
一般セッション | 一般セッション | [GS] J-9 自然言語処理・情報検索

[2L4-J-9] 自然言語処理・情報検索: 変換と生成

座長: 小林 一郎(お茶の水女子大学) 評者: 岡嶋 穰(NEC)

2019年6月5日(水) 15:20 ~ 16:40 L会場 (203+204 小会議室)

[2L4-J-9-01] LSTM-RNNを用いた文体変換手法

○下地 健太¹、森田 和宏¹、泓田 正雄¹ (1. 徳島大学大学院)

15:20 ~ 15:40

[2L4-J-9-02] 新聞記事データを用いたテキスト平易化

小藤 直紀¹、○難波 英嗣¹、竹澤 寿幸¹ (1. 広島市立大学)

15:40 ~ 16:00

[2L4-J-9-03] キーワード条件つき変分 Autoencoderによる広告文生成

○福田 宏幸¹ (1. 株式会社 電通)

16:00 ~ 16:20

[2L4-J-9-04] 条件付き変分オートエンコーダと単語分散表現による絵文字の生成

○山口 篤季¹、藤田 桂英¹ (1. 東京農工大学)

16:20 ~ 16:40

LSTM-RNN を用いた文体変換手法

A sentence style conversion method using LSTM-RNN

下地 健太^{*1}
Kenta Shimoji

森田 和宏¹
Kazuhiro Morita

泓田 正雄¹
Masao Fuketa

¹ 徳島大学大学院先端技術科学教育部
Department of information Science and Intelligent Systems, Tokushima University

Abstract: This paper describes a sentence style conversion method using recurrent neural network with long short-term memory cells (LSTM-RNN). In the proposed method, LSTM-RNN is used to learn direct style sentences vectorized by one-hot expressions. Then, the sentence end expression of the distal style sentence is removed and the vectorized one is input into the learned model. The next word is predicted until sentence ends, and the obtained word vector sequence is added to the end of the input vector sequence. A direct style sentence is converted by decoding the generated vector sequence into the form of the natural language. We experimented to evaluate the accuracy of the proposed method. As a result, it turned out that a sentence style can be converted by the method.

1. はじめに

近年、文書の電子化が進んでおり、コンピュータを使った文書作成の機会が増加している。電子媒体の文書にもアナログ媒体のものと同様にヒューマンエラーがしばしば見られる。その一つとして文体の統一に関する誤りが挙げられる。さらに、文体の統一の中でも多くの誤りが発生するのが、同一文書における常体文(「だ」「である」調)と敬体文(「です」「ます」調)の混在である。文書作成においてこのような誤りを訂正する作業は必要不可欠である。しかし、文書の作成者が自ら訂正作業をおこなうことは、その内容を理解しているため、客観性に欠ける。また、人手による訂正作業では文書作成と同様に一つの見落としも無いという保証はなく、文書が長くなればなるほどヒューマンエラーは増えていく。そこで、コンピュータによる文体変換システムが必要とされる。本稿では LSTM-RNN を用いた文体変換手法について述べる。また、提案手法の変換精度評価のために実験をおこなう。

林らの研究[1]では、人手で作成した変換ルールを用いた手法で書き言葉から話し言葉への文体変換をおこなっている。しかし、この手法ではルールでの対応が困難な場合があり柔軟な変換がおこなえないことが問題であった。本研究ではルールでは難しい場合に対応するために、LSTM-RNN を用いることにより、それぞれの文体に特有の表現や単語の並びを学習させ、未知の入力に対し生成した学習モデルの単語予測を利用して文体の変換をおこなう。

2. LSTM-RNN

RNN(Recurrent Neural Network)は時系列データの学習に特化した深層学習手法の一つである。時系列データとは、ある要素が時間的順序を追って並んでいるデータのことであり、文章、声、映像、株価情報等がこれに当たる。そのため、主に自然言語処理や音声言語処理等の分野で利用されている。内部にループ構造を持つことが特徴であり、これにより前のステップの情

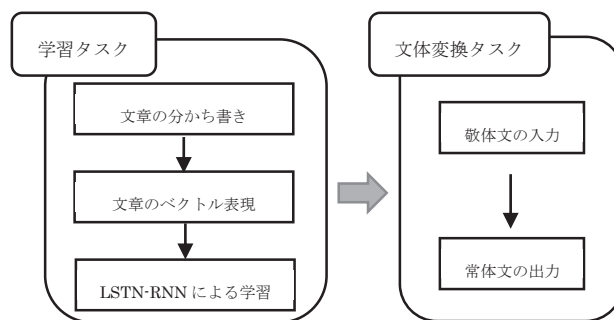


図 1: 提案手法

報を後続のステップに渡すことが可能となっている。また、本研究では RNN に加え、LSTM(Long Short-Term Memory)[2]を使用する。LSTM は学習に必要な情報の取捨選択をおこなうことにより、時系列データの長期依存性を学習することを可能にする RNN のユニットの一つである。LSTM を中間層に用いることにより、通常の RNN が持つ逆誤差伝播による学習における勾配消失に問題を解決し、より高い精度で時系列予測をおこなうことができる。本研究では、入力した文字列の次に来るであろう単語を予測するために RNN を用い、自然言語という比較的データ長の長い時系列データを扱うため、LSTM を採用する。本稿では LSTM を中間層に用いた RNN を LSTM-RNN と呼ぶこととする。

3. 提案手法

学習データに常体文のみを使用することにより、常体文の単語の並びを予測することができるので、これを利用して敬体文を常体文に変換する。これに関する提案手法の流れを図 1 に示す。また、本章では提案手法のそれぞれのステップについて解説する。

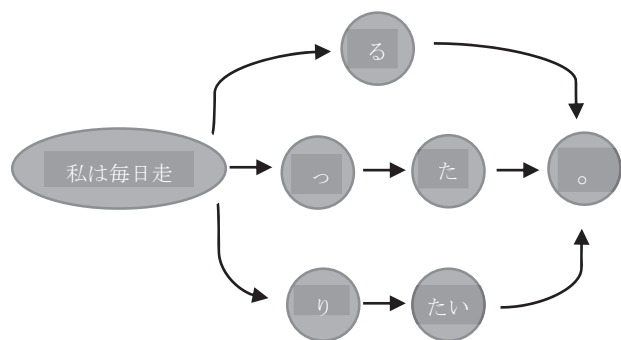


図 2:文末予測の例

3.1 学習タスク

本節では文章を学習しモデルを生成するタスクについて解説する。

まず、学習データには常体文のみを用いるため、常体文のみを収集する必要がある。敬体文には助動詞「です」、「ます」のどちらかが使われることが主であるため、これらの助動詞が含まれない文という条件で常体文のみを収集した。そして、学習に用いる文章に対し形態素解析をおこない、分かち書きをおこなう。ここで、動詞については、「走る」、「走り(ます)」のように同じ意味を表していても敬体文と常体文で活用が異なる場合があるため、形態素をさらに語幹で分割する。前述の例の場合、「走/る」、「走/り」のように分割する。形態素解析には MeCab[3]を使用し、語幹はその動詞の原形と前方一致する部分とする。その後、One-hot 表現を用いて単語のベクトル化をおこなう。One-hot 表現とは、1つのビットにだけ 1(High)を、他のビットには 0(Low)を割り当てるベクトル表現である。これにより、文章を単語ベクトルの集合として表現する。最後に、ベクトル化した文章について 1文毎に LSTM-RNN を用いて単語の順番を学習し、モデルを生成する。

3.2 文体変換タスク

本節では入力文を 3.1 節で生成した学習モデルにより文体変換をおこなうタスクについて解説する。

敬体文から常体文への変換が目的であるため、文体変換時の入力データには敬体文を用いる。3.1 節と同様にして入力となる敬体文に対し分かち書きをおこなった後、文末表現を除去する。文末表現は文の最後に現れる名詞もしくは動詞の語幹の次の文字から句点までとする。例えば、「私は毎日走ります。」という文の文末表現は「ります。」なので、除去後の文字列は「私は毎日走」となる。そして、これに対しベクトル化をおこない 3.1 節で生成した学習モデルに入力する。次に来る確率の高い文末語までの単語の組み合わせを複数生成し、その中で最も文全体としての生成確率が最も高いものを出力とする。例の場合、「私は毎日走」に続く単語の組み合わせとして、「る。」「った。」「りたい。」が候補となり、その中から「る。」が出力として選ばれる。(図 2 参照)これを元の入力文の末尾に追加し、できたベクトル列を自然言語の形に変換することで常体文を得る。前述の例では「私は毎日走る。」という常体文が得られる。

表 1:実験結果

| | ○ | △ | × |
|-----|----|----|----|
| 分類数 | 23 | 59 | 18 |

表 2:変換例

| 入力文 | 出力文 | 評価 |
|---------------------|--------------------|----|
| 実状の説明に適しているかも知れません。 | 実状の説明に適しているかも知れない。 | ○ |
| まじめな顔をして言いました。 | まじめな顔をして言った。 | ○ |
| 笑わせるのには事を欠きませんでした。 | 笑わせるのには事を欠いた。 | △ |
| まるで見当つかないのです。 | まるで見当つかないのか。 | △ |
| 完全に落第でした。 | 完全に落第る。 | × |
| これがまた家中の大笑いでした。 | これがまた家中の大笑いがある。 | × |

4. 評価実験

提案手法の精度を確認するために評価実験をおこなった。学習データには、読売新聞記事内の常体文を 26,366 文使用した。テストデータには青空文庫[4]の作品内の敬体文を 100 文使用し、提案手法を用いて文体変換をおこなった。本実験では埋め込み層、2つの LSTM 層と全結合層からなるニューラルネットワークを使用した。LSTM 層の次元数は 200 次元とし、損失関数には活性化関数には softmax 関数を用いた。評価は○(正しい)、△(常体に変換できているが元の敬体文の意味情報が保持されていない)、×(文の意味が通っていない)の三段階評価を人手によりおこなった。

評価実験の結果を表 1 に示す。8 割以上が意味に破綻のない文として出力された。本実験における変換例を表 2 に示す。評価が△のものについて、1 つ目の「笑わせるのには事を欠きませんでした。」は「笑わせるのには事を欠かなかった。」と変換されるべきであるが、提案手法では学習した単語の順番を元に予測をおこなうため、変換の過程で否定文であるという情報が失われている。また、2 つ目についても同様に肯定文であるという情報が失われている。この問題を解決するには、入力文の種類が何であるかを判定し、それと同様の種類の出力文を生成する必要があると考える。×のものについては、学習データの不足が原因であると考えられるため、十分な量のコーパスを用意することが必要である。

5. おわりに

本稿では、LSTM-RNN を用いた文体変換手法の提案をおこなった。100 の敬体文に対し本手法を適用することにより、精度評価をおこなった。その結果 8 割以上が意味に破綻がない文として出力されたが、変換の過程で入力文の意味情報が失われる場合があるという課題が明らかになった。

今後は今回の結果を踏まえ、課題に関して改良をおこなうことで変換精度を向上させていく予定である。

参考文献

- [1] 林由紀子, 松原茂樹. 自然な読み上げ音声出力のための書き言葉から話し言葉へのテキスト変換. 情報処理学会研究報告, pp. 49-54, 2007
- [2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [3] T. Kudo. Mecab : Yet another part-of-speech and morphological analyzer. <http://taku910.github.io/mecab/>
- [4] 青空文庫. <https://www.aozora.gr.jp/>

新聞記事データを用いたテキスト平易化

Text simplification using newspaper articles

小藤 直紀
Naoki KOTO

難波 英嗣
Hidetsugu NANBA

竹澤 寿幸
Toshiyuki TAKEZAWA

広島市立大学 情報科学部
School of Information Sciences, Hiroshima City University

Automatic text simplification attempts to automatically transform complex sentences into their simpler variants without significantly changing the original meaning. Several researches on automatic text simplification have conducted based on a large-scale monolingual parallel corpus. However, it is costly to manually construct a parallel corpus for text simplification. Therefore, we investigate automatic construction of a large-scale simplified corpus for Japanese from newspaper database corpora. In this paper, we examined several methods for sentence alignment of texts with different complexity levels. Using the best of them, we sentence-align the Mainichi newspaper and Mainichi newspaper for elementary students, thus providing large training materials for automatic text simplification systems.

1. はじめに

現在、日本には 200 万人以上の在留外国人や、旅行者など多種多様な人種、国籍の外国人が日本に滞在している。近年、そのような人々に向けて「やさしい日本語」(<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ1a.htm>)というものが考案されている。「やさしい日本語」とは、外国人にもわかりやすい日本語であり、地震などの災害が発生した際の情報伝達や、行政窓口の外国人対応など様々な状況で扱われている。難解な日本語を平易化し、情報伝達を容易にする「テキスト平易化」は、今後さらに重要なタスクになると考えられる。そこで本研究では、日本語テキストを平易化する手法を提案する。

近年、テキスト平易化を同一言語内の翻訳問題と考え、機械翻訳の枠組みで入力文から平易な同義文を生成する研究が盛んに行われている。その代表的な手法に、難解なテキストと平易なテキストからなる大規模な単言語パラレルコーパスを用いたテキスト平易化が挙げられる。しかし、その多くは英語で作成された大規模パラレルコーパスを用いた手法である。そこで、本研究では毎日新聞と毎日小学生新聞の記事データをそれぞれ難解なテキスト、平易なテキストとして対応付けを行い、大規模パラレルコーパスを作成する。それを用いて文の平易化を行うことにより、円滑な情報提供の手助けをすることが本研究の目的である。

論文の構成は以下のとおりである。2 章では関連研究を紹介する。3 章ではテキスト平易化システムの構築について述べる。4 章では評価実験とその結果、考察を述べる。5 章で本論文をまとめる。

2. 関連研究

本研究の関連研究として、2.1 節では、やさしい日本語について、2.2 節では、英語のパラレルコーパスによるテキスト平易化について、2.3 節では、平易なコーパスを用いないテキスト平易化について、2.4 節では、SCDVを用いた文書分類について、それぞれ述べる。

2.1 やさしい日本語

近年、外国人にも理解が容易な「やさしい日本語」を用いる取り組みが盛んであり、災害時の緊急連絡や、マニュアルなどに用いることで被害の減少が見込める。原文からやさしい日本語へ変換する技術に関して、例えば梶原らが作成した語彙平易化システム[梶原 17]などが挙げられる。

Maruyama らは日本語を対象にした自動平易化に用いる言語資源が少ないことを挙げ、「やさしい日本語」という観点で日本語に関する言語資源構築を行った[Maruyama 18]。原テキストには small parallel enja: 50k En/Ja Parallel Corpus for Testing SMT Methods が用いられている。Maruyama らはこのコーパス内にある 5 万文すべてに対して、2,000 語の平易な語で構成される語彙集合「やさしい語彙」を用いて文単位の書き換え、やさしい語彙の追加、削除を行うことでやさしい日本語コーパスを作成した。本研究では新聞記事コーパスを用いることにより、より精度の高いパラレルコーパスの構築を目指す。

2.2 英語のパラレルコーパスによるテキスト平易化

English Wikipedia と Simple English Wikipedia をコンパラブルコーパスと考え、ここから抽出された単言語パラレルコーパスを用いたテキスト平易化手法が提案されている。Zhu らは English Wikipedia と Simple English Wikipedia から得られた大規模な並列データセットを使用した統計的機械翻訳による手法をベースとした、ツリーベースの翻訳モデル(TSM)によるテキスト平易化モデルを提案している[Zhu 10]。Zhu らは Wikipedia の言語リンクをたどり、English Wikipedia と Simple English Wikipedia から 65,133 対の対応付けを行った。その後、JWPL を用いて特定のタグを削除、両テキストからプレーンテキストを抽出した。本研究では毎日新聞および毎日小学生新聞を用いる。対応付けには記事番号や「ある」、「いる」などの特徴を得られない動詞などを除いたベクトルを作成し、コサイン類似度を用いて対応付けを行う。

2.3 平易なコーパスを用いないテキスト平易化

梶原[梶原 18]らは生コーパスのみからテキスト平易化のための疑似パラレルコーパスを自動構築するフレームワークを提案している。生コーパスから無作為に抽出した2つの文に対して

連絡先:

小藤 直紀, 難波 英嗣, 竹澤 寿幸

広島市立大学 情報科学部

〒731-3194 広島市安佐南区大塚東三丁目 4 番 1 号

{koto, nanba, takezawa}@ls.info.hiroshima-cu.ac.jp

タスクに応じた品質推定を行い、一定以上の尤度を持つ文対を疑似パラレルコーパスとして抽出する

梶原らが、リーダビリティ推定で得た難解なサブコーパスと平易なサブコーパスを用いて文間類似度を計算し、英語のテキスト平易化のための疑似パラレルコーパスを構築した。本研究では、毎日新聞と毎日小学生新聞をそれぞれ難解なテキスト、平易なテキストとしてパラレルコーパスを構築する。毎日新聞と毎日小学生新聞で対応するテキストの難易ははっきりとしているので、リーダビリティを推定することなく、日本語の平易化のためのパラレルコーパスを作成することができる。

2.4 SCDV を用いた文書分類

Word2Vec のベクトル空間に対して GMM(Gaussian Mixture Models)でクラスタリングを行い、各単語がどのトピックに属しているのか考慮したベクトル空間に修正することで、単語ベクトルから文書ベクトルを作成する際の精度向上を目的とする手法 SCDV[Mekala 17]が他クラス文書分類問題において、他の手法と比べて優れていることが示されている。本研究でも SCDV が有効であると考え、記事間の類似度を計算する際の、特徴付きベクトルの作成に用いる。

3. テキスト平易化システムの構築

3.1 節ではテキストの平易化手順について述べる。3.2 節では、パラレルコーパスの構築手順である記事単位のアライメント取得手法について、3.3 節では、文単位のアライメント取得手法について、3.4 節では、記事間の対応付けについて、3.5 節では、テキスト平易化について述べる

3.1 平易化手順

テキスト平易化手順として、毎日新聞および毎日小学生新聞の記事を難解なテキストおよび平易なテキストとして対応付けてパラレルコーパスを構築する。新聞記事同士を対応付けするにあたり、記事単位の対応付けをおこなった後、対応している記事内で文単位の対応付けを行い、アライメントをとる。その後、翻訳作業を行う。

3.2 記事単位の類似度判定

本研究ではコサイン類似度を用いて記事間の類似度を計算した。コサイン類似度計算に用いる各記事ベクトルは TF-IDF により取得する。また、単語の意味をベクトルとして表現する word2vec をクラスタリングし、文書の特徴を得る SCDV という手法が近年盛んに用いられる。本研究の平易化テキストに対しても、文字の一致率だけでは無く、文字の意味的特徴を考慮することが必要であると考え、SCDV を用いる。

3.3 文単位の類似度判定

難解な文と平易な文の同義性を評価し、対応付けするために本研究では、単語分散表現のアライメントに基づく、3 つの文間類似度の計算手法[Song 15]と Word Mover's Distance[Rubner 98]を本タスクに用いる。以下にそれぞれの計算手法を記載する。

(1) Average Alignment

全ての単語の組み合わせについて単語間類似度を計算、平均して文間類似度 $S_{ave}(x, y)$ を求める。 $\phi(x_i, y_j)$ はコサイン類似度である。

$$S_{ave}(x, y) = \frac{1}{|x||y|} \sum_{i=1}^{|x|} \sum_{j=1}^{|y|} \phi(x_i, y_j)$$

(2) Maximum Alignment

文に含まれる各単語に対して最も類似度が高い単語を選択し、それらの単語のみ計算した単語間類似度 $\phi(x_i, y_j)$ を平均して $S_{asym}(x, y)$ を求める。 $S_{asym}(x, y)$ は非対称なスコアなので、それらの平均値を用いて対象な文間類似度 $S_{max}(x, y)$ を計算する。

$$S_{asym}(x, y) = \frac{1}{|x|} \sum_{i=1}^{|x|} \max_j \phi(x_i, y_j)$$

$$S_{max}(x, y) = \frac{1}{2} (S_{asym}(x, y) + S_{asym}(y, x))$$

(3) Hungarian Alignment

2 文を単語をノードとする 2 部グラフとして考え、一対一の単語アライメントに基づく文間類似度を定義する。文 x に含まれる各単語 x_i に対して Hungarian 法[Kuhn 95]によって文 y 中の単語 $h(x_i)$ を選択し、それらの $|x|$ 個の単語の組み合わせについて計算した単語間類似度を平均した文間類似度 $S_{hun}(x, y)$ を求める。

$$S_{hun}(x, y) = \frac{1}{\min(|x|, |y|)} \sum_{i=1}^{|x|} \phi(x_i, h(x_i))$$

(4) Word Mover's Distance

$\psi(x_u, y_v)$ は単語 x_u と単語 y_v の間の単語間非類似度(距離)を表し、本研究ではユークリッド距離を用いる。また A_{uv} は文 x 中の単語 x_u から文 y 中の単語 y_v への輸送量を表す行列であり、 n は語彙数、 $freq(x_u)$ は文 x 中での単語 x_u の出現頻度である。

$$S_{wmd}(x, y) = 1 - WMD(x, y)$$

$$WMD(x, y) = \min \sum_{u=1}^n \sum_{v=1}^n A_{uv} \psi(x_u, y_v)$$

$$\sum_{v=1}^n A_{uv} = \frac{1}{|x|} freq(x_u)$$

$$\sum_{u=1}^n A_{uv} = \frac{1}{|y|} freq(y_v)$$

3.4 記事間の対応付け

3.2 節で述べた記事単位の類似度を計測する際に、同じ内容の毎日新聞記事と毎日小学生新聞記事の日付に何日のずれがあるか調査した。調査の結果、前後1日より2日、前後2日より3日と対応している記事間の日付は近いほど対応数は増加していた。また、ある日付の毎日小学生新聞が、その日付の毎日新聞および、その日付より新しい毎日新聞と対応していることは極めて少なかった。人手による対応データは、毎日新聞から1日遅れで毎日小学生新聞の記事と対応している記事が記載されることが多い結果となっていた。

3.5 テキスト平易化

3.2 節および 3.3 節で提案した手法を毎日新聞および毎日小学生新聞に適用し、日本語テキスト平易化パラレルコーパスを自動構築する。構築したパラレルコーパスは、形態素解析器を用いて単語区切りを入れた後、機械翻訳システムを適用することで、日本語テキストの平易化システムを構築することができる。従来のテキスト平易化研究では、機械翻訳システムとして統計翻訳器が利用されることが多かったが、本研究でも、統計翻訳パッケージのひとつである *cicada*(<http://att-astrec.nict.go.jp/product/cicada/>)を用いる。この他、近年、機械翻訳分野で活発に研究されている *seq2seq* モデルをはじめとしたニューラル機械翻訳も利用して平易化システムを構築する。なお、本研究では、テキスト平易化パラレルコーパスの構築に主眼を置いているため、評価は記事単位の類似度判定および文単位の類似度判定についてのみ行う。

4. 実験

本章では、本研究で行った実験条件、実験結果、またその考察について、4.1 節では記事単位の類似度判定実験、4.2 節では文単位の類似度判定実験について述べる。

4.1 記事単位の類似度判定実験

(1) 実験条件

実験データおよび評価

毎日新聞、毎日小学生新聞ともに 2016 年の 1 年間分を使用して、人手により記事の対応付けを行い、記事単位の類似度判定実験に用いた。人手により作成された 856 対の記事単位の対応は A 評価: 顕著に対応が取れる B 評価: 対応を取れる という観点で対応付けが行われている。A 評価のものだけを用いて評価した場合と、A 評価と B 評価の両方を用いた場合の 2 通りで実験を行う。なお、評価には平均精度(average precision)を用いる。

比較手法

本研究では、以下の比較手法に基づき記事単位の類似度判定実験を行う。

- TF-IDF: 記事中の各単語の重み付けに TF-IDF 手法を用い、毎日小学生新聞の 1 記事に対して、その記事の日付当日および、その日付から前後 4 日分の毎日新聞の記事との類似度を計算
- SCDV: SCDV を用いてその記事の日付当日および、その日付から前後 3 日分までの毎日新聞記事との類似度を計算

(2) 実験結果

評価結果を表 1 に示す。表 1 より、対象記事に対して以前の記事だけでなく、TF-IDF を用いたアライメントの評価と SCDV を用いたアライメントの評価を比較したとき、TF-IDF を用いたアライメントのほうが優れていることがわかる。

表 1 記事単位の対応付け(平均精度)

| 手法 | A 評価 | A+B 評価 |
|--------|--------|--------|
| TF-IDF | 0.9206 | 0.9042 |
| SCDV | 0.8734 | 0.8520 |

(3) 考察

本節では 4.1.2 節の実験結果に対する考察を述べる。結果より、SCDV よりも TF-IDF を用いて作成した記事ベクトルがより優れていた。正解データと比較した際に、TF-IDF は正しいアライメント結果を示し、SCDV では誤ったアライメント結果を示した例を以下に示す。正解データの毎日小学生新聞記事を図 1、SCDV のアライメント結果である毎日新聞記事を図 2 に示す。毎日小学生新聞での読み仮名は省略している。

<毎日小学生新聞 2016 年 1 月 5 日, 記事番号 7 番>

第92回東京箱根間往復大学駅伝競走が2、3日、神奈川県箱根町―東京・大手町の往復10区間217・1キロメートルで行われ、青山学院大学が往路と復路を制して、2年連続2度目の総合優勝を果たしました。総合記録は10時間53分25秒。

図 1 毎日小学生新聞記事本文(一部抜粋)

<毎日新聞記事(2016 年 1 月 3 日, 記事番号 77 番>

前回大会に続いて、トヨタ自動車が後半に実力を発揮した。群馬県の上州路を舞台に1日に行われた「ニューイヤー駅伝第60回全日本実業団対抗駅伝競走大会」。トヨタ自動車が2連覇を達成して通算優勝回数も5位タイの「3」とした。

図 2 毎日新聞記事本文(一部抜粋)

図 1 と図 2 は記事内容が異なり、チーム名などの名詞は違うが、順位の表記や記事の構成、その他の用語は正解データである毎日新聞記事(2016 年 1 月 4 日, 記事番号 46 番)と非常に似ていた。SCDV の特徴および利点はテキストに対してクラスタリングを行い、各単語がどのトピックに属するかを加味したベクトル空間を作成できる点であるが、本研究で用いている新聞記事は元々各記事の内容に統一性があり、似た形式の記事が見られるため、SCDV を用いることによる利益が、十分に得られなかったことが原因と考えられる。

4.2 文単位の類似度判定

本節では、3.2 節で挙げた文単位の類似度判定実験に関する実験条件、実験結果、またその考察について述べる。

(1) 実験条件

実験データおよび評価

2016 年度の毎日新聞および毎日小学生新聞の対応している記事内で、文を人手により対応付けたものを文単位の類似度判定実験に用いた。文は句点で改行されており、先頭に記事番号と行番号が与えられ整理されている、この作業も人手により行った。人手により作成された 2,813 対の文単位の対応は A 評価: 顕著に対応が取れる B 評価: 対応を取れる、という観点で対応付けが行われている。A 評価のものだけを用いて評価した場合と、A 評価と B 評価の両方を用いた場合の 2 通りで実験を行う。なお、評価には平均精度(average precision)を用いる。

実験手法

本実験において、文単位のアライメントは 3.3 節で述べた単語分散表現のアライメントに基づく 4 種類の文間類似度計算手法、Average Alignment, Maximum Alignment, Hungarian Alignment, Word Mover's Distance を用いて文間の類似度を判定する。A+B 評価の正解データを用いる。

(2) 実験結果

4 種類のアライメント手法に対して、A 評価の正解データのみを用いた場合と、A 評価と B 評価の正解データ両方を用いた場合の平均精度による評価結果を以下の表 2 に示す。今回の実験では Maximum Alignment が A 評価の正解データのみを用いた場合、A 評価と B 評価の正解データ両方を用いた場合のいずれでも最も優れた精度を示している。

表 2 文単位のアライメント評価(平均精度)

| 手法 | 平均精度 | |
|----------------------|---------------|---------------|
| | A 評価 | A+B 評価 |
| Average Alignment | 0.8483 | 0.8407 |
| Maximum Alignment | 0.9802 | 0.9766 |
| Hungarian Alignment | 0.9658 | 0.9617 |
| Word Move's Distance | 0.6109 | 0.5912 |

(3) 考察

本節では 4.2.2 節の実験結果に対する考察を述べる。評価の結果より、Maximum Alignment が最も優れた平均精度を示している。以下に正解データと Maximum Alignment の結果が異なっている毎日新聞および日小学生新聞の、記事本文の抜粋を記載する。id は年月日、記事番号、行を表す。

<毎日新聞>

<id=160105178-1>東京都中央卸売市場「築地市場」(中央区)で5日、新春恒例の初競りが行われ、青森県大間産の200キロのクロマグロが、昨年の3倍超の1400万円(1キロ当たり7万円)で競り落とされた。

<毎日小学生新聞>

<id=S160106002-1>今年11月に移転する築地市場(東京都中央区)で5日、「最後の初競り」が行われました。

<毎日小学生新聞>

<id=S160106002-2>青森県大間産のクロマグロ200キログラムに、昨年の約3倍にあたる1400万円(1キログラム当たり7万円)の値段がつきました。

人手による正解データでは、<id=160105178-1>に対して<id=S160106002-2>が対応している。これは、互いに含まれている単語と文脈を考慮して対応付けたと考えられる。一方、Maximum Alignment のアライメント結果では<id=160105178-1>に<id=S160106002-1>が対応付けられている、Maximum Alignment が分割された文のどちらが十分に文脈を汲み取れているかの判断を誤ったことが原因と考えられる。しかし全体的には非常に高い精度を示しているのも、また事実である。

5. おわりに

本研究では、毎日新聞と毎日小学生新聞から構築されたパラレルコーパスによるテキスト平易化手法を提案した。パラレルコーパスの構築手順である記事単位の類似度判定実験では TF-IDF による評価結果が SCDV を用いた評価結果より優れた結果となり、文単位の類似度判定実験では、4 つのアライメント手法のうち Maximum Alignment による評価結果が最も優れた結果を示した。以上のアライメント手法により 30,940 文対からなる平易化のパラレルコーパスを構築した。

参考文献

[Kuhn 95] Harold William Kuhn: The Hungarian Method for the Assignment Problem, Naval Research Logistics Quarterly, Vol.2, pp. 83-97, 1995.

[Maruyama 17] Takumi Maruyama and Kazuhide Yamamoto: Simplified Corpus with Core Vocabular, Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC), 2018.

[Mekala 17] Dheeraj Mekala, Vivek Gupta, Bhargavi Paranjape, and Harish Karnick: SCDV : Sparse Composite Document Vectors Using Soft Clustering over Distributional , Representations, Proceedings of EMNLP 2017, pp. 659-669, 2017.

[Rubner 98] Yossi Rubner, Carlo Tomasi, and Leonidas Guibas: A Metric for Distributions with Applications to Image Databases, Proceedings of the 6th International Conference on Computer Vision, pp. 59-66, 1998.

[Song 15] Yangqiu Song and Dan Roth: Unsupervised Sparse Vector Densification for Short Text Similarity, Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1275-1280, 2015.

[Zhu 10] Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych: A Monolingual Tree-based Translation Model for Sentence Simplification, Proceedings of the 23rd International Conference on Computational Linguistics, pp.1353-1361, 2010.

[梶原 17] 梶原 智之, 小町 守: Simple PPDB: Japanese, 言語処理学会第 23 回年次大会, pp. 529-532, 2017.

[梶原 18] 梶原 智之, 小町 守: 平易なコーパスを用いないテキスト平易化, 自然言語処理, Vol.25, No.2, pp. 223-249, 2018.

キーワード条件つき変分 Autoencoder による広告文生成

Keyword Conditional Variational Autoencoder for advertising headline generation

福田 宏幸^{*1}
Hiroyuki Fukuda

^{*1}株式会社 電通
DENTSU INC.

Recently, fast and massive advertising headline production is highly demanded by growing number of digital ad. Many types of headline generation systems have been developed. But, most of these systems generate headlines systematically by rules and lack generation variety. On the hands, the systems which generate headlines almost randomly satisfy such variety but these headlines are not relevant to ad objective. Until now, it is still difficult to satisfy both variety and relevancy. To this end, we propose Keyword Conditional Variational Autoencoder for advertising headline generation. We regulate generation process by relevant keyword while keeping variety by randomly selected input hidden variables. In our experiment, It can generate variety of headlines and obtain headlines which include a relevant keyword.

1. 背景

近年、広告制作の実務においては、デジタル広告のシェア拡大に伴い、早く、大量に、広告文(いわゆるキャッチコピー)を制作することが求められている。デジタル広告においては、掲載単価が安いことに加え、入稿の手続きも WEB 上で簡単に出来ることから、次々と新しい広告を投入することが推奨されている。また、このように大量の広告を制作して、運用していくことで、広告の鮮度を保つことができ、同時に、どのような広告文が消費者の心を掴むのかという知見も得ることができる。結果として、制作の現場への作業負担が増大しており、AI を活用した負担の軽減が求められている。

これまでに、ルールベースを中心として様々な広告文生成の手法が開発されてきた[Yamane 14 等]が、生成手法が単純である程、確実な生成が期待できる一方で、バリエーションに限られるという欠点があった。また、人間がルールを設定している為、人間の発想の範疇を超えた、セレンディピティのようなアイデアが出る可能性も少ないと考えられる。

そこで、本研究では、近年研究が盛んになってきている深層生成モデルの一種である変分 Autoencoder (VAE)を用い、バリエーションに富んだ創発的な広告文の生成を目指す。一方で、広告には商品を売るという目的があり、広告文と商品との強い関連性が求められる為、ランダムな文生成ではなく、キーワードによって条件づけられた広告文生成を提案する。

2. 関連研究

近年、深層生成モデルの研究は、深層学習の研究の中でも特に活発に研究がなされているトピックの1つである。特に変分 Autoencoder (VAE) [Kingma 13]や敵対的生成ネットワーク (GAN) [Goodfellow 14]が発表されて以来、関連論文の数が急増している。

これら生成モデルの研究は、画像の生成を対象としたものが中心で、それらが実用的な精度になりつつある一方で、短文に対する生成モデルの研究は、まだ発展途上と考えられている。手法としては、画像同様に VAE を用いたもの[Bowman 15]と、GAN を用いたものがある。GAN においては、画像に比べ、自

然言語の離散的な性質から、Discriminator の勾配が逆伝播しないという課題があり、強化学習の Policy Gradient の手法を用いた SeqGAN [Yu 17]や、誤差関数に MMD による誤差を加えた TextGAN [Zhang 17]など、様々な工夫を凝らした研究がある。また、亜種としてテンプレート文を元に編集ベクトルによって文を生成するモデルも存在する[Guu 17]。

3. 提案手法

広告文と商品との関連性を高める為、訴求内容(価格や機能やイメージ)や訴求ターゲットで条件づける方法が考えられるが、本論文ではキーワード(単語)によって生成を条件づける手法を提案する。広告文のノウハウ本には、「キャッチコピーは、キーワード探し」というような記述もあり、商品特徴や伝えたいメッセージを短文に表現するために、キーとなる単語の選択が重要と考えられている。

具体的には、まずキーワードが選択され、そのキーワードを条件として広告文が生成されたと仮定し、条件つき変分 Autoencoder モデルを適応する。今回の実験では、簡便性の為、キーワードは名詞のみとした。

Keyword:ビール



広告文:暑い夏に、うまいビール。

図1 提案手法におけるキーワードによる条件づけ

3.1 データ

独自に収集した過去の新聞・雑誌の広告文のうち、15 単語以内の広告文およそ 30 万文を使用した。前処理として、MeCab および mecab-ipadic-NEologd を用いて形態素解析をし、id 化を行った。語彙数は 1 万語で足りし、残りは未知語として処理した。データセットを 8:2 に分割し、8 割の訓練データで学習した後、2 割の validation データの精度(誤差)によって、ハイパーパラメータチューニングを行った。

3.2 モデル

VAE のモデルは、[Bowman 15]をベースにし、Encoder および Decoder の双方にキーワードを条件として入力した。Encoder

と Decoder は共に 1 層 LSTM を使い、隠れ層の次元は 200 に設定した。入力文および入力キーワードの単語は共に 200 次元のベクトルの埋め込み表現として入力した。

上記によるネットワークの概略図 2 に示す。推論時には、入力キーワードとランダムにサンプリングされた z を Decoder に入力することで、新たな広告文を生成した。また、ビーム探索も用いている。最適化には Adam を使い validation の精度によって、推論で使用する Epoch を決めた。

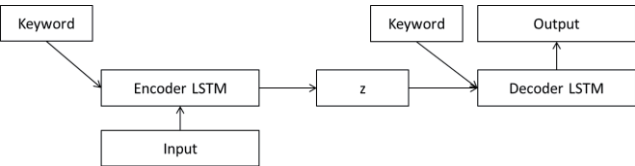


図 2 提案手法のネットワーク概略図

4. 実験

4.1 広告文生成例

前述のモデルを学習することにより、実際に得られた広告文の例を図 3 に示す。広告出稿が比較的多い業種である、トイレタリー・飲料・自動車の広告作業を想定し、「シャンプー」「ビール」「燃費」の 3 つをキーワードとして入力した。今回の実験では、各キーワードに対して 10000 文を出力した。変分 Autoencoder は、正しい文を潜在空間上に確率的にマッピングする為、比較的正しい文を生成しやすいと言われている。但し、実際には広告文になっていないものや、そもそも文として成立していないものも多く、人間が目視で選別している。一方で、「低燃費旅、妻対策」など、人間がなかなか思いつかないような言葉の組合せも出現しており、バリエーションの豊富さにも伺える。また、これらの広告文の類似表現が、元のデータに無いことも確認している。

| Keyword | 生成広告文 |
|---------|--------------------------------|
| シャンプー | 「時短シャンプー」そして、 |
| | シャンプーリゾート。 |
| | 春が気持ちいいシャンプーです。 |
| | がんばるシャンプーが、ここに。 シャンプー0 円ライフ |
| ビール | ビールは、美味しく、楽しめる! |
| | No.1 ビールの贈り物なら |
| | 新ブランドビールで生まれ変わりつづける |
| | ビールが時代の街づくりを |
| | 春のクラシックビールはもっと楽しんで! |
| 燃費 | 低燃費コンパクト |
| | 「燃費が自由自在」をサポート! |
| | 低燃費のヒミツ。 |
| | 低燃費旅、妻対策 |
| | 低燃費がヒロインなのには!? |

図 3 生成された広告文の例

4.2 キーワード出現率

本研究のもう1つの目的は、商品との関連性の高い広告文の生成であり、その為に、キーワードによる条件付けモデルを提案

した。ここでは、その結果の検証として、生成された広告文中のキーワードの出現率を計算した。結果、キーワードを入力していても、必ずしも生成文中にキーワードが出現する訳ではないが、元データにおける出現率に比べて約 50〜200 倍程度の大幅な出現率の向上が確認された。

| キーワード | 生成文での出現率 | 元データでの出現率 |
|-------|----------|-----------|
| シャンプー | 8.96% | 0.04% |
| ビール | 2.88% | 0.06% |
| 燃費 | 3.64% | 0.03% |

図 4 提案手法によるキーワードの出現率の比較

5. 結論および考察

本論文では、キーワードを入力条件とした変分 Autoencoder による広告文生成を試みた。結果、元データでの出現率に比べて高い確率で、目的のキーワードが入った広告文を生成することが可能となった。また、得られる広告文にはバリエーションがあり、時には人間が思いつかないようなセレンディピティ的な広告文も生成された。

一方で、必ずしもキーワードが含まれない場合があったり、文として成立していない出力も多く存在した。通常、深層生成モデルの研究では 100 万文〜1000 万文以上のデータセットが用いられるが、今回は広告文というデータの特殊性から、利用可能なデータが少なく、精度に影響したと考えられる。また、通常の条件付き変分 Autoencoder は、ユニークなクラスラベルを元に条件付けを行うが、今回の場合は1つの広告文に複数のラベルがついている状態である。結果、より複雑な構造の学習を強いることになり、今回の LSTM による学習器がその構造を十分に捉えられなかったものと思われる。

今後は、文として成立している広告文の精度を上げていくと共に、広告においては、様々なタイプの文生成が要求される為、訴求内容や訴求ターゲット等の他の条件においても効率的な広告文の生成ができるように研究を進めていきたい。

参考文献

[Yamane 14] 山根 宏彰, 萩原 将文: ウェブにおけるユーザの嗜好を反映したキャッチコピー自動生成システム, 日本感性工学会論文誌, 2014.

[Kingma 13] D. P. Kingma and M. Welling: Auto-encoding variational Bayes., arXiv, 2013.

[Goodfellow 15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio.: Generative adversarial nets. In Advances in Neural Information Processing Systems, 2014.

[Yu 17] Yu, Lantao, Zhang, Weinan, Wang, Jun, and Yu, Yong.: Seqgan: sequence generative adversarial nets with policy gradient., In AAAI, 2017.

[Zhang 17] Yizhe Zhang, Zhe Gan, Kai Fan, Zhi Chen, Ricardo Henao, Dinghan Shen, Lawrence Carin: Adversarial Feature Matching for Text Generation, arXiv, 2017

[Bowman 13] Bowman, Samuel R., Vilnis, Luke, Vinyals, Oriol, Dai, Andrew M., Jozefowicz, Rafal, and Bengio, Samy: Generating sentences from a continuous space. CoRR., 2015.

[Guu 17] Kelvin Guu, Tatsunori B. Hashimoto, Yonatan Oren, Percy Liang: Generating Sentences by Editing Prototypes., arXiv, 2017.

条件付き変分オートエンコーダと単語分散表現による絵文字の生成

Gerenerating Emoji with Conditional Variational Autoencoders and Word Embedding

山口 篤季^{*1} 藤田 桂英^{*2}
Atsuki Yamaguchi Katsuhide Fujita

^{*1}東京農工大学工学部

Faculty of Engineering, Tokyo University of Agriculture and Technology

^{*2}東京農工大学大学院工学研究院

Institute of Engineering, Tokyo University of Agriculture and Technology

Emoji are among the most widely used communication tools worldwide. Because the number of emoji increases every year and there are 82 face emoji, it might be difficult for users to select an appropriate emoji immediately. Moreover, it is troublesome to continue designing new emoji. Therefore, the aim of the present study is to generate an emoji based on input text automatically to facilitate easier communication and eliminate the process of designing new emoji. The proposed model employs conditional variational autoencoders (CVAE), quasi-recurrent neural networks (QRNN) as the text encoder, and the pre-trained word vector GloVe to the embedding layer connected to the text encoder. In the experiments described herein, it will be observed that the proposed method can generate an emoji that corresponds to an input caption, and output image quality is improved using GloVe.

1. はじめに

絵文字は 1999 年に日本で初めて公開され、携帯電話の普及とともに瞬間に人気となった。2010 年代になると、世界中でスマートフォンが普及し、絵文字は Unicode によって標準化され、広く使われるようになった。絵文字には、顔型の絵文字 (Smiley) から、食べ物、車両、建物に至るまで様々な種類が存在し、絵文字は我々の日常のコミュニケーション活動において重要な役割を果たしている。

絵文字が世界中で広く使われるようになるにつれて、絵文字を対象とした研究が盛んに行われるようになってきた。[1] は、Twitter 上における絵文字とテキスト間の関係性を調査し、どのような種類の絵文字がテキストベースのツイートに最も共起するかを予測した。この研究は、絵文字が世界中で広く扱われているにもかかわらず、自然言語処理の研究題材としてほとんど注目が集められてこなかったことに着目し、絵文字のもつ潜在的な意味を解析することを目的としていた。また、[3] は全ての Unicode の絵文字に関する単語分散表現を作成し、公開した。実験結果では、自然言語を対象とした単語分散表現と同様に、絵文字においても絵文字の意味の加減算が可能なが示され、テキスト解析において絵文字を考慮することの重要性が示唆された。

一方、絵文字の数は年々増加しており、顔型の絵文字に限っても 82 種類存在することから、ユーザが適当な絵文字を即座に選択することは困難である。さらに、新たな絵文字を適宜デザインし、追加する作業は大きな労力を必要とする。敵対生成ネットワークを始めとした機械学習の生成モデル分野が発展してきていることから、入力のカプションに対して対応する絵文字を自動生成することは、実現可能なタスクである。

本研究に応用可能な関連研究として “Text-to-image task” が有名である。これは、入力テキストに基づいて画像を生成するタスクである。このタスクについては数多くの先行研究が存在する一方で、text-to-image synthesis task を応用した絵

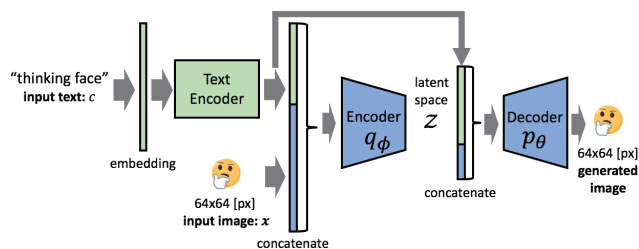


図 1: CVAE ベースの提案手法

文字生成に関する先行研究はほとんど存在せず、次の二つの先行研究のみが挙げられる。[8] と [6] は、深層畳み込み敵対生成ネットワーク (DCGAN) [7] を絵文字生成に活用し、入力のカプションから絵文字を生成することに成功した。しかし、これらの先行研究は実験に使用しているデータセットに関する説明や内容が不十分であり、実装に関する記述が省かれているため、再現性に欠ける研究となっている。

そこで、本研究では絵文字を入力文から自動生成するモデルを提案する。これにより、ユーザのコミュニケーション活動の促進が期待されるほか、新たに絵文字をデザインする手間を省くことができる。提案手法は、CVAE [9] と事前学習済み単語ベクトル GloVe [5] を活用している。CVAE を活用することで、従来手法の GAN ベースのモデルよりも高速でかつ安定した学習を実現できる。さらに、事前学習済みの単語ベクトル GloVe を活用することで出力画像の精度に関して、定量的評価により示す。

以下に、本論文の構成を示す。まず、絵文字を入力文から自動生成するモデルを提案する。次に、ベースラインと提案手法の定性的評価の結果を示す。最後に、本論文のまとめと今後の課題を示す。

連絡先: 山口 篤季, 東京農工大学工学部情報工学科, 東京都小金井市中町 2-24-16, y-atsuki@st.go.tuat.ac.jp

- loudly crying face
- a sad face with tears streaming down both cheeks
- this face is distraught and inconsolable

図 2: データセットの画像とキャプションの対応の一例

2. 提案手法

図 1 は提案するモデルを示している。このモデルは、CVAE をベースモデルとして活用し、GloVe を埋め込み層に用い、さらに QRNN [2] をテキストエンコーダとして利用している。モデルは式 (1) で定義される変分下限 \mathcal{L} を最大化することで学習できる。

$$\mathcal{L}(\mathbf{x}, c; \theta, \phi) = \mathbb{E}_{q_{\phi}(z|\mathbf{x}, c)} [\log p_{\theta}(\mathbf{x}|c, z)] - D_{KL}(q_{\phi}(z|\mathbf{x}, c) || p_{\theta}(z)) \quad (1)$$

単語分散表現

単語分散表現は自然言語処理の研究において、単語の特徴量を抽出するのに広く活用されている。提案手法は 300 次元の事前学習済み単語ベクトル GloVe [5] を活用している。事前学習済みの単語ベクトルを活用することで、モデルは入力文の意味をより効率的に捉えることができるようになり、出力画像の質が向上することが期待される。

テキストエンコーダ

テキストエンコーダには QRNN [2] を用いる。QRNN は畳み込みニューラルネットワーク (CNN) [4] をベースにしたモデルであり、再帰ニューラルネットワーク (RNN) の構造を CNN で再現したものである。RNN とは異なる特徴として、CNN をベースとしたモデルのため並列処理に特化していることが挙げられる。このため、LSTM ベースのネットワークと同等の性能を示しながらも、より高速な演算を実現している。

3. 評価実験

3.1 データセット

データセットは絵文字画像とキャプションデータで構成され、合計 260 個のキャプションデータが存在する。絵文字画像は EmojiOne から収集し、キャプションデータは、Unicode の定める CLDR Short Name と Emojipedia の対応する絵文字の説明文からなる。絵文字あたりのキャプション数は、平均値で 3.17、中央値で 3 となっている。また、一文あたりの単語数は平均値で 9.33、中央値で 8 である。図 2 に、絵文字画像データとキャプションデータの対応を示した例を示す。

3.2 実験設定

データセットの 10% をテストセットとした。提案手法は Adam にしたがって最適化され、学習率とバッチサイズはそれぞれ、0.001 と 26 に設定した。学習は 200 エポック実行した。

3.3 定量的評価

提案手法を CNN ベースの分類器^{*1} による出力画像の分類精度と、Inception スコア [10] の 2 つの評価指標に基づいて定量的に評価した。各指標は 100 回ずつ計算を行った。表 1 は 3

*1 CNN ベースの分類器を実装し、82 種類の絵文字を分類できるようにデータセットで学習させた。データセットのうち 10% をテストセットとした。分類器は 20 エポックの時点で、学習用セットにおいて、0.995、テストセットにおいて、1.00 の精度を記録した。

| Model | Accuracy | IS | Runtime (s) |
|----------------|--------------|-------------|-------------|
| CVAE w/ GloVe | 0.869 | 1.05 | 20.2 |
| CVAE w/o GloVe | 0.262 | 1.02 | / |
| Cond-DCGAN | 0.885 | 1.32 | 33.6 |
| Dataset | / | 1.38 | / |

表 1: 3 モデル間の定量的比較結果: “Accuracy” は CNN ベースの分類器の分類精度を示している。“IS” は Inception スコアの略である。太字は提案手法 (CVAE w/ GloVe) と従来手法 (Cond-DCGAN) 間でマン・ホイットニーの U 検定により有意な差が認められたことを表す。 ($p < .05$)

モデル間の定量的比較結果を示している。なお、DCGAN による従来手法の実装が公開されていないため、本論文では [7] による DCGAN の実装を基に、条件付き DCGAN を実装して比較対象とした^{*2}。

提案手法における GloVe の有無が与える影響 表 1 より、二つの指標において、事前学習済みの単語ベクトルを用いたモデルは、事前学習済みの単語ベクトルを用いないモデルよりも優れていることが確認できる。CNN ベースの分類器による分類精度に着目すると、比較対象の二つのモデル間では 0.6 以上の有意な差がある。したがって、事前学習済みの単語ベクトルを用いたモデルの大半の出力画像は、入力キャプションに従った画像を出力できている。一方、Inception スコアにおける二つのモデル間の差は 0.04 よりも小さく、データセットの値よりも 0.3 以上の差がある。これは、二つのモデルの出力画像が共に色彩面のノイズを含み、ぼやけていることが原因である。**提案手法と従来手法の比較** CNN ベースの分類器による分類精度は、提案手法と従来手法で有意ではあるが 2% 程度の差しか見られない。一方、従来手法の Inception スコアは、データセットのスコアとかなり近い数値を示しており、提案手法と従来手法の Inception スコアには 0.27 の大きな差が生じている。対照的に、エポック単位での実行時間に着目すると、提案手法は従来手法よりも 1.67 倍高速に動作している。また、従来手法は生成精度を最大化するのに 1400~1500 エポック程度を要するが、提案手法は 200 エポックで十分である。以上から、提案手法は従来手法よりも画質の観点では劣るが、入力キャプションに従った画像をより高速に生成できることが示された。

3.4 潜在表現の可視化

テキストエンコーダ (QRNN) と事前学習済み単語ベクトル GloVe を用いて条件付けられた、エンコーダネットワークがどのように入力画像を潜在変数に落とし込んでいるのかを、潜在表現を可視化することにより検証した。図 3 は、GloVe を用いて学習を行ったエンコーダモデルと、GloVe を用いないで学習をしたエンコーダモデルの出力の散布図を示している。図 3 において、同種の絵文字は、事前学習済み単語ベクトルの有無に関わらず、互いに近い位置にプロットされている。したがって、埋め込み層とテキストエンコーダは、入力キャプションの意味を捉えて、同種の絵文字に対して似たような数値を持つベクトルを生成できている。図 3(1) にプロットされている絵文字の意味に着目すると、 x の値が小さければ小さいほど、より “positive” な絵文字がプロットされている。対照的に、 y の値が小さければ小さいほど、より “negative” な絵文字が現れている。さらに、“positive” や “negative” の指標では分類の難しい絵文字は、原点: $(x, y) = (0, 0)$ 付近にプロットさ

*2 実装は <https://github.com/gucci-j/emoji-gan> に公開した。

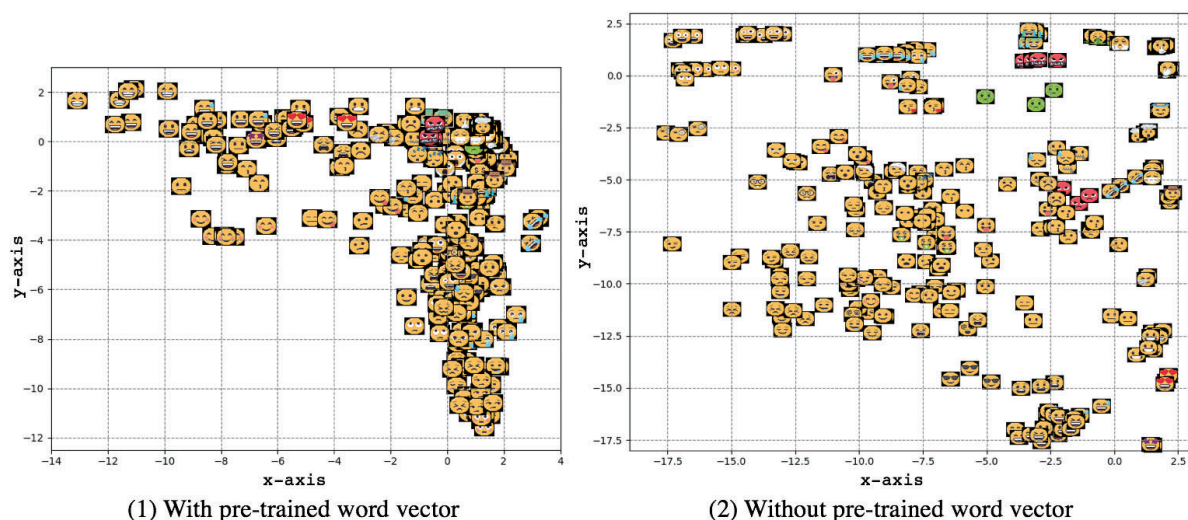


図 3: エンコーダネットワークの出力散布図: 全データセットを学習済みのエンコーダネットワークに入力し、その出力を散布図上にプロットした。なお、マーカーを対応する入力の絵文字に置き換えた。

れている。一方、図 3(2) においては、各絵文字はカテゴリ毎に分類されているものの、図 3(1) のような大域的な絵文字の関係性を見出すのは困難である。図 3(1) と (2) の散布図における絵文字の大域的な分布の違いは、事前学習済み単語ベクトルの影響によるものが大きい。以上の実験結果より、分布が異なる絵文字間の大域的な関係性を含むときには、出力画像のノイズが軽減されるものと示唆される。

4. おわりに

本研究では、入力のキャプションに対応する絵文字を生成できる CVAE ベースのモデルを提案した。また、事前学習済みの単語ベクトル GloVe を活用することで出力画像の画質が向上することが示された。

今後の展望としては、未知のキャプションデータに対して、適切な絵文字を合成できるようにモデルを改良することが挙げられる。これにより、モデルが豊富な種類の絵文字を生成できるようになり、平文でのコミュニケーションがより促進されることが期待される。

参考文献

- [1] Francesco Barbieri, Miguel Ballesteros, and Horacio Saggion. Are emojis predictable? In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pp. 105–111. Association for Computational Linguistics, 2017.
- [2] James Bradbury, Stephen Merity, Caiming Xiong, and Richard Socher. Quasi-recurrent neural networks. In *arXiv*, Vol. abs/1611.01576. 2016.
- [3] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. emoji2vec: Learning emoji representations from their description. In *Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, pp. 48–54. Association for Computational Linguistics, 2016.
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. 2012.
- [5] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.
- [6] Marcel Puyat. Emotigan : Emoji art using generative adversarial networks. <http://cs229.stanford.edu/proj2017/final-reports/5244346.pdf>, 2017.
- [7] Alec Radford, Luke Metz, and Soumith Chintala. Un-supervised representation learning with deep convolutional generative adversarial networks. In *arXiv*, Vol. abs/1511.06434. 2015.
- [8] Dianna Radpour and Vivek Bheda. Conditional generative adversarial networks for emoji synthesis with word embedding manipulation. In *arXiv*, Vol. abs/1712.04421. 2017.
- [9] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems 28*, pp. 3483–3491. 2015.
- [10] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.