

フォロー別フィルタによるツイートフィルタリングに関する研究

山村 悟

Twitter は、リアルタイム性が高くユーザの嗜好を反映した有益なツイートが多く投稿されており、新たな情報メディアとして注目されている。Twitter では、ユーザは他のユーザを自由にフォローすることができ、フォローしたユーザ (followee) のツイートはユーザのタイムラインに表示される。しかし、followee が投稿するツイートが全てユーザの選好ツイートであるとは限らない。また、ユーザの選好は followee によって異なることが考えられる。例えば、「雨が降ってきた」という気象に関するツイートは、同じ地域に住む followee が投稿した場合に限り選好であると考えられる。Twitter は投稿の手軽さから頻繁にツイートが投稿される傾向があるため、選好でないツイートが表示されることで選好のツイートが埋没し、タイムラインを閲覧する際のユーザの負担を大きくするという問題がある。

本研究では、フォロー別に異なる学習型のフィルタを作成することで、followee に対するユーザの選好を考慮したフィルタリング手法を提案する。followee のツイートだけを用いてフィルタを学習する場合、フォロー直後には十分な学習データが存在しないためフィルタが上手く機能しない。そこで提案手法では、まず分類済みの文書集合を用いた初期学習を行い、ユーザの判定した followee のツイートを用いて逐次的に追加学習を行う。追加学習の手法として、重み付き追加学習、階層フィルタリング、並列フィルタリングという 3 つの手法を考案した。

予備実験として、方式の異なる 4 つのベイジアンフィルタを実装し、フィルタリングの性能を比較した。各方式における最適なパラメータを実験により定め、600 件のツイートをフィルタリングした結果、F 値が 0.934 と最も高かった NaiveBayes 方式を提案法に採用した。

追加学習手法の比較実験として、各手法で 500 件のツイートを追加学習した後で、300 件のツイートをフィルタリングした結果、F 値が 0.871 と最も高かった重み付き追加学習が提案手法に適していることを明らかにした。

今後の課題として、重み付き追加学習における重みの動的な決定が挙げられる。

(指導教員 佐藤 哲司)