

読者レビューでの評価情報に基づく漫画の推薦

村松 拓実*

町田 翔*

河野 一志*

延澤 志保*

*東京都市大学知識工学部

1 はじめに

近年、漫画雑誌の銘柄や漫画作家の増加に伴い、漫画の単行本の新刊点数は年々増加の傾向にあり、2012年には1年間で29,394種類の単行本が出版されている[1]。漫画の種類は多種多様で、ユーザーごとに漫画作品の好みが変わり、漫画作品によっても絵柄やストーリーなど評価される箇所が異なるため、ユーザーは自分の好みに合った漫画作品を見つけるのが困難である。

2 関連研究

村瀬らは漫画の概要に出現する単語がどれだけ類似しているかを比較することによって設定や舞台、ストーリーの似た漫画作品を推薦する手法を提案している[2]。しかし、漫画作品によって「絵が上手い」や「キャラクターがかっこいい」など評価される箇所が異なる。そのため評価に関する情報にも着目する必要がある。

現在、ブログやレビューサイトなどで読んだ漫画についてのレビューを投稿している読者が多く、レビューからその漫画の評価情報を得ることができる。そこで本稿では、読者によって投稿された漫画作品のレビューに基づいた漫画作品の推薦手法を提案する。

評判分析の分野では、商品やサービスの機能や特徴を表す語句を評価属性、評価視点などと呼ぶ。例えば食品などでは値段や味がそれらにあたる。またこれらの評価基準となる語に対して、良い、悪いなどの評価を表す語を評価表現と呼ぶ。本稿では前者を評価視点、後者を評価表現と呼ぶ。

3 提案手法

本研究では漫画作品の推薦のため、漫画レビュー中の評価視点と評価表現に基づき作品の類似度を測ることにより、評価される箇所が似た作品を推定する。

本システムは漫画作品名を入力データとし、漫画レビューサイトと通販サイト上から抽出した評価表現と評価視点を利用したベクトル空間モデルを用いて作品同士の類似度を測定し、推薦を行う(図1)。

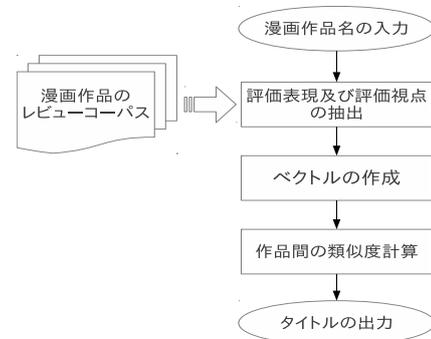


図1: 推薦システムのフローチャート

3.1 評価表現の抽出及び評価視点の抽出

評価視点に対する意見や評価に関する記述には「絵が上手い」の評価視点と評価表現が文中で主語述語の修飾関係にあることが多い。本手法では類似度計算に用いるベクトルを作成するために、まず評価表現を抽出し、評価表現と係り受け関係にある名詞を評価視点として抽出する。

評価表現の抽出には、平山らが評価表現の抽出に利用していた、評価表現に用いられる品詞を利用する。表1に評価表現抽出に利用した品詞のパターン[3]を示す。まず、入力されたレビューを1文ずつ読み込み係り受

表1: 評価表現の抽出に利用する品詞[3]

名詞(形容動詞語幹)	形容詞
名詞(ナイ形容動詞語幹)	動詞(係り先の次の語が形容詞の場合)

け解析器 CaboCha¹を用いて係り受け情報を取得する。次に形態素解析器 MeCab²を用いてレビュー文を形態素解析することで得られた品詞情報を元に、評価表現の品詞パターンに当てはまる語を評価表現として抽出する。評価表現を抽出した後に、抽出された評価表現の係り元である名詞を評価視点として抽出するが、単純に名詞のみを抽出すると評価視点としてふさわしくない語が多く抽出されてしまうためこれらの語は評価視点として抽出しない。名詞のうち、非自立、数、形容動詞語幹、接尾、代名詞、副詞可能、人名、地域といった語は評価視点の説明に使われていることが多いため、評価視点として扱わない。

3.2 ベクトルの作成

それぞれの漫画作品の類似度を計算するために、作品ごとに評価視点を素性としたベクトルを作成する。本

Comic Book Recommendation Based on Evaluation Information in Reader Reviews.

Takumi Muramatsu*, Sho Machida*, Kazushi Kouno*, and Shiho H.Nobesawa*.

* Faculty of Knowledge Engineering, Tokyo City University

¹CaboCha, <https://taku910.github.io/cabocha/>

²MeCab, <http://taku910.github.io/mecab/>

研究では評価視点に着目することと評価視点の極性に着目することの有効性を確認するため、素性の値が異なる2種類のベクトルを用意し、2パターンの類似度計算を行う。ベクトルの素性は、その単語が出現する文書数である $df(\text{Document Frequency})$ の値が2以上の評価視点とし2種類のベクトルで共通している。df値が1の単語はその漫画作品固有の単語と考えられるため、その単語を素性としてその他の漫画作品との比較する意味が無い。そのためdf値が2以上のものを素性とした。

1つ目のベクトル (V_1) は評価視点に着目することの有効性を確認するため、評価視点の出現回数の割合を素性の値とする。漫画作品 d の素性 k の出現回数を Cnt_{dk} とし素性 k の出現回数の割合の値 $Cratio_{dk}$ は、漫画作品 d のレビュー中の素性の全出現回数が N の場合、式(1)で定義する。

$$Cratio_{dk} = \frac{Cnt_{dk}}{\sum_{i=1}^N Cnt_{di}} \quad (1)$$

2つ目のベクトル (V_2) の素性の値は、評価視点の極性が作品推薦へ有効か否かを確認するため、評価視点とポジティブな評価表現との係り受けの割合とする。漫画作品 d の素性 k と係り受け関係にあるポジティブな評価表現の個数を Pos_{dk} 、ネガティブな評価表現の個数を Neg_{dk} として、素性の値 $Pratio_{dk}$ を式(2)と定義する。評価表現の極性の判別には日本語評価極性辞書(用言編, 名詞編) [4]を使用した。

$$Pratio_{dk} = \frac{Pos_{dk}}{Pos_{dk} + Neg_{dk}} \quad (2)$$

3.3 コサイン類似度の計算

作成したベクトルを用いて、ユーザが指定した漫画作品のベクトルとすべての漫画作品のコサイン類似度を求め、計算された値を基準とし推薦の可否を決定する。類似度が高いほど入力された作品と類似していると判断し、推薦作品として出力する。ベクトルの素性の個数を M としたとき、漫画作品 d_i と漫画作品 d_j の類似度 $sim(d_i, d_j)$ をコサイン類似度計算を用いて求める。1つ目の、評価視点の出現回数を素性の値としたベクトルを類似度計算に用いるとコサイン類似度は式(3)で表される [5]。

$$sim(d_i, d_j) = \frac{\sum_{k=1}^M Cratio_{ik} * Cratio_{jk}}{\sqrt{\sum_{k=1}^M Cratio_{ik}^2} * \sqrt{\sum_{k=1}^M Cratio_{jk}^2}} \quad (3)$$

4 評価と考察

4.1 実験

実験対象は、漫画レビュー.com¹、作品データベース²、Amazon.com³に投稿されている漫画作品レビューとす

¹漫画レビュー.com, <http://www.mannagareview.com/>

²作品データベース, <http://sakuhindb.com/>

³Amazon.com, <https://www.amazon.co.jp/>

る。各漫画作品150件から200件のレビューのテキストをコーパスとした。コーパスは5ジャンル計18作品とする。各作品についてそれぞれの作品についてその作品との類似度の平均より高いものを推薦、類似度の平均値より低い作品は推薦しない

本手法の有効性を確認するため、各作品ごとに推薦された作品が妥当か否かを3人の被験者により評価した。推薦が妥当か否かを評価した。

4.2 実験結果

評価視点の出現回数の割合を値にしたベクトル V_1 、評価視点の好評価の割合を値としたベクトル V_2 について、本手法の推薦結果の妥当性を評価した。評価において推薦が妥当と判断された割合を正解率とする(表2)。

表2: 推薦の正解率

	スポーツ	バトル	ギャグ	恋愛	S&G	全体
V_1	69%	72%	80%	85%	57%	72%
V_2	70%	72%	62%	63%	41%	63%

S&G: サスペンス, ギャンブル

V_1 で推薦した場合、全体で72%と高い正解率であった。 V_2 の場合は V_1 に比べ正解率は少し落ちてしまった。S&Gのジャンルに属する漫画作品は正解率が低い結果となった。S&G作品のレビューの特徴として、話の内容を細かく説明しているレビューが多い特徴がある。そのため評価表現と係り受け関係にある名詞として特徴的な語が抽出されにくくなり、各漫画作品との類似度に差が出なかったため正解率が低くなったと考えられる。 V_1 に関してはギャグ、恋愛のジャンルの正解率が高い結果となった。ギャグや恋愛は異なるジャンルにも物語の一部として描かれていることが多く、実験結果では異なるジャンルであってもギャグ要素や恋愛要素を含む作品を推薦できたため正解率が上がったと考えられる。

5 まとめ

本研究では、漫画レビュー中の評価情報に基づくことで、高い精度で漫画作品の推薦を行うことができた。 V_1 に関してギャグ、恋愛作品の正解率が高かったことから作品の推薦に対して評価視点との評価表現の係り受け関係に着目することの有効性を確認できた。

参考文献

- [1] 電通総研 情報メディア白書, 2012.
- [2] 村瀬尊好, 柘和佑, 安藤友晴, “マンガの概要に基づく作品推薦システム”, 第11回情報科学技術フォーラム, pp.319-325, 2012.
- [3] 平山拓央, 湯本高行, 新居学, 高橋豊, “属性評価モデルに基づく商品評価の抽出と提示”, DEIM Forum, F-2-5, pp.1-7, 2011.
- [4] 小林的ぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.2, pp.203-222, 2005.
- [5] 小柴健太郎, “語句の出現頻度に着目したアニメの自動推薦”, 東京都市大学卒業論文, 2016.