

表紙と内容の関連に基づく書籍推薦システム

親泊広直^{†1} 井村誠孝^{†2} 岸野文郎^{†3}

近年、電子書籍の普及に伴い、表紙デザインの印象の重要性が高まっており、表紙の好みに基づいて書籍を推薦できれば書籍を探す際の新たな指標になると考えられる。本研究では、ユーザの表紙に対する好みを利用して、類似の読書嗜好を持つ他のユーザを発見し、ユーザが好むと推測される書籍を推薦する手法を提案する。また、表紙画像と書籍内容の双方から得られる特徴量の関係を解析し、提案手法により推薦可能な書籍の条件について検討する。

Book recommendation system based on the relevant cover and contents

HIROMASA OYADOMARI^{†1} MASATAKA IMURA^{†2}
FUMIO KISHINO^{†3}

In recent years, with the spread of electronic books, the impression of cover design becomes more important. The recommendation of books based on the preference of the cover design is considered as one of the new tools for selection of the book. In this research, we propose a method for recommending a book presumed to be liked by users. The proposed system finds a reviewer which has the same preference as the user and recommends the book on which the reviewer set a high valuation. Furthermore, we analyze the relationship between features obtained from the cover image and ones from book content and investigate the conditions of the possible recommendation books by the proposed system.

1. はじめに

近年, iPad などの電子書籍閲覧端末の急速な普及により, 電子書籍コンテンツ市場が拡大している [1]. 電子書籍は書店のように実際に手にとって選ぶことができないため, 表紙デザインの印象によって書籍を購入する「ジャケ買い」が注目されており, 表紙デザインの変更によって売り上げ成果を出している事例もある [2]. 購入する書籍の選択において, 書籍の表紙デザインの重要性が高まっていると考えられ, 表紙の好みに基づいて書籍を推薦できれば, ユーザが書籍を探す際の新たな指標になると考えられる. 書籍の内容が, 購入者が表紙から予想した内容に反していても, 好みには合うこともあり, ユーザにとって読書の新たな境地を開ける可能性がある.

本研究では, 表紙の好みを利用して同じ嗜好を持つユーザを発掘し, ユーザの好みに合った書籍を推薦する手法を提案する. ユーザは表紙の見た目と書籍の好みを入力し, システムはユーザと好みが似ている他のレビュアーを探索する. 探索により発見された好みの合うレビュアーが高く評価した書籍を推薦する. これまでに筆者らが, 有効性を検証するため評価実験を行った結果, 本システムによってユーザにとって読みたい書籍を推薦できる可能性が示唆されている [3]. 本稿では, 推薦された書籍を実際に読んで評価してもらうことで更に詳細なシステムの評価実験を行

った結果を報告する. また, 推薦する書籍は内容を読んで評価されたデータを基にしているが, 表紙の好みとの関連性は検証されていない. 表紙画像データと本文のテキストデータの関係を解析することで書籍と表紙の関連性について検討する.

2. 関連研究

これまでにも, 書籍の雰囲気や印象に基づいて, ユーザに書籍を推薦するシステムの研究が行われている. 新宅らは, ユーザの状況や気分に適した書籍を提示することを目指して書籍の文字数や本文中のセリフの割合などの書籍本文の固有データから書籍の雰囲気や印象の判定を行っている [4]. 書籍の雰囲気や印象の中で最も基本的なものである明るさ・暗さを対象に, 様々なジャンルの書籍から明るいイメージ (すばらしい, 満ち足りた, 軽やかな等) ・暗いイメージ (陰しい, 災難, 恨む等) の単語を抜粋し, 書籍の本文に含まれる単語の一致数で書籍の明るさ・暗さの判定を行っている. また, 垣内らは書誌情報のひとつである「あらすじ」から感性情報の抽出を行い, 書籍の印象を定量化する手法を提案している [5]. 萱田らは, レビューからユーザが書籍に対して感じた印象を抽出し類似性を比較することにより, 書籍の内容に依存しない推薦手法を提案している [6]. 梶山らは, 書籍の表紙が購入の大きな指標になっていると考え, 読者の意見を反映させた電子書籍表紙画像の自動生成を目指し, 書籍内容や感想文のテキストから表紙色を抽出する手法を提案している [7].

上記のように, 書籍推薦手法の研究は数多くされている. しかし, 書籍の雰囲気や印象の定量化にとどまっ

^{†1} 関西学院大学大学院 理工学研究科
Graduate School of Science and Technology, Kwansei Gakuin University

^{†2} 関西学院大学 理工学部
School of Science and Technology, Kwansei Gakuin University

^{†3} 関西学院大学 感性価値創造研究センター
Research Center for Kansei Value Creation, Kwansei Gakuin University

おり、ユーザの好みや気分に合わせて推薦をするまでには至っておらず、発展途上にあるといえる。

3. 書籍の表紙を利用した書籍推薦システム

本研究では、表紙を見て選択した書籍を高く評価しているレビュアー群を探索し、そのレビュアー群が読んで満足した書籍を推薦するシステムを構築する。概要を図 1 に示す。

本システムの推薦手順は以下の通りである。まず、画面上に書籍の表紙のみを 10 冊提示する。次に、ユーザは画面上に表紙のみが表示された 10 冊の中から、その時の気分で好みの書籍を 3 冊選んでもらい、それぞれにどの程度気に入ったか・読みたいかなどで 1~5 の評価値をつけてもらう。その後、その 3 冊にユーザと同じ評価を与えている他のレビュアー群をレビュー共有サイトから探索する。具体的には、サイト上で、ユーザが選択した 3 冊全てに、ユーザと同じ評価値をつけているレビュアー群を抽出する。抽出されたレビュアー群が評価 5 をつけている書籍を全て抽出する。抽出した書籍の中で重複する書籍があれば、それら全てを推薦候補とする。なお、重複する書籍がなければ、抽出されたレビュアーのひとりでも 5 点を付けている書籍は全て推薦候補とする。そして、システムはそれらの推薦候補内からランダムに 1 冊の書籍を選択し、ユーザに推薦する。

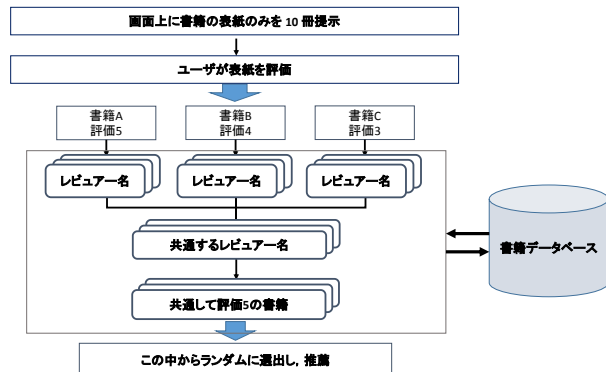


図 1 システムの推薦手順

4. 実装

本節では提案するシステムの実装について述べる。本システムでは、ウェブ上の書籍レビュー共有サイト「ブクログ [8]」を用いて、同じ嗜好を持つレビュアーを発掘する。ユーザに対してははじめに提示する書籍は、ブクログの 2009 年からの 5 年分のユーザの登録数の年間ランキング上位 5 冊から、重複しているものを除いた計 19 冊の中からランダムに選択する。本システムでは、レビューの件数が少ない書籍は、好みの類似したレビュアーの探索に支障をきたすため、レビューの件数が多いランキング上位の書籍を用いており、最大レビュー数 3756 件、最小 621 件となっ

ている。

本システムの使用における実験参加者の操作手順は以下の通りである。図 2 のように、提示された書籍の下にそれぞれチェックボックスと 1~5 の数値を選択できるセレクトボックスがあり、気に入った書籍 3 冊にチェックを付け、それぞれにどの程度気に入ったのか 1~5 の評価を付けてもらう。そして、下部の「決定」ボタンをクリックすることで、レビュアーの探索を開始する。その際、3 冊選択されていないなどの不備があれば、「3 冊チェックしてください」とメッセージが出現し、次の作業に進めないようになっている。推薦する書籍が決定すると、図 3 のように新たにウィンドウが開き、その書籍の表紙画像が提示される。

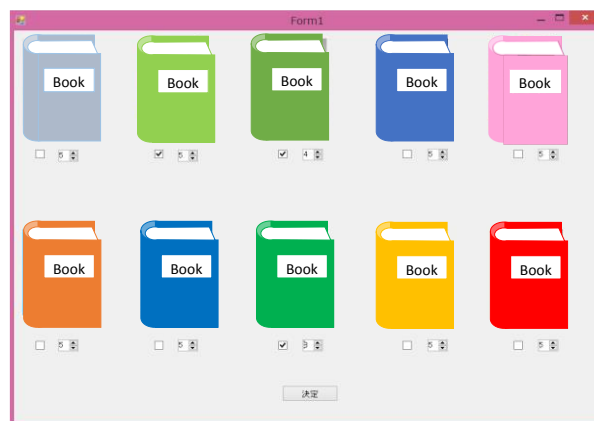


図 2 入力画面



図 3 推薦画面

5. システム評価実験

提案手法の妥当性を評価するため、システムにより推薦された書籍を実際にユーザに読んでもらう実験を行った。まず、推薦システムにより実験参加者の好みに合うとされる書籍を推薦し、推薦された書籍の表紙のみから判断される意外性の有無などについてアンケートを行う。次に、表紙とあらずじから判断される評価も回答してもらう。数日後に実験参加者に書籍を与え、任意のペースで読んでもらい、読了後にアンケートを行う。アンケートは、「1. 全

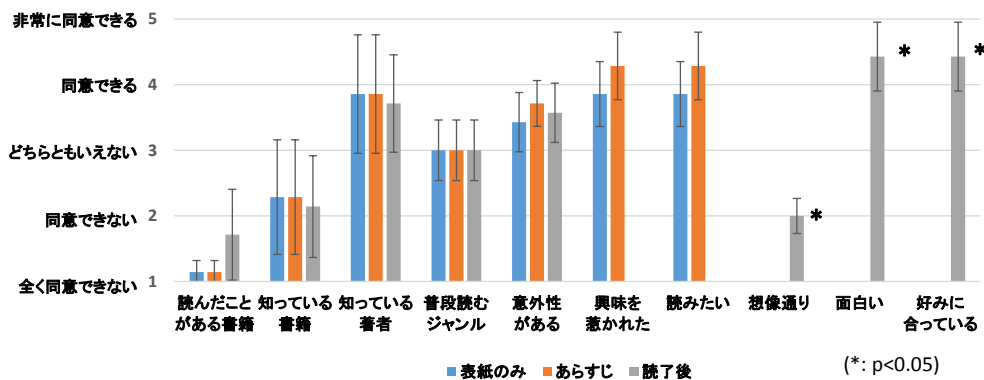


図4 アンケート結果

く同意できない」から「5. 非常に同意できる」の5段階評価とした。実験参加者は、22～24歳の男性7名である。

アンケート結果を図4に示す。図4より、「意外性がある」、「興味を惹かれた」、「読みたい」の項目において、「表紙のみ」と「あらすじ」の条件間で大きな違いが見られないことから、表紙とあらすじから受ける印象には関連性があると考えられ、主観評価では表紙と内容が関連している可能性が示唆された。また、読了後の「想像通り」の平均値は2.0(標準偏差0.53)と低い一方で、「面白い」と「好みに合っていた」の平均値は4.43(標準偏差1.05)と高評価であった。ピアソンのカイ二乗検定適合度検定の結果、共にp値は0.02となり有意差が見られた。このことから、意外性がありつつもユーザが読んで満足できるような推薦ができていると考えられ、表紙の好みによってユーザが満足できる書籍を推薦できる可能性が示唆された。

6. 表紙と内容の関連性検証実験

提案システムでは、推薦する書籍は内容を読んで評価されたデータを基にしているが、書籍の内容と表紙の好みとの関連性は検証されていないため、表紙の画像処理と書籍の内容のテキストマイニングに基づき、関係を調査する。表紙画像の特徴量として取得するデータは、表紙に使用されている文字の大きさや色の割合とした。書籍の内容としては、書籍の本文の内容が要約されているあらすじを使用した。本節では、表紙の色割合とあらすじのテキストデータの関連性について検証した結果を報告する。

6.1 色割合解析

表紙に含まれる色の割合を表紙色割合とし、あらすじのテキストマイニングとの相関関係を調査する。表紙色割合の抽出方法は、画像を各成分8bitに減色加工後、各色の割合を測定し、有彩色10色(赤, 橙, 黄, 黄緑, 緑, 青緑, 青, 青紫, 紫, 赤紫), 無彩色3色(白, 灰, 黒)の計13色に分類した。13色は日本カラーデザイン研究所により提案されたカラーイメージスケール[9]を参照した。カラーイメージスケールは、各色をトーン分類し代表色130色に

細分化しており、色の分類にはこれを利用する。代表色とは配色イメージの世界を代表する単色を意味する。色割合の測定は、以下の手順により行った。

- (1) 代表色の全色に対し、出現回数を0に設定する。
 - (2) 表紙画像のあるピクセルに対し、RGB値を算出する。
 - (3) (2)で算出されたRGB値の出現回数を1加算する。
 - (4) 全てのピクセルが走査されるまで、(2)～(4)を繰り返す。
 - (5) 出現した全てのRGB値に対し、代表色130色に一番距離の近い色を選択し、出現回数を合算する。以下の式を用いて代表色*i*との距離 L_i を決定する。
- $$L_i = \sqrt{(r - R_i)^2 + (g - G_i)^2 + (b - B_i)^2}$$
- r, g, b は算出されたRGB値を示し、 R_i, G_i, B_i は代表色*i*のRGB値を示す。出現した全てのRGB値に対し、距離を求め、距離が最小の代表色を含む色を距離の近い色とする。
- (6) 各色の出現回数を表紙画像の全ピクセル数で除した値を、表紙画像に含まれる色の割合とする。

あらすじのテキストマイニングは、抽出したあらすじを形態素解析し、13色の印象語とマッチングすることで行う。印象語は、カラーイメージスケールで定義されている形容詞180語に対し、日本色彩研究所による色彩連想語調査資料[10]を活用して、786語の単語に拡張したものである。色彩連想語は、赤であれば「リンゴ」、「熱い」のように、色から連想する言葉や事物を挙げたものである。

調査対象書籍は2009年からの6年分のユーザのお気に入り登録数の年間ランキング上位10冊から重複しているものを除く、計47冊とする。

表紙色割合とあらすじのテキストマイニングで得られた色の割合に対して、ピアソンの相関係数を算出した。47冊中、強い相関(0.4以上)を示したものが12冊あった。強い相関を示した書籍5冊と弱い相関を示した書籍5冊のあらすじと表紙上位3色と割合、相関係数を表1に示す。この結果より、書籍によっては関連があると思われる

表1 色割合解析結果の例

	タイトル	あらすじ (%)			表紙 (%)			相関係数
		1	2	3	1	2	3	
強い 相関	植物図鑑	青 3.8	紫 3.8	赤 1.9	白 35.9	黒 30.7	灰 18.5	-0.440
	もし高校野球の女子マネージャーがドラッカーの『マネジメント』を読んだら	青 2.3	赤 1.2	黒 1.2	青 39.0	緑 16.3	青紫 14.8	0.512
	天地明察	赤 3.0	黒 3.0		黒 51.6	緑 31.7	灰 13.0	0.481
	スティーブ・ジョブズ I	白 1.7	灰 0.8		白 60.7	黒 17.8	灰 12.3	0.903
	少女	黒 3.7	赤 1.9		黒 38.7	黄 28.6	灰 13.6	0.697
弱い 相関	1Q84 BOOK 1	黄 1.9	灰 1.5	黒 1.2	白 71.5	緑 10.6	黄 9.2	0.009
	恋文の技術	赤 1.5	青 1.5		白 65.1	赤 14.6	灰 9.9	-0.046
	あるキング	黒 1.6	赤 0.8	青 0.8	白 94.8	黒 2.5	灰 1.3	0.238
	これからの「正義」の話をしよう ——いまを生き延びるための哲学	赤 0.3	灰 0.3		白 79.9	灰 9.1	黒 9.1	-0.188
	20歳のときに知っておきたかったこと スタンフォード大学集中講義	青 1.4	黄 0.7	緑 0.7	灰 71.1	白 21.0	青 7.9	-0.141

ことがわかる。

強い相関の書籍に共通する特徴を発見するため、表紙に最も使用されている色があらすじ内の印象語に最も出現している色と一致している確率を正答率とし、強い相関を示した書籍とそれ以外の書籍でそれぞれ正答率を求めた。強い相関の書籍では 80%、それ以外の書籍では 0%となり、正答率に大きな違いがあることが分かった。このことから、使用する書籍によっては表紙と内容の関連性がある可能性が示唆された。表紙に最も使用されている色が、あらすじ内の印象語に最も出現している色である書籍のみを提案システムではじめに提示する書籍とすれば、表紙と内容の関連性がある推薦ができると考えられる。また、今回用いた調査方法以外の指標によって新たな関連性を発見できる可能性もあるため、検証の余地がある。

7. おわりに

本稿では、表紙と内容の関連性に基づく書籍推薦システムを提案し、構築した推薦システムを用いて具体的なシステムの評価実験と提案手法により推薦可能な書籍の条件についての検証を行った。評価実験より、意外性がありつつもユーザが読んで満足できるような推薦ができたと考えられ、表紙の好みによってユーザが満足できる書籍を推薦できる可能性が示唆された。また、書籍の表紙と内容の関連性について検証するため、表紙色割合とあらすじのテキストマイニングにより得られた色の割合との相関関係を調査した。その結果、表紙と内容の関連性がある可能性が示唆されたことから、表紙に最も使用されている色が、あらすじ内の印象語に最も出現している色である書籍のみを提案システムではじめに提示する書籍とすれば、表紙と内容の関連性がある推薦ができると考えられる。

今後は、表紙と内容の関連性の更なる調査として文字色

やフォント、デザインのシンプルさなど表紙に関する他の特徴を用いることで、新たな関連性の発見を試みる。その後、大規模な調査を目指すため、Amazon API などを用いて必要なデータの収集を行えるウェブサイトを構築していく。

参考文献

- [1] ICT 総研: 2014 年度 電子書籍コンテンツ市場動向調査, <http://www.ictr.co.jp/report/20141015000069.html>
- [2] 朝日新聞: 文庫本「ジャケ買い」『人間失格』は『デスクノート』風, http://www.asahi.com/culture/news_culture/TKY200709180102.html
- [3] 親泊 広直, 岸野 文郎: 表紙情報の好みに基づく書籍推薦と読了後の評価, 電子情報通信学会 2015 年総合大会講演論文集, A-15-23 (2015).
- [4] 新宅 彩加, 岸野 文郎, 石原 のぞみ, 中島 康祐, 藤田 和之, 伊藤 雄一: 書籍固有の情報を用いた書籍の明るさ判定, 電子情報通信学会総合大会講演論文集, p. 268 (2012).
- [5] 垣内 将希, 高岡 幸一, 灘本 明代: 感性パラメータを用いた書誌情報からの図書推薦手法の提案, 第 4 回データ工学と情報マネジメントに関するフォーラム, A9-4 (2012).
- [6] 萱田 真史, 村井 聡一, 牛尼 剛聡: レビューを利用した読者の印象に基づく書籍推薦手法, 第 4 回データ工学と情報マネジメントに関するフォーラム, A5-1 (2012).
- [7] 梶山 朋子, 越前 功: 読者の印象の色彩化による表紙画像生成の評価, 電子情報通信学会技術研究報告. IE, 画像工学 Vol.112, No.189, pp.75-79 (2012).
- [8] 株式会社ブックログ: ブックログ, <http://booklog.jp/>
- [9] 小林 重順: カラーイメージスケール改訂版, 講談社 (2001).
- [10] 日本色彩研究所: バックナンバー, <http://www.jcri.jp/JCRI/hiroba/COLOR/color-backnumber.htm>