**Minecraft Project Proposal**

Searching by MCTS & Q-learning and Neural Networks

**Shaohong Tian** shaohongt@student.unimelb.edu.au

**Jiahao Yu** jiahaoy@student.unimelb.edu.au

**Guo Zhou** zhoug2@student.unimelb.edu.au

**Ziwei Li** [ziweil6@student.unimelb.edu.au](mailto:ziweil6@student.unimelb.edu.au)

**Experimental Design**

**TITLE**

The Effect of the performance of an agent used Q-learning in a maze on scale Q-learning to choose an action to perform

**HYPOTHESIS**

If the Q-learning of the agent can choose an action more like human choose a direction in a maze, then the performance of agent will be better than others.

**INDEPENDENT VARIABLE**

The performance of an agent used Q-learning in maze

**LEVELS OF INDEPENDENT VARIABLE AND NUMBERS OF REPEATED TRIALS**

|  |  |  |  |
| --- | --- | --- | --- |
| Two directions(forward and turn left)(Control) | Three directions(forward, turn left and jump) | Four directions(front and turn back, turn left and turn right) | Five directions(front and turn back, turn left and turn right, jump) |
| 3 times | 3 times | 3 times | 3 times |

**DEPENDENT VARIABLE AND HOW MEASURED**

Change the domain of actions, measured the time of the agent achieves the goal. And we will use three different maps to test them 3 times, so the result will be more general.

**CONSTANTS**

1. Using the same map at each time.

2. Using the same CPU processing power.

3. Using the same learned value and learning rate.

**TITLE**

The Effect of the performance of an agent used different number of Q-learning in maze

**HYPOTHESIS**

If the agent use one Q-learning to choose an action and use another to evaluate whether current selection is best in a maze, then the performance of agent will be better than only use one Q-learning to choose an action.

**INDEPENDENT VARIABLE**

The performance of the agent in maze

**LEVELS OF INDEPENDENT VARIABLE AND NUMBERS OF REPEATED TRIALS**

|  |  |
| --- | --- |
| One Q-learning | Two Q-learning |
| 3 times | 3 times |

**DEPENDENT VARIABLE AND HOW MEASURED**

Comparing to one Q-learning, we will initialize two deep Q-learning, Q1 and Q2. Using both Q1 and Q2 in estimating the supervision signal- Q1 used for select action and Q2 used to evaluating current selection, measured the time of the agent achieves the goal. And we will use three different maps to test them 3 times, so the result will be more general.

**CONSTANTS**

1. Using the same map at each time.

2. Using the same CPU processing power.

3. Using the same learned value and learning rate.

4. Using the same domain of actions.