

# Empirical Analysis of Bubble Sort on Two Personal Computers using Performance Visualization and Mann Whitney U Test

Dipankar Das<sup>1</sup>, Neelanjan Manna<sup>2</sup>, Spandan Poddar<sup>3</sup>, Raunak Dasgupta<sup>4</sup>

<sup>1</sup>Assistant Professor, BCA Department, The Heritage Academy

<sup>2,3,4</sup> Student, BCA Department, The Heritage Academy

**Abstract:** The study presents a statistical analysis and visualizations of the Bubble sort algorithm in the worst case implemented using R on two different PCs. The data size starts from 100, goes up to 4000 with an interval of 100. The scatter plots of the worst case performances of Bubble sort for both the computers under study reveal approximate smooth patterns with few jitters (in both the cases). The study also shows the said performances closely follow quadratic trends with  $R^2$  values in both the cases  $\geq .99$ . The Mann Whitney U test which had been employed to study whether the performances were same across categories of computers showed that the distribution of the run time was same across categories of computers (Significance of Mann Whitney U test = .108).

**Keyword:** Bubble sort, Worst case, Performance Visualization, Mann Whitney U test, Quadratic trend

## I. INTRODUCTION

The Bubble sort algorithm is one of the naïve sorting algorithms in which the adjacent values are compared and exchanged. The worst case time complexity of Bubble sort is  $O(n^2)$ . In the past, the Bubble sort algorithm was studied by many researchers in different ways. Some had visualized the performances of the Bubble sort [1][3]. Comparisons of performance of Bubble sort with other sorting algorithms were carried out by many researchers [2][4][5][6]. Improved versions of Bubble sort were proposed by some researchers [7][8][9].

In the present study, the researchers have analysed the worst case run times of Bubble sort against the data sizes on two personal computers to unearth a simple question i.e. “Whether the worst case performances of Bubble sort are similar in both the computers under study” using both visualization and statistical techniques. The authors have considered the worst case run times of Bubble sort against the data sizes as the “Performance” of the said algorithm in the worst case.

## II. OBJECTIVES OF THE STUDY

- Scatter plot visualization of the performances of the Bubble sort (*worst case run times*)
- Quadratic trend line visualization of the Bubble sort (*worst case run times*)
- Identifying whether the distribution of the worst case run time of the Bubble sort is same across the categories of computers

## III. HYPOTHESES

### A. Null hypotheses

- $H_{01}$ : The worst case performances of the Bubble sort in both the computers under study are non linear
- $H_{02}$ : The worst case performances of the Bubble sort in both the computers under study are quadratic in nature
- $H_{03}$ : The distribution of the worst case run time is same across categories of computer

### B. Alternative Hypotheses

- $H_{11}$ : The worst case performances of the Bubble sort in both the computers under study are linear
- $H_{12}$ : The worst case performances of the Bubble sort in both the computers under study are not quadratic in nature
- $H_{13}$ : The distribution of the worst case run time is not same across categories of computer

## IV. METHODOLOGY

### A. The Configurations of The Computers Under Study

- Computer 1: Atom Quad Core, 2 GB RAM

2) Computer 2: AMD Sempron Dual Core, 4 GB RAM

**B. Algorithm implementation:**

1) Using R

**C. Data size**

1) Data\_size = 100; Data\_size ≤ 4000; Data\_size = Data\_size + 100

**D. Data Collection**

- 1) Ten (10) number of observations i.e. the worst case run times in seconds of the Bubble sort for each data size for each of the computers had been noted
- 2) The average run times for each data size for each of the computers had been calculated

**E. Visualization**

- 1) Scatter plots of the worst case performances of the Bubble sort for both the computers had been developed where data size was taken as x – axis and average run times was taken as y – axis
- 2) Quadratic trend lines had been drawn on the performance scatter plots for both the computers
- 3) Box plot of Mann – Whitney U test had been developed

**F. Analysis**

- 1) Mann : Whitney U test had been performed to test the distribution of the worst case run times
- 2) The null hypothesis of the Mann : Whitney U test is as follows: The distribution of the worst case run time is same across categories of computer (i.e.  $H_{03}$ )
- 3) The decision rule is as follows: if the significance of the Mann : Whitney U test < .05 then the null hypothesis is rejected

**V. DATA ANALYSIS**

The scatter plots of the worst case performances of Bubble sort for both the computers are shown below (Fig. 1 & Fig. 2).

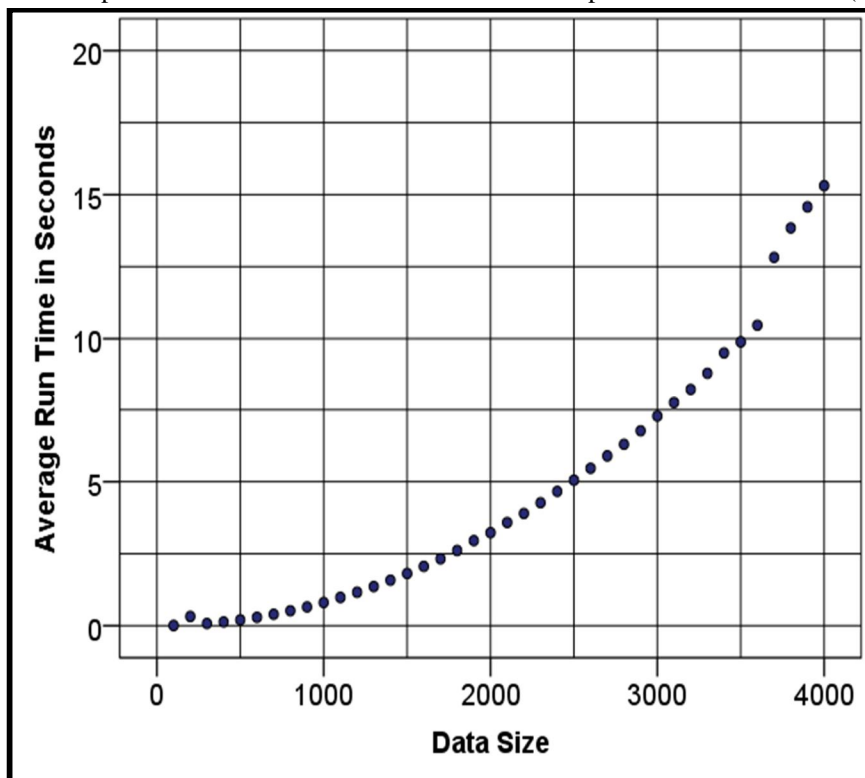


Fig. 1 Scatter plot of the performance for Computer 1

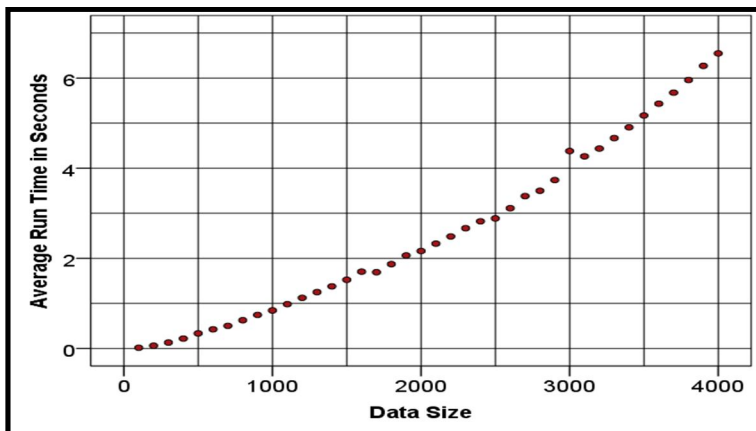


Fig. 2 Scatter plot of the performance for Computer 2

In case of Computer 1 (Fig. 1), the visual observation clearly shows a break in the smooth pattern from data size three thousand seven hundred (3700) onwards, whereas in case of Computer 2 (Fig. 2) an approximate smooth pattern is present with few deviations e.g. one thousand seven hundred (1700), three thousand (3000).

To contemplate the said performances of both the computers, quadratic trend lines are drawn on the scatter plots (Fig. 3 & Fig. 4).

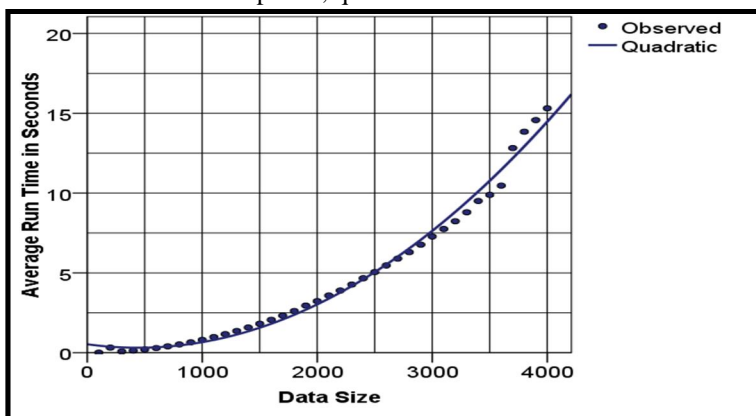


Fig. 3 Scatter plot and quadratic trend curve of the performance for Computer 1

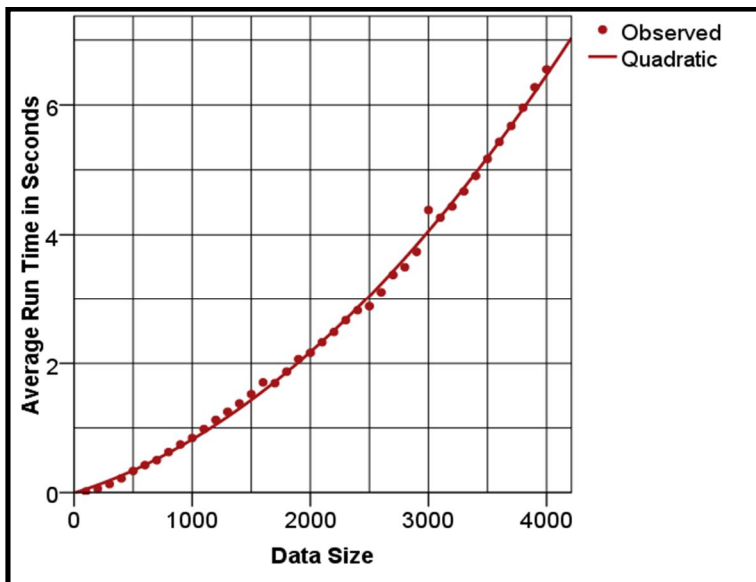


Fig. 4 Scatter plot and quadratic trend curve of the performance for Computer 2

From Fig. 3, we had observed that the performance of Bubble sort in the worst case for Computer 1 closely follows the quadratic trend line ( $R^2 \geq .99$ ). Similarly, from Fig. 4 we had observed that the performance of Bubble sort in the worst case for Computer 2 closely follows the quadratic trend line ( $R^2 \geq .99$ ).

#### A. Result of Mann-Whitney U test

The asymptotic significance of the Mann-Whitney U test performed on readings of both the computers is .108. Note that the significance level is .05. The box plot for visualizing the result is given below (Fig. 5):

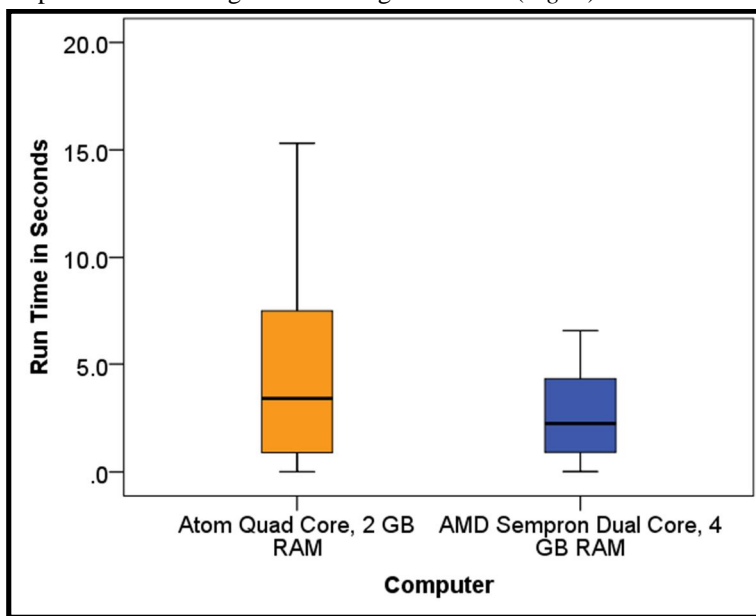


Fig. 5 Box plot of Mann-Whitney U test

### VI. CONCLUSION

From the above figures (Fig. 1 & Fig. 2) we observe that the scatter plots for both the computers under study, exhibits non linear trends. Similarly through visual inspections of the Fig. 3 & Fig. 4, we observe that the performance of the Bubble sort in the worst case strongly follows quadratic trends. The result of the Mann Whitney U clearly suggests that the distribution of run time is same across categories of computer. Therefore, the researchers have accepted all the null hypotheses ( $H_{01}$ ,  $H_{02}$  and  $H_{03}$ ) in this particular case. This particular study has given rise to a question which is as follows: whether this phenomenon is true for this case only. Finding a suitable answer to this question will be our future scope of study.

### REFERENCES

- [1] Das, D., Das, P., Dey, R., & Modak, S. (2016). Performance Visualization of Bubble Sort in Worst Case Using R Programming in Personal Computer. International Journal of Recent Trends in Engineering & Research, 2(10), 186-190.
- [2] Das, D., Das, P., Dey, R., & Modak, S. (2017). Comparative Analysis of Gnome Sort and Bubble Sort in the Worst Case using Performance Visualization in Personal Computer. International Journal for Research in Applied Science & Engineering Technology, 5(V), 589-593.
- [3] Das, D (2017). Visualization of Performance of Bubble Sort in Worst Case in Personal Computer using Polynomial Curve Fitting Technique. International Journal of Advanced Research in Computer Science, 8(1), 34-36.
- [4] Dhillon, A., & Singh, A. (2012). Analysis of Software Metrics for Bubble Sort and Selection Sort. International Journal of Computer Applications & Information Technology, 1(1), 38-40.
- [5] Popli, S., Talwar, A., & Gupta, S. (2014). A COMPARISON BETWEEN THREE SORTING ALGORITHMS BASED UPON THE TIME COMPLEXITY. International Journal of Advanced Technology in Engineering and Science, 02(01), 643-647.
- [6] Chandrawat, Y. S., & Rathore, Y. (2015). A COMPARATIVE STUDY OF SORTING ALGORITHM BASED ON THEIR TIME COMPLEXITY. INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY, 4(10), 334-337.
- [7] Khairullah, M. (2013, July). Enhancing Worst Sorting Algorithms. International Journal of Advanced Science and Technology, 56, 13-26.
- [8] Appiah, O., & Martey, E. M. (2015). Magnetic Bubble Sort Algorithm. International Journal of Computer Applications IJCA, 122(21), 24-28.
- [9] Mundra, J., & Pal, B. L. (2015). Minimizing Execution Time of Bubble Sort Algorithm. International Journal of Computer Science and Mobile Computing, 4(9), 173-181.