

# 分散表現技術を用いた 日本語ツイートの感情分類の比較実験

佐藤 一輝

## 要旨

Twitter はツイートと呼ばれる 140 文字以内の短文を投稿できるマイクロブログサービスである。一般的なブログや他の SNS と比較し、気軽にツイートを投稿できるという性質から、利用者の感情が表れやすいと考えられる。またこれまでに、ツイートを対象とした感情分類の研究も数多く行われており、英文ツイートを対象としたものや、感情極性（正か負の二値）を対象としたものなどが代表的である。

本論文では、日本語ツイートの感情分類実験の結果を報告する。具体的には、word2vec や Paragraph Vector, それらの拡張である ScoreSent2Vec や文字情報をベースとした tweet2vec など、各種の分散表現技術を含むベクトル化技術を日本語ツイートへと適用することでツイートをベクトル化し、K-近傍法やランダムフォレスト, SVM を用いて「喜」,「好」,「安」,「哀」,「厭」,「怖」,「怒」,「恥」,「昂」,「驚」の 10 種類の感情と 10 種類以外の感情と無感情を含む「その他」を加えた 11 クラスを対象とした感情分類モデルの構築を行った。構築した各モデルの性能評価実験を通じ、特に分類精度の観点から、それぞれのベクトル化手法や分類器が、どの感情の分類において有効かを検証した。