

# 病気症状の事実性解析に向けたツイートされる病気症状の多様性の調査 Investigation of the Diversity of Tweeted Disease Symptoms for Factual Analysis of Disease Symptoms

安藤 樹<sup>†</sup>      安藤 一秋<sup>‡</sup>  
Tatsuki Ando    Kazuaki Ando

## 1. はじめに

近年、医療分野に自然言語処理を応用する研究が注目されている。たとえば、SNS を対象とした研究として、自身のツイート内容をもとに「うつ病」を診断する研究[1]や、インフルエンザに関するツイートを収集してインフルエンザの流行度合を推定する研究[2]、感染症のみを対象として病気の事実性を解析する研究[3]など、対象を「特定の病気」に限定した研究が中心である。しかし、SNS 上では、様々な病気や症状に関する内容が発信されており、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。そこで、当研究室では、特定の病気や感染症であるか否かを問わず、いつ、どこで、どのような病気症状がツイートされているのかを収集・分析し、地域別・時系列別に可視化するシステム[4]の構築を進めている。このシステムにより、特定のエリアで特定の症状に関するツイートが急激に増加している場合、そのエリアで何か問題が発生していることが検知可能である。新型コロナウイルスへの感染についても、様々な症状が発症することが徐々に明らかになった経緯もあり、病気だけでなく、病気症状にも注目することは意義があると考えられる。

当研究室の先行研究[4]では、代表的な 14 種の病気症状の一般的な表現を含むツイートを事実性解析の対象に設定していたため、対象外となる病気症状や症状表現が多数存在していた。そこで本稿では、病気症状の様々な患者表現を収集・分析している研究[5]における、一般的な表現から専門的な表現、また、表記ゆれといった多様な表現に対応できる事実性解析の実現に向け、実際にツイートされる病気・症状表現の多様性や特徴などについて調査する。

## 2. ツイートの収集・調査方法

本研究では、MEDNLP が公開している患者表現辞書[6]を利用して、ツイートを収集して調査に利用する。

### 2.1 対象とする病気症状

本稿では、患者表現辞書にある頭痛や腹痛といった、痛みに関する病気症状に注目する。患者表現辞書の痛みに関する 15 種類の症状例を表 1 に示す。これらの病気症状の内、頭痛、腹痛、胸痛、眼痛、耳痛、関節痛、咽頭痛に関するツイートを本稿での調査対象に設定し、収集する。また、患者表現辞書内の病気症状のうち、表現数の多い発熱についても調査対象に加える。なお、足部痛、手痛、手指痛、顔面痛、肩部筋痛、腰痛及び全身痛については、病気ではなく、怪我に起因する事例が多いことから、本研究での調査対象から除外する。また、患者表現辞書の圧痛の表

<sup>†</sup> 香川大学 大学院 創発科学研究科, Graduate School of Science for Creative Emergence, Kagawa University

<sup>‡</sup> 香川大学 創造工学部, Faculty of Engineering and Design, Kagawa University

表 1 患者表現辞書の痛みに関する病気症状

頭痛	腹痛	胸痛	眼痛	耳痛
関節痛	咽頭痛	足部痛	手痛	手指痛
顔面痛	肩部筋痛	腰痛	全身痛	圧痛

表 2 調査対象とする各症状の検索キーワード

病気症状	検索キーワード		
頭痛	・頭が痛い	・あたまがいたい	・あたまが痛い
	・頭がいたい	・頭いたい	・頭痛
	・頭イタ	・あだまいでえ	・頭がガンガン
	・頭がズズズキ		
腹痛	・腹が痛い	・腹がいたい	・おなか痛い
	・腹いた	・腹痛	・腹部が痛い
	・下痢	・脇腹が痛い	・わき腹が痛い
	・おなかいたい		
胸痛	・胸が痛い	・胸がいたい	・むねがいたい
	・胸痛	・胸がチクチク	・胸元が痛い
	・胸が苦しい	・むねいた	・心臓が痛い
	・わき腹が痛い		
眼痛	・目が痛い	・めが痛い	・眼が痛い
	・目がいたい	・眼がいたい	
耳痛	・耳が痛い	・耳がいたい	・耳痛
関節痛	・関節が痛い	・関節がいたい	・関節痛
	・節が痛い	・ふしが痛い	
咽頭痛	・喉が痛い	・のどが痛い	・喉元が痛い
	・喉痛い	・のど痛い	
発熱	・発熱	・熱がある	・熱が出
	・熱出た	・微熱	・体が熱い
	・からだ熱い	・熱っぽい	・ねつがある
	・ねつ出た		

現例として、「圧迫される」や「軽い痛み」などがあるが、これらは痛みの対象となる部位が特定できないため、調査対象から除外する。

### 2.2 検索キーワードの設定

患者表現辞書には、病気症状（標準病名）に対して、様々な出現系が登録されている。たとえば、頭痛の場合、63 種の出現系が登録されている。調査における時間的コストの面から、本稿での調査では、出現系が 10 通りを超える場合、Google 検索のヒット数をもとに対象とする出現系を絞りこみ、各症状において最大 10 種の症状表現を検索キーワードに利用する。本稿での調査に利用する症状ごとの検索キーワードを表 2 に示す。

## 2.3 調査方法

本稿では、一般的な表現から専門的な表現、また、表記ゆれといった多様な表現に対応できる事実性解析の実現に向け、実際にツイートされる病気・症状表現の多様性や特徴などについて調査する。

設定した検索キーワードでツイートを検索し、症状別のツイート数を調査すると共に、患者表現辞書の表現数との関係を考察する。また、各症状を含むツイートを、明らかに病気症状によらないもの（以降、ノイズツイートと呼ぶ）とそれ以外（病気に関係するもの、関係する可能性があるものなど）に人手で分類し、症状別の特徴を確認する。ここで、ノイズツイートとは、怪我による痛みや心情的な痛みに関するツイート、病気症状に関する情報や意見のみが書かれたツイートのことである。

## 3. ツイートの収集結果と調査結果

### 3.1 ツイートの収集結果と考察

表 3 には、6 月 6 日~6 月 12 日までの 1 週間にツイートされた病気症状ごとの収集ツイート数を示す。表 3 に示す結果より、頭痛に関するツイートが最も多く、また、関節痛に関するツイートが最も少ないことがわかる。

次に、ツイート数の多い病気症状と少ない病気症状を患者表現辞書の表現数の視点で確認すると、患者表現辞書にある表現数が多い症状ほど、ツイート数も多くなる傾向が確認できた。

表 3 各病気症状を含むツイートの収集結果

病気症状	患者表現辞書にある表現数	1週間でのツイート数	1日当たりの平均ツイート数
頭痛	42	193,513	27,644
腹痛	54	104,736	14,962
胸痛	69	36,419	5,202
眼痛	6	24,030	3,432
耳痛	5	6,659	951
関節痛	4	5,345	763
咽頭痛	4	7,418	1,059
発熱	62	52,421	7,488

### 3.2 ツイートの調査結果と考察

収集したツイートに対して、ノイズツイートを人手で判定した結果を表 4 に示す。表中のノイズツイート数および割合は、病気症状ごとに 1 日当たり 100 件、1 週間で 700 件のツイートに対するノイズツイート数とその割合を示す。

表 4 から、胸痛と耳痛において、ノイズツイートの占める割合が高いことがわかる。また、発熱、咽頭痛、腹痛についてはノイズツイートの占める割合が低いことがわかる。

胸痛と耳痛のノイズツイートの例としては、「昨晚見た夢がすごく胸が苦しい夢だった」や「耳が痛いお言葉ですね」といった心情を表現するための比喩として症状が使われている場合が多い傾向を確認した。また、眼痛や関節痛のノイズツイートの例としては、「コンタクトが乾いて目が痛い」や「昨日、転んだせいであちこちの足の関節が痛い」といった怪我による痛みに関するツイートが多い傾向を確認した。頭痛のノイズツイートの例としては、「今朝、天井にぶ

つけて頭痛い」や「今日予定詰まり過ぎて、考えただけで頭痛くなる」といった怪我や心情に関するツイートが多い傾向を確認した。

一方、ノイズツイートの占める割合が少ない腹痛や咽頭痛、発熱のノイズツイートの例としては、「コロナワクチンの影響として腹痛や吐き気・喉の痛み・発熱などが多いみたい」や「コロナになるとやっぱり喉痛いのかな? なかったことないから分かん」といった新型コロナウイルスの症状やワクチンの副反応に関する情報のツイートが多い。

以上のことから、病気症状によってノイズツイートに特徴があることを確認した。また、ノイズツイートの占める割合が高いものは事実性解析の難易度も高い傾向にあると考えられるため、今後は、さらなる調査を継続する。

表 4 ノイズツイートの判定結果

病気症状	ノイズツイート数[件]	ノイズツイートの割合[%]
頭痛	169	23.4
腹痛	108	15.4
胸痛	329	47.0
眼痛	262	37.4
耳痛	311	44.4
関節痛	267	38.1
咽頭痛	102	14.6
発熱	95	13.5

## 4. おわりに

本稿では、多様な病気症状の事実性解析の実現に向けて、ツイートされる病気症状の多様性や特徴などについて調査した。調査の結果、病気症状ごとのツイート数は、頭痛に関するツイートが最も多く、患者表現辞書にある表現数が多い病気症状ほどツイート数も多くなる傾向を確認した。また、胸痛と耳痛についてはノイズツイートの占める割合が高く、発熱や咽頭痛、腹痛についてはノイズツイートの占める割合が低いこと、病気症状ごとに、怪我や心情といったノイズツイートの特徴が異なることを確認できた。

今後は、ツイートされる病気症状の多様性に対応した事実性解析の実現に向けて、病気症状ごとのツイート数やノイズツイートの特徴の調査を継続し、解析モデルに利用する素性について検討する。

### 参考文献

- [1] 玉井森彦, 米澤(深谷) 拓吾, 川西直, 長谷川晃朗, 武内良男, 田近亜蘭, 小川雄右, 古川壽亮, “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] 北川善彬, 小町守, 荒牧英治, 岡崎直観, 石川博, “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.
- [3] 松田紘伸, 吉田稔, 松本和幸, 北研二, “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.
- [4] 安藤翼, 安藤一秋, “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.
- [5] 西谷実紘, 矢田竣太郎, 若宮翔子, 荒牧英治, “生成アプローチによる患者表現の標準化”, JAMI & JSAI AIM 合同研究会資料, pp.5-01-5-07, 2021.
- [6] 患者表現辞書, <https://sociocom.naist.jp/patient-dic/>