

# 公衆衛生サーベイランスのための SNS 上で投稿される 多様な病気症状の事実性解析

安藤 樹<sup>†</sup> 安藤 一秋<sup>‡</sup>香川大学大学院 創発科学研究科<sup>†</sup> 香川大学 創造工学部<sup>‡</sup>

## 1. はじめに

近年、自然言語処理を医療分野に応用する研究が注目を集めている[1-5]。これらの研究の多くは、ユーザが現在の状況をリアルタイムで共有する Social Networking Service (SNS) を対象としている。例えば、SNS を対象とした研究として、自身の投稿内容から「うつ病」を診断する研究[1, 2]や、インフルエンザに関する投稿を収集してインフルエンザの流行を推定する研究[3, 4]などがある。これらの研究は、特定の病気のみを対象としたものであるが、SNS 上には様々な病気や多様な症状が投稿されているため、これらの情報を活用することで、新たな知見が得られる可能性がある。

米国の疾病管理予防センター (CDC) は、公衆衛生サーベイランスを「公衆衛生の実践の計画、実施および評価に不可欠な健康に関連するデータの継続的で体系的な収集、分析、解釈」と定義している。したがって、公衆衛生サーベイランスの目的の 1 つは、未知・既知の病気の発生を検知し、予防策を提案することである。そこで我々は、病気の流行を早期に検知するため、様々な病気症状に注目する。そして、SNS 上における症状の発症数の日々の変化から病気の流行を検知することを目指す。

本研究では、公衆衛生サーベイランスの一つとして、COVID-19やインフルエンザなどの感染症の流行状況や頭痛・腹痛といった各種症状の発症状況を観測・分析し、都道府県別・時系列別に可視化するシステム[6, 7]の実現を目的とする。提案システムにより、各種病気症状の流行・発症状況の迅速な把握、病気症状の特徴分析、未知の病気に対する流行の早期検知などが可能となる。提案システムは、症状を含む投稿が投稿者自身の症状の発症を意図したものであるかどうかを判定する事実性解析モジュール[6]と、投稿者の居住地を推定する位置推定モジュール[7]で構成される。

症状を発症した人は、様々な表現を用いて投稿する。そこで本研究では、患者が普段使用する症状表現と標準的な病名を対応付けた患者表現辞書[8]を利用する。辞書内の 11 種の標準病名に対する 86 種の症状表現を対象に、これらの症状表現を含む SNS 上の投稿に対し、事実性解析する手法を提案する。本稿では、RoBERTa モデル、JMedRoBERTa モデル、PubMedBERT モデルを用いた事実性解析モデルを構築し、これらのモデルの判定性能を評価する。

## 2. データセットの構築

事実性解析のデータセットは、様々な病気症状に関する多様な表現を含む投稿を用いて構築する必要がある。そこで、患者表現辞書の標準病名のうち、SNS 上でよく使われる 11 種の標準病名を選定し、これらを抽出対象とする。表 1 に選定した標準病名を示す。

患者表現辞書の各標準病名には、患者が実際に使用する様々な症状表現が収録されている。例えば、「頭痛」の場合、63 種の症状表現が収録されている。しかし、あまり利用されない表現も含まれているため、検索エンジンのヒット数に基づいて対象とする症状表現を絞り込むことにした。Google 検索 API へのリクエスト数の関係上、辞書に登録されている表現数が 10 を超える場合は、ヒット数の大きい 10 表現に絞り込み、表 1 の各標準病名について、合計 86 種の症状表現を選択した。そして、86 表現を利用して、Twitter (X) から 11 種の標準病名 (以降、病気症状) に対して、各 5,000 件のツイートを収集した。

症状表現を含む投稿が投稿者自身の症状の発症を意図している場合は正例、そうでない場合は負例としてそれぞれ 1,000 件からなるデータセットを構築した。

表 1 対象とした病気症状 (標準病名)

頭痛	腹痛	胸痛	眼痛	耳痛	関節痛
咽頭痛	発熱	めまい	動悸	嘔吐感	

## 3. 事実性解析モデル

本稿で性能を評価する 3 つの事実性解析モデルについて述べる。全てのモデルは、最終層に 1 次元畳み込み層と線形層を組み込んだモデル構成をとり、2 値分類タスクとしてファインチューニングする。

### 3.1. RoBERTa モデル

RoBERTa モデルは、早稲田大学が訓練した roberta-large-japanese<sup>1)</sup> モデル (以降、RoBERTa モデル) を用いて構築する。この事前学習済みモデルの構築には、日本語 Wikipedia にある約 1,800 万文を用いている。

### 3.2. JMedRoBERTa モデル

医療ドメインに特化した事前学習済み言語モデルの 1 つとして、杉本ら[9]は、約 1,100 万文の日本語医学論文データを用いた JMedRoBERTa を公開している。これは、RoBERTa のフルスクラッチ学習により構築されたモデルである。本稿での JMedRoBERTa モデルは、これを用いて構築する。

### 3.3. PubMedBERT モデル

別の医療ドメインに特化した言語モデルとして、PubMed の抄録と PubMed Central の記事全文を用いて事前学習させた PubMedBERT[10] がある。本稿での PubMedBERT モデルは、これを用いて構築する。

Factuality Analysis of Disease Symptoms Posted on SNS for Public Health Surveillance

<sup>†</sup>Tatsuki Ando · Graduate School of Science for Creative Emergence, Kagawa University

<sup>‡</sup>Kazuaki Ando · Faculty of Engineering and Design, Kagawa University

<sup>1)</sup><https://huggingface.co/nlp-waseda/roberta-large-japanese>

#### 4. 実験設定

本研究では、予備実験により、RoBERTa モデルと JMedRoBERTa モデルは、バッチサイズ 2,000 で 12 万 5000 ステップ学習し、PubMedBERT モデルはバッチサイズ 256 で 100 万ステップ学習したモデルを利用する。また、Optimizer には AdamW を、損失関数にはバイナリクロスエントロピーを採用する。

#### 5. 実験結果

各病気症状に対して、それぞれのモデルでの実験結果を表 2 に示す。11 種の病気症状のうち、10 種の病気症状で JMedRoBERTa モデルが最も良い F1 値を得た。また、全症状におけるモデル別の F1 値の平均は、RoBERTa モデルでは 0.710、JMedRoBERTa モデルでは 0.750、PubMedBERT モデルでは 0.720 であり、JMedRoBERTa モデルが最も高い分類性能となった。JMedRoBERTa モデルは、日本語の医療論文を事前学習に使用しているため、最も高い分類性能が得られたと考えられる。

また、PubMedBERT モデルにおいて、F1 値が 0.7 を下回る「胸痛」、「耳痛」、「動悸」、「嘔吐感」に対して、JMedRoBERTa モデルでは F1 値が大幅に向上していることを確認した。エラー分析の結果、PubMedBERT モデルでは、日本語特有の負例症状を誤判定する傾向が確認できたが、日本語に特化した JMedRoBERTa ではこの点が改善できていた。

「関節痛」については、JMedRoBERTa モデルよりも PubMedBERT モデルの方が 1.4 ポイント高い F1 値となった。この理由は、ファインチューニングに用いた学習データにおいて、負例の多様性が不足していることが影響していると考えられる。今後、学習データを増やすことで、JMedRoBERTa モデルのさらなる性能向上が見込める。

#### 6. おわりに

本稿では、ツイートされる病気・症状を地域別・時系列別に可視化するシステムを実現するため、患者表現辞書に含まれる 11 種の病気症状を含むツイートを対象に、RoBERTa、JMedRoBERTa、PubMedBERT を用いた 3 つの事実性解析モデルを構築し、これらのモデルの性能を評価した。全症状におけるモデル別の F1 値の平均は、RoBERTa モデルでは 0.710、JMedRoBERTa モデルでは 0.750、PubMedBERT モデルでは 0.720 となり、JMedRoBERTa モデルが最も高い分類性能を得た。今後は、詳細なエラー分析を実施すると共に、データセットにおける負例のバリエーションを増やして再評価する。

#### 参考文献

- [1] 玉井他, “うつ傾向推定に向けたツイート内容の解析法についての一検討”, 言語処理学会第 22 回年次大会発表論文集, pp.385-388, 2016.
- [2] A. U. Hassan, et al., “Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression”, Proc. of 2017 Int. Conf. on Information and Communication Technology Convergence (ICTC), 3 pages, Oct. 2017.
- [3] 北川他, “インフルエンザ流行検出のための事実性解析”, 言語処理学会第 21 回年次大会発表論文集, pp.218-221, 2015.
- [4] S. Wakamiya, et al., “After the Boom No One Tweets, Microblog-based Influenza Detection Incorporating Indirect Information,” Proc. of Int. Conf. on Emerging Databases, 2016.

表 2 判定結果

病気症状	モデル	precision	recall	F1 score
頭痛	RoBERTa	0.787	0.732	0.760
	JMedRoBERTa	<b>0.832</b>	<b>0.756</b>	<b>0.794</b>
	PubMedBERT	0.805	0.755	0.780
腹痛	RoBERTa	0.745	0.721	0.733
	JMedRoBERTa	<b>0.793</b>	<b>0.746</b>	<b>0.770</b>
	PubMedBERT	0.787	0.712	0.750
胸痛	RoBERTa	0.687	0.645	0.666
	JMedRoBERTa	<b>0.732</b>	<b>0.721</b>	<b>0.727</b>
	PubMedBERT	0.704	0.686	0.695
眼痛	RoBERTa	0.703	0.698	0.701
	JMedRoBERTa	<b>0.761</b>	<b>0.732</b>	<b>0.747</b>
	PubMedBERT	0.768	0.696	0.732
耳痛	RoBERTa	0.691	0.702	0.697
	JMedRoBERTa	<b>0.721</b>	<b>0.681</b>	<b>0.701</b>
	PubMedBERT	0.732	0.659	0.696
関節痛	RoBERTa	0.701	0.685	0.693
	JMedRoBERTa	0.757	0.732	0.745
	PubMedBERT	<b>0.774</b>	<b>0.743</b>	<b>0.759</b>
咽頭痛	RoBERTa	0.723	0.703	0.713
	JMedRoBERTa	<b>0.768</b>	<b>0.742</b>	<b>0.755</b>
	PubMedBERT	0.752	0.683	0.718
発熱	RoBERTa	0.731	0.721	0.726
	JMedRoBERTa	<b>0.754</b>	<b>0.742</b>	<b>0.748</b>
	PubMedBERT	0.723	0.692	0.708
めまい	RoBERTa	0.783	0.742	0.763
	JMedRoBERTa	<b>0.851</b>	<b>0.763</b>	<b>0.807</b>
	PubMedBERT	0.846	0.752	0.799
動悸	RoBERTa	0.683	0.671	0.677
	JMedRoBERTa	<b>0.753</b>	<b>0.717</b>	<b>0.735</b>
	PubMedBERT	0.658	0.623	0.641
嘔吐感	RoBERTa	0.683	0.672	0.678
	JMedRoBERTa	<b>0.731</b>	<b>0.712</b>	<b>0.722</b>
	PubMedBERT	0.661	0.638	0.650

- [5] 松田他, “Twitter を用いた病気の事実性解析及び知識ベース構築”, 人工知能学会第 30 回全国大会論文集, pp.2C5-OS-21b-4. 2016.
- [6] 安藤他, “ツイートされる病気症状の可視化に向けた病気症状の事実性解析のための素性検討”, 情報科学技術フォーラム講演論文集, pp.139-140, 2019.
- [7] M. Matsumoto, et al., “A Deep Learning Model of Estimating User’s Place of Residence Using Tweets and Weather Information”, Proc. of 2022 IEEE Asia-Pacific Conf. on Computer Science and Data Engineering, 6 pages, 2022.
- [8] 患者表現辞書, <https://sociocom.naist.jp/patient-dic/>
- [9] 杉本他, “JMedRoBERTa: 日本語の医学論文にもとづいた事前学習済み言語モデルの構築と評価”, 言語処理学会第 29 回年次大会発表論文集, pp.707-712, 2023.
- [10] Y.Gu, et al., “Domain-specific language model pretraining for biomedical natural language processing,” ACM Transactions on Computing for Healthcare, Vol.3, No.1, pp.1-23, 2021.