

EPUB 小説データからの表紙の自動生成

川口 晴会[†] 長尾 悠真^{††} 鈴木 伸崇^{†††}

† 筑波大学 情報学群 知識情報・図書館学類 〒305-8550 茨城県つくば市春日 1-2
 †† 筑波大学 大学院図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2
 ††† 筑波大学 図書館情報メディア系 〒305-8550 茨城県つくば市春日 1-2
 E-mail: †{hkawaguchi,ynagao}@klis.tsukuba.ac.jp, ††nsuzuki@slis.tsukuba.ac.jp

あらまし 本研究では選書の支援となると考え、EPUB形式の小説データからその内容の印象に沿った表紙を自動的に生成するシステムを提案する。表紙を構成する要素のうち色とフォントに着目し、小説の本文を解析してその印象に沿った色とフォントを求める。色については色と単語を対応させた色彩データベースを用いて推薦を行い、形容詞と色が対応付けられた色彩データベースを、Word2Vecを用いて動詞と名詞も対象となるように拡張した。フォントについては各フォントと書籍本文の印象を表した感性ベクトルを使用し、コサイン類似度が高いフォントを推薦する。フォントの画像と感性ベクトルの対を教師データとして機械学習を行い、フォントから感性ベクトルの推定を行う。本文テキストの感性ベクトルは形容詞対に類似した単語の出現頻度によって推定を行う。

キーワード 電子書籍, 色抽出, フォント推薦

1. はじめに

電子書籍と電子書籍リーダーの普及により、近年多くの電子書籍が利用可能となっている。それと同時に、テキストファイルを電子書籍のフォーマットに変換するツールが数多く登場し、任意のテキストを電子書籍として作成し読むことができるようになった。例えば、著作権の消滅した作品を提供する青空文庫[1]の作品や、オンライン小説をダウンロードして電子書籍に加工することが考えられる。また、小説投稿サイトなどの登場によって気軽に小説を公開できる環境になっており、作者自身が電子書籍として配布することも可能である。このような電子書籍の作成における主な問題として、多くの場合に適切な表紙が与えられていないことがある。表紙は書籍の印象を読者に伝える重要な要素であり、選書の際にも有用である[2]。しかし、書籍の内容に合った表紙を作成するには手間やコストがかかり、特に未読の小説を電子書籍に加工する場合、内容を反映した表紙を作成することは極めて困難である。

そこで、本研究では電子書籍の主要なフォーマットであるEPUB(Electronic PUBlication)[3]を対象とし、EPUB形式の小説データからその内容の印象に沿った表紙を自動的に生成するシステムを提案する。表紙を構成する要素のうち色とフォントに着目し、小説の本文を解析してその印象に沿った色とフォントを求める。色については色と単語を対応させた色彩データベースを用い、フォントについては各フォントと書籍本文の印象を表した感性ベクトルを使用して推薦を行う。得られた色を表紙の背景色とし、得られたフォントでタイトルや著者名を描画する。

小説の印象を伝えるために色に着目した先行研究では、書籍の本文とレビューの形容詞から抽出した色を表紙として画像を生成したもの[2]や、書籍本文のうち情景描写のみを対象として色抽出を行っているもの[4]が存在する。文献[2]は入力にレ

ビューも用いているが、その選択基準については考察されていない。しかし、実際に購入サイトなどで書かれているレビューは一個人によって書かれているため内容の偏った意見や感想が書かれている可能性があり、そもそもレビューがない書籍も少なからず存在する。また、形容詞のみを対象として色を抽出しているが、その他の単語も色と結びつく可能性があり小説全体の印象とは言い難い。一方、本研究では、書籍本文のみを入力として形容詞に加え名詞と動詞からも印象を抽出する。より多くの単語を色と関連付け、網羅性の高い色抽出を目指す。文献[4]は色抽出の対象となる単語数が少ないため、入力に用いる本文が短い場合に一致する単語が存在しない可能性があるという欠点がある。フォントの印象に関する先行研究としては、入力されたテキストに合うフォントを推薦するもの[5]や、フォントを印象に基づいて分類するもの[6]がある。文献[5]は被験者実験のアンケートに用いられたフォントのみが推薦の対象となっている。また、[6]は、フォントから抽出した太さや文字間隔のような特徴量を機械学習にかけてフォントを分類している。一方、本研究ではDeep Learningを用いることにより、利用者の所持する全ての日本語フォントが推薦の対象となり、人手で特徴量を抽出することなく未知のフォントに対しても感性ベクトルの推定が行える。

2. 諸定義

本章では、EPUBの概要、および提案手法で用いるWord2VecとCNNについて述べる。

2.1 EPUBの概要

EPUBは国際電子出版フォーラム(International Digital Publishing Forum)が策定したフォーマットである。2011年にEPUB3が最終勧告され、2016年現在での最新のバージョンとなっている。EPUBの主な特徴として(1)XMLやHTML、CSSなどの広く普及したWebの標準技術がベースとなってい

ること, (2) オープンで利用料が発生しないフリーなフォーマットであること, (3) 多様な閲覧環境を前提にしたアクセシブルなフォーマットであること, の3点が挙げられる. そのため, 今後も主要なフォーマットであり続ける可能性が高い [7]. EPUB のファイル構造の例を図1に示す.

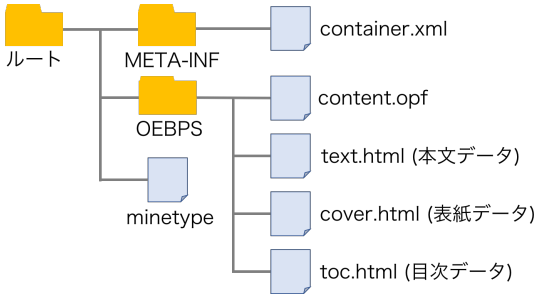


図 1: EPUB のファイル構造の例

メディアタイプを記述する minitype ファイル, メタ情報を格納する META-INF フォルダ, 本文や表紙などの書籍の内容を格納する OEBPS フォルダからなる. EPUB は, このようなファイルを zip でアーカイブしたものである. 提案手法では, OEBPS フォルダ内の本文テキストを解析し, 本文の印象に沿った表紙の HTML ファイルを生成する.

2.2 ニューラルネットワークの概要

Word2Vec と CNN で用いられるニューラルネットワークについて述べる.

ニューラルネットワークとは, 高度な情報処理のために, 脳内のニューロン (神経細胞) の働きをコンピュータ上で実現しようとしたものである. Word2Vec と CNN では階層的なニューラルネットワークが用いられている. 基本的な構造を図2に示す. ネットワークは入力層と中間層, 出力層からなり, CNN

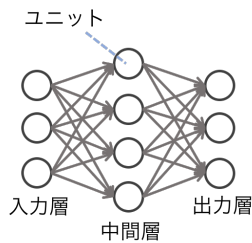


図 2: 階層的なニューラルネットワークの例

では中間層は任意の数になる. 中間層のユニットは, 前の層の複数のユニットから入力した情報に対し何らかの処理を行い, 次の層のユニットに情報を出力する. 第 n 層のユニットを $u_i (i = 1, 2, \dots, I)$, 第 $n+1$ 層のユニットを $u_j (j = 1, 2, \dots, J)$ とする. b_j をバイアス, ユニット u_i から u_j の重みを w_{ji} とする. ユニット u_i の出力値を式 (1) のように定義する. $n = 1$ のとき u_i は入力層のユニットであり, x_i は入力ベクトルの値である.

$$\begin{cases} x_i & (n = 1) \\ z_i & (n > 1) \end{cases} \quad (1)$$

ユニット u_j における処理は式 (2) のようになる.

$$S_j = \begin{cases} b_j + \sum_{i=1}^I w_{ji}x_i & (n = 1) \\ b_j + \sum_{i=1}^I w_{ji}z_i & (n > 1) \end{cases} \quad (2)$$

この S_j をもとに, ユニット u_j の出力値は活性化関数 f を用いて式 (3) のように表される.

$$z_j = f(S_j) \quad (3)$$

活性化関数は一般的にシグモイド関数などが使用されるが, タスクや層の位置に応じて選択する必要がある.

2.2.1 Word2Vec の概要

Word2Vec とは, Mikolov らによって提唱されたニューラルネットワーク [8] を実装したツールである. テキストコーパスを入力すると, コーパスの共起関係に基づいて任意の次元のベクトルに圧縮する. これにより, 単語同士の類似度をコサイン類似度によって表すことができる. 学習において使用するモデルは, Continuous Bag-of-Words (CBow) と Skip-gram の2つがある. 提案手法では, [9] において意味的にベクトルの精度が高いとされる Skip-gram を用いた.

Skip-gram とは, 文字列 w_1, w_2, \dots, w_n の t 番目の単語を w_t とすると, その周辺の単語 $w_{t-j}, w_{t-j+1}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+j}$ を予測するニューラルネットワークである. 入力層は単語の one-hot なベクトルとなっており, 次元数はコーパスのユニークな単語数 T になる. ここで, c を中心の単語 w_t に対して前後の単語の予測をどの距離まで行うかを定める定数とする. 学習の際は, 入力層に w_t の one-hot ベクトルを入れ, 出力層に $w_{t+i} (i = -c, -c+1, \dots, c-1, c)$ を入れることを繰り返す. 中間層は任意の次元数 k であり, 入力層から中間層への重みは T と k の積である. 中間層では活性化関数を通さないが, 出力層ではソフトマックス関数を使用する. ソフトマックス関数は合計が 1 になるよう正規化されており, 出力値は 0 から 1 の値をとる. これを利用し, 出力値をその単語が正解である確率として捉え, 正解である w_{t+i} の確率が高くなるようにネットワークの重みを調節する. 学習後のモデルにおける中間層の出力を, 単語のベクトルとして使用する.

2.2.2 CNN の概要

CNN (Convolutional Neural Network) とは, 多層のニューラルネットワークを用いた Deep Learning の一種であり, 主に画像認識で用いられている. 通常のニューラルネットワークにおいては隣接層のユニットは全て結合している. しかし, CNN の畳み込み層とプーリング層は特定のユニットのみが結合を持っているため, 高い精度で画像の特徴を抽出することが可能になっている.

一般的な CNN の構造として, 入力から畳み込み層とプーリング層のペアを複数回繰り返す. 出力層の前は全結合層が配置される. 畳み込み層 (Convolution layer) では, 任意の大きさのフィルタを入力画像に重ね合わせ, フィルタを 1 画素ずつ移動させて画像の濃淡パターンを検出する. 種類が異なる複数の

フィルタを用いることによって、画像の様々な特徴を獲得することが可能になる。プーリング層 (Pooling layer) は、畳込み層で獲得した特徴の位置感度を低下させる働きがあり、画像内で対象とする特徴が若干変化した場合でもプーリング層の出力は変化しない。畳込み層から出力された画像を正方形の画素に分割し、一般的にはそれぞれの画素集合の中で最大の値を出力する (MaxPooling)。

図3と図4に畳込み層とプーリング層の例を示す。

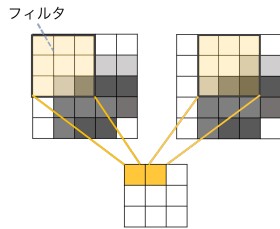


図3: フィルタの畳込みの例

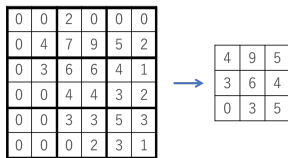


図4: MaxPooling の例

3. 提案手法

提案手法では EPUB ファイル内の本文テキストを形態素解析し、書籍内容の印象を反映した色とフォントを推薦する。また、推薦された色を背景色とし、フォントでタイトルと著者名を描画する表紙データを作成する。

色とフォントについては別のアプローチで推薦を行うため、本章ではそれぞれの手法について説明する。

3.1 色抽出の概要

本文テキストからの色抽出では、Word2Vec により拡張した色彩データベースを用いて色を想起させる単語を特定する。単語の出現頻度を考慮して色のスコアを算出し、色相ごとにスコアを合計する。そして、合計したスコアが最上位となった色相において最もスコアの大きい色を抽出色とする。

3.1.1 色彩データベース

色彩データベースは、日本カラーデザイン研究所により作成されたカラーイメージスケール [10] を利用して構築する。カラーイメージスケールでは、この代表色 130 色に対し、人々の感性を捉えるために有効な形容詞 180 語を定義している。ここでは仮に、180 語の形容詞を色彩形容詞と呼ぶ。また、色彩形容詞に対して代表色が使用される頻度を 1 から 5 の整数値で表してある。

代表色の RGB 値、色彩形容詞の語幹、使用頻度の 3 属性を格納したものを色彩データベースとする。色彩データベースの例を表 1 に示す。

表 1: 色彩データベースの例

RGB 値	色彩形容詞	使用頻度
255,236,255	淡い	2
255,152,136	あどけない	5
255,156,0	快活	4

3.1.2 色彩データベースの拡張

従来の色彩データベースは限られた形容詞しか色と対応づけられていないが、実際は形容詞以外の単語も色を連想させる可能性がある。形容詞に加え動詞と名詞も色を連想させ、意味が類似している単語は同一の色を想起させると仮定して、提案手法では 2 つの段階で色彩データベースの単語を拡張する。まず、Weblio 類語辞典 [11] を用いて色彩形容詞の類語を収集する。次に、色彩形容詞と類語に類似した単語を、Word2Vec においてコサイン類似語が高い単語を抽出することで収集する。

Word2Vec のコーパスには青空文庫の本文データを用いた、Mecab を用いた形態素解析により形容詞、動詞、名詞のリストを作成し、コーパスとしている。Skip-gram において入力単語の前後 5 単語を予測するように学習を行い、100 次元のベクトルを作成した。また、コーパス中で出現回数が 20 回に満たない単語は切り捨てて学習を行っている。

色彩形容詞と類語に対してコサイン類似度が 0.7 以上の単語を類似度が高い単語として抽出を行った。しかし、Word2Vec は入力単語に対して周辺の単語を予測するというモデルであるため、対義語もコサイン類似度が高くなる場合がある。例えば、「私は猫が好きである」と「私は猫が嫌いである」という文の「好き」と「嫌い」は対義語であるが、同じ文脈で登場してしまうのでベクトルとして表す際には類似度が高くなってしまふ。対義語同士が同じ色を想起させることはほとんどないと考えられる。そこで、あらかじめ色彩形容詞と類語の対義語を Weblio 対義語・反対語辞書から取得し、もしコサイン類似度が高い単語として対義語が抽出された場合には取り除く処理を行っている。

使用頻度の算出について述べる。Weblio で収集した単語には、収集する際に使用した色彩形容詞と等しい RGB 値と使用頻度を与える。Word2Vec で収集した単語には、色彩形容詞と類語と等しい RGB 値を与え、使用頻度にはその単語のコサイン類似度と使用頻度との積を与える。色彩データベースの拡張の例を図 5 に示す。

以上の処理によって、約 88,000 の単語、RGB 値、使用頻度の対応づけが得られた。

3.1.3 色のスコアの算出方法及び色の抽出方法

色抽出においては本文テキストを形態素解析し、色彩データベースと一致した単語の使用頻度の値をスコアとして RGB 値ごとに足し合わせる。より精度の高い色抽出を行うために 2 つの処理を行う。

否定語に続く語の除去

本文テキストに対する形態素解析では、「楽しくない」のような文は「楽しく」と「ない」に分割されて「楽しく」が色彩

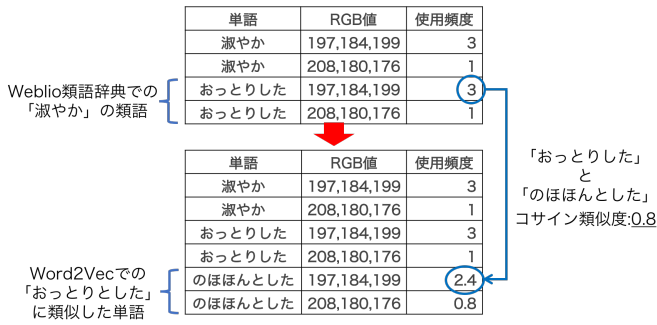


図 5: 色彩データベースの拡張の例

データベースに一致する可能性がある。活用形が未然形、連用形の単語が「ない」「ぬ」などの否定語に続く場合、直前の単語を取り除いている。

頻出単語の正規化

提案手法では Word2Vec を用いて機械的に単語を取得しているため、多義語などについては考慮されていない。そのような単語が複数回登場すると、色抽出の結果において意図しない方向に影響を与えてしまう可能性がある。そこで、常用対数を用いて単語の出現頻度に応じて正規化を行う。具体的には、色彩データベース上で単語 w が RGB 値 w_r と使用頻度 w_f を持つとすると、本文テキストでの単語 w の出現回数 w_k を取得し、単語 w の w_r に対するスコアを $w_f \times \log_{10}(w_k + 1)$ で表す。

以上の 2 つの処理をまとめると次の通りになる。ある色 C_i に対する色のスコアを S_i とし、本文テキスト中に出現した色彩データベースと一致する単語集合を W_i 、そのうち否定語に続く語の集合を N_i とする。このとき、 S_i は式 (4) によって得られる。

$$S_i = \sum_{w \in W_i \setminus N_i} (w_f \times \log_{10}(w_k + 1)) \quad (4)$$

算出した色のスコアを用いて、[4] を参考にして全 130 色のスコアを考慮した抽出法を用いる。まず、130 色を色相を基準にして 11 のグループに分ける。有彩色（赤、青、黄など）の場合は 1 つのグループが 12 の色を持ち、無彩色（黒、白など）の 10 色を 1 つのグループにまとめる。グループごとに色のスコア S_i を合計し、その合計スコアを昇順に並び替え、上位グループの中で最大のスコアを持つ色を抽出色とする。

3.2 フォント推薦の概要

フォント推薦には感性イメージをベクトルとして表した感性ベクトルを用いる。PC にあるフォントファイルの各フォントと本文テキストの感性ベクトルを推定する。フォントについては、被験者実験によって得たフォントと感性ベクトルの対を教師データとして学習を行い、作成された学習器を用いて感性ベクトルを推定する。本文テキストについては、出現する単語によって感性ベクトルを算出する。本文テキストの感性ベクトルとコサイン類似度が最も高い感性ベクトルを持ったフォントを推薦し、表紙に使用する。

3.2.1 感性ベクトル

感性ベクトルとは、感性イメージを表すベクトルであり、「古い」「新しい」などの形容詞対の尺度を要素とする。形容詞対の尺度を感性値とすると感性ベクトルの次元数は感性値、つまり形容詞対の数になる。本研究では、[5] を参考に選択した表 2 の 12 の形容詞対を用いる。また、それぞれの感性値は 1 から 7 の値をとり、表 2 で左が感性値 1、右が感性値 7 の印象の形容詞である。

表 2: 感性ベクトルに用いる形容詞対

古い	新しい
暗い	明るい
重厚な	軽快な
やわらかい	かたい
穏やかな	激しい
繊細な	大胆な
下品な	上品な
つまらない	楽しい
汚い	美しい
冷たい	暖かい
静かな	騒がしい
憎い	可愛い

3.2.2 フォントの感性ベクトルの推定方法

フォントの感性ベクトルを推定するために、本研究では感性ベクトルを推定する推定器を構成する。推定器は、被験者実験により収集した感性値を CNN で学習することにより構成する。実験に使用したフォントは PC にあらかじめインストールされているフォントや、フリーフォントなどから収集した日本語フォント計 171 個である。被験者は男性 16 名、女性 18 名の合計 32 名であり、平均年齢は 21.2 歳であった。実験の手順としては、被験者に 171 のフォントで表記された無意味文字列を提示し、フォントの印象を 12 の尺度を用いて 7 段階で SD 法で評価させた。被験者に提示した回答画面の例を図 6 に示す。

問1

安以字衣於あいうえおアイウエオ

* 1 2 3 4 5 6 7

古い ○ ○ ○ ○ ○ ○ ○ 新しい

* 1 2 3 4 5 6 7

暗い ○ ○ ○ ○ ○ ○ ○ 明るい

図 6: 被験者実験の回答画面の例

171 のフォントを 4 グループに分け、1 フォント当たり 8 名の被験者の回答を得た。各フォントの感性値の平均を算出し、これを被験者実験で得られたフォントの感性ベクトルとする。

被験者実験に用いたフォントは数が限られており、それ以外の日本語フォントも無数に存在する。そこで、未知の日本語フォントであっても感性ベクトルの推定が行えるように、被験者実験で得たデータを教師データとしてCNNによって学習を行う。フォントについては、被験者実験に用いたフォントで描画した文字の画像を用いる。フォントの特徴が現れやすいと考え「永」「あ」という2文字を使用して、90pointで115pixel×210pixelの白背景の画像に描画して作成した。入力データとなるフォントの画像の例を図7に示す。この画像を1pixelずつグレースケールに応じて0から255の数値に変換した数値列を学習の際に入力する。



図7: フォントの画像の例

学習のモデルは、CNNを用いてフォントの自動識別を行っている[12]を参考にして作成した。畳み込み層とプーリング層のペアを2回繰り返し、3回畳み込んだ後に全結合させ、シグモイド関数を通して300のユニットに出力し、線形出力関数を通して12のユニットに出力する。この12のユニットの入力値が、感性ベクトルの要素となる。フィルタの数を30、畳み込み層とプーリング層の大きさは共に3×3pixelとする。学習時の損失関数は平均二乗誤差、最適化関数はRMSpropを用い、教師データの感性ベクトルとフォントの画像から予測した感性ベクトルの平均二乗誤差が少なくなるように学習を進める。このモデルを深層学習ライブラリであるKerasで実装し、全171のデータのうち7割を訓練データ、3割を評価データとして50回(50epoch)の学習を行った。epoch毎の平均二乗誤差を図8に示す。

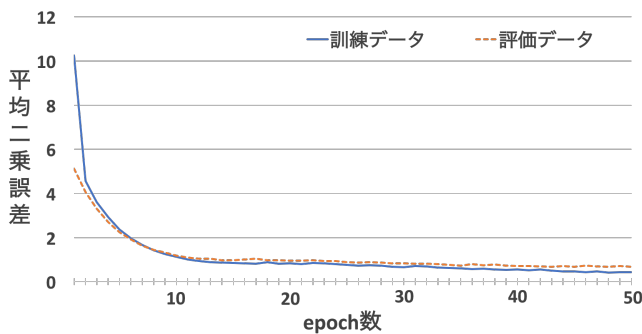


図8: フォントの感性ベクトルについての学習経過

学習済みのモデルの重みを用いることで、フォントの画像を入力するとフォントの特徴に応じた感性ベクトルを得ることができる。

3.2.3 本文の感性ベクトルの推定方法

本文テキストの感性ベクトルについては、本文に登場する単語と感性データベースの単語との一致によって算出する。ここでは仮に、感性ベクトルで用いる形容詞を感性形容詞とする。

感性データベースの作成方法について述べる。意味が類似した単語は似た感性イメージを持つと仮定して、感性形容詞と意味が近い単語を収集する。色彩データベースの拡張と同様に、Weblio類語辞典とWord2Vecを用い、感性形容詞とその類語に対してWord2Vecにおいて抽出を行う単語のコサイン類似度は0.7以上とした。また、類語の数が非常に少ない感性形容詞については日本語WordNet[13]も用いて類義語を収集した。感性データベースには単語、感性形容詞、スコアを3属性を格納する。スコアには感性形容詞と類語には1、Word2Vecで収集した単語にはコサイン類似度を与える。以上の処理によって約4,400の単語、感性形容詞、スコアの対応づけが得られた。感性データベースの例を表3に示す。

表3: 感性データベースの例

単語	感性形容詞	スコア
楽しい	楽しい	1
心踊る	楽しい	1
楽しむ	楽しい	0.745
陽気	楽しい	0.698

感性データベースを用いて感性ベクトルを算出する方法について説明する。本文テキストを形態素解析し、感性データベースと一致した単語のスコアを感性形容詞ごとに足し合わせる。その際、色抽出で行った2つの処理を同様に行う。つまり、本文テキスト中で否定語に続く語を除去し、単語の出現回数取得して常用対数によって正規化した値を感性形容詞ごとに合計する。これらの処理によって得た感性形容詞のスコアを用い、対になった感性形容詞のスコアの比を考慮して感性ベクトルを算出する。ただし、感性形容詞には異なる数の単語が対応づけられており、スコアの合計も大きく異なるので、単語が多く対応づけられている感性形容詞のスコアは大きくなりやすくなると考えられる。そこで、正規化のためにあらかじめ感性データベースにおける感性形容詞ごとのスコアの合計を算出し、対になった感性形容詞において合計スコアを使用する。具体的には、感性データベースにおいて感性値7の感性形容詞の合計スコアを D_i ($i = 1, 2, \dots, 12$)、感性値1の合計スコアを N_i とし、 N_i/D_i をあらかじめ保持する。感性形容詞ごとのスコアを算出する過程で感性値7の感性形容詞のスコアと N_i/D_i の積をとる。以上の処理を行った上で、感性ベクトルの感性値を v_i 、感性値 v_i にあたる感性形容詞対のスコアを K_i 、そのうち感性値7の感性形容詞のスコアを S_i とする。このとき、感性値 v_i は式(5)によって求められる。

$$v_i = 1 + \left(6 \times \frac{S_i}{K_i} \right) \quad (5)$$

しかし、ある感性形容詞対に類似した単語がほとんど本文テキスト中に出現しない場合には式(5)では不十分である。例えば、「古い」と「新しい」に類似した単語がそれぞれ0回と1回出現した場合に感性値は7になる。しかし、出現した単語が少ないならばテキストはその感性形容詞対についての感性イメージは弱いと考えられる。そこで、感性形容詞対に類似した単語

の出現回数が少ない場合にはベクトルのスケールを小さくする処理を行う。本文テキストの単語で感性データベースと一致した単語数を n とすると、ある感性形容詞対に類似した単語の出現回数が少ないとは、それら単語の出現回数が $n/12$ に満たないことをいう。出現回数が少ない場合、感性値 v_i にあたる感性形容詞対の出現回数を $n/12$ で割り、これを出現割合 p_i とする。感性値 7 の感性形容詞に類似した単語が感性値 1 のものより多く出現した場合、感性値 v_i は式 (6) によって求められる。

$$v_i = 4 + \left(3 \times p_i \times \frac{(S_i - K_i/2)}{K_i/2} \right) \quad (6)$$

感性値 1 の感性形容詞に類似した単語が多く出現した場合には、感性値 1 の感性形容詞のスコアを T_i とし、式 (7) によって算出する。

$$v_i = 4 - \left(3 \times p_i \times \frac{(T_i - K_i/2)}{K_i/2} \right) \quad (7)$$

4. 評価実験

本章では、提案手法の CNN を用いた学習モデルの評価と、色抽出とフォント推薦に関する評価実験について述べる。

4.1 CNN による学習の評価

提案手法の CNN の学習モデルを評価するために 10 分割交差検定を行う。具体的には被験者実験で得た 171 のデータをシャッフルし 10 分割する。その 10 グループのうち、1 つのグループを評価データ、残りの 9 グループを訓練データとして 50epoch の学習を行い、1epoch 毎に評価データの平均二乗誤差を記録する。評価データに用いるグループを変えながら同様の学習を 10 回行い、平均二乗誤差の平均を算出する。以上の手順で得られた平均二乗誤差を図 9 に示す。

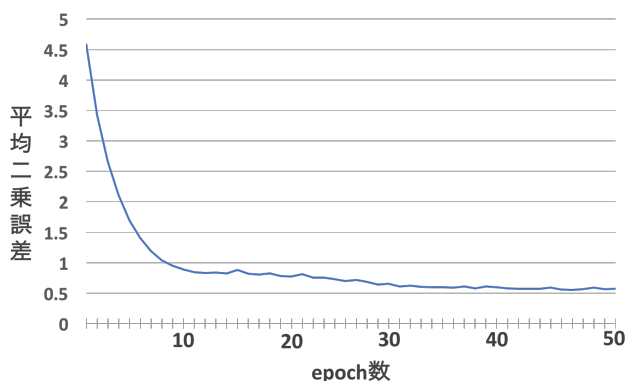


図 9: 交差検定での学習経過

交差検定では全てのデータを 1 度は評価データとして用いるので、学習モデルが過学習を起こすか、それとも未知のデータにも対応できる汎化性能が優れているかを見ることができる。提案手法の学習モデルは、評価データの平均二乗誤差がおおよそ 30epoch 目から約 0.6 で安定する。これは、未知の日本語フォントに対してもおおよそ妥当な感性ベクトルを予測できることを示している。

4.2 色抽出とフォント推薦の評価

提案手法により抽出した色と推薦されたフォントが、テキスト内容の印象を反映しているかを検証する実験を行った。色とフォントを別のアプローチで推薦するため、色とフォントそれぞれが妥当かについて評価する。評価方法としては、提案手法で推薦された 5 つの色やフォントを含む 10 の選択肢を提示して、推薦された色やフォントが選択されるかを検証した。対象は大学生 4 名、本実験で使用するテキストは『走れメロス』『羅生門』『山月記』の 3 作品である。

4.2.1 選択肢の作成と実験方法

まず色の選択肢の作成方法について説明する。カラーイメージスケールによると、有彩色（赤、青、黄など）は色相とトーンの違いを持ち、順に似た色相を並べると色相環ができる。10 色相環の例を図 10 に示す。

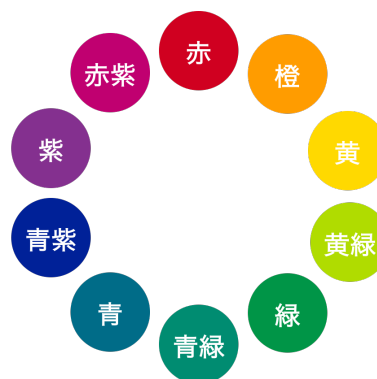


図 10: 10 色相環

それぞれの色相には明暗、派手、地味というように共通した色の調子（トーン）がある。トーンは明暗の違いを生む明度と派手・地味の違いに関わる彩度との相互の関係のもとに成り立つ。実験では色相と色調をもとに色の選択を行う。提案手法で色のスコアを算出し色相ごとのスコアを昇順して上位 5 位までの色相グループを代表グループとし、それぞれのグループで最大スコアを持つ色を抽出色とする。5 つの抽出色の属する色相と異なる色相の色で、色抽出の際に色ごとのスコアを昇順にし中央より下位になる色を他の 5 つの選択肢とする。また、5 つの選択肢については可能な限り、抽出色と異なる色調になるよう選ぶ。被験者には同一ディスプレイで 10 色を提示する。

フォントについては、被験者実験に用いた 171 のフォントから書体が偏らないように選んだ 23 フォントと新しく収集した 9 フォントの合わせて 32 フォントを推薦の対象となるフォントとした。選択肢の作成も色と同様に行い、フォントにおいては明朝体やゴシック体のような書体を基準にして選択する。提案手法でのフォント推薦の結果、コサイン類似度を昇順にした際の上位 5 フォントを推薦フォントとして、他の 5 つの選択肢が似ないように書体が異なるものから選ぶ。被験者には 10 のフォントを用いて、それぞれテキストのタイトルを描画して提示する。

以上の選択肢を用いて、テキストを読んでもらった上で、被験者にテキスト内容の印象に近いものを 3 つ選択してもらう評

価実験を行った。

4.2.2 実験結果

3つのテキストに対する提案手法による抽出色と推薦フォントを表4に示す。評価実験で被験者が抽出色・推薦フォントを選択した場合は1から5を、それら以外を選択した場合は×を記し、表5、表6、表7に示す。

表4: 抽出色と推薦フォント

テキスト		抽出色	推薦フォント
走れメロス	1	255,229,122 (淡黄色)	Corporate Logo Medium
	2	255,88,80 (ぼら色)	BT 12G Smooth Regular
	3	0,0,0 (黒)	ロゴたいぶゴシック Condense
	4	176,143,119 (茶ねずみ)	Honoka Antique-Kaku
	5	197,184,199 (うすいろ)	メイリオ
羅生門	1	0,0,0 (黒)	BT 12G Smooth Regular
	1	255,229,122 (淡黄色)	メイリオ
	3	208,0,32 (赤)	BT 10G Dot Regular
	4	179,186,200 (浅藍ねずみ)	Honoka Antique-Kaku
	5	255,168,40 (あんず色)	鉄瓶ゴシック
山月記	1	199,199,187 (明灰色)	ほのか明朝
	2	197,184,199 (うすいろ)	源ノ角ゴシック Regular
	3	255,217,0 (黄)	いろはマル
	4	255,213,159 (白茶)	Rounded M+ 1p light
	5	179,186,200 (浅藍ねずみ)	C4 ユニバーサルビュー DSP EL

表5: 『走れメロス』の評価実験の結果

被験者	選択色	回答一致率	選択フォント	回答一致率
被験者 A	2, 4, 5	100%(3/3)	1, ×, ×	33%(1/3)
被験者 B	1, 3, ×	66%(2/3)	3, ×, ×	33%(1/3)
被験者 C	1, 2, ×	66%(2/3)	3, ×, ×	33%(1/3)
被験者 D	1, ×, ×	33%(1/3)	4, ×, ×	33%(1/3)

表6: 『羅生門』の評価実験の結果

被験者	選択色	回答一致率	選択フォント	回答一致率
被験者 A	3, 4, ×	66%(2/3)	1, ×, ×	33%(1/3)
被験者 B	1, 3, ×	66%(2/3)	2, 4, 5	100%(3/3)
被験者 C	1, 3, ×	66%(2/3)	3, ×, ×	33%(1/3)
被験者 D	1, 3, ×	66%(2/3)	1, 3, ×	66%(2/3)

表7: 『山月記』の評価実験の結果

被験者	選択色	回答一致率	選択フォント	回答一致率
被験者 A	1, 3, 4	100%(3/3)	3, 4, ×	66%(2/3)
被験者 B	1, ×, ×	33%(1/3)	1, 2, 3	100%(3/3)
被験者 C	2, 3, ×	66%(2/3)	1, 2, 5	100%(3/3)
被験者 D	1, 4, 5	100%(3/3)	1, 2, ×	66%(2/3)

4.2.3 考察

評価実験の結果から、提案手法では抽出色は印象に沿う場合が多いが、フォントについてはテキストによって沿う場合と沿わない場合に分かれることが示された。回答一致率が低い被験者が確認された『走れメロス』『羅生門』について考察する。

『走れメロス』は色についておよそ60%ほどの回答一致率になったが、フォントの回答一致率は33%となり推薦フォント以外のフォントがより選択される結果になった。色において抽出色以外で2名の被験者に選択されたのは明るい黄緑色であり、主人公であるメロスが作品後半で走っている様子に影響を受けた可能性がある。色の抽出過程では黄緑色は『賑やか』『陽気』などとの単語と一致していたが、一致する単語数が少なかったため上位に推薦されなかった。フォントについては、推薦フォント以外の行書体が全ての被験者に選択されていた。これは『走れメロス』は主人公が走っているイメージが強いため、文字の面がつなり、すばやい印象を抱く行書体を選択したと考えられる。『走れメロス』に出現する単語から推定した感性ベクトルを表8に示す。

表8: 『走れメロス』の感性値

感性形容詞	感性値
古い - 新しい	4.0
暗い - 明るい	3.477
重厚な - 軽快な	4.0
やわらかい - かたい	4.423
穏やかな - 激しい	3.555
繊細な - 大胆な	3.234
下品な - 上品な	2.769
つまらない - 楽しい	3.883
汚い - 美しい	1.993
冷たい - 暖かい	4.927
静かな - 騒がしい	3.035
憎い - 可愛い	3.344

『走れメロス』のフォントの回答一致率が低くなった原因として考えられるのは、感性ベクトルで使用する形容詞に「走っている」ことを表現するものが含まれていなかったことである。また、物語は主人公が親友との友情を示して終わるので「美しい」と感じる読者もいる可能性があるが、推定した感性ベクトルでは「汚い」に近い感性値になっている。小説の導入から中盤までは、主人公が王に怒る場面や自身を「醜い」と感じる場面が多くあったためである。現在の推薦手法では小説全体に出現した単語を同様に扱うため、このようなことが起こりやすいと考えられる。

『羅生門』は、色の回答一致率はおよそ60%だったが、フォントについては100%と33%の被験者がおり差がある結果になった。フォントでは、推薦フォント以外で隷書体が3名の被験者に選択されていた。隷書体の感性ベクトルは「古い」に近い感性値を持つが、『羅生門』から推定した感性ベクトルは「古い-新しい」の感性値は4であった。小説中にそれぞれに類似した単語が1つも出現しなかった結果、「古い」「新しい」に関してはどちらにも近くない感性値になっている。隷書体を選択した被験者は『羅生門』の作品全体から「古い」印象を受けた可能性がある。今回作成した感性データベースでは単語が「古い」に類似しているかのみを指標として用いており、「古い」という意味を持たない単語から古い印象を受ける場合に感性値をうまく算出できない問題があると考えられる。

5. むすび

本研究では、書籍内容の印象を反映した表紙を生成することを目的として、テキストデータから色とフォントを推薦する方法を提案した。その結果、色の推薦においては印象に沿う色を概ね推薦できるが、フォントについてはテキストによって印象に沿う場合とそうでない場合に分かれることが示された。今後、意味が類似した単語の収集方法や感性ベクトルに使用する形容詞、テキストからの感性ベクトルの算出方法を改善することで、より印象を反映した表紙を生成することができると考えられる。

文 献

- [1] “青空文庫”, 青空文庫, <http://www.aozora.gr.jp/>, (参照 2016-12-4).
- [2] 飯田拓也, 梶山朋子, 大内紀知, 越前功, “読者の印象を反映させた書籍表紙画像生成のための色抽出手法,” 電子情報通信学会論文誌, vol.J97-D, no.1, pp.75-84, 2014.
- [3] International Digital Publishing Forum, “EPUB”, International Digital Publishing Forum, <http://idpf.org/epub>, (参照 2016-11-29).
- [4] 佐野泰成, “小説データにおける情景描写を考慮した色抽出手法,” 筑波大学, 卒業論文, 2015.
- [5] 飯場咲紀, 土斐崎龍一, 坂本真樹, “テキストの感性イメージを反映した色彩・フォント推薦,” 日本バーチャルリアリティ学会論文誌, vol.18, no.3, pp.217-226, 2013.
- [6] Peter O'Donovan, Janis Libeks, Aseem Agarwala, Aaron Hertzmann, “Exploratory Font Selection Using Crowdsourced Attributes,” ACM Transactions on Graphics (Proc. SIGGRAPH), vol.33, no.4, 2014.
- [7] 高瀬拓史, “EPUB 概説:電子出版物と Web 標準,” 情報管理, vol.57, no.9, pp.618-628, 2014.
- [8] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean, “Distributed representations of words and phrases and their compositionality,” In Advances in neural information processing systems, pp.3111-3119, 2013.
- [9] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient estimation of word representations in vector space,” ICLR Workshop, 2013.
- [10] 小林重順, カラーイメージスケール改訂版, 講談社, 2001.
- [11] “Weblio 類語辞典,” <http://thesaurus.weblio.jp/>, (参照 2016-12-12).
jwiki/latest/, (参照 2016-12-12).
- [12] Z. Wang, J. Yang, H. Jin, E. Shechtman, A. Agarwala, J. Brandt and T. Huang, “DeepFont: Identify Your Font from

- An Image,” ACM International Conference on Multimedia (ACM MM), pp.451-459, 2015.
- [13] “日本語 WordNet,” <http://nlpwww.nict.go.jp/wn-ja/>, (参照 2017-2-15).