

論文内容要旨

報告番号	甲 先 第 414 号	氏 名	板東 浩二
学位論文題目	深層ニューラルネットワークを用いた Twitterユーザの興味推定に関する研究		
<p>内容要旨</p> <p>本論文は、Twitterユーザの興味推定を実施するため、複数の深層ニューラルネットワークを用いて分析を行った研究の成果をまとめたものであり、次の5章により構成されている。</p> <p>第1章では、本研究の背景と目的、論文の構成について述べている。</p> <p>第2章では、Twitterを含むSNSのデータ分析に関する関連研究について示し、本研究の位置づけを述べている。</p> <p>第3章では、深層ニューラルネットワークを用いたTwitterユーザの興味推定について提案手法を述べている。本研究では、ユーザのプロフィール情報を使わず、Twitterの連続したツイートからユーザの興味（趣味）を推定する手法を提案する。1つのツイート自体には多くの情報が含まれず、一部のツイートはユーザの趣味に関連しないケースが多いため、信頼性を高めるために複数の連続したツイートから特徴を抽出する。ツイート単位で分散表現の平均ベクトルを抽出し、そのベクトルについて時系列情報を失わずに学習できる、深層学習の一手法である再帰ニューラルネットワークと畳み込みニューラルネットワークを用いる。Twitter上の連続した投稿内容を学習素性としてニューラルネットワークによりユーザの趣味カテゴリを推定する。</p> <p>第4章では、提案手法を用いた評価実験の結果と考察について述べている。分類に用いる連続ツイート数を3~15と可変にし、RNN、LSTM、GRU、CNNの4種類の深層ニューラルネットワークで6つの趣味カテゴリ（game, gourmet, music, craft, art, sports）へのツイート分類実験を行った。各ニューラルネットワークの学習モデルとして、fastTextによるWikipedia記事を事前学習済みの単語分散表現（300次元）を素性に用いた。また、ベースライン手法としてBag of Wordsを特徴量としたRandom Forestと、サポートベクターマシンを用いた分類を実施した。実験の結果、いずれのニューラルネットワークでもベースライン手法を上回る精度を確認し、LSTMを用いて連続ツイート数6のときに最大の推定精度が得られた。一方で、分類対象となるカテゴリ同士の関係性（artとmusicなどが近いなど）の問題、分類に用いる連続ツイートに趣味が含まれない問題などがカテゴリ分類失敗の要因となることが分かった。</p> <p>第5章では、本研究のまとめと今後の研究課題について述べている。プロフィール内容とツイート内容の類似度に基づき、ツイートの取捨選択をすることによる精度改善を試し、分散表現を既存のものから、学習データに含まれるツイートで追加学習させることでより趣味カテゴリ推定に適した分散表現を作成することなどを予定している。</p>			