市民のツイートを利用した分散表現に基づく都市別特徴の分析

安藤 有生

市民の移動は，都市ごとに異なった傾向にあり，時間とともに変化していくものである．そのような特徴をもつ市民の移動を把握することは，自治体や旅行代理店において，市民の需要を把握するうえで重要である．しかし，国勢調査では最新の調査結果を得るのに数年を要することや，宿泊旅行統計調査では都市ごとにまとめられた調査結果が公開されていないことから，市民の移動を把握することは困難である．そこで，本研究では，リアルタイム性を持つ Twitter のデータを使用し，都市ごとに市民の移動の傾向を分析する．

本研究では，都市ごとに収集した Twitter ユーザを用いて，地名と移動に関する動詞が出現しているツイートを，市民の移動先を表すツイートとして収集する．ユーザの居住地は，ユーザのプロフィールに記述されている情報を利用して判定している．市民の行動や認識の違いにより行き先となる地名を含むツイートの表現が異なると考えられるため，市民の移動の傾向は，地名の単語の分散表現を比較することにより明らかにすることができる．そこで，提案手法では，市民の移動先を表すツイートを，skip-gram を用いて単語の分散表現を生成するための学習データとし，多くの人がツイートしている行き先の単語を可視化する．可視化は，地名に対する市民の行動や認識を表す 2 つの対義語と地名との単語の分散表現の類似度の差分を利用して行う．さらに，クラスタリングを行うことで，市民の移動の傾向が似ている都市を分類する．

提案手法の有効性について検証するために，市民の移動の傾向を分析する実験を行った．実験では，8 つの都市ごとに収集した Twitter ユーザについて，5 ヶ月分の全都市で合計 108,507 件のツイートを用いた．都市ごとに，市民の移動の傾向は異なることから，同じ地名でも単語の分散表現は都市ごとに異なる特徴をもつ．本実験では，その都市の市民が，地名に関して，「行く場所か帰る場所か」，「行きたい (帰りたい) 場所か，実際に行った (帰った) 場所か」を明らかにする．そのために，都市ごとに，多くの人が行き先としてツイートしている 20 件の地名に対する単語の分散表現と，動詞「行く」と「帰る」，希望を表す助動詞「たい」と過去・完了を表す助動詞「た」の単語の分散表現とのコサイン類似度を z スコアにより正規化して差分を計算することで，20 件の地名を 2 次元に可視化する．さらに，可視化した地名のクラスタリングを行うことで，共通する傾向にある行き先を分類する．実験の結果，人気のある観光地が「行きたい」地名として可視化され，居住している都市名や帰省先の地名が「帰った」地名として可視化されたことで，単語の分散表現のコサイン類似度に基づく可視化の有効性を示した．また，クラスタリングについては，純度が平均 0.675 となっており，観光地や居住地など，行き先としての目的が明確な地名を分類できることを示した．

今後の課題としては，季節ごとに収集したツイートを用いることにより，季節によって異なる市民についての分析を検討している．また，神社，寺院，美術館といった施設名について，都市ごとの傾向を分析する手法や，ハッシュタグ (#) 等を用いて分析する手法を検討している．

（指導教員　関 洋平）