

# Twitter ミュート機能におけるキーワード拡張手法の提案

関根 萌芽<sup>†</sup> 佐藤 哲司<sup>††</sup>

<sup>†</sup> 筑波大学情報学群 〒305-8550 茨城県つくば市春日 1-2

<sup>††</sup> 筑波大学図書館情報メディア系 〒305-8550 茨城県つくば市春日 1-2

E-mail: †{sekine20,satoh}@ce.slis.tsukuba.ac.jp

**あらまし** Twitter では多種多様な話題に関する投稿がなされているため、楽しみにしていたコンテンツのネタバレや苦手な話題に関するツイートを意図せず見ってしまう可能性がある。回避方法として設定したキーワードを含むツイートを非表示にするキーワードミュート機能があるが、ツイートの中には設定したキーワードを含まないが、そのキーワードに関連する内容のツイートも存在する。本研究では、特定の話題に関するツイートを網羅的にミュートするための関連語を見つけることを目的とする2つのキーワード拡張手法を提案する。提案手法は、特定のキーワードを含むツイートを分析し拡張を行う手法と、キーワードに関連の強いユーザのツイートを分析し拡張を行う手法からなる。拡張して得られた関連語によって、特定の話題をどの程度網羅的にミュートできるかを適合率と再現率を用いて評価した結果、2つの提案手法ともに mute-words の拡張によってミュート性能を向上できることが明らかとなった。また、話題によって有効な手法が異なることも確認された。

**キーワード** Twitter, ミュート, キーワード拡張

## 1 はじめに

Twitter<sup>1</sup>の普及に伴い、多くの人が気軽に自分の考えをリアルタイムに発信し、他者と意見を共有することが可能となった。また、企業や芸能人などが運用する公式アカウントも多く存在し、情報源の一つとしても利用されている。Twitter は誰もが気軽に自分の意見を投稿できるため、Twitter のユーザは苦手とする話題に関する投稿や、楽しみにしていたコンテンツのネタバレを意図せず見ってしまうこともある。このような状況を回避するための方法として、Twitter ではキーワードミュート機能が提供されている。キーワードミュート機能では、あらかじめ見たくないと考えたキーワードを設定することで、そのキーワードを含むツイートをタイムラインや通知からミュート、すなわち非表示にすることができる。本論文では、キーワードミュート機能に設定するキーワードを mute-words と呼ぶ。

ここで、ある mute-words をキーワードミュート機能に設定したとしても、ツイートの中には mute-words を含んでいないが mute-words に関連するツイートも存在する。例えば新型コロナウイルスに関する話題をミュートしたいと考え、「コロナ」をキーワードミュート機能に設定としたとする。しかし、「PCR 検査を受ける」といったツイートが存在したとき、「コロナ」に関連するツイートであることが分かるが、「コロナ」というキーワードを含んでいないためミュートはされない。したがって、特定の話題に関するツイートを網羅的にミュートするためには、関連語も設定する必要があると考えられる。一方で、「コロナビール」のように、「コロナ」というキーワードを含むが、新型コロナウイルスとは関係ない話題もミュートされてしまう、といった問題点も存在する。本研究では、関連語を複数提示

し、ユーザ自身にキーワードミュート機能に設定するキーワードを選択してもらうことを想定しているため、考慮しないものとする。

本研究では、特定の話題に関するツイートを網羅的にミュートするための関連語を見つけることを目的とした mute-words の拡張手法を提案する。提案手法では、mute-words を含むツイートを分析し拡張を行う。mute-words を含むツイートに同時に出現しやすい語は関連語である可能性が高いと考えられる。また、Twitter では、テレビ番組のスタッフが運用し、そのテレビ番組の告知や裏話などを投稿する公式アカウントや、趣味垢と呼ばれる一般アカウントも存在する。趣味垢とは、好きなアイドルやゲームなど特定の話題に関するツイートを主としているアカウントのことである。また、本論文では企業や芸能人ではなく一個人が運用するアカウントを一般アカウントと呼ぶ。特定の話題に関するアカウントが存在することに着目し、mute-words に関するツイートを頻繁に行うユーザ、すなわち mute-words に関連の強いユーザのツイートを分析し拡張を行う。mute-words に関するツイートを頻繁に行うユーザが積極的に使用する語は関連語である可能性が高いと考えられる。

本論文では、第2章で先行研究を概観し、本研究の位置づけを明らかにする。第3章で提案手法について説明し、第4章では実験結果を示し、考察を述べる。最後に第5章で結論と今後の課題について述べる。

## 2 関連研究

本研究は、特定の話題に関するツイートを網羅的にミュートするための関連語を見つけることを目的とし、mute-words を含むツイートと mute-words に関するツイートを頻繁に行うユーザに着目した拡張手法を提案するものである。

1: <https://twitter.com>

Twitter 上の特定の話題に関するツイートを検知する研究として、Golbec [1] は、ネタバレを防止するため、Wikipedia<sup>2</sup>からあるドラマやスポーツの試合に関するキーワードを自動で抽出し、そのキーワードが登録されているリストを作成している。リストからテキストのネタバレを検知し、ミュートボタンを実装している。齊藤ら [2] は、スポーツ等の試合経過・結果に関するネタバレを防止するため、番組に関するハッシュタグを付加したツイートを「実況ツイート」とし、SVM によるネタバレ判別分類器を作成することによって、ネタバレであるツイートの推定を試みている。田島ら [3] は、アニメに関するネタバレを防止するため、ネタバレデータセットを構築し、連続する単語の一部の正規化や登場人物の名前の一般化などの事前処理を行った上で単語ベクトルを作成し、SVM を用いてネタバレであるツイートを推定している。

mute-words の拡張手法として、投稿者である Twitter ユーザに着目する手法も考えられる。ツイートに出現する語やユーザの推薦に関する研究として、渡邊ら [4] は、Twitter ユーザのツイートからそのユーザに特徴的な語を抽出し、その語を含むツイートを収集・分析することで関連語辞書を作成し、ユーザに特徴的な語と関連語辞書を用いて、ユーザの嗜好を分析しユーザの興味を表す語を抽出している。牧岡ら [5] は、久米ら [6] が提案した被推薦ユーザのツイートからスポーツや音楽などの興味のあるカテゴリを抽出し、そのカテゴリとツイートからユーザ間の類似度を算出することで趣味嗜好の合うユーザを推薦するシステムに加え、被推薦ユーザが関心のあるキーワードも考慮したユーザ推薦システムを提案している。

このように Twitter 上のネタバレを防止するために、特定の話題に関するネタバレツイートを推定する研究やユーザの嗜好に合った語やユーザを抽出する研究は行われているが、キーワードミュート機能に着目し、特定の話題を網羅的にミュートするための関連語を見つけることを目的とした研究は知られていない。

### 3 mute-words の拡張手法の提案

どのような語が実際に mute-words として使用されているかは重要な観点である。複数の Twitter ユーザを対象に行った、キーワードミュート機能の利用状況の調査を 3.1 で説明する。

予備調査の結果を受け、本研究では mute-words を拡張する以下の 2 つの手法を提案する。

(1) mute-words を含むツイートに、同時に出現しやすい語を関連語とみなし拡張を行う手法

(2) mute-words に関するツイートを頻繁に投稿するユーザが、積極的に使用する語を関連語とみなし拡張を行う手法

1 つ目の手法では、mute-words を含むツイートを収集し、収集したツイートの分析を行い、文書内に出現する単語の重要度を求める tf-idf 法を用いた拡張を行う。

2 つ目の手法では、mute-words に関するツイートを頻繁に投稿するユーザ、すなわち mute-words に関連の強いユーザを見

つける必要がある。まず、mute-words を含むツイートを収集し、そのツイートを投稿したユーザを抽出する。このユーザは、少なくとも 1 度は mute-words に関するツイートをしているため、mute-words に関連するユーザであるとみなす。このユーザ群を分析候補のユーザとし、分析を行い抽出した特に mute-words に関連の強いユーザを分析対象ユーザとする。次に、分析対象ユーザのツイートを収集し、分析を行う。一般的にユーザは 1 つの話題だけでなく複数の話題についてツイートを行うことが多いことから、分析対象ユーザのツイートをトピック分類することで、mute-words に関連の強いトピックに関するツイートのみを抽出する。最後に抽出したツイートの分析を行い、tf-idf 法を用いた拡張を行う。

mute-words を含むツイートの収集と分析対象ユーザの抽出を行う処理について 3.2 で説明する。続いて、mute-words を含むツイートの分析を行い tf-idf 法を用いた拡張を行う処理について 3.3 で説明する。最後に分析対象ユーザのツイートの収集と分析を行い tf-idf 法を用いた拡張を行う処理について 3.4 で説明する。

#### 3.1 キーワードミュート機能の利用状況に関する予備調査

Twitter におけるキーワードミュート機能を利用しているユーザが、どのような語を mute-words としているのかを明らかにするため、クラウドソーシングサービス Lancers を用いてアンケート調査<sup>2</sup>を行った。実施期間は 2020 年 5 月 26 日から 5 月 30 日の 5 日間である。回答人数は 85 人、mute-words の累積数は 266 個、種類数 233 個であった。

得られた mute-words を人手で集計した結果、設定されている mute-words は表 1 のように分類できることがわかった。また、mute-words を設定することで期待する効果を得られたかを 4 件法でアンケート調査し、平均値を満足度とした。

本研究では、表 1 のうち満足度の低かった項数 1 の政治や災害、感染症など多くのユーザが投稿するツイートに関する単語、項数 3 のスポーツやドラマ、漫画、映画の内容・結末などに関する単語に絞って研究を進めていく。

表 1: mute-words の分類結果

項数	概要	満足度
1	政治や災害、感染症など多くのユーザが投稿するツイートに関する単語	3.05
2	〇〇さんがいいねしました、拡散希望など タイムラインを埋めてしまう投稿に関する単語	3.20
3	スポーツやドラマ、漫画、映画の内容・結末などに関する単語	3.00

#### 3.2 分析対象ユーザの決定手法

mute-words に関連の強いユーザを見つけるため、まず、TwitterAPI を用い mute-words を含むツイートの収集を行う。件数は 1 日最大 2,000 件とし、9 日分のツイートを取得する。次に収集したツイートから投稿者を抽出し、抽出した投稿者を分析候補のユーザとする。分析候補のユーザから以下の 3 つの要素を用いてスコアを算出し、上位のユーザは mute-words に特に関連の強いユーザとみなし分析対象ユーザとする。

2 : <https://ja.wikipedia.org>

2 : 筑波大学図書館情報メディア系における倫理審査の通知番号: 第 20-11 号

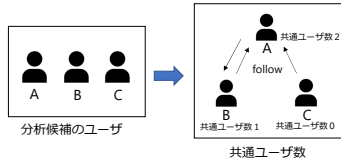


図 1: 共通ユーザ数の算出例

- 共通ユーザ数
- プロフィールに mute-words を含むかどうか
- mute-words を含むツイートの頻度

まず共通ユーザ数を取得する。あるキーワードを呟いているユーザ群から共通してフォローされているユーザは、そのキーワードに興味を持っているユーザたちの間で着目されているユーザである [5] という考えに基づいて、本研究でも、ある分析候補のユーザが他の分析候補のユーザにフォローされている人数を共通ユーザ数とする。例えば、図 1 の分析候補のユーザ A の共通ユーザ数は 2、B の共通ユーザ数は 1、C の共通ユーザ数は 0 である。

次にプロフィールを取得する。Twitter において、プロフィールはユーザの経歴や趣味、好きなものを記載することが多く、他のユーザはプロフィールを見ることでそのユーザがどのようなツイートをするか予測することができる。したがって、プロフィールに mute-words を含む場合は、そのユーザは mute-words に関するツイートを発する可能性が高いと考えられる。mute-words を含む場合はプロフィール値を 1、含まない場合は 0 とする。

共通ユーザ数とプロフィール値から各分析候補のユーザのスコアを算出する。ユーザ  $i$  のスコアは以下の計算式で算出する。

$$score_i = \alpha \frac{mutual_i - \min(mutual)}{\max(mutual) - \min(mutual)} + (1 - \alpha)profile_i \quad (1)$$

$mutual_i$  はユーザ  $i$  の共通ユーザ数、 $\max(mutual)$ 、 $\min(mutual)$  はそれぞれ最大共通ユーザ数と最小共通ユーザ数を指す。 $profile_i$  はプロフィール値を指す。 $\alpha (0 \leq \alpha \leq 1)$  は重み係数である。予備実験として、 $\alpha$  の値を 0.0 から 1.0 まで 0.1 刻みで変化させ、上位のユーザを手で確認し、比較したところ、 $\alpha = 0.8$  において特に mute-words に関連の強いユーザを抽出できたため、重み係数の値を  $\alpha = 0.8$  とした。

最後に mute-words を含むツイートの頻度を取得する。mute-words を含むツイートの頻度が高いユーザは、他のツイートも mute-words に関するツイートである可能性が高いと考えられることから、式 (1) で求めたスコア上位 20 人のユーザの最新ツイートを 2,000 件ずつ取得し、一日のうちの mute-words を含むツイートの平均回数を求める。

式 (1) で求めたスコアと mute-words を含むツイートの頻度から各分析候補のユーザの総合スコアを算出し、上位 9 人を分析対象ユーザとする。ユーザ  $i$  の総合スコアは以下の計算式で算出する。

$$score'_i = \beta score_i + (1 - \beta) \frac{N_i - \min(N)}{\max(N) - \min(N)} \quad (2)$$

$N_i$  はユーザ  $i$  の 1 日あたりの mute-words を含むツイートの平均数を指す。 $\max(N)$ 、 $\min(N)$  はそれぞれ 1 日あたりの mute-words を含むツイートの平均数の最大値と最小値を指す。 $\beta (0 \leq \beta \leq 1)$  は重み係数である。予備実験として、 $\beta$  の値を 0.0 から 1.0 まで 0.1 刻みで変化させ、上位のユーザを手で確認し、比較したところ、 $\beta = 0.8$  において特に mute-words に関連の強いユーザを抽出できたため、重み係数の値は、 $\beta = 0.8$  とした。

### 3.3 mute-words を含むツイートの分析

3.2 で取得した mute-words を含むツイートを日付毎に分ける。ツイートからハッシュタグを抽出した後、MeCab を用いてツイートの形態素解析を行い、名詞を抽出する。Twitter は新語や固有表現が出現しやすいため、辞書には週 2 回以上更新されるオープンソフトウェアであり、新語や固有表現に強いという特徴と持つ mecab-ipadic-NEologd を用いる。抽出したハッシュタグと名詞を tf-idf 法の tf 値と tf-idf 値で重みづけを行い、それぞれ上位 X 語を mute-words の関連語とする。

### 3.4 分析対象ユーザのツイートの分析

3.2 で決定した分析対象ユーザ 9 人の最新ツイートを最大 2,000 件ずつ取得する。取得したツイートをトピック分類し、mute-words に関連の強いトピックのツイートを分析する。

#### 3.4.1 トピック分類

トピック分類は、python のライブラリ gensim にある LDA モデルを用いて行う。前処理として、ユーザのツイートからハッシュタグを抽出した後、形態素解析を行い、名詞を抽出する。

1 ユーザのツイートを 1 文書とし、2 文書未満にしか出現しない低頻度語はノイズとして除去する。トピック数は、2 から 20 まで 1 刻みで変化させ、トピックモデルの指標として用いられる perplexity と coherence を参考にし、トピック数を決定する。

最も mute-words と関連が強いと考えられるトピックは、分類したトピックから人手で選択する。

#### 3.4.2 ツイートの抽出と分析

3.4.1 で選択したトピックに関するツイートを抽出するため、トピックの単語の出現確率を用いて各ツイートのスコアを算出し、閾値以上のツイートをそのトピックのツイートとし、分析を行う。まず、各ツイートからハッシュタグを抽出した後、形態素解析を行い、名詞を抽出する。抽出したハッシュタグと名詞の出現確率の和を求める。このとき、ツイートが長い、すなわち単語数が多いとスコアが高くなることから、各ツイートのスコアを全ツイートのスコアの最大値で割り、スコアの値を 0 から 1 に補正する。閾値以上のツイートをそのトピックに関するツイートとする。

選択したトピックに関するツイートからハッシュタグを抽出した後、形態素解析を行い、名詞を抽出する。抽出したハッシュタグと名詞を tf-idf 法の tf 値と tf-idf 値で重みづけを行い、それぞれ上位 X 語を mute-words の関連語とする。

予備実験として、ツイートの点数の閾値を 0.0 から 1.0 まで 0.1 刻みで変化させ、抽出したツイートを人手で確認したとこ

表 2: データセット

	西武ライオンズ	コロナ	鬼滅の刃
期間	2020/10/29-11/06		
ツイート件数 (件)	5,318	17,705	17,122
分析候補のユーザ数 (人)	3,177	15,838	15,376

る, 0.0 と 0.1 では, mute-words と関係のないツイートも存在したが, 0.2 からは, mute-words と関連するツイートを抽出することができた. そのため, 閾値は, 0.2 とした.

## 4 評価実験

### 4.1 データセット

分析するためのツイートを収集するため, mute-words を決める. 本研究では, 表 1 の項数 1 の政治や災害, 感染症など多くのユーザが投稿するツイートに関する単語として新型コロナウイルスを指す「コロナ」, 項数 3 のスポーツやドラマ, 漫画, 映画の内容・結末などに関する単語として野球チームの「西武ライオンズ」, 項数 1・3 両方に関するキーワードとして漫画 (アニメ) タイトル「鬼滅の刃」を mute-words として実験を行う.

3.2 で述べた手順で各 mute-words を含むツイートの収集を行う. 重複するツイートは削除した. データセットの詳細は表 2 に示す. 「コロナ」と「鬼滅の刃」は, 1 日に 2,000 件以上のツイートを投稿されていたため, 1 日の上限として定めた 2,000 件を収集することができたが, 「西武ライオンズ」は, 1 日のツイート件数は平均 600 件程度であったため, ツイート件数は少なくなっている.

### 4.2 トピック分類の結果

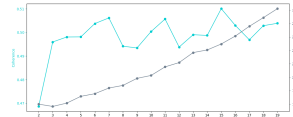
#### 4.2.1 トピック数の決定

4.1 のデータセットに対して 3.4.1 で述べた手順でトピック分類を行った.

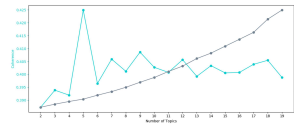
トピック数を検討するために求めた各 mute-words の Perplexity と Coherence をそれぞれ図 2 に示す. 横軸はトピック数, 右縦軸は Perplexity, 左縦軸は Coherence の値を示している. Perplexity はモデルの予測精度の評価指標であり, 値が低いほど良いモデルであると見なされる. Coherence はトピックの品質を測る評価指標で, 値が高いほどトピックの品質が良いと見なされる. そのため, Perplexity と Coherence の差の大きさを参考にトピック数を検討する.

「西武ライオンズ」は, 図 2a から Perplexity と Coherence の差が大きいトピック数 3, 4, それぞれでトピック分類を行い, 各トピックの出現確率上位の語を手で確認した. トピック数 3 の方がより上位の語に違いが見られ, 分類されていると考えられるためトピック数は 3 とした.

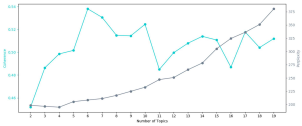
「コロナ」は, 図 2b から, Perplexity と Coherence の差が大きいのはトピック数 5 であることが分かる. トピック分類を行い, 各トピックの出現確率上位の語を手で確認したところ, 上位の語に違いが見られ, 分類されていると考えられるためトピック数は 5 とした.



(a) 「西武ライオンズ」



(b) 「コロナ」



(c) 「鬼滅の刃」

図 2: トピック数

表 3: 「西武ライオンズ」の各トピックの出現確率上位 20 語

topic0	topic1	topic2
#seibulions	#seibulions	#seibulions
#埼玉西武ライオンズ	#埼玉西武ライオンズ	#埼玉西武ライオンズ
選手	選手	西武
投手	投手	選手
西武	西武	投手
ライオンズ	試合	プロ野球
試合	戦	ロッテ
戦	ライオンズ	楽天
先発	チーム	試合
#プロ野球	監督	戦
ロッテ	先発	野球
プロ野球	#プロ野球	監督
監督	プロ野球	開催
チーム	#中村剛也	東京
#中村剛也	結果	新型コロナウイルス
勝利	楽天	#プロ野球
ファン	ロッテ	球団
#森友哉	メットライフドーム	開幕
辻	ファン	ライオンズ
球団	#栗山巧	チーム

「鬼滅の刃」は, 図 2c から, Perplexity と Coherence の差が大きいのはトピック数 6, 7, それぞれでトピック分類を行い, 各トピックの出現確率上位の語を確認したところ, どちらも上位の語に違いの見えないトピックが存在し, 分類できていないと考えられる. 次にトピック数 4 でトピック分類を行い各トピックの出現確率上位の語を確認したところ, 上位の語に違いが見られ, 分類されていると考えられるため, トピック数は 4 とした.

#### 4.2.2 分析するトピックの決定

最も mute-words に関連の強いトピックを見つけるため, 各 mute-words のトピック分類結果の出現確率上位の単語を手で確認した. 各 mute-words のトピック分類結果の出現確率上位の 20 語をそれぞれ表 3, 表 4, 表 5 に示す.

表 3 から, 「西武ライオンズ」の全てのトピックは, 主に以下の要素で構成されていることが分かる.

- 西武ライオンズを示す単語
  - 西武ライオンズに所属する選手名や監督名
  - 他の野球チーム名, 「開催」や「開幕」など野球界全体に関する単語
  - 「投手」や「先発」などの野球用語, 勝ち負けなどの試合内容・結果に対する単語
  - 「ファン」などの観戦者に関する単語
- 西武ライオンズを示す単語と西武ライオンズに所属する選手

表 4: 「コロナ」の各トピックの出現確率上位 20 語

topic0	topic1	topic2	topic3	topic4
感染	感染	感染	感染	感染
確認	発表	確認	確認	確認
合計	確認	#新型コロナウイルス	米	新型コロナウイルス
詳細	新型コロナ	詳細	新型コロナ	発表
#新型コロナウイルス	米大統領選	感染者数	米大統領選	米
発表	新型コロナウイルス	国内	バイデン	米大統領選
国内	バイデン	合計	新型コロナウイルス	バイデン
感染者数	影響	発表	発表	影響
新型コロナ	国内	新型コロナ	トランプ氏	新型コロナウイルス
点	米	点	影響	トランプ氏
新規感染者	死亡	米大統領選	対策	警戒
新型コロナウイルス	トランプ氏	新型コロナウイルス	コロナ	必要
速報	対策	バイデン	逮捕	コロナ
全国	女性	死亡	警戒	県内
北海道	コロナ	米	死亡	対策
死亡	#新型コロナウイルス	感染者	台風	死亡
感染者	全国	新規感染者	全国	逮捕
米大統領選	逮捕	トランプ氏	東京	見直し
神奈川県	公開	影響	可能性	台風
バイデン	感染者	東京都	女性	可能性

表 5: 「鬼滅の刃」の各トピックの出現確率上位 20 語

topic0	topic1	topic2	topic3
JUMPSHOP	JUMPSHOP	鬼滅の刃	鬼滅の刃
鬼滅の刃	発売	セット	グッズ
発売	鬼滅の刃	JUMPSHOP	受注
写真	公開	発売	紹介
入場	新型コロナ	グッズ	発売
商品	商品	商品	イラスト
入店	写真	フィギュア	JUMPSHOP
店舗	確認	受注	受付
税	登場	キャラクター	#鬼滅の刃
お客様	感染	缶バッジ	缶バッジ
登場	受注	写真	劇場版
受注	グッズ	竈門炭治郎	新型コロナ
来店	店舗	感染	ufotable
発表	入場	編	公開
大阪	大阪	#鬼滅の刃	ufotableWEBSHOP
原作	開催	発表	感染
新型コロナ	紹介	紹介	カフェ
グッズ	開始	登場	登場
フェア	キャラクター	福豆	特典
感染	発表	予定	確認

名や監督名を上位により多く含むトピックを選択することで、特に「西武ライオンズ」に関するツイートを抽出することができると考えられる。上位 20 語を確認したとき、トピック 0 とトピック 1 は西武ライオンズを示す単語と西武ライオンズに所属する選手名や監督名を同数程度含んでいるため、上位 50 語までの単語を手で確認した結果、トピック 0 の方が多く含んでいたため、分析するトピックは、トピック 0 とした。

表 4 から、「コロナ」の全てのトピックは、主に以下の要素で構成されていることが分かる。

- 新型コロナウイルスを示す単語
- 「感染」や「対策」などの新型コロナウイルスに対する単語
- 新型コロナウイルスの感染者がでている国や都道府県、地域
- アメリカの大統領選挙に関する単語

新型コロナウイルスを示す単語と「感染」や「対策」などの新型コロナウイルスに対する単語を上位により多く含むトピックを選択することで、特に「コロナ」に関するツイートを抽出することができると考えられる。上位 20 語を確認したとき、トピック 0 とトピック 1、トピック 2 は新型コロナウイルスを示す

単語と「感染」や「対策」などの新型コロナウイルスに対する単語を同数程度含んでいるため、上位 50 語までの単語を手で確認した結果、トピック 2 に最も多く含まれていたため、分析するトピックは、トピック 2 とした。

表 5 から、「鬼滅の刃」の全てのトピックは、主に以下の要素で構成されていることが分かる。

- 鬼滅の刃を示す単語
- 鬼滅の刃の制作会社やグッズを販売する店舗名
- 「受注」や「缶バッジ」などグッズに関する単語
- 「公開」や「劇場版」など映画に関する単語
- 登場人物の名前
- 新型コロナウイルスに関する単語

鬼滅の刃を示す単語と登場人物の名前を上位により多く含むトピックを選択することで、特に「鬼滅の刃」に関するツイートを抽出することができると考えられる。上位 20 語を確認したとき、鬼滅の刃を示す単語と登場人物の名前はトピック 2 に最も多く含まれていたため、分析するトピックは、トピック 2 とした。

選択したトピックに対して、3.4.2 で述べた手順でトピックに関するツイートの抽出を行った。ツイート件数は「西武ライオンズ」は 5,280 件、「コロナ」は 1,686 件、「鬼滅の刃」は 742 件であった。

### 4.3 正解データセット

提案手法において拡張した関連語から、mute-words に関するツイートをどの程度網羅的にミュートできるか評価を行うため、正解データセットを作成する。正解データセットは、mute-words に関連するツイートと mute-words に関連しないツイートで構成される。mute-words に関連するツイートを適合データとし、mute-words に関連しないツイートを非適合データとする。

#### 4.3.1 適合データ

mute-words に関連するツイートのデータを作成する。特定の話題についてツイートをすると、ハッシュタグと呼ばれる検索およびタグ付けを可能とするキーワードをツイートに付与することが多いため、「#[mute-words]」というハッシュタグをキーワードとして TwitterAPI を用いてツイートの収集を行う。収集したツイートから、「#[mute-words]」を削除する。また、mute-words を含むツイートを削除した。

収集したツイートが mute-words に関連するツイートであるかの判別を手で行った。以下に当てはまるツイートを適合データとする。

- ツイートのテキストが mute-words に関連する
  - ツイートのテキストは mute-words に関連すると判断できないが、リツイートと呼ばれる他人のツイートを引用できる機能を用い、その内容が mute-words に関連する
  - ツイートのテキストは mute-words に関連すると判断できないが、URL の内容を見ると mute-words に関連する
- 作成した適合データは、「西武ライオンズ」は 576 件、「コロナ」は 1,125 件、「鬼滅の刃」は 1,344 件となった。

### 4.3.2 非適合データ

mute-words に関連しないツイートのデータを作成する。TwitterAPI を用い、日本語で投稿されたツイートの中から無作為に 2,000 件収集した。収集したツイートから mute-words に関するツイートを人手で削除し、残ったツイートを非適合データとした。各 mute-words の適合データと同数を非適合データからランダムで抽出し、組み合わせたものを正解データセットとする。

## 4.4 評価実験

### 4.4.1 評価尺度

以下の 2 つの提案手法に対して、mute-words に関するツイートをどの程度ミュートできるか評価を行う。

- 提案手法 1: mute-words を含むツイートに、同時に出現しやすい語を関連語とみなし拡張を行う手法
- 提案手法 2: mute-words に関するツイートを頻繁に投稿するユーザが、積極的に使用する語を関連語とみなし拡張を行う手法

評価尺度として、適合率 (Precision) と再現率 (Recall) を用いる。作成した正解データセットの各ツイートに関連語を含む場合は mute-words に関連するツイートであると仮定する。予測結果が正解とされるのは、正解データセットの適合データ内のツイートを mute-words に関連するツイートであると予測したときである。適合率は mute-words に関連するツイートであると予測したものに対する正解率 (=mute-words に関連するツイートであると予測した結果が正解であった件数/mute-words に関連するツイートであると予測した全件数)、再現率は適合データ内のツイートをどの程度検知できたかという検知率 (=mute-words に関連するツイートであると予測した結果が正解であった件数/適合データの全件数) を表す。

キーワードミュート機能を利用するユーザにとって、mute-words に関連しないツイートをミュートしてしまうことより、mute-words に関するツイートを網羅的にミュートすることが重要であると考え、本研究では再現率に特に注目して評価を行う。

適合率と再現率は、関連語の数を、10, 20, 30, 50, 70, 100, 150, 200 と変化させ、それぞれの語数に対して算出する。また、関連語数内の tf 値の語数と tf-idf 値の語数の割合も、0:1 から 1:0 まで 1 刻みで変化させそれぞれ算出する。

### 4.4.2 結果

各 mute-words のそれぞれの語数の最も再現率の高い結果を得られた tf 値と tf-idf 値の割合における再現率とそれに対応する適合率の値を表 6, 表 7, 表 8 に示す。提案手法によらず、どの mute-words も再現率は関連語数が増えるにつれ高い値が得られた。適合率は再現率が高くなるにつれ低くなっていく傾向にあった。提案手法 1 において最も再現率が高い値は、「西武ライオンズ」は 0.815, 「コロナ」は 0.930, 「鬼滅の刃」は 0.903 である。提案手法 2 において最も再現率が高い値は、「西武ライオンズ」は 0.930, 「コロナ」は 0.745, 「鬼滅の刃」は 0.809 である。

また、提案手法 1 と提案手法 2 の再現率を比較したとき、「西

武ライオンズ」は関連語数が 50 のときを除き提案手法 2 の方が高く、「コロナ」と「鬼滅の刃」は全ての関連語数において提案手法 1 の方が高かった。

表 6: 「西武ライオンズ」の適合率と再現率

語数	提案手法 1		提案手法 2	
	適合率	再現率	適合率	再現率
10	0.992	0.677	0.969	0.719
20	0.965	0.716	0.947	0.747
30	0.960	0.756	0.954	0.801
50	0.896	0.825	0.923	0.817
70	0.872	0.836	0.907	0.871
100	0.828	0.848	0.865	0.885
150	0.828	0.881	0.817	0.925
200	0.815	0.892	0.784	0.930

表 7: 「コロナ」の適合率と再現率

語数	提案手法 1		提案手法 2	
	適合率	再現率	適合率	再現率
10	0.949	0.521	0.956	0.429
20	0.885	0.657	0.947	0.546
30	0.861	0.705	0.936	0.563
50	0.819	0.791	0.934	0.576
70	0.792	0.822	0.915	0.617
100	0.764	0.858	0.858	0.646
150	0.723	0.908	0.844	0.711
200	0.693	0.930	0.854	0.745

表 8: 「鬼滅の刃」の適合率と再現率

語数	提案手法 1		提案手法 2	
	適合率	再現率	適合率	再現率
10	0.979	0.593	0.895	0.142
20	0.945	0.693	0.855	0.165
30	0.920	0.733	0.844	0.268
50	0.883	0.769	0.856	0.396
70	0.868	0.807	0.845	0.477
100	0.834	0.842	0.800	0.566
150	0.781	0.874	0.817	0.750
200	0.743	0.903	0.809	0.784

各 mute-words において、最も高い再現率が得られたとき、mute-words に関連のあるツイートであるが、予測できなかったツイートは以下のような特徴があることが分かった。

- 西武ライオンズ
  - 関連語にない選手名や名前
  - 「2-3」のような点数結果
- コロナ
  - テキストは「やばい！」など一言のみで、「コロナ」に関する URL や画像が添付されている
- 鬼滅の刃
  - 関連語にない登場人物の名字や名前

- 登場人物の名前を平仮名にしている
- 登場人物のセリフ
- テキストは「見た！」など一言のみで、「鬼滅の刃」に関する画像が添付されている

#### 4.5 考察

##### 4.5.1 トピック分類の結果における考察

4.2.2 から、トピック分類の結果、全てのトピックは各 mute-words に関連し、表 3, 表 4, 表 5 から、全てのトピックに共通する単語も複数存在するという結果が得られた。このような結果になった要因として、分析対象ユーザをあげる。各 mute-words の分析対象ユーザを手で確認したところ、全ての分析対象ユーザは mute-words に関連するユーザを抽出することができていた。すなわち、分析対象ユーザは全員 mute-words に関連するツイートを行うユーザである。さらに、3.4.1 の処理で 1 ユーザのツイートを 1 文書とし、2 文書未満にしか出現しない低頻度後はノイズとして除去していることから、残った単語の多くは mute-words に関連する単語になり、かつ mute-words に関連する単語の頻度が特になくなっていく可能性が考えられる。そのため、全てのトピックが mute-words に関連するトピックになったと考えられる。

しかし、4.2.2 でトピックによって各要素の単語の数は異なることが分かっていることから、各トピックに何かしらの特徴がある可能性が考えられるため、最もトピック数の多い「コロナ」の各トピックに特徴はあるか考察を行う。4.2.2 では、出現確率上位 50 語を確認したが、ここでは上位 200 語を手で確認した。「コロナ」の全てのトピックは主に以下の要素で構成されていた。

- 「感染」や「対策」、「症状」など新型コロナウイルスに対する単語
- 新型コロナウイルスの感染者がでている国や都道府県、地域
- 「中止」や「学校」など新型コロナウイルスによる影響を受けた状況・施設等に関する単語
- 日本の政治家の名前や政治に関する単語、アメリカの大統領選挙に関する単語

各トピックの特徴として、トピック 0 は新型コロナウイルスに対する単語が上位に多く、次に新型コロナウイルスの感染者がでている国や都道府県、地域が多く存在した。トピック 1 は新型コロナウイルスに対する単語が上位に多く、次に新型コロナウイルスによる影響を受けた状況・施設等に関する単語が多く存在した。トピック 2 は新型コロナウイルスに対する単語が上位に多く、他はどれかに特化することなく全ての要素に関する単語がほぼ均一に存在した。トピック 3 は全体的に日本の政治家の名前や政治に関する単語、アメリカの大統領選挙に関する単語が多く存在した。トピック 4 は日本の政治家の名前や政治に関する単語、アメリカの大統領選挙に関する単語が上位に多く、次に新型コロナウイルスによる影響を受けた状況・施設等に関する単語が多く存在した。

このことから、本研究では「コロナ」を網羅的にミュートす

ることを目的としトピックを選択したが、「コロナ」という話題の中でも特に日本の政治やアメリカの大統領選挙に関する話題をミュートしたいと考えるユーザが存在したとき、トピック 3 を選択することでよりユーザの期待する範囲でのミュートをできる可能性も考えられる。

##### 4.5.2 評価実験の結果における考察

再現率に着目し、提案手法の有効性について考察する。4.4.2 で述べたように「西武ライオンズ」は関連語数 50 のときを除いて提案手法 2 の方が高い精度の結果が得られた。対して、「コロナ」と「鬼滅の刃」は提案手法 1 の方が高い精度の結果が得られた。このような結果になった要因として大きく 2 つ考えられる。

1 つ目は、提案手法 2 における分析対象ユーザである。各 mute-words の分析対象ユーザを確認すると、どのアカウントも mute-words に関するツイートを行っていたが、「西武ライオンズ」は「西武ライオンズ」を好きな個人が運用する一般アカウントが多いのに対して、「コロナ」と「鬼滅の刃」はニュース記事のツイートを行う公式アカウントやエンタメ情報を発信する公式アカウントが多く見られた。一般的に、ニュースやエンタメ情報を発信するアカウントのツイートは、ツイートにより詳しい情報が書かれている記事の URL を載せているものが多い。そのため、ツイートのテキストは URL 先の内容を簡単に説明するものが多く、得られる情報量は少ない傾向がある。したがって、「コロナ」と「鬼滅の刃」の提案手法 2 において精度が低くなったと考えられる。一方で、「西武ライオンズ」に多く存在した西武ライオンズのファンである一般アカウントはいわゆる趣味垢であり、西武ライオンズの試合状況に関する詳細や感想のツイートをすることが多くなっていた。1 ツイート毎の情報量が多いことから提案手法 2 において高い精度の結果が得られたと考えられる。

このように、分析対象ユーザの特徴に違いがでた要因について、mute-words の特徴をあげる。「コロナ」と「鬼滅の刃」のように表 1 の項数 1 の特徴を持つ mute-words は、メディアに大きく取り上げられる傾向にあるため、年代・趣味嗜好を問わず多くのユーザによってツイートが投稿される。そのため、ニュースやエンタメ情報を発信するアカウントといった、mute-words に関連するが、他の話題に関してもツイートをを行うアカウントが抽出されたと考えられる。一方で、「西武ライオンズ」のように表 1 の項数 3 の特徴を持つ mute-words は、mute-words の話題を主とする趣味垢が存在する可能性が高いと考えられるため、mute-words に関するツイートを頻繁に行うユーザを多く抽出できたと考えられる。ただし、「鬼滅の刃」のように社会的に大ヒットし、メディアに大きく取り上げられるものは例外になると考えられる。

2 つ目はツイート数である。提案手法 1 において分析するツイート数は「西武ライオンズ」が 5,318 件、「コロナ」は 17,705 件。「鬼滅の刃」は 17,122 件となっており、提案手法 2 において分析するツイート数は「西武ライオンズ」は 5,280 件、「コロナ」は 1,686 件、鬼滅の刃は 742 件であった。「西武ライオンズ」のツイート数は提案手法 1 と提案手法 2 でほぼ変わらない

いが、「コロナ」と「鬼滅の刃」は大きく異なる。提案手法2において「コロナ」と「鬼滅の刃」はツイート数が少なかったため、分析するための情報量が足りず、提案手法1に比べ精度が低くなった可能性があると考えられる。

提案手法2において分析するツイート数は、分析対象ユーザのツイートのトピック分類をした後、最も mute-words に関連の強いと考えられるトピックに関するツイートを抽出したことにより少なくなったと考えられる。したがって、必要なツイート件数を定め、その件数に達するまで分析対象ユーザを増やすなどの工夫をすることで、より精度があがる可能性があると考えられる。また、提案手法2において精度の低かった「コロナ」と「鬼滅の刃」の分析するツイート件数の内訳を確認する。「コロナ」の1ユーザあたりのツイート件数で最も多い件数は1,058件、最も少ない件数は10件となっていた。「鬼滅の刃」の1ユーザあたりのツイート件数で最も多い件数は291件、最も少ない件数は4件となっていた。どちらも1ユーザあたりのツイート件数の差が大きく見られた。本研究で用いた tf-idf 値には文書ごとの単語数の差による影響が大きく出てしまうという欠点が存在することから、1ユーザごとのツイート件数の差が大きい、すなわち文書ごとの単語数の差が大きいため、tf-idf 値に影響がでてしまい、精度が低くなってしまった可能性も考えられる。しかし、「西武ライオンズ」の分析するツイート件数の内訳を確認すると、1ユーザあたりのツイート件数で最も多い件数は1,967件、最も少ない件数は26件と差は大きい、高い精度が得られている。そのため、精度が上がらない可能性も考えられるが、tf-idf 法で使用した単語の出現頻度と単語の逆文書頻度に加え、文書内の総単語数に関する情報を用いることで、文書の単語数を考慮する BM25 という手法を用いることで、より高い精度の結果が得られる可能性も考えられる。

次に、mute-words に関連するが、予測できなかったツイートについて考察する。4.4.2で述べたように、mute-words に関連するが予測できなかったツイートの特徴を見つけることができた。これらの特徴を持つツイートも予測できる関連語を見つけることが今後と課題となるが、「西武ライオンズ」と「コロナ」の特徴にある、テキストは短文だが、URL や画像が添付されているツイートに関しては、テキストではなく URL 先や画像を見ることによって mute-words に関係していると判断できるため、関連語によって予測することが難しいと考えられる。また、「西武ライオンズ」と「鬼滅の刃」の特徴にある、関連語にない選手名や名前、関連語にない登場人物の名前や名前では、関連語に名字はあるが名前はなかったため、予測できなかったツイートも存在する。そのため、人物名は名字と名前どちらか片方だけでなく、両方を関連語とすることでより精度があがる可能性があると考えられる。

## 5 おわりに

本研究では、Twitter のキーワードミュート機能に着目し、特定の話題に関するツイートを網羅的にミュートするための関連語を見つけることを目的とした2つの mute-words 拡張手法を

提案した。キーワードミュート機能とは、設定したキーワードを含むツイートを非表示(ミュート)にする機能である。本論文では、設定するキーワードのことを mute-words と呼ぶ。提案手法は、ツイート中に出現する mute-words との共起語を手がかりとする手法と mute-words を頻繁に投稿するユーザを手がかりとする手法からなる。

第一の提案手法では、mute-words を含むツイートを収集し分析することで拡張を行っている。第二の手法では、mute-words に関するツイートを頻繁に投稿するユーザ、つまり mute-words に関連の強いユーザを見つけるため、mute-words を含むツイートから投稿者を抽出し、分析を行うことで、特に mute-words に関連の強いユーザを抽出した。抽出したユーザを分析対象ユーザとし、そのユーザのツイートを分析することで拡張を行った。

拡張を行い得られた関連語を適合率と再現率を用い評価実験を行った結果、高い精度の結果を得ることができた。しかし、mute-words によって有効な提案手法は異なり、その要因として第二の手法で抽出した分析対象ユーザの特徴、またはツイート数が関わっていると考えられた。

今後の課題としては3つ挙げられる。1つ目は情報量の少ない分析対象ユーザの抽出が挙げられる。この課題の解決には、分析対象ユーザの抽出手法を改善する必要があると考えられる。2つ目は第二の手法において分析するためのツイート数が少なかったことが挙げられる。この課題の解決には、必要とするツイートの件数を設定し、その件数に達するまで分析対象ユーザの数を増やすなどの工夫が必要だと考えられる。3つ目は3つの mute-words でしか実験を行っていないことが挙げられる。本研究の有効性をより確かにするため、上記の課題を解決した上で、異なる mute-words による実験を行う必要があると考えられる。

## 文 献

- [1] Jennifer Golbeck. The twitter mute button: a web filtering challenge. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2755–2758, 2012.
- [2] 齊藤令, 寺田実. Twitter の実況ツイートを利用したタイムライン上のネタバレ情報検知. 情報処理学会第78回全国大会講演論文集, Vol. 2016, No. 1, pp. 539–540, 2016.
- [3] 田島一樹, 中村聡史. Twitter におけるアニメのネタバレツイート判定手法の提案. 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), Vol. B5-4, pp. 1–8. 電子情報通信学会, 2016.
- [4] 渡邊恵太, 加藤昇平. Twitter における語の関連性に着目したユーザ興味語抽出手法の提案. 人工知能学会全国大会論文集 第26回全国大会 (2012), p. 1B1R37. 人工知能学会, 2012.
- [5] 牧岡竜人, 平川豊, 大関和夫. 興味領域とキーワードを考慮した twitter ユーザ推薦に関する考察. 情報処理学会第79回全国大会講演論文集, Vol. 2017, No. 1, pp. 467–468, 2017.
- [6] 久米雄介, 打矢隆弘, 内匠逸. 興味領域を考慮した twitter ユーザ推薦手法の提案と評価. 情報処理学会研究報告. ICS, [知能と複雑系], Vol. 2015, No. 1, pp. 1–8, 2015.