

漫画の感想を含むツイートの構成要素の判別

宋 亮玄

現在、Web 漫画や電子コミックといったインターネット上で読める漫画やそれらを掲載する Web サイトやアプリの利用が拡大している。そして、電子コミックや Web 漫画ではコメント機能が多くのユーザーに利用されており、コメント機能にはコメントが評価順や投稿時間順に並べられていることが多いが、投稿されたコメントが何について書かれているかやどのような意見が書かれているかを分類できれば、読者がより充実して web 漫画や電子コミックのコメント機能を利用することの支援が期待できると考えた。そこで、本研究では、漫画に対する読者のコメントを分類して、Web 漫画や電子コミックの読者のコメント機能の利用を支援する手法を提案する。

提案手法では、漫画の感想を含むツイートを収集して、アノテーションを行い、データセットを作成する。そして、作成したデータセットを用いて各属性ごとに、事前学習済みモデルの BERT を訓練して分類実験を行う。アノテーションを行う属性は、「ストーリーに関する感想が含まれているか否か (有・無)」、「ストーリーに関する感想の極性 (肯定・否定・中立・意見無し)」、「キャラクターに関する感想が含まれているか否か (有・無)」、「キャラクターに関する感想の極性 (肯定・否定・中立・意見無し)」である。漫画のコメントの中にはそのストーリーに関係する感想やその話の回で注目して描かれたキャラクターについてコメントで書かれることが多いため、本研究の感想の分類ではストーリーとキャラクターに注目した。また、漫画の感想を含めたツイートとして、多種多様なテーマの作品を掲載している週刊少年ジャンプの感想が書かれたツイートに注目した。

分類実験で使用するツイートデータは Twitter API を用いて収集を行う。ツイートの収集日は 2022 年 5 月 16 日と 2022 年 5 月 23 日に行った。収集したツイートは集英社発行の週刊少年ジャンプ 24 号と 25 号のハッシュタグである「#wj24」と「#wj25」を含むツイートである。データセット作成のためのアノテーション作業では、自分を含めた作業員 3 人で行い、アノテーションの一致度を計測した。全ての属性において、Fleiss' Kappa 係数を用いた作業員間の一致度は 0.8 以上となった。BERT を用いたツイートの分類実験は、5 分割交差検証で行い、分類実験の精度は「ストーリーの有無」において 0.772、「ストーリーの極性」において 0.612、「キャラクターの有無」において 0.838、「キャラクターの極性」において 0.642 となった。また、ストーリーとキャラクターの感想の極性の属性値において f 値の結果に差が確認できて、予測ラベルの不正解例からどのような表現が上手く分類されていないかや作成したデータセットの属性値にばらつきがあったこともあり、アノテーション作業での課題も確認できた。

(指導教員 関 洋平)