

マイクロブログにおける投稿パターンを考慮した コミュニケーションスキルの分類手法の提案

石川 将吾

イベントや企画のために、ソーシャルメディアを利用できれば、多くの参加者を集められる。その一方で、知らない人と繋がるためには、その人のコミュニケーションスキルを判定する必要がある。

本研究では、マイクロブログへの投稿時間帯や他のユーザとのメンションに基づく交流を手がかりにして、ユーザのコミュニケーションスキルを分類する手法を提案する。

提案手法では、ユーザの発信したツイートの内容を参照して作成した正解データを用いて、特定の時間帯の投稿頻度や、ツイートの種類を投稿パターンとして抽出し、未知のユーザに対して機械学習を用いたコミュニケーションスキルの分類を行う。また、相互情報量を用いて、コミュニケーションスキルごとに特徴のある投稿時間帯を選択し、機械学習に使用する素性とした。

投稿パターンを素性とした、ユーザのコミュニケーションスキルの分類の有効性を示すために、ベースラインとの比較実験を行った。実験では、コミュニケーションスキルの高いユーザ、低いユーザ、どちらでもないユーザの3つの正解ラベルを付与した各147ユーザを対象とした。提案手法は、相互情報量を用いて計算した特徴のある投稿パターンと、投稿時間帯についての投稿数の平均と分散、Twitterの居住地を表す属性であるPlaceNameまたはプロフィール欄への地名の表記の有無を素性とした。Random Forestを用いた実験の結果、提案手法では、F値の平均が全体で0.47、正解率が0.44となり、コミュニケーションスキルの高いユーザについては、F値が0.52となった。また、ベースラインとして、ユーザに対して各ラベルを、ランダム付与した結果、F値の平均が全体で0.30、正解率が0.30になった。さらに、投稿時間帯についての投稿数の平均と分散を素性として使わない場合、F値の平均が全体で0.37、正解率が0.36と低下した。結果について調査したところ、コミュニケーションスキルの低いユーザの投稿時間帯についての投稿数の分散が他のユーザよりも大きくなり、特にリツイートにおける差が顕著であった。このことから、コミュニケーションスキルの低いユーザは、ツイートの投稿時間帯にばらつきがあると分かる。

以上の実験を通し、投稿パターンを用いたコミュニケーションスキルの分類の有効性が確認できた。また、投稿時間帯についての投稿数の分散が、コミュニケーションスキルの分類に有効なことを明らかにした。一方で、コミュニケーションスキルの低いユーザには、ツイートに人を中傷する言葉や、性的と捉えられる言葉が特徴語として現れるため、素性として利用することが考えられる。

今後の予定としては、コミュニケーションスキルごとの特徴語を素性として追加して、正解率を高めることを検討している。

(指導教員 関 洋平)