UNIVERSITÀDEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
BIOENGINEERING, ROBOTICS AND SYSTEM ENGINEERING

Research Track 2

Third Assignment

Professor:
Carmine Recchiuto

Authors: Amirmohammad Saberi, Seyed Alireza Mortazavi, Reza Davoudi Beni
Student ID's: S5840276 , S6136275 , S6115320

ABSTRACT

This report compares the performance of two algorithms designed for a Research Track 1 course assignment by Seyed Alireza Mortazavi and Amirmohammad Saberi. The algorithms enable a robot to pair golden tokens located in an arena. The primary objective of the study is to evaluate the effectiveness of the two scripts under different conditions, such as varying the quantity of tokens and their generation radius. By analyzing the algorithms' adaptability to different spatial configurations, we aim to assess their overall efficacy in achieving the intended task.

HYPOTHESES

Hypothesis Testing is a statistical tool that is usually used in scientific research to reach a conclusion about the parameters of a data set (mean, variance, etc.).

The hypothesis for a given problem is described by two competing hypotheses: **Null hypothesis (H0)** which states that the two methods that are being compared are equally good (based on the evaluation methods applied later).

On the other side, **Alternative hypothesis (Ha)** is defined which states that one method is superior in comparison to the other method. If one of these hypotheses is true, the other one is false, and vice versa. One method to prove that the alternative hypothesis is true, and one method is better than the other method is to reject the null hypothesis. Consequently, as a first step in analyzing the two codes in this report, the null and alternative hypotheses should be defined as follows:

Null Hypothesis (H0): The results obtained from Amirmohammad code and Seyed Alireza code have the same performance and they are like each other. In other words, we cannot determine which algorithm is faster than the other one in gathering the boxes and the average time for this task is equal in both algorithms ($\mu A = \mu B$).

Alternative hypothesis (Ha): Seyed Alireza's algorithm is faster than Amirmohammad algorithm in gathering the boxes next to each other under different circumstances. Consequently, the average time it takes for my robot to gather all the boxes is more in comparison with Seyed Alireza's code ($\mu A > \mu B$).

In the first step to test the null hypothesis, sample test data must be gathered. The number of test data should be enough so that it gives proper, accurate, and trustworthy results. There are different methods to check and compare the performance of two algorithms, such as Z-test, T-test, two sampled T-test, paired T-test, etc. There are some basic preconditions for each method so that we can use that specific method for analysis. Z-test is usually used when the population and its mean and standard deviation are available. To use the T-test, at least one of the characteristics of the population is known (mean, standard deviation). In the case of this report, the population is not available, and we are just comparing two different algorithms that complete a task. For this analysis, two sampled T-tests and paired T-test can be useful.

Experiment and Results

To obtain the test data for the analysis, the same environments are created for both algorithms. Each algorithm is tested in different environments with 6 and 8 golden boxes to check how fast it gathers all the boxes and how accurate each algorithm is in detecting the boxes and gathering them. The algorithms are executed 5 times in each environment configuration which makes an overall 40 simulations for each algorithm. To create the exact same environment for each algorithm and validate the comparison, we did not use random generation of the boxes. Consequently, the positions of the boxes are the same in an environment with the same number of boxes. We have decided to apply a paired T-test on the extracted data.

We could have also used a two sample T-test if we simulated both algorithms 40 times with the same number of randomly generated golden boxes in the environment.

This method should be implemented as follows:

1. Calculate the difference ($d_i = y_i - x_i$) between the two observations on each simulation pair of two algorithms, making sure to distinguish between positive and negative differences.
2. Calculate the mean difference ($\bar{d}$).
3. Calculate the standard deviation of the differences, s_d, and use this to calculate the standard error of the mean difference ($SE(\bar{d}) = \sqrt{\ } s_d n$).
4. Calculate the t-statistic, which is given by $T = \bar{d} / SE(\bar{d})$. Under the null hypothesis, this statistic follows a t distribution with n - 1 degree of freedom.
5. Use tables of the t-distribution to compare your value for T to the t_{n-1} distribution. This will give the p-value for the paired t-test.

In Table 1, the results of Amirmohammad code and Seyed Alireza's code simulation success are presented under different circumstances. Both codes were completely successful in detecting and gathering the boxes in all cases that were investigated.

DATA COLLECTION

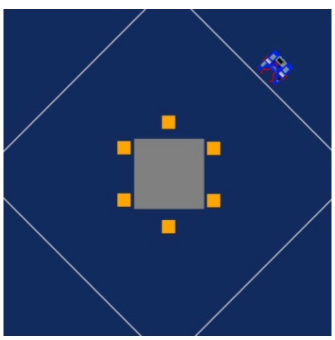


Figure 1: set radius to 0.5

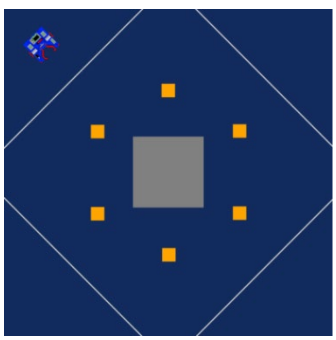


Figure 2: set radius to 1.5

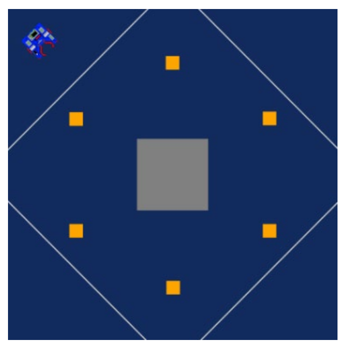


Figure 1: set radius to 2

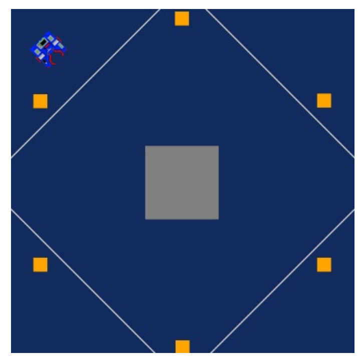


Figure 1: set radius to 3

We conducted experiments with different spatial configurations by adjusting the token generation radius and quantity. For each configuration, we ran both algorithms and collected paired measurements of their performance. Each algorithm was executed five times per configuration, resulting in a total of 40 different states. This sample size allowed us to apply the central limit theorem, which states that the sampling distribution of the mean will approximate a normal distribution given a sufficiently large sample size. However, since the true population variance is unknown, the sampling distribution of the mean follows a t-distribution, which accounts for the uncertainty in the population variance estimate.

Experimental Configurations

Radius = 0.5: 5 and 8 tokens

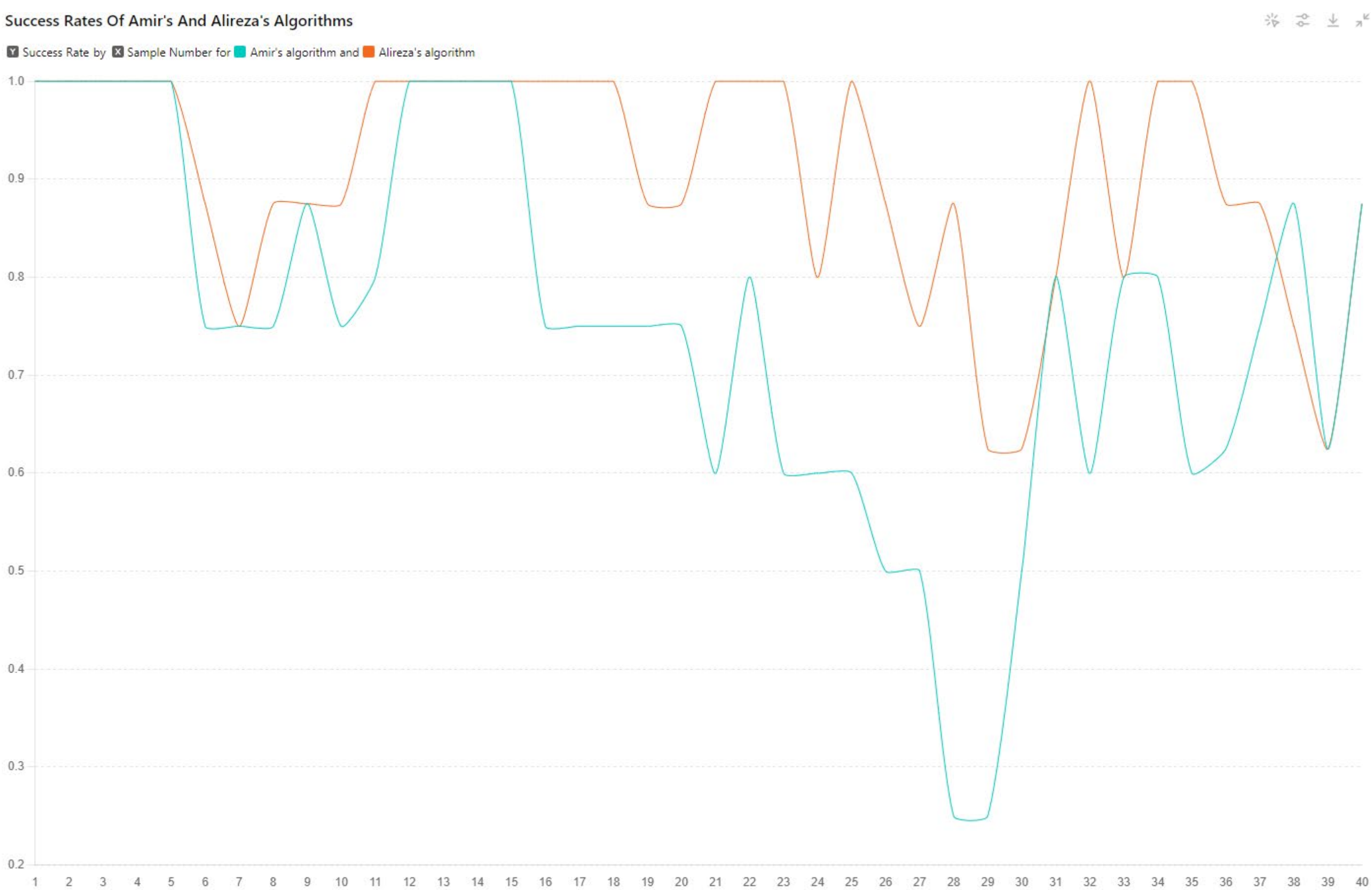
Radius = 1.5: 5 and 8 tokens

Radius = 2.0: 5 and 8 tokens

Radius = 3.0: 5 and 8 tokens

			Success rate		
Sample number	Radius	Number of Tokens	Amir's algorithm	Alireza's algorithm	Difference
1	0.5	6	1	1	0
2	0.5	6	1	1	0
3	0.5	6	1	1	0
4	0.5	6	1	1	0
5	0.5	6	1	1	0
6	0.5	8	0.75	0.875	0.125
7	0.5	8	0.75	0.75	0
8	0.5	8	0.75	0.875	0.125
9	0.5	8	0.875	0.875	0
10	0.5	8	0.75	0.875	0.125
11	1.5	6	0.8	1	0.2
12	1.5	6	1	1	0
13	1.5	6	1	1	0
14	1.5	6	1	1	0
15	1.5	6	1	1	0
16	1.5	8	0.75	1	0.25
17	1.5	8	0.75	1	0.25
18	1.5	8	0.75	1	0.25
19	1.5	8	0.75	0.875	0.125
20	1.5	8	0.75	0.875	0.125
21	2	6	0.6	1	0.4
22	2	6	0.8	1	0.2
23	2	6	0.6	1	0.4
24	2	6	0.6	0.8	0.2
25	2	6	0.6	1	0.4
26	2	8	0.5	0.875	0.375
27	2	8	0.5	0.75	0.25
28	2	8	0.25	0.875	0.625
29	2	8	0.25	0.625	0.375
30	2	8	0.5	0.625	0.125
31	3	6	0.8	0.8	0
32	3	6	0.6	1	0.4

33	3	6	0.8	0.8	0
34	3	6	0.8	1	0.2
35	3	6	0.6	1	0.4
36	3	8	0.625	0.875	0.25
37	3	8	0.75	0.875	0.125
38	3	8	0.875	0.75	-0.125
39	3	8	0.625	0.625	0
40	3	8	0.875	0.875	0



Results of the Analysis

1. Mean of the difference: $\bar{d}$ = 0.154375

The mean of the differences is calculated by taking the average of all the differences between the paired observations of the two algorithms.

2. Standard deviation of the difference: S_p = 0.169169114

The standard deviation of the differences measures the variability or spread of the differences around the mean difference.

3. Standard error of the difference: $SE(\bar{d})$ = 0.026747985

The standard error of the mean difference provides an estimate of the variability of the mean difference from the true population mean difference.

4. T-value: T = 5.77146268

The T-value is a ratio that compares the mean difference to the standard error of the mean difference. It indicates how many standard errors the mean difference is away from zero.

Conclusion

The paired T-test results showed that the calculated t-value (5.77146268) exceeded the critical t-value (Comparing the calculated t-value to the critical t-value at a chosen significance level (5%)) at a significance level of α=0.05 with 39 degrees of freedom (df = N - 1, where N represents the number of paired observations). Consequently, the null hypothesis was rejected, indicating a significant difference in the performance of the two programs when the token count and radius settings are varied.