Reinforcement Learning CIA Assignment

Use Case: Movie Recommendation System

Introduction:

In a movie recommendation system, the model's objective is to iteratively better itself on the choices/recommendations it provides to like-minded users regarding the kind of movies they would prefer to watch. Usually, we use machine learning for building recommendation systems such as these but in this approach, Reinforcement Learning presents a different perspective to what kind of data is fed to the model and how it is trained in an unconventional manner.

Problem Statement:

To expose the model to user choices of movie selections and accordingly make it understand the kind of movies that similar people opt for making better recommendations

Proposed Approach:

1. There will be a platform where users can sign up and choose their selection of movies from the portal. This data will be fed to the model in the backend for training purposes. Initially after the model has been exposed to substantial amount of data, it starts recommending movies to the other users based on parameters like common genre, actors in the movie, rating, etc.

2. There will be a system to check if the user selects a movie based on the recommendation provided by the model, then a positive reward is assigned else a penalty is given. This will be done for a set of recommendation. Suppose the model suggests 5-6 movies, we will have to keep count of how many the user preferred to actually watch.

3. Coming to k-arm bandits, each of the arm can be kept to assess each of the specific parameters for determining potential matches in movie choices. Like if the bandit has 3 arms, then the first arm could keep track of the genre, the second arm for rating and the third arm for common actors in the films.

4. The first arm in charge of the genre could be given more importance as genre plays a more important role and the collective weightage of all the 3 arms could be computed iteratively to come to a set of recommendations which can be presented to the user.

5. This bandit can also be potentially extended to involve more arms and more complex respective functions for the actions to be taken and accordingly the reward-penalty policy can also be modified.