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

三林 亮太 *　山本 岳洋 *　大島 裕明 *

概要.　本論文では，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を提案する．ラップバトル体験システムでは，ユーザの音声を入力に，返答となるラップを音声で出力する．ラップバトル体験システムを実現するため，GPT-2 と BERT を用いたバース生成手法を提案する．提案手法では，ラップバトルにおけるライム（韻）とアンサー（返答）を考慮してバース生成をおこなった．結果として，たとえば「今回の WISS は静岡県 難しいお茶っ葉のさじ加減」という入力に対して,「静岡の俺がここで勝ち上がるぜ よしここで寝とけ 寝てても起こすぜ西軍共同 俺がもちろん感動」というバースが得られた．

1　はじめに

ラップバトルとは 2 名 [1]のラッパーがバース（即興ラップ）で対話する競技である．たとえば，図 1 に示すように，先攻のラッパーの「お前はできないライム作成 俺は忙しい博士学生」というバースに対して，後攻のラッパーは「博士学生ならば院で覚醒しろ　はっきり決めようぜ白か黒」という返答のバースを返す．これを数回繰り返し，どちらのラップが優れていたかを競う競技である．

バースには，ライムとアンサーを含む必要がある．ライムとは直感的には，ある単語に対して母音列が同一の語のことである．たとえば，図 1 のバース 1 に示す「作成」と「学生」は母音列が両者とも「AUEI」と同一であり，ライムであるといえる．アンサーとは，相手の発言をふまえた返答をすることである．たとえば，図 1 のバース 2 に示す「博士学生ならば院で覚醒しろ」というアンサーは，相手の発言であるバース 1 の「俺は忙しい博士学生」をふまえて返答しているので，アンサーといえる．

このような，ライムとアンサーを考慮したバース生成手法は，著者らの知る限り存在しない．ラップバトルにおけるバース生成の研究としては，Wu ら [5] と Savery ら [3] の 2 件しかなく，これらの研究では，ライムとアンサーは同時に考慮されていない．そこで，本研究では，ラップバトルにおけるライムとアンサーを考慮したバース生成手法を提案する．

2　バース生成

バース（即興ラップ）は図 2 のように 3 つのステップで生成する．第 1 のステップでは，GPT-2 モデルを用いて，相手の発言に対する返答の文を生成する．第 2 のステップでは，新たに用意した GPT-2 モデルを用いて，生成した返答の文から次の文を生成する．さらに，生成した文を入力に，次の文を生成する．この工程を繰り返して生成した文を並べることでバースを構成する．第 3 のステップでは，生成したバースの末尾に対して，BERT モデルを用いて，ライムを付与する．これらの工程により，ライムとアンサーを考慮したバースを生成する．

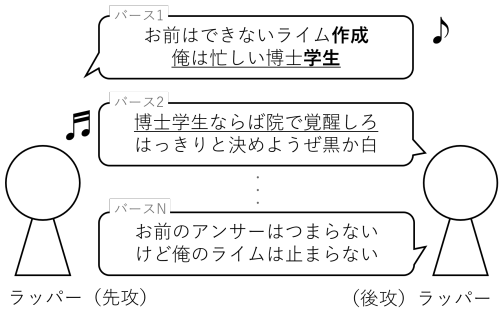

図 1. ラップバトルのイメージ

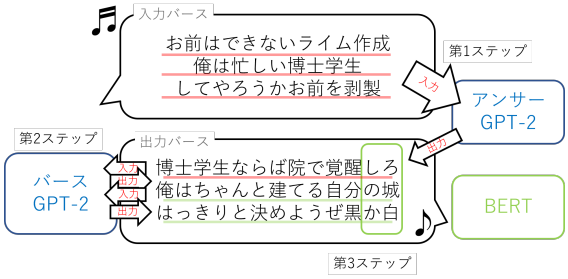

図 2. 提案手法の概要図

2.1　バース内の文生成

バース内の文生成には GPT-2 [2] を用いた．GPT-2 は Transformer [4] の Decoder を用いた自然言語

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

* 兵庫県立大学

[1] 3 対 3 など，複数人でおこなう場合もあるが，1 対 1 の 2 名が主流である．

文生成モデルである．本研究では，バース内の文を2種類に分けて生成した．1つ目は，バースの1文目で相手に対する返答をおこなう，アンサー文である．2つ目は，バース内をつなぐ文である，バース文である．この2種類の文を分けて生成するために，それぞれ個別の GPT-2 モデルを用いて生成をおこなった．

アンサー文の生成は，図 2 に示すように，入力バースと返答バースの1文目をペアにしてファインチューニングした．バース文の生成は，図 2 に示すように，返答バースの *n-1* 文目から *n* 文目をペアにしてファインチューニングした．

GPT-2 のファインチューニングには，rinna 株式会社の事前学習モデルを用いた．ファインチューニングをするにあたって，ラップバトルの文字起こしコーパスを作成した．コーパスは，YouTube に投稿されているラップバトル動画を文字起こしして作成した．文字起こしはクラウドソーシングにておこない，620 件のラップバトルからなるコーパスを作成した．

2.2 ライム付与

Nikolov らは生成済みのラップに対して，末尾の語をライムに置き換えることで，ライムを考慮したラップを生成した [1]．本研究では Nikolov らの手法を参考にし，GPT-2 で生成したバースに対してライムを付与した．まず，生成したバースに対して，末尾を特殊トークンの [MASK] に置き換える．置き換えた [MASK] を BERT モデルで予測し，すべての予測した語に対して，母音が一致しているかを判別する．すべての文に対して，母音が一致するようなライムが見つかれば，末尾の [MASK] と置き換える．

BERT のファインチューニングには，東北大学の事前学習モデルを用いた．ラップバトルコーパス内のすべての文に対して，末尾のトークンを [MASK] に置き換え，予測するファインチューニングをおこなった．

3 ラップバトル体験システム

ラップバトル体験システムは，音声ベースでラップバトルがおこなえる Web サービスである．2節で提案したバース生成手法と音声入出力 API を用いて実装した．システムは図 3 に示すように音声を入力し，音声のラップを出力する．

音声入力には Mozilla 社の Web Speech API を使用した．本 API は音声をリアルタイムでテキストに変換する API である．音声出力には CoeFont Cloud API を利用した．CoeFont Cloud API はテキストを特定の音声に変換する API である．本研究では，ラップに特化した音声を提供しているマチーデフ氏の CoeFont を利用した．

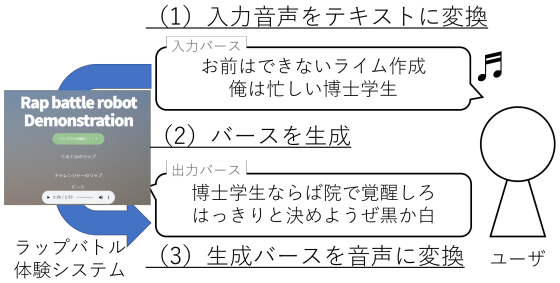

図 3. ラップバトル体験システムの例

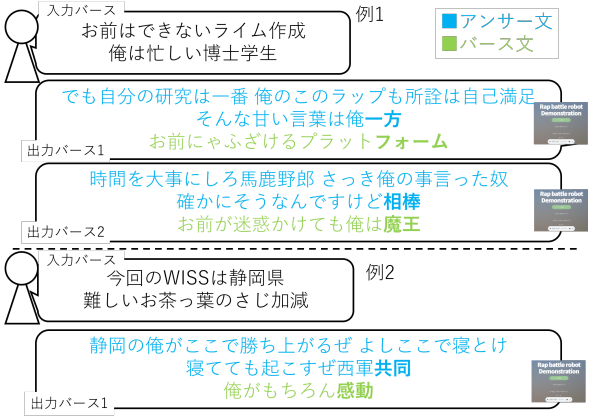

図 4. ラップバトル体験システムの出力

3.1 結果

ラップバトル体験システムの出力結果を図 4 に示す．結果として，いくつかの入力バースに対して，アンサーとライムが考慮されたバースが出力されていることがわかる．例1の出力バース1では，アンサーは入力バースの「忙しい」をふまえた,「時間を大事にしろ馬鹿野郎」の部分である．ライムは「相棒」と「魔王」であり，これらは「AOU」で母音が一致している．例1の出力バース2では，アンサーは入力バースの「博士学生」をふまえた,「研究が一番」の部分である．ライムは「一方」と「フォーム」であり,「OU」で母音が一致している．また，入力バースを例2のような異なる内容に変更しても，アンサーとライムが考慮されていることがわかる．アンサーは入力バースの「静岡県」をふまえた,「静岡の俺が」の部分である．ライムは「共同」と「感動」であり,「OU」で母音が一致している．

4 まとめ

本論文では，GPT-2 と BERT モデルを用いたバース生成手法の提案と，ラップバトル体験システムについて述べた．結果として，いくつかの入力に対して，ライムとアンサーを考慮したバースが生成できるようになった．

謝辞

本研究の一部は JSPS 科学研究費助成事業 JP21H03775，JP21H03774，JP18H03494，JP21H03554 による助成を受けたものです．ここに記して謝意を表します．

参考文献

[1] N. I. Nikolov, E. Malmi, C. Northcutt, and L. Parisi. Rapformer: Conditional Rap Lyrics Generation with Denoising Autoencoders. In *Proceedings of INLG'20*, pp. 360–373, 2020.

[2] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 2019.

[3] R. Savery, L. Zahray, and G. Weinberg. Shimon the Rapper: A Real-Time System for Human-Robot Interactive Rap Battles. *arXiv preprint arXiv:2009.09234*, 2020.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin. Attention is All You Need. In *Proceedings of NIPS'17*, pp. 6000–6010, 2017.

[5] D. Wu and K. Addanki. Learning to rap battle with bilingual recursive neural networks. In *Proceedings of IJCAI'15*, pp. 2524–2530, 2015.