

登場人物の感情表現文の抽出による 物語のハイライト生成

三木 建哉[†] 松本 和幸[†] 吉田 稔[†] 北 研二[†]
[†] 徳島大学理工学部

1. はじめに

電子書籍の普及により数多くの文学作品をインターネット上で読むことが可能になった。しかし、読者がそれの中から読む作品を選ぶことは容易ではない。本稿では物語について登場人物の感情が表現されている文を抽出することによる指示的要約の手法を提案する。

2. 既存研究

横野が感情表現とそれに関係する文を抽出することによる物語の要約手法を提案している[1]。物語の地の文について登場人物の動作文が重要であると、中でも感情が描写されている文を重要文として抽出している。

本稿で提案する手法では、感情表現の出現文をスコア付けすることで、要約文の短縮化を目指している。

3. 提案手法



図1. 提案手法の流れ

3.1 登場人物の推定 登場人物は文中に動作の主語として出現していると考えられる。登場人物の推定はKNP¹による格解析の結果と動詞項構造シソーラス辞書[2]の格情報を用い、動作主に当たる語を抽出することで行う。

3.2 感情表現の抽出 各文で感情表現が行われているかどうかは、感情表現辞典[3]に掲載されている感情語との一致により判定を行う。表記ゆれへの対処としてJuman++²を用いて感情語と本文の両方を形態素ごとに分割し、代表表記を照らし合わせて一致を判定する。

3.3 登場人物と感情の対応付け 感情語に対応する登場人物の推定は、3.1の結果をもとにして作成した動詞と登場人物を関連付けたデータを用いて行う。KNPの係り受け解析結果から感情語と係り受けの関係がある動詞を探し、その動詞に対応する登場人物を感情語と対応付ける。

3.4 要約 感情表現辞典では感情語が喜, 好, 昂, 安, 驚, 怒, 哀, 怖, 恥, 厭の10分類で掲載されている。この分類をもとにして抽出した文を登場人物ごとに10次元のベクトルで表現する。例えば、「好」で登録され

ている感情語が1回登場する文は「好」に対応する要素の値を1とし、2回登場する文は2とする。また、感情表現辞典には1つの感情語に複数の感情が登録されているものがあつたため、そのような語に対しては後に行う処理への影響が小さくなるように1語のベクトルの大きさが1になるように値の調整を行う。

このベクトルから登場人物ごとに感情の大きさ・変化・強まりをそれぞれ式1, 2, 3により算出する。 \mathbf{v} は付与したベクトル, l は行番号である。

$$\text{大きさ: } M_i = \|\mathbf{v}_i\|_2 \quad (1)$$

$$\text{変化: } C_i = \frac{\|\mathbf{v}_i - \mathbf{v}_{i-1}\|_2}{l_i - l_{i-1}} \quad (2)$$

$$\text{強まり: } S_i = \frac{\mathbf{v}_i \cdot \mathbf{v}_{i-1}}{\|\mathbf{v}_i\|_2 \|\mathbf{v}_{i-1}\|_2} \quad (3)$$

文に対して登場人物ごとに算出した感情の大きさ・変化・強まりの絶対値を足し合わせた値で文のスコアを算出し、文をランキングする。

4. 提案手法の評価

既存手法として、1文中のTF-IDF値の和でランキングした結果と比較を行う。小説は青空文庫³の「杜子春」を使用し、各手法の上位10文と事前に人手で選択した物語中のハイライト文(26文)の一致度を適合率で、ランキングとして上位に正解データが存在するかを平均適合率で評価する。結果を表1に示す。

表1. 「杜子春」における各手法の上位10文での比較結果

	適合率	平均適合率
提案手法	0.4	0.62
TF-IDF	0.3	0.79

5. 今後の課題

会話文での感情表現は話者の感情であると考えられるため、会話文中の感情語については主語の推定方法を変更する必要がある。

参考文献

- [1] 横野光, “登場人物の感情表現に着目した物語要約”, 言語処理学会第13回年次報告, 2007
- [2] 竹内ほか, “日本語コーパスに対する動詞項構造シソーラスの概念と意味役割のアノテーション”, 言語処理学会第19回年次大会, 2013
- [3] 中村明, “感情表現辞典”, 東京堂出版, 1993

¹ <https://nlp.ist.i.kyoto-u.ac.jp/?KNP>

² <https://nlp.ist.i.kyoto-u.ac.jp/?JUMAN%2B%2B>

³ <https://www.aozora.gr.jp/>