

## 複数の属性を組み合わせたソーシャルメディアからの市民意見抽出

石田 哲也

行政の政策や接客業のサービスを向上させるには、市民によるフィードバックを反映させることが必要不可欠である。しかし、パブリックコメント制度や市民公募制度等の既存の行政への市民参加の方法については、参加者の偏りや少なさという問題が自治体から指摘されている。また、コールセンターの設置や顧客満足度調査の実施等の、接客業が顧客の声を獲得する既存の方法では、コストの面で問題がある。そこで本研究では、偏りの無い多くの市民意見を獲得するため、Twitter のつぶやきからの自動的市民意見抽出を目的とする。

しかし、Twitter のつぶやきに現れる市民意見は多岐にわたり、従来の意見分析研究のような意見の極性や感情の種類といった特定の観点のみの分析では不十分である。そこで本研究では、深層学習モデルを用いてつぶやきを複数の観点から分析し、極性、意見タイプ、地域依存性等の複数の属性をつぶやきに自動で付与する。そして、付与した属性を組み合わせて指定することで、多様な市民意見を整理して抽出する。この際、従来の意見分析研究における極性に加えて、アプレイザル理論を用いて、対象に着目した意見のタイプをつぶやきに付与する。また、つぶやきの地域依存性の属性を付与することで、具体的な地名や施設名等を含む、市民の生活環境に根差した意見を抽出する。さらに、市民意見が社会の状況の影響によって推移することに着目し、時系列ごとの市民意見の出現頻度の分析と、時系列を指定した市民意見抽出を行う。

本研究では、BERT をファインチューニングすることで各属性の分類モデルを構築する。さらに、関連する複数のタスクを同時に学習することで精度向上が実現可能な Multi-task learning を適用することで、各属性の分類に対する有効性を検証する。

つぶやきからの市民意見の自動抽出のため、はじめに分類モデルの教師データとなる市民意見分析コーパスを人手で作成する。本研究では、日本で最も多くの市民が暮らす横浜市の市民のつぶやきを用いて、「保育園」と「飲食店のテイクアウトサービス」についての市民意見分析コーパスを作成する。作成した市民意見分析コーパスは、全ての属性において Fleiss の  $\kappa$  係数を用いた判定者間一致度が 0.6 を超えたことから、判定者に依存しない妥当な基準で作成されたことが示された。また、5 分割交差検証を用いた分類精度 (Accuracy) が、多くの属性で 0.8 を超え、高精度な市民意見抽出が可能であることが示された。分類精度が 0.692 と、全属性のうち唯一 0.7 以下であった保育園ドメインの意見タイプ属性については、Multi-task learning によって分類精度が 0.752 まで向上し、t 検定 (有意水準 5%, 両側検定) で有意差が認められた。

時系列ごとの市民意見の出現頻度の分析では、本研究における市民意見抽出が、社会の状況や自治体の対応等の影響による市民意見の出現頻度の推移を捉えられることを実験により明らかにした。また、つぶやきに自動で付与された複数の属性を組み合わせ、さらに投稿時期によって条件を指定して市民意見を抽出することで、行政に反映可能な市民意見や、飲食店の商品の評価を自動で抽出できることについても、実験により明らかにした。

(指導教員 関 洋平)