

Twitter における学術関連情報発信者の特徴分析と分類

Classification and feature analysis of Twitter users posting scientific related information

学籍番号：201221577

氏名：荒川 唯

Yui ARAKAWA

Twitter はマイクロブログの一種で、140 字以内にツイートと呼ばれる記事を投稿するサービスである。人がベースとなり、非対称なゆるいつながりを持っており、新たな情報収集ツールとしても利用できるといえる。Twitter の特性を活かして情報収集を行うにはフォロワーアカウントの選択が重要である。Twitter 研究においてユーザーに関する研究はなされている。本研究では特定の有益な情報として学術関連情報に着目し、学術関連情報を発信するユーザーの特徴の分析を行った。本研究の目的は Twitter 上における学術関連情報発信者の属性やツイートの特徴を明らかにすることである。特徴が分かることによって、フォロワーの判断基準や自動分類のための基礎的な知見になると考えられる。

分析では学術関連の情報を発信すると想定されるアカウントを専門家アカウントと定義し、属性の調査およびツイートのテキスト分析、自動分類実験を行った。データは大学教員のアカウント 119 アカウント 11,557 ツイートを取得、比較のためのアカウントとしてパブリックタイムラインからランダムにアカウントの抽出を行い、119 アカウントを取得、そこから 11,900 ツイートを収集し、非専門家アカウントとして用いた。分類実験の分類器にはランダムフォレストを適用し、評価および重要な特徴量の抽出を行った。

特徴分析では、相対出現頻度 (BOW) において語尾に関する語に違いが見られるなど専門家アカウントにはテキストの特徴があることが分かった。Twitter の特徴としては URL の相対出現頻度が比較的多く、@におけるフォロー/フォロワー関係において対象アカウントが自身はフォローしていないフォロワーと@のやりとりを行う特徴が見られた。専門家/非専門家の分類実験では BOW, BOW+文字数, BOW+Twitter の特徴量, BOW+文字数+Twitter の特徴量を用いたいずれの実験でも精度、再現率、 F 値 9 割を超える性能を得られた。特徴量の重要度の抽出では Twitter の特徴量とし@におけるフォロー/フォロワー関係が有効であることが示唆された。

研究指導教員：芳鐘 冬樹

副研究指導教員：池内 淳