

EPUB 小説データからの表紙の自動生成

川口 晴会

電子書籍と電子書籍リーダーの普及により、近年多くの電子書籍が利用可能となっている。それと同時に、テキストファイルを電子書籍のフォーマットに変換するツールが数多く登場し、青空文庫や Web 小説などの任意のテキストを電子書籍として読むことができるようになった。このような電子書籍の作成における主な問題として、多くの場合に適切な表紙が与えられていないことがある。表紙は書籍の印象を読者に伝える重要な要素であり、選書の際にも有用である。しかし、書籍の内容に合った表紙を作成するには手間やコストがかかり、特に未読の小説を電子書籍に加工する場合、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。

そこで、本研究では電子書籍の主要なフォーマットである EPUB を対象とし、EPUB 形式の小説データからその内容の印象に沿った表紙を自動的に生成するシステムを提案する。表紙を構成する要素のうち色とフォントに着目し、小説の本文を解析してその印象に沿った色とフォントを求める。色については色と単語を対応させた色彩データベースを用いて推薦を行う。提案手法では、130 語の形容詞と色が対応付けられた色彩データベースを、Word2Vec を用いて動詞と名詞も対象となるように拡張した。本文テキストを形態素解析し、色彩データベースに一致する単語を出現頻度に応じてスコアに変換する。色相に基づいて抽出された色をグループ分けし、スコアの合計が最も大きいグループの最上位の色を推薦する。フォントについては各フォントと書籍本文の印象を表した感性ベクトルを使用する。フォントで描画された文字列を提示し、「新しい - 古い」のような複数の形容詞対を用いて SD 法で回答を得る被験者実験を行った。被験者実験によって得られた値を感性ベクトルとし、フォントの画像と感性ベクトルの対を教師データとして Deep Learning の一種である CNN で学習を行う。学習したモデルを用いてフォントから感性ベクトルの推定を行う。また、本文テキストの感性ベクトルは形容詞対に類似した単語の出現頻度によって推定を行う。推定されたフォントの感性ベクトルと本文テキストの感性ベクトルの類似度を求め、最も類似度の高いフォントを推薦する。得られた色を表紙の背景色とし、得られたフォントでタイトルや著者名を描画する。

提案手法によって抽出された色と推薦されたフォントがそれぞれテキスト内容の印象を反映しているか、また生成した表紙が妥当であるかを評価する被験者実験を行った。その結果、色抽出においてはテキスト内容の印象に必ずしも沿わない色が抽出される場合もあるが、フォントの推薦においては概ね被験者の印象に沿うものを推薦できることを確認した。

(指導教員 鈴木伸崇)