Pharmahacks 2025
Error 404 Team Name Not Found

Multiple Sequence Alignment

Genevieve Laprade

Sarah Ford

Rewant Chauhan

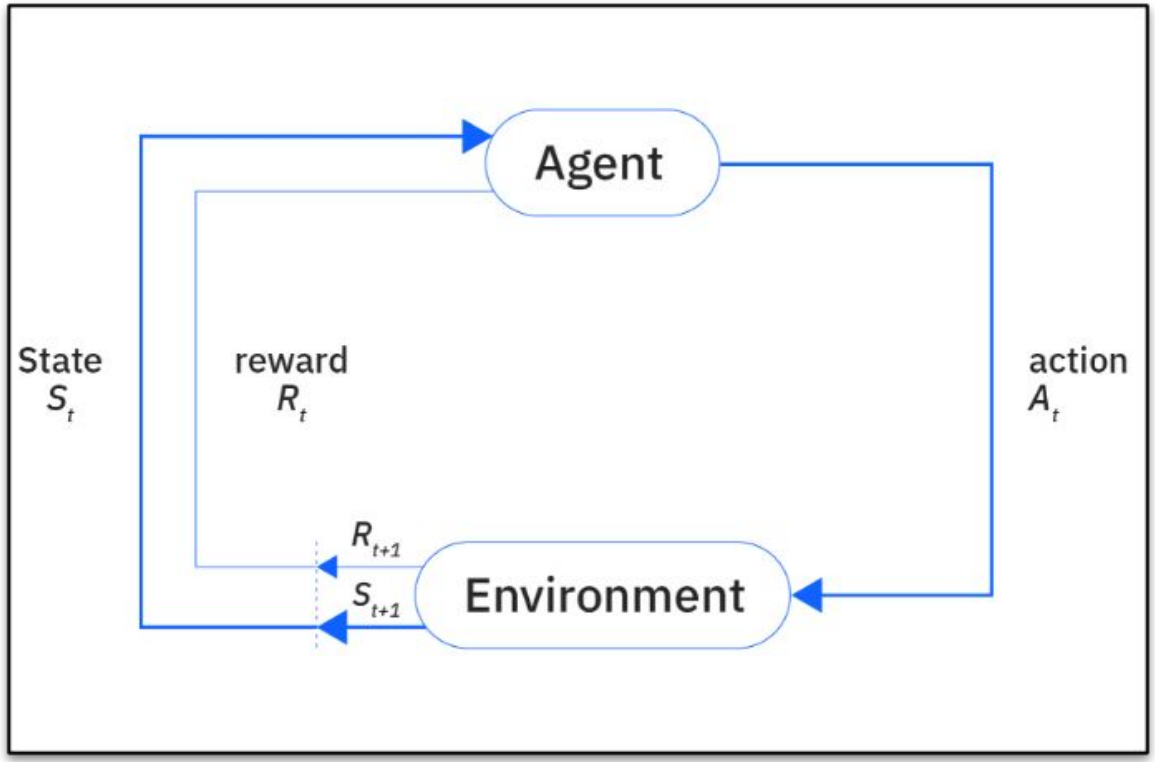
Brandon Perestrelo

01 Explorations

Reinforcement Learning

Agent: Model
Action: Step
Environment: Sequence Alignment
State: Score calculation for current alignment
Reward:
<u>Positive</u>
Score(New) > Score(Old)
- Step is made

<u>Negative</u>
Score(New) < Score(Old)
- Step is not made
- Model is penalized

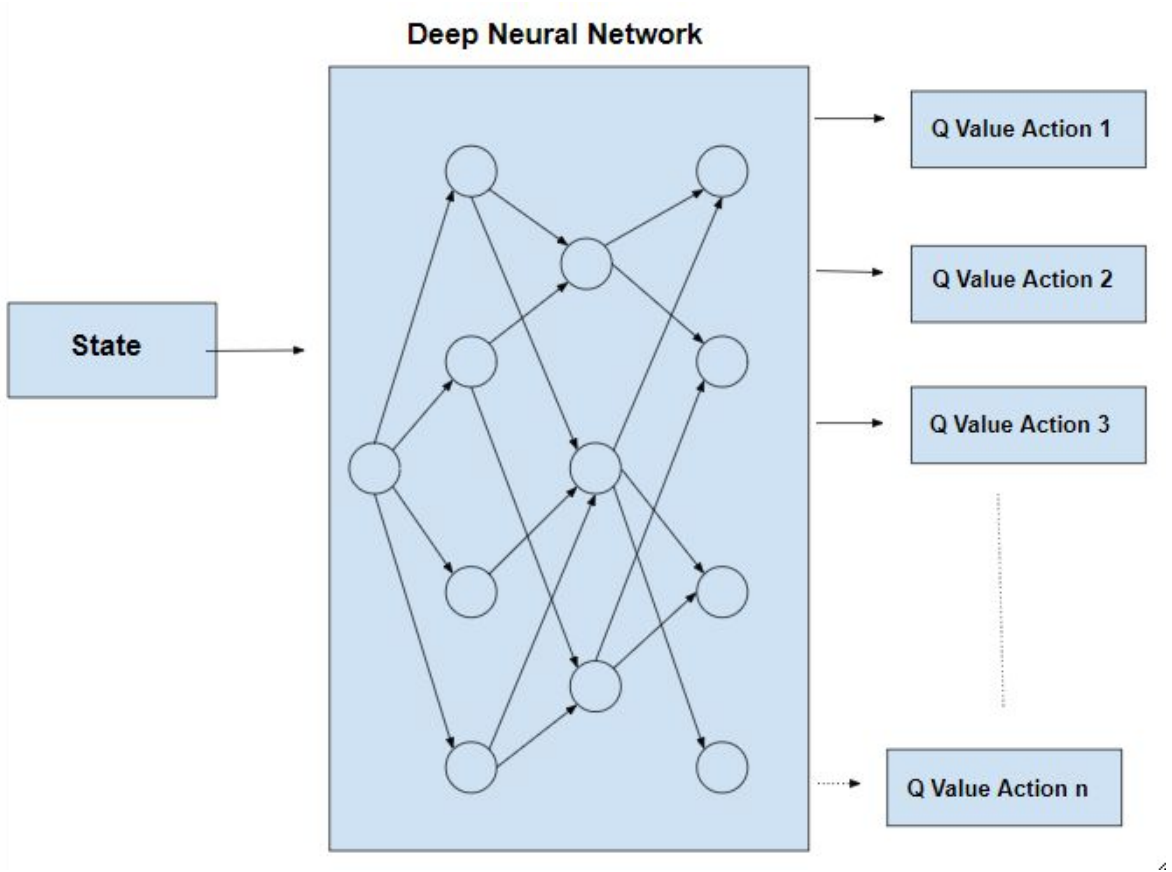


Challenge: Exponential number of possible steps

Model: Deep-Q Learning

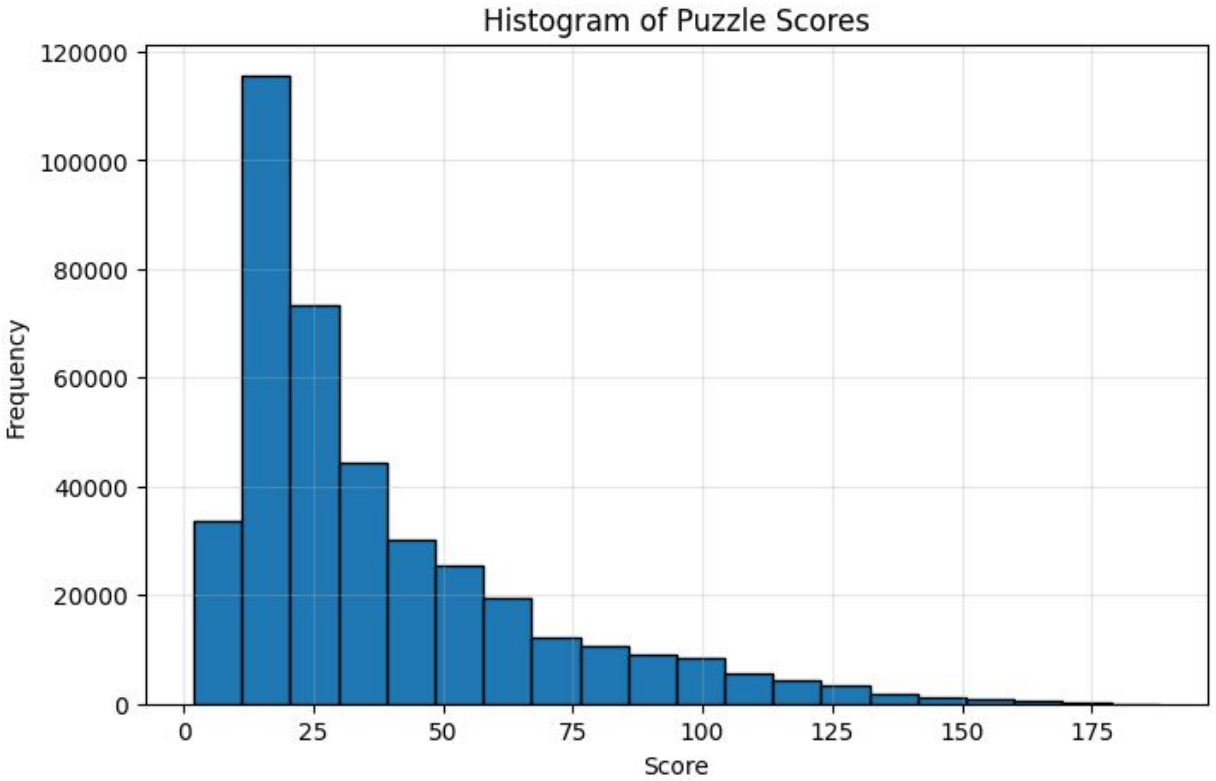
Mathematical approximation of all the possible steps given a state

Neural Network chooses the best action (step)



Source: https://www.geeksforgeeks.org/deep-q-learning/

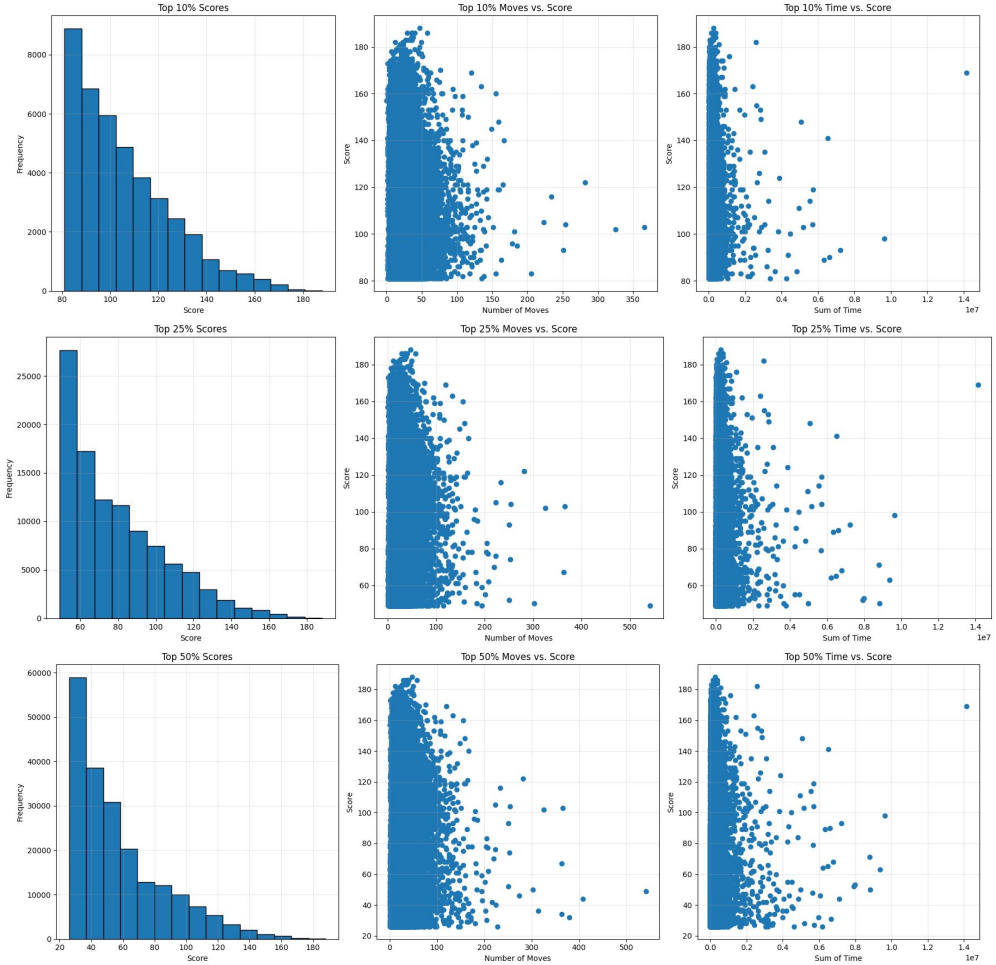
Visualize Data Spread



Trend's from Top Scorers

No strong trends between number of moves or gameplay time and high scores

:(



Model Theory

Accepted Pairs

-	C
T	G
G	T
A	C
A	C
A	C
-	T
T	G

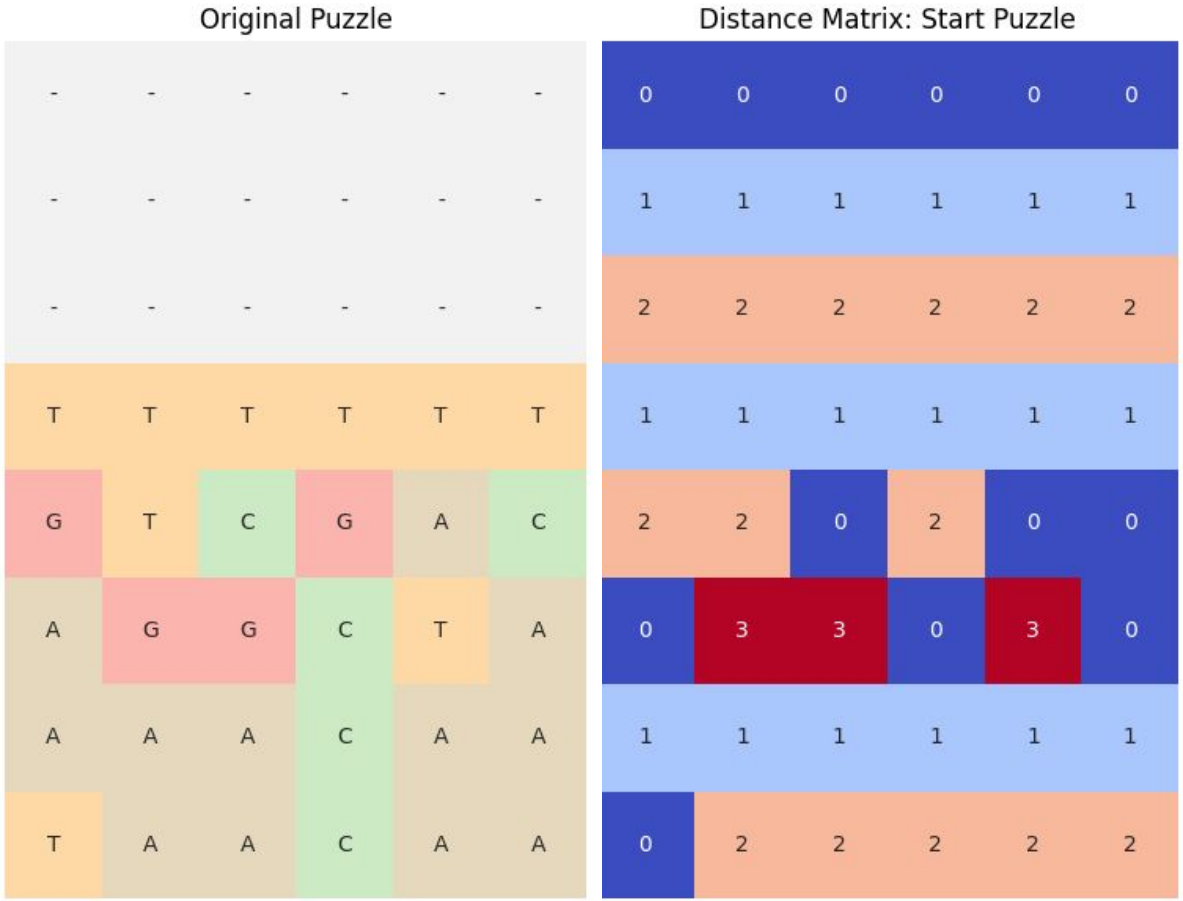
Original Puzzle

-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
T	T	T	T	T	T
G	T	C	G	A	C
A	G	G	C	T	A
A	A	A	C	A	A
T	A	A	C	A	A

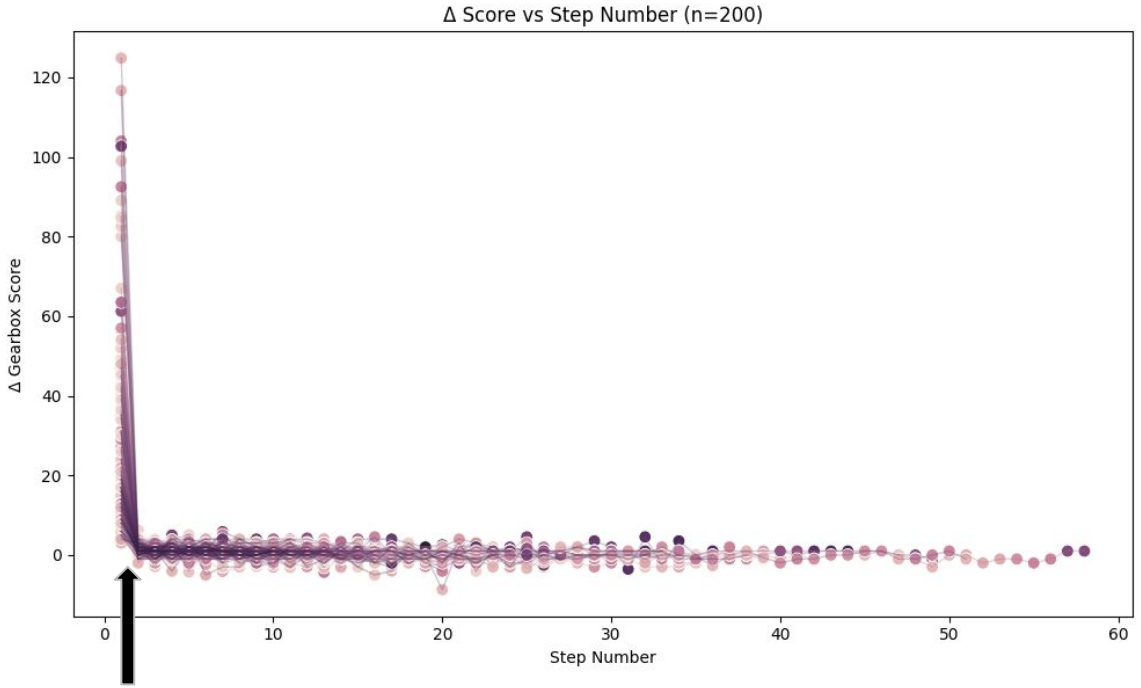
Distance Algorithm

Calculate how far each base pair is from alignment

Spatial awareness task component enhances model accuracy and efficiency

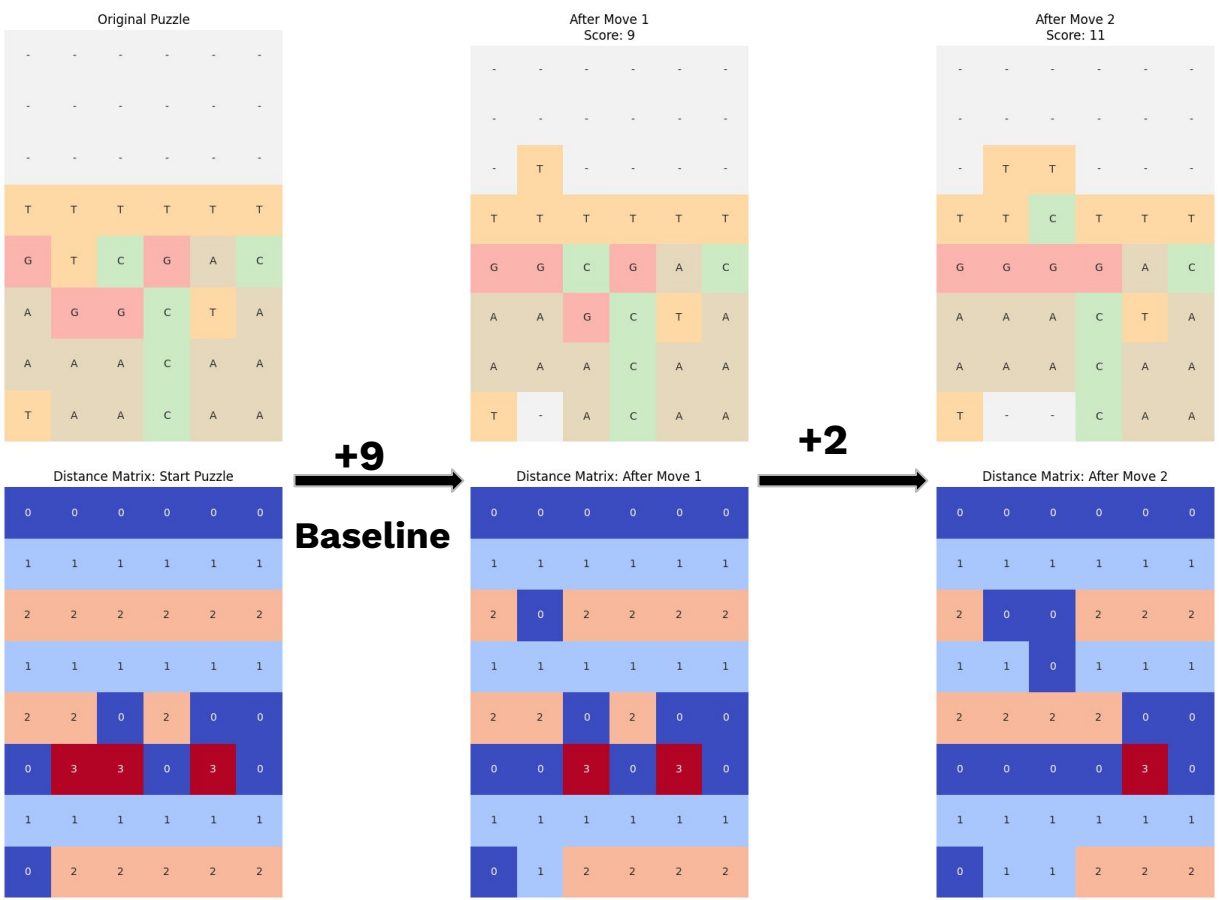


Importance of first steps



Baseline gearbox

Algorithm input: Correlate changes in distance matrix to gearbox score



Model + Task

Adding the distance algorithm task as a parameter in the neural network helps refine learning to make it more MSA-specific

Features:
- Steps
- Accepted Pairs
- Start
- Distance Matrix

Target Variable:
- Score

Goal:
- Model solution score should beat test solution score

Moves variable omitted for simplicity

Model

Adding the distance algorithm task as a parameter in the neural network helps refine learning to make it more MSA-specific

Features:
- Steps
- Accepted Pairs
- Start
- Distance Matrix

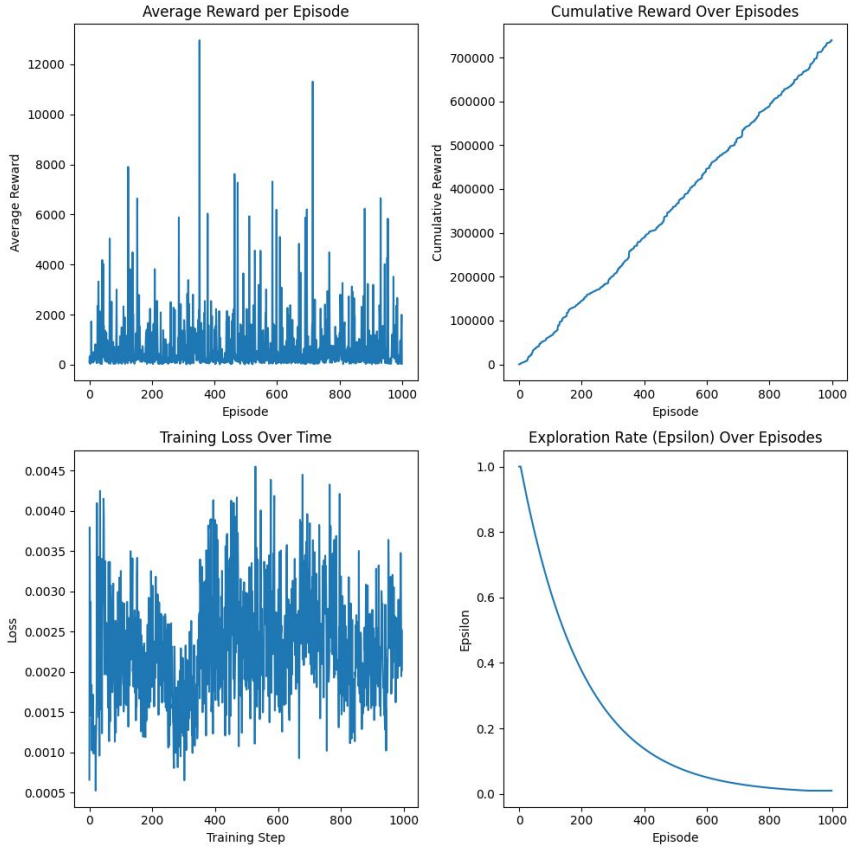
Target Variable:
- Score

Goal:
- Model solution score should beat test solution score

Moves variable omitted for simplicity

02
Results

Evaluation Metrics



Evaluation Metrics

Model Metrics currently in queue as of submission deadline

03
Conclusions

Biological Significance

- This algorithm is a starting point for small sequences
- Much more complicated to apply to large sequences
- Could be refined to be more biology-specific with the integration of the spatial task algorithm

Future Directions

- Increasing models layers and episodes
- Evaluate carbon footprint
- Use Leave-One-Out cross validation to refine algorithm
- Expand on distance matrix calculations to refine efficiency