

Name: Bhavesh Kumar B

Reg No:21011101031

Recommendation System using K-Arm Bandit algorithm

Introduction to problem Statement:

I have selected to work with music recommendation system since it offers a robust approach in addressing this problem using K- Arm bandit algorithm and provides a well-balanced trade-off between recommendations (exploration) and discovering new favourites (exploitations).

Implementation:

Using K-Arm bandit algorithm is to maximize the total reward by balancing exploitation and exploration, where:

Arm=song or a playlist

Pulling an arm=recommending a song and observing their reaction.

Reward:

play=+1, skip=-1, like=+2, dislike=-2.

- At the outset, the system treats each song as an arm with an initial reward estimate, often set to zero or a small random value. The system also tracks how often each song is recommended.
- Song Selection:
 - Greedy Strategy: The system randomly selects a song with a probability ϵ (exploration) and selects the song with the highest estimated reward with a probability of $1 - \epsilon$ (exploitation).
- Reward Update: After user reacts to the song, the system updates its reward estimate using the formula: $Q_t(a) = (\text{sum of rewards when a taken prior to } t) / (\text{number of times a taken prior to } t)$.
- The recommendation process is repeated, continually refining the system's understanding of which songs users prefer, thereby improving the quality of recommendations over time.
- In the initial phase model, system focusses mainly on exploration.
- As the system learns about their interactions, it updates it rewards for each song.
- In the end after many iterations, the system gradually shifts towards recommending songs with higher rewards and still exploring to ensure that everything is captured.

Example Scenario:

Consider a music recommendation system with 10 songs. Initially, each song has an equal chance of being recommended. As users interact, the system identifies that Songs A and B are frequently liked (+2 reward), while Song C is often skipped (-1 reward). The system begins to favour Songs A and B but occasionally explores other songs to avoid missing better alternatives.