COVID-19 ワクチン接種に対する市民意見の立場分類のためのコーパスの構築

榊原　智仁

COVID-19 禍において，個人が意見や感想を自由に発信できる SNS 上では特に COVID-19 ワクチンの是非について活発に議論されてきた．しかし，SNS 上ではユーザの興味のある情報を優先的に推薦する構造や，自分と似た意見を持つユーザに囲まれやすいといった特徴があり，ユーザの持つ考えが先鋭化しやすい．このような要素から SNS では意見分極が生じ，各ユーザの持つ COVID-19 ワクチンへの立場に対して影響を与えていると推察される．そこで本研究は，SNS 上の COVID-19 ワクチン接種に対する市民意見の形成を明らかにすることを目的とした立場分類コーパスを構築し，それを用いた分類モデルを作成し，評価する．

本研究では，まず 2021 年末時点の COVID-19 ワクチン第 1 回目接種率に差があり人口規模が似ている 2 都市を対象として市民意見の収集を Twitter 上で行う．収集期間は 2021 年 1 月 1 日から同年 12 月 31 日までの 1 年間とした．この際，COVID-19 ワクチン接種に関係のあるツイートのみを収集するためにワクチンの銘柄等の特定のクエリを含み，他の感染症に対するワクチンはクエリで排除することを条件とする．収集したツイート数は札幌市で 84,886 件，福岡市で 48,286 件であった．収集したツイートの中からどちらかの都市特有の意見などのバイアスがかからないようにするため両都市で 8,000 件ずつ同数を抽出し，COVID-19 ワクチン接種に関する 3 属性のアノテーションを行う．ここでは，客観性を担保するために著者以外に複数人の作業協力者で実施する．このようにして構築したデータセットを用いて，各属性について BERT をファインチューニングし分類モデルを作成後，実際に分類を行い，結果に基づいて妥当性と改善点を考察することでデータセットの評価を行う．

対象とする 2 都市は札幌市と福岡市で，アノテーションする属性は 3 つである．分類数はそれぞれ 2 値，4 値，3 値である．著者と 4 人の作業協力者でアノテーションを行い，それぞれの属性について Fleiss の κ 係数が 0.6 を超え，データの信頼性が確認された．

構築したデータセットを用い，各属性についてファインチューニングした BERT で 5 分割交差検証による評価を行った．しかし，「COVID-19 ワクチン接種に関する話題への適合性」の「適合」の再現率が 1.00，「COVID-19 ワクチン接種に対する立場の強さ」の「強」の F 値が 0.00 になるなど不均衡データが原因の問題が起きた．それらに対してダウンサンプリングとアップサンプリングを行うことで，それぞれ 0.632，0.560 となり改善された．

COVID-19 ワクチン接種率に影響を及ぼしうる「COVID-19 ワクチン接種に対する立場」の「反対」ラベルについては F 値が 0.462 であり，ランダムで分類するよりは精度が高いという結果となった．

（指導教員　関 洋平）