4W-05

深層学習による自己回帰モデルを用いた俳句生成器の評価

平田 航大 †
北海道大学 大学院情報科学院 †

横山 想一郎 ‡
北海道大学 大学院情報科学研究院 ‡

山下 倫央 §
北海道大学 大学院情報科学研究院 §

川村 秀憲 ¶
北海道大学 大学院情報科学研究院 ¶

1 はじめに

芸術分野における創作活動は，人間固有の欲求である知的好奇心や想像意欲を源とする人間特有の活動である．創作活動を人工知能に行わせるという取り組みは，知能とは何か，人とは何かという問いに対する重要な糸口を含んでいると考えられている [1].

本稿では，日本で古くから親しまれてきた文章による芸術作品である俳句を対象とし，深層学習による自己回帰言語モデルを用いた生成を行う．俳句は，音数が 5・7・5 の 17 音で構成されること，「季語」と呼ばれる単語を一つだけ含むことなどの制約を基本とする定型詩であり，詠み手の情景や心情が表現される．俳句の鑑賞者が情景や心情を読み取る際には，時代背景や鑑賞者個人の持つ知識が大きな役割を果たすため，複数の鑑賞者が良いと考える俳句は必ずしも同一ではなく，こうした過程の理解は人工知能による創作における課題である．一方で，情景や心情を把握することが難しい俳句に対する評価は多くの鑑賞者の間で共通している．しかし，こうした観点での生成文評価は文章生成モデルで一般的に用いられる生成文の流暢さとは異なり，俳句データで文章生成モデルを訓練した際の性能は明らかでない．

こうした背景から，本稿では俳句データを用いて言語モデルを訓練し俳句生成器としての評価を行う．言語モデルの流暢さの指標であるパープレキシティに加え，モデルの生成文が俳句の制約を満たす割合を用いてモデルを評価する．さらに，情景や心情を想像することのできる俳句が生成される可能性を検証するため，俳句経験のある人間によるラベル付きデータを作成し，言語モデルにより算出される尤度と比較する．

2 関連研究

深層学習による自己回帰モデルが文章生成において高い性能を示すことが知られ，代表的な構造として再帰的ニューラルネットワークの一種である Long-Short Term Memory (LSTM) を用いた AWD-LSTM [2] および Transformers のデコーダ部を用いた GPT-2 [3] が挙げられる．特に，GPT-2 は対象のデータセットの学習に先立ち，Wikipedia の文章をはじめとした膨大な公開データセットによる事前学習を行うことで，高い性能が得られることが知られている．

表 1　学習に用いるデータセット

データセット名	作品数	総文字数
青空文庫データセット	16,222 作品	約 2.2 億字
俳句データセット	504,068 句	約 6,500 万字

表 2　俳句テストデータに対するパープレキシティ

モデル名	パープレキシティ
AWD-LSTM	55.22
GPT-2	33.69
GPT-2（+ 青空文庫）	30.58

3 実験

3.1 実験設定

AWD-LSTM および GPT-2 を用いた次の 3 種類の生成モデルを俳句データセットにより訓練し，得られた俳句生成器を評価する．各モデルには，データセットの文章を文字ごとに分割し 6,542 種類のトークンを割り当てた整数列が入力される．

AWD-LSTM　3 層の AWD-LSTM により構成され，2,200 万のパラメータ数を持つモデルである．先行研究で精度向上が確認された fine-tune と呼ばれる技法 [2] を適用し，俳句データセットのみを学習する．

GPT-2　1.1 億のパラメータを持ち，GPT-2 の提案論文 [3] で GPT-2 Small と呼ばれるモデルである．事前学習を行わず，俳句データセットのみを学習する．

GPT-2（+ 青空文庫）　先述の GPT-2 と同じ構造を持つが，青空文庫データセットによる事前学習の後に，俳句データを学習する．

データセットを表 1 に示す．俳句データセットはインターネットで公開されている俳句を収集して用いる．GPT-2（+ 青空文庫）の事前学習では青空文庫データセッ

Evaluation of a haiku generator with autoregressive models based on deep learning
† Kodai Hirata, Graduate School of Information Science and Technology, Hokkaido University
‡ Soichiro Yokoyama, Faculty of Information Science and Technology, Hokkaido University
§ Tomohisa Yamashita, Faculty of Information Science and Technology, Hokkaido University
¶ Hidenori Kawamura, Faculty of Information Science and Technology, Hokkaido University

表3　俳句の制約を満たすモデルの生成文の割合

モデル名	音数	音節	季語数	切れ字数	未知語	非類似句	全条件
AWD-LSTM	15%	33%	43%	95%	81%	95%	3%
GPT-2	28%	54%	63%	95%	94%	98%	12%
GPT-2（+ 青空文庫）	24%	46%	61%	96%	96%	98%	9%
学習データ	33%	60%	70%	96%	96%	-	19%

ト[*1]を 8：1：1 に分割し，学習，検証，テストデータとする．俳句データセットの学習時は 5 分割交差検証を実施し，異なる 5 つのランダムシードで学習を行う．それぞれ検証データに対する損失が最小のモデルを評価する．俳句テストデータに対するパープレキシティを表 2 に示す．

3.2 実験結果

3.2.1 俳句の制約を満たす生成文の割合による比較

各モデルの生成文 1,000 個に対し，俳句の制約に対応する次の 7 条件を満たす割合を評価する．形態素解析には MeCab，辞書に UnDic 2.3.0 を用いる．

音数　文字列を形態素に分解したのち各形態素の音数を取得し，累計の音数が 17 音になること．

音節　文字列を形態素解析したときに先頭から 5 音目もしくは 12 音目をまたぐ形態素が存在しないこと．

未知語　文字列を形態素解析した際に，未知語と判定される形態素を含まないこと．

季語数　インターネット上の季語データベース[*2]から収集した 8,642 語のうち，文字列が 1 語のみを含むこと．

切れ字数　文字列を形態素解析したときに切れ字と判定される単語の出現回数が 1 回以下であること．

非類似句　俳句データセットの学習データとの編集距離の最小値が 5 よりも大きいこと．

全条件　上記の 6 点の条件をすべて満たすこと．

表 3 に示す結果から，AWD-LSTM に比べ，GPT-2 の生成文が全条件を満たす割合は学習データに近く，青空文庫データセットによる事前学習により低下することがわかる．これらの制約は機械的に判定可能であるため，事前学習による割合の低下は，大量の文章を生成し制約を満たす生成分を選別することで容易に対応可能である．

3.2.2 ランク付きデータに対する AUC による比較

2019 年 6 月，2021 年 11 月にマルコボ．コム[*3]と共同で開催したイベントにおいて収集された，計 16,200 句のラベル付き俳句データセットを用いる．

これらのイベントにおいては，言語モデルによる生成文に対し，俳句経験者が俳句とし成立し得るかという点でラベルを付与した．1 名あたり 300 の文字列から，「並選」30 句程度，「特選」5 句程度として情景や心情を伝える俳句として適切な文字列を選択するよう教示を与えた．2019 年のイベントでは LSTM による生成モデルによる文字列を，2021 年のイベントでは GPT-2 による文字列を用いた．

表4　2019 年のイベントでの俳句に対する AUC

	並選以上	特選以上
LSTM	0.60	0.61
GPT-2	0.63	0.64
GPT-2（+ 青空文庫）	0.65	0.65

表5　2021 年のイベントでの俳句に対する AUC

	並選以上	特選以上
LSTM	0.52	0.49
GPT-2	0.55	0.54
GPT-2（+ 青空文庫）	0.56	0.55

言語モデルが算出した尤度とラベルの間で AUC を算出した結果を表 4，表 5 に示す．GPT-2（＋青空文庫）が算出する尤度が，俳句経験者がラベルを付与した俳句に対して高い尤度を算出する傾向にあることがわかる．このモデルは「湖の渦にあなたす浜ぐもり」のように情景の想像が困難な俳句に対して低い尤度を正しく算出するものの，「コスモスの花のぬくもりありにけり」にような，情景を俳句として表現した意図の把握が困難な場合は，高い尤度を算出する傾向にあった．

4 まとめ

本稿では青空文庫データ，俳句データで AWD-LSTM と GPT-2 を訓練し，生成文が制約を満たす割合とラベル付き俳句データに対する AUC から俳句生成器としての評価を行った．実験の結果から事前学習を行う GPT-2 が最も優れ，情景の想像が可能な俳句の生成について一定の性能を持つことを確認した．

参考文献

[1] 川村秀憲，山下倫央，横山想一郎: 人工知能が俳句を詠む:AI 一茶くんの挑戦，オーム社 (2021).

[2] Merity, S., Keskar, N. S. and Socher, R.: Regularizing and optimizing LSTM language models, *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings* (2018).

[3] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I.: Language Models are Unsupervised Multitask Learners, Technical report.

[*1] https://github.com/aozorabunko/aozorabunko
[*2] http://www.haiku-data.jp/kigo.php, 7 月 7 日閲覧
[*3] https://www.marukobo.com