

漫画のあらすじから抽出した 登場人物の関係を考慮したエピソード推薦

涌井 春那[†]公立ほこだて未来大学[†]奥野 拓[‡]公立ほこだて未来大学[‡]

1 はじめに

近年、漫画の楽しみ方として「推し」の登場人物を中心に楽しむことがトレンドであり、多くの人が検索エンジンを使用して推しとそれに関わる登場人物のエピソードを検索している。推しの登場エピソードは簡単に見つかるが、推しとそれに関わる登場人物のエピソードの検索は適切なキーワードが分からず難しい。そこで本研究では、エピソードごとのあらすじから登場人物を抽出し、登場人物の関係を考慮したエピソードを推薦するシステムを構築する。

2 関連研究

田中らは、人物関連図によって小説の内容の想起を支援するシステムを構築している [1]。人物関連図を描画するために、人物 A と人物 B の関連度を、ある文における人物 A から見た人物 B との関係と、人物 B から見た人物 A との関係の総和として求める。前者の関係は、ある文と前後の文のいずれかの主語が人物 A かつその文中の主語、述語、目的語のいずれかに人物 B が存在するとき関連していると見なす関数によって求め、後者の関係も同様に求める。この手法では、人物判定の正誤と言い換え表現をユーザが修正する必要がある。

宮川らは、類似した人物関係が描かれる他

の漫画を検索するために、関係と性格を相関図で表示するシステムを構築している [2]。Wikipedia から漫画のキャラクタ概要文を取得し、関係辞書を用いて関係を抽出している。関係辞書とは、友人や親友、ライバルなど人物間の関係を表す語を登録した辞書である。関係辞書は、RELATIONSHIP^{*1}と連想類語辞典を用いて構築している。キャラクタ概要文から修飾・非修飾関係にある語を抽出し、関係辞書内の語と一致する語と固有名詞を含む文節をキャラクタの関係として収集し、相関図を作成している。

西原らは、物語テキストから半自動で構築したパターンと人手で構築したパターンを用いた関係抽出手法を提案している [3]。3 項関係 P1-R-P1 と、2 項関係 P1-P2&R を抽出するパターンを構築し、マッチした登場人物と関係辞書に含まれる関係語から、人物関係を得る。P1, P2 はそれぞれ人名を、R は関係語 (人物間の関係を表す語)、P2&R は人物と関係の両方を表す語を示す。例えば、2 項関係のとき「太郎の父」の「太郎」が P1、「父」が P2&R で、3 項関係のとき「太郎の弟の次郎」の「太郎」が P1、「弟」が R、「次郎」が P2 である。関係語は、角川類語新辞典と日本語語彙大系から人手で選別し、関係辞書に追加している。

3 提案手法

エピソード推薦に用いる関係と関連度を以下の手順で求める。テキスト形式の漫画のあらす

Episode Recommendation Considering Character Relationships Extracted from Manga Episode Synopsis

[†] Haruna Wakui, Future University Hakodate

[‡] Taku Okuno, Future University Hakodate

^{*1} <https://vocab.org/relationship/>

表1 Wikipediaの概要文から構築したパターン

パターン
P1のRである(P2 P2とP3)
P1の~Rである~P2
P1とP2(の は)R.
P1(の は を が)R・P2
P1は~R・P1
R・P1(~ 、)P2
P1・P2R
P1のR~P2
P1(が は が~ という~人物が~)P2のR
P1(は の)RのP2
P1のRのP2(数字)R

じ、キャラクタ概要文、登場人物リストを用いる。利用するあらすじの条件は、人物関係に関する記述が豊富であることである。最初に、あらすじ、概要文に記述されている人物名を登場人物リストに従ってフルネームに名寄せする。次に、形態素解析で利用する辞書に、登場人物リストの人物名を追加する。その後、形態素解析であらすじと概要文から登場人物と関係を抽出し、関連度を算出する。関係を、Wikipediaの概要文から構築したパターン(表1)をマッチングすることで抽出する。表1において「~」は任意の文字列を示す。パターンのRである関係語を、宮川らが用いているRELATIONSHIPと西原らが用いている角川類語新辞典の語彙分類体系に加えて、日本語WordNetから選び、関係の抽出に用いる関係辞書に追加する。関連度は、田中らの手法を用いて算出する。本研究では、あらすじ毎に文の数(N)が異なるため、関連度に1/Nを掛けて正規化する。求めた関係の中からユーザが選び、その人物間の関連度が高いエピソードを推薦する。

4 評価実験

4.1 目的と方法

関係抽出・関連度算出の精度を評価することを目的とした実験を実施する。関係抽出では、Wikipediaの概要文から構築したパターンと、西原らのパターンを比較・評価する。対象とする作品は「名探偵コナン」*2である。あらすじ

は「アニヲタ Wiki(仮)」*3、概要文はWikipediaの「名探偵コナン」の記事、登場人物リストは「全事件レポート編纂室」*4から取得する。実験では、「少年・青年マンガ10選」*5に含まれる漫画(対象とする作品を除く)のWikipediaの概要文からパターンを構築した。

4.2 結果と考察

同じ関係辞書で、同じテキストを用いた結果、西原らのパターンの正解率は22.94%、Wikipediaから新たに構築したパターンは60.42%となり、関係抽出の正解率が37.48ポイント向上した。

不正解の内訳として、抽出誤りや、同じ人物関係の抽出があった。抽出誤りの多くは、パターンの任意の文字列に含まれる人名や代名詞などとの関係が抽出されていないことが原因であるため、パターンを見直す必要がある。

5 まとめ

本研究では、漫画のあらすじから抽出した登場人物の関係を考慮したエピソードの推薦を目指し、関係抽出と関連度を算出した。今後は、関係の抽出精度の向上を目指す。

参考文献

- [1] 田中翔太郎, 岡部誠, 尾内利紀夫: 物語の内容想起支援インタフェースの開発, 情報処理学会第55回プログラミング・シンポジウム予稿集, Vol. 2014, pp. 7-15 (2014).
- [2] 宮川菜奈, 松下光範: 異なる作品間のキャラクタの関係を比較するための相関図可視化システム, DEIM2022, No. F23-5 (2022).
- [3] 西原弘真, 白井清昭: 物語テキストを対象とした登場人物の関係抽出, 言語処理学会第21回年次大会発表論文集, pp. 628-631 (2015).

*3 <https://w.atwiki.jp/aniwotawiki/>

*4 <https://websunday.net/conandb/>

*5 <https://booklive.jp/feature/index/id/manga100>

*2 青山剛昌: 名探偵コナン, 小学館 (1994).