

# 俳句生成を目的とした言語モデルに対する Attention 機構の導入

横山 想一郎<sup>1,a)</sup> 米田 航紀<sup>2,b)</sup> 山下 倫央<sup>1,c)</sup> 川村 秀憲<sup>1,d)</sup>

概要：本研究では、俳句の生成を目的として、現代俳句を教師データとして LSTM に基づく俳句の言語モデルの開発を進めている。言語モデルの汎化性能の向上のため、LSTM へ Attention を導入した。また、俳人の詠んだ俳句および形態素をランダムに入れ替えた俳句を教師データとして、言語モデルが生成した俳句を評価する LSTM を開発した。本稿では、俳句の生成器と評価器の詳細を説明するとともに、生成器に導入した Attention 機構が生成俳句に与える影響を検証する。

## Neural Language Model with Attention for Haiku Poems Generation

### 1. はじめに

俳句は世界最短の定型詩とされ、5音・7音・5音の計17音を基本として、その中に季語や強く言い切る働きを持つ切れ字（「や」、「かな」、「けり」）などを含みながら、言葉巧みに写実や情景を表現する。作者が五感で感じたものや思い描いた情景を俳句に詠んだとき、読み手が作者の感じたものを心の中で再現できることが重要視される。俳句に含まれる季語とは、特定の季節を表す言葉であり、本情と呼ばれる背景的な状況や意味が作者と読み手の間で共有されていることが前提とされ、作者から読み手へ情景を伝達する機能を持つ。人が持つ知識や感情を含んだ情報を日本語の文章などの形でエンコード・デコードする人工知能技術は現状では確立されておらず、こうした課題を含む点で俳句の生成は良い課題である。

コンピュータにより俳句を生成するアプローチとして、俳句に用いられるフレーズやフレーズの関係性についてのデータベースを辞書をもとに構築し、ユーザーが与えた単語との関係性に基づいてフレーズを選択する方法が提案されている [1]。フレーズやその関係性のデータベースを構築

する際に、ウェブサイトの日本語文から共起頻度などを計算することで、より多様な語彙をもとに俳句を生成する研究も行われている [2]。また、俳句データを用いてニューラルネットワークによる言語モデルを学習するアプローチも試みられている [3]。ニューラルネットワークによる言語モデルは、共起頻度に基づくデータベースと比べフレーズの関係性をより柔軟に表現可能であることが期待される。一方で、先行研究においては、俳句を学習した言語モデルは、パープレキシティのみにより評価されており、音数などの俳句の定型詩としての制約など、生成された俳句の質に関する評価を行う余地がある。本稿では、ニューラルネットワークによる俳句の言語モデルを様々な構成のもとで学習し、パープレキシティによる言語モデルの評価とともに、生成される俳句の質を評価することで、俳句の生成に適した言語モデルの検討を行う。

また、人間が俳句を詠むときに、いくつかの候補となる俳句を作句した後に、それぞれの俳句を評価し、最終的に1つの俳句を選択するプロセスを経ることがある。こうしたプロセスを代表して、近代の著名な俳人である高浜虚子は、「選は創作なり」という言葉を残している。本稿では、俳句の言語モデルとは別に、生成された俳句を入力として評価値を出力するニューラルネットワークを学習し、こうしたプロセスの再現とより質の良い俳句の生成を目指す。

本稿は以下の通り構成される。第2章において俳句の生成及び評価を行う二つのニューラルネットワークにより構成される俳句の生成システムを説明する。第3章では、特

<sup>1</sup> 北海道大学大学院情報科学研究科  
〒060-0814 北海道札幌市北区北14条西9丁目

<sup>2</sup> 北海道大学大学院情報科学研究科  
〒060-0814 北海道札幌市北区北14条西9丁目

a) yokoyama@complex.ist.hokudai.ac.jp

b) yoneda@complex.ist.hokudai.ac.jp

c) tomohisa@complex.ist.hokudai.ac.jp

d) kawamura@complex.ist.hokudai.ac.jp

に俳句の言語モデルについて、いくつかの構成のもとで学習を行い、得られる言語モデルおよび生成される俳句の質について評価し、第4章で結論を述べる。

## 2. 提案する俳句の生成システム

本稿で提案する俳句の作成システムは、大きく分けて以下の二つのモジュールにより構成される。

- 俳句生成のための言語モデル
- 生成俳句を評価するモデル

俳句生成のための言語モデルは、過去に俳人の読んだ俳句により言語モデルを学習し、得られた言語モデルを使用して俳句を生成する。俳句を評価するモデルは、俳句を入力として評価値を出力するニューラルネットワークを持ち、言語モデルが生成した俳句から最終的に出力する俳句を選択するために用いられる。

### 2.1 俳句生成のための言語モデル

本稿ではニューラルネットワークによる言語モデルを利用して俳句を生成する。インターネットから収集された俳句データをもとに、言語モデルを学習する。得られた言語モデルをもとに、任意の文章に続く単語の出現確率が可能となる。そこで、文頭文字を言語モデルに対する初期の入力として、出現確率に応じた乱数による単語の選択を、終端文字が選ばれるまで繰り返すことにより、俳句を作成する。こうした手順を、生成する俳句の個数だけ繰り返す。

### 2.2 生成俳句を評価するモデル

俳句の評価モデルは、個々の俳句に対する評価値を学習データとして与え、生成俳句に対する評価値を推定するニューラルネットワークにより構成される。ニューラルネットワークに適切な学習データを与えることで、日本語としての解釈の容易さや、専門家の視点による俳句としての平均的な評価を推定するモデルが得られることが期待される。例えば、言語モデルが学習した個々の俳句について俳句の専門家による評点を付与し、ニューラルネットワークの学習データとすることで、専門家による俳句の評点を推定する深層モデルが学習可能と期待される。しかし、俳句データの入手先として有力な俳句の作品集やコンテストの入選作品などには、一般的に評価の高い俳句のみが掲載されており、専門家の視点で高い評価を得ることの出来なかった俳句のデータを既存のデータベースから機械的に収集することは困難である。このため、こうした学習データの作成には大きな手間がかかる。

本稿では、推定器の学習データを、言語モデルの学習データに含まれる俳句の文章を正例とし、俳句データベースに含まれる俳句から任意の二つの形態素を乱数を用いて選択し、出現する位置を交換することにより作成した文章を負例として作成する。こうした学習データを与えること

表 1 俳句生成を行う LSTM の学習パラメータ

| パラメータ名     | 設定値     |
|------------|---------|
| LSTM 層数    | 2       |
| LSTM ユニット数 | 128     |
| 最適化手法      | Adam[4] |
| 学習率        | 0.001   |

表 2 学習した言語モデルのパープレキシティ

| 分かち書き手法       | Attention | テストデータ | 訓練データ |
|---------------|-----------|--------|-------|
| 文字            | なし        | 41.4   | 39.3  |
| 文字            | あり        | 42.5   | 35.9  |
| Sentencepiece | なし        | 164.5  | 126.1 |
| Sentencepiece | あり        | 181.1  | 106.9 |

により、日本語としての解釈が容易な俳句に対して高い評価値を出力する推定器が得られると考えられる。正例は言語モデルの学習データを流用することができ、負例も機械的な操作により作成可能であるため、比較的 low コストで実施可能である。

## 3. 俳句の生成システムの学習

### 3.1 俳句生成のための言語モデル

言語モデルを学習する際に使用するデータは、インターネット上のデータベース\*1から収集した。「(」等の読みがなを表す記号を含む俳句は学習データとして不適として除外し、最終的に訓練データとして 444,437 句、テストデータとして 1,024 句を使用した。

ニューラルネットワークによる言語モデルとしては、Long Short-Term Memory(LSTM) によるもの [5] や、LSTM に Attention 機構を加えたもの [6]、言語翻訳のタスクで優れた成績を示す Transformer モデルの構造を模したものの [7] が提案されている。本稿では LSTM ニューラルネットワークにより言語モデルを構成した。俳句を単語ごとに分割する分かち書きの方法として、文字ごとに分割するものと、Sentencepiece[8] により頻出文字列を語彙に追加し、語彙数の増加と引き換えに平均単語数を削減したものの二つを比較する。また、LSTM の構成に関しても、Attention 機構を持つものと持たないものの二種類を比較し、合計四通りの手法について比較する。Sentencepiece の語彙は訓練データ中に出現する全ての文字を被覆し、語彙数を 8000 として、訓練データをもとに作成した。その他のパラメータは予備実験により表 1 の通り決定した。

得られた言語モデルのパープレキシティを表 2 に示す。分かち書き手法が異なる場合、語彙数が異なるため、パープレキシティの値から言語モデルの優劣を比較することは出来ない。分かち書き手法によらず、訓練データに対しては Attention 機構がパープレキシティを小さくする一方で、テストデータに対しては Attention 機構はパープレキ

\*1 俳句例句データベース, <http://taka.no.coocan.jp/a5/cgi-bin/HAIKUreikuDB/ZOU.htm>

表 3 句頭に一般的に見られる形態素の分類

| 大分類 | 中分類  |
|-----|------|
| 名詞  | 普通名詞 |
|     | 固有名詞 |
|     | 数詞   |
| 形状詞 | 全て   |
| 動詞  | 一般   |
| 形容詞 | 一般   |

シティを大きくすることがわかる。ここから、今回の実験設定においては、Attention 機構の導入は過学習につながると言える。

言語モデルにより生成された俳句を評価する。俳句の質を客観的に評価することは困難であるため、形態素解析器を用いて俳句に含まれる音数などの定量化が可能な指標を計算し、訓練データとそれぞれの言語モデルで生成された俳句群との間で比較する。こうした指標の値が訓練データに近いほど、訓練データの持つ俳句の特徴を捉えた、質の高い俳句が生成されていることが期待される。先に述べた4通りの設定により得られた言語モデルを用いて、それぞれ一万句ずつ俳句を生成し、その性質を評価する。形態素解析器として MeCab を、その辞書として現代書き言葉 UniDic を利用した。

俳句の質を検討するための指標について述べる。言語モデルには、俳句を構成する単語間の関係性を捉えることが期待され、訓練データと同一の俳句のみを出力することは好ましくない。このため、生成された俳句について、訓練データの俳句群との最小のレーベンシュタイン距離が6以上の俳句を非類似句と定義する。次に、生成された非類似句の中から、未知語を含むと判定された俳句の割合、および、俳句を構成する各形態素の音数を辞書により求めたとき、その総和が俳句の制約である17音を満たす俳句の割合を計算する。さらに、17音で構成されると判定された俳句について、5音・7音・5音の3区間に分割することを考え、これらの区間をまたぐ形態素が存在する俳句を句またがりと定義する。また、句またがりでない俳句に関して、3区間の先頭に位置する形態素のうち、表3に示す分類に当てはまらないものが一つでも存在する場合は、該当する俳句を一般的でない句頭を持つ俳句と定義する。以上の定義は、定型詩としての俳句の制約や、一般的に見られる特徴を反映したものである。形態素解析器が正しく働かない場合や、俳句の表現技法として意図的に制約を破る場合があるため、訓練データに関してもこうした指標を必ずしも満たさないが、生成された俳句の性質を見る上で有用と考えられる。

言語モデルにより生成された俳句や訓練データに含まれる俳句が、先に述べた指標を満たす割合を、表4に示す。非類似句は、どの言語モデルにより生成された俳句においても、一貫して高い割合を示した。その他の指標については、

分かち書き手法の違いが Attention 機構の有無よりも大きな影響を与える傾向にあり、分かち書きに Sentencepiece を利用し、Attention 機構を含まない言語モデルが、訓練データに近い俳句を生成していることがわかる。

### 3.2 生成俳句を評価するモデル

評価モデルを学習する際に使用するデータは、言語モデルと同様に、インターネット上のデータベースから収集し、「(」等の読みがなを表す記号を含む俳句を除外したデータを用いる。訓練データとしては、言語モデルと重複する444,437句の収集された俳句を正例とし、負例は学習の度に正例の俳句からランダムに形態素を交換することにより作成する。テストデータとしては、言語モデルと重複する1,024句を正例として使用し、それぞれの俳句についてランダムに形態素を交換することで作成した1024句を負例とする。

推定器は Bidirectional LSTM (BLSTM) により構成され、俳句全体の文字列を入力した上で、評価値を実数値により出力する。推定器に入力する俳句には、言語モデルのときに用いられたのと同様の Sentencepiece を用いた分かち書きを施した。俳句の全単語を入力した後の BLSTM の隠れ層の値をを足し合わせ、1層の全結合層でソフトマックス層に入力することで、2分類を行う。その他の学習パラメータは5に示す通りである。

学習により得られた推定器のテストデータに対する正解率は80.7%であった。前節で最も学習元の俳句データに近い性質を示した、Sentencepiece による分かち書きを利用し Attention 機構を含まない言語モデルが出力した9,880句の非類似句に対して、評価モデルを適用し、評価値の小ささが上位の10句を表6に示す。単語の交換が起きた確率を評価値としているため、小さい値ほど良い評価と考えることのできる。評価値が最も小さい俳句は、3文字のみから構成されており、俳句として成立していないため、本来は高い評価値が推定されるべきである。これは、評価モデルの学習データに、こうした文字数が極端に異なる俳句が含まれていないためと考えられる。この他の俳句に対する妥当性の評価は、今後の課題である。

## 4. おわりに

本稿では、言語モデルによる俳句の生成および俳句の評価の機能を実装した、俳句の生成システムを提案した。俳句の生成には言語モデルの学習において一般的な LSTM を用い、インターネットから収集された40万句ほどの俳句データを用いて学習を行った。得られた言語モデルに対してはパープレキシティの計算とともに、生成された俳句に対する音数など用いた評価を行った。その結果、今回の実験設定においては、語彙を構成する方法として Sentencepiece を利用し、Attention 機構を用いない LSTM による言語モ

表 4 学習した言語モデルにより生成された俳句の性質

| 分かち書き手法       | Attention | 非類似句  | 未知語なし | 音数 17 音 | 句またがり | 一般的でない句頭 |
|---------------|-----------|-------|-------|---------|-------|----------|
| 文字            | なし        | 99.2% | 93.4% | 38.9%   | 66.5% | 73.9%    |
| 文字            | あり        | 98.9% | 93.0% | 39.7%   | 65.9% | 73.7%    |
| Sentencepiece | なし        | 98.8% | 93.5% | 41.6%   | 63.1% | 71.8%    |
| Sentencepiece | あり        | 98.7% | 93.7% | 41.0%   | 64.5% | 72.5%    |
| 訓練データ         |           | -     | 96.2% | 54.8%   | 48.1% | 60.7%    |

表 5 俳句評価値の推定器の学習パラメータ

| パラメータ名     | 設定値   |
|------------|-------|
| LSTM 層数    | 2     |
| LSTM ユニット数 | 50    |
| 最適化手法      | Adam  |
| 学習率        | 0.001 |

表 6 俳句評価値の推定例

| 俳句              | 評価値    |
|-----------------|--------|
| 卯波や             | 0.0066 |
| やはらかき涙のほたる五月かな  | 0.0135 |
| 揚花火しきりに広くなりけり   | 0.0146 |
| 透きとほる鱧の荅を食べてをり  | 0.0152 |
| 沙羅の花のながるる闇の夜明かな | 0.0157 |
| 初鏡連れて書きたる牛一匹    | 0.0169 |
| 神の留守胸三つ出て大いなる   | 0.0171 |
| 芋の花になりたる如く遊びけり  | 0.0173 |
| 抱き上げて火の恋しくて大根漬  | 0.0178 |
| 地の底にかなしき梯子秋驟雨   | 0.0179 |

デルが、最も訓練データに近い性質の俳句を生成することを示した。更に、単語の交換の有無を学習データの評価値として、俳句の評価モデルを学習し、その結果を示した。

今後の課題として、生成された俳句を専門家とともに検討することによる、評価モデルの出力の妥当性の検証が挙げられる。また、特定のモチーフに基づく俳句の生成や選択も課題といえる。モチーフを指定する方法としては、システムにモチーフを指す単語を入力する方法の他に、画像に対するキャプションの生成の研究を応用して、風景を写した画像を入力とすることが考えられる。

## 参考文献

- [1] Tosa, N., Obara, H. and Minoh, M.: Hitch Haiku: An Interactive Supporting System for Composing Haiku Poem, *Entertainment Computing - ICEC 2008* (Stevens, S. M. and Saldamarco, S. J., eds.), Berlin, Heidelberg, Springer Berlin Heidelberg, pp. 209–216 (2009).
- [2] Rzepka, R. and Araki, K.: Haiku Generator That Reads Blogs and Illustrates Them with Sounds and Images, *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, AAAI Press, pp. 2496–2502 (2015).
- [3] Wu, X., Klyen, M., Ito, K. and Chen, Z.: Haiku Generation Using Deep Neural Networks, 言語処理学会第 23 回年次大会発表論文集, pp. 1133–1136 (2017).
- [4] Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *International Conference on Learning Representations (ICLR)* (2015).

- [5] Sundermeyer, M., Schlüter, R. and Ney, H.: LSTM Neural Networks for Language Modeling, *INTERSPEECH* (2012).
- [6] Daniluk, M., Rocktäschel, T., Welbl, J. and Riedel, S.: Frustratingly Short Attention Spans in Neural Language Modeling, *CoRR*, Vol. abs/1702.04521 (online), available from <http://arxiv.org/abs/1702.04521> (2017).
- [7] Radford, A.: Improving Language Understanding by Generative Pre-Training (2018).
- [8] Kudo, T. and Richardson, J.: SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing, *CoRR*, Vol. abs/1808.06226 (online), available from <http://arxiv.org/abs/1808.06226> (2018).