International Journal of Recent Trends in Engineering & Research
ISSN(ONLINE) : 2455-1457
IMPACT FACTOR : 3.344

Visualization of Binary Search in Worst Case Using Spline Interpolation Curve Fitting in Personal Computer

Dipankar Das
Assistant Professor, The Heritage Academy, Kolkata, India

Abstract—The paper aims to visualize the performance of binary search algorithm in the worst case scenario in a personal computer (laptop) using spline interpolation curve fitting. The researchers have chosen Linux operating system and OpenJDK runtime environment for simulating the binary search algorithm in the worst case and have run the java code of binary search for data size one thousand to twenty thousand with an interval of five hundred. For each data size one thousand observations have been noted. To eliminate and/or minimize the effect of outliers (if any) from the observations for each data size three different approaches have been employed in this paper. These approaches are (i) calculation of mean execution time for each data size, (ii) identifying the largest cluster of execution time for each data size by using Two-step clustering and finding the mean value of the largest cluster for each data size and (iii) identifying the center of the largest cluster of execution time for each data size by using K-means clustering where K = 2. At the end, three different spline interpolation curves have been obtained for each of these three cases and we observe that all of them display different patterns.

Keywords—Two-Step clustering, K-Means clustering, Binary search, Interpolation, Spline, Curve fitting

I. INTRODUCTION

From our early days, we use visualization to communicate our ideas with others. It is one of the most effective ways of communication. In this form of communication, we use images to communicate our ideas in a very subtle way. In many areas e.g. science, education, engineering visualization have a lot of applications [1]. The data visualization is the pictorial or graphical representation of data. The data visualization is described as the effort to understand the significance of data by placing it in a visual context [2]. According to Bostock (2014), the algorithm visualizations are very unusual; the designers experiment with novel forms for better communication and this is enough reason to study them [3]. In the present study, the researchers have tried to visualize the performance of binary search algorithm in the worst case scenario in a personal computer (laptop) using spline interpolation curve.

II. RELATED WORK

The related works pointed out that the analysis of binary search algorithm had been conducted by the researchers in many ways. The researchers had carried out (i) comparative analysis between linear search and binary search [4],[5],[6],[7], (ii) comparative analysis between linear search, binary search and interpolation search [12] and (iii) comparative analysis between the execution times of binary search on two personal computers [8]. Polynomial curves were fitted to the execution times of binary search in the worst case in a personal computer for different data sizes to analyze the binary search algorithm [10]. Curve estimation techniques were employed for visualization of the performance of binary search [11]. At the same time, modified binary search algorithm [9] had also been proposed by some researchers.

III. OBJECTIVES OF THE STUDY

To visualize the performance of binary search in the worst case in a personal computer (laptop) using Spline interpolation curve fitting.

IV. METHODOLOGY

The experimental dataset had been generated by running the java code of binary search algorithm in the worst case scenario for different data sizes (starting from data size = *1000* to data size = *20000* with an interval of *500*) in a personal computer (laptop) having hardware configuration of Intel(R) Core(TM)2 Duo CPU T5870 @2.00 GHz. For each data size the researchers had collected one thousand (*1000*) observations and altogether thirty nine thousand (*39000*) observations had been noted (*1000* for each data size * *39* data sizes). The entire experiment had been conducted on Ubuntu 12.04.4 LTS (Linux) operating system and OpenJDK runtime environment having Java version 1.6.0_36. The observed execution times of binary search in worst case scenario were in nano-seconds. The researchers had calculated mean execution time (M1) of binary search in the worst case for each data size. Two-Step clustering algorithm and K-means clustering algorithm (where k = *2*) were also employed in this research work to identify the largest cluster for each data size. The (i) mean value (M2) of the largest cluster (using Two-Step clustering) for each data size and (ii) cluster center (M3) of the largest cluster (using K-means clustering, where k = *2*) for each data size had been noted. Three (*3*) different scatter plots had been generated which were as follows: M1 *versus* data size, M2 *versus* data size and M3 *versus* data size. Three (*3*) spline interpolation curves were drawn passing through the following datasets: M1 *versus* data size, M2 *versus* data size and M3 *versus* data size. SPSS version 20 software had been used for data analysis and spline interpolation curve fitting.

V. DATA ANALYSIS & FINDINGS

The spline interpolation curve of M1 *versus* the data size is given below:

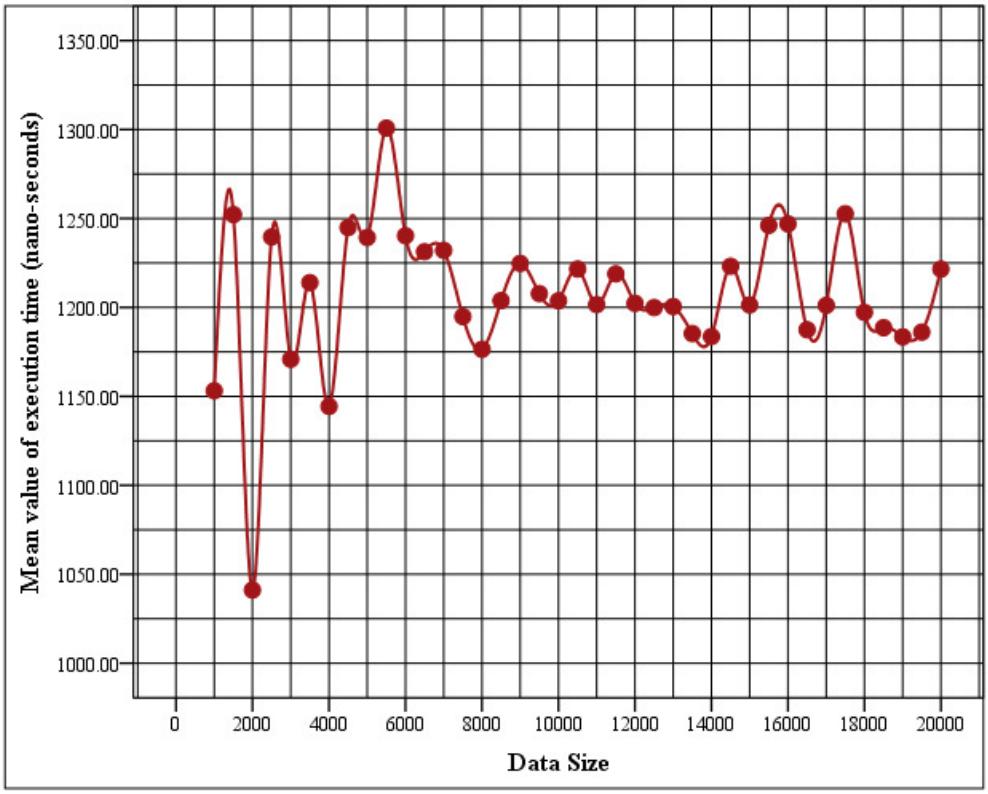

Figure 1. Spline interpolation curve of M1 versus data size for Binary Search in the worst case

In the above figure, the dark red circles represent the data points and the dark red curve represents the spline interpolation curve of M1 *versus* data size for binary search in the worst case. It has been observed from the above figure (Figure 1) that majority of the data points (*32* data points) lies between one thousand one hundred seventy five (*1175*) nano-seconds to one thousand two hundred fifty (*1250*) nano-seconds. The figure shows a non linear trend. It has been observed that only one (*1*) point (for data size *2000*) is showing relatively low value in comparison with other values.

The spline interpolation curve of M2 *versus* the data size is given below:

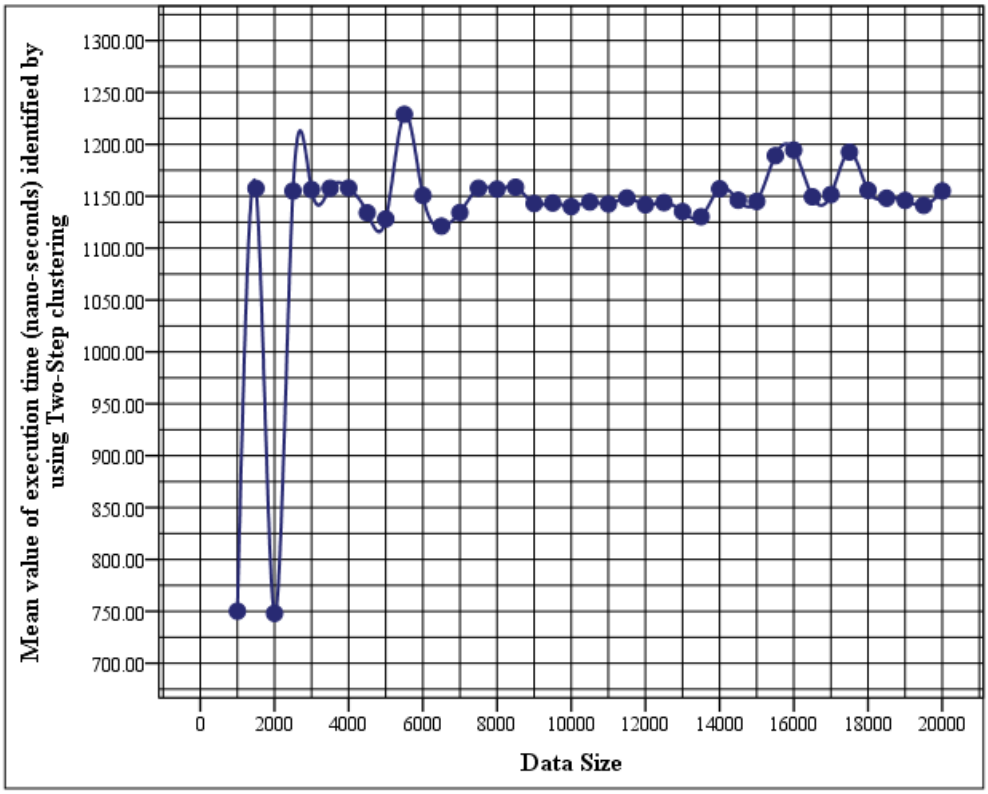

Figure 2. Spline interpolation curve of M2 versus data size for Binary Search in the worst case

In the above figure, the dark blue circles represent the data points and the dark blue curve represents the spline interpolation curve of M2 *versus* data size for binary search in the worst case. It has been observed from the above figure (Figure 2) that majority of the data points (*32* data points) lies between one thousand one hundred twenty five (*1125*) nano-seconds to one thousand one hundred seventy five (*1175*) nano-seconds. The figure shows a non linear trend. It has been also observed that two (*2*) data points (for data size *1000* and *2000*) are displaying relatively very low values in comparison with other values.

The spline interpolation curve of M3 *versus* the data size is given below:

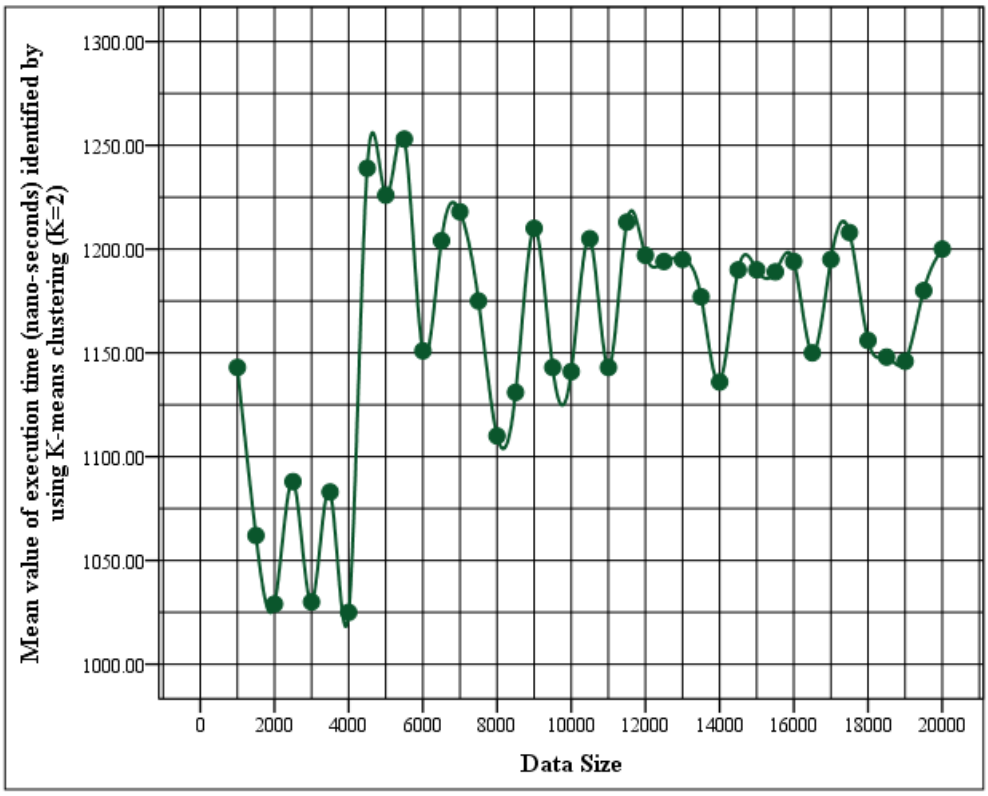

Figure 3. Spline interpolation curve of M3 versus data size for Binary Search in the worst case

In the above figure, the dark green circles represent the data points and the dark green curve represents the spline interpolation curve of M3 *versus* data size for binary search in the worst case. It has been observed from the above figure (Figure 3) that majority of the data points (*29* data points) lies between one thousand one hundred twenty five (*1125*) nano-seconds to one thousand two hundred twenty five (*1225*) nano-seconds. The figure shows a non linear trend. It has been observed that six (*6*) data points (for data size *1500*, *2000*, *2500*, *3000*, *3500* and *4000*) are displaying relatively low values in comparison with other values.

VI. CONCLUSION

The main objective of this paper is to visualize the performance of the binary search in the worst case generated by simulating the binary search algorithm in a personal computer (laptop). For achieving this objective the researchers have used several techniques which are as follows: (i) data generation, (ii) mean value calculation for each data size and (iii) drawing spline interpolation curve passing through the data points. The data has been generated by running java code of binary search in the worst case scenario and noting down the execution time of the said code in nano-seconds for different data sizes (*1000* to *20000* with an interval of *500*). For each data size we have noted one thousand (*1000*) execution times. In this paper, the researchers have used three (*3*) different techniques to eliminate and/or minimize the effect of outliers (if any) from the observations of the execution time for each data size. The first technique used by the researchers is finding mean value of the execution time (mean value of *1000* observations) for each data size. The second technique is finding mean value of the largest cluster of execution time for each data size by using Two-step clustering algorithm and the last one is finding cluster center of the largest cluster of execution time for each data size by using K-means clustering algorithm where k = 2. At the end, three (*3*) spline interpolation curves are obtained for these three (*3*) data sets. From these curves we observe that all the three (*3*) curves are showing different patterns. Therefore, it may be concluded that if we try to find mathematical equations of these data points we are sure to get different results. This entire research work throws some open questions to us *e.g.* what will be the mathematical equations of the

best curves that can be fitted to these three datasets, will all of them lies in the same mathematical family, what will be goodness of fit statistics of the best fit curves, can we effectively reduce the range of y axis by using K-means algorithm and increasing the value of k. Finding answers to these questions will certainly remain our future scope.

REFERENCES

1. Visualization (computer graphics). (n.d.). Retrieved July 01, 2016, from https://en.wikipedia.org/wiki/Visualization_(computer_graphics)
2. What is data visualization? - Definition from WhatIs.com. (n.d.). Retrieved July 01, 2016, from http://searchbusinessanalytics.techtarget.com/definition/data-visualization
3. Bostock, M. (2014, June 26). Visualizing Algorithms. Retrieved July 01, 2016, from https://bost.ocks.org/mike/algorithms/
4. Kumari, A., Tripathi, R., Pal, M., & Chakraborty, S. (2012). Linear Search Versus Binary Search: A Statistical Comparison For Binomial Inputs. *International Journal of Computer Science, Engineering and Applications (IJCSEA)*, *2*(2). Retrieved October 10, 2015, from http://www.airccse.org/journal/ijcsea/papers/2212ijcsea03
5. Sapinder, Ritu, Singh, A., & Singh, H.L. (2012). Analysis of Linear and Binary Search Algorithms. *International Journal of Computers & Distributed Systems, 1*(1).
6. Parmar, V. P., & Kumbharana, C. (2015). Comparing Linear Search and Binary Search Algorithms to Search an Element from a Linear List Implemented through Static Array, Dynamic Array and Linked List. *International Journal of Computer Applications*, *121*(3). Retrieved October 10, 2015, from http://search.proquest.com/openview/a9b016911b033e1e8dd2ecd4e7398fdd/1?pq-origsite=gscholar
7. Pathak, A. (2015). Analysis and Comparative Study of Searching Techniques. *International Journal of Engineering Sciences & Research Technology, 4*(3), 235-237.
8. Das, D., Kole, A., Mukhopadhyay, S., & Chakrabarti, P. (2015). Empirical Analysis of Binary Search Worst Case on Two Personal Computers Using Curve Estimation Technique. *International Journal of Engineering and Management Research*, 5(5), 304 – 311. Retrieved November 10, 2015, from http://www.ijemr.net/DOC/EmpiricalAnalysisOfBinarySearchWorstCaseOnTwoPersonalComputersUsingCurveEstimationTechnique%28304-311%29.pdf
9. Chadha, A. R., Misal, R., & Mokashi, T. (2014). Modified Binary Search Algorithm. *arXiv preprint arXiv:1406.1677*.
10. Das, D. (2016, June). Polynomial Curve Fitting of Execution Time of Binary Search in Worst Case in Personal Computer. *International Journal of Recent Trends in Engineering & Research,* 02(06), 245-250.
11. Das, D., Kole, A., & Chakrabarti, P. (2015, November). Sample Based Visualization and Analysis of Binary Search in Worst Case Using Two-Step Clustering and Curve Estimation Techniques on Personal Computer. *International Research Journal of Engineering and Technology, 02*(08), 1508-1516.
12. Roy, D., & Kundu, A. (2014). A Comparative Analysis of Three Different Types of Searching Algorithms in Data Structure. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(5). Retrieved October 10, 2015, from http://www.ijarcce.com/upload/2014/may/IJARCCE6C%20a%20arnab%20A%20Comparative%20Analysis%20of%20Three.pdf

