物理シミュレータによる自己検証ループを用いた、高信頼な自然言語-ROS2 コマンドデータセットの構築手法

嶋中雄大†1 菅谷みどり†1

概要： 本稿は、大規模言語モデル（LLM）を用いた自然言語によるロボット制御の信頼性向上を目的とし、そのための高品質な学習データセットを自律的に構築する新手法を提案する。LLM の応用は、物理法則を無視した不正確なコマンドを生成する「ハルシネーション」という深刻な課題に直面している。この課題を解決するため、本研究では物理シミュレータ「Gazebo」を検証機構としてループに組み込んだ「シミュレータによる代理評価フレームワーク」を構築した。本フレームワークは、LLM が生成した ROS2 コマンド候補をシミュレータ上で実行し、その物理的な成否を客観的基準で自動判定する。これにより、成功したコマンドペアからなる高品質な SFT（教師ありファインチューニング）用データセットと、成功・失敗の対からなる DPO（直接的選好最適化）用データセットを同時に生成する。本アプローチは、ロボット制御におけるデータセット構築コストを大幅に削減し、安全で信頼性の高いロボット AI の実現を加速させるものである。特に、本手法で生成されたデータセットは、リソース制約のある組込みシステム上での動作を想定した軽量な推論モデルの育成を可能にする点で、実践的な価値を持つ。

キーワード：大規模言語モデル, ロボット制御, 物理シミュレーション, データセット構築

1. はじめに

近年、人間とロボットが協調する社会の実現に向け、専門家でないユーザーでも自然言語を用いて直感的にロボットを操作できる技術への需要が高まっている。大規模言語モデル（LLM）は、その卓越した言語理解能力から、この課題を解決する鍵として期待されている。

しかし、LLMを物理世界で動作するロボットに直接応用するには、三つの大きな障壁が存在する：

1. **ハルシネーション問題**:
汎用的なテキストデータで学習したLLMは、ロボットの物理的制約や制御システムの厳密な規約を理解しておらず、予測不能あるいは危険な動作を引き起こすリスクが常に存在する。

2. **データセット不足**:
LLMを特定のロボットタスクに適合させるための、高品質な「自然言語指示」と「正解コマンド」のペアからなる専門的なデータセットが絶対的に不足しており、その人手による構築コストは極めて高い。

3. **リソース制約**:
そもそも、巨大なLLMを直接ロボットに搭載することは、多くの組込みシステムが持つ計算能力や消費電力の制約から非現実的である。

そこで本研究では、これらの課題を解決するため、物理シミュレータを検証機構として活用し、高信頼なデータセットを低コストかつ自律的に構築するフレームワークを提案する。これにより、最終的にエッジデバイス上で動作可能な、小型で高効率な制御モデルの学習基盤を構築することを目指す。

2. 提案手法

本研究では、**「シミュレータによる代理評価フレームワーク」**を提案する。本フレームワークは、LLMによるデータ生成と、Gazeboシミュレータによる物理的検証を繰り返すことで、データセットの品質と規模を自律的に向上させる。

処理フロー
処理フローは以下の通りである
候補生成:
一つの自然言語指示に対し、LLMが複数の多様なROS2コマンド候補を生成する。
物理検証:
生成された各コマンドを、Gazeboシミュレータ上のロボットで実行する。ROS2で実装された監視ノードが、コマンド実行前後のロボットの状態（位置・姿勢）を観測する。
成否判定:
観測された状態と命令の意図を比較し、タスクごとにあらかじめ定義された成功基準に基づ

†1 芝浦工業大学
Shibaura Institute of Technology

き、客観的に成否を判定する。例えば、単純な移動タスクでは「目標位置との最終的なユークリッド距離誤差 < 5cm」といった基準を設定する一方、将来的により複雑なマニピュレーションタスクを扱う際には、「対象オブジェクトの正しい把持状態」や「目標配置位置・姿勢との誤差」といった多角的な基準を導入することを想定している。

データセット蓄積:
検証結果に基づき、データセットを構築する。

SFT用データセット:

検証に成功した< 自然言語指示, ROS2コマンド>のペアは、そのまま高品質なSFTデータとして`positive_samples.jsonl`に蓄積される。

DPO用データセット:

同一指示に対して得られた成功例をchosen、失敗例をrejectedとし、DPO学習用の選好ペアを構築する。

3. 関連研究

本研究は、(1)LLMのロボット制御応用、(2)シミュレーションの活用、(3)選好学習、という三つの研究分野の交差点に位置する。

LLMのロボット制御応用:
GoogleのPaLM-SayCanに代表される研究[1]は、LLMを「高レベルなタスクプランナー」として活用し、行動計画を生成するアプローチで大きな成功を収めている。しかし、これらの研究は主にタスクレベルの計画に焦点を当てており、各行動を実行するための低レベルな制御コマンド自体の物理的な妥当性を検証する機構は持たない。

シミュレーションの活用:
ロボティクス分野では、Sim-to-Realの文脈で、Domain Randomizationなどの手法[2]を用いてシミュレーションデータを生成し、実世界で頑健に動作するポリシーを学習させる研究が盛んに行われてきた。しかし、これらの多くは画像認識や状態からの強化学習が主目的であり、本研究のように「自然言語」と「物理的成果」を結びつけるためのグラウンディング機構としてシミュレーションを活用するアプローチは未だ開拓されていない。

ロボティクスにおける選好学習:
DPO[3]やRLHFといった選好学習は、人間の評価者がロボットの複数の動作を見て「どちらが良いか」をラベリングすることで、より人間に好まれる振る舞いを学習させる。このアプローチは効果的である一方、人間のアノテーションコストが膨大であり、スケーラビリティに課題を抱える。本研究は、この人間の評価者を「物理シミュレーションの客観的な成否判定」で代替する点に新規性がある。

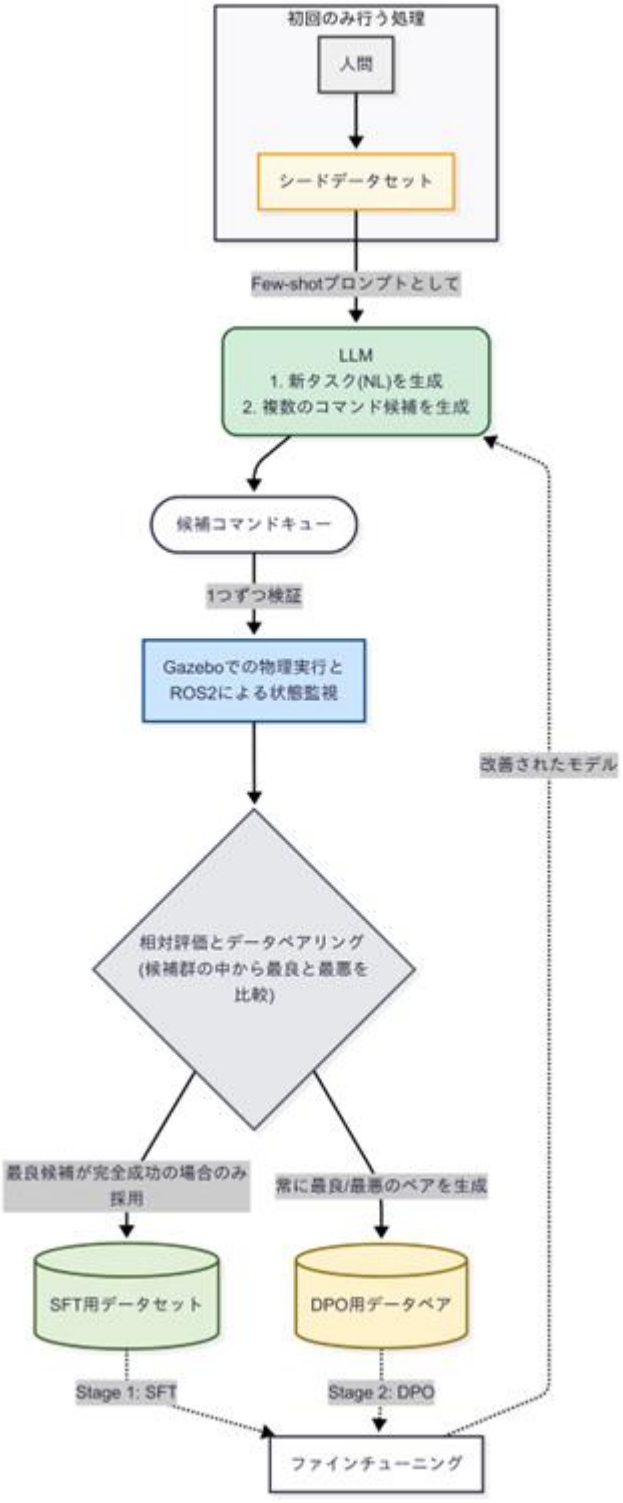

図 1. 提案処理フロー

4. 評価と実験

提案手法の有効性を実証するため、以下の評価実験を計画している。

実験目的:
提案手法で生成したデータセットが、未検証のデータセットに比べ、LLMのファインチューニング性能を向上させることを示す。

実験設定:
比較モデル:
(A) ベースLLM（ファインチューニングなし）
(B) LLMが生成しただけの未検証データでSFTを行ったモデル
(C) 本手法による検証済みデータでSFTおよびDPOを行ったモデル

評価指標:
20種類の未知の自然言語指示に対するタスク達成率をシミュレーション上で評価する。

期待される結果:
本手法で学習したモデル(C)が、他のモデルに比べてタスク達成率を大幅に向上させることが期待される。これは、物理的に検証されたデータが、モデルの信頼性を向上させる上で極めて有効であることを示唆する。

5. 結論

本研究の最も重要な貢献は、組込みシステム分野に対して、LLMの恩恵を現実的な形で提供する道筋を示した点にある。大規模な計算資源を必要とする汎用LLMそのものではなく、本フレームワークで生成した特化型データセットを用いてファインチューニングされた軽量モデルは、ロボットに搭載される典型的なエッジデバイス上での推論を可能にする。これは、AIの推論処理をクラウドからエッジへと移行させ、応答性や自律性の高いロボットシステムを実現するための基盤技術となる。

今後の課題は二点ある。

第一に、本研究の枠組みはシミュレーション上の成功を前提としており、実環境との差異、いわゆるSim-to-Realギャップが依然として存在する。このギャップを緩和するため、シミュレーション環境の物理パラメータや外観を変化させるDomain Randomization のような既存技術の導入や、少数の実機試行結果をデータセットにフィードバックする適応ループの構築が求められる。第二に、マニピュレーションなど、より状態遷移が複雑で多角的な成功基準を要するタスクへの本フレームワークの拡張が挙げられる。

6.参考文献

[1] A. Brohan, N. Brown, J. Carbune, et al., "Do As I Can, Not As I Say: Grounding Language in Robotic Affordances," *arXiv preprint arXiv:2204.01691*, 2022.

[2] J. Tobin, R. Fong, A. Ray, et al., "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 23-30.

[3] R. Rafailov, A. Sharma, C. Mitchell, et al., "Direct Preference Optimization: Your Language Model is Secretly a Reward Model," *arXiv preprint arXiv:2305.18290*, 2023.