

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*

神谷美希[†] 當間愛晃[†] 琉球大学工学部情報工学科[†]

1 はじめに

インターネットの普及により、通販サイトから手軽に小説が購入できるようになった。また、これらと並行して [1] のような電子媒体での小説が普及し、現在では通販サイトなどを通じて 10 万冊以上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い場合、自分の求めている小説を見つけるのは難しい。単純な検索システムで「泣ける本」など、あやふやな情報のみの検索をおこなうと検出される本は多くなる。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 [2]。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 [3] がある。[3] では、人名、その人物の「年齢」「性別」「職業」「身体的特徴」を抽出し、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

レビューなどを用いると人気度に依存してしまい選別しにくい小説がでてきてしまうため、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。検索システムの考案を図 1 に示す。はじめにユーザーからどのような本を探しているのかを「主人公と仲間が悪と戦うストーリー」など自然文で入力

してもらい、そこから入力された自然文のキーワードを抽出し、プロットと比較を行うことで小説の数を絞り込む。この時点で対象となる小説の数が多すぎる場合には、数を絞るために「主人公は女性がいいですか」のような質問をユーザーに提示する。ユーザーは「はい」「いいえ」「どちらでもかまわない」から選ぶ形で回答する事により、対象小説を絞り込む。対象小説が十分な数、例えば 3 冊程度になった時点でその結果を提示する。

このシステムは予め作成されたプロットを用い、ユーザー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を進めている。

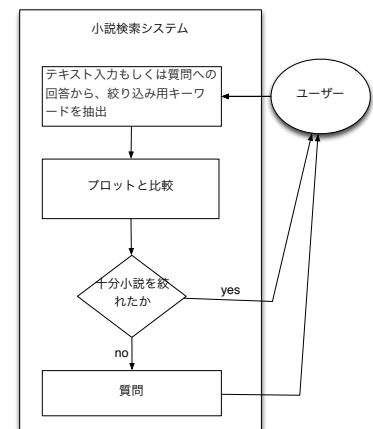


図 1: 検索システムの構成案

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。プロットの完成イメージ図が図 2 である。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」はあらかじめ抽出出来るものとみなし、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

*A Basic Study of Plot Building for Novel Search System

[†]Miki KAMIYA, Naruaki TOMA, The Department of Information Engineering, University of the Ryukyus

た。

場所	登場人物	登場人物同士の関係	できごと
あるところ	おばあさん、おじいさん、桃太郎	おじいさん $\xrightarrow{\text{夫婦}}$ おばあさん おじいさん $\xrightarrow{\text{育てる}}$ 桃太郎 おばあさん $\xrightarrow{\text{育てる}}$ 桃太郎	桃を拾う 桃から男の子が生まれる
あるところ	おじいさん、おばあさん、桃太郎、鬼	おじいさん $\xrightarrow{\text{夫婦}}$ おばあさん おじいさん $\xrightarrow{\text{育てる}}$ 桃太郎 おばあさん $\xrightarrow{\text{育てる}}$ 桃太郎 桃太郎 $\xrightarrow{\text{退治}}$ 鬼	鬼退治へ出かける
旅の途中	桃太郎、犬、雉、猿	桃太郎 $\xrightarrow{\text{仲間}}$ 犬、雉、猿	犬、雉、猿が仲間になる
鬼ヶ島	桃太郎、犬、猿、雉、鬼	桃太郎 $\xrightarrow{\text{仲間}}$ 犬、雉、猿 桃太郎 $\xrightarrow{\text{退治}}$ 鬼 犬 $\xrightarrow{\text{退治}}$ 鬼 猿 $\xrightarrow{\text{退治}}$ 鬼 雉 $\xrightarrow{\text{退治}}$ 鬼	鬼を退治

図 2: 桃太郎のプロットイメージ図

2.3 敵か否かの判断

敵同士は攻撃をし、それを表す動詞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、その一文のすべての動詞を抽出、登場人物同士の特徴とする。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、「駆け回る」と「寝る」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、抽出した動詞を [4] より、攻撃を表す動詞と比較し、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。

$$\frac{\text{攻撃を表す動詞とマッチした動詞数}}{\text{固定の登場人物二人に関する一場面すべての動詞数}} \times 100 \quad (1)$$

2.4 結果および考察

青空文庫より [5] を 4 つの場面に分け、最後の場面での登場人物同士の敵か否かの判断を行った結果が図 3 となる。結果から「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」が一番敵である可能性が高くなり、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「桃太郎が殴り、おじいさんとおばあさんは助かりました。」という一文では桃太郎、おじいさん、おばあさんの各登場人物同士の動詞は「殴る、助かる」になり、おじいさんとおばあさんに関係ない「殴る」(攻撃を表す動詞)が入っているためスコアが高くなる点。二人以上の登場人物が一文に出現した場合、その一文の動詞をすべて抽出、特徴としてしまい「犬が鬼に噛み付き、猿は鬼に引っ掻く。」という文から「犬・鬼」「猿・鬼」「犬・猿」それぞれが「噛み付く」と「引っ掻く」を特徴としてしまい、敵であると判断されやすくなるという点。敵判断のキーワードとなる [4] の攻撃を

登場人物	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
計算結果	14	14	14	7	13	4

登場人物	桃太郎 犬	桃太郎 猿	桃太郎 鬼	犬 猿	犬 鬼	猿 鬼
計算結果	0	0	0	5	7	7

図 3: 敵か否かの判断

表す動詞だけでは少なかった点があげられる。

3 今後の予定

今回の結果から、動詞のみを使い、敵か否かを判断するのは難しいことがわかった。そのため、係り受けを使い、誰が誰に対しておこなったことかを考慮したい。また、「A する」という文の場合、動詞である「する」より「A」の部分が必要になる。これより戦いに関する名詞などをキーワードとして増やす。また、登場人物たちが現れる文から、「A は B を C した」などの動作主、受動者が確認できる場合、「C した」の部分が登場人物同士の関係性に使用できないか実験を行い、「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 [4] 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] 青空文庫:<http://www.aozora.gr.jp/>
- [2] 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)
- [3] 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)
- [4] 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
- [5] 楠山正雄 桃太郎:http://www.aozora.gr.jp/cards/000329/files/18376_12100.html