

## 市民のツイートを行政課題別に集約するための関連語の抽出

南澤 亜樹

市政や行政において、市民の意見を反映させることは重要であり、そのために自治体はアンケート等で意見を収集している。しかし、参加できる市民が限られたり、コストがかかるという問題があることから、本研究では、誰でも簡単に情報発信ができる Twitter におけるツイートを対象にする。一方で、ツイートのはりは多様であり、行政に反映させるためには、行政が対応すべき課題（以下、行政課題とする）に適合するツイートを分類して集約する必要がある。

行政課題に適合する市民のツイートを収集するには、単純には行政課題を示す単語での収集が考えられるが、ツイートは短文であり、直接行政課題を示す単語を明記するとは限らず、十分な数の行政課題に関するツイートを集約できない。本研究では、幅広く市民のツイートを抽出するために、対象となる行政課題の関連語を抽出する手法を提案する。

提案手法では、市民のツイートを多く収集するために、分散表現を用いて抽出した類似単語から関連語候補を抽出する。しかし、行政課題名が類似している場合、関連語候補も類似してしまい、行政課題に適合したツイートを収集する際にノイズが含まれてしまう。本研究では、他の行政課題名との類似度の比較や、関連語候補に対して分散表現を用いて抽出した類似単語に行政課題名が含まれているかで関連語候補のフィルタリングをして、関連語を抽出する。

提案手法の有効性を検証するために、比較手法との比較実験を行った。比較手法は、相互情報量を用いて抽出された単語とする。実験データは、水戸市民、つくば市民と推定したユーザの1年間（2016年11月1日～2017年10月31日）のツイートとする。ツイート数は、水戸市民では、3,663,155件、つくば市民では、8,714,082件である。行政課題としては、「幼稚園」と「保育園」、「電車」と「バス」の二組について、評価を行った。

結果として、提案手法で抽出した有用な関連語の割合が、全ての行政課題において比較手法を上回ることを確認した。また、提案手法で抽出した有用な関連語の割合の平均値が、比較手法を優位に上回ることを確認するために、有意水準1%で対応のある片側t検定を適用したところ、 $p=0.008$ で、有意差を確認できた。

提案手法で抽出した関連語を利用して集約したツイートについて、行政課題に対する適合性を評価した結果、多くの関連語について行政課題別に適合するツイートを集約できた。集約したツイートを確認すると、行政課題名のみでは抽出が難しいツイートも集約することができ、提案手法の有効性を確認できた。一方で、提案手法で抽出された関連語の中には、異なる行政課題に適合するツイートを集約してしまう単語があった。これは、同じ部局に関連する行政課題全般に共通して現れる関連単語が、単語の類似度の大小比較を用いるだけではフィルタリングをすることができなかったことが原因と考えられる。

今後の課題として、市民のツイートのうち、行政課題についての意見が含まれているものに限定をすることで、行政課題に反映しやすい市民意見を抽出することについて検討する。

(指導教員 関 洋平)