

Twitter ミュート機能におけるキーワード拡張手法に関する研究

関根 萌芽

多くの人が自分の考えや見聞きしたことを気軽に発信し他者と共有する Twitter の普及に伴って、Twitter のユーザは苦手とする話題に関する投稿や、楽しみにしていたコンテンツのネタバレ投稿を意図せず見ってしまうこともある。このような状況を回避するための方法として、あらかじめ見たくないとするキーワードを設定することで、そのキーワードを含むツイートをタイムラインや通知からミュート、すなわち非表示にすることができるキーワードミュート機能が提供されている。本論文では、キーワードミュート機能に設定するキーワードを `mute-words` と呼ぶ。

キーワードミュート機能に、ある `mute-words` を設定したとしても、ツイートの中には `mute-words` を含んでいないが `mute-words` に関連する内容のツイートも存在する。そこで本論文では、特定の話題に関するツイートを網羅的にミュートするための関連語を見つけることを目的とした 2 つの `mute-words` の拡張手法を提案する。1 つ目の提案手法は、`mute-words` を含むツイートに同時に出現しやすい語を関連語とみなし拡張を行う手法である。2 つ目の提案手法は、`mute-words` に関するツイートを頻繁に投稿するユーザが使用する語を関連語とみなし拡張を行う手法である。

具体的には、第 1 の提案手法では、指定された `mute-words` を含むツイートを収集し、`tf-idf` 法を適用して拡張することで関連語を抽出する。第 2 の提案手法では、指定された `mute-words` を含むツイートを頻繁に投稿するユーザを抽出した後に、抽出したユーザを分析対象ユーザとして、そのユーザのツイートを `tf-idf` 法を適用して拡張することで関連語を抽出する。分析対象ユーザの抽出に際しては、ツイート本文だけでなくユーザプロフィール等も考慮することで抽出の精度向上を図る。

実験では、予め指定する `mute-words` ごとに関連するツイート群と関連しないツイート群とを正解データセットとして人手で作成し、`mute-words` を拡張し得られた関連語を含むツイートの件数から適合率と再現率を評価した。`mute-words` として「西武ライオンズ」と「コロナ」、「鬼滅の刃」を指定し、実験を行った結果、2 つの提案手法ともに `mute-words` の拡張によってミュート性能を向上できることが明らかとなった。また、話題によって有効な手法が異なることも確認された。この原因として、第 2 の提案手法で抽出した分析対象ユーザの投稿傾向が影響していることが考えられるが、どのような話題で違いが出るのかを明らかにするには、更に多くの `mute-words` で実験するなどが今後の課題である。

(指導教員 佐藤 哲司)