登場人物の感情表現に着目した物語要約

横野 光 †
† 岡山大学大学院自然科学研究科
kara@momo.it.okayama-u.ac.jp

1 はじめに

インターネットの普及によって我々は膨大な量のテキスト情報を入手することが可能になった．その種類は論説文や新聞記事などのようなものから，ブログの記事や日記など多岐にわたる．文芸作品に関しても，電子出版の普及や，著作権の切れた文学作品を電子テキスト化し公開している青空文庫[1]のようなインターネット電子図書館，また，アマチュアが公開している小説など多数存在している．

このように公開されている多くの物語の中から読みたいものを探すことは手間がかかる作業である．その負担を軽減する際に，要約はユーザにとって役に立つ情報となりうる．しかし，全ての物語に要約が存在しているわけではなく，人手で要約を作成することは労力を必要とする．そのため計算機による要約の自動生成は有用である．

物語の自動要約に関する研究には，物語中の命題間の関係に着目したもの [1] や隣接文間の結束性に基づいたもの [2] などがある．また，その他の要約手法としては Lehnert の提案した plot unit によるものがある [3]．これは物語の構造の中心は登場人物の感情的な反応と感情の状態，それらの因果関係にあるという考えに基づいている．この要約手法では物語中の出来事に対して登場人物が良い感情を抱いたか悪い感情を抱いたかという情報を付与し，それらを組み合わせて物語の構造を組み立て，要約を作成する．しかし，出来事に対して登場人物がどのような感情を抱いたかと推定することは困難である．

本稿では Lehnert のモデルを参考に物語において登場人物の感情の変化が示されている部分は他の部分に比べて重要であるという仮説を立て，これに基づいた要約手法を提案する．

2 提案手法

本手法では物語の展開に重要な箇所を抽出することで要約を生成する．

物語の展開に重要な役割を果たすのは登場人物の動作文である [4][2]．よって，物語中の登場人物の動作文を全て抽出すれば，それは一種の要約であるといえる．しかし，全ての動作が物語にとって重要であるというわけではなく，これによって得られた要約は冗長なものになることがある．より要約率の高い要約を得るためには，動作文の中でもより重要とみなせるものだけを抽出するということが考えられる．

本稿では物語中のある地の文が重要かどうかの判定に，その文に登場人物の感情が描写されているかどうかを利用する．登場人物の感情が表現されている箇所は物語において重要であるとみなし，その箇所を要約として抽出する．また，可読性を考慮して，物語で場面が転換することを示す表現も要約に含める．

抽出した文だけでは文章としてのまとまりが悪く脈絡のない要約が生成されることがある．このことを考慮して，抽出した文に対して結束性があると判断される周辺の文も要約に取り入れる．

本手法の流れを以下に示す．

Step 1. 登場人物推定
Step 2. 照応解析
Step 3. 重要文抽出
Step 4. 隣接文の抽出

2.1 登場人物推定

登場人物はその物語に固有の表現であるものが多く，それらは単語として一般的な辞書に登録されていない．また，『さるかに合戦』に登場する “うす” のように辞書に登録されている語が登場人物であっても，それが人名でない場合もある．

したがって，登場人物の推定に際して人名辞書などの既存の辞書だけに頼った方法では不十分といえる．

本稿では登場人物と考えられる文字列を自動的に辞書に追加することで物語固有の表現に対応する．そして，その辞書を利用して登場人物として出現する表現を推定する．

具体的な手順を以下に示す．

物語の地の文において複数回出現する文字列のうち，以下の条件を満たすものを登場人物候補とする．

- 平仮名，カタカナの小文字から始まっていない
- 句読点，カギ括弧，“！”，“？” を含まない
- 文字列中に接続詞，連体詞，代名詞を含まない
- 複数の形態素で構成されている
- 文章中に “は” を接続した形で出現する

得られた候補の中で同頻度で出現した他の候補文字列の部分文字列であるものは候補から削除し，残った登場人物候補を固有名詞として形態素辞書に追加する．

登場人物は物語において重要な役割を担っているため，主題として取り上げられることが多く [5]，また，登場人物は意志を持つものであるため，意志性のある動作文の動作主として現れることが多い．この傾向を利用して，登場人物として出現しているかどうかを推定する．

物語の地の文の形態素解析を行い，主題化する機能を持つ副助詞 “は” を後接して複数回出現する

[1] http://www.aozora.gr.jp
[2] 参考文献 [5] による

名詞で，その文の述部となる動詞に一つでも意志性のある動作を示すものがあれば，その名詞を登場人物と推定する．

意志性のある動作を示す動詞の判定には日本語和語動詞 LCS[6] を利用した．これは日本語和語動詞を中心に約 1000 個の動詞に関してその語彙概念構造を記述を試みたものである．このうち，実際に LCS が付与されている延べ 742 項目の動詞を利用した．

2.2 照応解析

文章では照応や省略が用いられることがあるため，登場人物の動作などを示す文であっても，登場人物の表現そのものが文に出現しないことがある．

本稿では要約として抽出する文の判定に登場人物の表現が文中に出現するかどうかを利用するので，全ての照応に対してその先行詞の同定を行うのではなく，主題の省略に対してと，人物を指す代名詞である “彼” と “彼女” に対してのみ先行詞の同定を行う．

主題の省略は前文と主題が一致しているときに起こりやすい．よって，ガ格，またはハ格がない文に対しては，直前の地の文の主題を先行詞とする．

人称代名詞 “彼”，“彼女” が文中に出現したとき，その前の地の文の先行詞候補から先行詞を選ぶ．先行詞候補は前文の名詞である．複文の主節に代名詞が出現した場合は従属節の名詞も先行詞候補に含む．

選ぶ際の先行詞のなりやすさの順位は以下のように決定する [7].

ハ格＞ガ格＞ニ格＞ヲ格＞その他

“彼” は男性を示す人物を，“彼女” は女性を示す人物を先行詞とする代名詞であるので，先行詞候補のうち登場人物であるものに対しては優先順位を上げる．これらの値については実験的に決定した．

ある登場人物表現に対して，それを先行詞とする代名詞は “彼” か “彼女” のどちらかに決定される．したがって，物語の全ての地の文に対する照応解析が終了したとき，登場人物表現を先行詞とする代名詞の数を数え上げ，“彼” と “彼女” の両方が含まれていた場合，その数の多い方を正解とし，間違った先行詞推定をしている照応に関しては，その先行詞候補を省いて照応解析をやり直す．

2.3 重要文抽出

2.3.1 感情描写文の抽出

物語の重要な出来事の中心となる文として，登場人物の感情を描写している文を抽出する．

感情を描写している文かどうかは，感情表現辞典 [8] に掲載されている感情表現を含んでいるかで判定する．感情表現辞典は近代や現代の小説 808 作品で使用されている喜怒哀楽といった基本的な感情表現に加え，臨時的な精神状態の表現までを収録したものである．本稿ではこの辞書に掲載されている語句編の延べ 2136 項目を感情表現の抽出に利用した．

文が感情表現を含んでいて，その文のハ格に現れる名詞が登場人物であるか，また，文中の感情表現が係っている文節に登場人物が含まれている動作文であれば．その文を重要文として抽出する．

さらに，文章の脈絡を考慮して，抽出した感情表現の生起原因と考えられる文も要約に取り入れる．

直後の文に “なぜなら” のように理由を示す接続詞が使われていれば，それを抽出する．そのような文がなければ，直前の動作文を生起原因とみなして，その文を重要文として抽出する．

2.3.2 場面転換表現の抽出

要約文の可読性を考慮し，物語中での場面転換を示す文とその場面において登場人物がはじめて出現する動作文も要約に含める．

場面転換を示す表現としては，作者が明示的に定めた章の区切りの他に，登場人物の場所の移動や時間の経過の表現が考えられる．

本稿では作者が明示的に定めた章の区切りと，変化を現す動詞 “なる” が述部に用いられている文を場面転換の表現として採用した．

物語中に複数行の空白があれば，それを作者の意図的な場面転換とみなして，空白の後にあるはじめの一文を要約に取り入れる．

動詞 “なる” を用いた文に関しては，形式段落のはじめの文で，その文中に名詞+助詞 “に”+動詞 “なる” という表現が出現していた場合，それを抽出する．

2.4 隣接文の抽出

抽出した重要文だけでは，抽出文に “その” や “この” といった指示詞などが含まれていた場合，それが何を指すのかという情報が抜けてしまい，話題のまとまりの欠けた不完全な要約になってしまう．

そこで局所的なまとまりを考慮して，抽出した文と連接関係がある文も要約として抽出する．連接関係は 2 文間に含まれる以下の言語形式で求められる [9].

- 接続語句
- 指示語
- 助詞・助動詞
- 同語反復・言い換え
- 応答詞

本稿では，このうち接続語句と指示語にのみ着目して連接関係を判定する．

抽出した文を s_i とし，物語の地の文において s_i の直前の文を s_{i-1} とする．以下の条件のいずれかを満たせば s_{i-1} を s_i と連接関係のある文として要約に含める．

- s_i に指示詞が含まれている
- s_i に “そこで”，“すると” など展開を示す接続詞が使われていれる

これによって得られた文に対しても同様に前文との連接関係の判定を行い，連接関係がないと判定されるまで繰り返す．

3 評価

本手法の有効性を検証するために，青空文庫に公開されている小説を入力として与え，本モデルを適用して生成した要約文に対して，その要約率と要約文の網羅性と可読性に対する評価を行った．

物語には様々な種類のものが存在する．それらを一律に扱って評価するのは困難である．本稿では，現在において最も数の多い現代小説を主な対象と考える．しかし，そのほとんどが著作権の問題で自由に参照できる形で公開されていないので，現代小説に形式の近い新字新仮名で書かれた近代小説を評価対象にした．

3.1 実験と結果

本稿での要約生成では，形態素解析にMeCab[3]を，係り受け解析にCaboCha[4]を利用した．

評価に使用した小説は以下の3作品である．

作品 a. 菊池寛,『俊寛』
作品 b. 宮沢賢治,『風の又三郎』
作品 c. 芥川龍之介,『杜子春』

要約率

表1に各作品の要約の要約率を示す．比較対象として，物語中の登場人物が現れる動作文を抜き出したものを使用した．

作品名	動作文のみ	提案手法
作品 a	68.8%	34.9%
作品 b	63.8%	43.5%
作品 c	48.3%	44.2%

表 1: 要約率

本手法によって生成された要約の方が登場人物が出現する動作文のみを抜き出した要約に比べて高い要約率の要約文が得られた．

新聞記事の要約に比べると本手法での要約率は高くないが，物語は新聞記事に比べて要約として抽出すべき箇所の判定が困難であり，このことを考慮すれば，得られた要約の要約率は低くないと考えられる．

要約文の網羅性

要約文が物語中の重要な場面を抽出できているかどうかについて，人手で作成された要約と比較を行った．作品 a の俊寛が島に残されて憤る場面や作品 c の杜子春が地獄で母親が苦しむ様を見て叫ぶ場面など，作品中で重要とされる場面が抽出することができた．

抽出できなかった主な場面を表2に示す．この分析は3.2節で行う．

[3]http://mecab.sourceforge.net/
[4]http://www.chasen.org/~taku/software/cabocha/

作品名	抽出できなかった主な場面
作品 a 作品 b 作品 c	・俊寛の島での新しい生活の様子 ・三郎が転校する ・杜子春が再びお金持ちになる ・馬となった父母が杜子春の前に連れられてくる

表 2: 抽出できなかった主な場面

要約文の可読性

4人の被験者に生成した要約文を読んでもらい，主観的に判断して文間のつながりがおかしいと思った箇所を指摘してもらった．

可読性の評価結果を表3に示す．平均文間隔は要約文中にどのくらいの間隔でつながりが悪いと感じた箇所が出現するかを示したものである．

作品名	箇所数	要約文数	平均文間隔
作品 a	7	156	22.2
作品 b	17	298	17.5
作品 c	12	69	5.75

表 3: 要約文の可読性の評価

対象とした物語の構造の複雑さや文章の長さに対して，得られた要約文の可読性はそれほど低くなかった．

3.2 考察

3.2.1 要約文の網羅性に関する考察

抽出できなかった場面についての考察を述べる．

人手による要約に含まれ，本手法によって生成された要約には含まれなかった表現は主に以下の2種類である．

- 場面転換の表現
- 台詞文

場面転換の表現に関しては，本稿では登場人物の場所の移動や，時間経過の表現を考慮に入れなかったため，これらの表現は抽出できなかった．

『俊寛』で抽出できなかった表現の周辺の要約文を示す．

(1). 俊寛は跪いて，目に見えぬ何物かに，心からの感謝を捧げたかった．
(2). とうとう，俊寛はその五尺を越ゆる大魚を征服してしまう．

(1) と (2) の間には，別の日に俊寛が釣りを始めるという表現があり，これは場面の転換を示していると考えられる．

場面転換の表現の抽出には述語の動詞に移動を表すものが使われているかどうかや，“次の日” など時間を表す表現が使われているかどうかで判定ができると考えられる．しかし，場面の転換は場所の移動や時間の経過以外でも示される．(3) は『杜子春』の要約中の一文である．

(3). 杜子春はその翌日から，忽ち天下第一の大金持に返りました.

これは杜子春の状況が変化したことによって場面転換を示している．本稿では場面転換の表現として登場人物の変化を表す動詞である “なる” に着目したが，(3) のようにその動詞が使われていなくても変化を表している文があり，登場人物の状態変化を表す動詞を考慮することでこれらの表現が抽出できると考えられる．

この他の抽出できなかった表現に，物語の重要な要素が登場人物の台詞として現れるものがある．

『風の又三郎』の要約では，最後は以下の文で終わっている．

(4). 「たいへん早いですね．あなたがた二人で教室の掃除をしているのですか.」先生がききました.

本文ではこの後にある先生の台詞で三郎が転校していったことが説明される．これはこの物語でも特に重要な要素であるが，本手法では登場人物の台詞の内容を考慮していないために抽出できなかった.

3.2.2 可読性に関する考察

複数の被験者によって文間のつながりが悪いと指摘された箇所についての考察を述べる．

つながりが悪いと指摘された原因には，場面転換表現の欠如などの要約として抽出されなければいけない表現がなかったというものの他に連接関係の判定誤りが挙げられる．

例えば,『風の又三郎』の要約に以下のような文章がある.

(5). 「はい.」先生は嘉助を指さしました.

(6). 「わあ，うまい，そりゃ，やっぱり又三郎だな.」(以下略)

本文における (5) と (6) の間には,

- 「高田さん名はなんて言うべな.」
- 「高田三郎さんです.」

という 2 人の会話がある．これが要約中に含まれていないために (5) と (6) はつながりが悪いと感じられた．これは連接関係の判定に会話を考慮していないのが原因である.

この他に，照応詞を含む文が要約に含まれていて，その先行詞を含む文が要約に含まれなかったためにつながりが悪いと指摘された箇所があった.

(7) は『杜子春』の要約の一文である.

(7). 杜子春の声には今までにない晴れ晴れした調子が罩っていました.

本文における (7) の前には以下に示す杜子春の台詞がある.

- 「何になっても，人間らしい，正直な暮しをするつもりです」

(7) 中の “杜子春の声” はこの台詞を指している．この台詞は “杜子春の声” に対する先行詞であると考えられるが，(7) 中に表層として指示詞などを含んでいないため，連接関係がないと判定され，その結果台詞が抽出されず，つながりが悪いと判定された.

これらに対しては，より詳細な照応解析を行い先行詞との関係を求めることで解決できると考えられる.

4 おわりに

本稿では，感情表現に着目した物語の要約手法を提案し，その評価を行った.

動作文のみを抜き出したものに比べて，本手法によって生成された要約の方が要約率が高く，登場人物の感情に着目することで物語における重要な要素を絞り込むことができることが分かった．しかし，本手法では場面転換の表現や重要な要素を含む台詞を抽出することができなかった.

今後の課題としては，これらの要約に必要な要素の抽出を行うとともに，抽出した要約文の縮約や統合などによる要約率の向上が挙げられる.

参考文献

[1] 村上聡, 上之薗和宏, 榎津秀次, 古宮誠一: 物語の自動要約, *The 18th Annual Conference of the Japanese Society for Artificial Intelligence* (2004).

[2] 山本悠二, 増山繁, 酒井浩之: 小説自動要約のための隣接文間の結束性判定手法, 言語処理学会第 12 回年次報告 (2006).

[3] Lehnert, W. G.: Plot units and narrative summarization, *Cognitive Science*, Vol. 5, pp. 293–331 (1981).

[4] Longcre, R. E.: Two hypotheses regarding text generation and analysis, *Discourse Processes*, Vol. 12, pp. 413–460 (1989).

[5] 泉子・K・メイナード: 談話分析の可能性, くろしお出版 (1997).

[6] 加藤恒昭, 畠山真一, 坂本浩, 伊藤たかね: 日本語和語動詞に関する語彙概念構造辞書構築の試み, 言語処理学会第 11 回年次報告 (2005).

[7] 石崎雅人, 伝康晴: 談話と対話, 東京大学出版会 (2001).

[8] 中村明: 感情表現辞典, 東京堂出版 (1993).

[9] 永野賢: 文章論総説, 朝倉書店 (1986).