

人とのつながりに着目した気分障害の背景の都市別分析

米丸 周吾

人と人とのつながりが弱い地域では、コロナ禍により、社会的な孤立化が一層深刻化しつつある。本研究では、精神の健康を保つ上で重要な市民とのつながりを、定量化に用いるためのリソースとして十分な検討が行われていないツイートに基づき定量化することで、都市別につながりの強さを定量化することを目指す。また、市民同士のつながりと、疾病の国際統計分類である ICD-10 において定義されているうつ病や双極性障害などの病気を対象とする「気分障害」との関係性を、地域に着目し分析することで、気分障害に影響を与える人同士のつながりについて明らかにすることを目的とする。

市民とのつながりを定量化するために、本研究ではソーシャルキャピタルと呼ばれる概念に基づき、親族、友人、職場関係者とのつながりとイベントなどへの参加を表す指標を定義する。次に、定義した各指標に関連するツイートを都市単位で抽出し、つながりに対するユーザの肯定的あるいは否定的な評価を表す「つながりに対する評価」や、実世界かネットのつながりかを表す「つながりの形態」といったラベルを深層学習モデルを用いて付与する。そして、付与したラベル数をもとに地域ごとの指標の評価値を計算する。

本研究では、都市別のデータセットに加え、都市や指標を横断したデータセットを用いて、大規模事前学習言語モデルをファインチューニングし、各ラベルの分類モデルを構築することで、各ラベルの分類に対する有効性を検証する。大規模事前学習言語モデルとしては、RoBERTa と BERT を採用し、その分類性能を比較する。

各ラベルの分類のため、気分障害の患者数増加率が異なる都市を対象に各指標についてのデータセットを、複数の判定者により人手で構築した。構築したデータセットは、全ての都市と指標において Fleiss の κ 係数を用いた判定者間一致度が 0.6 を超えたことから、判定者に依存せず、妥当な基準で構築されたことが示された。また、都市を横断して訓練することに加え、RoBERTa を用いることで、5 分割交差検証を用いた分類精度（マクロ平均 F 値）が多くの上ラベルにおいて BERT と比較して有意に向上することが示された（有意水準 5%、両側検定）。特に、人同士のつながりであるかどうかを判断する「つながりの情報の有無」においては、いずれの都市・指標においても 0.8 を越える精度で分類できることを明らかにした。

ラベル数と気分障害の患者数との時系列相関分析では、2020 年におけるオンライン上のイベントの参加に関するツイート数と気分障害の患者数との間で負の相関が確認でき、ツイートの内容から、Zoom を用いたオンライン飲み会や交流会などのイベントへの参加が、市民の精神に好影響を与えた可能性があることを示した。また、仕事が再開し始める 2020 年の 7 月から 12 月の期間において、職場関係者とのつながりに対する否定的な評価に関するツイート数と患者数との間で正の相関がみられ、職場での新型コロナウイルス感染症への対応に対する不満が、気分障害に影響を与えた可能性があることを示した。

（指導教員 関 洋平）