

# 機械学習を用いた川柳の面白さの予測

太田聖三郎<sup>1</sup> 河原大輔<sup>1</sup> 野村理朗<sup>2</sup>

<sup>1</sup> 早稲田大学理工学術院 <sup>2</sup> 京都大学

{ota-seizaburo@akane.,dkw@}waseda.jp, nomura.michio.8u@kyoto-u.ac.jp

## 概要

川柳は日本の伝統文芸の一つである。詩や俳句に関する機械学習を用いた研究は散見されるが、川柳に関するものはない。本研究では、川柳を構成する要素を様々な指標に分解し、それらの予測を組み合わせることで、より高精度な川柳評価を予測する手法を提案する。クラウドソーシングを用いて川柳データセットを構築し、BERT-like モデルをファインチューニングすることより川柳の評価を予測する。それに加え、川柳をデータに事前学習モデルの追加学習を行い、川柳の特徴を理解した SenryuBERT を構築する。

## 1 はじめに

近年、人工知能による創作活動はめざましい発展を遂げている。自然言語処理においては詩や俳句などの生成研究 [1, 2] がある。その一方で、それらの理解に関する研究はほとんどされていない。それらの理解には創作物の持つ面白さや独創性などが関わってくる。面白さは人のコミュニケーションの中でも大きな役割を担っている。つまり、人と計算機のコミュニケーションを円滑にするためには、計算機による面白さの理解が重要な要素となる。

本研究では川柳という文芸を対象とし、人が川柳の面白さをどのように理解しているか、また、計算機がそれを理解できるかを分析する。川柳は俳句と同様に五・七・五の音数を持ち、元来は俳諧連歌の長句が独立したものである。短句に対しての前句付けが流行したことにより生まれた文芸であり、口語的で風刺的な内容が特徴である。川柳の面白さは人が理解するにおいても難しく、これを計算機に理解させるのはさらに難しいタスクである。この問題に取り組むにあたって、川柳の面白さを構成する要素を細分化し、判断しやすい指標を設けることで、川柳の総合的な面白さの予測を目指す。川柳の面白さを定量分析するために、クラウドソーシングによ

て各指標をアノテーションする。得られたアノテーションをもとに言語モデルのファインチューニングを行い、川柳を入力し川柳の総合的な評価を予測するタスクを行う。川柳の評価を直接予測する場合と、細分化した指標の予測を組み合わせる場合の精度の差を比較する。

## 2 関連研究

**詩・俳句生成** 川柳について機械学習を用いた研究は無い。しかし、川柳に通ずるものとして、近年では詩や俳句の生成 [1, 2] が行われている。Hitsuwari ら [2] は人手評価において、人工知能が生成した俳句と人間が作った俳句との見分けがつかないほど完成度の高い川柳の生成を実現している。

**BERT/RoBERTa** 近年の自然言語処理研究の飛躍的な進歩には BERT [3] の登場が大きく関わっている。BERT の学習は事前学習とファインチューニングからなる。事前学習は自己教師あり学習であり、主な学習方法は Masked Language Model (MLM) である。MLM は入力文の一部のトークンをマスクし、マスクされたトークンを予測するタスクである。予測されるトークンのうち 80% は [MASK] トークンに置き換えられ、10% はランダムな他のトークンに置き換えられる。残りのトークンは元のトークンのままで残される。これは、ダウンストリームタスクにおいて [MASK] トークンが出現しないことを考慮したためである。Liu ら [4] は BERT を改良したモデル RoBERTa を提案し、データセットの増大や動的なマスクングなどにより精度を向上させた。

**大喜利データセット** 中川ら [5] は大喜利を複数の指標に分解し、クラウドソーシングを用いて各指標に対しアノテーションを行うことで、定量分析可能な大喜利データセットを構築している。

## 3 川柳データセットの構築

本研究では川柳を構成する要素を複数のわかりやすい指標に分解し、定量分析を行うために川柳デー

タセットを構築する。

### 3.1 川柳データの収集

川柳は川柳投稿サイト「まるせん」<sup>1)</sup>より提供を受けたものを用いる。このデータは川柳、お題、作者を含む581,336句からなる。

### 3.2 川柳データの前処理・フィルタリング

本研究では、川柳の評価をクラウドソーシングを用いてアノテーションする。アノテーション対象の川柳を選別するために、川柳の前処理・フィルタリングを行う。投稿川柳にはしばしば絵文字を含んでいるものがある。これらは実験で用いる言語モデルの語彙に含まれていないため削除する。また、上記サイトではお題を作者が自ら設定し投稿するため、不適当なお題やイレギュラーなお題が散見される。そのため、クラウドソーシングにかかる川柳については、運営によりあらかじめお題が指定されている「コンテスト」の川柳を採用する。コンテストはサイト内で定期的に開催され、その度に10句ほどの入選作品が選出されるため、ある程度の川柳の質が担保される。データ収集時点までに開催されたコンテスト304回のうち、時事的なお題(例: VAR)や長いお題(例: 最近の若いもん)を除き名詞や形容詞を中心に250題について、フィルタリングを行った。各お題について入選作品+それ以外を合わせ20句ずつを選出し、合計5,000句の川柳を抽出した。

### 3.3 川柳の面白さの細分化と評価のアノテーション

前節で述べた5,000句の川柳に対し、クラウドソーシングを用いてアノテーションを実施する。プラットフォームはYahoo!クラウドソーシング<sup>2)</sup>を用いる。川柳の解説書[6,7]を参考に、川柳を8つの指標に細分化し、それぞれの指標について5段階評価を行ってもらおう。細分化した指標を下記に示す。

1. 句がお題に沿っている (along)
2. 場面をイメージしやすい・わかりやすい (imaginable)
3. 句に対比の構図がある (contrast)
4. あたりまえ・ありきたりなことは言っていない (usual)
5. 不適切な表現が含まれていない (appropriate)

1) <https://marusenryu.com/>

2) <https://crowdsourcing.yahoo.co.jp/>

6. 語のリズム感・軽快さがある (rhythmic)
7. 言い回しに独創性がある (unique)
8. 体験や体感への想いが感じ取れる・余韻がある (experience)

これらの項目に、

- 総合的に良い川柳である (overall)

を加え、合計9個の指標について、1句あたり10人のクラウドワーカーを雇い、その平均値を正解ラベルとして付与する。クラウドワーカーにはお題、川柳が与えられる。

## 4 SenryuBERTの構築

既存の言語モデルは新聞やウェブテキストで学習しているため、川柳の風刺的な文脈や語順などのリズム感を理解できない可能性がある。この問題に取り組むために、川柳データを活用し、言語モデルの追加学習を行うことでSenryuBERTを構築する。

### 4.1 ランダムトークン置換 (RTR)

2節で述べた通り、BERT[3]のMLMでは、予測するトークンのうち、80%を[MASK]トークン、10%をランダムな他のトークン、10%をそのままのトークンにして学習を行う。川柳のような短い系列を入力とする場合、[MASK]トークンと異なり、ランダムなトークンに置き換えると、BERTの特徴量はランダムなトークンの影響を大きく受けるため、文脈が破綻する可能性がある。これを防ぐため、通常のMLMに加えて、ランダムなトークンの置き換えをしない学習も行い比較する。

### 4.2 追加学習

川柳データを用いた言語モデルの追加学習は、早大RoBERTa-base<sup>3)</sup>をベースとして行う。学習は表1に示す4種類の条件で行う。モデル3の場合は入力を“お題 [SEP] 川柳”とし、その他のモデルでは“川柳”とする。また、学習時にマスクされるトークンにお題は含まれないようにする。実験で使用するハイパーパラメータを付録Bに示す。

### 4.3 結果

各モデルの[MASK]トークンの予測例を表2に示す。例には条件1, 2, 3とベースラインの早大

3) <https://huggingface.co/nlp-waseda/roberta-base-japanese-with-auto-jumanpp/>

表1 追加学習における学習条件設定

条件	[MASK]	同じ	ランダム	お題付き
1	80%	10%	10%	無し
2	90%	10%	0%	無し
2'	80%	20%	0%	無し
3	80%	10%	10%	有り

表2 [MASK] トークンの予測例 (入力: “歯ブラシと一緒に [MASK] あの”). BL はベースライン (RoBERTa) を表す.

条件	ランク1		ランク2		ランク3	
	トークン	確率	トークン	確率	トークン	確率
BL	の	0.109	ない	0.058	いる	0.045
1	使う	0.191	使った	0.060	捨てる	0.038
2	消えた	0.441	消える	0.034	捨てた	0.027
3	シャンプー	0.116	消えた	0.049	帰った	0.028

RoBERTa-base の結果を示す. 付録 C に他の予測例を掲載する. 定量的な評価は次節のダウンストリームタスクにて行う.

**定性的評価** ベースラインの RoBERTa では助詞や関係のない名詞・助詞が予測され, 正解と近い単語の出力は見られない. 条件3のお題と川柳を入力として学習したモデルではベースラインほど遠くはないものの, 正解に近い予測はほとんど見られなかった. 条件1と2のモデルはどちらもある程度のニュアンスが捉えられており, 正解に近い予測ができています. 条件2のモデルでは生起確率が他と比べて高くなっており, これはランダムトークンによる文脈の揺れが抑えられたためと思われる.

## 5 川柳構成要素の分析・予測

本研究では, 川柳の面白さという曖昧な要素を直接予測する場合と, 3節で示した8つのわかりやすい指標に分解してから予測する場合の精度を比較する. また, 予測した各指標のスコアから総合評価を予測するタスクを行い, 総合評価の予測精度の比較を行う. まとめると下記の3つの実験を行う.

1. 川柳から総合評価の予測
2. 川柳から各指標の評価の予測
3. 各指標の評価から総合評価の予測

実験1, 2には既存のモデルや4節で構築したモデルなどの BERT-like モデルを使用し, 実験3には決定木アルゴリズムの LightGBM [8] を使用する.

### 5.1 実験設定

実験1, 2では事前学習済み BERT-like モデルを回帰問題としてファインチューニングすることで, 川

表3 総合評価の予測結果

モデル	Pearson (ave)	Pearson (max)
BERT <sub>BASE</sub>	0.353	0.373
RoBERTa <sub>BASE</sub>	0.314	0.412
SenryuBERT	0.246	0.419
- RTR (80%)	0.418	<b>0.490</b>
- RTR (90%)	0.350	0.462
+ お題	<b>0.420</b>	0.456

柳から総合評価や各指標の評価を予測する. 4節で構築した SenryuBERT に加え, BERT や RoBERTa での学習を行う. モデルの詳細と学習に用いたハイパーパラメータは付録 A, B に示す. 学習に使用するデータセットは3節で構築した川柳データセットを用いる. データセットは訓練用4,000句, 検証用500句, テスト用500句に分割する. 川柳はお題ありきであるため, モデルへの入力は “川柳 [SEP] お題” とする. 各指標についてファインチューニングを10回行い, 精度を平均値と最大値で算出する. モデルの評価には Pearson の相関係数を用いた.

実験3ではまず, 実験1, 2から最良のモデルを選択し, 訓練データの各指標を予測する (自動ラベルと呼ぶ). 各指標の自動ラベルから LightGBM を用いて総合評価を予測し, 実験1のスコアと比較する. LightGBM の学習に用いたハイパーパラメータを付録 B に示す. 訓練データに自動ラベルを付与する手順は, まず, 訓練用データを10分割する. そのうち9個のデータ (3,500句) を用いて学習したモデルで, 残りの1個のデータに対して各指標のスコアを予測しラベルとする. これを10個の分割全てで行い, 4,000句に自動ラベルを付与する.

### 5.2 実験結果と考察

**川柳から総合評価の予測** 川柳から直接総合評価を予測した際の評価結果を表3に示す. 3節で提案した SenryuBERT は RoBERTa をベースとしたモデルであるが, RoBERTa ベースのモデルに関してはすべて BERT よりも最大値で良い精度となったが, RoBERTa と SenryuBERT の間に大きな差はみられなかった. しかし, SenryuBERT において RTR を行わなかった場合や, お題を入力に加えて学習した場合は精度の向上が見られた. 最良のモデルは SenryuBERT において, RTR を行わず, [MASK] トークンの割合を80%とした場合で, SenryuBERT とは最大で約0.07の差が見られた. ただし, お題を入力に加えて学習した SenryuBERT 以外のモデルは精度

表4 各指標の評価の予測結果

モデル	Pearson (ave)	Pearson (max)	Pearson (ave)	Pearson (max)	Pearson (ave)	Pearson (max)
	along		imaginable		contrast	
SenryuBERT	0.092	0.227	0.280	0.322	0.154	0.324
- RTR (80%)	0.178	<b>0.330</b>	0.357	0.422	<b>0.307</b>	<b>0.407</b>
- RTR (90%)	<b>0.208</b>	0.300	<b>0.359</b>	<b>0.430</b>	0.251	0.373
	usual		appropriate		rhythmic	
SenryuBERT	0.093	0.177	0.124	0.242	0.092	0.287
- RTR (80%)	0.158	<b>0.216</b>	<b>0.246</b>	<b>0.352</b>	<b>0.260</b>	0.330
- RTR (90%)	<b>0.185</b>	0.213	0.244	0.320	0.214	<b>0.333</b>
	unique		experience			
SenryuBERT	0.182	0.279	0.227	0.412		
- RTR (80%)	<b>0.325</b>	0.352	0.423	0.476		
- RTR (90%)	0.279	<b>0.363</b>	<b>0.471</b>	<b>0.496</b>		

表5 各指標における川柳評価の予測例と比較(川柳: “札束を 数え快感 銀行員”, お題: “現実逃避”)

手法	along	imaginable	contrast	usual	appropriate	rhythmic	unique	experience	overall
正解ラベル	2.9	3.9	2.7	2.8	3.5	3.3	3.0	3.4	2.9
直接予測	-	-	-	-	-	-	-	-	3.264
提案手法	3.366	4.038	3.585	3.495	3.434	3.197	3.074	3.714	3.212

が安定せず、データの順序などの影響を受けやすいことが分かった。表5に総合評価を川柳から直接予測する場合と提案手法との比較例を示す。提案手法が僅かに良い予測ができてることがわかる。表5に示す指標は3節で示した指標に対応している。

**川柳から各指標の評価の予測** 川柳を入力として3節に示した各指標の評価を予測した結果を表4に示す。表4にはSenryuBERTのファインチューニングによる結果と、総合評価の予測にて高精度であったRTRを行わないモデルの結果を示す。いずれの指標においても、RTRを行わないモデルはSenryuBERTの精度を上回る結果となった。この内、総合評価の予測より高い精度となったのは、体験や体感への想いが感じ取れる・余韻がある(experience)のみであった。

**各指標の評価から総合評価の予測** 本実験では、川柳から各指標の評価の予測において精度の高かったSenryuBERT-RTR(90%)のモデルを用いて訓練データに自動ラベルを付与し、LightGBMの学習を行う。川柳から直接総合評価を予測した結果と、各指標のスコアの予測から総合評価を予測した結果を表6に示す。データセットの各指標の正解ラベルをもとに総合評価を予測する場合には、非常に高い精度で予測ができています。これは、細分化した8個の指標が、人が川柳の面白さをどう理解しているかをうまく表す基底になっていると理解できる。自動

表6 総合評価の予測結果の比較

モデル	Pearson
LightGBM (正解ラベル)	0.8912
LightGBM (自動ラベル)	0.4760
SenryuBERT (直接予測)	0.4618

ラベルで学習した場合には、正解ラベルと比べて大幅に精度は下がるが、川柳から直接予測する場合と比べると、0.01ほど高い精度が得られ、提案手法がやや高い精度で予測できていることがわかる。自動ラベルの精度があまり高くない原因として、experienceしかoverallの精度に勝っていないことが考えられる。

## 6 おわりに

本研究では、現状取り組まれていない川柳という文芸を理解するタスクに機械学習を用いて取り組んだ。川柳から直接総合評価を予測するアプローチをベースラインとして、川柳を相対的にわかりやすい指標に細分化し、それぞれの項目の予測を組み合わせることで総合的な評価を予測する手法を提案した。提案手法はベースラインに比べて若干の精度の向上が見られた。

今後、川柳の面白さの理解を目指すには更なる精度の向上が必要である。また、提案手法を活かし、川柳の生成タスクに取り組んでいきたい。

## 謝辞

川柳データを提供いただいた川柳投稿サイト「まるせん」に感謝する。本研究は JSPS 科研費 JP22H01103, JP21H04901 の助成を受けて実施した。

## 参考文献

- [1] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Jozefowicz, and Samy Bengio. Generating sentences from a continuous space. 2015.
- [2] Jimpei Hitsuwari, Yoshiyuki Ueda, Woojin Yun, and Michio Nomura. Does human-ai collaboration lead to more creative art? aesthetic evaluation of human-made and ai-generated haiku poetry. **Computers in Human Behavior**, Vol. 139, p. 107502, 2023.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [4] 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, 2019.
- [5] 中川裕貴, 村脇有吾, 河原大輔, 黒橋禎夫. クラウドソーシングによる大喜利の面白さの構成要素の分析. 言語処理学会年次大会発表論文集, Vol. 25, pp. ROMBUNNO.B3-2 (WEB ONLY), 2019.
- [6] 新家完司. 川柳の理論と実践. 新葉館出版, 2011.
- [7] 野林正路. 詩・川柳・俳句のテキスト分析. 和泉書院, 2014.
- [8] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. **Advances in neural information processing systems**, Vol. 30, pp. 3146–3154, 2017.

## A 学習に使用したモデル

5 節の実験 1,2 で使用した BERT-like モデルの詳細を表 7 に示す。

表 7 学習に使用した事前学習モデルの詳細

model	corpus	#params	#dimension	vocabulary size
BERT <sub>BASE</sub> (cl-tohoku/bert-base-japanese-whole-word-masking)	Wikipedia	110M	768	32,000
RoBERTa <sub>BASE</sub> (nlp-waseda/roberta-base-japanese-with-auto-jumanpp)	Wikipedia & CC-100	125M	768	32,000

## B ハイパーパラメータ

表 8 に、4 節の実験で使用したハイパーパラメータを示す。また、表 9 には、5 節の実験 1,2 で使用したハイパーパラメータを示す。表 10 に示すのは、5 節の実験 3 で使用したハイパーパラメータである。

表 8 RoBERTa の追加学習に使用したハイパーパラメータ

hyperparameter	
learning rate	1e-5
batch size	16
#epoch	50
#warm up steps	50000

表 9 BERT-like モデルのファインチューニングに使用したハイパーパラメータ

hyperparameter	
learning rate	5e-5
batch size	8
#epoch	15
#warm up steps	500

表 10 LightGBM の学習に使用したハイパーパラメータ

hyperparameter	
metric	l2 (mse)
#leaves	31
learning rate	0.01
feature fraction	0.7
bagging fraction	0.8
bagging frequency	5
min child samples	20
#epoch	15

## C SenryuBERT の予測例

表 11 に、SenryuBERT による [MASK] トークンの予測例を示す。

表 11 [MASK] トークンの予測例 (入力: “[MASK] の財布の中身 500 円”)

条件	ランク 1		ランク 2		ランク 3	
	トークン	確率	トークン	確率	トークン	確率
BL	彼氏	0.039	付録	0.031	雑誌	0.020
1	ブランド	0.770	金持ち	0.013	母さん	0.007
2	休日	0.071	ブランド	0.060	コンビニ	0.031
3	コンビニ	0.061	プレゼント	0.043	誕生日	0.034