

公衆衛生分野におけるオープンソース・インテリジェンスの有効性の検証

安納住子*、杉田暁**
 *上智大学、**中部大学

1. はじめに

日本の気温上昇率は世界平均気温の上昇率（気象庁の解析では 100 年当たり 0.74℃）よりも大きく(1)、また、大都市では都市化による気温の長期的な上昇傾向がみられる(2)。それゆえ、将来の日本においては、温暖化や気温上昇による熱ストレスが増加し、特に高齢者の熱中症リスクが増加することが予想されている。

日本の省庁による熱中症対策においては、救急搬送サーベイランス（搬送者情報：性別、年齢区分、傷病程度、発生場所）および公表、予防に係る普及啓発等が行われている。しかしながら、予防のためには、熱中症の自覚症状をもつ人を早期に探知し、迅速な公衆衛生的対応を可能にするイベントベースサーベイランスが望ましいとされている(3)。近年、ソーシャル・ネットワーキング・サービス（Social Networking Service：SNS）に投稿される発言は、イベントベースサーベイランスの役割を果たすことが期待されている。

本研究は、熱中症に関連する Twitter の投稿文、救急搬送サーベイランスデータをもとに深層学習を用いて、ツイートの事実性判別、熱中症発生リスクの予測、公衆衛生学分野におけるオープンソース・インテリジェンス（Open Source Intelligence：OSINT）の有効性を検証することを目的とする。

2. 方法

対象地域

熱中症による救急搬送人員数が、東京都、埼玉県、大阪府に次いで第 4 位(2022 年調べ) (4)の愛知県名古屋市を対象地域とした。

フレームワークの概要

本研究の第一の目的は、熱中症発生の早期検知を実現するイベントベースのサーベイランスを構築することである。本研究では、早期検知するための 2 段階のフレームワークを提案する（図 1）。

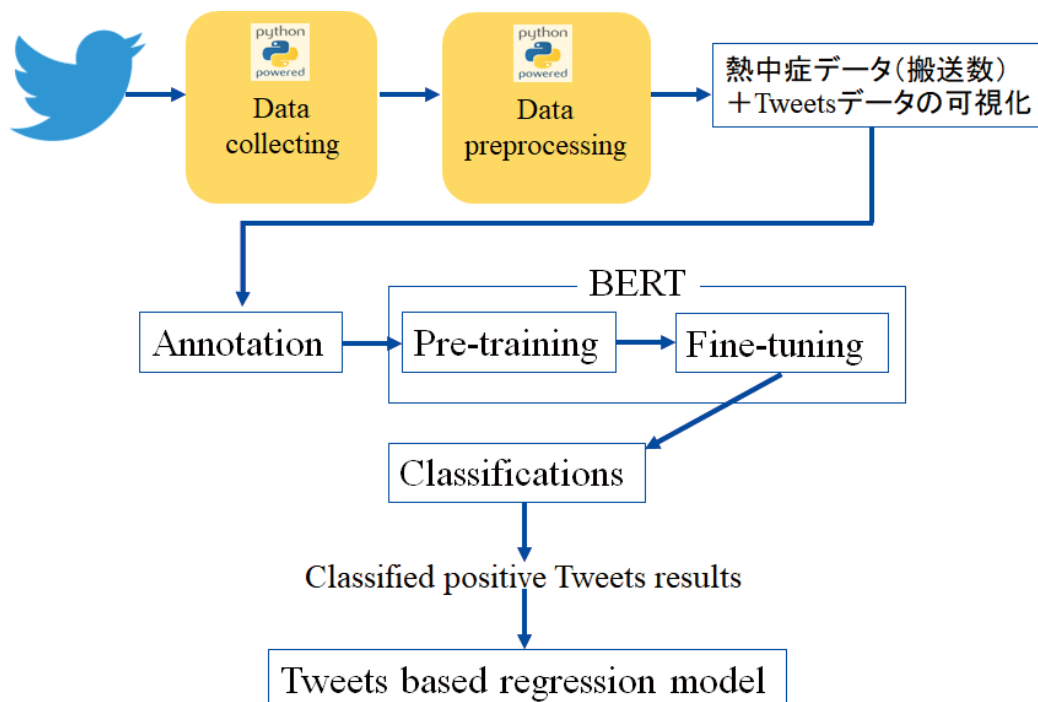


図 1. 研究フレームワークの概要

[1] フレームワークの最初のフェーズでは、先ずツイートを収集し、そのデータを前処理する。次に、自然言語処理モデル：BERT (Bidirectional Encoder Representations from Transformers) を適用して、ツイートの事実性判別を行う。

[2] 第2フェーズでは、分類結果に対して、提案モデル：自己回帰モデル (Transformer)、ベクトル自己回帰モデル (Transformer) を用いて、時系列予測を行う。さらに、Transformer の有効性を検証するため、LSTM (Long Short-Term Memory) をベースラインとして予測精度の比較を行う。

データ収集

本研究に使用する2種類のデータを次の通りに収集した。

2017/4/22 から 2022/9/30 までに収集された町区レベルの熱中症による救急搬送人員に関するデータを名古屋市消防局から取得した(5)。

Twitter API v2、BEARER_TOKEN を用い、上記の期間に合わせて2017/4/22 から 2022/9/30 までに投稿された熱中症に関連する投稿文を収集した。

データ前処理

収集した投稿文のデータに対して、クリーニングと正規化の前処理を行うことにより、記号などのノイズを除去した。

モデル

前処理したツイートの事実性判別を行うため、BERT モデルを用いる。また、愛知県名古屋市の熱中症発生分布を時系列で予測するため、Transformer(6)を用いる。Transformer モデルは、自然言語処理にも適用されているが、インフルエンザの時系列予測にも応用され、高精度での予測が確認されている(7)。

3. 結果

日毎に数え上げた熱中症データ（搬送数）（図2上）と、「暑い or あつい」のキーワードを含む投稿文数（図2下）を時系列に可視化した結果を図2に示す。図から、日毎の熱中症搬送数と「暑い or あつい」のキーワードを含む投稿文数との間に関連がみられた。

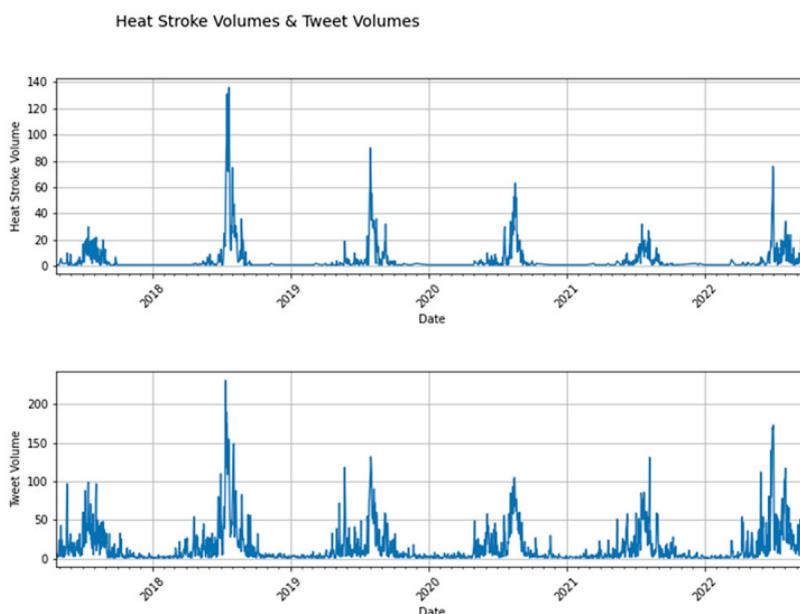


図2. 熱中症搬送数/日とツイート数/日

投稿文のいくつかを以下の図3に示す。コラムHには前処理前の投稿文、コラムIには前処理後の投稿文が格納されている。#以降の文全てがクリーニングされていた。また、記号が完全に除去できておらず、さらに、「あついー」が「ついー」に処理されていることが判明した。

	E	F	G	H	I	J	K	L	M
123	Nagoya City Naka W	136.9076886	35.1608036	おはようございます。#名古屋市中区 体感的にはカラッとするような?昨日までのジメジメが少ないと思う。けど、、暑いですからね。水分補給して夏バテしないように今日もステキな一日を#イマソラ#企業公式相互フォロー https://t.co/6dLN3Dmi1m	おはよう				
124	Nagoya City Moriyar	136.9950872	35.2172183	@miyorin0208 おは👼天使 可愛い キュン👼真					

	E	F	G	H	I	J	K	L
938	Nagoya-shi Midori-k	136.9707947	35.0851921	あ! (´д`)!!(´_`)/_/`\'/ /`\'` /`\'`'				
939	Nagoya City Atsuta	136.8979869	35.129583	本日よりお盆マンス	本日 盆 マンスリー			

	E	F	G	H	I
3701	Nagoya-shi Minato-	136.8543525	35.0878966	あついー	ついー
3702	Nagoya-shi Nakagav	136.875559	35.1539414	@yunn_mm 今日って	今日 暑い

図3. 前処理前後の投稿文

4. 考察

日毎の熱中症搬送数と「暑い or あつい」のキーワードを含む投稿文数との間に関連がみられたことから、ツイートの有効性が示唆された。

先行研究では#から始まるハッシュタグがクリーニングされていたことから、本研究でも実施したが、#以降の文全てがクリーニングされていた。ハッシュタグを含め必要な文も含まれることから、ハッシュタグの必要性を再考するとともに、プログラムの修正が課題である。また、クリーニングしきれていない記号を含め、誤ったクリーニングについても、追加調査の上、プログラムの修正が必要である。

5. まとめ

今後は、修正プログラムにより前処理したデータに対してアノテーション作業を行い、アノテーションされた正解ラベル付きのデータセットとBERTモデルを用いて、ツイートの事実性を分類するモデルを作成する予定である。また、分類結果と熱中症データ（搬送数）を用いて熱中症発生リスクを時系列で予測するモデルの作成、さらに、熱中症早期警戒システムの構築を目指す。

6. 謝辞

本研究の一部は、中部大学問題複合体を対象とするデジタルアース共同利用・共同研究IDEAS202206の助成を受けたものです。また、名古屋市消防局様からは、熱中症救急搬送に関するデータを提供して頂き感謝申し上げます。

参考文献・データ

- (1) 気象庁, 世界の年平均気温偏差の経年変化 (1891~2022 年),
https://www.data.jma.go.jp/cpdinfo/temp/an_wld.html
- (2) Nishat Tasnim Toosty, Aya Hagishima, Ken-Ichi Tanaka, (2021) Heat health risk assessment analysing heatstroke patients in Fukuoka City, Japan. PLoS ONE 16(6): e0253011.
<https://doi.org/10.1371/journal.pone.0253011>
- (3) WHO, A guide to establishing event-based surveillance, isbn 978 9 9061 3 1 3.
- (4) 総務省消防庁, 熱中症による救急搬送状況 (令和 4 年) 「都道府県別救急搬送人員 (昨年比)」 (グラフ),
<https://www.fdma.go.jp/disaster/heatstroke/post3.html>
- (5) 名古屋市消防局.
- (6) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, (2017) Attention is all you need. arXiv:1706.03762v5 [cs.CL]
- (7) Neo Wu, Bradley Green, Xue Ben, Shawn O'Banion, (2020) Deep transformer models for time series forecasting: the influenza prevalence case. CoRR arXiv:2001.08317.