

小説作品に対する読者嗜好性の定量化

日大生産工 (学部) ○小林 千草 日大生産工 関 亜紀子

1. はじめに

今日、書籍の電子化が広がると共に書籍コンテンツの検索や推薦が行われるサービスの場が多く存在している。これらのサービスでは、ジャンルや著者などのメタ情報を用いた内容ベースフィルタリングや、同じ嗜好を持つユーザの履歴情報を用いた協調フィルタリングによって推薦が行われている。1) また、レビューを共有できるサービスも増えており、利用者は推薦されたコンテンツの概要や読者レビューなどの情報を選書の参考にしている。特に小説では新しいジャンルや著者の作品を読み始めたい場合に読者レビューが重視されていると考えられる。

そこで本研究では、小説の読者レビューに着目し、これらに含まれる「面白い」や「残念」などの感情表現を、その小説を読んだ際の満足感として定量化する方法を検討する。そして、ユーザが嗜好する小説には共通の満足感が存在するという仮定のもとで、コンテンツ推薦に活用する方法を検討する。

2. 作品における満足感の抽出

読者レビューに含まれる感情表現を用いて、それぞれの小説作品についての満足感を定量化する。分析手順を図1に示す。まず、各小説作品の読者レビューをテキストデータとして収集し、形態素解析を行い、品詞に分解する。次に、品詞の中から感情語を抽出し、感情ごとに出現頻度を集計する。その後、各感情の代表的な感情語を形態素解析で得られた感情語の中から選出し、代表語とする。



図1：作品満足感の分析手順

*喜

喜び or 面白い or 笑い or 幸せ or 楽しい

*哀

空しい or 泣く or 悲しい or 孤独 or 悲痛

図2：コーディングルールの一例

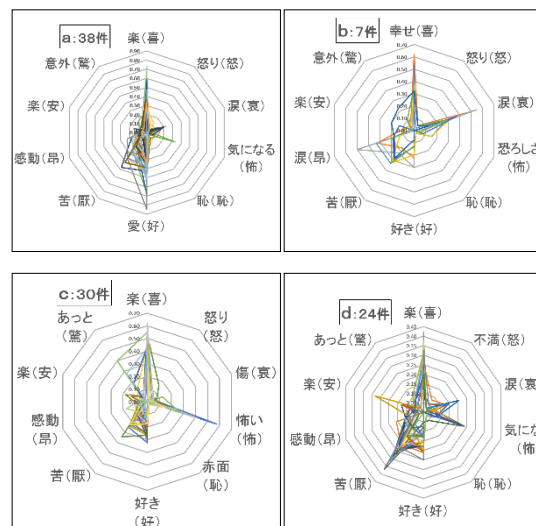


図3：「エンターテイメント」の例

分析用の読者レビューは、読書家によるレビュー交換が多数行われている読書メーター2)の投稿内容から、各作品40件ずつ用いる。収集した読者レビューは日本語の軽量解析用ツールであるKH Coderを用いて形態素解析を行う。抽出する感情語は感情表現辞典3)に掲載されている「喜」「怒」「哀」「怖」「恥」「好」「厭」「昂」「安」「驚」の全10種類とし、図2のようにコーディングルールを設定して各感情の出現頻度をクロス集計する。

「エンターテイメント」のジャンルに所属する99冊を対象にクロス集計した結果、全体で明確な特徴は得られなかった。そこで、クラスター分析により4つのグループに分類して再度分析を行った。各グループの代表語をラベルとしてグラフ化したのが図3である。この結果から、図3のaとdのように同じジャンルでも頻出する感情の分布に違いがあることが分かる。また、図3のcとdでは同じ感情の「怒」

Quantitative of Readers Preference for Novels

Chigusa KOBAYASHI and Akiko SEKI

でもそれぞれ表示された代表語に違いがあることが分かる。このことから、従来のジャンル検索では同ジャンルに属する作品の満足感の違いを区別できていないと考えられる。

3. 読者満足感の定量化

ユーザが小説作品を選別する上で重視する満足感を明らかにするため、各ユーザの既読かつ高評価の小説に対する読者レビューの感情表現を分析する。その後主成分分析を行うことでそのユーザの満足感を構成する成分を抽出する。

表1は、あるユーザが既読とした高評価の小説10件を用いたクロス集計結果である。その結果を用いてユーザの満足感を視覚化したのが図4である。また、表1の結果を主成分分析したのが図5である。

図4、図5の結果からこのユーザの嗜好に合う特徴的な満足感には「喜」「哀」「昂」「怖」の4つだと考えられる。また図5の結果から、第一主成分の高い値にはシリアスな作品が存在し、低い値にはハートフルな作品が存在する傾向がある。一方、第二主成分の高い値にはコミカルな作品が存在する。このことから、このユーザの好みは3つのグループに分けことが読み取れる。そこで、この得られたユーザの満足感の成分と作品における満足感の成分を考慮した小説を推薦することで、ユーザの満足感に合った小説が推薦されると考える。

表1：クロス集計結果

	*喜	*怒	*哀	*怖	*恥	*好	*厭	*昂	*安	*驚
夜は短し歩けよ乙女	0.50	0.00	0.03	0.03	0.08	0.33	0.18	0.05	0.00	0.00
旅猫リポート	0.50	0.03	0.68	0.00	0.00	0.30	0.23	0.53	0.03	0.03
ジェノサイド	0.30	0.05	0.10	0.08	0.00	0.13	0.10	0.18	0.05	0.10
六番目の小夜子	0.30	0.00	0.00	0.20	0.03	0.20	0.13	0.15	0.05	0.05
ガリレオの苦悩	0.53	0.03	0.05	0.05	0.00	0.20	0.28	0.03	0.05	0.03
陰日向に咲く	0.44	0.00	0.08	0.00	0.00	0.13	0.13	0.18	0.00	0.13
フライ・ダディ・フライ	0.68	0.00	0.08	0.15	0.00	0.30	0.10	0.13	0.03	0.03
親指さがし	0.14	0.00	0.00	0.42	0.00	0.17	0.06	0.03	0.00	0.06
クレイドゥ・ザ・スカイ	0.18	0.00	0.08	0.03	0.00	0.05	0.18	0.08	0.05	0.03
ネバーランド	0.50	0.00	0.13	0.13	0.03	0.25	0.35	0.08	0.08	0.00

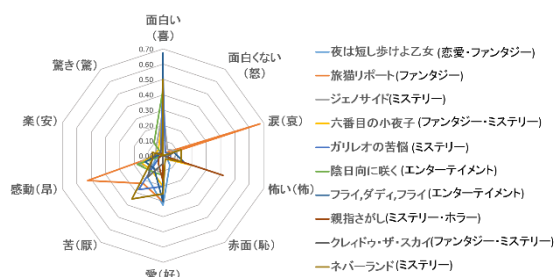


図4：既読かつ高評価の小説10件分析結果

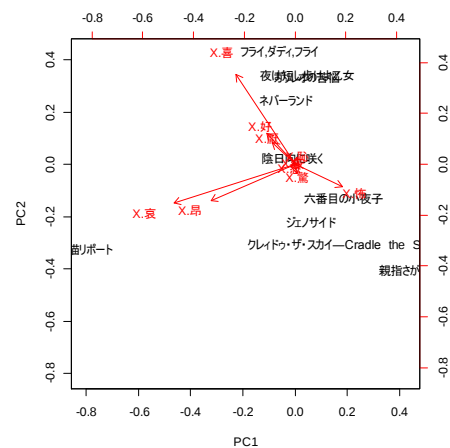


図5：主成分分析結果

4. 実験方法および評価方法

提案手法を用いたコンテンツ推薦の有効性を検証するために、既存システムとの比較実験を行う。実験には本研究で推薦された小説作品とAmazonで推薦された小説作品とを比較する。

本研究で推薦する作品とAmazonで作品を選んだ際に推薦される作品についてあらずじとレビューを読み、「興味あり」か「興味なし」かを判断し、どちらがより多く自身の嗜好に合った作品が推薦されたかを評価する。

5. まとめと今後の課題

本稿では、ジャンルごとの小説作品における満足感の定量化と読者における満足感の定量化の分析を行った。これにより、従来のジャンル検索では同じジャンル内の作品において満足感の違いを区別できない可能性があることがわかった。また、読者満足感における分析ではあるユーザによる嗜好の特徴的な感情を抽出することができた。

今後は、多くのユーザで実験を試み、その結果を元に目的であるジャンルや著者を問わずユーザの嗜好に合った小説作品を推薦するシステムの実現を目指す。

参考文献

- 1) 蔣淵舒, 信頼度を考慮した協調フィルタリングに基づく書籍の推薦方式の提案, 第75回全国大会講演論文集 (2013) p. 675 – 677
- 2) 「読書メーター」
<http://book.akahoshitakuya.com/>
(2014/9/23 アクセス)
- 3) 中村明, 感情表現辞典, 東京堂出版(1993)