ResearchTrack2

Assignment 3: statistical analysis of a robot simulation

Cappellini Matteo (4822622)

1 Introduction

The task I was given was to perform a statistical analysis on the first assignment from the *ResearchTrack1* course, considering two different implementations: mine, and a solution from one of my colleagues. The purpose of the simulation was testing which one performs better, when the number of silver and golden tokens in the environment is changed.

2 Experiment description

The objective of this experiment was to observe the behavior of two different algorithms when the number of tokens in the environment was changed and determine which performed better. In particular, the simulations were conducted with a number of tokens ranging from 3 to 9, for both silver and golden tokens.

To be able to extract statistically significant data, each simulation was repeated 5 times for every token configuration, resulting in a total of 35 simulations for both algorithms. Both scripts were tested in the same arena under identical conditions to ensure reliable results.

To understand which one had better performances I measured the time needed for each simulation to be completed by utilizing the *time* library and keeping track of each result.

3 Hypothesis

To begin with the analysis of this experiment it is necessary to formulate the ***null hypothesis***, which is the initial assumption that I want to test and the ***alternative hypothesis***, which is the claim that I will accept if the null hypothesis turns out to be wrong. Here are stated the two assumptions:

- ***Null hypothesis:*** the two algorithms produce very similar results and it is not possible to define which one has better performances overall.
- ***Alternative hypothesis***: the two algorithms are indeed significantly different in terms of performance and we can identify which one is better.

To test these hypothesis I will compare the two data sets with the ***two sample T-Test*** with a significance level of 5%.

4 Results

In the following table I will show all the results that I got from all the simulations; the column entitled *Simulation A* contains the values obtained from my algorithm while the ones from *Simulation B* are the results of the algorithm developed by my colleague Manuel Delucchi.

Test	Tokens	Simulation A	Simulation B
1	3	59.56	71.43
2	3	59.82	72.41
3	3	60.06	72.43
4	3	60.36	72.92
5	3	60.87	73.42
6	4	88.72	84.58
7	4	88.74	85.51
8	4	90.44	87.21
9	4	92.36	87.55
10	4	93.25	89.52
11	5	80.42	106.02
12	5	80.55	106.92
13	5	83.21	107.02
14	5	88.87	107.92
15	5	89.49	108.35
16	6	116.81	88.05
17	6	122.12	89.43
18	6	122.73	90.67
19	6	124.65	92.96
20	6	126.03	98.54
21	7	113.69	113.57
22	7	116.94	114.16
23	7	120.64	116.18
24	7	126.05	118.58
25	7	128.35	124.33
26	8	128.66	115.01
27	8	130.26	119.23
28	8	137.96	122.51
29	8	138.93	128.28
30	8	144.08	128.66
31	9	174.75	264.66
32	9	272.49	269.21
33	9	317.23	305.2
34	9	338.41	334.01
35	9	436.93	338.06

Table 1: Simulations results

From *Figure 1* we can observe the comparison between the average completion time of the two algorithms.

While it is difficult to say which algorithm is more efficient by looking at this bar graph, it still provides us with valuable insights into their behavior. As we can observe, my code, was significantly faster when dealing with 6 tokens, but slower inside scenarios involving a smaller amount of tokens. However, interestingly, when analyzing an environment with a high number of tokens, the behavior of both algorithms appears to be quite similar.

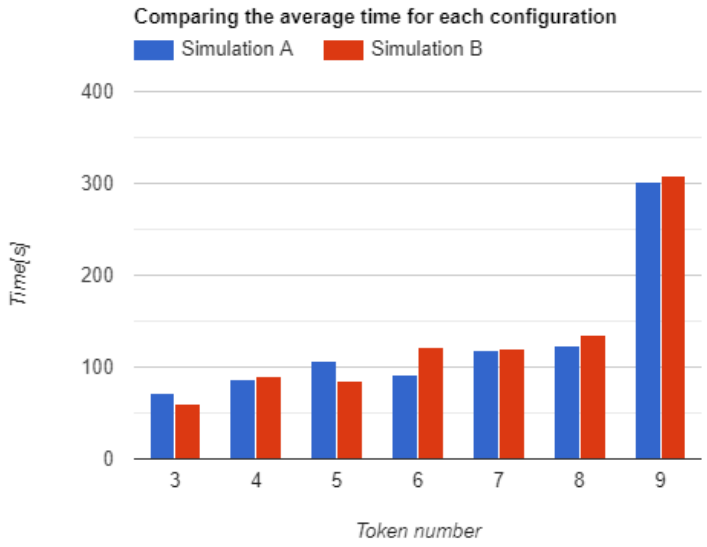

Figure 1: Average completion time comparison

To get results more significant than these, we can finally conduct a ***two-sample T-Test***, thanks to the Matlab function *ttest2*. This function takes two data sets as input, along with a specified significance level (in this case, 0.05), and returns a boolean value to signal whether to reject the null hypothesis (1) or not (0) and the *p-value*, which represents the probability of observing the data, given that the null hypothesis is true.

I was able to obtain some concluding results, such as:

- **H = 0**
- **p-value = 0.722**

5 Conclusions

After running a T-Test for each configuration, the obtained results consistently showed a value of H equal to 0 and the p-value significantly larger than 0.05.

This means that the null hypothesis cannot be rejected, therefore, based on the collected data, it can be concluded that the two algorithms exhibit similar behavior, and no significant difference can be established to determine which one is superior.