

# ソーシャルリスニングによる データサイエンス分野の話題傾向抽出方式

## Topic Trends Extraction Method in the Data Science Field by Social Listening

関戸惟久<sup>1</sup> 岡田龍太郎<sup>1</sup> 中西崇文<sup>1†</sup>

IKU SEKIDO<sup>1</sup> RYOTARO OKADA<sup>1</sup> TAKAFUMI NAKANISHI<sup>1</sup>

**概要** 本稿では、ソーシャルメディアからユーザの生の声を分析してマーケティングに利用する戦略であるソーシャルリスニングを応用し、データサイエンス分野を対象とした直近の話題の傾向の抽出方式について示す。本方式では、有名なデータサイエンティストの Twitter アカウントをいくつか選定し、そのアカウントにある直近のツイートから、話題となるキーワードをそれらのアカウントで出現している単語の共通点と差異により抽出することが可能である。本方式によって、直近のデータサイエンス分野の現状と今後の動向について把握することが可能になる今後、本方式により、データサイエンス分野と他分野との直近の話題の関連性を見つけることが可能になり、データサイエンスが適用されていない他分野への応用方法を模索する一助になりうる。

**キーワード** ソーシャルメディア コンテキストアウェアネス Web サービス  
スマートコミュニティ コンシューマシステム

### 1. はじめに

近年、インターネット産業の発展によって、SNS（ソーシャル・ネットワーキング・サービス）という利用者同士が交流することができるサービスが生まれた。近年における SNS の利用者数は、全世界で 38 億人と膨大なユーザーが存在している。SNS 利用者の増加によって、各ユーザーが発する膨大なデータを収集、分析することで、あらゆる発見が得られるという、ソーシャルリスニングが、重要視されている。

SNS 上のものに限らないが、技術の進歩によって、膨大なデータを収集、分析する学問として、データサイエンスという学問が、近年注目を浴びている。しかし、データサイエンスは、比較的新規な学問である。よって、データサイエンスは、今後の動向が予測することが難しくなっている。

本稿では、ソーシャルリスニングによるデータサイエンス分野の話題傾向抽出方式について示す。本方式は、ソーシャルメディアからユーザの生の声を分析してマーケティングに利用する戦略であるソーシャルリスニングを応用し、データサイエンス分野を対象とした直近の話題の傾向の抽出方式について示す。本方式は、有名なデータサイエンティストの Twitter アカウントをいくつか選定し、そのアカウントにある直近の Tweet から、話題となるキーワードをそれらのアカウントで出現している単語の共通点と差異により抽出することが可能である。

本方式によって、直近のデータサイエンス分野の現状と今後の動向について把握することが可能になることで、データサイエンス分野と他分野との直近の話題の関連性を見

つけることが可能になり、データサイエンスが適用されていない他分野への応用方法を模索する一助となりうるということが考えられる。

本稿は、次のように構成される。2 節では、関連研究について示す。3 節では、本方式であるソーシャルリスニングによるデータサイエンス分野の話題傾向抽出方式について示す。4 節では、本方式を実現する実験システムを実現し、実験により本方式の有効性の検証をする。5 節で、本稿をまとめる。

### 2. 関連研究

本節では、本方式に関連する研究について挙げる。テキストデータを用いて話題傾向抽出について多数の研究がされてきた。ここでは、特に SNS を用いた話題傾向抽出についての関連研究を述べる。

浦野ら[2]は、SNS 情報を対象に、話題が遷移する傾向についての解析を行った。この研究では、SNS 情報を取りこむ際、ブログ情報の記事情報を考慮することとしている。ブログの情報から記事情報を取り込む際、特徴語を抽出している。この特徴語を抽出する方式として、TF-IDF を用いることで特徴語の抽出を実現している。本方式による特徴語の抽出は、本稿で述べる提案方式の特徴語抽出と類似している。

木原ら[3]は、膨大なユーザを抱える SNS 上から、有意義な話題を抽出することについての研究を行った。この研究では、分析する SNS を SNS の中でもユーザが多く、情報発信や情報交換が容易であるという観点で、Twitter を分析対象と設定した。Twitter を分析の対象にするという点が、稿で述べる提案方式の対象とする SNS と類似している。

<sup>1</sup> 武蔵野大学 データサイエンス学部 データサイエンス学科

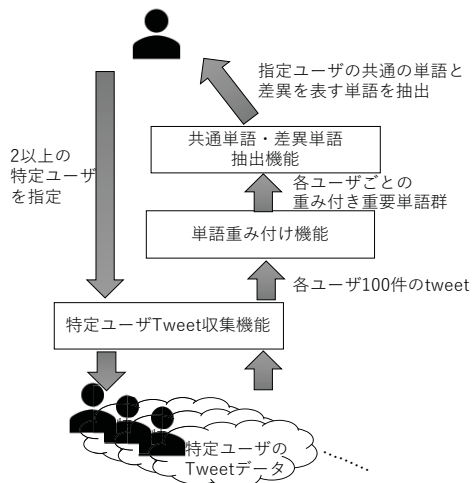


図 1 本方式の全体図

分析方法に関しては、ベクトルの同士の距離を測り分析しているという点は、本稿とは、単語のベクトルに注目しているという点で異なる。

### 3. データサイエンス分野の話題傾向抽出方式

本節では、提案方式である、ソーシャルリスニングによるデータサイエンス分野の話題傾向抽出方式について述べる。

#### 3.1 全体像

本節では、提案方式であるソーシャルリスニングによるデータサイエンス分野の話題傾向抽出方式の全体像について示す。本提案方式の全体像を図 1 に示す。本研究の目的は、新規概念について、今後の動向を予測することができる手法として、ソーシャルリスニングを用いた話題傾向抽出方式を実現することである。ここで言うソーシャルリスニングとは、Twitter や Facebook 等のソーシャルメディア (SNS) 上で発信された情報を、収集・分析することを指す。本研究では、対象の概念の専門家の SNS 上の呟きの内容を、新規概念の今後の動向と考えと仮定する。本稿では、データサイエンス分野を対象にした。データサイエンスの専門家であるデータサイエンティストの SNS 上で呟きについてテキストデータとして取得する。そのテキストデータを複数用意し、単語ごとに区切り、各単語に対して TF-IDF を用いて単語ごとの重要度を出す。その中で重要度の高い単語を抽出し、各テキストデータから持ってきた単語と比較し、その共通点と差異で分析を行う。テキストデータを単語ごとに区切る方法として、形態素解析を用いた。形態素解析は、文章を単語の最小単位 (品詞ごとなど) に分類することをいう。本方式では、形態素解析を行うことで、SNS 上の呟きに含まれる会話文や口調に関係なく TF-IDF を行うことができる。TF-IDF とは、文書における単語ごとの重要度を表す方法である。本方式で、TF-IDF を用いることで、多数の単語が出ているデータに対しても、単語ごと

の重要度出し、重要度の高い単語を抽出することができる。そして抽出された単語に対して共通点、差異を用いて分析することができるようになる。本方式では、ソーシャルリスニングを用いた分析を行う媒体を、Twitter とした。Twitter は、日本で人気がありアメリカに次ぐ、2 番目に大きなマーケットと言われている。そのように日本で人気がある点、Twitter は 140 字までのショートテキストを投稿できる SNS で、本方式で用いるテキスト分析を適用しやすいという点から、対象とする媒体を Twitter とした。

3 節の構成について述べる。3.2 節では、本方式を用いるユーザの条件についての特定のユーザの Tweet 抽出方式について述べる。3.3 節では、形態素解析、TF-IDF を用いた単語重み付け機能について述べる。3.4 節では、TF-IDF によって重みつけられた単語の中から、重要単語として抽出されたものについて、共通・差異を用いて分析する、共通単語・差異単語抽出機能について述べる。

#### 3.2 特定ユーザの Tweet 抽出機能

本節では、特定のユーザの抽出について述べる。この機能は、二つのステップによって実現される。一つ目は、対象とする特定ユーザの選択について述べる。二つ目は特定ユーザの Tweet を抽出する機能について述べる。

##### 3.2.1 対象とする特定ユーザの選択

データサイエンスの話題傾向を掴むためには、最新の内容を含むテキストデータであり、かつ、データサイエンス分野の専門家による発信されたテキストデータが必要であると考えられる。本稿では、特にデータサイエンス分野の専門家とされるアカウントから発信された Tweet のテキストデータを対象とする。本方式では、データサイエンティストの中でも、知名度がある人を対象とするため、Twitter アカウントのフォロワー数の最低人数を 1000 人程度に設定した。

##### 3.2.2 特定ユーザの Tweet 抽出機能

対象とする特定ユーザの選択によって対象とされたユーザに対して、Twitter の Tweet を抽出する機能について述べる。本方式では、Twitter から Tweet を抽出する機能を、TwitterAPI を用いて抽出を行う。本稿では、対象の新規概念の今後の動向を予測することが目的のため、最新情報を取得できる最新 Tweet について 100 件に対して抽出を行う。

#### 3.3 単語重み付け機能

本節では、形態素解析、TF-IDF を用いた単語重み付け機能について述べる。この機能は、二つのステップによって実現される。一つ目は、前節で述べた特定ユーザの Twitter 上の最新 100 件の Tweet に対して、文章を品詞ごとに分ける形態素解析機能について述べる。二つ目は、形態素解析の結果を用いて、単語ごとの重みをつけるために用いる TF-IDF 機能について述べる。

##### 3.3.1 形態素解析機能

TwitterAPI を用いて所得した最新 Tweet の 100 件のテキ

ストデータについて、形態素解析を用いる。形態素解析は、文書を単語の最小単位(品詞ごとに)分類するものである。本稿では、形態素解析を行うエンジンとしてオープンソースである MeCab を用いて行う。単語の重要度を出し、その中で重要度が高い単語を抽出して分析するため、文書中における、助動詞や助詞等の単体では意味を持たない単語や一人称等の単体で意味を持つが、文書中では大きな意味を持たない単語を、削除する。このような単語の削除のためのストップワードリストとして、SlothLib の日本語ストップワードリスト[4]を用いた。本リストは助詞、助動詞、文書中でよく使われる大きな意味を持たない単語をまとめたリストである。このリストを用いて、形態素解析を行なった文書に対して、リストに該当する単語の除外を行う。

### 3.3.2 TF-IDF 機能

形態素解析を行なった文書に対して、単語の重要度を表すため、TF-IDF を用いる。本方式で重み付けを TF-IDF を用いて行うことで、本方式で使用する IDF 値を、ユーザごとに設定することで、ユーザごとの単語の重要度を出すことができる。ユーザの最新 100 個の Tweet 全体に対して、各ユーザ固有の単語の重要度を測ることができる。

### 3.4 共通単語・差異単語抽出機能

本節では、共通単語・差異単語抽出機能について述べる。この機能は二つのステップによって実現される。一つは、抽出された単語の中から共通点を用いて重要単語を抽出する共通単語抽出機能、もう一つは、抽出された単語の中の差異を用いて重要単語を抽出する、差異単語抽出機能である。3.4.1 節では、共通単語抽出機能について示す。3.4.2 節では、差異単語抽出機能について示す。

#### 3.4.1 共通単語抽出機能

TF-IDF を用いた単語重み付け機能によって抽出された単語の中から共通点を用いて、重要単語を抽出する共通単語機能について述べる。共通単語抽出機能は、前節で示した重み付けされた単語の中から、重みが大きい単語 (TF-IDF 値が高い単語) を抽出する。重要単語同士が対象とするアカウントの中で何人に共通しているかどうかを調べる。対象概念の専門家からの同一の重要単語が抽出されているかどうかを比べ、そしてその共通単語を抽出することで、新規概念にとって重要であるとする共通単語を抽出する。該当単語が、どのアカウントで出現しているかをカウントし、共通して出現している単語を重要単語として抽出する。これにより、データサイエンスにおける共通した新規概念を獲得することが可能となる。

#### 3.4.2 差異単語抽出機能

重要度が高く、どのアカウントにも共通して出現する単語については、その分野における新規概念として捉えることが可能であるが、特にデータサイエンス分野においては、様々な分野との連携が考えられることから、データサイエンティストによって、使われる単語が異なることもある。

このようにアカウント固有の言葉を抽出するために、差異単語を抽出する。具体的には、TF-IDF による重みが大きく、かつ、共通単語抽出機能で抽出されなかった単語を差異単語として抽出する。これにより、それぞれのデータサイエンティストにおける使われる単語の違いや広がりを観察することが可能となる。

## 4. 評価実験

本節では、本手法の評価実験について述べる。4 節の構成について述べる。4.1 節では、実験環境について述べる。4.2 節では、形態素解析, TF-IDF を用いた単語重み付け機能について述べる。4.3 節では、共通単語・差異単語抽出機能について述べる。

本実験で使用した、データサイエンティストは前節で示した、対象とする特定ユーザの選択をもとに選択した。本実験で使用するデータサイエンティストの人数を、4 人とした。

### 4.1 実験環境

本節では、実験環境について述べる。3 節で示したシステムを実装した。本方式はデータサイエンス学問についての抽出方式である。よって Tweet 抽出機能の対象となるデータサイエンティストを選ぶ。抽出条件については、3.2.1 節で示したとおり、データサイエンティストであること、Twitter アカウントを保有していること、データサイエンティストとして有名であることを判断するため、Twitter フォロワー数が 1000 人を超えていることである。また本実験で使用するデータサイエンティストの人数を 4 人とした。以上の条件で抽出された人物は、表 1 のようになる。4 ユーザに対して、本実験を行う。

3.3.1 節で示した形態素解析機能を実装するにあたって、MeCab を使用する。

### 4.2 形態素解析, TF-IDF を用いた単語重み付け機能

本節では、形態素解析, TF-IDF を用いた単語重み付け機能についての実行例について示す。この方式は 3 節で示したとおり、MeCab を用いた形態素解析、そして TF-IDF の 2 つの段階からなる。

3 節で示した形態素解析機能のとおり、MeCab を用いた形態素解析を、最新 Tweet(2021 年 1 月)100 件について実行した。Tweet には会話で使うような言葉や、1 単語としては意味をなさないようなものが存在していた。余分なテキストを削除するため、3 節で示した、SlothLib の日本語ストップワードリストを用い削除を行った。ストップワードリストを用いることで、余分なテキストを削除することができた。

表 1 対象ユーザのアカウント情報

| 対象ユーザ           | Twitter<br>アカウント | Twitter<br>フォロワー数 |
|-----------------|------------------|-------------------|
| TJO             | @TJO_datasci     | 44,700            |
| かめ              | @usdatascientist | 17,000            |
| マスク・ド・<br>アナライズ | @maskedanl       | 14,900            |
| ヤエリ             | @yaesuri_man     | 2,868             |

表 2 TF-IDF 機能の実験結果

|                              | 単語   |
|------------------------------|--|
| <b>TJO</b>                   | 東大、運動、生理学、東京、画像、ロンドン、<br>コロナウイルスワクチン、英主、水道、道理、<br>感染、福岡、電話、華僑、輸入、病床、グー<br>グルドライブ、データセンター、電気、物理、<br>歯科、主治医、歯医者、手洗い、飲食、リモ<br>ート、緊急事態宣言、日本、陽性、検査、栃<br>木、駅、台風、エッフェル塔、ガウスマルコ<br>フ、テレワーク、goto キャンペーン、罰金、<br>残業、エンジニア、年収、雪、携帯、デパー<br>ト、幕末、社会、記録、学術、小学生、琵琶<br>湖、機械学習、サイエンティスト、年末年始、<br>京都、ワクチン、株、薬剤、アストラゼネカ、<br>古典、藤子 f 不二雄、無料、ミノタウロス、編<br>集   |
| <b>かめ</b>                    | Python、インスタ、サイエンティスト、サイエ<br>ンス、職、エラー、テクノロジー、お金、職<br>業、テックナード、居酒屋、ランチ、丸の内、<br>講師、お年玉、コロナ、slack、子ども、教育、<br>課題、資料、パワポ、テレビ、ドラマ、twitter、<br>先生、米国、仮想通貨、政府、金融、政策、<br>経済、ブログ、クリスマス、投資、学校、統<br>計、matplotlib、excel、seaborn、公園、富士通、<br>挑戦、コンフォートゾーン、素人、ベストセ<br>ラー、スタバ、ラー、高専、年明け、所属、<br>友達、布教、ツイート、ポイント、変数、帰<br>国、ウイルス、世界、株、デジタル、部署、<br>就職、深層、仕事、ジャーナル、特許、暇、<br>無料、週間、学問、酒井、恩返し、配信、書<br>籍、習得、小学校、問題、日常、日本語、英<br>語、読書、筋トレ、試験、受験、コンテナ |
| <b>マスク・<br/>ド・アナ<br/>ライズ</b> | 書籍、市場、研究、文中、ブローリー、稲作、<br>道場、販売、日経、教室、有隣堂、書店、ヨ<br>ドバシアキバ、選手、ラム、飯、防衛、賽銭、<br>初詣、アルコール、プロテイン、祝い酒、キ<br>ン肉マン、座談、文字数、画像、登壇、企業、  |

|            |   |
|------------|---|
|            | ハッキング、情報、軍団、ワニ、爬虫類、プ<br>ロジェクト、本、池袋、本店、献立、機械学<br>習、ビジネス、営業、撮影、心霊、プログラ<br>マー、アントフィナンシャル、真髓、ファシ<br>リテーショ、業界、言語、国内、メディア、<br>運営、ダンス、トレンド、エンジニア、若者、<br>大城、技術評論者、電子、毒、日本、大使館、<br>車、闇、社長、ワイドショー、YouTube、ブッ<br>ク、渋谷、名古屋、都内、twitter、紙、デッド<br>プール、フォールアウト、女子高、マック、<br>SDGS、確定申告、配送、電話、課題解決、文<br>化、慣習、配信、工場、地理、ノジマ、ラノ<br>ベ、クリスマス、お年玉、プレゼント、鬼滅<br>の刃、税金、大学、官公庁、インフルエンサ<br>ー、イキリデータサイエンティスト、無料、<br>ガイアの夜明け、カンブリア宮殿、詐欺、テ<br>レビ東京、ベンチャー、縄張り、業務、集客、<br>プロモーション |
| <b>ヤエリ</b> | 数理、統計、新卒、株価、仮想通貨、機械学<br>習、小学生、漢字、暗記、民主、福利、厚生、<br>ビジネスマン、python、筋トレ、スポーツ、<br>書籍、セキュリティ、言語、データベース、<br>面接、ツイート、kaggle、教師、ミーティン<br>グ、物理、カレー、国会、百貨店、twitter、献<br>立、在宅、大学、文法、将棋、東大、囲碁、<br>地理、地球、本、電話、教科書、メダリスト、<br>google、工場、宗教、投資、おむつ、ビール、<br>マイク、カメラ、音声、健康診断、音、画像、<br>プログラミング、京大、無料、動画、日本、<br>科学、医療、タレントデータサイエンティ<br>スト、兵庫、神戸、新聞、クリスマス  |

形態素解析機能は、SlothLib の日本語ストップワードリス  
トを使用することのみで良いと判断した。

続いて TF-IDF 機能についてだが、先ほどの形態素解析  
の結果を用いて、3.3.2 節の TF-IDF 機能をもとに、IDF 値  
をユーザごとに設定を行った。TF-IDF を用いることで、重  
み付けを行うことができた。重み付けを行った結果、重要  
単語を抽出することができた。抽出された重要単語を表 2  
にまとめた。

#### 4.3 共通単語・差異単語抽出機能

本節では、共通単語・差異単語抽出機能を用いた分析につ  
いて示す。3.4 節の方式の通り、単語の共通点・差異を示す  
システムを実装した。本実験は、三ステップにより構成され  
る。一つ目は共通単語抽出機能の実行例であり、二つ目は  
差異単語抽出機能の実行例である。

##### 4.3.1 共通単語抽出機能の実験

本節では、共通単語抽出機能の実行例について述べる。

表3 共通単語抽出機能の実験結果

|      | 共通単語                                       |
|------|--|
| 4人共通 |  |
| 3人共通 | 日本, 機械学習, 物理, 地理, 仮想通貨, 株, 書籍, 画像, twitter |
| 2人共通 | 東大, 電話, 言語, 宗教, プログラミング, 医療, 生理学, 深層学習     |

表4 差異単語抽出機能の実験結果

|             | 単語  |
|-------------|---|
| TJO         | 東大, 運動, 生理学, 東京, 画像, ロンドン, コロナウイルスワクチン, 英主, 水道, 道理, 感染, 福岡, 電話, 華僑, 輸入, 病床, グーグルドライブ, データセンター, 電気, 物理, 歯科, 主治医, 歯医者, 手洗い, 飲食, リモート, 緊急事態宣言, 日本, 陽性, 検査, 栃木, 駅, 台風, エッフェル塔, ガウスマルコフ, テレワーク, goto キャンペーン, 罰金, 残業, エンジニア, 年収, 雪, 携帯, デパート, 幕末, 社会, 記録, 学術, 小学生, 琵琶湖, 機械学習, サイエнтиスト, 年末年始, 京都, ワクチン, 株, 薬剤, アストラゼネカ, 古典, 藤子不二雄, 無料, ミノタウロス, 編集   |
| かめ          | Python, インスタ, サイエнтиスト, サイエンス, 職, エラー, テクノロジー, お金, 職業, テックナード, 居酒屋, ランチ, 丸の内, 講師, お年玉, コロナ, slack, 子ども, 教育, 課題, 資料, パワポ, テレビ, ドラマ, twitter, 先生, 米国, 仮想通貨, 政府, 金融, 政策, 経済, ブログ, クリスマス, 投資, 学校, 統計, matplotlib, excel, seaborn, 公園, 富士通, 挑戦, コンフォートゾーン, 素人, ベストセラー, スタバ, ラー, 高専, 年明け, 所属, 友達, 布教, ツイート, ポイント, 変数, 帰国, ウイルス, 世界, 株, デジタル, 部署, 就職, 深層, 仕事, ジャーナル, 特許, 暇, 無料, 週間, 学問, 酒井, 恩返し, 配信, 書籍, 習得, 小学校, 問題, 日常, 日本語, 英語, 読書, 筋トレ, 試験, 受験, コンテナ |
| マスク・ド・アナライズ | 書籍, 市場, 研究, 文中, ブロリー, 稲作, 道場, 販売, 日経, 教室, 有隣堂, 書店, ヨドバシアキバ, 選手, ラム, 飯, 防衛, 賽銭, 初詣, アルコール, プロテイン, 祝い酒, キン肉マン, 座談, 文字数, 画像, 登壇, 企業, ハッキング, 情報, 軍団, ワ  |

|     |   |
|-----|---|
|     | ニ, 爬虫類, プロジェクト, 本, 池袋, 本店, 献立, 機械学習, ビジネス, 営業, 撮影, 心霊, プログラマー, アントフィナンシャル, 真髓, ファシリテーション, 業界, 言語, 国内, メディア, 運営, ダンス, トレンド, エンジニア, 若者, 大城, 技術評論者, 電子, 毒, 日本, 大使館, 車, 闇, 社長, ワイドショー, YouTube, ブック, 渋谷, 名古屋, 都内, twitter, 紙, デッドプール, フォールアウト, 女子高, マック, SDGS, 確定申告, 配送, 電話, 課題解決, 文化, 慣習, 配信, 工場, 地理, ノジマ, ラノベ, クリスマス, お年玉, プレゼント, 鬼滅の刃, 税金, 大学, 官公庁, インフルエンサー, イキリデータサイエンティスト, 無料, ガイアの夜明け, カンプリア宮殿, 詐欺, テレビ東京, ベンチャー, 縄張り, 業務, 集客, プロモーション |
| ヤエリ | 数理, 統計, 新卒, 株価, 仮想通貨, 機械学習, 小学生, 漢字, 暗記, 民主, 福利, 厚生, ビジネスマン, python, 筋トレ, スポーツ, 書籍, セキュリティ, 言語, データベース, 面接, ツイート, kaggle, 教師, ミーティング, 物理, カレー, 国会, 百貨店, twitter, 献立, 在宅, 大学, 文法, 将棋, 東大, 囲碁, 地理, 地球, 本, 電話, 教科書, メダリスト, google, 工場, 宗教, 投資, おむつ, ビール, マイク, カメラ, 音声, 健康診断, 音, 画像, プログラミング, 京大, 無料, 動画, 日本, 科学, 医療, タレントデータサイエンティスト, 兵庫, 神戸, 新聞, クリスマス  |

4.2 節で行った単語重み付け機能によって抽出された単語を, 3.4.1 節で示したとおり対象とするアカウントの中で何人のユーザに共通している単語であるか調べる. 共通した単語をまとめたものが表3になる. 表3を見ると, 「機械学習」や「深層学習」といったデータサイエンスに使われる手法が見られるのと同時に, 「物理」や「地理」, 「仮想通貨」「株」などのデータサイエンスとはあまり関連のない単語が見られた.

#### 4.3.2 差異単語抽出機能の実験

本節では, 差異単語抽出機能の実行例について述べる. 4.2 節で行った単語重み付け機能によって抽出された中から, 共通単語抽出機能で抽出されなかった単語を差異単語として抽出する. 差異単語として抽出された単語をまとめたものが表4になる. 表4を見ると, 新型コロナウイルス関連の単語や地名などの単語が見られる. またアカウント

によって抽出される単語の特徴の差が見られた。

## 5. まとめ

本稿では、ソーシャルリスニングによるデータサイエンス分野の話題傾向抽出方式について述べた。本方式は、対象となる新規概念に関する専門家を対象として、SNS 上での最新 Tweet に対して分析を行うことで、今後の動向を予測することができる手法として、ソーシャルリスニングを用いた話題傾向抽出方式を実現した。

また、本方式を検証するための実験システムを構築し、実際にデータサイエンスを対象にし、本方式によってデータサイエンスの今後の話題を抽出できることを確かめる実験を行い、データサイエンスの話題傾向を予測することができた。

今後の課題として、本稿ではデータサイエンスに対しての方式のみの提案であったが、他の新規概念に対しても本方式の実現を行う。また本方式を用いた、アプリケーションを用いた、対象物の話題傾向をつかむコンテンツの提供等の検討が挙げられる。

## 参考文献

- [1] 浦野史隆, 六井淳, SNS 間の話題遷移を推定する自己組織化マッピング. FIT2017(第 16 回情報科学技術フォーラム). 2017-09-05
- [2] 木原大志, 白木原渉, 大石哲也, 越村三幸, 藤田博, 長谷川隆三, Twitter の時系列解析による注目話題の抽出. 情報処理学会第 74 回全国大会. 2012-03-06.
- [3] SlothLib の日本語ストップワードリスト, <<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NL/Filter/StopWord/word/Japanese.txt>> (参照 2021-05-07)