Evaluating MILLION: Language-Based Meta-RL in Various Simulation Environments

Rex Imperial | Baldemar Guajardo
Department of Computer Science, College of Engineering and Computer Science

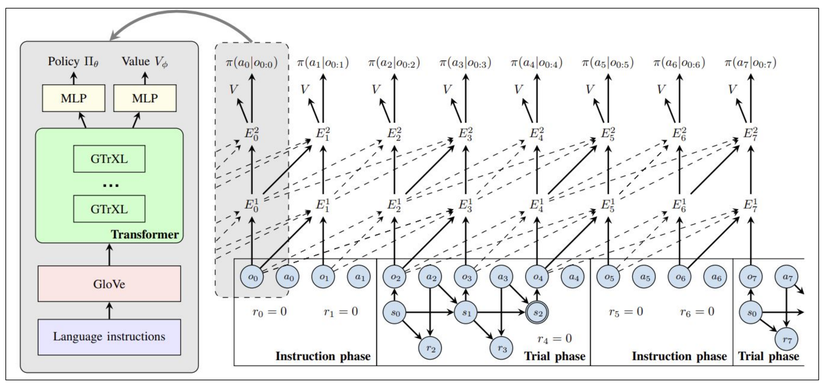
BACKGROUND

Meta-reinforcement learning (meta-RL) seeks to equip agents with the ability to quickly adapt to new tasks, mirroring human-like learning. Traditional meta-RL algorithms often struggle with diverse task sets, as they rely heavily on reward signals, which can be sparse or ambiguous. Bing et al. (2022) address this limitation by incorporating natural language instructions into the meta-RL framework. Their approach, Meta-Reinforcement Learning using Language Instructions (MILLION), leverages language to provide richer task information, enabling more efficient learning and improved generalization in robotic manipulation tasks. Our research explores the potential of extending this language conditioned meta-RL approach beyond its original application in robotics, testing its applicability and effectiveness in new environments.

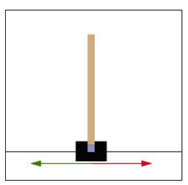
PURPOSE AND HYPOTHESIS

Purpose:
This research aims to evaluate the generalizability of language conditioned meta-RL, specifically the MILLION algorithm, by extending its application beyond robotic manipulation tasks to novel environments. The goal is to determine if the benefits of incorporating language instructions for efficient task learning and adaptation observed in robotic contexts can be replicated in different domains.

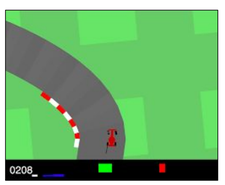
Hypothesis:
It is hypothesized that leveraging language instructions for task interpretation will enable more effective learning and adaptation in new RL environments. Specifically, the algorithm is expected to demonstrate improved sample efficiency and generalization performance compared to traditional RL methods that rely solely on reward signals, even in tasks with different underlying dynamics and state spaces.



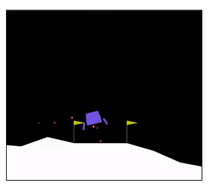
MILLION Algorithm Architecture



CartPole Environment



Car Racing Environment

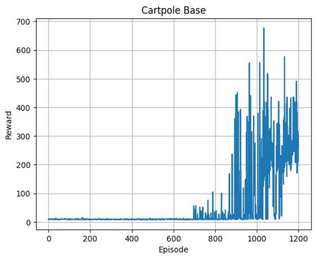


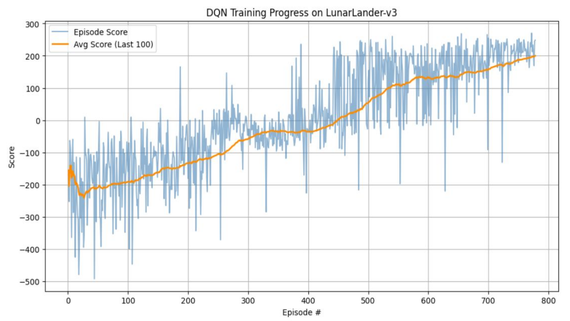
Lunar Landing Environment

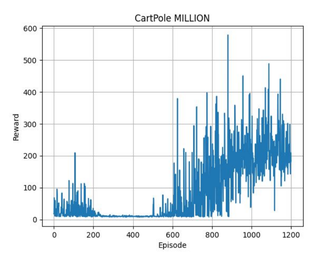
MATERIALS AND METHODS

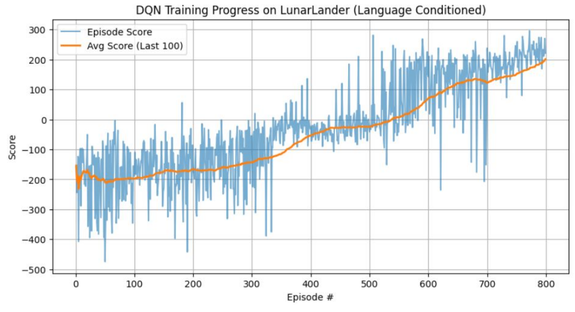
This project involved modifying base reinforcement learning (RL) models for several classic control environments with components of the MILLION algorithm, primarily the encoding of language instructions. The following steps were taken:

1. **Base RL Environments:** Three distinct RL environments were selected to provide a range of control challenges:
   - CartPole: A simple balancing task with discrete actions.
   - Car Racing: A continuous control task with a complex state space.
   - LunarLander: A more complex control task with discrete actions and a focus on efficient landing.
2. **Base RL Model Implementation:** For each environment, a standard deep reinforcement learning model was implemented. This served as the baseline for comparison.
3. **Language Instruction Encoding:** Inspired by the MILLION algorithm, a module was developed to encode language instructions relevant to each environment's task.
   - For CartPole, instructions like "Balance the pole", “Keep pole upright”, or "Avoid falling" were encoded.
   - For Car Racing, instructions such as "Drive fast," "Stay on track," or "Slow down before turns" were used.
   - For LunarLander, instructions like "Land safely," "Stay upright," or "Avoid crashing" were employed.
   - The specific encoding method involved techniques like word embeddings.
4. **Model Modification:** The base RL models were modified to incorporate the language instruction encoding. This involved feeding the encoded instructions into the model's policy network, allowing it to condition its actions on the given language.
5. **Training and Testing:**
   - Each model (base and modified) was trained on its respective environment.
   - Performance was evaluated by comparing the base models to the language-conditioned models. Key metrics included training speed and rewards/epoch.
6. **Analysis:** The results were analyzed to assess the impact of language instructions on the RL agent's learning and performance in each environment. This analysis focused on determining if the benefits observed in robotic manipulation, such as improved sample efficiency and generalization, transferred to these non-robotic control tasks.









RESULTS

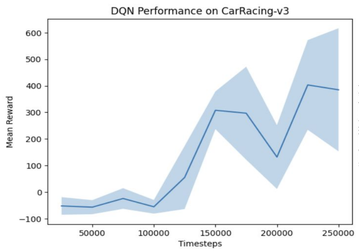
Language-conditioned (MILLION) vs. baseline RL models were assessed in CartPole, LunarLander, and CarRacing.

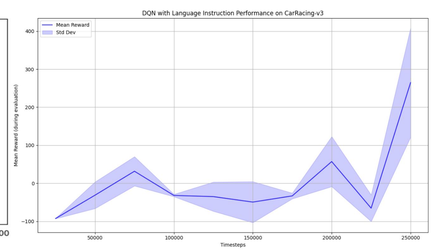
CartPole:
- Baseline: Gradual learning, stable after 800-900 episodes; high reward variance.
- MILLION: Faster path to higher average rewards, quicker stability, greater consistency.

LunarLander:
- Baseline DQN: Learned proficiently after ~500 episodes.
- Language DQN: Robust learning; potentially higher peak scores & smoother ascent (efficient refinement).

CarRacing:
- Baseline DQN: Learned, but with high fluctuations & less defined peak; inconsistent strategy.
- Language DQN: Learned, with a more pronounced peak reward (~250k timesteps); language guided to more consistent high-reward behavior.

Summary: Results suggests language instructions offer tangible benefits: accelerated learning, higher/more stable rewards, and sometimes superior peak performance.





CONCLUSIONS

This research assessed if language-conditioned meta-RL (MILLION) generalizes to diverse non-robotic environments. Our hypothesis was that language would enhance learning.

Results across CartPole, LunarLander, and CarRacing indicate language conditioning positively impacts learning. Language-conditioned models often achieved higher rewards, steadier learning, or reached proficiency more efficiently. This supports the idea that language provides richer task information, improving sample efficiency and generalization beyond robotics.

Further quantitative analysis would strengthen these conclusions. This study shows the promise of extending language-conditioned meta-RL to broader applications.

BIBLIOGRAPHY

Bing, Z., Koch, A., Yao, X., Huang, K., & Knoll, A. (2022). Meta-Reinforcement Learning via Language Instructions. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 5985-5991. (Semantic Scholar Link, arXiv Link)