

キャラクターロボットの動作生成に向けた大規模言語モデルの性能比較 Comparison of Large Language Model for Character Robot Motion Generation

小林 稜¹⁾ 佐藤 紘基¹⁾ 山内 翔¹⁾ 鈴木 恵二¹⁾
Ryo Kobayashi Koki Sato Sho Yamauchi Keiji Suzuki

1 はじめに

ロボットの社会的な利用拡大には、「人に愛されるロボット」の開発が必要不可欠である。「人に愛されるロボット」とは、人の関心を惹き、多様で継続的な行動をとるロボットである。人と密接に関わるロボットにおいて、ロボットの容姿がロボットの使われ方や継続的な利用期間に大きな影響を与えることが知られている。好感を持たれるデザインの特徴として、ベビースキーマやネオテニーと呼ばれる哺乳類の赤ん坊が持つ性質を持つことが知られている。こうしたデザインのロボットは、耐久性や運動性能の確保、奇抜なデザインに対する行動設計の難しさから、実現が困難な場合があった。しかし、山内・鈴木が提案した 3D プリンタを用いたロボット設計手法により、デザインを優先して設計することが可能となった [1]。

本研究では、3D プリンタを用いて、人に好感を持たれるデザインのキャラクターロボットを作成し、動物の動きを深層学習し、設計したロボットの動作生成システムを構築することで、より人に好かれるロボットの開発をおこなうことが目的である。ロボットの動作生成について、従来の研究では、動物の動きを模倣することで性能を向上する研究 [2] はあるが、人に好かれる・可愛がられるロボットの動作生成について、研究しているものはない。

関連研究として、オオカミの受容的行動をロボットに応用し、人に親和感を与えるかの心理実験 [3] がある。この実験では、オオカミの社会行動のうち、親和感を演出する受容的行動に着目し、ロボットにおいても人間の行動に対して、受容的な行動をおこなうことで親和感を与えるか、というものである。結果は、動物の一部行動が人に好意的な印象を与えることがわかった。

そこで、本研究では、ロボットの動作生成の動作基盤として、動物の動きを深層学習し、デフォルメすることで、人に好かれる行動を生成できると考えた。深層学習には、先読みし、穴埋め推論が可能である大規模言語モデルを採用した。

2 提案手法

図 1 に提案システムの全体像を示す。

動物の動きを深層学習するにあたり、本研究では、猫に着目する。猫は、好感を持たれる特徴をもつ哺乳類で、人と関係が近く、毛づくろいや首をかしげるなどのコミュニケーション行動が多く存在する。また、オンライン上に猫のデータが多くあることから、動作生成の基盤として猫をあげた。事前学習では、猫の動画から、猫の骨格を抽出し、骨格を 3D 化する。そして、それぞれの関節の角度の動きを収集する。収集した猫の関節の動きを学習できるように言語化を行い、猫の動きを学習する。ファインチューニングでは、3D プリンタで作成し

1) はこだて未来大学.Future University Hakodate

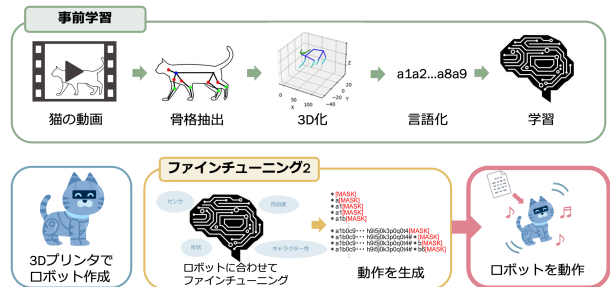


図 1 提案システムの全体像

たロボットのセンサ、形状、自由度、キャラクター性を考慮したファインチューニングをおこなう。そして、デフォルメ化して動作を生成することで、人に好かれるロボットを開発できると考えた。

本研究では、提案システムのうち、基盤となる事前学習部分について研究を実施し、深層学習に用いる大規模言語モデルの検証をおこなう。

3 大規模言語モデル

本研究では、深層学習に用いる大規模言語モデルとして、ALBERT[4] と gMLP[5] の 2 つのモデルによる実験をおこない、提案手法の学習モデルとしての有効性を検証する。

ALBERT は BERT[6] を軽量化した Transformer による双方向の Encoder を用いた自然言語処理モデルである。BERT は計 11 種類の自然言語タスクで評価したところ、全タスクで SoTA を達成した高精度なモデルである。しかし、学習が難しいことや学習に時間がかかること、NSP タスクの有効性が明確でないなどの問題があった。ALBERT はパラメータ数の削減と共有により、モデルの軽量化と学習速度の高速化を可能にした。また、SOP タスクを導入したことで、ALBERT は BERT を超えて SoTA を達成したモデルである。

gMLP は、多層パーセプトロンとゲート機構で構成された機械学習モデルである。ALBERT や BERT のような Attention 機構を持つ Transformer 型アーキテクチャとは異なり、Attention 機構を持たない MLP 型アーキテクチャである。BERT と性能を比較したところ、性能にほぼ差がないことが確認された高精度なモデルである。

以上から、2 つのモデルの推論能力をロボットに応用することで、違和感なく生物学的に破綻しない動作生成を際限なくおこなうことができると考えられる。

4 動作言語の定義

猫の動きを言語化するにあたって、本研究では独自の言語を定義した (以下、動作言語と呼ぶ)。表 1 にその概要を示す。動作言語は猫の各関節のピッチ、ロール、ヨーに対してアルファベットを付与し、その動きを数字として記述する。各関節の可動域を数字の数、等分する。そして、各関節位置に最も近い等分した位置に置換

し、離散化をおこない、数字として記述する。また、言語性を向上する試みとして、シーケンスの先端と末端を示すシンボルを設定した。加えて、変化した関節のみを羅列した 1 行を 1 シーケンスとする。

表 1 動作言語の概要

アルファベット (a から z)	各関節のピッチ, ロール, ヨーのいずれかを示す
数字 (0 から 9)	関節の角度の位置を示す
*	シーケンスの先頭を示す
#	シーケンスの末端を示す

5 動作言語の検証

ALBERT と gMLP が本研究で定義した動作言語が学習可能であるか検証をおこなった。図 2 にエポック数 10 で学習した際の損失関数の推移を示す。

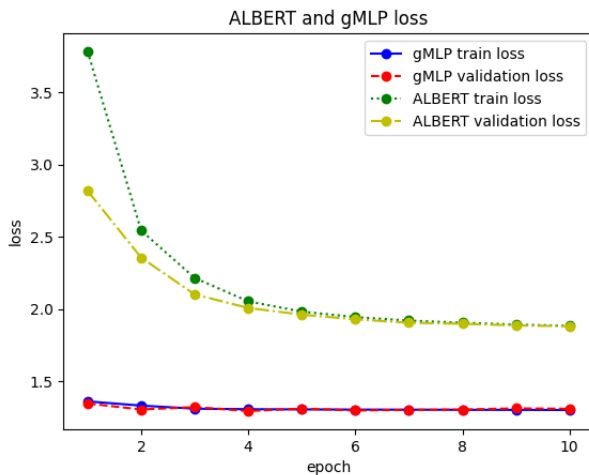


図 2 学習時の損失関数の推移

図 2 の結果から、ALBERT はエポック数が増えるにつれて、損失関数の数値が段階的に下がり、収束していることが確認できる。対して、gMLP は損失関数の推移が小さいものの ALBERT よりも損失関数の数値は低いことがわかった。ALBERT はエポック 10 が validation loss が最も小さく、gMLP はエポック 4 が validation loss の最も小さいモデルであることがわかった。

また、シーケンスを生成し、定義した言語形式に沿った生成が可能であるか検証をおこなった。生成には、穴埋め推論を用いて、最も高い推論候補を継ぎ足していくことで生成をおこなった。5,000 文字分生成し、終端文字を区切りとして分割してシーケンスとした。各シーケンス数は ALBERT が 138 個、gMLP が 142 個となった。そのシーケンスのうち、定義した言語形式として不適であるもののグラフにした。シーケンスが不適であるものの内訳として、定義した言語形式に沿っていないもの、一つのシーケンスに重複した関節があるもの、ある関節の関節位置が前回と同じ関節位置を示しているものを対象にし、言語形式に沿った生成ができるかを確認した。その結果を以下に示す(図 3)。

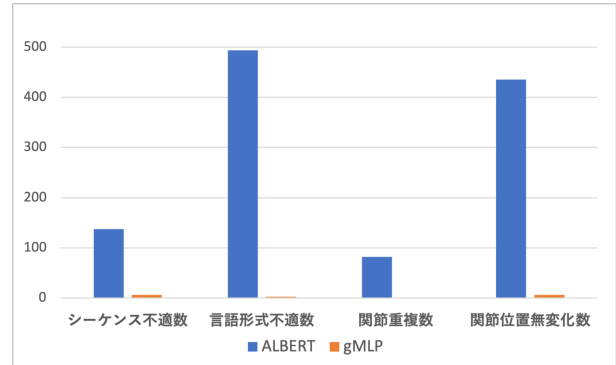


図 3 生成したシーケンスの検証結果

図 3 の結果から、gMLP は言語形式に沿った生成ができていたことが確認できた。対して、ALBERT はほとんどが形式に沿っていないことが確認できた。これは、学習に用いたデータがランダムデータであることが考えられ、特性をうまく学習できなかった可能性が考えられる。また、シーケンス生成において、推論候補の中から言語形式に沿ったものと抽出するなどの工夫をすることで、より形式に沿った生成が可能であると考えられるため、検証していく必要がある。

6 おわりに

本研究では、キャラクターロボットの動作生成に向け、大規模言語モデルをロボットに応用した動作生成システム構築を提案した。使用する大規模言語モデルとして、ALBERT と gMLP をあげ、本研究で定義した動作言語が ALBERT と gMLP で学習可能であるか検証をおこなった。今回の結果では、gMLP が ALBERT より優れた結果となった。

今後、猫の動作データで学習をおこない、生成したシーケンスが学習データの性質を有しているかを検証する。そして、ALBERT と gMLP での性能を比較する。また、ロボットシミュレーション環境を構築し、生成したシーケンスの動作を可視化し、違和感のない動作を生成しているか確認する。

参考文献

- [1] 山内翔, 鈴木恵二. 結晶構造を参考にした内部構造と 3d プリンタによるモデル形状を優先したロボット設計. 電気学会論文誌 C (電子・情報・システム部門誌), Vol. 139, No. 9, pp. 1051-1058, 2019.
- [2] Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Edward Lee, Jie Tan, and Sergey Levine. Learning agile robotic locomotion skills by imitating animals. In *Robotics: Science and Systems*, 07 2020.
- [3] 中田亨, 佐藤知正, 森武俊, 溝口博. ロボットの対人行動による親和感の演出. 日本ロボット学会誌, Vol. 15, No. 7, pp. 1068-1074, 1997.
- [4] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*, 2019.
- [5] Hanxiao Liu, Zihang Dai, David So, and Quoc V Le. Pay attention to mlps. *Advances in Neural Information Processing Systems*, Vol. 34, pp. 9204-9215, 2021.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.