

マスク化言語モデル RoBERTa を用いた俳句の評価

花野 愛里咲[†]
北海道大学 工学部[†]

横山 想一郎[‡]
北海道大学 大学院情報科学研究所[‡]

山下 倫央[§]
北海道大学 大学院情報科学研究所[§]

川村 秀憲[¶]
北海道大学 大学院情報科学研究所[¶]

1 はじめに

近年、人工知能技術を用いた創作活動が盛んに行われており、その対象は、絵画、音楽、文章と多岐に渡っている [1]。現状の課題として、鑑賞者が良いと思う作品を生成することだけでなく、その中から鑑賞者が良いと思う作品を選ぶことも挙げられている。

本稿では、創作の対象として、短い文章で情景や心情を表現するという制約を持つ俳句を取り上げる。俳句の創作においては、詠み手が情景や心情を俳句で表現する過程や俳句から情景や心情を想像する仕組みを理解し、再現することが課題である。

人工知能技術を俳句に適用した先行研究として、深層言語モデルを用いて数千万句の俳句を生成した事例が報告されている [2]。数十万句の俳句データ、深層言語モデル、大規模な計算機リソースを駆使することで、人間が俳句であると解釈できる俳句候補群の生成がある程度は可能になってきた。一方で、俳句候補群から受け手に最も受け入れられる句を選句するという事は容易ではない。選句においては、特定の個人からの高い評価や多くの読者の支持を得られる俳句を選ぶためには、個人の選好・属性、時代背景やその他のメディアを含めた創作の流行を踏まえる必要がある。そのため、現状では選句に関しては多くの課題が残っている。

このような背景を踏まえて、本稿では深層言語モデルを用いて俳句候補群から俳句として認識できる候補を選択する俳句評価器を開発する。俳句評価器としてマスク化言語モデル RoBERTa [3] を採用し、人間が作った俳句を正例として、俳句ではない文字列を負例として識別するタスクを適用することで、俳句として成立する候補を選択する機能の獲得を目指す。

2 俳句評価器の構築

本研究では、人間が作った俳句としてインターネットから収集した俳句を正例として、俳句ではない文字列として生成した俳句の定義を満たす文字列を負例として用いる。これらを識別するタスクを RoBERTa で学習することで俳句評価器を構築する。多くの自然言語処理のタスクで SoTA を獲得した BERT よりも、文書分類タスクにおける高い精度を達成した RoBERTa を採用した。

俳句ではない文字列である負例として、以下の3種類のデータセットを生成する。

- 交換データセット：正例の俳句に含まれる名詞や季語、助詞、動詞、上五、下五のいずれかを同じ種類の別の単語に交換して生成する。例えば、正例の「みちのくの青田に降りる山の雲」という俳句の「青田」という季語をランダムに選ばれた季語「ざくろ」に交換することで「みちのくのざくろに降りる山の雲」という文字列を生成する。
- 散文データセット：青空文庫の作品に含まれる有季定型句を満たす部分文字列を抽出して生成する。青空文庫に含まれる既存の俳句との重複を避けるため、最小編集距離が5以下の文字列を除外する。例えば、青空文庫の作品の中の「春の来るのがどのくらい祝福であるかをお察する。」という文から季語である「春」を含み17音になる「春の来るのがどのくらい祝福で」という文字列を抽出する。
- ランダムデータセット：青空文庫の作品に含まれる単語をランダムに抽出し、17音になるように結合して生成する。作成した文字列の中から単語を1つランダムに選択し、その単語と同じ音数の季語に置き換える。例えば、青空文庫の作品から合計17音になるようにランダムに選択した名詞「家族連れ」、接続詞「しかし」、形容詞「程近い」、「はかない」を結合する。次にこれらの4つの単語からランダムに選択された「家族連れ」を同じ音数の季語「冬の雲」に置き換えて「冬の雲しかし程近いはかない」という文字列を生成する。

Evaluation of haiku using masked language model RoBERTa

[†] Arisa Hanano, School of Engineering, 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1 各データセットに含まれる俳句数

	学習データ	検証データ
俳句データセット	407,439 句	46,195 句
交換データセット	407,439 句	46,195 句
散文データセット	66,000 句	7,593 句
ランダムデータセット	407,439 句	46,195 句

表2 各データにおけるラベルを正しく識別した割合

学習データの負例	早期停止	俳句	交換	散文	ランダム
交換		47.6%	88.9%	76.1%	85.3%
散文		98.5%	1.5%	83.7%	60.1%
ランダム		98.2%	4.2%	17.3%	79.8%
交換+散文①		45.1%	89.3%	96.7%	98.4%
交換+散文②		63.1%	78.2%	66.0%	83.9%
交換+散文③		61.7%	80.6%	65.0%	86.6%
交換	○	76.8%	60.3%	26.0%	48.6%
散文	○	98.4%	1.3%	81.7%	59.4%
ランダム	○	98.8%	2.9%	11.0%	77.3%
交換+散文①	○	77.2%	60.1%	90.9%	86.4%
交換+散文②	○	85.8%	48.6%	31.1%	49.3%
交換+散文③	○	86.3%	51.3%	29.2%	52.6%

各データセットに含まれる俳句数を表1に示す。表1中の俳句データは正例としての俳句数、その他は負例の俳句数である。

3 実験

3.1 実験設定

学習データの負例として下記の6パターンを設定する。正例と各パターンの負例を識別するタスクを学習した6つのモデルを作成し、識別性能を比較する。

- 交換データセットのみ
- 散文データセットのみ
- ランダムデータセットのみ
- 交換+散文①：交換データ 203,719 句と散文データ 66,000 句を結合。
- 交換+散文②：正例、交換データ、散文データの割合が1:1:1(各データ 135,813 句)になるように作成。散文データにはランダム選択によるオーバーサンプリングを適用。
- 交換+散文③：正例、交換データ、散文データの割合が2:1:1(俳句データ 203,719 句、交換データ・散文データは各データ 101,859 句)になるように作成。散文データにはランダム選択によるオーバーサンプリングを適用。

過学習を防ぐ早期停止を導入し、検証データに対して損失が最も低いモデルを早期停止のモデルとした。

3.2 実験結果

検証データに対する識別結果を表2に示す。表中の数値は、各データセットに対して俳句・非俳句の2クラス分類を行い、正しく識別できた割合を示している。例えば、見出しが「俳句」の列は、各モデルが俳句データに対して、俳句であると正しく識別できた割合である。

早期停止を用いた場合は、用いていない場合と比較して、正例の俳句データに対してはすべての条件において正しく識別する割合が向上し、負例に対してはすべての条件において正しく識別する割合が低下した。

正例である俳句を正しく識別した割合が高かったモデルは、早期停止に関わらず、負例として散文データのみ、またはランダムデータのみを用いたモデルである。ただし、これらのモデルでは交換データを負例として識別する割合が低い。また、負例として用いた交換、散文、ランダムの各データに対して、正しく負例と識別した割合が高かったモデルは交換+散文①である。学習データの負例が1種類だけのモデルよりも負例に対しては正しく識別する割合が高い。ただし、交換+散文①のモデルは正例の俳句データに対して、他のモデルよりも識別精度が低い。

実験結果から、学習データに用いる負例によって、正例と負例の識別の精度に大きな差が出る事が確認された。俳句として成立する候補を選択する場合には、既に大量の俳句候補群が生成可能であることを前提とすると、負例の識別精度が高い交換+散文①を負例として学習したモデルを用いることが望ましい。

4 まとめ

本稿では、実際の俳句データセットを正例とし、青空文庫データを用いて生成した俳句ではない文字列を負例とするデータセットを生成し、マスク化言語モデル RoBERTa を用いた俳句として成立する文字列を選択する俳句評価器を構築した。実験の結果、この評価器によって、俳句とみなすことができない候補を高い精度で識別可能であることが確認された。

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, July 2021.
- [2] 横山想一郎, 山下倫央, 川村秀憲. 深層学習を用いた俳句の生成と選句. 人工知能, Vol. 34, No. 4, pp. 467-474, 2019.
- [3] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized BERT pretraining approach. *arXiv*, 2019. cite arxiv:1907.11692.