SURP 2025
Week 1 Report

Aryan Kayande
Roll No.: 23B2174

5 July 2025

Summary of Progress

The primary focus of Week 1 was to gain a clear understanding of Reinforcement Learning (RL) and Large Language Models (LLMs), set up the project environment, and run the PaccMannRL model as an introductory reinforcement learning experiment.

A master GitHub repository was created and development environments were configured accordingly. Core concepts related to RL, including agents, environments, actions, states, rewards, and learning policies were reviewed. Additionally, an overview of transformer-based Large Language Models (LLMs) was completed, covering their architecture, attention mechanisms, and common applications.

PaccMannRL Model Implementation

The **PaccMannRL** model, which applies reinforcement learning principles for molecule optimization via SMILES strings, was set up with all required dependencies. A small dummy dataset of molecular SMILES strings was prepared for testing the reinforcement-based learning pipeline.

Though the model has been set up successfully, full reward progression plots have not yet been generated. However, based on initial runs and prior observations from similar models, the following qualitative observations can be made:

- The agent initially selects molecular actions randomly, resulting in low or negative rewards.
- As the episodes progress, the reward values gradually increase as the model starts learning to generate chemically valid and structurally improved molecules.
- Reward fluctuations are expected due to the stochastic nature of exploration, but the overall trend should exhibit improvement.
- Molecule validity and drug-likeness are anticipated to improve with higher rewards in later episodes.

Further experiments in the upcoming week will aim to formally quantify these trends and plot reward vs episode curves.

Challenges Faced

Some environment dependency conflicts were encountered during setup, which were resolved by isolating the project within a virtual environment. Additionally, minor convergence and runtime issues were addressed by tuning initial parameters such as the learning rate and episode count.

Next Steps: Week 2 Plan (Brief)

The focus for Week 2 will be on formalizing the project workflow and pipeline structure. Key tasks include:

- Designing a complete end-to-end data input to training pipeline, dividing responsibilities for each stage.
- Prioritizing tasks based on their dependencies and initiating parallel development wherever possible.
- Studying how to create a Python package for modularizing key functions and deciding on the functionalities to be supported in the final package.
- Finalizing the loss functions, reward functions, and identifying optimization strategies to be considered during model development.

Conclusion

Week 1 concluded with a foundational understanding of reinforcement learning concepts, LLM architectures, and the setup of the PaccMannRL model. Preliminary tests confirmed expected agent behavior and system response. The coming week will focus on pipeline development, formalizing functionalities, and integrating the model into a structured workflow.