

概要と読者レビューに基づく漫画特徴抽出

白石 絵里奈* 田村 亮介* 浅沼 爽汰* 白井 聡一* 藤田 和成* 町田 翔* 延澤 志保*

*東京都市大学知識工学部

*東京都市大学大学院工学研究科

1 はじめに

国内で流通する漫画のタイトル数は、2010年の時点で、出版・印刷分で29,364点、電子配信分で10万点以上に及ぶ[1]。これだけ膨大な作品数の中から、読者の好みに合う作品を探すことは、自動でも人手でも難しい。本研究では、漫画推薦の一助としての漫画特徴抽出として、作品の概要と読者の興味をそれぞれ反映するような、作品の特徴を表す重要な語句を自動で抽出する手法を提案する。

2 漫画推薦手法についての関連研究

漫画の推薦では、例えば、商用サイト等で活用されている協調フィルタリング等の手法が考えられる。これは読者の購入傾向等を基に推薦を行うため、人気作品が出力されやすい。一方、人気がない、購入履歴がない作品等は推薦されづらく、推薦対象の偏りを生む恐れがある。

本研究では、購入傾向ではなく、漫画の内容や特徴のみを用いて漫画の推薦を行う手法を検討する。

村瀬らは、内容の類似度に着目して、作品の概要文の類似性に基づいて推薦を行う手法を提案した[2]。村瀬らの手法では、推薦対象候補の漫画と、読者の好みの漫画のストーリーのマッチングをとることで漫画の推薦を行う。

これに対して、漫画の推薦に際しては、内容よりもジャンルや話の展開、絵柄、キャラクター等が反映されるべきとの考え方がある。例えば、「高校生が部活動で努力を重ね優勝する」という内容の作品でも、少年漫画と少女漫画では絵柄が異なり、ギャグや恋愛要素の有無等で読者の好みは分かれる。村松らは、レビュー文から作品の評価を表す単語を抽出し、これをベクトルとして用いて作品間の類似度を算出することで漫画の推薦を行った[3]。村松らの手法では、ジャンルの同じ漫画は類似しているとの仮定に基づき、出力結果はジャンルを考慮して評価している。

3 概要とレビューの双方を考慮する漫画特徴抽出

漫画の推薦では、読者のニーズに合う作品か否かを判断するのに十分な情報が必要である。しかし、漫画の好みを言葉で表現することは一般に難しい。少女漫画等といったジャンルは括りが大き過ぎ、作品全体からの絞り込みには使えても、候補作品の絞り込みには向かない。同じ作品を好む読者も、ストーリーやキャラクター、絵柄等、着目点が同じとは限らない。

そこで本研究では、作品の特徴語を概要とレビューの双方からそれぞれ抽出することで、作品の内容やジャンルと、読者視点での評価項目とを合わせて提示する手法を提案する。本研究では、作品の内容を示す特徴語を「概要キーワード」、作品の読者視点での評価に関連する特徴語を「評判キーワード」と呼ぶ。概要キーワードおよび評判キーワードの抽出に際して、形態素解析器はMecab¹を、辞書はmecab-ipadic-NEologd²を利用した。

3.1 概要キーワード

村瀬らの用いたWikipediaでの漫画作品の記述[2]には、簡単な説明(図1)に続いて、概要やあらすじ、キャラクター一覧等が含まれる。図1の例では、「ちはやふる」と

ウィキペディア

ちはやふる

『ちはやふる』は、末次由紀による日本の少女漫画、およびそれを原作としたテレビアニメ、実写の映画、『BE・LOVE』（講談社）において2008年2号号から連載中。競技かるたに没頭する少女の青春を描く。アニメ化及び実写映画化された他、現実社会の競技かるた漫遊にも影響を及ぼした[4]。

図1: Wikipedia上の漫画作品の紹介例

いう漫画が少女漫画であること、競技かるたや青春を題材としていること等が冒頭で簡潔に説明されており、その後、あらすじ等具体的な内容が記述されている。例えば先の例(図1)では、「競技かるた」等が概要キーワードとして期待される。このように概要キーワードは名詞が主と考えられるため、本研究では、名詞一般、名詞-サ変接続、名詞-固有名詞-一般のうち、出現回数が3回以上の単語を概要キーワードとして抽出する。

Feature Extraction Based on Wikipedia Description and Reader Reviews

Erina Shiraiishi*, Ryosuke Tamura*, Sota Asanuma*, Souichi Shirai*, Kazumasa Fujita*, Sho Machida*, and Shiho H. Nobesawa*.

* Faculty of Knowledge Engineering, Tokyo City University

* Graduate School of Engineering, Tokyo City University

¹Mecab, <http://taku910.github.io/mecab/>.

²mecab-ipadic-NEologd, <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>.

3.2 評判キーワード

読者の興味を反映する評判キーワードの抽出には、村松らの手法 [3] 同様、レビュー (図 2) を利用する。

- ・ニッチな事象を描いているわりに、ストーリーにどんどん惹き込まれ、面白くてたまらない。
- ・全身全霊をかけて何かに打ち込むことを力強く表現してくれている。

図 2: 漫画作品のレビュー例

図 2 の例では、例えば「惹き込まれる」や「面白い」等が評判キーワードとして期待される。そこで本研究では、形容詞-自立、形容詞-非自立を評判キーワードの候補とする。形容詞は活用語のため、基本形で集計を行う。

4 漫画特徴抽出実験結果

本研究では、「ちはやふる」と「ハイキュー！」の 2 作品を対象として、特徴抽出実験を行った (表 1)。概要キ

表 1: 2 作品の特徴の比較

	ちはやふる	ハイキュー！
Wiki 概要 抜粋	末次由紀による日本の少女漫画、およびそれを原作としたテレビアニメ、実写の映画。競技かるたに没頭する少女の青春を描く。現実社会の競技かるた浸透にも影響を及ぼした。	古館春一による高校バレーボールを題材にした日本の漫画作品。
概要 キーワード	新, かるた, 高校, クイーン, 優勝, 対戦, 出場, A 級, 予選, 全国大会, 競技かるた, 団体戦, 大会, 運命, 想い, 名人戦, 宣言, 個人戦, 決勝戦, 相手, 勝利, 準決勝, 部員, 二人, 代表, 昇格, 小学校, 再会, チーム, 目標, 夢, 前半, 進学, 自分, 札, 告白, 決勝, 苦戦, 引退, 防衛, 部室, 敗退, 入部, 対決, 吹奏楽部, 春, 修学旅行, 自身, 試合, 指導, 再戦, 永世名人, モデル	音, 駒, 高校, 日向, 予選, 代表, 試合, 学園, 勝利, 梟, 練習試合, 谷, 合宿, 入部, 参加, 攻撃, セッター, インターハイ, コース, チーム, 中学時代, 大会, 対戦, 相手, 準決勝, 工, 戸, 決定, 強豪, 遠征, リベロ, エース, 練習, 流川, 目, 復帰, 美学, 波, 南, 沢, 対応, 全国大会, 善, 選抜, 糸, 初戦, 出場, 寺, 最後, 公式戦, 繋, 苦戦, 強化合宿, ライバル, ブロック, フルセット, バレーボール, トライアウト, チームメイト, コート, 23 巻, 1 年生
評判 キーワード	おもしろい, 強い, あつい, すごい, はやい, 素晴らしい, うまい, たのしい, 切ない, 美しい, やすい	おもしろい, すごい, あつい, はやい, 素晴らしい, 強い, 楽しい, カッコいい, 待ち遠しい, やすい, 嬉しい, うまい

ワードは Wikipedia の概要やあらすじから抽出した。評判キーワードは Amazon.com やレビューサイトにおける、漫画のレビュー文から抽出した。

作品の説明のため掲載した Wikipedia の冒頭説明部分 (表 1) では、この 2 作品はジャンルも話の内容も異なり、類似性は見られない。

概要キーワードを見ると、この 2 作品はどちらも高校校部活動をテーマに、大会での勝利をチームで目指す内容であることが窺える (表 1)。大きく異なるのは、「ちはやふる」は競技かるた「ハイキュー！」はバレーボールとい

う点である。しかし、競技対象の違いに絡むキーワード以外には、概要キーワードに大きな差は見られない。例えば「テニスの王子様」という作品では、同様に「全国大会」「優勝」等の語が並ぶのに混じって、「超人技」「個性的」等の語が多く見られ、作品の方向性が先の 2 作品とは異なることが見て取れる。

評判キーワードを見ると、どちらも「おもしろい」「強い」等、似た語が並ぶことがわかる (表 1)。これはどの作品でも同じわけではなく、例えば「テニスの王子様」では「おもしろい」が多い等、現れるキーワードやその頻度に違いがある。

このように、概要キーワードは作品の内容やジャンル、評判キーワードでは作品の印象や感想等が抽出可能であり、その双方を合わせて提示することで、作品の特徴を直観的につかむことが可能である。

Wikipedia の利用については、評判キーワードの抽出に不向きなことの他に、もう 1 点問題がある。Wikipedia は作品によって記述量が大きく異なり、記述の多い作品は説明やあらすじが詳細に記述されている等、内容にも差が見られる。レビューも同様に作品ごとに量が異なるが、量が増えることで記述が詳細になるわけではない点で、Wikipedia と異なる。この問題を解決するには、Wikipedia 以外の概要情報を積極的に追加する等、データの検討を進める必要がある。

5 まとめ

本研究では、漫画の推薦に役立つような、作品の特徴と読者の興味点の提示を目的として、キーワードとなる語句の推定を行った。キーワードとなる語句の推定には、作品の概要とレビューを用いた。概要は、作品の設定や題材が客観的に書かれており、作品の類似の判断材料になりうる。一方レビューは、設定や題材からは読み取れない、実際の読者の着眼点や作品の評価が書かれている。概要とレビューの両方を用いることで、作品の特徴を表す語句と、読者の興味を具体的に反映する語句を推定した。概要から推定した語句を概要キーワード、レビューから推定した語句を評判キーワードとし、作品ごとに提示した。

参考文献

- [1] 一般財団法人デジタルコンテンツ協会, “漫画制作・流通ガイド”, 経済産業省, 2015.
- [2] 村瀬 尊好, 柊 和佑, 安藤 友晴, “マンガの概要に基づく作品紹介システム”, 第 11 回情報科学技術フォーラム, pp.319-325, 2012.
- [3] 村松 拓実, 町田 翔, 河野 一志, 延澤 志保, “読者レビューでの評価情報に基づく漫画の推薦,” 情報処理学会第 79 回全国大会, no.3Q-07, pp.537-538, 2017.