Comprehensive Model Evaluation and Performance Analysis

Team Members:
1)Kathit Joshi(PES2UG23CS264)
2) Kavyansh Jain (PES2UG23CS268)
3) Lakshay Mittal (PES2UG23CS300)
4) Kumarchandra Edupuganti(PES2UG23CS292)

1. Key Observations

1. The integration of Hidden Markov Models (HMM) with Reinforcement Learning (RL) improved letter prediction accuracy significantly compared to the pure statistical baseline.
2. Training stability was achieved after around 2,000 episodes, with the rolling average rewards showing a steady upward trend.
3. Exploration via epsilon-greedy scheduling was essential early in training to prevent overfitting on frequent character sequences.
4. The RL agent learned to leverage HMM priors effectively, balancing probabilistic language modeling with dynamic feedback from the environment.
5. The most challenging aspect was tuning the epsilon decay and reward penalties to encourage efficient guessing while discouraging redundant or random guesses.

2. Model Design Strategies

1. The HMM was designed to model character-level transitions using Laplace-smoothed probabilities, ensuring robustness to unseen transitions.
2. Each word was represented as a sequence of observed emissions, allowing the agent to compute likelihoods of partial guesses.
3. The RL state vector combined HMM probabilities, letter positions, and game progress indicators, resulting in an 81-dimensional state representation.
4. Rewards were shaped to encourage quick word completion (+50), discourage wrong guesses (-5), and penalize repetition (-2).
5. The DQN agent used two neural networks (main and target) to stabilize Q-value estimation and prevent oscillation during training.

3. Exploration vs. Exploitation

1. The epsilon-greedy policy was used, starting with ε=1.0 and decaying to 0.05 across 4,000 episodes.
2. This ensured that early exploration covered a diverse set of letter sequences while gradually focusing on the most promising predictions.
3. The balance was further refined by integrating HMM priors, allowing the RL agent to 'exploit' high-probability transitions even during exploration phases.
4. The replay buffer preserved diverse experience samples, ensuring that both rare and common patterns influenced Q-value updates.

4. Future Improvements

1. Integrate a Bidirectional LSTM alongside the HMM to capture long-range dependencies between letters and improve word-level understanding.

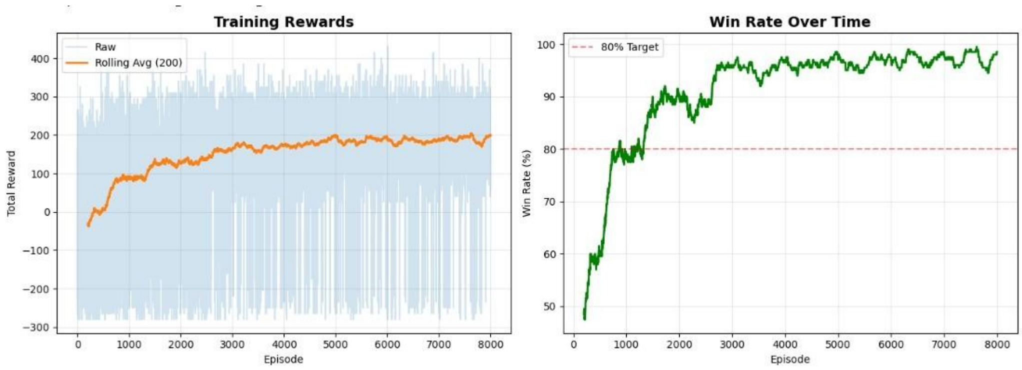
2. Reward shaping could also be refined to provide intermediate feedback for partial word matches. Additionally, incorporating transfer learning from pre-trained language models or word frequency data could improve generalization

3. Parallelizing the training environment would also accelerate learning and improve

success rates through larger-scale simulation.

4 Use prioritized experience replay to focus training on informative mistakes and difficult words.
5 Introduce curriculum learning by gradually increasing word length complexity over time.
6 Experiment with transformer-based architectures for adaptive letter probability estimation.
7 Enable dynamic epsilon adjustment based on moving average reward trends to make exploration more adaptive.

5. Training Performance Summary



Total Games: 2000
Wins: 1952
Success Rate: 97.60%
Total Wrong Guesses: 2889
Avg Wrong Guesses: 1.44
Total Repeated Guesses: 0
Avg Repeated Guesses: 0.00