

1. 背景
2. 提案手法
3. 実験結果
4. 今後の課題

機械学習を用いた X 上の炎上予測モデルの 提案

**Proposal for a Machine Learning-Based Flare-Up Prediction Model
on Platform X**

山本 藤也 (Touya Yamamoto)
u220067@st.pu-toyama.ac.jp

富山県立大学 情報システム工学科

November 26, 2024

1. 背景

2/12

SNS と炎上問題

- SNS の普及により、情報発信が容易になった一方で、炎上や偽情報拡散といった問題が増加。
- 特に政治的話題での炎上リスクが高く、迅速な予測が求められる。

目標

- 炎上ポストを事前に予測し、問題発生を防ぐ。
- 機械学習モデルを用いた二段階アプローチを提案。

2.1 提案手法の概要

3/12

二段階アプローチ

- 第一段階: 拡散予測モデル
→投稿が広範囲に拡散されるかどうかを評価
- 第二段階: 感情分析モデル
→拡散が予測されたポストが「positive」または「negative」の感情を持つか

2.2 特徴量について

4/12

特徴量

- 特徴量については、取得したデータをそのまま使用した特徴量が 5 種類、本文やリポスト数から作成した特徴量を 5 種類

表 2: アカウントとテキストデータの特徴量

特徴量	内容
author-created-at	アカウントが作成されてからの経過時間
created-at	投稿の投稿日時
followers-count	アカウントがフォローしているアカウントの数
following-count	アカウントをフォローしているアカウントの数
tweet-count	アカウントが投稿した数

表 3: 作成したデータの特徴量

特徴量	内容
text-length	投稿のテキストの長さ
day-of-week	投稿した日時の曜日
hour-of-day	投稿時間帯
hourlycount	時間差分 (1 時間)
tenminutecount	時間差分 (10 分間)
countsdata	10、30、60、120 分後のツイート数

2.3 拡散予測モデル

5/12

拡散予測モデル

- 特徴量: アカウント作成日、フォロワー数、投稿時刻など
- モデル: Transformer を用いた二値分類

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

表 1: 混同行列 (拡散予測モデル)

		正解ラベル	
		拡散しない	拡散する
二値分類モデルの判別結果	拡散しない	TP	FP
	拡散する	FN	TN

$$F1 = \frac{2(Precision \cdot Recall)}{Precision + Recall}$$

2.4 拡散予測モデルの結果

6/12

モデルの評価結果

- 正解率 (Accuracy) : 0.999
精度 (Precision) : 0.971
再現率 (Recall) : 0.825
F1 値 : 0.892
- 判定速度に関しては, 170,176 件のデータを 3.145 秒で処理

表 4: 二値分類モデルの精度評価値 (拡散予測モデル)

Acc	Pre	Rec	F1
0.999	0.971	0.825	0.892

2.5 感情予測モデル

7/12

拡散予測モデル

- 特徴量: 投稿のテキスト長、投稿曜日、時間帯など
- モデル: RoBERTa を使用した二値分類

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

表 5: 混同行列 (感情分析モデル)

		正解ラベル	
		positive	negative
二値分類モデルの判別結果	Positive	TP	FP
	Negative	FN	TN

$$F1 = \frac{2(Precision \cdot Recall)}{Precision + Recall}$$

2.6 感情予測モデルの結果

8/12

モデルの評価結果

- 正解率 (Accuracy) : 0.728
精度 (Precision) : 0.616
再現率 (Recall) : 0.664
F1 値 : 0.719

表 6: 二値分類モデルの精度評価値 (感情分析モデル)

Acc	Pre	Rec	F1
0.728	0.616	0.664	0.719

3.1 炎上予測実験

9/12

実験概要

- 拡散予測: 合計 1,066,039 件の投稿（拡散あり: 313,647 件、拡散なし: 752,392 件）
- 感情分析: 合計 710,067 件の投稿（Positive: 355,116 件、Negative: 354,951 件）
学習データとして 90%、テストデータとして 10%を使用。

表 7: データセット一覧

キーワード	収集期間	ポスト数	オリジナルポスト数
G7 汚染水	2023/01/05 ~ 2023/05/28	10,961	994
G7 BRICs	2023/01/01 ~ 2023/05/28	149,282	11,595
G7 LGBT	2023/01/01 ~ 2023/06/07	538,103	13,981
2022 年沖縄知事選挙	2022/08/31 ~ 2022/09/16	955,096	25,779
#ローソンのロールケーキ食べたい	2022/08/10 ~ 2022/08/25	342,115	871
au 障害	2022/06/30 ~ 2022/07/06	70,603	10,433
#台風 14 号	2022/09/12 ~ 2022/09/26	417,975	51,034

表 8: データセットの概要

キーワード	関連するイベント
G7 汚染水	G7 と 福島原発処理水をめぐる意見の盛り上がり
G7 BRICs	G7 と BRICs の GDP をめぐる長期的な意見の盛り上がり
G7 LGBT	G7 と LGBT に関する意見の盛り上がり
2022 年沖縄知事選挙	2022 年 9 月 11 日に投票が行われた沖縄知事選挙
#ローソンのロールケーキ食べたい	企業広告
au 障害	2022 年 7 月に発生した KDDI の通信障害
#台風 14 号	2022 年 9 月に小笠原近海で発生した台風

3.2 炎上予測実験

10/12

実験結果

- 政治的な話題: 炎上するポストの割合が高い (例: 「G7 LGBT」では 0.744%)
- 一般的な話題: 炎上の割合が低い (例: 「#ローソンのロールケーキ食べたい」では 0%)
→政治的な投稿は、感情的議論を引き起こしやすい傾向が示唆

表 9: 炎上予測実験結果

キーワード	拡散すると予測	拡散しないと予測	炎上すると予測	炎上しないと予測	炎上するポストの割合
G7 汚染水	1	993	1	0	0.101
G7 BRICs	17	11,578	17	0	0.147
G7 LGBT	104	13,877	104	0	0.744
2022 年沖縄知事選挙	120	25,659	61	59	0.237
#ローソンのロールケーキ食べたい	3	868	0	3	0.000
au 障害	6	10,429	0	6	0.000
#台風 14 号	27	51,007	7	20	0.014

結論

■ ボットの影響

政治的な話題では主に人間による投稿が中心。一般的な話題では、ボットが投稿の拡散に大きく寄与（例：「#ローソンのロールケーキ食べたい」で拡散率 100%がボットによるもの）

表 10: 拡散予測モデルにおけるボットによるツイートの割合

キーワード	拡散すると予測	ボットによるポスト数	ボット拡散率
G7 汚染水	1	0	0.000
G7 BRICs	17	1	5.882
G7 LGBT	104	16	15.385
2022 年沖縄知事選挙	120	17	14.167
#ローソンのロールケーキ食べたい	3	3	100.000
au 障害	6	1	16.667
#台風 14 号	27	5	18.519

表 11: 炎上予測モデルにおけるボットによるツイートの割合

キーワード	炎上すると予測	ボットによるポスト数	ボット炎上率
G7 汚染水	1	0	0.000
G7 BRICs	17	1	5.882
G7 LGBT	104	16	15.385
2022 年沖縄知事選挙	61	8	13.112
#ローソンのロールケーキ食べたい	0	0	0.000
au 障害	0	0	0.000
#台風 14 号	7	2	28.571

4.1 今後の課題

12/12

改善点

- モデル精度向上（特に感情分析モデル）。
- 時系列データの活用。
- データセットの拡充。