

大規模言語モデルを用いた日本語ラップ創作支援システム

廣晴輔^{1,a)} 塚田浩二¹

概要: 近年ラップ (HIPHOP) は世界的に人気のある音楽ジャンルであり, J-POP など別ジャンルにおいても, 歌詞の一部にラップや押韻表現を取り入れる例が増えている. ラップ歌詞 (Verse) の作成は, 意味的な表現に加えて, 押韻表現 (Rhyme) を組み合わせる必要があるため, 難易度が高い. そこで, 本研究では Verse 生成の初級者を対象として, 任意のキーワードを入力として, 大規模言語モデルを介して, 意味表現・押韻表現を備えた Verse を生成する創作支援システムを提案する. キーワードの入力やシステムの提示する歌詞候補からの選択を対話的に行うことで, 歌詞のテーマやユーザの意思を反映した一貫性のある Verse の生成を目指す.

1. 背景と目的

アメリカ 3 大音楽賞の一つであるビルボード・ミュージック・アワード 2022^{*1}のトップアーティストにラッパーの Drake が選出されたように, 近年世界的にラップ (HIPHOP) は人気のある音楽ジャンルである. また J-POP など別ジャンルにおいても, 歌詞の一部にラップや押韻表現を取り入れる例が増えている. 一方, ラップ歌詞 (Verse) の作成は, 意味的な表現に加えて, 押韻表現 (Rhyme) を組み合わせる必要があるため, 難易度が高い.

図 1 に通常の歌詞作成と Rhyme 表現を用いた歌詞作成手順を比較した例を示す. このように, 「押韻を考慮」するプロセスが増えることで, 歌詞作成の難易度が高くなっていると考えられる.

そこで, 本研究では Verse 作成の初級者を対象として, 大規模言語モデルを用いて, 意味表現・押韻表現を備えた Verse を生成する創作支援システムを提案する. キーワードの入力やシステムの提示する歌詞候補の選択を対話的に行うことで, ユーザの意思を反映した一貫性のある Verse の生成を目指す.

2. 関連研究

ラップの自動生成に関する研究は英語圏を中心に多数行われている.

Nguyen ら [1] は, ラップ歌詞や韻を踏んだ単語のデータ

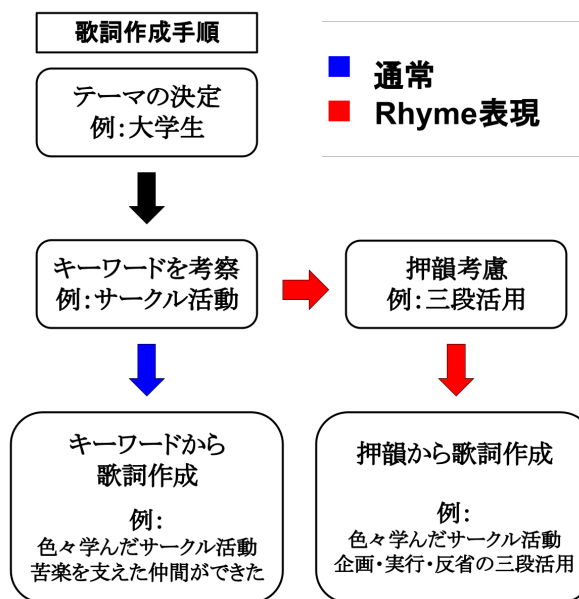


図 1 歌詞作成手順の比較

ベースを活用して, 既存の歌詞を入力として新しい歌詞を生成する手法を提案した. Barbieri ら [2] はマルコフ過程を用いてラップ歌詞の半自動生成システムを実装した.

また LSTM モデルを使用した研究も行われている. LSTM (Long Short Term Memory) モデルとは再帰ニューラルネットワークの一種であり, 短期的な情報と長期的な情報の両方を効果的に保持することができるため, 自然言語のような複雑なデータ構造に適したモデルとして広く使用されている. Potash [3] らは LSTM モデルを作成し, 対象にしたラッパーに似ている歌詞の自動生成に取り組んだ. また Yu ら [4] は LSTM モデルを改良して, 歌詞のみ

¹ 公立はこだて未来大学

^{a)} b1020189@fun.ac.jp

^{*1} <https://www.billboardmusicawards.com/2022/05/winners-announced-for-the-2022-bbmas/>

でなく楽曲の自動生成にも取り組んだ。

Wentink[5]は韻密度に着目して、ラップ歌詞の自動生成モデルを評価する手法を提案した。またNikolovら[6]は入力した文章をラップ調に変換する手法、Rapformerを提案した。例えば、新聞記事をラップ調に変換する事例を紹介している。

日本語ラップの生成に関する研究として、三林ら[7]はRapformerを拡張して、日本語ラップバトルを想定したVerse生成システムを構築した。日本語の文章を入力すると、Rapformer等を用いて応答文を作成し、更に双方向のエンコード表現を可能にした自然言語処理モデルBERT(Bidirectional Encoder Representations from Transformers)[8]を用いて韻の付与を行うシステムである。さらに、三林ら[9]はGPT-2とBERTを用いたラップバトル体験システムも開発している。これはユーザが音声入力したVerseに対して、システムで返答Verseを生成し、それを音声で出力するシステムである。GPT-2は大規模言語モデル(Large Language Model:LLM)の一種であり、大量のテキストデータから学習された、非常に大規模なパラメータセットを持つ。本研究でも、大規模言語モデルとBERTを用いたVerse生成システムを提案するが、歌詞の創作支援に注目する点が異なる。歌詞の創作においては、一曲を通して一貫したテーマが重要となるため、システムから提示したVerse候補をユーザに都度選択させ、ユーザの意思を反映しやすくするように工夫する。

3. 提案

本研究ではユーザが入力したキーワードを基に、大規模言語モデル(LLM)を用いて日本語ラップのVerseを生成する創作支援システムを提案する。

図2にシステムの概要図を示す。

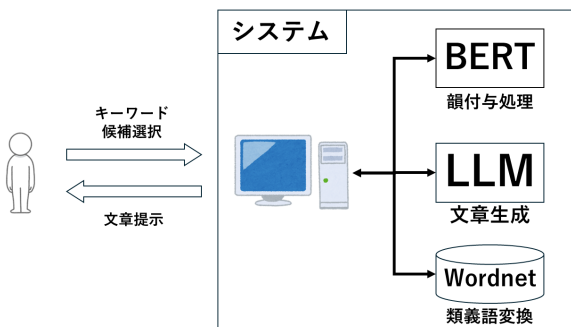


図2 システムの概要図

BERTは双方向のエンコード表現を可能にした事前学習モデルであり、文章内の一部に対してMASKトークンを付与して、単語の置き換え処理を行う学習方法がある。

図3にシステムのシーケンス図を示す。現在の実装では

図中のm(小節目数)は7, n(候補数)は3としている。

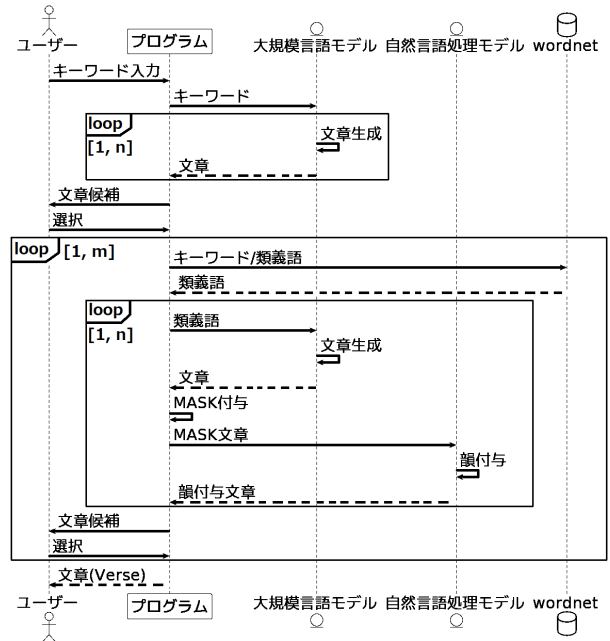


図3 システムのシーケンス図

以下にシステムの流れを記述する。

- (1) ユーザはシステムにキーワードを入力する。
- (2) システムはLLMを用いて、1小節目の文章候補をn個生成し、ユーザが選択する。(なお、選択した候補をユーザが変更することもできる。)
- (3) システムは日本語Wordnetを用いてキーワードから類義語を取得する。
- (4) システムは取得した類義語を新たなキーワードとしてLLMを用いて、2小節目の文章候補をn個生成する。
- (5) 2小節目の文章候補の文末にMASKトークンを付与する。
- (6) システムはBERTを用いて、1小節目の文章の文末と母音が一致するように2小節目の文章候補のMASKトークンを置き換える。
- (7) ユーザは生成された文章候補から一つを選択する。(なお、選択した候補をユーザが変更することもできる。)こうした処理をm回繰り返すことで、Verseを生成する。キーワードの入力や一小節目毎に歌詞の選択を求めることで、ユーザの意思に沿った形で、テーマに一貫性のあるVerseの生成を目指す。

またユーザとシステムの対話的な創作支援を行うためGUIを実装することで、生成された文章の変更や、自身の歌詞入力等をサポートしつつ、初級者でも扱いやすいように工夫する。

4. 実装

ここでは、システムの実装について説明する。

まず、メインプログラムはPythonで実装している。プ

プログラムの動作としては、ユーザからの入力を受け、LLM や BERT, Wordnet との橋渡し役を担っている。

LLM として OpenAI 社から提供されている API を使用してファインチューニングを行った gpt-3.5-turbo モデルを使用している。日本で活躍しているラップアーティスト・ラップグループから約 800 件の歌詞データをスクレイピングし、日本語で構成された小節のみをデータセットとしてファインチューニングを行った。

データセットの作成手順は、まず全ての歌詞データを 1 小節ごとに分解した。次に分解した歌詞データから日本語・数字以外の文字が含まれる小節を全て取りいて形態素解析を行い、小節内のランダムな名詞を取得した。これらの名詞を用いて、小節毎に以下のような構造のレコードを作成した。なお、【名詞】には抽出された名詞が、【小節】には各小節の全文が記載される。

- {" messages": [{" role": " user", " content": " 【名詞】 をキーワードとして 1 小節のラップを出力" }, {" role": " assistant", " content": " 【小節】" }] }

合計 7611 レコードのデータセットを構築した。OpenAI の API では処理するテキストの最小単位を Token として定めているが、ファインチューニングを行う場合データセットは仕様上*2 16,385Token 以下である必要があり、構築したデータセットからレコード数を限定する必要があった。よって、曲やアーティストに偏りが出ないように 300 レコードを抽出した。このようにして、ファインチューニングした LLM を文章生成等に使用する。

本研究ではファインチューニング後の LLM を、入力キーワードや類義語からの文章生成に使用する。

BERT を基盤としたさまざまなモデルが提案されており、本研究では、東北大学自然言語処理研究グループから公開されている bert-base-japanese-whole-word-masking モデル*3 を使用する。本研究では類義語から生成された文章の一部を MASK として BERT に入力することで、母音の一致度から韻を付与した単語で MASK を置き換える。

キーワードの類義語変換を行うために国立研究開発法人情報通信研究機構が公開している日本語 Wordnet[10] を使用する。日本語 Wordnet は日本語の概念辞書である。ユーザが入力したキーワードや一つ前に得られた類義語を、Wordnet で検索し、キーワードに紐づいた概念から類義語を複数出力する。

システムの GUI は Tkinter を用いて実装を行った。(図 4)。まず、ユーザはキーワードを入力する。次に LLM から生成された文章がボタンとして表示され、一つを選択す

ることで画面下部のテキストボックスに文章が反映される。ここでユーザは文章を任意に書き換えることができ、確定ボタンを押下すると画面右側のキャンバスにテキストが反映される。また類義語を新たなキーワードとして LLM で生成された文章を画面上に表示する。なお、更新ボタンを押下することで、LLM を用いて文章を再生成することもできる。

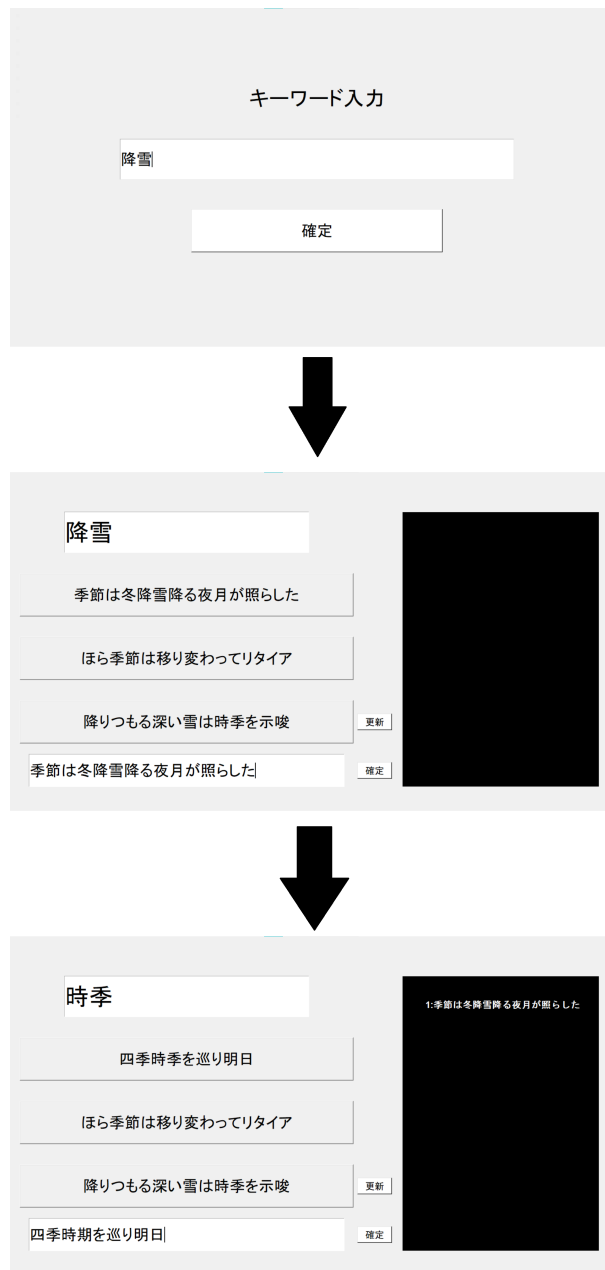


図 4 システムの動作例

5. 議論

表 1 に、「未来」というキーワードからシステムを用いて対話的に生成した Verse の出力例を示す。これは、明らかな誤字を訂正し、文章の長さも整えて対話的に生成した例である。

*2 <https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset>

*3 <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

表 1 未来をキーワードにした対話的な出力例

小節数	出力 verse
1	諦める勇気なんてないだけでもえぐる未来がくる [8]
2	信じる未来 (Future) どこにだってフューズム
3	取るも捨てるも後あとの自己言及
4	オチが添えりゃ此後が大変だ安シングル [9]
5	未来は金がじゃぶじゃぶ汚れた足跡描いた配り人数
6	降り積もる終雪どんな現場か知らんとおまえが口入 [10]
7	街の子供孫にも語るぜこの話に対して
8	流すテープあの頃の後皆一緒にて

6小節目の「口入」という単語は、5小節目の「配り人数」から「ku・nyu・u」という音節で生成されているが、「ku・ti・i・re」という音節で7小節目の単語「対して」の生成に利用されている。このような複数音節を有する単語への対応も検討していく。

6. まとめと今後の展望

本研究では Verse 作成の初級者を対象として、任意のキーワードを入力として、大規模言語モデルを介して、意味表現・押韻表現を備えた日本語 Verse を生成する創作支援システムを作成した。

今後の展望としてモデルの再学習・韻の付与や類義語選定手法の改善を進める。また、歌詞生成の初心者を対象とした評価実験を行う。Rhyme 表現を用いた歌詞作成を本システム有り/無し状態で体験してもらい、システムの効果や使いやすさを検証する。さらに、生成された Verse 自体に対する印象評価も実施する。

参考文献

- [1] Hieu Nguyen and Brian Sa, Rap lyric generator, Stanford, CS224N Final Projects, 2009.
- [2] Gabriele Barbieri, Francois Pachet, Pierre Roy, Mirko Degli Esposti, Markov constraints for generating lyrics with style, In Proceedings of the 20th European Conference on Artificial Intelligence (ECAI'12), pp.115-120, 2012.
- [3] Peter Potash, Alexey Romanov, Anna Rumshisky, Ghostwriter: Using an LSTM for automatic rap lyric generation, EMNLP' 15, pp.1919-1924, 2015.
- [4] Yi Yu and Simon Canales, Conditional LSTM-GAN for melody generation from lyrics, "ACM Transactions on Multimedia Computing, Communications, and Applications", Volume 17, Issue 1, Article No.:35, pp.1-20, 2021.
- [5] Matthias Wentink, Creating and Evaluating a Lyrics Generator Specialized in Rap Lyrics with a High Rhyme Density, TScIT 38, February 3, 2023.
- [6] Nikola I Nikolov, Eric Malmi, Curtis Northcutt, Loreto Parisi, Rapformer: Conditional rap lyrics generation with denoising autoencoders, International Natural Language Generation Conference 2020(INLG' 20), pp.360-373, 2020.
- [7] 三林亮太, 橋口友哉, 山本武洋, 大島裕明, ラップバトル

におけるライムの意味類似性を考慮したバース生成システム, DEIM Forum 2021 F14-5, 2021.

Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, North American Chapter of the Association for Computational Linguistics 2019(NAAACL' 19), pp.4171-4186, 2019.

三林亮太, 山本武洋, 大島裕明, GPT-2 と BERT を用いたバース生成によるラップバトル体験システム, WISS 2021 論文集, 2021.

Francis Bond, Timothy Baldwin, Richard Fothergill and Kiyotaka Uchimoto, Japanese SemCor: A Sense-tagged Corpus of Japanese, The 6th International Conference of the Global WordNet Association (GWC-2012), 2012.