4M-5

小説データの本文分析による選書支援手法の提案

藤田　佳恵 †　　鈴木　伸崇 ‡

† 筑波大学 情報学群 知識情報・図書館学類　　‡ 筑波大学 図書館情報メディア系

1　はじめに

近年，様々なサイトで気軽に書籍が購入できるようになった．それらのサイトは，選書を支援する機能を提供しており，代表的なものとして感想投稿と五つ星評価がある．しかし，それらの機能にはいくつかの問題がある．感想投稿は，投稿者の感情に左右されやすく，過度に批判的な感想や書籍と関係ない意見が書かれることがある．また，投稿内容が小説の内容を示唆していることがあり，ネタバレの危険性がある．五つ星評価は，人気度などは判断できるが，書籍の方向性などは把握できず，また評価の付いていない書籍の判断は行えない．このような問題を生じることなく，書籍の印象を読者に伝えることのできる手法が実現できれば有用と考えられる．

本稿では，個人の感情に左右されず，かつ文章によるネタバレを起こさずに小説の印象をユーザに提示する手法を提案する．本手法では，小説データに対して，Plutchik 理論 [4] に基づいて「喜び⇔悲しみ」「恐怖⇔怒り」「好意⇔嫌悪」の 3 つの軸ごとに印象値を算出しグラフとして提示する．作成するグラフは，折れ線グラフとレーダーチャートの２つである．前者は小説の印象値の推移をグラフ化したものであり．小説全体の流れを把握することができる．後者は，3 つの軸に対して小説の印象値 (小説を通しての平均) と，折れ線グラフの変化度 (グラフの上下動の激しさ) を表す数値を表示する．ただし本稿では，紙面の都合上レーダーチャートに関する説明は省略する．

関連研究として，梶山らは，小説の内容に合った色を算出する手法を提案している [1]．村井らは，興味喚起部分の試し読みを行う手法を提案している [3]．

2　提案手法

2.1　感情辞書

感情辞書とは,「単語が文章に与える影響を数値化したもの」であり，文献 [2] の手法を参考に作成した．この手法は新聞記事を対象としたものであり，両極の印象を持った軸を設定し，そのどちらに近いかが単語ごとに 0 から 1 の範囲で印象値として表される．本手法では，Plutchik 理論 [4] に基づき「喜び⇔悲しみ」「好意⇔嫌悪」「恐怖⇔怒り」の 3 つの軸を用いている (印象値は，左が 1 で右が 0 に対応する)．文献 [2] の手法では，印象値の算出の際，軸の両極ごとに基準となる印象語を設定する必要がある．本手法では，小説特有の幅広い表現に対応するため，類語辞典 (Weblio 等) を参考に 20 個の印象語を選定した．青空文庫の日本文学から 200 件を選定して MeCab で形態素解析し，出現する単語の印象値を算出して辞書を構築した．

2.2　折れ線グラフ

折れ線グラフは，小説を分割して個々の区間の印象値を求めグラフとして表したものである．これにより，小説を通しての印象値の推移を表現できる．図 1 のように小説を 19 個の区間に分けて，個々の区間の印象値を計算する．各区間の印象値は，次のようにして求める．まず，3 つの印象軸に対して，次の 2 つの値を計算する．(1) 感情辞書を参照してその区間に出現する単語の印象値の平均を求める．(2) 小説全体に出現する単語の印象値の平均を求める.「(1) の値 – (2) の値」が，当該区間の当該軸に対する印象値となる．

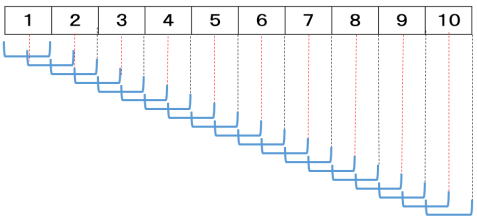

図 1: 小説データの分割方法

例として,『トロッコ』の折れ線グラフを図 2 に示す．この作品は，主人公の良平が冒険を求めてトロッコに乗って遠出する話である．良平は作業員の許可を貰い嬉々としてトロッコに乗るが，日が暮れた後にたった一人で泣きそうになりながら長い距離を走って帰る．以下，図 2 の四角で囲まれた部分に着目して説明する.「喜び⇔悲しみ」軸 (黄色) に注目すると，中盤の喜びはトロッコを楽しく押している場面で，嬉々とした様子が見られる.「恐怖⇔怒り」軸 (青色) は最後に良平が感じている恐怖がグラフの上昇としてみることができる.「好

Proposed method to help select the book by analyzing the text of the book

†Yoshie FUJITA　‡Nobutaka SUZUKI

†College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba

‡Faculty of Library, Information and Media Science, University of Tsukuba

Copyright ©2014 Information Processing Society of Japan.
All Rights Reserved.

意⇔嫌悪」軸 (緑色) は，中盤で喜びが強く現れた場面で，トロッコを押すことを許してくれた作業員への好意がグラフに現れている．更に，終盤の帰る場面では，人恋しく思っている場面であり，その部分からグラフの終盤に好意が表れていると考えられる．

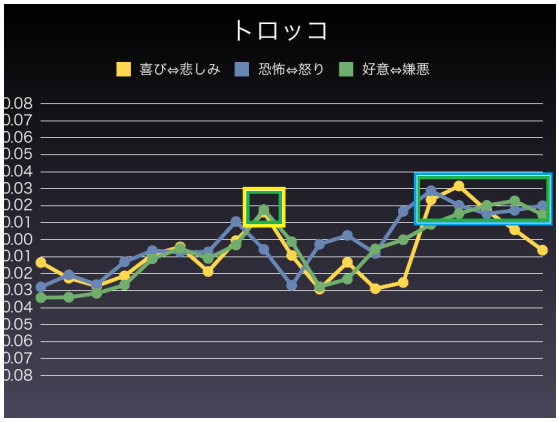

図 2: 『トロッコ』の折れ線グラフ

3 評価実験

大学生 9 名を対象に，小説の本文と本手法で作成したグラフを見比べて，両者の印象がどの程度一致しているかを 5 段階 (1 から 5) で判断してもらった．値が大きいほど一致度が高いことを示している．利用した小説は，『羅生門』『蜘蛛の糸』『手袋を買いに』『走れメロス』の 4 作品である．この 4 作品を選んだ理由は，授業でもよく取り上げられており，多くの人に読まれている可能性が高いからである．各小説の折れ線グラフに対して，特に起伏が激しく特徴的と考えられる部分 2 箇所を A・B という形で抜き出し，その部分で強く表れている感情と小説との一致度を調べた．実験に利用した『羅生門』のグラフを図 3 に示す (左の赤丸が A，右の赤丸が B)．なお，実験協力者には，グラフの他，小説の本文データを随時参照可能とし，A と B に該当する部分の文章には色づけして正確な位置が分かるようにしている．

表 1: 平均一致度

タイトル	喜び⇔悲しみ	恐怖⇔怒り	好意⇔嫌悪	調査人数/平均一致度
羅生門 A	2.11	3.11	2.22	9 人
羅生門 B	2.78	3.56	-	2.76
蜘蛛の糸 A	3.14	2.28	-	7 人
蜘蛛の糸 B	4.14	3.86	3.28	3.34
手袋を買いに A	4.33	-	-	6 人
手袋を買いに B	-	2.50	3.67	3.50
走れメロス A	4.22	-	4.11	9 人
走れメロス B	-	3.78	-	4.04
平均	3.45	3.18	3.32	-

結果を表 1 に示す．3 個の軸に対する 16 個の平均一致度のうち，11 個で 3 を上回っており，グラフと小説の印象は概ね一致していると考えられる．今回の評価実験では『走れメロス』の一致度が最も高くなっており，この小説には「印象語が文中に出現するような喜怒哀楽が表現されている」「一人称視点で書かれている」といった傾向が見られた．一方,『羅生門』の一致度が低くなったのは,「複雑な表現で感情が書かれている」「情景描写が多い」「複数人視点で書かれている」といった傾向があるためと考えられる．また,「恐怖⇔怒り」軸は小説内で同時に出現する場面があり，この軸自体の再検討が必要であると考えられる．更に，出現頻度が非常に低い単語に対して偏った印象値が設定されていることが確認できため，感情辞書を作成するときの小説数を増やす等の対応が必要であると考えられる．

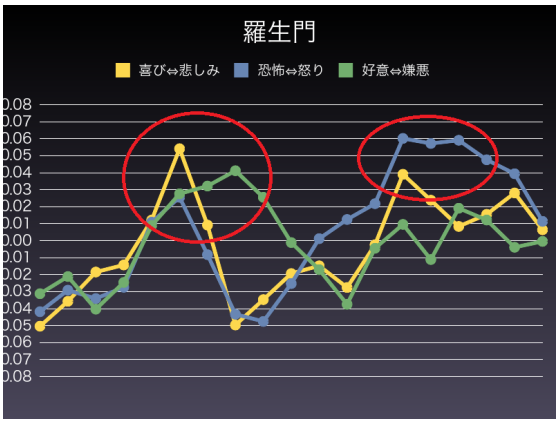

図 3: 評価実験で用いた『羅生門』のグラフ

4 むすび

本稿では，折れ線グラフを用いて小説の印象の推移を示す手法を提案した．今後の課題として，まず，複数人の感情が混合してる小説・情景描写の多い小説を考慮する必要がある．次に，グラフの両極の感情が同時に出現する文章を考慮する必要がある．最後に，出現頻度の低い単語の印象値を改善する必要がある．

参考文献

[1] 梶山，越前，“読者の印象の色彩化による表紙画像生成の評価”，信学技報, 122(46)，pp.199–204, 2012.

[2] 熊本，河合，田中，“新聞記事を対象とするテキスト印象マイニング手法の設計と評価”，信学論．94-D(3), 2011.

[3] 村井，牛尼，“電子書籍小説の選別における興味喚起度 MAP の有効性の評価”，情処研報, 155(2)，pp.1–8, 2012.

[4] R. Plutchik, ”The Nature of Emotions”, American Scientist, 89, pp.344–350, 2001.

Copyright ©2014 Information Processing Society of Japan.
All Rights Reserved.