

小説データの本文分析による選書支援手法

藤田 佳恵

近年、ネット通販や電子書籍として気軽に書籍が購入できるようになった。それと同時に、サイトごとに様々な形で選書を支援する機能が付いている。代表的なものとして、感想投稿と五つ星評価がある。これらの機能は様々なサイトが有しているが、必ずしも十分とはいえない。感想投稿は、一個人によって書かれているため、その感情に左右されやすく、また、批判的な感想や書籍と関係ない意見が書かれていることがある。また、ネタバレの問題があるため、選書には必ずしも適さない。五つ星評価は、人気度などを判断するには良いが、書籍の方向性などはわからない。また、知名度が低く評価が付いていない書籍は全く判断することができない。このように、インターネット上での書籍購入が進む中で、選書のためのシステムはいまだ不十分である。そのため、本研究では個人の意見に左右されず、また文字によるネタバレがおきない新たな選書支援手法を提案する。

今回提案する手法では、まず感情辞書を作成し、その辞書のデータから小説の印象を算出しグラフ化する。ここで、感情辞書とは単語が小説に与える影響を数値化したものであり、本研究では両極のイメージを持つ「喜び⇔悲しみ」「恐怖⇔怒り」「好意⇔嫌悪」の3つの各軸において辞書を作成した。そして、約40個の小説データに対して3つの軸ごとに印象値を算出した。本研究で作成するグラフは小説の内容を可視化した折れ線グラフと、小説全体が持つ印象値と決定係数を利用したレーダーチャートの2つである。折れ線グラフは小説を19個に区切り、印象値の推移をグラフ化した。小説の流れを知ることにより有効であり、かつ文字によるネタバレが発生しない。したがって、読んだことのない小説のイメージをつかむことに適していると考えられる。レーダーチャートは、3つの軸に対して小説全体の印象値と、折れ線グラフの変化量を表す決定係数を利用した数値を表示した。複数の小説を比較する時に有効であり、小説の特徴を示すことも出来ると考えられる。

提案手法による効果を検証するため、大学生を対象に、本手法で得られた折れ線グラフに関する評価実験を行った。4冊の小説の折れ線グラフから起伏が激しい部分を抽出し、実際の小説のイメージとどれくらい一致しているかを評価したところ、評価対象である16か所の感情に対して11か所が概ね一致しているという結果が得られた。

以上の結果により、本システムで作成した折れ線グラフは小説内の印象を表すのに有効であると考えられる。軸の設定・感情辞書の作成・グラフ化という各段階でさらなる検証を行い、レーダーチャートを含めたグラフの効果的な見せ方を研究することで選書支援に対する精度の向上が期待できる。

(指導教員 鈴木伸崇)