

機械学習を用いた川柳の面白さの予測 Predicting the Quality of Senryu using Machine Learning

太田 聖三郎[†], 河原 大輔[†], 野村 理朗[‡]

Seizaburo Ota, Daisuke Kawahara, Michio Nomura

[†] 早稲田大学, [‡] 京都大学

Waseda University, Kyoto university

ota-seizaburo@akane.waseda.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。川柳を使った事前学習モデルやクラウドソーシングを用いた川柳データセットを構築し、BERTのファインチューニング（パラメータの微調整）により川柳の評価を予測する。

キーワード:川柳, BERT, 言語モデル, ファインチューニング

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。理解には創作物の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重要である。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大事な要素となる。

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、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って各指標をアノテーション（データのラベル付け）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、川柳に関連して、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を比較し、それらの見分けがつかないほど完成度の高い俳句が生成されていることを示した。

BERT/RoBERTa 近年の自然言語処理研究はBERT [3] の登場により大きく進歩した。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」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。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ける川柳については、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で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う。

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)
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)

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

8. 体験や体感への想いが感じ取れる・余韻がある (experience)

上記の項目に、

- 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、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、1 人あたり 10 句を評価してもらう。

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、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 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、学習時にお題はマスクされない。

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

表 2 [MASK] トークンの予測例（入力：“歯ブラシと一緒に [MASK] あのだ”）。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。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し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より高くなっており、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 モデルをファインチューニングすることで、川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。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。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

表 3 総合評価の予測結果

モデル	Pearson (ave)	Pearson (max)
BERT _{BASE}	0.353	0.373
RoBERTa _{BASE}	0.314	0.412
SenryuBERT	0.246	0.419
- RTR (80%)	0.418	0.490
- RTR (90%)	0.350	0.462
+ お題	0.420	0.456

インチューニングを 10 回行い、精度を平均値算出する。モデルの評価には Pearson の相関係数を用いた。

実験 3 では、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。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。SenryuBERT は RoBERTa をベースとしている。SenryuBERT, RoBERTa 共に、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いたことがわかる。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

表 4 各指標の評価の予測結果

モデル	Pearson (ave)	Pearson (ave)	Pearson (ave)	Pearson (ave)
	along	imaginable	contrast	usual
SenryuBERT	0.092	0.280	0.154	0.093
- RTR (90%)	0.208	0.359	0.251	0.185
	unique	experience	unique	experience
SenryuBERT	0.124	0.092	0.182	0.227
- RTR (90%)	0.244	0.214	0.279	0.471

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

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

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。川柳から総合評価を直接、また各指標のスコアの予測から予測した結果を表 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。

文献

- [1] Bowman, S. R., Vilnis, L., Vinyals, O., Dai, A. M., Jozefowicz, R., & Bengio, S. (2015) Generating sentences from a continuous space.
- [2] Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023) Does human-ai collaboration lead to more creative art? aesthetic evaluation of human-made and ai-generated haiku poetry. *Computers in Human Behavior* **139**, 107502.
- [3] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019) *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. (Association for Computational Linguistics, Minneapolis, Minnesota), pp. 4171-4186.
- [4] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019) Roberta: A robustly optimized bert pretraining approach.
- [5] 中川裕貴, 村脇有吾, 河原大輔, & 黒橋禎夫. (2019) 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* **25**, ROMBUNNO.B3-2 (WEB ONLY).
- [6] 新家完司. (2011) *川柳の理論と実践*. (新葉館出版).
- [7] 野林正路. (2014) *詩・川柳・俳句のテキスト分析*. (和泉書院).
- [8] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017) Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* **30**, 3146-3154.