D-011

# 教師あり学習による信憑性の低いツイートの収集
# Collecting Untrustworthy Tweets through Supervised Learning

藤浦 礼奈[1)]　鈴木 優[1)]
Reina Fujiura　Yu Suzuki

## 1　はじめに

本稿では，コロナウィルス蔓延などの社会問題に関するツイートの中から，信憑性が低いツイートを収集する方法を提案する．本研究では，多数の人が納得できる情報源を作成することを最終的な目標としているため，モデルで判定されたツイートのラベルを，人手で再度判定する．モデルにより正確に判定できないツイートに対して，人手で再度判定することにより，モデルの境界線をより正確なものにできると考えたためである．しかし，モデルの精度を理解しなければ，精度を高める方法を判明させることはできない．そこで本研究では，信憑性が低いツイートを自動的に収集・分類する過程に焦点を当てている．

## 2　関連研究

Wikipedia の情報の質について述べた我々の研究 [1] では，Wikipedia の情報の質を測定する方法をまとめている．その中でも半自動的な質の測定方法である Mizzaro[2] は，論文の査読を自動的に行う方法を提案している．

Mizzaro の研究では，論文の質を評価するために，全体的な品質を示す一つのスコアを測定する方法を提案している．論文と購読者は，それぞれスコアを持っている．論文のスコアは，読者の判断により動的に更新される．購読者のスコアは，著者ならば自分の書いた論文の評価で，読者ならば判断力で，など，購読者自身の行動により更新される．また，論文と購読者は，スコアの更新に影響を与える，安定値という価値を持っており，安定値はスコアの更新に影響を与える．また，安定値は信頼できるかどうかを推定している．それぞれのスコアはリンクしており，論文のスコアは，その論文を読むかどうかを決定するために使用できる．

Mizzaro の研究では，評価対象と評価者全てに値を与え，それぞれのスコアの値の変動がリンクし，動的に更新されている．しかし，我々の研究では人手の介入はさせるが，評価対象や評価者に値を割り振ることなく，評価対象のみに人の意見を反映させている点が異なる．

## 3　提案手法

図 1 において，提案手法の全体図を示す．

(1)集めたツイートを信憑性が低い，高い，関係無いのうちどれに属するかラベル付けをする．ラベル付けしたツイートを単語の出現回数による特徴ベクトルに変換し，これらを訓練データとしたモデルを構築する．3.1 節で説明する．

(2)ラベル付けしていないツイートを特徴ベクトルにしたものを，作成したモデルに入れて判定する．3.2 節で説明する．

(3)モデルで判定してラベルが付与されたツイートを，人手を用いてラベルが正しいかどうかを判定する．間違ったラベルが付与されていたら，正しいラベルを再度付与し，(1) でラベル付けしたツイートに追加し，再度モデルを構築する．3.3 節で説明する．

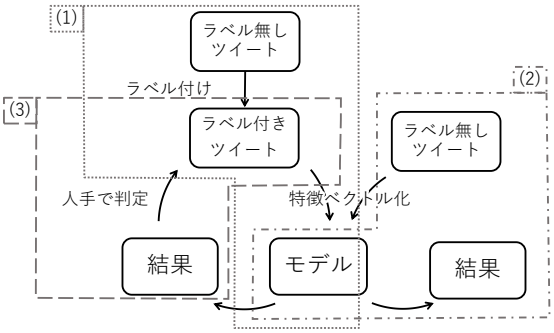


図 1　提案手法の流れ

### 3.1　モデルの構築

モデルを構築するために，ツイートに対して信憑性が低い，高い，関係無いの 3 種類のラベルを付与する．以下このツイート群を $T$ とする．次に，それぞれのラベルが付与されたツイート数が同数となるようにツイートを収集する．

次に，$T$ に含まれるツイートに対して形態素解析を行う．名詞と動詞だけを形態素解析により取り出したが，これはツイートの特徴を表すのではないかと考えたためである．このようにして得られた単語から，特徴ベクトルに変換する．特徴ベクトルは (1) 式のように定義する．

$$\vec{v} = [t_1, t_2, \cdots, t_n] \tag{1}$$

$\vec{v}$ はツイートの特徴ベクトル，$t_i$ はツイートに含まれる単語 $i$ の出現回数である．

特徴ベクトルにした $T$ を訓練データとして用いて，SVC (Support Vector Classification) によりモデルを作成する．SVC を用いた理由は，教師あり学習のカテゴリ分類ができて，かつデータ数が少なくても実装可能であると考えたためである．また，他の分類器を用いても同じように実装することができるが，どの分類器が適切であるかどうかは今後比較検討する必要がある．

### 3.2　分類

テストデータとして，訓練データに用いていない $T$ を特徴ベクトルにしたものを，作成したモデルに入れて判定する．テストデータであるツイートそれぞれに，分類器によってどのラベルが付与されたかを集計する．この結果，信憑性が低いと正しく判定されたツイートは，警告のラベルを付与して表示する．

### 3.3　人手によるラベルの判定

多数の人が納得できる情報源を作成するために，モデルに入れて分類されたツイートのラベルが，本当に正しいのかを人手で判定する．付与されたラベルが正しければそのままにし，間違っていたら正しいと思うラベルを再度付与する．人手で判定されたツイートは，モデルを

1)　岐阜大学 工学部 電気電子・情報工学科

表 1　ツイート例

| 信憑性 | 例 |
|---|---|
| 低い | 愛知県の緊急事態宣言を未だに解除しないのはコロナのせいでは無い |
| 高い | 埼玉県志木市でコロナ患者さんが確認。近いところで危険。 |
| 関係無い | 飲みの予定が 8 件くらいできたので落ち着いたらみんな飲もうね。コロナ〜〜 |

構築するためのツイートとして，再利用する．

今回の実験では，この節に記述された実装は行っていない．

## 4　評価実験

4 章では，SVC により信憑性が低いツイートかどうかを正しく判定できるかどうかを確かめることを目的とし，3.1 節と 3.2 節で述べた部分の実装を行った．

### 4.1　実験方法

本実験で用いるツイート群 $T$ を第一著者により人手で 13 件ずつ集めた．収集したツイートの一部を表 1 に示す．信憑性が低いツイートは，一般的に公開されている情報では無いもの，また事実関係が明確でないものとした．信憑性が高いツイートは，一般的に公開されている情報であるもの，またツイートの発信者の情報に関するものとした．関係無いツイートは，それ以外のものとした．例えば一番上のツイートについて，愛知県が緊急事態宣言を解除しないこととコロナウィルスとの関係が無いことは情報公開されていないため，信憑性が低いと判定している．一方で，二番目のツイートは埼玉県志木市でコロナウィルスに罹患した患者が発生したことは公開されているため，信憑性が高いと判定している．

次に，$T$ のそれぞれのツイートに対して MeCab を用いて形態素解析をする．形態素解析をしたツイートの Bag of Words を計算し，それぞれのツイートを単語の特徴ベクトルに変換する．

特徴ベクトルに変換したツイートを，訓練データとしてモデルを作成する．このとき，SVC で用いるパラメータは，ラベル付けしたツイートからそれぞれ 2 件ずつを検証データとし，グリッドサーチで求めた．

モデルの精度を計測するために，Leave-one-out を用いた．検証データとして使用していないツイートのうち一つを除いて学習を行い，正しいラベルが付与されている割合を調べる．

### 4.2　実験結果

$T$ のうち合計 6 件を検証データとし，パラメータ決定のために用いた結果，最適なパラメータは $C$ が 100，gamma 値が 0.001 であった．また検証データに対する最も高いスコアは 0.50 であった．グリッドサーチを用いて算出した最適なパラメータを基に，Leave-one-out 用いて，表 2 のように判定割合を算出した．判定割合は (2) 式のように定義した．

$$a(l_c, l_p) = \frac{b(l_p)}{N(l_c)} \tag{2}$$

$a$ は判定割合，$l_p$ はモデルにより判定されたラベル，$l_c$ は正しいラベル，$b(l_p)$ はモデルによりラベル $l_p$ と判定されたツイート数，$N(l_c)$ は正解ラベルが $l_c$ であるツイート数である．つまり，信憑性が低いツイートが

表 2　実験結果

| | | ラベル | | |
|---|---|---|---|---|
| | | 低い | 高い | 関係無い |
| 推定結果 | 低い | 0.18 | 0.36 | 0.45 |
| | 高い | 0.09 | 0.64 | 0.27 |
| | 関係無い | 0.09 | 0.36 | 0.55 |

11 件あり，そのうち 2 件が信憑性が低いと判定されたとき，推定結果とラベルが共に「低い」の部分の値は $2/11 = 0.18$ となる．

### 4.3　考察

信憑性が高いツイートと関係無いツイートは，比較的良い判定結果が得られた一方で，信憑性が低いツイートは正しく判定されず，ほとんどが高い，もしくは関係無いと判定されてしまった．これは，データ数が少ないため，入力に用いたデータに結果が左右されてしまったことが原因と考えられる．また，$T$ を特徴ベクトルにした際，ツイート内の単語の出現回数のみに着目していた．しかし，信憑性が低いツイートは，通常のカテゴリ分類で見られるように，特定の単語が入っていたら信憑性が低いと必ずしも判定することはできない．実際に，信憑性が低いと主観で判定したツイート内の単語には，名詞などには共通点がほとんど無く，言い切りの形や単語間の関係に，信憑性が低いと判定するための特徴があった．そこで今後はデータ数を更に増やし，ツイートを単語による出現回数の特徴ベクトルで表すのではなく，分散表現で表す．また，人手により再度判定されたツイートをモデルの訓練データとして与えることで，分類精度を向上させることができるのではないかと考える．

## 5　おわりに

本稿では，「コロナ」を含むツイートの中から，信憑性が低いツイートを収集する方法を提案した．ラベル付けしたツイートを基に SVC によりモデルを作成し，ラベル付けしていないツイートを分類する機能を実装した．また実験では，作成したモデルを用いて，信憑性が低いツイートと高いツイートと関係無いツイートを正しく判定できるかという実験を行った．信憑性が高いツイートと関係無いツイートの判定は比較的良い結果を得ることができた．しかし，本研究で用いた分類方法では，信憑性が低いにも関わらず，信憑性が低いと判定されないツイートが多く，現在の方法では課題が残る．また，本研究の最終目標である，多数の人が納得できる情報源の作成を実現するためには，実際に多数の人にモデルにより判定されたツイートを，再度判定してもらわなければならない．そこで今後は，単語間の関係性に注目した分類方法の研究を進めつつ，多数の人の判断基準を導入してより良い情報源の作成をする．

### 謝辞

本研究の一部は JSPS 科研費 18H03342，19H04221，19H04218，および大川情報通信基金の助成を受けたものです．

### 参考文献

[1] 鈴木優. Wikipedia における情報の質. 情報処理学会論文誌, Vol. 6, No. 4, pp. 46–58, 2013.

[2] Stefano Mizzaro. Quality control in scholarly publishing: A new proposal. In *Journal of the American Society for Information Science and Technology*, pp. 989–1005, 2003.