

解釈可能な文字の視覚情報と意味情報を考慮した日本語の文書解析

青木, 匠 / AOKI, Takumi

(出版者 / Publisher)

法政大学大学院理工学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 理工学研究科編

(巻 / Volume)

63

(開始ページ / Start Page)

1

(終了ページ / End Page)

7

(発行年 / Year)

2022-03-24

(URL)

<https://doi.org/10.15002/00025367>

解釈可能な文字の視覚情報と意味情報を考慮した日本語の文書解析

JAPANESE TEXT ANALYSIS CONSIDERING VISUAL AND SEMANTIC INFORMATION OF INTERPRETABLE CHARACTERS

青木 匠
Takumi AOKI
指導教員 彌富 仁

法政大学大学院理工学研究科応用情報工学専攻修士課程

While Asian languages use characters with visual information such as radicals and shapes of kanji characters, Japanese also uses characters without visual information such as hiragana and katakana. To take these features of Japanese into account, a text analysis model that considers semantic information as well as visual information of characters has been proposed, and has achieved high performance in text analysis. However, while the visual information of the characters can be interpreted as radicals and shapes, the semantic information cannot be interpreted. In this paper, we propose a text analysis model that considers highly interpretable visual and semantic information. The visual and semantic information of the character is separated and stored in each dimension of the low-dimensional representation of the character to enhance the interpretability. In addition, we focus on the fact that the acquisition of the meaning of characters in the semantic information depends on the training data. By combining the visual and semantic information, we expect to improve the robustness of the acquisition of the meaning of characters and to achieve high performance in text analysis. In sentiment analysis, we were able to interpret the semantic information of characters such as negative, positive, and user information. The combination of visual and semantic information improved the sentiment analysis task by about 1% and the recognize textual entailment task by about 0.3% compared to using only one of the two.

Keywords : Natural Language Processing, Sentiment Analysis, Recognize Textual Entailment, Interpretability, Character Embedding

1. はじめに

日本の漢字および中国語の文字には、部首や文字形状などといった視覚情報から文字の意味を獲得できる。漢字や中国語には、部首という文字の意味に関する共通する情報がある。水の意味を示す“さんずい”の部首を持つ漢字は“海”や“池”などといった水に関する意味を持つなど、文字の視覚情報は文字の意味を理解するための重要な役割を持っている。こうした文字の特徴があることから、視覚情報を考慮することで日本語における文書解析性能が向上すると期待されている。

視覚情報を考慮した文書解析手法が提案され、日本語や中国語の特徴を活かした文書解析性能が向上したことが報告されている [1, 2]。文字を画像として扱い、convolutional auto-encoder (CAE) [3] で文字の低次元表現を学習することで、視覚情報である文字形状や部首を獲得できる。視覚情報を獲得するために画像を用いることで、画像処理分野のデータ拡張手法が導入できるため、容易に文書解析の頑健性が向上できる。文字画像の一部を隠すデータ拡張手法 [4] を適用することで、文書解析の性能が向上したことが報告されている [5]。しかし、日本語には平仮名、片仮名といった視覚情報が少な

い文字が存在するため、文脈といった意味情報も考慮する必要がある。また、漢字には似た文字形状や同じ部首でも、意味が異なる文字が存在するため視覚情報だけでは文字の意味を獲得することは難しい。例として、“土”と“土”は似た形状、“漢”と“海”は同じ部首を持つがどちらも異なる意味である。

日本語において、文字の視覚情報だけではなく意味情報も考慮したことにより、さらなる文書解析性能の向上を達成する [5-7]。文字の視覚情報の獲得と文書解析を同時に学習する end-to-end モデルで実現している。文字の視覚情報を獲得するモデルが、文字画像の部首に注意が向いていることが確認されている [6]。文字表現の特徴空間上で、同じ視覚情報かつ意味情報を持つ文字が存在していることが確認されている [5]。日本語の視覚情報を持たない文字や同じ視覚情報でも異なる意味を持つ文字に対しても、意味のある情報を獲得したことで文書解析の性能が向上する。しかし、高い文書解析性能を達成した一方で、文字の低次元表現の解釈性に改善の余地が残されている。

低次元表現の解釈性を向上させる観点で、各次元独立した特徴を持つ表現にする手法が挙げられる [8, 9]。画像処理分野で入力画像の特徴を分離して、低次元表

現の各次元に格納する β -variational auto-encoder (β -VAE) [10] を文字画像に適応することで、文字の視覚情報を各次元に格納できたことが報告されている [8]. β -VAE は低次元表現が事前分布である正規分布に従う制約 β を強めることで、各次元独立した低次元表現となる。文字表現のある一つの次元に摂動を付加することで、次元に格納されている視覚情報に関係する文字に変化するため、各次元に格納されている視覚情報を分析できる。さらに、各次元に視覚情報が格納されていることに着目し、意味のある文字に変換するデータ拡張手法が提案されたことで、容易に文書解析の頑健性を向上できる。事前学習された単語表現 [11] に対して行列分解を行い、特定の次元範囲ごとに意味カテゴリを付与することで、高い解釈性の単語表現が獲得できる [9]. 単語の解釈性を向上させるだけではなく、従来の単語表現よりも低次元な単語表現でも同等以上の高い文書解析性能を達成できる。このように文字および単語表現の解釈性を向上させることで、低次元表現で高い文書解析性能を達成し、低次元表現に格納されている情報を分析することができる。

本研究は、視覚情報と意味情報を考慮した文字の低次元表現の解釈性を向上させ、高い文書解析性能を持つモデルを提案する。提案するモデルは2つあり、高い解釈性の意味情報を考慮した文書解析、高い解釈性の視覚情報と意味情報を組み合わせた文書解析モデルである。高い解釈性の意味情報のみを考慮した手法は、VAE のエンコードと文書解析モデルを同時に学習する end-to-end モデルである。VAE の特性により各次元が独立した低次元表現を獲得できるため、文字表現から意味を分析可能である。文書解析モデルに VAE の低次元表現が正規分布に従う制約が付与されることになるため、文書解析性能の向上に期待する。高い解釈性の視覚情報と意味情報を考慮したモデルは、VAE による視覚情報の獲得、VAE のエンコードと文書解析モデルによる意味情報の獲得を同時に学習する end-to-end モデルである。ここで、文書解析に inputs する文字の低次元表現は視覚情報と意味情報を組み合わせたものとしている。視覚情報と意味情報を組み合わせることで、片方の情報を用いたときで予測できなかったデータに対しても正確に予測することで、高い文書解析性能を期待する。

2. 方法

本研究では、視覚情報と意味情報を格納した文字表現の解釈性を高めつつ、文書解析を行うモデルを提案する。文字の視覚情報および意味情報の特徴が分離され、文字表現の各次元独立に格納されることで解釈性を向上させている。また、意味情報の解釈性を高めつつ文書解析を行うモデルも本研究の提案手法である。提案する手法の全体図を図 1 に示す。入力テキストの各文字を画像として扱い、VAE のエンコードで視覚情報と意味情報を示す文字表現を獲得する。文字表現の視覚情報は VAE デコード、意味情報は文書解析モデルにより学習が行われる。ただし、視覚情報と意味情報を組み合わせた文字

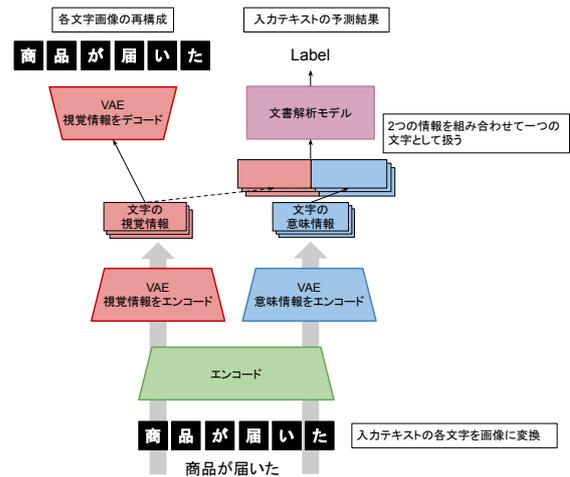


図 1 解釈性の高い視覚情報と意味情報を考慮した提案手法の全体図。

表現が一つの文字として文書解析が行われる。提案手法は大きく2つの構成が組み合わせており、視覚情報を考慮した文字表現の獲得、意味情報を考慮した文字表現と文書解析モデルを学習することである。以下の節でそれぞれを詳細に説明する。また、本研究では、文字表現の解釈性を定量的に評価するために情報理論を活用した評価指標を用いたため、本章で算出方法を説明する。

(1) 視覚情報を考慮した文字埋め込み表現を学習

文字を画像として扱い β -VAE で低次元表現を獲得することで、視覚情報を考慮しつつ、文字表現の各次元独立に部首や形状などの特徴を格納する。 β -VAE は画像処理分野の生成モデルとして提案された手法であるが、解釈性の高い低次元表現が獲得できるため、視覚情報を考慮した文字表現の獲得に使用されている [8].

(2) 文字の意味情報を獲得しつつ文書解析モデルを学習

入力テキストの各文字を画像化し、VAE のエンコードによる文字表現を獲得と文書解析を同時に end-to-end 学習することで、各次元が独立した意味情報を考慮した文字表現を獲得しつつ文書解析を行う。また、本研究では、文書解析モデルに attention 機構を用いた LSTM を使用し、文書解析結果の可視化を行う。

(3) 文字表現の解釈性の評価指標

本研究では、文字表現の解釈性を定量的に評価するために、情報理論を活用した手法 [12] を用いて“情報量”、“分離性”、“独立性”を算出する。

(4) 情報量

VAE による低次元表現は入力画像を圧縮したものであるため、入力画像と同じ情報を持つ必要がある。低次元表現 z の情報は入力画像 x と依存関係にあることがいえるため、

$$I(x, z) = H(x) - H(z|x) \quad (1)$$

と式ができる。 $I(x, z)$ が大きな値を取れば、入力画像と同じ情報を持つ低次元表現で、“情報量”が高いといえる。“情報量”の指標を文字表現に適応することで、文字

の意味が低次元表現に格納されていることが数値として確認することができる。

(5) 分離性と独立性

VAE は入力画像から特徴を抽出し、低次元表現の各次元独立に分離された特徴を格納する。分離された入力画像の情報が低次元表現の各次元に格納され、互いの次元同士で異なる情報を持つため、

$$I(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i}) = I(\mathbf{x}, (\mathbf{z}_i, \mathbf{z}_{\neq i})) - I(\mathbf{x}, \mathbf{z}_{\neq i}) \quad (2)$$

$$= I(\mathbf{x}, \mathbf{z}) - I(\mathbf{x}, \mathbf{z}_{\neq i}) \quad (3)$$

と式ができる。 i は低次元表現, $I(\mathbf{x}, \mathbf{z})$ は低次元表現の”情報量”, $I(\mathbf{x}, \mathbf{z}_{\neq i})$ は”分離性”を示す。 $I(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i})$ が大きな値を取れば, ”情報量”があり各次元に分離された特徴を持つ低次元表現で, ”分離性”が高いといえる。さらに, 各次元に格納されている分離された特徴が互いの次元同士独立していることを評価する場合,

$$\hat{I}(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i}) = I(\mathbf{x}, \mathbf{z}_i) - I(\mathbf{z}_i, \mathbf{z}_{\neq i}) \quad (4)$$

$$= I(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i}) - I(\mathbf{z}_i, \mathbf{z}_{\neq i} | \mathbf{x}) \quad (5)$$

と式ができる。 $I(\mathbf{x}, \mathbf{z}_i)$ は”情報量”, $I(\mathbf{z}_i, \mathbf{z}_{\neq i})$ は”独立性”を示す。 $\hat{I}(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i})$ が大きな値を取れば, ”情報量”があり各次元に独立した特徴を持つ低次元表現で, ”独立性”が高いといえる。しかしながら, $I(\mathbf{x}, \mathbf{z}_i)$ などを計算する上で, 低次元表現の特定の次元に”情報量”がないノイズとなる値が存在するため, 以下の重みを $I(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i})$ および $\hat{I}(\mathbf{x}, \mathbf{z}_i | \mathbf{z}_{\neq i})$ に追加することでノイズに対応する。ここで, L は低次元表現の次元数を示す。

$$\rho_i = \frac{I(\mathbf{x}, \mathbf{z}_i)}{\sum_{j=0}^{L-1} I(\mathbf{x}, \mathbf{z}_j)} \quad (6)$$

”分離性”および”独立性”の指標を文字表現に適応することで, 各次元に部首構造などの視覚情報や感情などの意味情報が分離され独立に格納されていることが数値として確認することができる。

3. 評価実験

本提案手法である解釈性の高い視覚情報と意味情報を考慮した文書解析, 解釈性の高い意味情報を考慮した文書解析モデルを日本語の文書を用いて, 文書解析性能を評価する実験を行った。比較手法として, 先行研究の解釈性の低い意味情報を考慮した文書解析 [5] と解釈性の高い視覚情報を考慮した文書解析 [8] を用いた。

(1) データセット

文字の意味情報として感情情報を獲得することに期待できる感情分析を行うために The Multilingual Amazon Reviews Corpus Japanese*1 を使用した。レビュー文からポジティブまたはネガティブを分類する 2 クラス分類のタスクである。元のデータセットはレビュー文に対してレビュースコア 1 から 5 が付与されているが, レビュースコアが 2 以下はネガティブ, 4 以上はポジティブとしてクラスを作成した [13]。訓練データは 32,000

件, 評価データは 5,000 件, テストデータは 5,000 件である。訓練データは元のデータ数が 200,000 件と大規模なデータなため, 全体の 20% を使用した。また, 各データは事前に分割されていたものを使用した。前処理として, 絵文字の削除を行った。

文字と文全体の意味を理解する必要のある含意認識タスクを行うために, 株式会社リクルートが提供する Japanese Realistic Textual Entailment Corpus*2 を用いた。仮説文 (Hypothesis) と前提文 (Premise) の 2 つ文があり, 仮説文から前提文の意味が言えるのかを評価するタスクで, 言える場合は含意する, いえない場合は含意しないの 2 クラス分類である。例として, 仮説文が”部屋から海が見える”, 前提文が”部屋はオーシャンビューで景色がよかったです”があるとする。前提文の”オーシャンビュー”から部屋から海が見えることがいえるため, 含意するとなる。訓練データは 19,602 件, 評価データは 5,602 件, テストデータは 2,831 件である。各データは事前に分割されていたものを使用した。

(2) 実験設定

入力文字画像は 32×32 pixel とし, 視覚情報を獲得するために VAE を使用した。テキストの可変長入力に対応するために LSTM を用いた。さらに, 文書解析で重要な文字に注目し解釈可能にするために attention を追加した。含意認識タスクは仮説文と前提文の 2 文入力であるため, それぞれの入力する LSTM を用意し, 2 つの出力結果を結合させ分類を行う。文字表現の獲得および文書解析を学習する際に, 最適化手法の adam [14] を用いて学習率は $1e-4$ とした。さらに, early stopping を 30 epoch に設定して, 評価データに対する損失関数が低下しない場合は学習を打ち止めた。

(3) 文書解析性能の評価

制約項の beta が 1 のとき, 感情分析と含意認識タスクのテストデータにおける正解率を表 1, 2 に示す。

+ 文書解析モデル	Accuracy [%]	
	10 次元	20 次元
意味情報	83.15	85.06
解釈性の高い視覚情報	83.77	85.43
解釈性の高い意味情報 (Ours)	86.30	86.89
解釈性の高い視覚情報と意味情報 (Ours)	86.18	87.82

表 1 感情分析タスクのテストデータにおける正解率。

+ 文書解析モデル	Accuracy [%]	
	10 次元	20 次元
意味情報	72.61	75.87
解釈性の高い視覚情報	72.58	75.40
解釈性の高い意味情報 (Ours)	75.33	76.56
解釈性の高い視覚情報と意味情報 (Ours)	75.16	76.96

表 2 含意認識タスクのテストデータにおける正解率。

文字表現が 10 次元のときは解釈性の高い意味情報を考慮した文書解析, 20 次元のときは解釈性の高い視覚情

*1 https://huggingface.co/datasets/amazon_reviews_multi

*2 <https://github.com/megagonlabs/jrte-corpus#references>

報と意味情報を組み合わせた文書解析が最も高い正解率となった。また、文字表現の次元数は 10 次元よりも 20 次元の方がどのモデルにおいても高い精度となった。

(a) Attention による文書解析結果の可視化

感情分析タスクにおいて、文字表現が 20 次元で beta=1 のときの attention による予測に重要な文字を可視化した結果を図 2 に示す。視覚情報と意味情報の片方および両方を用いた文書解析でも、感情に関係する文字に文書解析モデルが注目した。しかしながら、同じ文字に着目しているにも関わらずに、意味情報を考慮した文書解析で間違った結果があった。

含意認識タスクにおいて、文字表現が 20 次元で beta=1 のときの attention による予測に重要な文字を可視化した結果を図 3 に示す。視覚情報と意味情報の片方および両方を用いた文書解析でも、含意関係を認識する上で重要な文字に文書解析モデルが注目した。しかしながら、同じ文字に着目しているにも関わらずに、視覚情報を考慮した文書解析で間違った結果があった。

正解ラベル: Negative 予測ラベル: Negative
 以前楽天でこの画像と同じ商品購入しました。とても品質が良かったですが、少し高くて…今回アマゾンで安かったので同じ商品と思ったら全然違います。食品直接届けると説明書いていましたが、パッケージを開けるとシリコンにゴミがいっぱいついてあります。又透明のムラがひどいです。なによりもシンナーのような匂いがキツ過ぎて頭が痛くなりました…怖いです

(a) 解釈性の高い視覚情報

正解ラベル: Negative 予測ラベル: Positive
 以前楽天でこの画像と同じ商品購入しました。とても品質が良かったですが、少し高くて…今回アマゾンで安かったので同じ商品と思ったら全然違います。食品直接届けると説明書いていましたが、パッケージを開けるとシリコンにゴミがいっぱいついてあります。又透明のムラがひどいです。なによりもシンナーのような匂いがキツ過ぎて頭が痛くなりました…怖いです

(b) 解釈性の高い意味情報

正解ラベル: Negative 予測ラベル: Negative
 以前楽天でこの画像と同じ商品購入しました。とても品質が良かったですが、少し高くて…今回アマゾンで安かったので同じ商品と思ったら全然違います。食品直接届けると説明書いていましたが、パッケージを開けるとシリコンにゴミがいっぱいついてあります。又透明のムラがひどいです。なによりもシンナーのような匂いがキツ過ぎて頭が痛くなりました…怖いです

(c) 解釈性の高い視覚情報と意味情報

図 2 感情分析タスクにおいて、解釈性の高い文字の視覚情報と意味情報を片方および両方を考慮した文書解析で視覚情報を用いたときに間違った予測したときの attention による可視化結果。

(4) 文字表現の評価

文字表現が正規分布に従う制約を変化させたときに、最も解釈性の高い文字表現を評価する。

(a) 解釈性の定量的な評価

“情報量”は文字表現に文字の意味が格納されていることを示す。文字表現が正規分布に従う制約を変化させたとき、感情分析タスクで獲得する文字表現における“情報量”を図 4, 5 に示す。制約を強めるほど、文字表現の情報量は少なくなった。視覚情報は意味情報よりも、制約に大きな影響を受けていることが確認された。どの文字情報を考慮した場合でも、文字表現が 10 次元で beta=1 の時に最も高い“情報量”を持つことを確認した。

“分離性”は文字表現の各次元に文字情報が分離されていることを示す。文字表現が正規分布に従う制約を変化させたとき、感情分析タスクで獲得する文字表現における“分離性”を図 6 に示す。文字情報における視覚情

正解ラベル: Non-entailment 予測ラベル: Entailment
 premise: 鮑が美味しかったです。

hypothesis: 金目鯛の煮付けが美味しかったです。

(a) 解釈性の視覚情報

正解ラベル: Non-entailment 予測ラベル: Non-entailment
 premise: 鮑が美味しかったです。

hypothesis: 金目鯛の煮付けが美味しかったです。

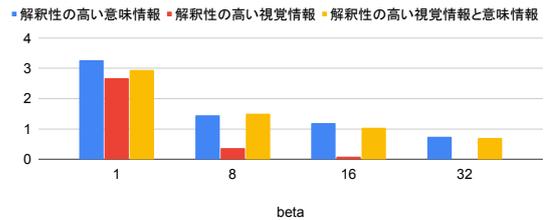
(b) 解釈性の高い意味情報

正解ラベル: Non-entailment 予測ラベル: Non-entailment
 premise: 鮑が美味しかったです。

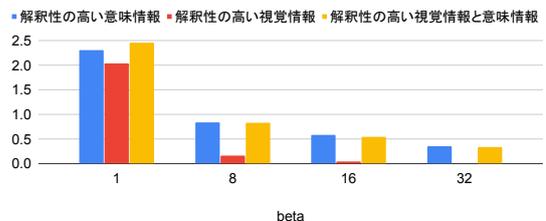
hypothesis: 金目鯛の煮付けが美味しかったです。

(c) 解釈性の高い視覚情報と意味情報

図 3 含意認識タスクにおいて、解釈性の高い文字の視覚情報と意味情報を片方および両方を考慮した文書解析で視覚情報を用いたときに間違った予測したときの attention による可視化結果。



(a) 10 次元

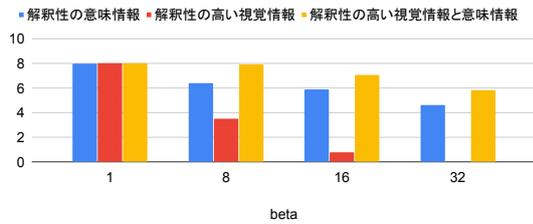


(b) 20 次元

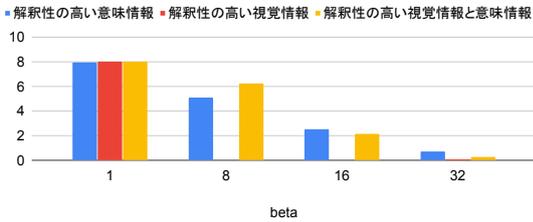
図 4 感情分析タスクで獲得する文字表現が 10, 20 次元のとき、視覚情報と意味情報の片方および両方を考慮した文字表現の各次元ごとの情報量を平均した値。

報は beta=8、意味情報は beta=32、視覚情報と意味情報を組み合わせた情報は beta=32 が最も高い“分離性”を持つことが確認できた。

“独立性”は文字表現の各次元に格納されている文字情報が独立していることを示す。文字表現が正規分布に従う制約を変化させたとき、感情分析タスクで獲得する文字表現における“独立性”を図 7 に示す。文字情報における視覚情報は beta=8、意味情報は beta=32、視覚情報と意味情報を組み合わせた情報は beta=32 が最も高い“独立性”を持つことが確認できた。



(a) 10次元



(b) 20次元

図5 感情分析タスクで獲得する文字表現が10, 20次元のとき、視覚情報と意味情報の片方および両方を考慮した文字表現の情報量。

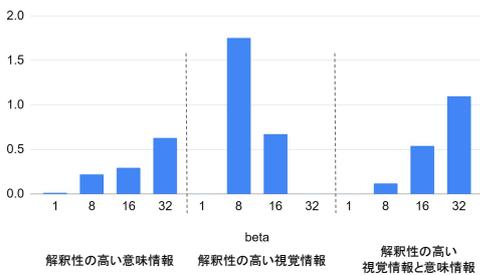


図6 感情分析タスクで獲得する文字表現が10次元のとき、視覚情報と意味情報の片方および両方を考慮した文字表現の情報量を考慮した分離性。

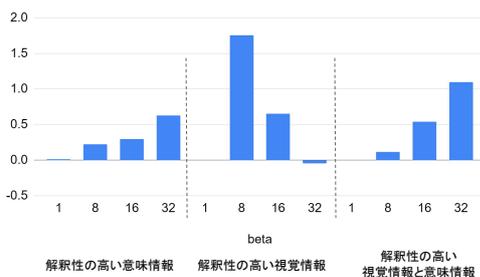


図7 感情分析タスクで獲得する文字表現が10次元のとき、視覚情報と意味情報の片方および両方を考慮した文字表現の情報量を考慮した独立性。

(5) 近傍法による文字表現の分析

文字表現の特徴空間上近くに存在する文字はで同じ意味を持つことを示すことため、k-近傍法により近傍文字を求めた。図5で最も高い情報量を持つ10次元でbeta=1のときの文字表現を使用した。

(a) 各次元に格納されている情報

感情分析タスクにおいて、情報量を考慮した“分離性”および“独立性”で最も高いスコアだった次元数とbeta

の値のときの文字表現を用いて、各次元に格納されている情報を分析した。解釈性の高い視覚情報は10次元でbeta=8、解釈性の高い意味情報は10次元でbeta=32のときである。ある特定の次元に対して摂動を付加した後の文字表現を対象として、k近傍法を用いて変化した文字の意味を確認した。視覚情報で行った近傍結果を図3に示す。

対象文字	近傍文字	
疑	摂動前	●, 浮, 擊, 券, 武, 唇, 般, 藥, 索, 爆
	4次元目に+3	遡, 客, 彫, 腐, 仇, 募, 雲, 誕, 牢, 裏
	5次元目に+3	浅, 竿, 涙, 貧, 金, 籠, 珙, 嵩, 受, 4
嬉	摂動前	川, 逮, 歲, 整, m, 姉, 進, 遮, 殺, 慰
	3次元目に-3	謎, 丸, 逮, 溪, 歲, 忘, 蠟, 飛, 嬉, 麟
	8次元目に+3	搦, 捲, 嬉, 探, 抒, 摺, 搔, 襪, 批, 援

表3 解釈性の高い視覚情報を考慮した文字表現の特定の次元に摂動を付加した文字を対象にしたときの上位20件の近傍文字。

視覚情報において、“疑”の摂動付加前は文字形状の情報だったのが、4次元目に摂動を付加した後は画数の多い文字形状、5次元目に摂動した後は“さんずい”に関係する文字になった。“嬉”の摂動付加前は“女”へんや文字形状の情報だったのが、3次元目に摂動を付加した後は“ごんべん”、8次元目に摂動を付加した後は“てへん”に関係する文字になった。文字形状や部首などの視覚情報が分離され各次元に格納されていることが確認できた。意味情報で行った近傍結果を図4に示す。

対象文字	近傍文字	
疑	摂動前	途, 悔, 陥, 詐, 票, ぼ, 熔, 折, km, 怪
	4次元目に+3	怪, 毯, 誰, 疑, 悔, 局, 盜, 騙, 不, 凹
	7次元目に+3	駄, 粗, 隈, 怪, 悔, 毯, 檜, 頭, 途, 衰
嬉	摂動前	旬, 暮, 喜, 嬢, 幸, 遽, 馴, 躍, 巧, 醬
	2次元目に+3	幸, 暮, 嬉, 那, 馴, 旬, 娠, 沸, 術, 申
	5次元目に-3	沸, 麗, 嬉, 喜, 嬌, 畏, 申, 晴, 癒, 議

表4 解釈性の高い意味情報を考慮した文字表現の特定の次元に摂動を付加した文字を対象にしたときの上位20件の近傍文字。

意味情報において、ネガティブ文字の“疑”の摂動付加前は、単なるにネガティブ情報だったのが、4次元に摂動を付加した後は凹むを示す“凹”や壊れるを示す“壊”といった商品の不備に関するネガティブな情報、7次元に摂動を付加した後は無駄を示す“駄”や後悔を示す“悔”といった商品を買わなければよかった感情のネガティブな情報に関する文字になった。ポジティブ文字の“嬉”の摂動付加前は、単なるポジティブ情報だったのが、2次元目に摂動した後は旦那を示す“那”や一人暮らしなどを示す“暮”といった感情に関するユーザー情報、5次元に摂動した後は素晴らしいを示す“晴”や愛用を示す“愛”といった商品に対する満足感に関する文字になった。ネガティブ、ポジティブなどとい

た感情情報や感情に関係するユーザー情報が分離され、各次元に格納されていることが確認できた。

4. 考察

(1) 文字表現と文書解析性能の関係性

意味情報における解釈性の有無で言語処理性能に違いがあるのか考察する。表 1, 2 により、意味情報を持つ文字表現において、解釈性の低い手法よりも高い手法のほうが、文字表現が 10 次元のときに感情分析タスクおよび含意認識タスクで約 3% 高い結果となった。解釈性のある手法は文字表現が正規分布に従う制約があり、言語処理モデルの過剰適合を防いでいると考えられる。

解釈性の高い視覚情報と意味情報を比較し、文書解析に与える影響を考察する。表 1, 2 により、解釈可能な文字表現において、文字表現が 10 次元にときに視覚情報よりも意味情報のほうが感情分析および含意認識タスクで 3% 高い結果となった。含意認識タスクの予測を可視化した図 3 から、含意関係に重要な“鯛”と“鮑”に着目しているにもかかわらず予測結果は間違っているため、視覚情報よりも意味情報のほうが文書解析性能に大きな関係があると考えられる。

本提案手法の解釈性の高い意味情報を考慮した手法と解釈性の高い視覚情報と意味情報を考慮した手法を比較する。図 1, 2 により、文字表現が 10 次元においては、解釈性の高い視覚情報と意味情報を結合させた文書解析よりも意味情報を考慮した文書解析が感情分析および含意認識タスクで約 0.2% 僅かに高い結果となった。視覚情報と意味情報を考慮した手法は、視覚情報と意味情報にそれぞれ 5 次元に格納されており、高い言語処理性能を実現するには不十分な情報だと考えられる。

すべての手法と文字表現の次元数を比較して、解釈性のある視覚情報と意味情報を考慮した手法が最も高い精度となった。視覚情報と意味情報を同時に扱うことで、視覚情報による文字の意味獲得に頑健性と意味情報の高い文書解析性能を達成するといった互いの情報を扱うことで得られることを組み合わせたためだと考えられる。感情分析と含意認識タスクにおける予測の可視化を示す図 2, 3 から、視覚情報と意味情報のどちらか一方の情報を扱ったときに間違った文書でも 2 つを扱うことで正解することができているためである。

(2) 文字表現の解釈性

表 4 により、摂動前の対象文字の近傍はネガティブやポジティブな情報を持つ文字となった。さらに、文字表現のある特定の次元に対して摂動を付加した後に、特徴空間上で近くに存在する文字を分析した結果、ネガティブにおける商品の不備や購入した後悔、ポジティブにおける満足感、感情に関係するユーザー情報に関する文字を取得することができた。感情分析における意味情報とはネガティブやポジティブ情報で、商品の不満や後悔、満足感などの意味があることが解釈することができた。さらに、意味情報が分離されたことで、感情に関係するユーザー情報も取得していることが確認できたといえる。

表 3 により、摂動前の対象文字の近傍は同じ部首や文字形状などといった持つ文字となった。さらに、文字表現のある特定の次元に対して摂動を付加した後に、特徴空間上で近くに存在する文字を検索した結果、異なる部首や文字形状の文字が得られた。部首を捉えているだけではなく、文字構造や“てへん”などといった部首の種類も捉えていると解釈することができる。先行研究も同様な結果を得られており、文字の視覚情報とは部首や文字形状だといえる。

5. まとめ

本研究では、解釈可能な意味情報を考慮した文書解析手法を提案した。文字の意味情報を分離し文字の低次元表現の各次元に格納することで解釈性を高めている。感情分析タスクによる実験結果により、文字の意味情報としてネガティブ、ポジティブの情報だけではなく、意味情報を分離したことにより、感情の要因となった情報が意味情報も格納していることが確認した。さらに、従来の解釈性の低い意味情報を考慮した手法よりも、約 3% の文書解析性能を達成した。また、本研究では、解釈性の高い視覚情報と意味情報を考慮した手法を提案した。2 つの情報を比較しつつ実験を行った結果、視覚情報は文字の意味獲得に頑健性があるが高い文書解析性能を実現できない、意味情報は高い文書解析性能を実現できるが文字の意味獲得は学習データに依存するため過剰適合することを考察した。感情分析および含意認識タスクによる実験結果から、2 つの情報を組み合わせることで片方の情報を用いた手法で予測が間違ったデータに対しても正解したことで約 1% 文書解析性能を向上することができた。

本研究分野の日本語の特徴を生かした文書解析手法には、文字の視覚情報に大きな改善の余地が残されている。画像処理分野で畳み込み層を使用せずに入力画像の特徴を捉え最先端の精度を達成した Vision Transformer (ViT) [15] が提案されている。ViT を文字画像に適用することで畳み込み層では取得できない視覚情報を獲得することで、高い文書解析性能を達成することに期待できる。

参考文献

- [1] F. Dai and Z. Cai, “Glyph-aware embedding of Chinese characters,” in *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, (Copenhagen, Denmark), pp. 64–69, Association for Computational Linguistics, 2017.
- [2] D. Shimada, R. Kotani, and H. Iyatomi, “Document classification through image-based character embedding and wildcard training,” in *2016 IEEE International Conference on Big Data (Big Data)*, pp. 3922–3927, 2016.
- [3] J. Masci, U. Meier, D. Cireşan, and J. Schmidhuber, “Stacked convolutional auto-encoders for hierarchical feature extraction,” in *Artificial Neu-*

- ral Networks and Machine Learning – ICANN 2011* (T. Honkela, W. Duch, M. Girolami, and S. Kaski, eds.), pp. 52–59, Springer Berlin Heidelberg, 2011.
- [4] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, pp. 13001–13008, 2020.
- [5] S. KITADA, R. KOTANI, and H. IYATOMI, “End-to-end text classification via image-based embedding using character-level networks,” in *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1–4, 2018.
- [6] F. Liu, H. Lu, C. Lo, and G. Neubig, “Learning character-level compositionality with visual features,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, (Vancouver, Canada), pp. 2059–2068, Association for Computational Linguistics, 2017.
- [7] M. Daif, S. Kitada, and H. Iyatomi, “AraDIC: Arabic document classification using image-based character embeddings and class-balanced loss,”
- [8] T. Aoki, S. Kitada, and H. Iyatomi, “Text classification through glyph-aware disentangled character embedding and semantic sub-character augmentation,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, (Suzhou, China), pp. 1–7, Association for Computational Linguistics, 2020.
- [9] K.-T. Liao, C.-S. Lee, Z.-Y. Huang, and S.-d. Lin, “Explaining word embeddings via disentangled representation,” in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, (Suzhou, China), pp. 720–725, Association for Computational Linguistics, 2020.
- [10] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework,” *In Proc. of ICLR*, 2017.
- [11] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Doha, Qatar), pp. 1532–1543, Association for Computational Linguistics, 2014.
- [12] K. Do and T. Tran, “Theory and evaluation metrics for learning disentangled representations,” in *International Conference on Learning Representations*, 2020.
- [13] P. Keung, Y. Lu, G. Szarvas, and N. A. Smith, “The multilingual Amazon reviews corpus,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 4563–4568, Association for Computational Linguistics, 2020.
- [14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” in *International Conference on Learning Representations*, 2021.