# TO DESIGN A EFFICIENT N-NARY SEARCH ALGORITHM

Mr. Mani Ram
Assistant Professor, Department of Computer Science, World Institute of Technology-Haryana

Abstract- As the amount of data is increasing over the tremendous rate, it is extremely viable to imply smart analysis N-Nary search algorithm plays a vital role in saving the execution time of a program in given data set. The paper presents a classical efficient algorithm for searching the object on computer base application various searching algorithm are available but this N-Nary searching algorithm provide less execution time for program.

Keywords- BS, SC, BCS , WCS, ASA, PSC, Smoothed score Distribution, ROC, AUC, H measure Curve, augmented Naïve Bayes, Logistic regression and ross Quinlan.

## 1.Introduction

A search **algorithm** is the step-by-step procedure used to locate specific data among collection of data. It is considered a fundamental procedure in computing. In computer science, when **searching** for data, the difference between a fast application and a slower one often lies in the use of the proper search **algorithm**.

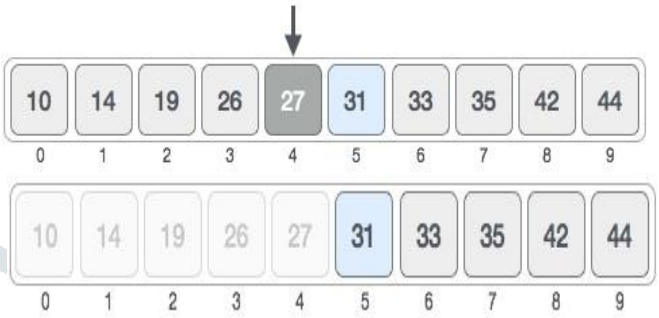
**Linear Search**:- A Linear search in C programming: The following code implements linear search (**Searching** algorithm) which is used to find whether a given number is present in an **array** and if it is present then at what location it occurs. It is also known as sequential search.


Fig.1.1. Linear Search

**Binary Search:-** For a binary search to work, it is mandatory for the target array to be sorted. We shall learn the process of binary search with a pictorial example. The following is our sorted array and let us assume that we need to search the location of value 31 using binary search.


First, we shall determine half of the array by using this formula − Mid= low+(High-Low)/2.

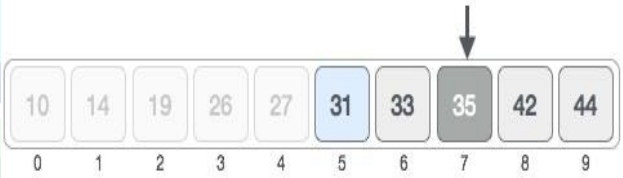
Here it is, 0 + (9 - 0 ) / 2 = 4 (integer value of 4.5). So, 4 is the mid of the array




We change our low to mid + 1 and find the new mid value again. We change our low to mid + 1 and find the new mid value again.

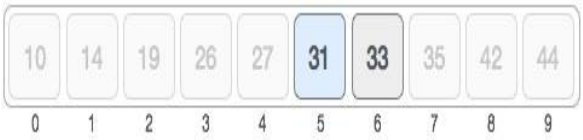
Low=Mid+1, Mid= low+(High-Low)/2.

Our new mid is 7 now. We compare the value stored at location 7 with our target value 31.

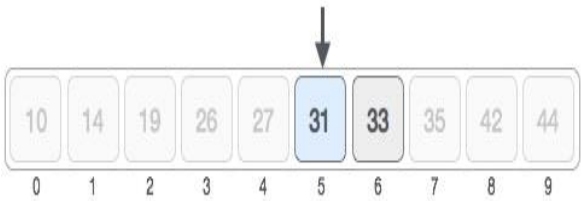



Our new mid is 7 now. We compare the value stored at location 7 with our target value 31.


The value stored at location 7 is not a match, rather it is more than what we are looking for. So, the value must be in the lower part from this location.




Hence, we calculate the mid again. This time it is 5.




We compare the value stored at location 5 with our target value. We find that it is a match.

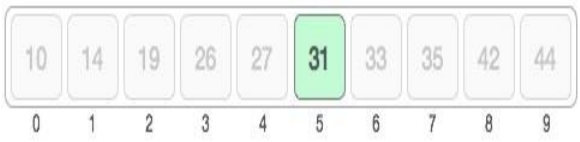



Fig 1.2. Binary Search

We conclude that the target value 31 is stored at location 5.

Binary search halves the searchable items and thus reduces the count of comparisons to be made to very less numbers.

**Ternary Search**

A **ternary search algorithm** is a technique in computer science for finding the minimum or

maximum of a unimodal function. A ternary search determines either that the minimum or

maximum cannot be in the first third of the domain or that it cannot be in the last third of the

domain, then repeats on the remaining two thirds. A ternary search is an example of a divide

and conquer algorithm.

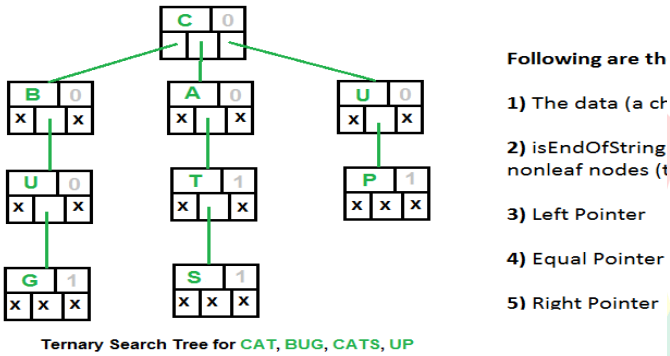



Fig.1.3. Ternary Search

## 2. Efficiency of searching

**Algorithm efficiency** A measure of the average execution time necessary for an **algorithm** to complete work on a set of data. **Algorithm efficiency** is characterized by its order. There are two main measure of efficiency of algorithm. It means time complexity and space complexity order .T under supervised classification scheme, the problem domain of binary classification is discussed in course of work. With the available training data set the model is trained that predicts the outcome into states i.e. 0 or 1

For search algorithms, the main steps are the comparisons of list values with the target value. Counting these for data models representing the ***best case***, the ***worst case***, and the ***average case*** produces the following table.

Table- 2.1

| **Case** | **Comparisons for N = 100,000** | **Comparisons in terms of** |
|---|---|---|
| **Best Case** (fewest comparisons possible) | 1 (target is first item in list) | 1 |
| **Worst Case** (most comparisons possible) | 100,000 (target is last item in list) | N |
| **Average Case** (avg number of comparisons) | 50,000 (target is middle item in list) | N/2 |

Where N-is the number of items in the list.

## 3. PROBLEM IDENTIFICATION

The domain of searching have many searching algorithm. These algorithm have problem of time taken during the execution of program or efficiency of algorithm. So efficiency of algorithm is improved by taking suitable action. Amongst the diverse applications of searching tend to formulate the efficient searching algorithm.

## 4.GENERAL PROCEDURE

The data set illustrated below in the **Table 2** is taken from uci repository that contains details of Bank customer contacted during marketing campaign upon which assessment needs to be done whether or not customer will subscribe a deposit based on training the machine and testing it on future entries. Nowadays marketing selling campaigns comprise an effective methodology to extravagate their business plans. Companies use a specific marketing scheme while contracting them to meet a specified goal. Communication with customers takes place through various channels. Contacts can be divided into inbound and outbound, depending on which side triggered the contact (client or contact centre), with each case posing different challenges (e.g., outbound calls are often considered more intrusive). It should be stressed that due to internal competition and current financial crisis, there are huge pressures for European banks to increase a financial asset.

Table 2.Raw data Set

| age | job | marital | education | default | balance | housing | loan | contact | day | month | duration | campaign | pdays | previous | poutcome | y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | unemployed | married | primary | no | 1787 | no | no | cellular | 19 | oct | 79 | 1 | -1 | 0 | unknown | no |
| 33 | services | married | secondary | no | 4789 | yes | yes | cellular | 11 | may | 220 | 1 | 339 | 4 | failure | no |
| 20 | student | single | secondary | no | 502 | no | no | cellular | 30 | apr | 261 | 1 | -1 | 0 | unknown | yes |
| 35 | management | single | tertiary | no | 1350 | yes | no | cellular | 16 | apr | 185 | 1 | 330 | 1 | failure | no |
| 30 | management | married | tertiary | no | 1476 | yes | yes | unknown | 3 | jun | 199 | 4 | -1 | 0 | unknown | no |
| 59 | blue-collar | married | secondary | no | 0 | yes | no | unknown | 5 | may | 226 | 1 | -1 | 0 | unknown | no |
| 35 | management | single | tertiary | no | 747 | no | no | cellular | 23 | feb | 141 | 2 | 176 | 3 | failure | no |
| 36 | self-employed | married | tertiary | no | 307 | yes | no | cellular | 14 | may | 341 | 1 | 330 | 2 | other | no |
| 78 | retired | divorced | primary | no | 229 | no | no | telephone | 22 | oct | 97 | 1 | -1 | 0 | unknown | yes |
| 39 | technician | married | secondary | no | 147 | yes | no | cellular | 6 | may | 151 | 2 | -1 | 0 | unknown | no |
| 41 | entrepreneur | married | tertiary | no | 221 | yes | no | unknown | 14 | may | 57 | 2 | -1 | 0 | unknown | no |
| 43 | services | married | primary | no | -88 | yes | yes | cellular | 17 | apr | 313 | 1 | 147 | 2 | failure | no |
| 39 | services | married | secondary | no | 9374 | yes | no | unknown | 20 | may | 273 | 1 | -1 | 0 | unknown | no |

The data set is converted to the categorical format .

## 4.CONCLUSION

The thesis finishes up with the way of comparing and analysing different models with help of parameters and plots. However while evaluating performance of classifier we generally talk about two types of misclassification costs, one is false alarms ( classified as positive but actually is negative) and other false positives( classified negative but in reality is positive).The study evolved by us has taken into account interrelated measures which are depicted as combination of both types of costs. For binary classification we need to minimize false positive as well as false negatives entries. The end user classifier need to trade-off between these two objectives (hence the weights for each of the costs).For such trade-offs ROC curve is one option. It has been scaled down to a value known as area under curve i.e. AUC for better comparisons amongst classifiers. However this AUC does not consider misclassification in a coherent way for which H measure has been dealt and taken in account .From our observations the value of AUC ,H measure as well as precision ranks high for ADA BOOST in comparison to other classification models which are deciding parameters of accuracy inference in our data set problem.

## 5. FUTURE SCOPE

In future work, we propose to collect more client based data, in order to check if high quality predictive models can be achieved without contact-based information. We also plan to apply the best DM models in a real setting, with a tighter interaction with marketing managers, in order to gain a valuable feedback. Appropriate analysis with advanced classification models shall improve time and cost involved in assessment.

## 6.REFERENCES

[1] P. Domingo's, "A Few Useful Things to Know about Machine Learning", Communications of the ACM ,Volume 55, Issue 10, October 2012.
[2] P. Langley, "Selection of relevant features in machine learning", Proceedings of the AAAI Fall Symposium on Relevance, New Orleans, LA: AAAI Press, 1994.
[3] M.Sahare and H.Gupta, "A Review of Multi-Class Classification for Imbalanced Data", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume 2, Issue5, September 2012.
[4]Y. Singh, P.K. Bhatia and O. Sangwan, "A Review of Studies on Machine Learning Techniques", International Journal of Computer Science and Security, Volume 1, Issue 1, 2007.
[5] G. Ratsch, "A Brief Introduction into Machine Learning", Friedrich Miescher Laboratory of the Max Planck Society, 2004.