

小説検索システムのための人物間関係推定に関する基礎研究*

神谷美希[†], 當間愛晃[‡], 赤嶺 有平[‡], 山田 孝治[‡], 遠藤 聡志[‡]

琉球大学大学院理工学研究科情報工学専攻[†], 琉球大学工学部情報工学科[‡]

1 はじめに

現在では通販サイトなどを通じて膨大な数の電子書籍が流通している。しかし、数が膨大なぶん読みたい小説のタイトルや著者が決まっていな場合、自分の求めている小説を見つけるのは難しい。そのため、ユーザーが求めている小説を簡単な質問を繰り返すことで探してくれる検索システムを構築することを最終目標としている。

小説の検索システムを構築するためにはまず、どのような小説であるか特徴づける必要がある。物語の読み手は登場人物が引き起こすできごとと心情の表象に応じて興味を感じる [1]。これより、登場人物による行動、登場人物同士の関係などを物語の進行にそって表したプロットを作成し、それを小説の特徴とする。登場人物の情報や相関図に関する研究として [2] がある。[2] では、人名、登場人物をノードとし人物関係をエッジとしたグラフによって人物相関図を表現している。ここでは関連度が高いほどノード間のエッジを短くしており、重要な登場人物が図の中央に来ようになっている。これを参考にし、プロットを作成したい。

本稿では、小説検索システム構築の手始めとなるプロット作成を目的とし、場所・登場人物・登場人物同士の関係性・できごとを用いたプロット作成の実験結果について報告する

2 提案手法

2.1 システム概要および本稿の狙い

小説の本文を利用し、求めている小説を検索できるシステムの構築を目指す。このシステムは予め作成されたプロットを用い、ユーザが入力した自然文から必要な情報を抽出し、抽出した情報とプロットを比較す

ることで推薦候補を絞り込む。候補が多数存在している場合には絞り込み用の質問を提示し、それに対する回答を利用して適当数に減るまで絞り込むことを想定している。現在はこの前段階であるプロット作成における敵対関係有無の判定に着手している。

2.2 プロット作成

プロットは各場面の「場所」、「登場人物」、「登場人物同士の関係性」、「できごと」の4項目の組み合わせで表現される。本稿では、小説の場面分けは時間、場所、出てくる登場人物の変化を考慮し手作業で行った。また、「登場人物」「場所」「できごと」はあらかじめ抽出できてるものとみなし、その前提で「登場人物同士の関係性」を「敵」かそれ以外かを判断するだけの問題設定とした。

2.3 敵か否かの判断実験 1

敵同士は戦いや攻撃を表す動作が文中に現れると考え、登場人物同士の動作を使用した。一文に登場人物が二人以上いる場合に限定し、その分における動詞を特徴として抽出する。また、あらゆる表示方法(引っ掻く、引っかく、ひっかく)に対応するため読みを使用した。例えば「犬は庭を駆け回り、猫はこたつで寝ました。」という一文からは犬と猫が登場人物とした場合、「かけまわる、ねる」が特徴となる。動詞の抽出を行うために Mecab を使用した。これを一場面の全文に行い、抽出した動詞の中から [3] の分類に基づき攻撃を表す動詞のみを抽出する。登場人物同士が敵か否かの判断を式 (1) を使い行う。数値が高いほど敵である可能性が高くなると考える。

$$\frac{\text{キーワードとマッチした単語数}}{\text{固定の登場人物二人に関する一場面すべての動詞数}} \times 100 \quad (1)$$

2.4 結果と考察 1

図1の「結果: 実験1」に示す。また、図1は赤文字が敵と判断するのが正しい項目、黒文字が敵ではないと判断するのが正しい項目となっている。実験1の結果より、「きじ・猿」が敵である可能性が一番高くなり、

*A basic study on a character relationship estimation for novel search system

[†]Miki KAMIYA, Information Engineering Course, Graduate School of Engineering and Science, University of the Ryukyus

[‡]Naruaki TOMA, Yuhei AKAMINE, Koji YAMADA, Satoshi ENDO, Department of Information Engineering, University of the Ryukyus

逆に敵同士と判断して欲しい「桃太郎・鬼」は敵である可能性が低くなった。原因として動詞や名詞の動作主、受動者を無視したことが原因と考える。例えば「犬が鬼にくいつき、猿は鬼に引っかきました。」など二人以上の登場人物が一文に出現した場合、その一文の動詞と登場人物名以外の名詞をすべて抽出、特徴としてしまい「犬・鬼」「猿・鬼」「犬・猿」それぞれが「くいつく、ひっかく」を特徴としてしまい、スコアが高くなる点。登場人物同士の特徴となる動詞とキーワードの動詞を比較するさい読みを使用したため、「轢く(ひく)」のように攻撃を表す動詞と「引く(ひく)」のような攻撃を表さない動詞がマッチする点。今回使用したキーワードのみだと少なすぎたという3点があげられる。

2.5 追加実験(実験2)

実験1の結果を踏まえ、キーワードと登場人物同士の特徴を比較する際に読みではなく基本形を使用し、さらに類義語辞典を利用し、さまざまな漢字とひらがなの組み合わせに対応するようにした。また比較対象となる登場人物同士の特徴を動詞のみではなく登場人物名以外の名詞を追加した。キーワードに[5]の競争・争い・戦争・攻防・勝敗・軍事から、名詞と動詞を増やした。実験1で使用した文章と同じもの、敵か否かを判断する式を式(2)を使用し実験をおこなった。

$$\frac{\text{キーワードとマッチした単語数}}{\text{固定の登場人物二人に関する一場面すべての動詞、名詞数}} \times 100 \quad (2)$$

2.6 結果と考察2

追加実験の結果を図1の「結果: 実験2」に示す。これより、実験1で敵である確率が高くなっていた「おじいさん・おばあさん」などの3項目が敵である確率が低くなり、「桃太郎・鬼」の的である確率が高くなり、改善が見られた。しかし、敵と判断して欲しい「桃太郎・鬼」より「きじ・犬」の敵である可能性が高くなっている。これは動詞や名詞の動作主、受動者を考慮することで改善されると考える。また、今回キーワードとマッチした単語は「つつく、攻め込む、降参、引っかく」のみであり、キーワードとなっていないが戦いや攻撃などに関する言葉を手作業で探した結果、逃げるなどの単語計5つ見つかった。これらはキーワードの不足も悪化要因に繋がっていることを意味している。

3 今後の予定

今回の結果から、動詞、名詞のみを使い、敵か否かを判断するのは難しいことがわかった。動作の動作主、受

人物	おじいさん おばあさん	おじいさん 桃太郎	おばあさん 桃太郎	きじ 犬	きじ 猿	きじ 鬼
結果: 実験1	11.11	11.11	11.11	7.14	12.5	4.00
結果: 実験2	0.00	0.00	0.00	3.70	3.26	4.04

人物	桃太郎 犬	桃太郎 猿	桃太郎 鬼	犬 猿	犬 鬼	猿 鬼
結果: 実験1	0.00	0.00	0.00	5.00	7.14	7.14
結果: 実験2	0.00	0.00	3.54	3.06	4.48	4.48

図 1: 敵か否かの判断

動者に関する問題解決のために係り受けや、キーワードとなる単語が出現している一文から動作主や受動者に関するパターンがないか調べていきたい。また、「敵」以外の関係性のパターンを増やしていきたいと考えている。関係検出漏れがあった場合、登場人物同士が現れる場面で関係性を確認できるパターンがないか検討しパターンを増やす。しかし、これだけだと重要ではない人物同士の関係が抽出され、プロットが複雑になりすぎる。このため、馬場ら [2] の研究を参考に、人物 A と人物 B の関連度の高さを算出し、関連度の高い人物同士が現れる文に注目し、優劣を付けてプロットに含めるか否かを判定したい。

参考文献

- [1] 大村彰道 監修, 秋田喜代美, 久野雅樹 編集: 文章理解の心理学 認知、発達、教育の広がりの中で (2001)
- [2] 馬場 こづえ, 藤井 敦: “小説テキストを対象とした人物情報の抽出と体系化”, 言語処理学会第13回年次大会発表論文集, pp.574-577(2007)
- [3] 五段活用-動詞-品詞の分類-Weblio 辞書: http://www.weblio.jp/parts-of-speech/五段活用_1
- [4] 楠山正雄 桃太郎: http://www.aozora.gr.jp/cards/000329/files/18376_12100.html
- [5] 国立国語研究所編: 分類語彙表 増補改訂版, 大日本図書,(2004)
- [6] 神谷 美希, 當間 愛晃: “小説検索システムのためのプロット作成に関する基礎研究”, 情報処理学会第77回全国大会講演論文集, 分冊2, pp.169-170(2015)