

地域特有の単語埋め込み表現を用いたイベント参加地域の推定

小久保 千裕

Twitter などのソーシャル・ネットワーキング・サービス (SNS) は多くの人に利用されており、ユーザはイベントに対する意見や感想を投稿する事が多い。そのため、イベント主催者は Twitter の投稿を分析することで、そのユーザの意見を取得することができる。しかし、同じ名称のイベントが複数地域で開催されている場合、どの地域で開催されたイベントに対するツイートであるかの判別は難しい。

本研究では、ツイートに現れる単語の地域差に着目し、イベント開催地域のツイートや地域 Web ページを用いて地域別単語埋め込み表現を作成し、既存の単語埋め込み表現と組み合わせることで地域特有の単語埋め込み表現を作成する。そして、作成した単語埋め込み表現を用いてツイート投稿者のイベント参加地域を推定する手法を提案する。

地域特有の単語埋め込み表現は、地域別単語埋め込み表現と既存の単語埋め込み表現 hottoSNS-w2v を組み合わせて用いることで作成する。地域別単語埋め込み表現は、イベント開催地域の住民と判定したユーザが発信したツイート、開催地域のイベントに関するツイート、開催地域の Web ページを用いて作成する。ツイートの埋め込み表現抽出には SCDV を用いる。hottoSNS-w2v は、日本語 SNS や日本語 Wikipedia などの大規模コーパスを入力として学習された単語埋め込み表現モデルである。

単語埋め込み表現を組み合わせる際には、地域別単語埋め込み表現には存在するが既存の単語埋め込み表現には存在しない未知語に着目した。まず、地域別単語埋め込み表現モデルを利用して、未知語と hottoSNS-w2v に存在する他の単語との類似度を計算する。次に、類似度の最も高い単語の hottoSNS-w2v における埋め込み表現を、未知語の単語埋め込み表現として新たに登録する。

実験では、仙台七夕まつりと湘南ひらつか七夕まつりを対象として、提案手法を用いた際のツイート投稿者のイベント参加地域の推定精度を評価し、既存の単語埋め込み表現のみを用いた手法と比較する。

実験の結果、既存のモデル hottoSNS-w2v に地域別単語埋め込み表現を新たに登録することにより、hottoSNS-w2v のみを単語埋め込み表現として用いた比較手法よりも F1-score において上回る分類精度が得られることを確認した。また、地域別単語埋め込み表現を作成する際はツイートのみをコーパスとして使用するよりも、地域 Web ページを共に用いることで、分類精度が向上することを確認した。

(指導教員 関洋平)