

EPUB 小説データからのイラスト推薦

石井 慧海

近年、多くの書籍が電子書籍として出版され、タブレット端末やスマートフォン等で動作する電子書籍リーダーが広く普及している。これにより、多くの人が手軽に電子書籍を利用できるようになっている。さらに、最近では無料の電子書籍作成ソフトも数多く公開されており、個人でも簡単に電子書籍を作成することができるようになっている。しかし、提供される電子書籍には表紙がない場合や、表紙があったとしても非常に簡易なデザインとなっており、イラストが何も描かれていない場合がある。表紙が用意されていない書籍に対して、それぞれの小説の内容に合致したイラストを作成するには多くの時間やコストを要し、特にイラストの作成に習熟していないユーザにとっては極めて容易でないと考えられる。さらに、未読の小説に対して内容を反映した表紙を作成することは極めて困難である。このような場合に、小説の印象を表すイラストを推薦することができれば有用であると考えられる。

そこで本研究では、電子書籍の主要なフォーマットとして使用されている EPUB (Electronic Publication) を対象とし、EPUB 形式の小説データから、内容に合ったイラストを表紙として推薦し、選書や電子書籍作成の支援を行うための手法を提案する。

提案手法では、まず、EPUB 形式のファイルから小説の文章を取り出して内容を解析し、Doc2Vec の学習モデルを用いてベクトル化する。また、小説のタイトルも小説の内容を表す重要な単語が含まれていることが多いと考えられるため、同様にベクトル化し文章のベクトルに加算する。次に、イラストについてもベクトル化を行う。本研究では、イラストコミュニケーションサービスである pixiv に投稿されたイラストに着目し、その中でも内容を表すタグが付属したイラストを対象とする。イラストのベクトル化は、イラストに付属したタグからイラストの印象を反映したもののみを抜き出し、リンクされているタグに関連するピクシブ百科事典に掲載された記事のテキストをもとに行う。ここで得られた小説のベクトルとイラストのベクトルとの間のコサイン類似度を算出し、類似度の高いものを小説の内容に合致したイラストとして推薦する。

提案手法をシステムとして実装し、5 編の小説を用いて評価実験を行った。その結果、提案システムを用いることにより、おおよその正解ラベルが付与されたイラストを上位に推薦することができ、提案手法の有用性が示唆された。

(指導教員 鈴木伸崇)