

# モザイクアプローチにより情報流出の危険性があるツイートの判別手法の提案

澤田 晃志<sup>†</sup> 塚田 晃司<sup>‡</sup>

和歌山大学システム工学部<sup>‡</sup>

## 1. はじめに

SNS は現代人にとって人との繋がりにおける大事な役割を持つ。普及率も凄まじく利用率の最も高い SNS である「LINE」は 90.3%，本研究で利用した「Twitter」は 42.3%となっている[1]。

同時に SNS 上では様々なトラブルが発生している。中でも個人の特特定は私生活にも影響が大きく、被害者以外の人間にも被害が広がることも考えられる。また個人の特特定は被害者本人がトラブルを意識していても中々防ぐことが難しい。その理由の1つとして直接的な個人情報でなくとも断片的な情報から個人情報を知ることによって個人の特特定を行えることが挙げられる。このことを利用して個人の特特定を行う手法が「モザイクアプローチ」である。そこでモザイクアプローチに着目し、SNS 上に存在するモザイクアプローチに利用され得る投稿を判別する手法を提案する。

本研究では匿名性と利用率の双方が高い SNS である「Twitter」を選び、モザイクアプローチに利用され得るツイートを学習させることで自動的な判別を行うシステムの開発と評価を行う。

## 2. 関連研究

SNS 上での個人の特特定に関する研究の中でも Twitter に着目した研究として[2]が挙げられる。この研究ではユーザの属性と様々な情報の相関関係から投稿時間とユーザの私生活の充実感の関係について分析を行っている。また個人情報漏洩の防止方法も提案している。しかし提案まででありモザイクアプローチについては言及していない。

また個人の特特定と様々な要因の関係について分析を行った研究[3]がある。10 因子の中から自己顕示欲と意図しない個人の特特定の経験は関連性が高いこと、その中でも学生や女性がより高いことが結果として表れている。

さらに SNS 上での自己情報の公開する範囲と意図しない個人の特特定の経験には明確な関係が表れず、多くのユーザが自己情報の公開する範

囲を制限していることが明らかになった。

個人の特特定に関する様々な研究がなされている一方でモザイクアプローチについて明確に言及している研究は見られなかった。

## 3. 提案手法

本研究では個人の特特定の影響が大きい方が良く、また利用率の高さも必要であることから匿名性と利用率の両方が高い Twitter を利用する。筆者が選出したキーワードを含むツイートと偏りを減らすために無作為に選んだツイートをモザイクアプローチに利用され得るものとそうでないものに分別し、前処理を行った後、訓練データとして教師あり学習を行う。本研究では Fasttext を用いて機械学習を行う。学習後、任意のツイートを判別させることでシステムの評価を行う。評価は再現率、適合率、正答率、F 値で行う。

また本研究は機械学習によるモザイクアプローチの対策を最終的な目的としており、今回はその前段階としてテキストのみを対象としたモザイクアプローチにおける機械学習の精度の研究を行う。

### 3.1 訓練データの概要

表 1 に選出したキーワードを示す。これらのキーワードは個人情報に繋がる可能性が高く、頻繁に使われると筆者が判断した単語である。大きく分けると時間、場所、状況の3つになっており、合計 18 個のキーワードを選出した。

またツイートの検索時には RT、リプライ、リンク付きツイート、画像付きツイート、ニュース記事のツイート、公式アカウントによるツイートを除く。

収集するツイートの件数は 1 個のキーワードにつき 50 件、偏りを防ぐためのキーワード無しの無作為に選んだツイート 600 件の計 1500 件のツイートを収集する。また収集した時期は 2023 年 1 月 1 日、2023 年 1 月 5 日、2023 年 1 月 9 日の 3 回に分けた。分けた理由は元旦と成人式のようなイベントのある日とそうでない日の比較、イベント同士の比較を行うためである。

Proposal of a method to classify tweets carrying risks of personal data leakage using a mosaic approach

<sup>†</sup>Koshi Sawada, Koji Tsukada

<sup>‡</sup>Faculty of Systems Engineering, Wakayama University

表1 訓練データ収集時のキーワード

時間	場所	状況
今朝	近所	外出
今夜	近く	旅行
明日	付近	出かける
昨日	近辺	帰り
今日	最寄り	行きつけ
最近	駅前	
近頃		

### 3.2 前処理

収集したツイートの本文に MeCab を利用して形態素解析を行い品詞ごとに分解する。また辞書は最近の単語にも対応する必要があると考えて mecab-ipadic-Neologd を使用する。この辞書は毎週 2 回自動的な更新を行う Web 上を言語資源とした MeCab 用のシステム辞書である[4]。

また訓練データに使用するツイートには著者が手動でラベル付けを行う。

### 4. 評価実験

サイズの比較として本来の 5 分の 1 のサイズである 300 件のツイートを訓練データとした評価実験を行う。訓練データに含まれていない任意のツイート 20 件を正しく分類できているかを評価する。使用したツイートは全てイベントの無い日のツイートである。本来のサイズでの実験は追加実験として後日行う。

また Fasttext のパラメータの学習数を 1000, 単語ベクトルの次元数を 100 にしたモデル 1, 学習数を 3000, 単語ベクトルの次元数を 200 にしたモデル 2, 学習数を 3000, 単語ベクトルの次元数を 100 にしたモデル 3 の 3 つを作成し比較する実験も並行して行う。

#### 4.1 実験結果

実験結果をモザイクアプローチに利用され得るツイートとそうでないツイートに分けて表 2, 表 3 に示す。また各ツイートに対しての各モデルの精度の平均を表 4 に示す。

実験結果から学習数, 単語ベクトルの次元数を増やすことで精度の上昇が見られる。またモザイクアプローチに利用され得るツイートの分類が正しく行えていないことが分かる。

表2 各モデルによるモザイクアプローチに利用されないツイートの分類結果

	モデル 1	モデル 2	モデル 3
正解	9	9	9
不正解	1	1	1

表3 各モデルによるモザイクアプローチに利用され得るツイートの分類結果

	モデル 1	モデル 2	モデル 3
正解	2	2	2
不正解	8	8	8

表4 各モデルによるツイートの分類精度

	モデル 1	モデル 2	モデル 3
危険性あり	0.9887	0.9938	0.9938
危険性なし	0.9627	0.9664	0.9662

### 4.2 考察

Fasttext のパラメータについて学習数, 単語ベクトルの次元数は増やすことで精度の上昇がみられた。今後もモデル 2 のパラメータ値をもとに他のパラメータを加えた改良が良いと考えられる。またモザイクアプローチに利用され得るツイートの分類が正しく行えていない原因として訓練データの偏りが考えられる。収集したツイートの 2 割ほどしかモザイクアプローチに利用され得るツイートは含まれていないため、モザイクアプローチに利用され得るツイートの分類にのみ影響が出たと考えられる。

### 5. おわりに

本稿では、モザイクアプローチにおける機械学習の精度の向上を目的に実験と評価を行った。

今後の課題として、イベントの有無が与える影響の比較、今回の実験でみられた訓練データの偏りの解消、これらを本来のサイズのデータで評価実験を行うことが挙げられる。

### 参考文献

- [1]総務省：令和 2 年度情報通信メディアの利用時間と情報行動に関する調査,入手先 <[https://www.soumu.go.jp/main\\_content/000765258.pdf](https://www.soumu.go.jp/main_content/000765258.pdf)>(参照 2023-01-03).
- [2]松下翔太ほか：Twitter 投稿文脈による個人情報推定と個人情報漏洩の防止方法の提案と試作, 研究報告インターネットと運用技術 (IoT), Vol.2011-IoT-12, No.9, pp1-6 (2011).
- [3]田崎洋祐, 稲葉緑：複数の SNS アカウント利用時における意図しない個人特定に関する要因分析, 研究報告コンピュータセキュリティ (CSEC), Vol.2020-CSEC-88, No.35, pp.1-8(2020).
- [4]Toshinori Sato : mecab-ipadic-neologd/README. ja.md at master – GitHub, 入手先 <<https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>> (参照 2023-01-05).