

単語分散表現におけるフォントスタイル情報の利用

原口 大地^{1,a)} 下田 和² 内田 誠一¹

概要: 本研究の目的はフォントを用いて単語分散表現をより明確に表現することである。単語分散表現を獲得できる有名な手法として Word2Vec が知られている。しかし、Word2Vec では“beautiful”と“ugly”のような対義語ペアに対して類似した分散表現が獲得されることがある。これは「意味の類似した単語は類似した文脈上で使用される」という分布仮説に基づいており、Word2Vec の内部では周囲の共起単語が類似している単語には類似した分散表現を与える処理が行われるためである。本研究では単語とその単語に使用されるフォントの関係を分析したところ、一部の単語間では使用されるフォントに差異があることを確認した。この結果を踏まえ、Word2Vec にフォント情報を組み込むことで、より明確な単語分散表現の獲得を目指す。

1. はじめに

フォントは読者に様々な印象を与える書体デザインである。そのため、ポスターや書籍表紙等の制作フローにおいて、フォントは専門家によって注意深く選択される。文献 [1] ではフォントとそのフォントが与える印象を評価し、印象からフォントの検索を可能にしている。

フォントと印象に関係性があることから、本研究では、フォントと単語（の意味）にも関係があると仮定する。例えば、図 1 に見られるように、“beautiful”という単語ではその意味を表現するような洗練されたフォントが使用されやすく、“ugly”ではその単語の意味同様の印象を与えるような醜い（風変わりな）フォントが使用されやすいと考えられる。このようにフォントの与える印象と単語の意味には一定の相関があると考えられる。

本研究の目的は、上記の仮定に基づき、単語分散表現にフォントスタイル情報を利用することである。単語分散表現を獲得できる有名な手法として Word2Vec [2], [3] が知られている。Word2Vec は「意味の類似した単語は類似した文脈上で使用される」という分布仮説に基づいて設計されており、内部では周囲の共起単語が類似している単語には類似した分散表現を与える処理が行われている。そのため、“beautiful”と“ugly”のような同様の文脈で用いられる対義語ペアに対して類似度の高い分散表現が獲得されることがある。また、前述の通りある単語において、その単語の



図 1: 書籍表紙の例。

意味や印象を表すフォントが使用されやすく、対義語ペアでは異なるフォントが使用されやすいと考えられる。そこで、Word2Vec にフォントスタイル情報を組み込み、再学習を行うことでより適切な分散表現の獲得、すなわち同義語ペアに対しては類似度の高い分散表現、対義語ペアに対しては類似度が低い分散表現の獲得を目指す (図 2)。

Word2Vec の単語分散表現を様々なモダリティで補正する既存手法として文献 [4] では単語の色を、文献 [5] では単語の音を用いて単語分散表現の更新を行うことにより、一定の単語において識別力のある単語分散表現の獲得に成功している。これらの先行研究の成功例から、単語に相関があるフォントスタイル情報を Word2Vec に組み込むことは、識別力の高い分散表現の獲得に寄与すると考えられる。

前述の目的に対し、本稿では次の二項目について実施した結果を報告する。第一に約 21 万冊の書籍表紙画像のタイトル部から収集された単語画像からフォントスタイル情報を取得した。書籍表紙はデザイナーによって慎重に作成されているため、それらの単語画像とそのフォントは意味のある相関関係を持つことが期待される。また、各単語に使用されるフォントの分布をフォントスタイル情報として

¹ 九州大学大

² 株式会社サイバーエージェント

³ 九州大学大学院システム情報科学研究院

^{a)} daichi.haraguchi@human.ait.kyushu-u.ac.jp

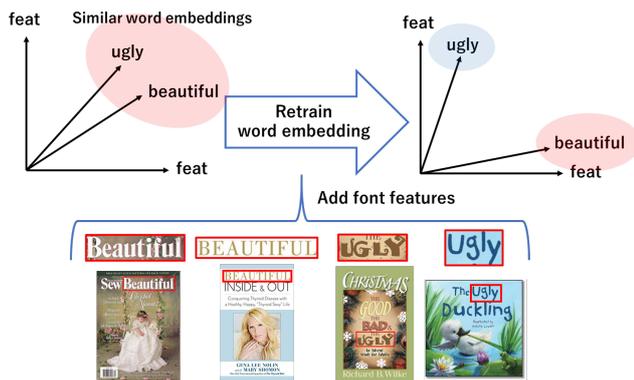


図 2: 本研究の概要.



図 3: 各フォントクラスの単語の合成画像例.

取得した. 本実験では, 上記のフォントスタイル情報を用いて, 単語とその単語に使用されるフォントの傾向の分析を行った. その結果, 一部の単語で使用されるフォントに顕著な傾向がみられた. また, 単語カテゴリごとに使用されるフォントの一定の傾向も確認することができた.

第二に, フォントスタイル情報を用いて単語分散表現の更新を行った. 図 2 に既に示したようにフォントスタイル情報を用いることで “beautiful” と “ugly” のような対義語ペアにおいて, より識別力のある分散表現の獲得が期待される. 一方で, フォントの相関関係がない単語も一定数存在すると考えられるため, 悪影響も想定される. そこで, 通常の word2vec の分散表現の改善例及び劣化例を示す.

本研究の主な貢献は以下の三点である.

- 著者が知る限り, 本研究は膨大な実単語画像を用いたフォントと単語の客観的な相関解析の初めての試みである.
- フォントスタイル情報を利用した全く新しい単語分散表現を提案する.
- 全ての単語においては単語分散表現の改善は見られなかったものの, フォントスタイル情報の利用は, 一部の同義語・対義語ペアに関して単語分散表現の識別力向上に貢献することを明らかにした.

2. 単語とフォントスタイル傾向の分析

本節では, 実画像を用い, 単語とその単語に使用されるフォントの傾向の分析を行った実験の結果について述べる.

2.1 書籍タイトルデータセットの生成

フォントと単語の関係を分析する対象として, 本実験では, 書籍表紙上のタイトル部分の単語を使用した. その理由は, 書籍表紙は, 専門家によってそのデザインが慎重に設計されるため適切なフォントが選択されており, それらの単語とフォントに一定の関係があると期待されるためである. このためには, 書籍表紙画像に対して, 文字検出及び文字認識を行いタイトル部分を抽出する必要がある. 詳細は以下の通りである.

本実験では 207,572 枚の書籍表紙画像^{*1}を用い, そのタイトル部の文字領域をフォントスタイル情報の傾向の分析に使用した. 書籍表紙中の文字領域の抽出には CRAFT[6] を使用し, その文字領域に対し文字認識を行った. なお, 抽出した文字領域の高さが 15pixel 以下の単語は除外した. 文字認識には文献 [7] で最も認識精度の高かったモデル TPS-ResNet-BiLSTM-Attn を使用した. その文字認識結果と書籍表紙のメタデータに含まれるタイトル情報を照合し, タイトルに相当する部分のみの単語を選出した. その後, stop words, 固有名詞, 数字, 複合語, 及び出現回数が 10 回未満の単語は除外した. また, lemmatizer^{*2} を使用して時制など単語の表記を統一した. 結果的に, 書籍表紙画像からは, 8,384 種類の単語が語彙として得られた. ただし, 後述のフォントスタイル情報を用いた単語分散表現の獲得では, Google News corpus を用いて単語分散表現が事前学習済みの Word2Vec を使用するために, その学習時に用いられていない単語 (特殊な固有名詞や造語等) は対象にできない. よって, 本実験を通して使用する単語は, Google News corpus による 300 万単語と書籍表紙タイトル部の単語の積集合を取り, 7,883 単語となった.

2.2 フォントスタイル情報の取得

各書籍タイトルのフォントスタイル情報を取得するために ResNet18[8] を用いて 6 クラスのフォントスタイル分類を行った.

2.2.1 フォント分類器を学習するためのデータセット

様々なフォントで印字した単語の合成画像を生成し, フォント分類器を学習するためのデータセットとして使用した. 合成画像の生成には SynthText[9] を使用した. 合成画像に用いたフォントは Adobe Font Folio^{*3} から 1,127 フォント, Ultimate Font Download^{*4} から 704 フォントを使用した. また, フォント分類のクラスは文献 [10] に従い, Serif, Sans-Serif, Hybrid, Script, Historical Script, Fancy

^{*1} <https://github.com/uchidalab/book-dataset>

^{*2} www.nltk.org/modules/nltk/stem/wordnet.html

^{*3} <https://www.adobe.com/jp/products/fontfolio.html>

^{*4} <http://www.ultimatefontdownload.com/>

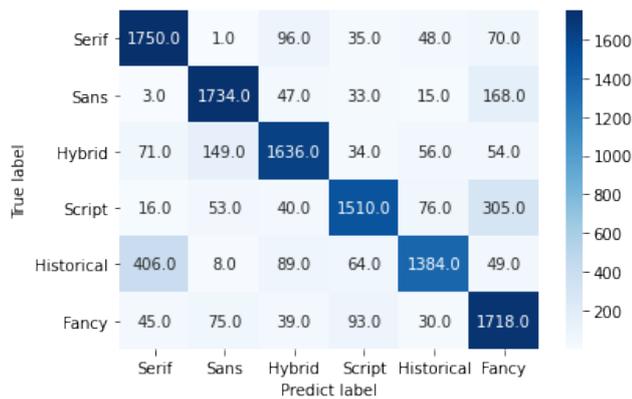


図 4: フォント分類のテスト結果に対する confusion matrix. Sans は Sans-Serif, Historical は Historical Script を表す.

の 6 クラスに分類した. Fancy はその他 5 クラスに属さないフォントをまとめたクラスである. 図 3 にフォントクラスごとの合成画像の例を示す.

単語の合成画像を各フォントクラスごとに 15,000 枚を学習, 各フォントクラスごとに 2,000 枚をバリデーションとテストに使用した. Serif, Sans-Serif, Hybrid, Script, Historical Script, Fancy のそれぞれに属するフォント数は 510, 314, 151, 74, 58, 704 フォントである. ただし, 学習, バリデーション, テストデータの各フォントクラスに含むフォント数は 8 : 1 : 1 の割合で分割している.

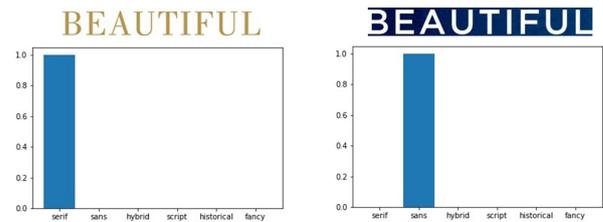
2.2.2 フォント分類

ResNet18 を用いてフォント分類を行った. 入力 は 上述のデータセットの画像であり, 正解ラベルは上述のフォントクラスの 6 クラスとなっている. ただし, 入力画像は convolution や pooling 処理で処理できるサイズを保つため, 高さが 100pixel になるようアスペクト比を保ちながら正規化を行っている. 本実験のモデルは global average pooling を含み, 画像サイズには依存しない学習が可能であるため入力画像の幅は任意とする. 学習における損失関数 L_f は, 学習に使用した各フォントクラスに属するフォント数の逆数を重みとして乗算した重み付きクロスエントロピー

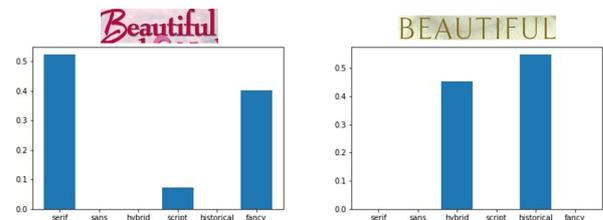
$$L_f = - \sum_c \frac{1}{N_c} \sum_{i=1}^{N_c} t_{c,i} \log y_{c,i} \quad (1)$$

を使用した. ここで N_c は学習データのフォントクラス c に属するフォント数である. また, $y_{c,i}$, $t_{c,i}$ は, それぞれフォントクラス c の入力 $x_{c,i}$ に対する出力値及び正解ラベルである.

上記の条件で学習後, SynthText によるテストデータのフォントクラス分類精度は 81.10%であった. 図 4 にテスト結果の混同行列を示す. 同図から Historical Script の画像における Serif への誤分類が多く存在することが確認できる. これは図 3(e) 下段で確認できるように, Historical



(a) ピークが 1 つのみの例.



(b) ピークが複数存在する例.

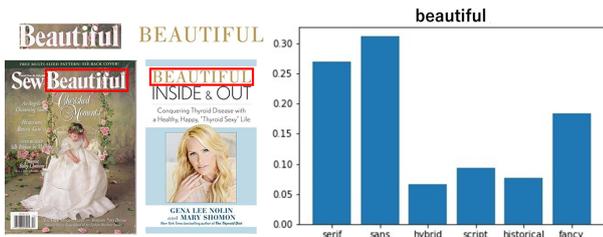
図 5: フォント分類モデルにおける softmax 関数の出力例. 横軸はフォントクラス (左から順に Serif, Sans-Serif, Hybrid, Script, Historical Script, Fancy).

Script には非常に Serif に類似したフォントを含んでいるため, 分類が困難であったと考えられる. 後述するように, フォントスタイル情報の分析には, 分類結果のクラスラベルではなく, 全クラス分の softmax 関数の出力値を用いるため, この問題点は緩和されると考えられる.

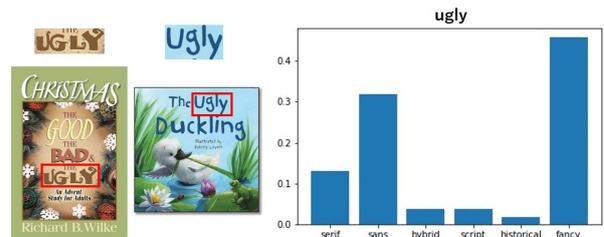
2.2.3 書籍表紙タイトル部のフォントスタイル情報

上述のフォント分類モデルを使用し, 書籍表紙タイトル部の単語のフォントスタイル情報の取得を行った. 図 5 に書籍表紙タイトル部に対してフォント分類モデルで推論し, 得られた最終層の softmax 関数の値を示す. 書籍表紙タイトル部でも十分にフォントスタイル情報の取得が可能であることが確認できる. 同図 (a) から Serif や Sans-Serif のピークが 1 つであり, 適切な分類結果であると確認できる. 一方で, 同図 (b) 左では 1 文字目とそれ以降の文字が異なっており, それらのフォントを考慮した softmax 値を出力している. また, 同図 (b) 右では Hybrid と Historical script のフォントクラスの境界が難しいものとなっておりピークが 2 つ存在することが確認できる. この結果から必ずしもフォントクラスを一意に定めることは困難であるとされる. そこで, 分析には softmax 関数の出力結果をフォントスタイル特徴として使用する.

求めたフォントスタイル特徴を書籍表紙タイトル部の分析に利用する. 書籍表紙タイトル部のフォントスタイル特徴を取得後, 同単語でそれらの平均を計算し, 各単語の平均フォントスタイル特徴を取得する. 図 6 に書籍タイトル部の単語の平均フォントスタイル特徴例を示す. 同図 (a) では, “beautiful” にはフォーマルなフォントの Serif 及び Sans-Serif が使用されやすいことが確認できる. 一方



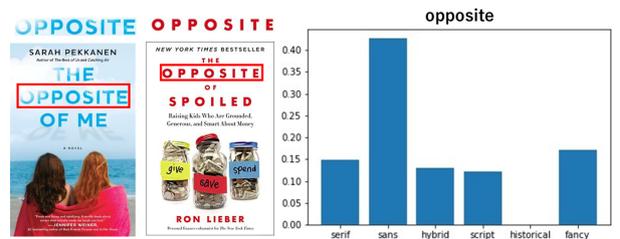
(a) "beautiful"



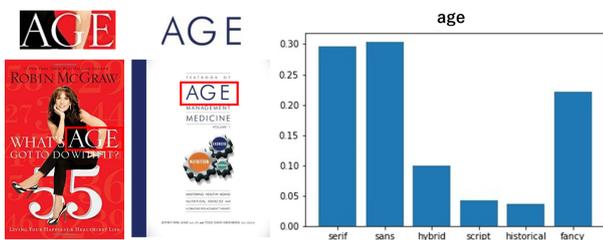
(b) "ugly"



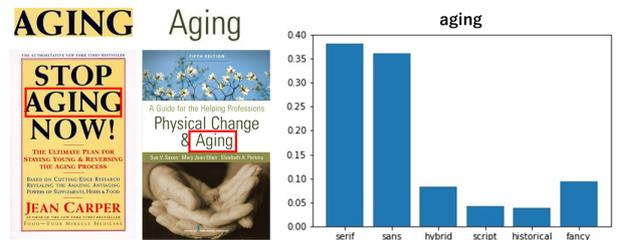
(c) "synonym"



(d) "opposite"



(e) "age"



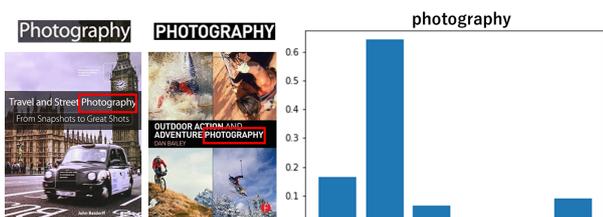
(f) "aging"



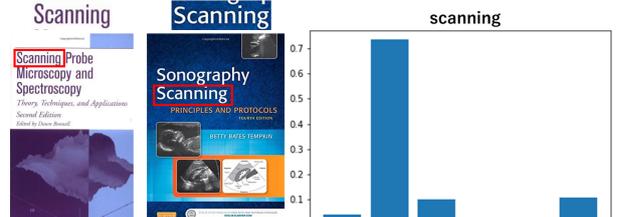
(g) "bike"



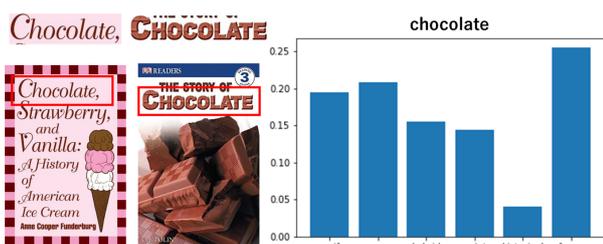
(h) "wheel"



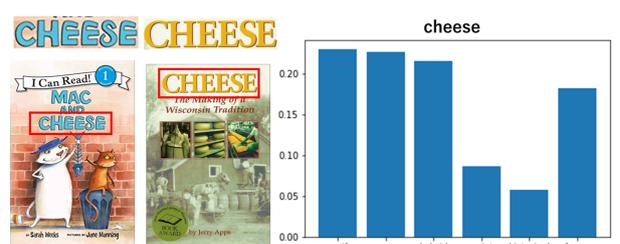
(i) "photography"



(j) "scanning"



(k) "chocolate"



(l) "cheese"

図 6: 書籍表紙タイトル部における単語の平均フォントスタイル特徴の例。

で、同図 (b) では、“ugly”にはインフォーマルな Fancy が使用されやすいことが確認できる。“beautiful”と“ugly”の意味が対照的であるだけでなく、使用されやすいフォントの傾向も対照的であることがわかる。同様に、同図 (c),(d) の“synonym”と“opposite”は対義語関係にあり、“synonym”では Serif, “opposite”では Sans-Serif と異なるフォントスタイル特徴が顕著であることが分かる。このような傾向が他の単語ペアにも期待される。加えて、同図 (e), (f) の“age”と“aging”は同義語関係にあり、ともにフォントスタイル特徴のピークは Serif と Sans-Serif の2つにあることが確認できる。このように、同義語ではフォントスタイル特徴が類似することが期待される。

2.3 各単語に使用されるフォントスタイルの傾向

各単語に使用されるフォントの傾向を分析するために横軸をフォントスタイル特徴の分散、縦軸を全単語の平均フォントスタイル特徴との距離で表したグラフを図7に示す。ここで、フォントスタイル特徴の分散を

$$\sigma_w^2 = \frac{1}{K} \sum_k (h_w^k - \mu)^2 \quad (2)$$

とする。ただし、単語 $w \in V$ における平均フォントスタイル特徴を $h_w = (h_w^1, \dots, h_w^K)$ とする。また、平均フォントスタイル特徴は softmax 値の平均であるため $\sum_k h_w^k = 1$ であり、 $\mu = \sum_k h_w^k / K = 1/K$ となる。また、平均フォントスタイル特徴との距離を

$$l_w = \|h_w - \bar{h}\| \quad (3)$$

とする。ここで、全単語の平均フォントスタイル特徴は $\bar{h} = \sum_w h_w / |V|$ 、全単語数は $|V|$ である。同図の各軸を平均で区切り、この軸を基準に赤色の領域を第一象限、黄色の領域を第二象限、緑色の領域を第三象限、青色の領域を第四象限とする。各象限に含まれる単語数はそれぞれ 1,938, 1,387, 3,556, 1,002 単語である。第一象限に属する単語が最も顕著なフォントスタイル特徴を持っており、そこにおよそ 25%の単語が存在することから、一部の単語では使用されるフォントに相関があると言える。

2.4 単語カテゴリごとのフォントの傾向

図8に単語カテゴリごとに4象限に含まれる単語の割合を示す。単語カテゴリは英語の概念辞書である WordNet の Lexnames^{*5}に従い、分類を行った。同図から「処理 (process)」に関する名詞群 (“noun.process”, 例えば “photography” や “scanning” などが含まれる) が第一象限に多く、使用されるフォントに偏りがあることが考えられる。図6(i), (j)に「処理」に関する名詞のフォントスタイル特

*5 wordnet.princeton.edu/documentation/lexnames5wn

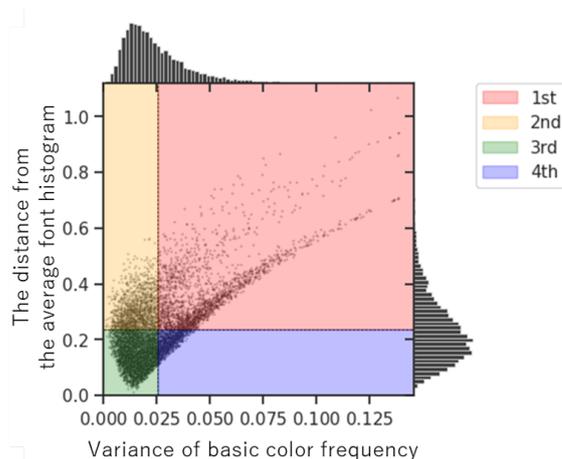


図7: 各単語に使用されるフォントの傾向分析のための分布。

徴の例を示す。これらの単語に使用されるフォントの多くは Sans-Serif であった。また、同図 (j) の “scanning” はサイエンスやメディカル系の書籍に多く含まれていた。

一方で、第一象限が少ない単語カテゴリとして「食べ物 (food)」に関する名詞 (“noun.food”) が見られる。図6(k), (l)に食べ物に関する単語のフォントスタイル特徴の例を示す。これらの例から、「食べ物」に関する単語群は、使用されるフォントが一意に定まりにくいことが分かる。

3. Word2Vec におけるフォントスタイル情報の利用

3.1 フォントスタイル情報を利用した単語分散表現の獲得

本実験では同義語・対義語それぞれのペアに対し、より識別力の高い分散表現を獲得するために、前章で得られたフォントスタイル特徴を用いて Word2Vec の単語分散表現の更新を行った。上述の書籍表紙のタイトル部から得られた 7,883 単語の分散表現のみ更新を行い、その他の単語については考慮しないことに注意されたい。上述の単語を入力とし、Word2Vec の再学習を行った。

図9に単語分散表現の更新方法の概要を示す。再学習では入力 x_w を入力単語 w の V 次元 one-hot ベクトルとし、出力は対応する単語 w の平均フォントスタイル特徴を表す $K = 6$ 次元のベクトルである。中間層の出力である N 次元ベクトル $W_I x_w$ は、入力単語 w の分散表現となる。以上より、ニューラルネットワークは次の損失関数 L を最小化するように学習する。

$$L = \sum_w \|h_w - f(W_O W_I x_w)\| \quad (4)$$

ここで、 $W_I \in \mathbb{R}^{N \times V}$ 及び $W_O \in \mathbb{R}^{K \times N}$ は学習される重み行列であり、 f は softmax 関数である。 W_I は Word2Vec で学習済みの重みで初期化し、 W_O はランダムな値で初期化する。ネットワークは単語分散表現 $W_I x_w$ と W_O から softmax 関数を介してフォントスタイル特徴 h_w の予測

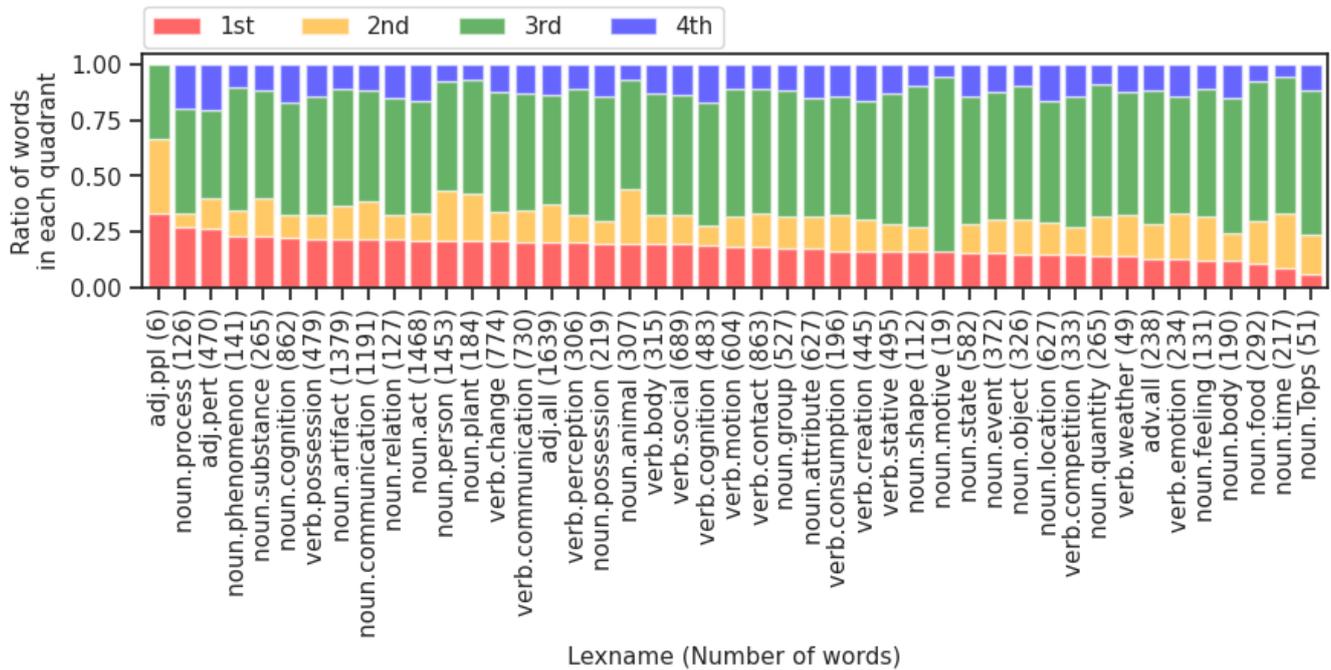


図 8: 単語カテゴリごとの 4 象限に含まれる単語の割合. カテゴリ名の括弧内の数値は同カテゴリに属する単語数を表す.

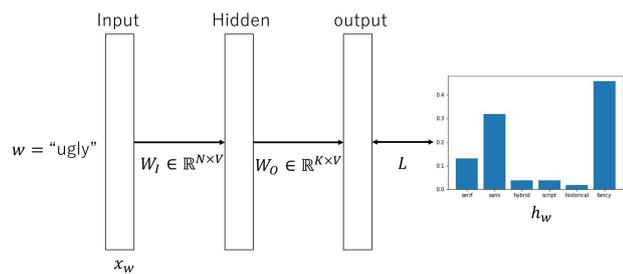


図 9: フォントスタイル特徴を用いた word2vec の単語分散表現の更新.

を行う. 以上の学習により, 異なる (類似した) フォントが使用されやすい単語間でより異なる (類似した) 単語分散表現の獲得が期待される.

3.2 同義語・対義語ペアの単語分散表現の変化

単語分散表現の更新後における同義語・対義語ペアの分散表現のコサイン類似度変化を図 10 に示す. ここで 2 つの単語 w_1, w_2 間において, 再学習前の単語分散表現のコサイン類似度を \bar{s}_{w_1, w_2} とし, 再学習後, すなわちフォントスタイル情報を利用したときの分散表現のコサイン類似度を s_{w_1, w_2} とする.

同図から同義語ペアでは類似度が向上したペアが類似度が低下したペアよりも多いことが確認できる. 一方で, 対義語ペアでは類似度が向上したペアが類似度が低下したペアよりも増加した. このことから, 必ずしもポジティブな結果, すなわち, 期待したような類義語ペアでは類似度を向上させ, 対義語ペアでは類似度を低下させる結果とはならなかった. 各単語ペアにおいて, 全体的にフォントス

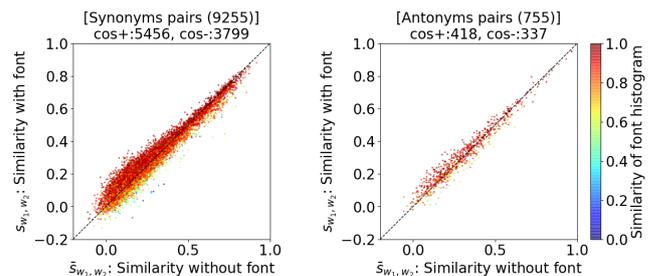


図 10: フォントスタイル情報の利用前後における同義語・対義語ペアの単語分散表現のコサイン類似度の変化. 各点の色は各単語ペアのフォントスタイル特徴の類似度を表している. また $\text{cos}+$ の数値はフォント情報の利用後にコサイン類似度が向上したペア数, $\text{cos}-$ は低下したペア数を示す.

イル特徴のコサイン類似度が高いもの (同図中の赤点) が多いことが要因の一つに考えられる. 同図からフォントスタイル特徴のコサイン類似度が比較的低いもの (青点や黄点) は, フォントスタイル情報を利用したとき単語分散表現の類似度も低くなっていることがわかる. このことから再学習に使用したモデルが適切に機能し, フォントスタイル情報が識別力の高い分散表現の獲得に寄与していると言える.

分散表現の更新により, 対義語ペアに関して分散表現の類似度が微量ながら向上してしまったものの, 大きな劣化を抑えつつ同義語ペアに関しては識別力が向上した. 図 11 に品詞ごとの分散表現の類似度分布の変化を示す. 同義語ペアでは名詞, 動詞, 形容詞で中央値の増加が確認できる.

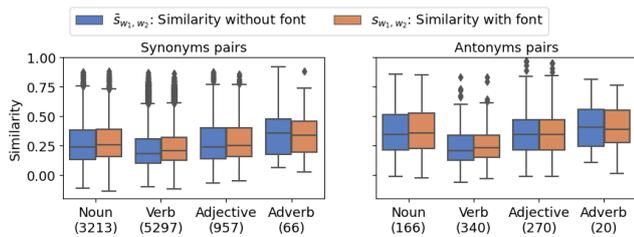


図 11: 品詞ごとのフォントスタイル情報の利用前後における同義語・対義語ペアの単語分散表現のコサイン類似度の変化。

このことからフォントスタイル情報によってポジティブな影響を受けたと考えられる。一方、副詞の最大値が大きく変化しており、フォントスタイル情報の利用前よりも類似度が小さくなっている。副詞は書籍タイトル部で中心の単語になることが少ないためタイトル部のその他の単語のフォントの影響を受けた可能性が考えられる。対義語ペアでは分布の広がりには大きな変化は見られないことから、分散表現の更新による大きな影響がなかったと考えられる。

4. おわりに

本報告では、実験環境下の単語に使用されるフォントの傾向の分析を行ったところ、一部の単語には使用されるフォントと単語に一定の傾向があることが分かった。その結果に基づき、フォントスタイル情報を用いて Word2Vec の単語分散表現の更新を行ったところ、一部の同義語・対義語それぞれのペアでより識別力の高い分散表現の獲得に成功した。

今後の課題として、フォントスタイル情報を他の要素と組み合わせることが考えられる。例えば、先行研究である文献 [4] では分散表現の更新に単語の色が使用されており、対義語に対してより識別力の高い単語分散表現の獲得に成功している。本研究では同義語で識別力のある分散表現が獲得できたことから、色とフォントを組み合わせると分散表現を更新することで同義語・対義語どちらに対しても識別能力の高い分散表現の獲得が期待される。

また、フォント分類における分類クラス数の再検討も今後の課題として挙げられる。現在は既存モデルの ResNet18 を用い、フォントの 6 クラス分類を行っている。本実験の結果では、同義語・対義語それぞれのペアにおける単語分散表現のコサイン類似度は高い傾向にあることが確認された。フォント分類のクラスを更に細分化することで、より顕著なフォント特徴を捉えることができ、同義語・対義語それぞれのペアにおける分散表現の類似度に大きな変化が見られることが期待される。

謝辞 本研究は JSPS 科研費 JP17H06100 の助成を受けた。

参考文献

- [1] Tianlang Chen, *et al.*, “Large-scale Tag-based Font Retrieval with Generative Feature Learning,” ICCV2019.
- [2] Tomas Mikolov *et al.*, “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.
- [3] Tomas Mikolov *et al.*, “Distributed Representations of Words and Phrases and Their Compositionality,” NeurIPS2013.
- [4] Masaya Ikoma *et al.*, “Effect of Text Color on Word Embeddings,” DAS2020.
- [5] Ashwin K Vijayakumar *et al.*, “Sound-word2vec: Learning word representations grounded in sounds,” arXiv preprint arXiv:1703.01720, 2017.
- [6] Youngmin Baek *et al.*, “Character Region Awareness for Text Detection,” ICCV2019.
- [7] Jeonghun Baek *et al.*, “What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis,” ICCV2019.
- [8] Kaiming He *et al.*, “Deep Residual Learning for Image Recognition,” arXiv:1512.03385, 2015.
- [9] Ankush Gupta *et al.*, “Synthetic Data for Text Localisation in Natural Images,” CVPR2016.
- [10] アイデア編集部, ABC のみほん:かたちで見分けるフォントガイド, 晋遊舎社, 2012.