

書籍タイトルと表紙画像を曖昧化し 興味を誘発するブックカバー作成システム

飯塚里帆^{†1} 中山祐之介^{†1} 橋田朋子^{†1}

概要：本研究では書籍のバーコードからタイトルを取得し、形態素解析と上位概念辞書 API によって名詞を抽象化した書籍タイトルを作成する。同時に表紙画像のメインカラーを抽出し、抽象化した書籍タイトルと組み合わせることで書籍情報を曖昧化したブックカバーを自動で作成するシステムを提案する。書店において、作品情報を見せない代わりに推薦文が印刷されたカバーをかけた状態で書籍を陳列する事例に筆者らは着目した。本稿では関連事例と具体的なシステム実装方法について報告する。

1. はじめに

筆者らは物事の曖昧化によりユーザの興味を喚起する可能性に注目している。例えば、物品の包装はその物品が何であるかを様々なレベルで曖昧にする仕組みであり、中身の解釈や受け取り方を変化させることができる。さらに最近では書籍や DVD などに関しても具体的なタイトルを隠して人手でキャッチコピーや推薦文のみを付与し、作品情報をやや曖昧化させるような試みも行われ、好評を博している [1][2]。書籍では一般的に表紙に印刷した書籍情報や装丁によって内容を印象的に伝えようとするが、一方で情報が過度に伝わることで読む前から内容が理解され書籍を読まない・買わないといった課題もあり、このような曖昧化の試みは今後需要が見込まれる。

以上のような背景を受け、本研究では、特に文庫本・新書の書籍情報と装丁を自動で曖昧にする仕組みを提案する。本システムは2つの機能からなる。1つ目は、日本語 WordNet [3] という概念辞書を用いることによりタイトルの一部を上位概念の単語に変換する機能であり、2つ目は、ComputerVision により装丁デザインの特徴的な色を抽出し、単純な画像を生成する機能である。これにより、図1に示すような書籍の本来のタイトル、装丁を大きく崩さずに自動で書籍の表紙を曖昧にすることが可能なシステムを実装する。本稿では本研究の関連事例、システムの実装について以下に述べる。



図1 (左)元の書籍(森絵都, アーモンド入りチョコレートワルツ, 角川書店, 2006.)
(右)本システムを用いて作成したカバーをかけた書籍

2. 関連研究

2.1 商品のパッケージを曖昧にする事例

商品パッケージを敢えて曖昧な表記にする事例として、NOT ジャケ借 [2]は作品を鑑賞した著名人がつけたキャッチコピーのみが見える状態でDVDを陳列する事例である。これらは推薦者自身が推薦文やキャッチコピーを考案することでDVDの情報を曖昧にしたものである。商品パッケージの中でもブックカバーに関するものとして、文庫X [1]は書店員の推薦文と値段のみが印刷されたブックカバーをかけて販売することで知名度を大幅に高めることに成功した事例である。さらにFUMIKURA [4]は読者が作品情報を入力し統一されたデザインのオリジナルブックカバーを印刷できるサービスである。これらの先行事例に対し、本研究はタイトルを概念辞書による自動抽象化処理によって抽象化する点や、オリジナルの書籍の装丁の痕跡を残しつつ抽象化するという点が異なる。

2.2 日本語 WordNet を用いた研究

日本語 WordNet [3]を用いた研究は以下のものが挙げられる。伊藤ら [5]は日本語 WordNet の上位下位概念を用いて単語を抽象化し、二文の関係性を推定している。眞田ら [6]は直上概念を決定する際に日本語 WordNet を利用している。本研究は書籍を抽象化する目的で利用している点で異なる。

3. システム

3.1 システム構成

本研究で提案するシステムの構成・処理を図2に示す。入力部に Web カメラ、バーコードリーダー (ビジコム BC-BR900L-W) を使用する。出力にはプリンタ (Canon PIXUS MG7530) を使用する。提案システムは書籍タイトルを抽象化する仕組みと表紙画像を抽象化する仕組みの2つによって構成されている。ISBN コードからのタイトル取得には、国立国会図書館サーチの API を使用する。タイトルの抽象化のために、Mecab による形態素解析、WordNet 上位概念辞書 API を使用する。表紙画像の抽象化には、K-means 法

^{†1} 早稲田大学
Waseda University

によるクラスタリングによって特徴色の抽出を行う。処理システムは Python3.7 を用いて実装する。

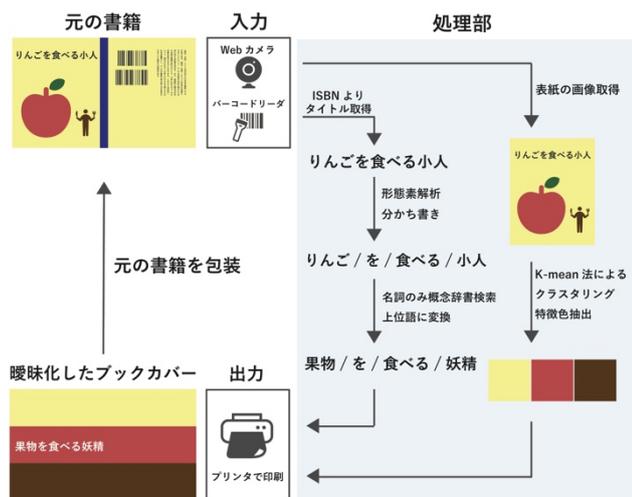


図2 システム構成・処理図

3.2 システム処理

3.2.1 入力部

バーコードから ISBN-13 を読み取り、国立国会図書館サービスの API に入力することで書籍のタイトルを取得する。表紙画像は Web カメラから取得する。

3.2.2 抽象化処理

表紙の抽象化処理はタイトルの抽象化と装丁デザインの抽象化によって構成される。

(1)タイトルの抽象化

取得したタイトルを Mecab で形態素解析する。解析した形態素から名詞のみを WordNet 上位概念辞書 API に入力する。上位概念がある名詞を 1 段階上の概念を出力することで抽象化されたタイトルを取得する。

(2)装丁デザインの抽象化

取得した画像から色のパターンを K-means 法によってクラスタリングし、3 色の RGB 値を抽出する。抽出した色を文庫本カバーサイズの画像に等分に配置する。最終的に抽象化されたタイトルと、抽象化された画像を合成する。

3.2.3 出力部

合成したタイトルと画像を文庫本のカバーサイズに合わせた A4 サイズの紙に印刷する。完成したブックカバーを元の書籍にかける。

3.3 動作確認

実際に提案システムを用いて市販の書籍のタイトルと表紙を抽象化したブックカバーの例を図3に示す。図3ではタイトルに関して、『缶詰ラベルコレクション』は『食品符帳集合』に変換された。『夜の朝顔』は『夜のつる植物』に変換された。表紙画像に関して、『缶詰ラベルコレクション』は代表的な #d97cdb, #3f1f3e, #4d67d4 の3色で構成された。『夜の朝顔』は代表的な #dcd9ce, #575c34, #bea981 の3色で構成された。

さらに暖昧化したカバーをかけた書籍を数名に見せたと

ころ「曖昧になっている」「意味が理解できる」「この書籍を読みたいと感じる」「作者名は表示していてもよい」という意見が得られた。これによって、書籍タイトルと表紙の曖昧化によってユーザの書籍への興味の喚起ができることが示唆された。



図3 (左)元の書籍(1枚目：公益社団法人日本缶詰協会、缶詰ラベルコレクション MADE IN JAPAN, 青幻舎, 2012. 3枚目：豊島ミホ、夜の朝顔, 集英社, 2009.) (右)抽象化したブックカバーをかけた書籍

4. まとめと今後の展望

本研究ではバーコードとカメラ画像から書籍タイトルと表紙画像を取得して抽象化したブックカバーを自動作成するシステムの提案と開発を行った。

今後の展望としては、大きく2つ挙げられる。1点目はタイトル変換システムの改良である。現段階では語尾活用を変換する必要のない名詞のみを変換する方式を採用しているが、自然な表現を維持したまま名詞以外の品詞を抽象化するシステムを構成することを目指す。もう1点は抽象化された表紙の装丁の改良である。表紙だけでなく背表紙にも抽象化された書籍タイトルを印字することで本棚に書籍をしまっている際にも抽象化された書籍タイトルが見える状態を目指す。また、文庫本や新書サイズにとどまらず、あらゆるサイズの書籍に対応できるブックカバー作成システムの開発を実現したい。

参考文献

- [1] 文庫 X. <https://www.shinchosha.co.jp/news/article/303/> (参照：2018-12-10).
- [2] NOT ジャケ借. <https://tsutaya.tsite.jp/feature/movie/campaign/not-jyakegari/index> (参照：2018-12-10).
- [3] Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi and Kyoko Kanzaki: Enhancing the Japanese WordNet. *The 7th Workshop on Asian Language Resources, in conjunction with ACL-IJCNLP 2009*, Singapore.
- [4] FUMIKURA. <http://www.usagi-g.net/fumikura/index.php> (参照：2018-12-10).
- [5] 伊藤直貴, HugoHernault, 石塚満. WordNetを用いた機械学習による談話構造解析. 情報処理学会第73回全国大会講演論文集, 2011, vol.1, p383-384.
- [6] 眞田和枝, 塩井隆円, 波多野賢治: 概念辞書グラフの経路に基づく語の直上概念決定法. 情報処理学会第77回全国大会講演論文集, 2015, vol.1, p.639-640.
- [7] George A. Miller. WordNet: A Lexical Database for English. In *Communications of the ACM*, Vol. 38, pp. 39-41, 195.