにおける個人情報収集の目的の公表が要求されており，この要求を満たした割合（すなわち，遵守率）は 86.7%であった．一方，法案 23 条 2 項では第三者提供に関する情報の公表が要求されており，遵守率は 4.9%であった．

推奨項目では，半数以上のプライバシーポリシーが条件事項を満たした法案は 5 件あったが，いずれもその遵守率は 20%を下回った．法遵守できていないポリシーは，ポリシー変更，共同利用，開示・訂正・利用停止等の請求手続きや事業者に関する記載が不十分である傾向があった．

参考項目では，半数以上のプライバシーポリシーが条件事項を満たした法案は 17 件あった．そのうち遵守率 70%を超えた法案は 4 件で，ファーストパーティによる収集・利用目的や セキュリティ対策，問い合わせ先 に関する記述が充実していた．一方で，遵守率 50%を下回った法案は 8 件あり，セキュリティ対策における委託先の管理や従業員の教育，データの正確性の確保，ユーザの権限における利用停止の権利や第三者提供停止の権利，請求に応じない

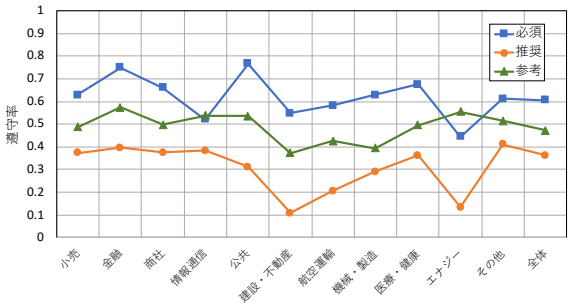


**図 5** 業界別の法遵守率

理由の説明，そしてデータ保管期間等の詳細な記述が不足していた．

## 5. 議論

### 5.1 ガイドラインとプライバシーポリシー

表 1 に示した業界別の個人情報取り扱いガイドラインにおいて，プライバシーポリシーで公表すべき事項を具体的に定めている業界では，それら項目の記載割合が高かった．これらガイドラインに従うことにより，プライバシーポリシーの記載内容の見直しや更新頻度が多くなったとともに，記載内容も充実したことが要因と考えられる．したがって，具体的なガイドラインの規程は法遵守を促進するために効果的であると考えられる．

### 5.2 業界特性とプライバシーポリシー

業界ごとのビジネスの性質や業務に関する法律が，プライバシーポリシーの記載の過不足に影響を与えていた．例えば，エナジー業界では共同利用を必須とする業務が存在するため，共同利用に関する記載量が増えたり，法律を引用した言葉の定義や説明をしたりしている企業も存在した．

### 5.3 本研究の制約と課題

本評価で使用した学習データセットは日本企業 64 社のコーポレートサイトのプライバシーポリシーのみを対象としたため，一部のカテゴリに属するデータが不足し，それらのカテゴリ分類の精度が下がった．正確に法遵守状況を評価するためには，カテゴリ分類の精度が十分に高い必要

があり，学習データセットの拡張は今後の課題である．

また，提案手法の値の検出では単語マッチングを用いたため，その検出精度はマッチングに用いた単語リストに依存する．本評価では学習データセットから TF-IDF を用いて特徴的な単語を抽出したが，値ラベルによって単語リストの単語数に差があった．誤検出や見逃しを最小限にするために，単語リストの拡張や更新は今後の課題である．

本評価では，現行法の要求事項に基づき論理式を作成し，法遵守率を算出した．個人情報保護法は三年おきに更新されるため，提案手法の論理式も更新が必要である．

## 6. 関連研究

### 6.1 プライバシーポリシーの内容分析

これまで多くの研究者がプライバシーポリシーを分析してきた．Wilson ら [11] は，英語のプライバシーポリシー 115 件を分析し，記載内容から 10 種類のカテゴリを特定した．これらプライバシーポリシーに対して，特定したカテゴリのラベル付けを行い，OPP-115 データセットとして公開している．Harkous らは上記 OPP-115 データセットおよび CNN を活用し，プライバシーポリシーの記載内容を分類する手法を提案しており [12]，その結果を可視化するツールを提案している [18].

上記の既存手法はいずれも英語のプライバシーポリシーを対象としているが，本研究ではこれら手法を参考に，日本語のプライバシーポリシーを分析している．既存の分類カテゴリ [11] を日本の法律向けに更新するとともに，日本語のプライバシーポリシー向けの新しい学習データセットならびに分類モデルを作成している．

### 6.2 プライバシーポリシーの法遵守

EU の GDPR を代表例に，プライバシー関連法規制の制定に伴って，これら法への遵守について調査されている．GDPR 施行前後に EU 内のウェブサイトを調査した結果，15.7%のウェブサイトがプライバシーポリシーを新たに規程・公表したとともに，72.6%が更新したと報告されている [19]．また，プライバシーポリシーの記載内容を分類し，その結果を GDPR の要求事項と比較することにより，法遵守状況を確認する調査も行われている [1], [7].

本研究は上記研究を参考に，日本の個人情報保護法の要求事項を論理式で表すことにより，国内企業のプライバシーポリシーにおける法遵守を評価している．

## 7. おわりに

本研究では，日本企業のプライバシーポリシーにおける個人情報保護法の遵守率を調査した．その結果，多くのプライバシーポリシーでファーストパーティによる個人情報の取り扱いやデータセキュリティに関する十分な記述を確認できたが，第三者提供に関する記述は不十分であることがわかった．また，各業界におけるガイドラインの有無やビジネスの性質に起因して，業界間で記載の過不足に差が生じることもわかった．これら業界ガイドラインの分析により，プライバシーポリシーで公表すべき内容の具体性が法遵守率の向上に寄与することを示唆する結果が得られた．各企業によるプライバシーポリシーの充実化は重要だが，個人情報保護委員会や各業界がガイドラインを整備することも全体の法遵守率向上のために重要であると考える．

### 参考文献


[1] T. Linden *et al.*, “The privacy policy landscape after the GDPR,” *Proc. Priv. Enhancing Technol.*, vol. 2020, no. 1, pp. 47–64, 2020.

[2] 総務省, “電気通信事業における個人情報保護に関するガイドライン.” https://www.soumu.go.jp/main_content/000507466.pdf.

[3] 個人情報保護委員会 and 金融庁, “金融分野における個人情報保護に関するガイドライン.” https://www.ppc.go.jp/files/pdf/kinyubunya_GL.pdf.

[4] 個人情報保護委員会 and 厚生労働省, “医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス.” https://www.ppc.go.jp/files/pdf/01_iryoukaigo_guidance3.pdf.

[5] 個人情報保護委員会 and 厚生労働省, “健康保険組合等における個人情報の適切な取扱いのためのガイダンス.” https://www.ppc.go.jp/files/pdf/03_kenpokumiai_guidance3.pdf.

[6] 総務省, “信書便事業分野における個人情報保護に関するガイドライン.” https://www.soumu.go.jp/main_content/000485167.pdf.

[7] S. Liu *et al.*, “Have you been properly notified? automatic compliance analysis of privacy policy text with GDPR article 13,” in *WWW*, pp. 2154–2164, 2021.

[8] B. Andow, “Htmltoplaintext.” https://github.com/benandow/HtmlToPlaintext.

[9] T. Kudo, “Mecab: Yet another part-of-speech and morphological analyzer.” https://taku910.github.io/mecab/.

[10] F. Inc., “fasttext.” https://fasttext.cc/.

[11] S. Wilson *et al.*, “The creation and analysis of a website privacy policy corpus,” in *ACL*, 2016.

[12] H. Harkous *et al.*, “Polisis: Automated analysis and presentation of privacy policies using deep learning,” in *USENIX Security Symposium*, pp. 531–548, 2018.

[13] mozilla, “Dnt.” https://developer.mozilla.org/ja/docs/Web/HTTP/Headers/DNT.

[14] M. Simon, “Apple is removing the do not track toggle from safari, but for a good reason.” https://www.macworld.com/article/232426/apple-safari-removing-do-not-track.html.

[15] 日本経済新聞, “日経会社情報 digital: 日経電子版.” https://www.nikkei.com/nkd/.

[16] Dun & Bradstreet, Inc., “D&b hoovers™.” https://www.dnb.com/products/marketing-sales/dnb-hoovers.html.

[17] 経済産業省, “ガス事業法.” https://elaws.e-gov.go.jp/document?lawid=329AC0000000051.

[18] H. Harkous *et al.*, “Polisis.” https://pribot.org/.

[19] M. Degeling *et al.*, “We value your privacy ... now take some cookies: Measuring the gdpr’s impact on web privacy,” in *NDSS*, 2019.