ISSN No. 0976-5697

Volume 8, No. 1, Jan-Feb 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

# Visualization of Performance of Bubble Sort in Worst Case in Personal Computer using Polynomial Curve Fitting Technique

Dipankar Das
Department of BCA
The Heritage Academy
Kolkata, India

***Abstract:*** Bubble sort is one of the simple sorting algorithms. In this study, the researchers had used polynomial curve fitting technique to visualize the performance of Bubble sort in the worst case in a personal computer. To identify the best fit (i) R square, (ii) Adjusted R square, (iii) Root mean square error, (iv) Akaike information criterion (AIC) and (v) Bayesian information criterion (BIC) had been used. The Bubble sort algorithm in the worst case had been implemented using C programming language and the algorithm had been run for data size two thousand five hundred (2500) to data size twenty thousand (20000) with an interval of five hundred (500). For each data size one hundred (100) observations (execution time in seconds) had been recorded and for each data size the median value of the observations (execution time in seconds) had been calculated. Thus, the researchers had calculated thirty six (36) data points (data size versus median value of execution time in seconds). The polynomial curve fitting had been tried and tested on these thirty six (36) data points (data size versus median value of execution time in seconds). In total twenty four (24) models starting from linear model to polynomial of degree 24 had been employed in this study to identify the best model and among these models "Polynomial of degree 2" model had been identified as the best model.

***Keywords:*** Polynomial curve fitting; Bubble sort; AIC; BIC; Performance visualization

## I. INTRODUCTION

Bubble sort is one of the simple sorting algorithms. It uses divide and conquer philosophy to do its job. In the worst case, the time complexity of Bubble sort is $O(n^2)$ [1][2]. We know that curve fitting is used to capture the trend in the dataset by assigning a single function across the entire range of data [3]. In this study, the researchers have used polynomial curve fitting technique to visualize the performance of Bubble sort in the worst case in a personal computer. To identify the best fit among all the tested models, we have employed R square, Adjusted R square, Root mean square error, Akaike information criterion (AIC) and Bayesian information criterion (BIC).

## II. LITERATURE REVIEW

Bubble sort algorithm has been studied by many researchers over the years. We have listed few of them below:

Alnihoud & Mansi (2010) in their research work had given enhanced Bubble sort algorithm [4]. Khairullah (2013) had presented enhancement of Bubble sort in the research work [5]. Sharma (2015) had proposed a new technique to enhance the performance of Bubble sort in the worst case [6].

Kapur, Kumar & Gupta (2012) had proposed an End to End Bi Directional sorting algorithm which used to be more efficient than Bubble sort, Selection sort and Insertion sort [7]. Brijwal, Goel, Papola & Gupta (2014) in their paper had presented a both ended sorting algorithm which was faster than Bubble sort algorithm [8].

Dhillon & Singh (2012) had analyzed Bubble sort and Selection sort on different software metrics (LOC, Execution time, Program length, Program vocabulary etc.) [9]. Popli, Talwar & Gupta (2014) had compared the Bubble sort, Insertion sort and Selection sort on the basis of time complexity [10]. Chandrawat & Rathore (2015) had compared five different sorting algorithms (Insertion, Selection, Bubble, Quick and Merge) in the following seven parameters – sorting approach, sorting type, in place, time complexity, algorithm type, stability and strategy [11].

Das, Das, Dey & Modak (2016) had analysed the Bubble sort algorithm in the worst case in a personal computer. In their study, they had implemented the Bubble sort algorithm using R programming and showed that the performance approximately followed quadratic curve [12].

## III. OBJECTIVES

- To identify and visualize the best polynomial curve which can be fitted to the performance of Bubble sort in the worst case (Data size versus median value of execution time in seconds) in the personal computer under study.

## IV. METHODOLOGY

### A. *Data Generation*

The Bubble sort algorithm in the worst case had been implemented using C programming language and the algorithm had been run for data size two thousand five hundred (*2500*) to data size twenty thousand (*20000*) with an interval of five hundred (*500*). For each data size, one hundred (*100*) observations (execution time in seconds) had been recorded and the median value of the observations (execution time in seconds) for each data size (*2500 – 20000* with an interval of *500*) had been calculated. In total, we had thirty six (*36*) data points under study.

### B. *Polynomial Curve Fitting and Visualization*

For the purpose of the study, we had considered data size (ds) as x – axis and median value of execution time in seconds (t_median) as y – axis. The researchers had used twenty four (24) numbers of polynomial models for the study.

For all the models we had calculated R square, Adjusted R square and Root mean square error (RMSE). We know that the

model which has high value of R square, Adjusted R square i.e. close to one (1) and low value of RMSE i.e. close to zero (0) fits the data well [13]. The models which had high value of R square and Adjusted R square along with low value of RMSE had been identified for further study. We had calculated both Akaike information criterion (AIC) and Bayesian information criterion (BIC) for all the identified models (which had been identified for further study). The best model was identified based on the following criteria: (i) the model which is having lowest AIC value [14][15] and (ii) the model which is having lowest BIC value [15].

After, the identification of the best models using both the information criteria, the curve had been drawn for the model for the purpose of visualization along with the scatter plots of the data points (ds versus t_median).

### C. *Hardware Used*

Intel(R) Core(TM)2 Duo CPU, 2.93 GHz, 1.99 GB RAM

### D. *Operating System*

Windows XP, Professional N, Version 2002, Service Pack 3

### E. *Software Used For Data Analysis*

R version 3.3.1 (2016-06-21)

## V. DATA ANALYSIS AND FINDINGS

The R square, Adjusted R square and RMSE of the polynomial models tested on the data set (ds versus t_median) are given in the following table (Table I).

Table I. R square, Adjusted R square and RMSE of the Polynomial models

| *Name of the Model* | *R square* | *Adjusted R square* | *RMSE* |
|---|---|---|---|
| Linear | 0.9596853 | 0.9584996 | 0.08206947 |
| Polynomial of degree 2 | 0.9998664 | 0.9998583 | 0.004725351 |
| Polynomial of degree 3 | 0.9998664 | 0.9998538 | 0.004724899 |
| Polynomial of degree 4 | 0.9998664 | 0.9998492 | 0.004724495 |
| Polynomial of degree 5 | 0.9998671 | 0.999845 | 0.00471188 |
| Polynomial of degree 6 | 0.9998676 | 0.9998402 | 0.004703041 |
| Polynomial of degree 7 | 0.9998844 | 0.9998556 | 0.00439377 |
| Polynomial of degree 8 | 0.9998851 | 0.9998511 | 0.004380644 |
| Polynomial of degree 9 | 0.9998908 | 0.999853 | 0.00427089 |
| Polynomial of degree 10 | 0.9998909 | 0.9998472 | 0.004269776 |
| Polynomial of degree 11 | 0.9998911 | 0.9998412 | 0.004265824 |
| Polynomial of degree 12 | 0.9998915 | 0.999835 | 0.004256675 |
| Polynomial of degree 13 | 0.9998948 | 0.9998327 | 0.004191642 |
| Polynomial of degree 14 | 0.9999022 | 0.999837 | 0.004042801 |
| Polynomial of degree 15 | 0.9999022 | 0.9998288 | 0.004042656 |
| Polynomial of degree 16 | 0.9999076 | 0.9998298 | 0.003928675 |
| Polynomial of degree 17 | 0.9999118 | 0.9998284 | 0.003839633 |
| Polynomial of degree 18 | 0.99993 | 0.9998559 | 0.003419605 |
| Polynomial of degree 19 | 0.9999318 | 0.9998508 | 0.003375959 |
| Polynomial of degree 20 | 0.9999373 | 0.9998538 | 0.003235331 |
| Polynomial of degree 21 | 0.9999415 | 0.9998537 | 0.003127313 |
| Polynomial of degree 22 | 0.9999417 | 0.999843 | 0.003121683 |
| Polynomial of degree 23 | 0.999944 | 0.9998368 | 0.003057438 |
| Polynomial of degree 24 | 0.9999444 | 0.9998232 | 0.00304685 |

From the above table (Table I) we observe that all the tested models are having very high values of R square and Adjusted R square along with low value of RMSE. Therefore, all the models have been selected for further explorations by using information criteria (AIC & BIC).

The AIC and BIC of the polynomial models tested on the data set (ds versus t_median) are given in the following table (Table II).

Table II. AIC & BIC of the Polynomial Models

| *Name of the Model* | *AIC* | *BIC* |
|---|---|---|
| Linear | -71.85004806 | -67.09949124 |
| Polynomial of degree 2 | -275.382988 | -269.0489123 |
| Polynomial of degree 3 | -273.3898839 | -265.4722892 |
| Polynomial of degree 4 | -271.3960384 | -261.8949247 |
| Polynomial of degree 5 | -269.5885498 | -258.5039173 |
| Polynomial of degree 6 | -267.7237413 | -255.0555898 |
| Polynomial of degree 7 | -270.6212945 | -256.3696241 |
| Polynomial of degree 8 | -268.8367052 | -253.0015158 |
| Polynomial of degree 9 | -268.6635996 | -251.2448913 |
| Polynomial of degree 10 | -266.6823851 | -247.6801578 |
| Polynomial of degree 11 | -264.7490575 | -244.1633113 |
| Polynomial of degree 12 | -262.9036454 | -240.7343803 |
| Polynomial of degree 13 | -262.0121448 | -238.2593607 |
| Polynomial of degree 14 | -262.6152948 | -237.2789918 |
| Polynomial of degree 15 | -260.6178725 | -233.6980506 |
| Polynomial of degree 16 | -260.6770413 | -232.1737004 |
| Polynomial of degree 17 | -260.3276749 | -230.2408151 |
| Polynomial of degree 18 | -266.6689935 | -234.9986147 |
| Polynomial of degree 19 | -265.593878 | -232.3399803 |
| Polynomial of degree 20 | -266.6573535 | -231.8199368 |
| Polynomial of degree 21 | -267.1022749 | -230.6813394 |
| Polynomial of degree 22 | -265.2320058 | -227.2275513 |
| Polynomial of degree 23 | -264.7292325 | -225.141259 |
| Polynomial of degree 24 | -262.9790207 | -221.8075283 |

From the above table (Table II) we observe that the model named "Polynomial of degree 2" is having lowest AIC and BIC values. Therefore, the "Polynomial of degree 2" model is identified as the best model among the 24 tested polynomial models in this case.

## VI. CONCLUSION

The visualization of the performance of the Bubble sort in the worst case in personal computer under study by using the "Polynomial of degree 2" curve which is the best polynomial curve identified among the twenty four (24) tested polynomial models is given below (Figure 1).

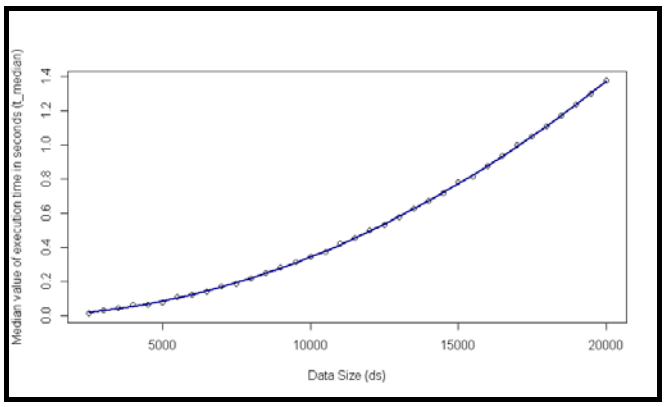


Figure 1. Visualization of the performance of the Bubble sort in the worst case in personal computer using "Polynomial of degree 2" curve

In this empirical study, we had used twenty four (24) polynomial models and found that out of these twenty four (24) models the "Polynomial of degree 2" (i.e. quadratic model) model best fitted the data points. The scope of this study is limited to the visualization of the performance of Bauble sort algorithm in the worst case in personal computer under study with the help of the best identified polynomial curve which in this case turned out to be "Polynomial of degree 2" curve. Here, we have not tested the dataset with any other types of curves other than polynomial curves. Therefore, the study does not reveal any information regarding the questions like whether

the performance of Bubble sort in the worst case can be best fitted to any other type of curve, can we develop any predictive model for the performance of Bubble sort in the worst case. At the same time, the finding of the study holds good for the data points under observation (data size from 2500 to 20000 with an interval of 500) in a particular hardware & software implementation platforms. Therefore, the study also does not tell us what will happen beyond this range of data size in the same platform (both hardware & software) or what will be the performance of the Bubble sort algorithm in the worst case in other platforms (both hardware & software) within this data range or beyond this data range. Getting answers to these questions will surely be our future scope of study.

## VII. REFERENCES


[1] *Q&A for Previous Year Questions - Subject: CPDS (B.Tech. I Year) - Subject Code: GR11A1003 - UNIT - V* [PDF]. (n.d.). Retrieved January 10, 2017, from http://www.it.griet.ac.in/wp-content/uploads/2014/08/UNIT-V_QA.pdf

[2] Sorting. (n.d.). Retrieved January 19, 2017, from https://www.cs.cmu.edu/~adamchik/15-121/lectures/Sorting%20Algorithms/sorting.html

[3] *Numerical Methods Lecture 5 - Curve Fitting Techniques* [PDF]. (n.d.). Retrieved January 10, 2017, from http://web.iitd.ac.in/~pmvs/courses/mel705/curvefitting.pdf

[4] Alnihoud, J., & Mansi, R. (2010). An Enhancement of Major Sorting Algorithms. *The International Arab Journal of Information Technology , 7* (1), 55-62.

[5] Khairullah, M. (2013). Enhancing Worst Sorting Algorithms. *International Journal of Advanced Science and Technology , 56*, 13-26.

[6] Sharma, V. (2015). A New Approach to Improve Worst Case Efficiency of Bubble Sort. *International Research Journal of Computer Science , 6* (2), 54-58.

[7] Kapur, E., Kumar, P., & Gupta, S. (2012). PROPOSAL OF A TWO WAY SORTING ALGORITHM AND PERFORMANCE COMPARISON WITH EXISTING ALGORITHMS. *International Journal of Computer Science, Engineering and Applications , 2* (3), 61-78.

[8] Brijwal, A., Goel, A., Papola, A., & Gupta, J. K. (2014). Both Ended Sorting Algorithm & Performance Comparison with Existing Algorithm. *International Journal of IT, Engineering and Applied Sciences Research , 6* (3), 4-9.

[9] 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.

[10] 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.

[11] 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.

[12] 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 , 02* (10), 186-190.

[13] Fit. (n.d.). Retrieved January 10, 2017, from https://in.mathworks.com/help/curvefit/evaluating-goodness-of-fit.html

[14] MAZEROLLE, M. J. (n.d.). APPENDIX 1: Making sense out of Akaike's Information Criterion (AIC): It...and interpretation in model selection and inference from ecological data [PDF]. Retrieved January 10, 2017, from http://avesbiodiv.mncn.csic.es/estadistica/senseaic.pdf

[15] Maydeu-Olivares, A., & Garcı´a -Forero, C. (2010). Goodness-of-Fit Testing [PDF]. Elsevier Ltd. Retrieved January 10, 2017, from http://www.ub.edu/gdne/amaydeusp_archivos/encyclopedia_of_education10.pdf