



PIO-NET を利用した消費者問題の傾向分析 ～テキストマイニングを用いた時系列データのトピック比較～

リサーチ・ディスカッション・ペーパー
RESEARCH DISCUSSION PAPER

谷岡 広樹、徳永 欽也、徳永 美和子、伊藤 友基、金子 時佳
宮本 麗子、小早川 優、小林 なずな、大原 海里

2023 年 9 月

消費者庁新未来創造戦略本部国際消費者政策研究センター
International Consumer Policy Research Center (ICPRC)
Strategic Headquarters for Frontiers of Consumer Policy
Consumer Affairs Agency, Government of Japan

本リサーチ・ディスカッション・ペーパーは、全て研究者個人の責任で執筆されており、消費者庁の見解を示すものではありません。
(問合せ先：088-600-0000)

本リサーチ・ディスカッション・ペーパーは、新未来創造戦略本部国際消費者政策研究センターの研究者等により行われた研究の成果を取りまとめたものです。学界、研究機関等の関係する方々から幅広くコメントを頂き、消費者行政における政策立案の基礎資料として役立つことを意図して発表しております。なお、全て研究者個人の責任で執筆されており、消費者庁の見解を示すものではありません。

PIO-NET を利用した消費者問題の傾向分析 ～テキストマイニングを用いた時系列データのトピック比較～

谷岡 広樹^{1),2)}、徳永 欽也⁴⁾、徳永 美和子⁴⁾、伊藤 友基³⁾、金子 時佳⁵⁾
宮本 麗子³⁾、小早川 優⁴⁾、小林 なずな³⁾、大原 海里³⁾

1) 徳島大学大学院創成科学研究科理工学専攻講師

2) 消費者庁新未来創造戦略本部国際消費者政策研究センター客員主任研究官

3) 消費者庁新未来創造戦略本部国際消費者政策研究センター特任研究員

4) 消費者庁新未来創造戦略本部国際消費者政策研究センター特任研究員（～2023年3月末）

5) 消費者庁新未来創造戦略本部国際消費者政策研究センター特任研究員（～2023年8月末）

概要

本稿は、2021年1月から2021年12月の1年間に受け付け、全国消費生活情報ネットワークシステム（PIO-NET）へ登録された相談内容のうち、新型コロナウイルス感染症関連の話題について抽出したものについて独自に分析し、消費者問題の傾向分析を行ったものである。PIO-NETの相談内容に含まれる話題の分析には、統計分析手法のトピックモデリングを用いる。話題の分類と特徴的なキーワードの抽出を行い、月ごとの新型コロナウイルス感染症関連の主な出来事について分析を行った。その結果、新型コロナウイルス感染症の流行が継続し、会社や家賃といった収入に関する相談や、ワクチン接種のような国の施策に関連する相談が目立った。同様の手法で2020年1月から2020年12月の1年間に受け付けた相談内容について分析を行った結果と比較した結果、新型コロナウイルス感染症の流行が始まったばかりの混乱期である2020年に多かったマスク等の商品の品不足関連の相談は2021年には目立たなくなり、送り付け商法等に関する相談や副業や会社、家賃といった経済的な理由に基づく相談が多くなったことが分かった。国の施策に関する相談についても、2020年に多かった給付金やGo To トラベル等の相談に代わり、2021年にはワクチン接種関連の相談が増加した。さらに、2020年1月から2021年12月の2年間の相談内容について、トピックモデリングで得られたトラブルキーワードを用いて、連続する2ヶ月のトピックの類似度を計算することにより、トピックの変動を分析し、新型コロナウイルス感染症に関連する主な出来事との関連を考察した結果、緊急事態宣言の発出の前後では、大きくトピックの内容が変化している等、国の施策が消費者の相談内容に少なからず影響を与えていることが分かった。

1 はじめに

2020年頃から社会問題となっている新型コロナウイルス感染症の話題は、全国の消費生活センター等及び国民生活センターが受け付けた相談情報を蓄積したデータベースである全国消費生活情報ネットワークシステム¹（PIO-NET; Practical Living Information Online Network System）[1]に登録された相談情報にも様々な形で影響があると考えられる。

文献[2]の2020年1月から2020年12月までの1年間に受け付け、PIO-NETへ登録された新型コロナウイルス感染症関連の相談情報について、テキストマイニング²の手法のひとつであるトピックモデリング³を適用した分析結果によると、個別の話題に関する報道のタイミングと、消費生活相談の増減に大きな傾向の違いはみられなかったが、4月下旬に施策が発表された特別定額給付金等について、相談は1か月遅れでトピックとして現れるなど、政策の発表から相談として現れるまで、若干のタイムラグがあることが示唆されたが、ワクチン接種の開始や行動に関する規制や緩和など、様々な政策が講じられてきた2021年以降の消費生活相談について、テキスト分析等により傾向分析を行った研究はなされていない。

そこで本研究では、2021年1月から2021年12月までの1年間に受け付け、PIO-NETへ登録された新型コロナウイルス感染症関連の相談情報について分析を行い、2020年1月から2020年12月までの1年間に受け付け、PIO-NETへ登録された新型コロナウイルス感染症関連の相談情報の分析結果との比較を行う。さらに、新型コロナウイルス感染症関連の相談内容について、連続する2ヶ月のトピックの類似度の変化を定量的に表すことでトピックの時系列的な変遷を明らかにする。新型コロナウイルス感染症のように、突発的に発生した事案に伴う消費者被害の早期把握は難しいが、本手法を用いてトピックの時系列的な変遷を把握することができれば、今後、突発的に事案が生じた際に、システム的に消費者問題の早期発見が実現できるのではないかと考えられる。

2 研究方法

2.1 分析対象データ

本研究では、2021年1月から2021年12月までの1年間に受け付け、2022年3月31日までにPIO-NETに登録されている相談情報のうち、「新型コロナウイルス感染症関

¹ 全国消費生活情報ネットワークシステム（PIO-NET）とは、国民生活センターと全国の消費生活センター等をネットワークで結び、消費者から消費生活センターに寄せられる消費生活に関する苦情相談情報（消費生活相談情報）の収集を行っているシステムである。地方公共団体は、消費生活に関する様々な相談を受け付ける窓口である消費生活センターや消費生活相談窓口を設置している。消費生活センターの消費生活相談員が消費者から受け付けた苦情や相談についての情報が記録されているのがPIO-NETである。PIO-NETには、相談者と契約当事者の情報（「年齢」、「性別」、「職業」等）や、「受付年月日」、「相談概要」などの項目が入力されている。

² テキストマイニングとは、定型化されていない文章の集合からなるテキストデータをフレーズや単語に分解して詳細に解析し、有用な情報を抽出する分析手法のことである。

³ トピックモデリングとは、文章が含む話題（トピック）の確率的なモデルを分析する手法である。

連」の相談情報 38,921 件⁴を対象とする。本相談情報については、国民生活センターから提供を受けた、商品・役務名、件名、相談概要に記載されているテキストデータのみを基に抽出し、独自の分析に使用した。

2.2 分析方法

図 2.1 を用いて分析の流れを説明する。まず、①キーワードの抽出で相談概要のテキストデータを対象に、形態素解析により単語（名詞・固有名詞）を抽出する。ここで取り出された単語のことをキーワードと呼ぶ。次に、②年間のキーワードの頻度を分析し、③月別のキーワードの分析で月別のキーワードの頻度の分析を行う。④年間のトピックモデリングでは、トピックモデルを用いてトピックとキーワードを抽出し、⑤月別のトピックモデリングで月別のトピックとキーワードを抽出する。⑥トピックの分析と⑦トピックの分類を行った後、⑧トピックの時系列変化の分析を行う。

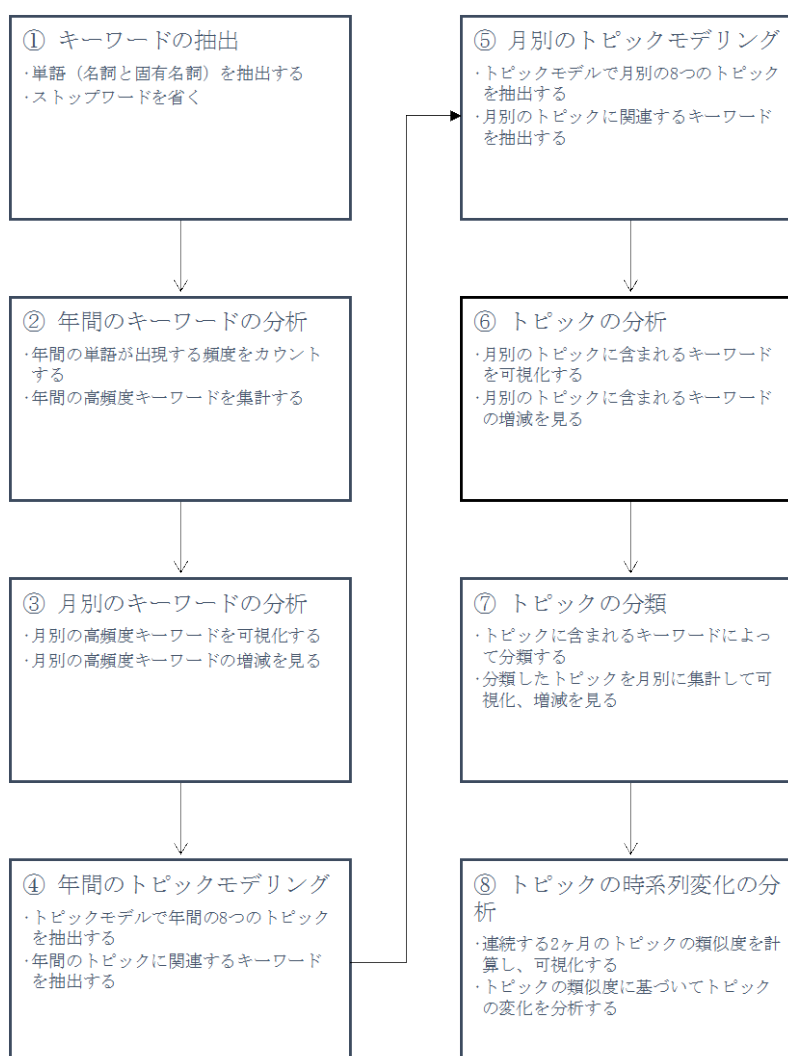


図 2.1：PIO-NET データの分析の流れ

⁴ 件名または相談概要のいずれかに文字列「コロナ」が含まれるもの、または指定ワードが「新型コロナ関連」の相談情報（2021年1月～12月受付）の件数。

3 キーワードの分析

3.1 キーワードの抽出

キーワードの分析では、まず PIO-NET データの相談概要のテキストデータを形態素解析⁵する。形態素解析には spaCy v2.3.5 (GiNZA, Sudachi) [3][4][5] を用いる。このとき、品詞が名詞と固有名詞の単語のみを採用する。また、分析に不適切と考えられる単語であるストップワード⁶を除いた結果をキーワードとする。ストップワードには SlothLib⁷ [6]の日本語ストップワードリスト 328 語に加え、出現頻度が 1 回の単位や指示語、代名詞等の単語及び新型コロナウイルス感染症関連の話題に共通の単語である「新型コロナ」「コロナウイルス」といった単語を追加して用いた。

表 3.1 及び図 3.1 は、ストップワードを除き、分析対象とした単語（キーワード）及びこのキーワードを含む文章の統計情報である。この表をみると分かるように、2020 年と 2021 年で比較すると、新型コロナウイルス感染症関連の相談件数、文章数及びキーワード数は減少したが、異なり語数は増加していた。このことから、新型コロナウイルス感染症関連の話題が収束しつつあるが、相談内容に多様性が高まった可能性が示唆される。次章以降では、この傾向を確かめながら、その具体的な変動内容について詳しくみていく。

表 3.1：PIO-NET（新型コロナウイルス感染症関連）の統計データ

データ種別	相談件数	文章数	キーワード数	異なり語数 ⁸
2020 年（新型コロナウイルス感染症関連）	90,064 件	878,677 文	16,041,662 語	28,234 語
2021 年（新型コロナウイルス感染症関連）	38,921 件	392,278 文	7,627,084 語	43,742 語

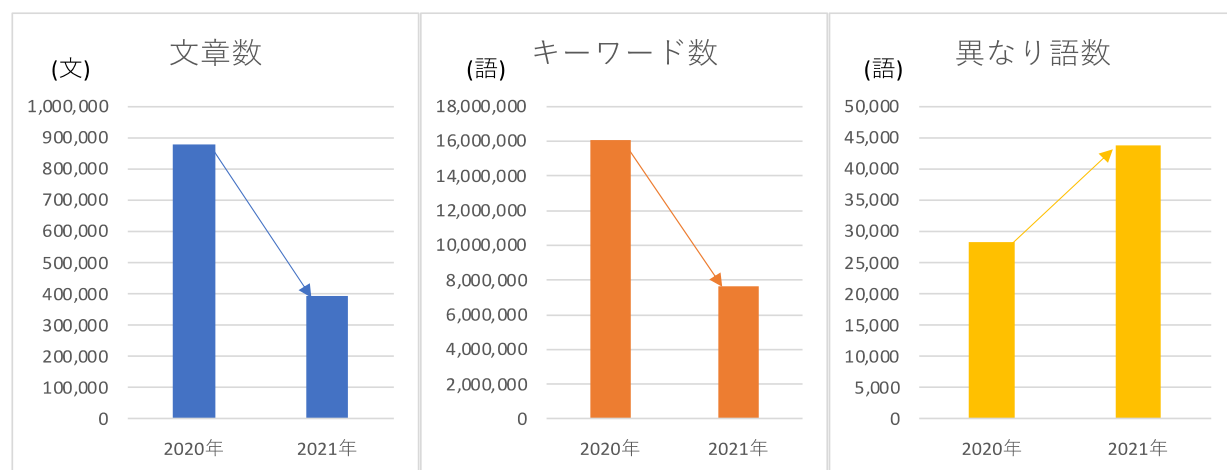


図 3.1：PIO-NET データ（新型コロナウイルス感染症関連）の統計データ(表 3.1)の変化

⁵ 形態素解析とは、文章を単語に分ち書きし、単語の品詞を推定する自然言語処理技術である。

⁶ ストップワードとは、分析に不要として除外する単語のことである。

⁷ SlothLib 日本語ストップワード: <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

⁸ 異なり語数は、解析結果から得られたキーワードの総数から重複するキーワードを省いた単語数である。

3.2 PIO-NET データのキーワード分析

新型コロナウイルス感染症関連の相談内容の傾向を調べるために、2021年1月から2021年12月までの相談内容に含まれる単語の出現頻度の上位20語を高頻度キーワードとして、図3.2に示す。これをみると、2020年に84,310件であった「マスク」は7,539件に減少したことが分かる。

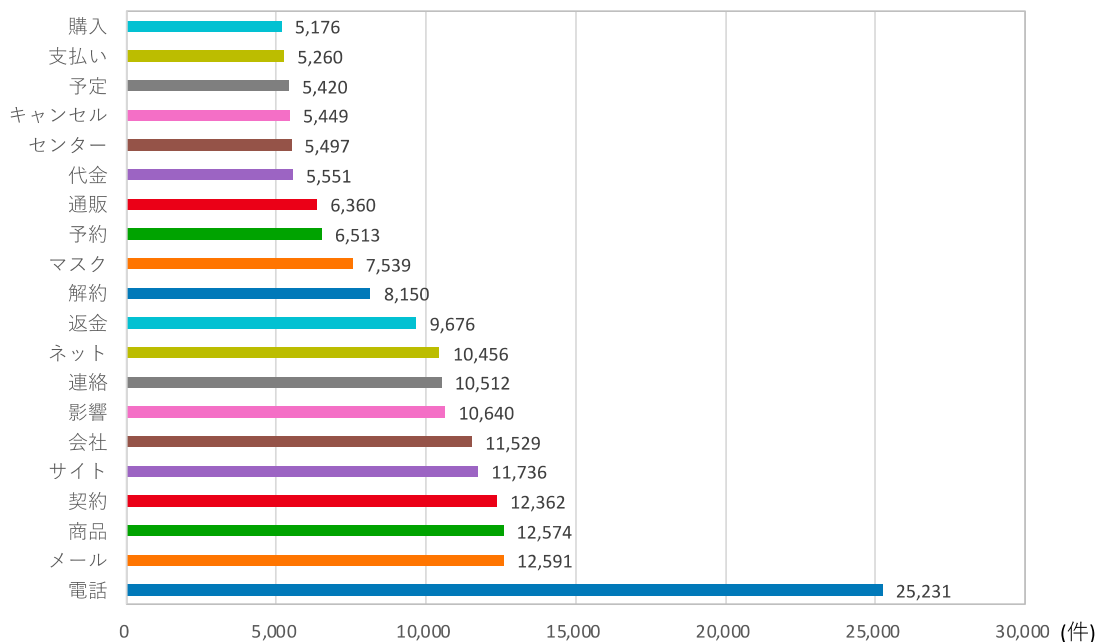


図 3.2 : 2021 年 PIO-NET データ内の高頻度キーワードの出現頻度

図 3.3 は 2021 年 1 月から 2021 年 12 月までの高頻度キーワードの出現頻度を積み上げたグラフである。出現頻度の総数は 2021 年 1 月から 2021 年 12 月にかけて、全体としては減少傾向にあるが、5 月と 8 月・9 月には一時的に増加している。

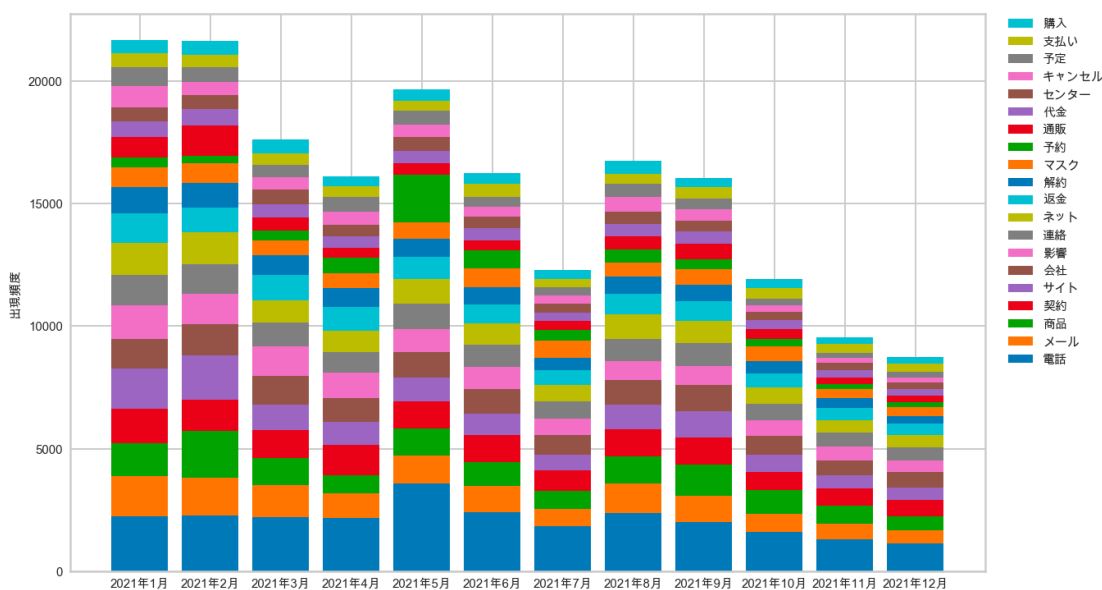


図 3.3 : 2021 年 PIO-NET データ内の高頻度キーワードの月別の出現頻度

5月に増加した理由について、2021年4月から高齢者を対象としたワクチン接種の予約受付が開始し、5月に「ワクチン接種の予約をしたいが、電話が繋がらない」といった相談が増加した背景（参照：図3.4）が考えられる。そして、8月・9月に一時的に増加している理由としては、7月末から8月に国内感染者が増加し、8月末に緊急事態宣言及び蔓延防止重点措置の対象地域が徐々に拡大していったこと（参照：表5.7、表5.8）が背景にあり、新型コロナに関連する相談が増えたのではないかと推測される。

図3.4は2021年1月から2021年12月の高頻度キーワードの出現割合を月別にみたグラフである。キーワード「予約」「電話」が5月頃に増加したことが分かる。

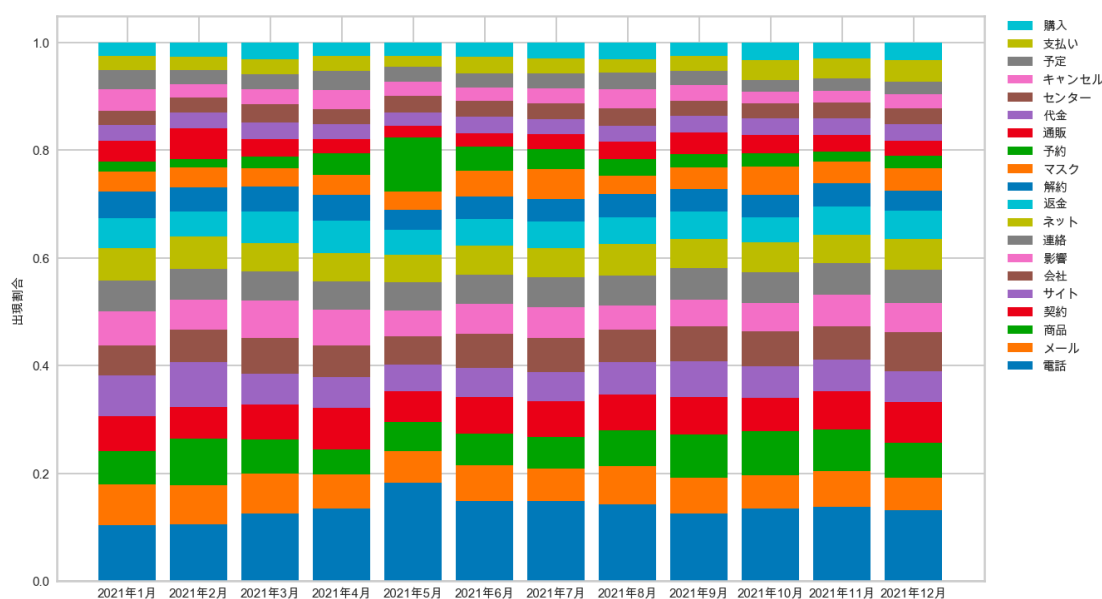


図3.4：2021年PIO-NETデータ内の高頻度キーワードと月別の出現割合

頻度順にまとめたキーワード分析では、大まかな変化はみて取れるが、どのような相談があったのかといった話題についての分析が困難である。そのため、次章以降では、より話題に着目した分析を行うため、トピックモデリングを用いた分析を行う。

4 トピックモデリング

本章では、PIO-NET データに含まれる話題（トピック）をトピックモデルの一種である潜在的ディリクレ配分法（LDA: Latent Dirichlet Allocation）[9][10] を用いて抽出する。トピックモデルとは、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したモデルである。トピックごとに単語の出現頻度分布を想定することで、トピック間の類似性やその意味を解析できる。図 4.1 はその概念を図示したものである。本研究では、単語に相当するものとして PIO-NET の相談概要から抽出したキーワードを用いる。

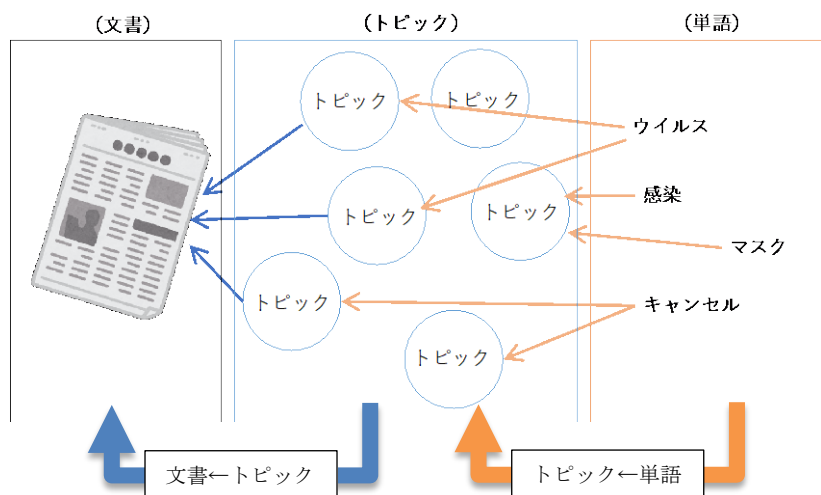


図 4.1 : LDA の概念図 (文献[2]より引用)

図 4.2 は、LDA をグラフィカルモデル⁹で表したものである。ここで N は単語数、 M は文書数を意味する。 φ は各トピックにどの単語が出現するかを表す多項分布であり、 K はトピックの数を意味する。 W は文書に含まれる単語であり、 Z は単語 W に割り当てられたトピック、 θ は各文書がどのトピックに属するかを表す多項分布である。また、 α と β は連続型の確率分布の一種であるディリクレ分布のパラメータである。

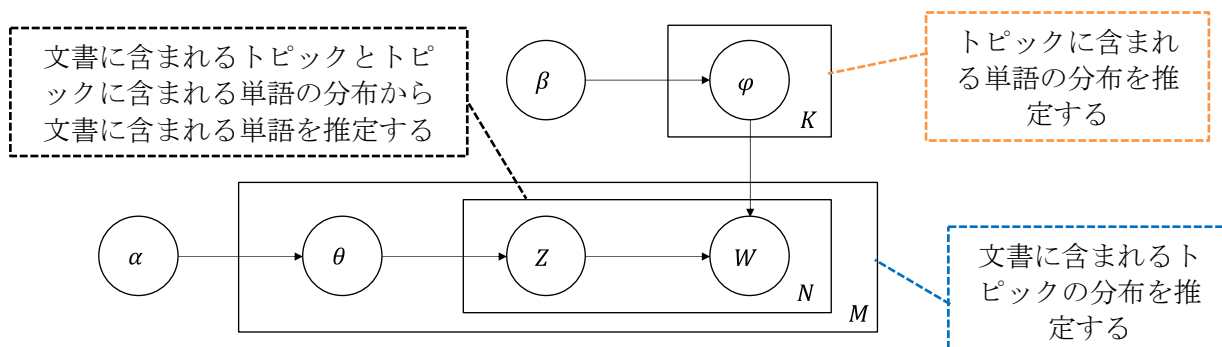


図 4.2 : LDA のグラフィカルモデル

なお、LDA の計算には、文書に含まれるトピックを推定するために単語とその頻度または重要度が必要となる。そのため、PIO-NET の相談概要から形態素解析により抽出されたキーワードの重要度に TF-IDF[11]を用いる。

⁹ グラフィカルモデルとは、対象となる事象を確率的なモデルとして視覚的に表現したものである。

4.1 PIO-NET データのトピックモデリング

PIO-NET データの 2021 年 1 月から 12 月までの 1 年間の相談情報について、相談概要 1 件を文書 1 件として計算し、LDA で分析した結果である 8 つのトピックを図 4.3 に示す。



図 4.3 : 2021 年 PIO-NET 新型コロナ関連のトピック (8 トピック)

各トピックのキーワードから、「結婚式」「契約」「解約」「チケット」「返金」は新型コロナウイルス感染症の拡大の影響でサービス等を受けられなくなった際のキャンセル料や返金に関する消費者トラブル、「商品」「サイト」「マスク」は通販等で購入した商品に関する消費者トラブルに関するトピック、「会社」「家賃」「収入」「返済」は収入減に関連して生じた問題についてのトピック、「ワクチン」は、ワクチン接種開始以降の予約や詐欺被害に関するトピックであると推測される。

なお、図 4.3 の中のトピックについて、大きく 4 つのトピック分類「サービス」「商品」「収入」「国の施策」「その他」のいずれかに関するトピックであると判断し、表 4.1 で示すような色の矩形で囲った。

表 4.1 : トピックの分類と矩形の色

トピック分類	矩形の色	キーワードの例
サービスに関するトピック	オレンジ	返金、キャンセル、解約等
商品に関するトピック	緑	商品、サイト、マスク、通販等
収入に関するトピック	青	仕事、家賃、返済、収入、会社等
国の施策に関するトピック	赤	緊急事態宣言、ワクチン、給付金等
その他のトピック	なし	電話、相談、対策、苦情等

4.2 PIO-NET データのトラブルキーワード

PIO-NET データのトピックとして表れた種類の相談が、1年間を通してどのような傾向であるか、各トピックに含まれるキーワードを用いて分析する。具体的には、2021年1月から2021年12月までの月別のトピック分析を行い、生成された8つのトピックについて、それぞれを構成するキーワードのうち生成確率が最大となる代表キーワード96語（8語×12ヶ月）から、重複を除いた30語を2021年のトラブルキーワードと定義する（表4.2）。

表4.2：2021年PIO-NETデータの月別トピックから抽出したトラブルキーワード

キャンセル	返済
支払い	会社
苦情	結果
工事	チケット
センター	収入
検査	緊急事態宣言
店舗	メール
予約	サイト
解約	返金
病院	契約
家賃	電話
広告	キャンセル料
対策	マスク
感染	商品
結婚式	ワクチン

2020年1年間（文献[2]）と比較して、なくなったトラブルキーワードには、航空券、給付金、消毒液、旅行、教室、定額、外国人、仕事、ポイント、トイレトペーパー、クレジットカード、キャンペーンなどがある。一時的に品不足となった消毒液やトイレトペーパーについての相談や、旅行会社都合で急遽取りやめになった旅行のキャンセル、給付金の受給についての相談については、2021年には目立たなくなっていることが分かる。

一方、2021年に新しく出てきたトラブルキーワードには、支払い、苦情、工事、センター、検査、店舗、予約、病院、返済、結果、収入、緊急事態宣言、ワクチンがある。PCR検査やワクチン接種の予約に関する相談や、緊急事態宣言の発出に伴う事案の相談が増加したことが分かる。結婚式やマスクといったトラブルキーワードは、2020年から2021年にかけて引き続き出現しているが、結婚式については、2020年は開催の延期に関する相談だったものが、2021年は延期の後にキャンセルや再延期をしたといった相談に移行した。マスクについては、2020年は入手できないことや通販に関する相談が多かったが、2021年は送り付け商法に関する相談に移行した。

5 PIO-NET データのトピック月別分析

PIO-NET データと社会情勢の変化を比較するため、2021年1月から2021年12月までを月別にトピックモデリングを用いて分析し、8つのトピックを、それぞれのトピックに含まれる代表的なキーワード20語で表した。また、色のついたキーワードは、トラブルキーワードである。なお、トピックを囲む矩形の色は表4.1の基準に基づく。また、各月のトピックについて、月別の新型コロナウイルス感染症関連の主な話題との関連を考察した。

2021年1月

2021年1月のトピックの代表的なキーワードとしては、「感染」「商品」「契約」「電話」「返金」「マスク」「広告」「キャンセル料」があった。商品やサービスについての契約やキャンセルについての相談が多かったことや、収入減を要因とする副業に関するトラブルなどもみられた。



図 5.1 : 2021年1月のPIO-NETのトピック

表 5.1 : 2021年1月の新型コロナウイルス感染症関連の主な話題

2021年1月
6日：国内の感染者初の1日6000人超
7日：「緊急事態宣言」1都3県に発出
11日：世界の感染者 累計9000万人超える
13日：「緊急事態宣言」11都府県に拡大
13日：外国人の入国を全面停止
16日：世界の死者数 200万人超える
17日：都内の自宅療養者1万人超に 第5波以来 感染急拡大で一気に増加
19日：国内の死者初の1日100人超、重症者1001人超える
20日：厚労省 ワクチン米製薬大手ファイザーと契約を正式締結と発表
27日：世界の感染者累計1億人超える

2021年2月

2021年2月のトピックの代表的なキーワードとしては、「返済」「対策」「マスク」「契約」「商品」「返金」「電話」「キャンセル料」があった。1月に引き続き旅行のキャンセル、結婚式などのイベントのキャンセルに関する相談の他、マスクの送り付けや給付金詐欺などに関連する相談がみられた。また、これらに関連し、キャンセルや返金に関する電話が繋がらないといった内容の相談も目立った。



図 5.2：2021年2月のPIO-NETのトピック

表 5.2：2021年2月の新型コロナウイルス感染症関連の主な話題

2021年2月
2日：「緊急事態宣言」10都府県は延長
3日：消費者庁 ワクチン接種に便乗した不審電話の注意呼びかけ
5日：消費者庁「コロナ禍における消費者被害防止キャンペーンを実施します」
9日：消費者庁「行政機関等をかたった"なりすまし"にご注意」
12日：消費者庁「新型コロナワクチン詐欺 消費者ホットライン」の開設について
13日：改正コロナ特措法 施行

2021年3月

2021年3月のトピックの代表的なキーワードとしては、「チケット」「契約」「マスク」「家賃」「返金」「電話」「キャンセル料」「商品」があった。2月に引き続き、イベント・旅行・商品のキャンセルと返金に関するトラブルが多かったことが分かる。また、マスクを送り付けられて高額な支払いを請求される相談も多くみられた。



図 5.3 : 2021年3月のPIO-NETのトピック

表 5.3 : 2021年3月の新型コロナウイルス感染症関連の主な話題

2021年3月
5日：「緊急事態宣言」1都3県で2週間延長(21日まで)
21日：「緊急事態宣言」1都3県が解除、2か月半ぶりにすべてが解除となる
25日：東京五輪 聖火リレー始まる
26日：消費者庁「研究用抗原検査キット及び抗体検査キット使用についての注意」
26日：消費者庁「新型コロナウイルス感染症の検査キットの販売事業者5社に対する行政指導」

2021年4月

2021年4月のトピックの代表的なキーワードとしては、「ワクチン」「キャンセル料」「返金」「契約」「電話」「病院」「マスク」「サイト」があった。この月から、新たなトピックとして「ワクチン」が表出している。このトピックに関する相談概要としては、新型コロナウイルス感染症のワクチン接種の電話予約が繋がらない、という内容のものが多くみられた。また、3月に引き続き、マスクが送り付けられて支払いを請求されたことについての相談もみられる。



図 5.4：2021年4月のPIO-NETのトピック

表 5.4：2021年4月の新型コロナウイルス感染症関連の主な話題

2021年4月
5日：「まん延防止等重点措置」3府県(大阪、兵庫、宮城)で適用開始
8日：厚労省 新型コロナウイルス感染症影響の「解雇」「雇い止め」見込み含め10万人超える
12日：「まん延防止等重点措置」3都府県(東京、京都、沖縄)で適用開始
12日：高齢者へのワクチン接種開始
20日：「まん延防止等重点措置」4県(埼玉、千葉、神奈川、愛知)で適用開始 10都府県に拡大
25日：「緊急事態宣言」4都府県(東京、大阪、兵庫、京都)に発出
26日：国内の死者 1万人超える
28日：変異ウイルス国内の感染者 1万人超える

2021年5月

2021年5月のトピックの代表的なキーワードとしては、「収入」「ワクチン」「契約」「商品」「電話」「結果」「緊急事態宣言」「マスク」があった。新型コロナウイルス感染症の影響で仕事がなくなり、クレジットカードの支払いができないことや、副業をしようとして高額商品を購入したことについての相談があった。このほか、4月に引き続きワクチン接種に関する相談や、PCR検査の検査キットに関連する相談も多くみられる。



図 5.5 : 2021年5月のPIO-NETのトピック

表 5.5 : 2021年5月の新型コロナウイルス感染症関連の主な話題

2021年5月
12日：「緊急事態宣言」2県(愛知、福岡)を追加し6都府県に拡大
12日：「まん延防止等重点措置」8道県に拡大
14日：農水省「Go To イート」期限を最長12月まで延長決定
16日：「緊急事態宣言」9都道府県に拡大
16日：「まん延防止等重点措置」10県に拡大
16日：消費者庁 ワクチン接種の予約代行詐欺に注意呼びかけ
17日：東京・大阪 ワクチン大規模接種 ネット予約受付開始
21日：「緊急事態宣言」沖縄県を追加し10都道府県に拡大
21日：米モデルナ製と英アストラゼネカ製の新型コロナウイルス感染症ワクチンを正式承認
24日：首都圏1都3県、関西3府県の高齢者を対象のワクチン大規模接種開始
26日：新型コロナウイルス感染症関連の倒産1500社超える
26日：厚労省 ワクチン接種601万人余 85人死亡「重大な懸念認められず」
27日：「緊急事態宣言」9都道府県で延長

2021年6月

2021年6月のトピックの代表的なキーワードとしては、「電話」「収入」「メール」「感染」「解約」「マスク」「ワクチン」「契約」があった。4、5月から引き続きワクチン接種に関連する相談が多かったことが分かる。また、緊急事態宣言の影響による、旅行のキャンセル、旅行や結婚式等のキャンセル無料期間を過ぎてしまったことに対する相談もみられた。



図 5.6：2021年6月のPIO-NETのトピック

表 5.6：2021年6月の新型コロナウイルス感染症関連の主な話題

2021年6月
8日：職場や大学などでのワクチン接種、申請受付開始
10日：大規模接種センター 対象を全国に拡大
20日：「緊急事態宣言」 沖縄県を除き解除、7都道府県は「まん延防止等重点措置」に移行
21日：ワクチン職域接種 本格的に開始、全国17大学でも
23日：自治体の大規模接種と職域接種、新規受付を一時休止
24日：ワクチン接種 国内高齢者の半数超が1回接種

2021年7月

2021年7月のトピックの代表的なキーワードとしては、「キャンセル料」「マスク」「ワクチン」「電話」「検査」「会社」「契約」「メール」があった。ワクチン接種、PCR検査、旅行のキャンセル、不正な請求、ローンの返済、習い事の解約、詐欺サイトといったこれまでも話題に上がった相談内容の他、新たに海産物の購入勧誘に関する相談がみられた。本トピックは、新型コロナウイルス感染症とは関係なく通年でみられるトピックであるが、徐々に新型コロナウイルス感染症に関連した話題として目立つようになってきた。



図 5.7：2021年7月のPIO-NETのトピック

表 5.7：2021年7月の新型コロナウイルス感染症関連の主な話題

2021年7月	
9日	東京五輪 1都3県の全会場で無観客開催へ
9日	「緊急事態宣言」東京4度目の発出、沖縄は延長
17日	五輪選手村で初の陽性者
23日	東京五輪 開幕
26日	ワクチンパスポート 各市町村で申請受付開始
28日	自宅療養者 全国で1万人超える
29日	国内の感染者 初の1日1万人超

2021年8月

2021年8月のトピックの代表的なキーワードとしては、「チケット」「会社」「結婚式」「ワクチン」「電話」「契約」「緊急事態宣言」「マスク」があった。新しい話題としては、無料のはずのワクチン接種に対する料金の請求や、ワクチン接種の予約代行を名乗る業者からの電話に対して、詐欺を疑う相談が増えた。このほか、在宅ワークが増えたことに伴うインターネット回線の契約切り替えに関する相談も増加している。



図 5.8：2021年8月のPIO-NETのトピック

表 5.8：2021年8月の新型コロナウイルス感染症関連の主な話題

2021年8月	
5日	世界の感染者 2億人超える
6日	国内の感染者 100万人超える
8日	「まん延防止等重点措置」8県を追加
8日	東京五輪 閉幕
13日	国内の感染者 初の1日2万人超
20日	「緊急事態宣言」13都府県に拡大、「まん延防止等重点措置」も16道府県に拡大
24日	東京パラリンピック 開幕
27日	「緊急事態宣言」21都道府県に拡大、「まん延防止等重点措置」12県に適用
27日	政府 イベント開催制限

2021年12月

2021年12月のトピックの代表的なキーワードとしては、「工事」「予約」「チケット」「支払い」「電話」「サイト」「マスク」「契約」があった。トピックとしては大きな変化はなかったが、徐々に資材の入荷遅延に伴う工事の遅れに関する相談が増えてきていることが分かった。



図 5.12：2021年12月のPIO-NETのトピック

表 5.12：2021年12月の新型コロナウイルス感染症関連の主な話題

2021年12月
1日：ワクチン3回目 接種始まる
6日：「ワクチンパスポート」で特典 試験的事業始まる 札幌など
20日：アプリでコロナワクチン接種証明始まる
22日：オミクロン国内で初の市中感染を確認
24日：「モルヌピラビル」新型コロナの飲み薬として正式に承認

6 PIO-NET データの時系列分析

6.1 トラブルキーワードの時系列分析

図 6.1 は 2021 年のトラブルキーワードの出現件数を積み上げたグラフである。特徴的なのは、4 月頃から「予約」「ワクチン」が増加し 5 月にピークとなっていることである。

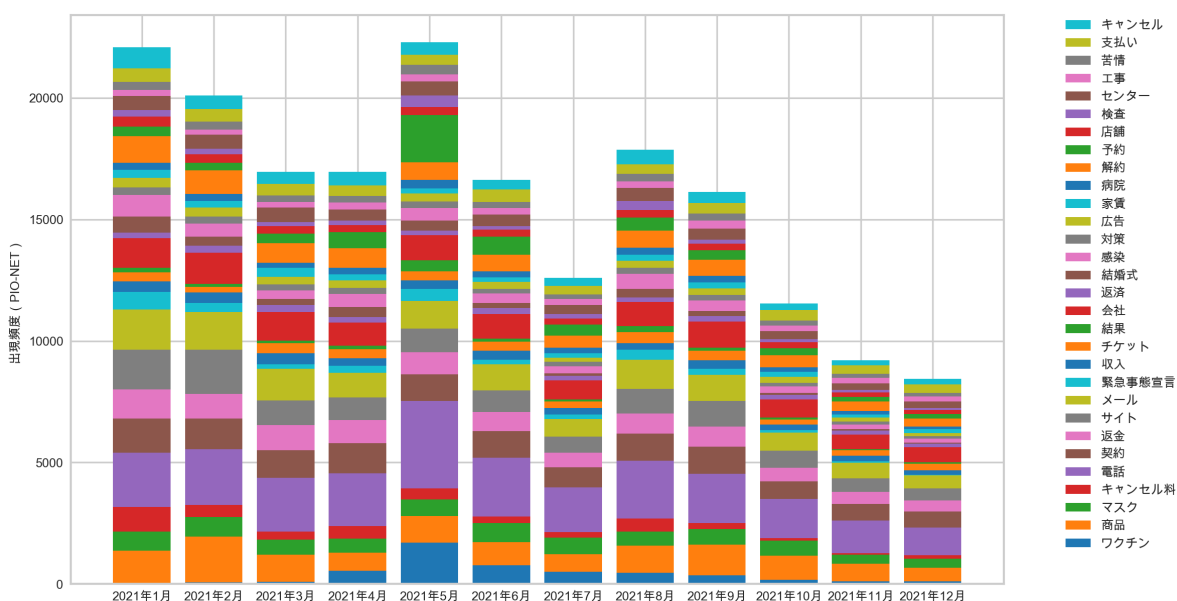


図 6.1 : 2021 年 PIO-NET データ内のトラブルキーワードと月別の出現頻度

図 6.2 は 2021 年のトラブルキーワードの出現割合を積み上げたグラフである。出現割合でも「予約」「ワクチン」が 4 月から増え始め、5 月にピークとなっていることが分かる。

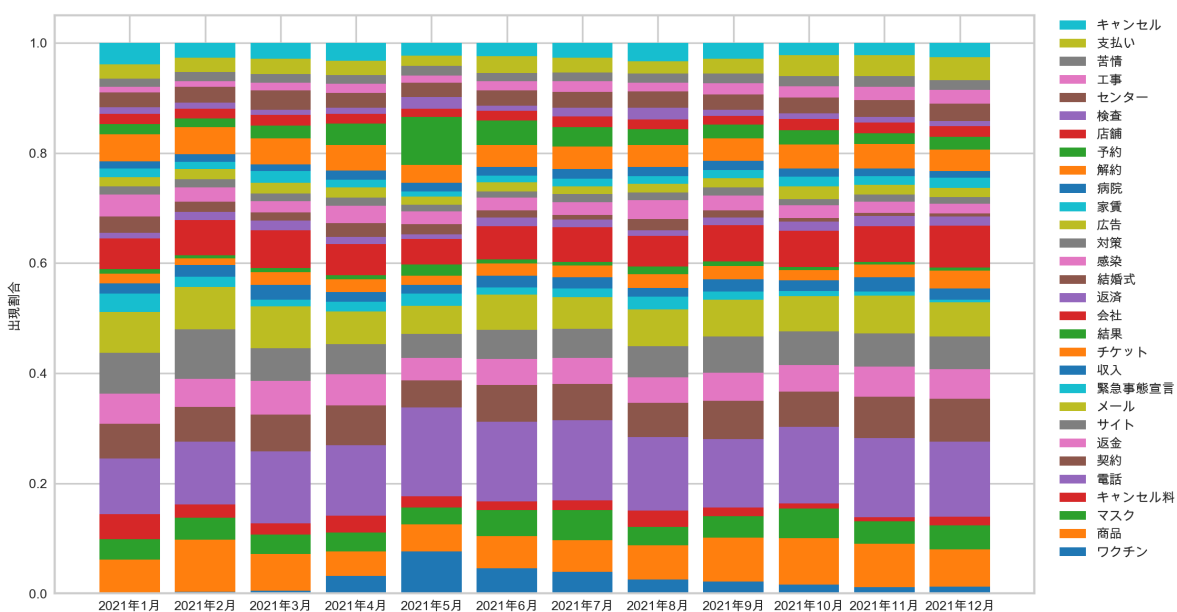


図 6.2 : 2021 年 PIO-NET データ内のトラブルキーワードと月別の出現割合

6.2 トピック分類の時系列分析

トピック分類の時系列分析として、本章では2020年及び2021年のPIO-NETデータに対して分析を行う。まず、図6.3では、文献[2]における2020年のPIO-NETデータから抽出されたトピックに対して、表4.1に基づいて分類したトピック分類の割合を月別に示す。年間を通してサービスと商品に関するトピック分類が半数以上を占める。5月以降に国の施策が増加しているのが分かる。

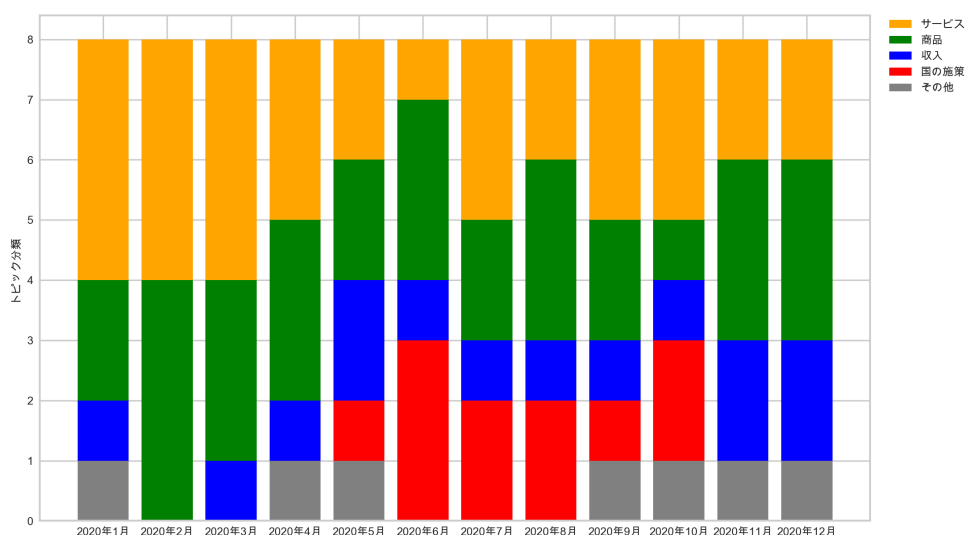


図 6.3 : 2020年 PIO-NET データ内の月別のトピック分類の割合

図6.4は2021年のトピック分類の割合を月別に示したグラフである。2020年と同様に年間を通してサービスと商品に関するトピック分類が半数以上を占める。

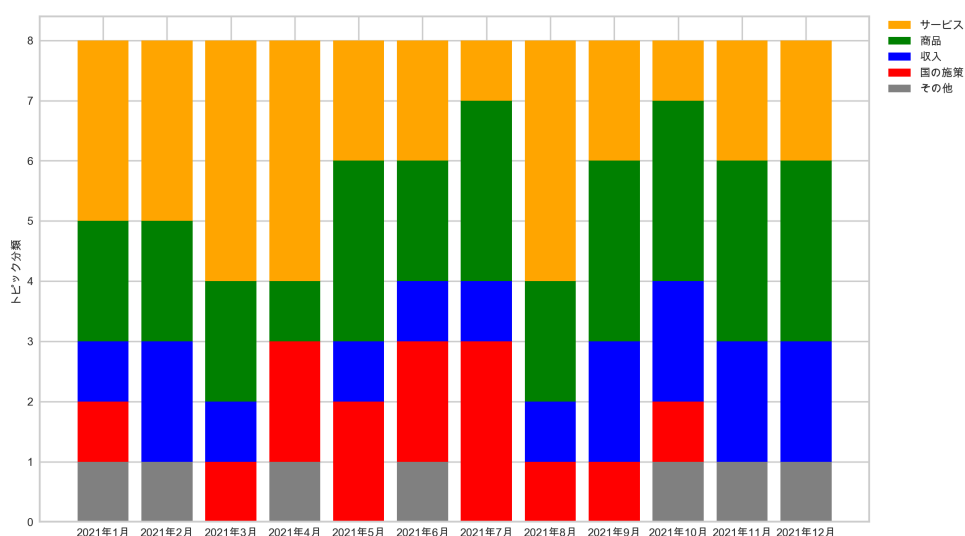


図 6.4 : 2021年 PIO-NET データ内の月別のトピック分類の割合

トラブルキーワードの時系列グラフ（図6.1、6.2）からは、特に多く相談があったトピックを代表するキーワードについて、その具体的な出現の傾向が分かった。また、

トピック分類の時系列グラフ（図6.4）からは、相談の大まかな内容の傾向が分かり、トピックそのものの分析とはまた違った傾向を読み取ることができる。

7 PIO-NET データのトピックの時系列変化

トピックの時系列変化の度合いを数値化して可視化するため、2020年1月から2021年12月までの2年間のPIO-NETデータから新型コロナウイルス感染症関連の相談を抽出し、月別のトピックを表すキーワードをベクトル表現で表すと、ベクトル間の類似度を計算することが可能となる。本稿ではトピックを表すキーワードのベクトル間の類似度として、コサイン類似度(Cosine Similarity)[12]、相関係数(Correlation Coefficient)[13]、最大情報量係数(MIC; Maximum Information Coefficient)[14]を採用し、月ごとのトピックの時系列変化を可視化する。

7.1 月別のトピックベクトル¹⁰

PIO-NETデータについて、1年ずつトピックモデリングを行い、月ごとに8つのトピックに分類すると、各トピックを代表するキーワード20語、合計160語の代表キーワードが得られる。代表キーワードから重複を除いた n 語について、各トピックに含まれる代表キーワードのディリクレ分布に従う確率をスコア t_1, t_2, \dots, t_n とする。重複するキーワードについてはスコアに加算し、月別($m = \{1, 2, \dots, 12\}$)のトピックベクトル T_m を式(7.1)で定義する。

$$T_m = [t_1, t_2, \dots, t_n] \quad (7.1)$$

ある月 m のトピックベクトル T_m と、その次月 $m+1$ のトピックモデル T_{m+1} の類似度については、ベクトル \vec{a} とベクトル \vec{b} の類似度関数 $f(\vec{a}, \vec{b})$ を用いて式(7.2)で定義する。

$$S_{m,m+1} = f(T_m, T_{m+1}) \quad (7.2)$$

このとき、ベクトル間の類似度関数 $f(\vec{a}, \vec{b})$ は、次節以降で説明するコサイン類似度、相関係数、最大情報係数を想定している。複数の類似度関数で得られる値の範囲が異なるため、 $S_{m,m+1}$ を \mathbf{x} としたとき式(7.3)のように標準化を行う。

$$\mathbf{x} \mapsto \frac{\mathbf{x} - \bar{\mathbf{x}}}{s} \quad (7.3)$$

ここで、 $\bar{\mathbf{x}}$ は \mathbf{x} の平均値、 s は \mathbf{x} の標準偏差である。

¹⁰トピックベクトルとは、トピックの代表キーワードをベクトル化したものである。

7.2 コサイン類似度(Cosine Similarity)

コサイン類似度(Cosine Similarity) [12]は、2つのベクトルがどのくらい似ているかを、2つのベクトルがなす角のコサイン(cosine)で表す尺度である。具体的には、図 7.1 に示すようにベクトル \vec{a} とベクトル \vec{b} のなす角 θ としたとき、ベクトル間の類似度を $\cos \theta$ で表す。コサイン類似度で取りうる範囲は、 $-1 \leq \cos \theta \leq 1$ となり、2つのトピックを表すベクトルのコサイン類似度が1に近いほど2つのトピックの内容は類似しているといえる。

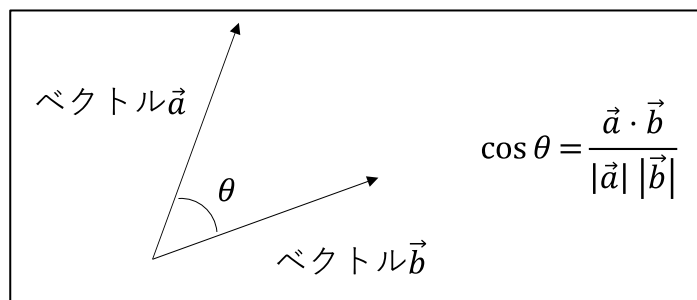


図 7.1 : コサイン類似度(Cosine Similarity)の計算例と図示

コサイン類似度に基づく類似度関数 $f_{\cos}(\vec{a}, \vec{b})$ は、ベクトル \vec{a} とベクトル \vec{b} を用いて式(7.4)で計算できる。

$$f_{\cos}(\vec{a}, \vec{b}) = \cos \theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} \quad (7.4)$$

ここで、 $\vec{a} \cdot \vec{b}$ はベクトル \vec{a} とベクトル \vec{b} の内積、 $|\vec{a}|$ はベクトル \vec{a} の大きさ、 $|\vec{b}|$ はベクトル \vec{b} の大きさを表している。

7.3 相関係数(Correlation Coefficient)

相関係数(Correlation Coefficient)[13]は、2つのベクトルに含まれる変数間の線形な関係の強弱を数値化したものである。相関係数の計算方法には、Pearson、Kendall、Spearman 等が提案されている。Spearman の相関係数は、変数の大小関係に基づく順位差を用いるため、Spearman の順位相関係数と呼ばれる。Spearman の順位相関係数は、正規分布を仮定しないノンパラメトリックな方法であり、取りうる値の範囲は $-1 \leq \rho \leq 1$ である。以下では、相関係数というと Spearman の順位相関係数を意味する。

図 7.2 は、 $A = [1, 2, 1, 0, 1, 2, 5, 3, 2, 4, 1, 4, 5]$ 、 $B = [3, 0, 0, 0, 2, 1, 2, 2, 2, 3, 0, 3, 5]$ について相関係数の計算例と分布を図示したものである。

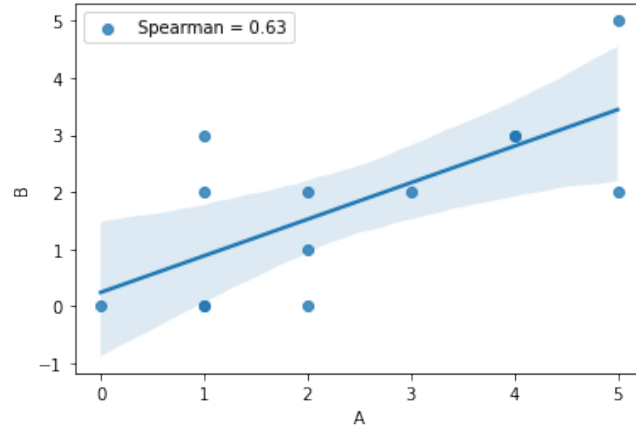


図 7.2 : 相関係数(Spearman の順位相関係数)の計算例と図示

相関係数を用いたベクトル \vec{a} とベクトル \vec{b} の類似度関数 $f_{corr}(\vec{a}, \vec{b})$ について、ベクトル \vec{a} とベクトル \vec{b} に含まれるキーワードをスコア順に並べ替えたとき、 D を対応する 2 つのキーワードの順位之差、 N を値のペアの数、同順位のある場合は、同順位の個数 n_x と n_y 、それらの順位を t_i, t_j ($i = 1, 2, \dots, n_x; j = 1, 2, \dots, n_y$) としたとき、式 (7.5) で計算できる。

$$T_x = \frac{N^3 - N - \sum(t_i^3 - t_i)}{12},$$

$$T_y = \frac{N^3 - N - \sum(t_j^3 - t_j)}{12},$$

$$f_{corr}(\vec{a}, \vec{b}) = \rho = \frac{T_x + T_y - \sum D^2}{2\sqrt{T_x T_y}} \quad (7.5)$$

7.4 最大情報量係数(MIC; Maximum Information Coefficient)

最大情報量係数(MIC; Maximum Information Coefficient)[14]は、2 つのデータの間を散布図で表し、その散布図の任意の空間をグリッドで区切ったとき、それぞれのグリッド内に含まれるデータ間の相互情報量のうち、最大となる値である。離散確率変数 X と Y の相互情報量は式(7.6)で計算される指標であり、2 つの確率変数の相互依存の尺度を表す数値である。

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x) p(y)} \quad (7.6)$$

図 7.3 で示すような例においても、コサイン類似度や相関係数では、変数間の線形な関係しか正しく数値化できなかつたが、グリッド分割し最大情報量係数を用いることで、非線形な関係性も数値化できる。

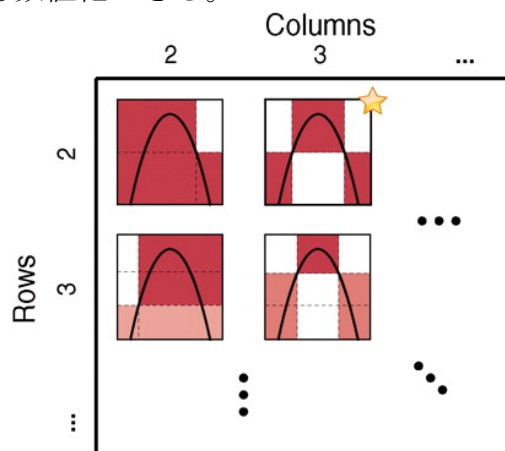


図 7.3 : 最大情報量係数(Maximum Information Coefficient)のグリッド分割の例 (文献[14]より引用)

ベクトル \vec{a} とベクトル \vec{b} をそれぞれ確率変数 A と B とし、分割されたグリッドの相互情報量を I_g としたとき、最大情報量係数を用いた類似度関数 $f_{mic}(A, B)$ は式(7.7)で表せる。

$$f_{mic}(A, B) = \max I_g(A; B) \quad (7.7)$$

7.5 トピックの時系列変化

連続する2ヶ月のトピックベクトルの類似度を用いて月別のトピックベクトルの時系列変化を図7.4に図示する。類似度が0より大きい場合は、トピックの変化が少ないといえる。逆に、0を下回る場合はトピックに変化があったといえる。最大情報量係数については、2021年3月から2021年4月にかけて大きくマイナスとなっており、コサイン類似度や相関係数では外れ値となるような局所的なトピックの変化に対しても変化があったと判断している可能性がある。

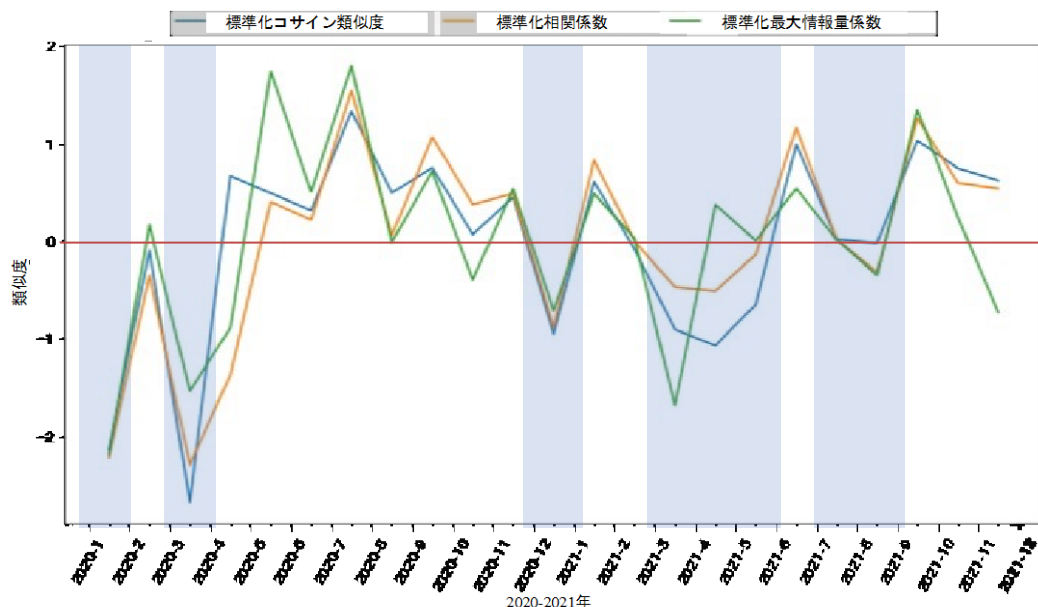
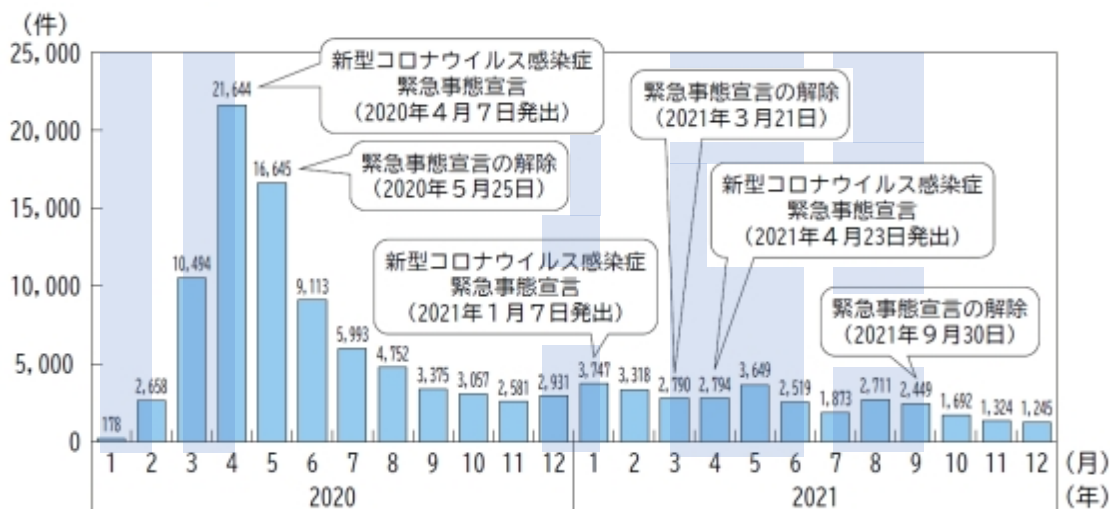


図 7.4 : 2020-2021 年新型コロナウイルス感染症関連トピックの時系列変化

比較のために、新型コロナウイルス感染症関連の相談件数の推移を図7.5に示す。トピックの時系列変化で変化がみられた時期には、緊急事態宣言が発出されていることが分かる。



(備考) 1. PIO-NETに登録された消費生活相談情報 (2022年3月31日までの登録分)。
2. 「新型コロナ関連」の相談。

図 7.5 : 2020-2021 年新型コロナウイルス感染症関連の相談件数の推移 (月別) (文献[8]より引用)

図 7.4 によると、2020 年 1 月から 2020 年 2 月、2020 年 3 月から 2020 年 4 月、2020 年 12 月から 2021 年 1 月、2021 年 3 月から 2021 年 5 月、2021 年 7 月から 2021 年 9 月にかけてトピックに変化がみられ、このタイミングで消費者の相談内容に変化がみられたことが分かる。図 7.5 によると、新型コロナウイルス感染症の話題が最初に出現した 2020 年 1 月から 2020 年 2 月を除いて、トピックに変化がみられた時期に新型コロナウイルス感染症に関連する緊急事態宣言が発出されていることが分かる。

また、月別のトピックの時系列変化と消費者庁の注意喚起の発出時期との関連について検討するため、2020 年 1 月から 2021 年 12 月までに発出された新型コロナウイルス感染症関連の注意喚起について表 7.1 に示す。

表 7.1：2020 年 1 月から 2021 年 12 月の消費者庁が発出した新型コロナウイルス感染症関連の注意喚起

<p>2020 年 4 月</p> <p>10 日：消費者庁「製造所等及び製造所固有記号の表示の運用について」</p> <p>10 日：消費者庁「米トレーサビリティ法の弾力的運用について」</p> <p>10 日：消費者庁「食品表示法に基づく食品表示基準の弾力的運用について」</p> <p>15 日：消費者庁「身に覚えのない商品の送り付けにご注意ください」</p> <p>17 日：消費者庁「公式 LINE アカウントの開設について」</p> <p>21 日：消費者庁「特別定額給付金を装った詐欺にご注意ください」</p> <p>28 日：消費者庁「新型コロナウイルス感染症給付金関連消費者ホットライン」の開設について</p>
<p>2020 年 5 月</p> <p>1 日：消費者庁「新型コロナウイルス感染症予防効果を標ぼうする食品について(注意喚起)」</p> <p>7 日：消費者庁「消費者団体の財務諸表等の提出に係る対応について」</p>
<p>2020 年 6 月</p> <p>5 日：消費者庁「新型コロナウイルス感染症に対する予防効果を標ぼうする商品等の表示に関する改善要請等及び一般消費者等への注意喚起について(第 3 報)」</p> <p>11 日：消費者庁「適格消費者団体等の書類提出に係る措置について」</p> <p>15 日：消費者庁「新型コロナウイルス感染症給付金関連消費者ホットライン」の開設について</p> <p>15 日：消費者庁「未利用食品の扱いについての更新情報」</p>
<p>2020 年 12 月</p> <p>25 日：消費者庁「新型コロナウイルス感染症の抗体検査キット使用後の注意」</p> <p>25 日：消費者庁「抗体検査キットの販売事業者 6 社に対する行政指導について」</p>
<p>2021 年 2 月</p> <p>5 日：消費者庁「コロナ禍における消費者被害防止キャンペーンを実施します」</p> <p>9 日：消費者庁「行政機関等をかたった"なりすまし"にご注意」</p> <p>12 日：消費者庁「新型コロナワクチン詐欺 消費者ホットライン」の開設について</p>
<p>2021 年 3 月</p> <p>26 日：消費者庁「研究用抗原検査キット及び抗体検査キット使用についての注意」</p> <p>26 日：消費者庁「新型コロナウイルス感染症の検査キットの販売事業者 5 社に対する行政指導」</p>

注意喚起の内容について詳しくみると、2020 年 4 月には特別定額給付金に関する詐欺への注意喚起がある。「給付金」に関する相談は、2020 年 4 月頃から増加し始め 6

月をピークに徐々に減少していることから、消費者からの相談件数の増加に先立って、注意喚起の発出を行っていることが分かる。この他、2020年12月に最初に発出された検査キットに関する注意喚起は、2021年3月に再び発出されている。5章の月別のトピックをみると、2021年の5月と8月に検査キットに関する相談がトピックとして表出しており、相談件数の増加に先立って注意喚起の発出を行っていることが分かる。

8 考察

2020年と2021年にPIO-NETに登録された「新型コロナウイルス感染症関連」の相談件数¹¹を比較すると、90,064件から38,921件に半減しており、図7.5をみても、減少傾向にあることが分かる。文献[2]によると、相談内容については、2020年は、「マスク」「トイレトペーパー」「消毒液」に代表されるように商品の入手に関する相談のほかに、イベントのキャンセルについての相談が多かったが、2021年以降は、「仕事」や「収入」といった経済的な理由に基づく相談や、Go To キャンペーン、ワクチン接種等の国の施策に関連する相談が増加した。また、資材や人材の不足が原因で、工事遅延の相談なども増加していた。

2020年と2021年の相談内容について、キーワードごとに分析すると、「マスク」については、両年とも、1年を通して一定の割合でトピックとして含まれたが、文献[2]によると、2020年は同月内でマスクのトピックが複数見られ、その中にはマスクの送り付け、品不足、品質への不満に関する相談が確認された。2021年では、送り付け商法等に関する相談が多くみられた。「副業」について、2020年ではトピックとして確認されなかったが、2021年1月頃から徐々に増えてきた。このことから、感染拡大初期ではなく、感染拡大から一定期間経過した後に経済的な理由に基づく相談が増加する傾向があるのではないかと示唆される。

特に大きい影響を与えたワクチン接種については、2021年2月に医療従事者向けの先行接種が開始しているが、2021年4月高齢者へのワクチン接種開始以降、ワクチン関連の相談が増加していることが明らかになった。また、「工事」については、コロナの影響による資材不足、人員削減が原因となる工事遅延の相談が2021年10月以降増加している。これらのキーワードの増加やトピックの出現状況から、緊急事態宣言の発出やワクチン接種開始は、国の施策の中でも、全ての人々の行動や心理に大きく変化を与えるものであるところ、トピックが大きく変化したのではないかと考えられる。

相談概要内の記載を確認したところ、「副業」というキーワードではなく「マルチ」「ネットワークビジネス」と記載している場合もあったので、相談概要の中には、関連語として取り扱うキーワードが潜んでいることについて、留意する必要があるのではないかと考えられる。また、国民生活センターからの注意喚起情報を受け取った相談員がこれに留意し、消費者からトラブルキーワードを聞き出せるようになることも

¹¹ 件名または相談概要のいずれかに文字列“コロナ”が含まれるもの、または指定ワードが「新型コロナウイルス関連」の相談情報（2021年1月～12月受付）の件数。

考えられる。このことから、消費者からの相談内容をいち早く、正確に把握するためには、相談員への情報提供も重要であるといえる。

2020年から2021年にかけて、連続する2ヶ月のトピックの変化をみたとき、トピックの変化の大きさについては、コサイン類似度と相関係数では大きな違いはみられなかったが、最大情報量係数とは傾向が異なる部分があった。具体的には、2021年3月から2021年5月のコサイン類似度及び相関係数の値には、緩やかに変化があったが、最大情報量係数の値は2021年3月から2021年4月に大きく変化し、2021年4月から2021年5月はあまり変化がなかった。これは、コサイン類似度及び相関係数がベクトル全体の傾向を比較するような指標であるのに対し、最大情報量係数が局所的な変化に敏感な指標であるため、2021年3月に「予約」や「ワクチン」といった新出のトラブルキーワードが出現していたことが影響していると考えられる。

9 おわりに

本稿では、2021年1月から2021年12月までの1年間に受け付け、PIO-NETへ登録された新型コロナウイルス感染症関連の相談情報について分析を行い、2020年1月から2020年12月までの1年間に受け付け、PIO-NETへ登録された新型コロナウイルス感染症関連の相談情報の分析結果との比較を行った。また、新型コロナウイルス感染症関連の相談内容について、連続する2ヶ月のトピックの類似度の変化をコサイン類似度、相関係数、最大情報量係数を用いて定量的に表すことで、トピックの時系列的な変遷を明らかにした。さらに、各月のトピックの時系列変化について、社会的な情勢と比較したところ、トピックに変化がみられたタイミングで緊急事態宣言が発出されていたことが明らかになった。

相談内容のトピックの変化について、コサイン類似度、相関係数、最大情報量係数を用いてトピック間の類似度を計算し、月単位のトピックの変化を数値化した結果、緊急事態宣言が発出された時期にトピックが大きく変化していることが分かった。緊急事態宣言の発出時期とは、つまり、新型コロナウイルス感染症が感染拡大した時期であり、この時期に消費生活相談の内容が変化しているということは、社会情勢の変化に伴う消費者への影響が少なくないことを示唆している。また、消費者庁の注意喚起後にトピックが変化する傾向がみられた。さらに、関連する相談の件数が増加している傾向がみられたことから、消費者庁が注意喚起を行う時期については適切ではあるが、注意喚起の発出後に相談件数が増加することを踏まえて、現状の注意喚起の方法と相談件数への影響について、さらなる調査が必要ではないかと考える。

新型コロナウイルス感染症のように、突発的に発生した事案に伴う消費者被害の早期把握は難しいが、本手法を用いてトピックの時系列的な変遷を把握することができれば、今後、突発的に事案が生じた際に、システム的に消費者問題の早期発見が実現できるのではないかと考えられる。ひとつの方法として、トピックベクトルの類似度を指標とするトピックの変化の検知手法は、突発的に生じた事案の検出に利用できる可能性がある。コサイン類似度や相関係数の値を用いてトピックベクトルの全体の傾向を分析しつつ、最大情報量係数の値と比較し、新しいトピックの出現を検知できる

可能性がある。今後は、トピックの分析手法やトピックの時系列分析の手法を用いて、消費者問題の早期発見を実現する手法とシステムの開発を進めたい。

参考文献

- [1] 独立行政法人国民生活センター (2021) 「PIO-NET の紹介」
<http://www.kokusen.go.jp/pionet/> (最終アクセス日: 2021年11月14日)
- [2] 谷岡 広樹, 徳永 欽也, 徳永 美和子, 近森 美麻子, 三谷 達也, 宮本 麗子, 山口 愛弓, 小早川 優, 安井 杏奈: PIO-NET を利用した消費者問題の傾向分析, 消費者庁新未来創造戦略本部国際消費者政策研究センター リサーチ・ディスカッション・ペーパー, 2022.
- [3] M. Honnibal and I. Montani: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing, 2017.
- [4] 松田寛, 大村舞, 浅原正幸: 短単位品詞の用法曖昧性解決と依存関係ラベリングの同時学習, 言語処理学会第25回年次大会, 2019.
- [5] K. Takaoka, S. Hisamoto, N. Kawahara, M. Sakamoto, Y. Uchida, Y. Matsumoto: Sudachi: a Japanese Tokenizer for Business, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), May 7-12, 2018.
- [6] 大島 裕明, 中村 聡史, 田中 克己: SlothLib Web 検索研究のためのプログラミングライブラリ, 日本データベース学会 letters, 日本データベース学会, Vol. 6, No. 1, pp. 113-116, 2007.
- [7] 消費者庁 (2021) 「令和3年版 消費者白書」
- [8] 消費者庁 (2022) 「令和4年版 消費者白書」
- [9] D. M. Blei, A. Y. Ng, and M. I. Jordan: Latent Dirichlet allocation. *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [10] 佐藤一誠, 奥村学: トピックモデルによる統計的潜在意味解析, コロナ社 (自然言語処理シリーズ), 2015.
- [11] G. Salton, A. Wong, and C.-S. Yang: A Vector Space Model for Automatic Indexing, *Commun. ACM*, Vol. 18, No. 11, pp. 613–620, 1975.
- [12] A. Huang: Similarity measures for text document clustering, *Commun. In Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand, Vol. 4, pp. 9-56, 2008.
- [13] C. Spearman: The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, Vol. 15, No. 1, 1904.
- [14] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher and P. C. Sabeti: Detecting Novel Associations in Large Data Sets, *Science*, Vol. 334, No. 6062, pp. 1518-1524, 2011.