IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Decision Tree Problem Solving Techniques: A Review

N. Sandhu[1*], S. Kumar[2]

[1*]Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India
[2] Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India

*Corresponding Author: sandhunikita0096@gmail.com, Tel.: +91-8109377730

Available online at: www.ijcseonline.org

Accepted: 24/Jul/2018, Published: 31/July/2018

Abstract— The problem of classification is one of the major problems associated with data mining. Numerous classification algorithms have been implemented, although there was hardly an algorithm that surpasses all individual algorithms with respect to the standard. A decision tree is a type of classification method in which the end result is the class to which the data belongs. There are many problems faced by decision tree and this paper considers two out of them. First problem faced by decision tree is finding the optimal solution, which is resolved by heuristic techniques more quickly and further efficiently than conventional techniques. Another problem facing a decision tree is scalability issue, which is solved by RainForest framework. RainForest framework considers the scalability problem and has different types of algorithms that work in different types of cases. This article provides a brief overview of the framework of RainForest and Heuristics and steepest ascent hill climbing which are utilized to overcome the scalability issues and the limitation of finding the optimal solution respectively.

Keywords—Decision tree, Classification, RainForest, Heuristics and Steepest ascent hill climbing.

I. INTRODUCTION

Decision tree is a coherent model that efficiently and consistently combines the sequence of the basic test, where each test compares a numerical attribute with a threshold value. The decision tree includes the attribute of the input and output attributes and vice versa for triggering by association between input and output attributes, using controlled methods [1]. Decision trees are the strongest models used in prediction [2]. The advantage of decision tree method is that decision trees can usually process all types of variables, even if they do not have values [3]. Decision trees are the most popular methods because of their high degree of tolerance of noisy data and handling of massive data set [4]. Classification of large data sets is an important problem in data mining. Classification in the field of data mining is the form of data analysis task that is used to extract models that describe data classes. The process of classification consists of two stages; Firstly, it is a learning process in which the data of training data are analyzed by a classification algorithm. The second stage is the use of the model for classification, and test data sets are used to assess the accuracy of the classification rules [5]. There are many problems that need to be considered during the building of the decision trees. We focus on the problem of finding the optimal solution and solving the issues of scalability. In the literature, many classification algorithms have been proposed. The study shows that, so far no algorithm has surpassed all other algorithms in terms of accuracy and comprehensive result. The RainForest framework is used to separate the scalability issues in decision tree design from central functions in terms of quality [6]. Heuristics are used to solve the problem of finding the optimal solution in large data space when classic methods cannot find a precise solution. Heuristics have many types of algorithms associated with it and steepest ascent hill climbing is a type of hill climbing heuristic algorithm.

Section I, contains an introduction to the decision tree, classification, and problems associated with the decision tree. In section II, we will look at three classes that can be used to solve the decision tree problems in terms of scalability and finding the optimal solution using the RainForest framework, heuristics and the steepest ascent of the hill climbing, and also to review some of the documents that used these classes to solve their problems. In section III, we present our conclusion and future work.

II. TECHNIQUES FOR SOLVING DECISION TREE PROBLEMS

We represent three classes that can be used to solve these problems that arise when building a decision tree. These classes do not solve all the problems associated with decision trees, but eliminate some problems.

II.I. RainForest framework

RainForest closes the gap between the constraints on the main data sets of algorithm memory in the machine learning and statistical documentation, and ascendable necessities of the mining framework. The principal dissimilarity betwixt the greedy top-down scheme and a finely-tied RainForest scheme, the second highlights the chief element, AVC-set. AVC-set permits you to split the tree structure of the scalability problems of the classification algorithms to determine the splitting criterion [6].

RainForest permits us to execute a variety of procedures, incorporating hybrid, repetitive and recursive perspectives, alongside their compromises. A number of cases as used to explain the AVC-set, which is used to define a specific node with different values for different attributes and classes. The set of all AVC-sets is known as the AVC-group of the node [7]. Contingent upon the quality of available main memory, there are three cases [6]:

(i) AVC-group of the root node chips in the main memory.

(ii) Only each separate AVC-set chips in the main memory, for the root node, except the AVC-group which cannot be placed in main memory.

(iii) Not even a single AVC-set of the root can be placed in the main memory.

The RainForest framework has three types of algorithms, they are: RF-Read, RF-Write, RF-Vertical and RF-Hybrid. For the root node, RF-Hybrid, RF-Read and RF-Write requires the AVC-group to fit in the main memory. The algorithm RF-Vertical works if each AVC-set fits in memory, but the full AVC-group does not work [6].

The recent literature review of the RainForest framework is discussed below:

Johannes Gehrke, et al. [6] proposed an integrated infrastructure to create workstations of decision trees that share scaling algorithms to form a decision tree for the key functions that define the tree standard in their article. This approach offers performance improvements of more than five times the Sprint algorithm which is the fastest scalable classification algorithm. However, unlike Sprint, the general algorithm necessitates a definite minimal quantity of main memory that is equivalent to a set of discrete values in the input column relations. If to store individual AVC- sets there is enough memory, so it is very likely that we will get a very good acceleration over Sprint; if the memory is enough to store all AVC-sets for the node, even better acceleration. The key concept is to note that the decision trees in the literature are based on the criteria of splitting them into the AVC-group of a relatively compressed node.

Yi Yang and Wenguang Chen [7], an optimized algorithm was proposed in the article inspired by RainForest. RainForest framework has various algorithms, of which the best is a hybrid betwixt the classic application of a recursive and repetitive application that uses more memory, but includes minimal write processes. Using more sophisticated criteria for switching betwixt the two algorithms, we can obtain improved performance even if all statistical information is stored in memory. The following aspects differs this work from RainForest:

(i) Different assumptions. In two papers, the various underlying assumptions were made.

(ii) Different data structures. To reduce the number of calculations, Taiga provides an additional layer for recording, to which the template belongs. The Taiga transposition is a further educational structure of data memory.

(iii) Different switching criteria algorithms. RF-Hybrid algorithm performs switching criteria and switches to RF-Write once AVC-group cannot be written into memory. On the contrary, Taiga-Hybrid algorithm utilizes a further complex principle for switching betwixt taiga-iterative and taiga-recursive.

Haixun Wang and Carlo Zaniolo [8] presented a new technique in which by storing histograms for pairs of attributes:

(i) Significant acceleration compared to traditional classifiers based on the separation of individual attributes and

(ii) The capability to construct classifiers that utilize a consecutive combination of values of pairs of non-categorical attributes as a separation criterion. In the following way, the primary aims were performed:

1) The CMP utilizes sampling methods to lessen the calculation of the Gini index. It also avoids the materialization of list of sorted attributes and, thus reducing the disk I/O;

2) Using a simplified version of the proposed valuation method in CLOUDS, CMP provides, eliminates the loss of precision generated by the sampling. However, dissimilar to CLOUDS, this removes the requirement for additional submissions in the data set to discover the accurate separation;

3) After each data set survey using prediction techniques, CMP allows for decision trees to grow more than one level;

4) CMP utilizes a linear set of attributes as a criterion of split, to create a better model than those that can be obtained from the SPRINT or CLOUDS.

ID3 and C4.5 suffers from a lack of bias toward multi-valued attributes, seeking to prefer unbalanced splits and creating sections that can lead to the formation of dense trees. Devashish Thakur, et al. [9] presented a document aimed at improving the decision tree algorithm for ID3 and C4.5 by simply changing the methods for selecting the attributes. The changes alter the calculation of the gain in the ID3 calculation and split the information in C4.5, and we get a decision tree that is higher than the classification accuracy. To accelerate the classification process, preprinting strategy and a RainForest approach were used in the background. The Oracle program is used to check the accuracy of the algorithm classification with the ID3 and C4.5 algorithm and it turned out that the proposed algorithm has a much higher accuracy.

Lixin Fu [10] used statistic trees to compute a data cube, and then built a decision tree on top of it. The new algorithm generates trees with the same prediction accuracy as existing decision tree algorithms, such as SPRINT and RainForest, but significantly improves performance. This article also provides a system architecture that easily integrates DBMS, OLAP, and data mining. A new classifier is proposed that extracts some of the calculated data cubes to customize decision trees for classification. Once the data cubes are computed by scanning the source data once and stored in the statistic trees, they are ready to respond to OLAP requests. Due to the combination of technologies, from data cubing and tree-based classification, the way of seamless integration of data mining and data systems of cubic systems is laid. An architectural design of such an integrated system is proposed.

Senjuti Basu Roy, et al. [11] have investigated navigation technique based on minimal methods for corporate database systems based on a phased search paradigm is proposed. The proposed methods dynamically suggest the faces for drilling into the database, so the cost of navigation is minimized. Boundaries are chosen based on their ability to quickly move to the most promising tuples, as well as the ability of the user to provide the desired values for them. It also provides solutions that can take into account the tuple offset introduced by any ranking function. They provide scalable and efficient implementation of solutions and present results that show the effectiveness and reliability of the solutions.

II.II. Heuristics

Heuristic search hypothesis - Solutions to problems are represented in the form of character structures. The system of physical symbols uses its intelligence to solve problems by searching, that is, by generating and gradually changing the structure of symbols until it creates a solution structure. In the serial heuristic search, the main question is always: What to do next? In the search for a tree, this question, in turn, has two components:

(i) From which node in the tree we will look for the next, and

(ii) What direction will we take from this node? [12]

Information useful in answering the first question can be interpreted as measuring the relative distance of various nodes from the target. Information useful in answering the second question, tells us which direction to look for [12]. There are many types of heuristic algorithms that are used to solve various problems that are encountered in different areas. One of them is the steepest ascent. It is part of the heuristic hill climbing algorithm.

The recent literature review of heuristics is discussed below:

X.-Z. Wang, et al. [13] presented an article in 2001, where three heuristic algorithms were analyzed and compared for the generation of fuzzy decision trees. The authors proposed one of them. The comparison is two-fold. One of them is an analytical comparison based on an extended choice of attributes and mechanisms of reasoning; another is an experimental comparison on the basis of the tree size and precision training. The motive of this document is to study the relative advantages and disadvantages of three heuristics and to present few practical recommendations on according to what to select the suitable guidance for a task. It was noted that in order to obscure ID3, it could generate the main strengths of the comparably compact tree algorithm without considerable arithmetic attempt. For the method of Yuan and Shaw, key strengths that the algorithm can adequately manage non-specificity that exists in the classification. For the proposed heuristics, key strengths of the algorithm can produce weighted fuzzy rules with great learning precision.

Peter D. Turney [14] presented a new algorithm for classification of costs, inexpensive classification using expensive tests, also known as ICET, also pronounced “ice tea”. ICET uses a genetic algorithm [15] for the evolution of the displacement population for the decision tree induction algorithm (modified version [16]). The following are the features of ICET:

(i) They are sensitive to the test cost.

(ii) They are sensitive to the classification evaluation cost.

(iii) Using a genetic search algorithm, greedy search is combined.

(iv) If the test value depends on the choice of the second test, the conditional costs can be handled.

(v) With delay as a result, tests are characterized by immediate test results.

Using five sets of real medical data, a series of three experiments were conducted to analyze ICET behavior. The main characteristics of the five algorithms in five sets of data were examined in the first group. ICET was found to be much less expensive than other algorithms. The second set of experiments examined the stability of ICET for various modifications of their input. The reliability of ICET is shown by the results. The search for ICET in motion is the third set of experiments. It turned out that the search can be improved by sowing the movement of the primary population.

Yudong Zhang, et al. [17] presented an article that proposed a new way for detecting spam. The purpose of this new method is to reduce the false positive mistake of incorrectly making new spam as spam. First, a method is used to select a function based on covers to extract important functions. Second, the decision tree was selected as a model with C4.5 classifier as a learning algorithm. Third, the cost matrix was introduced, giving different weights to two types of errors, namely, false positive errors and false negative errors. To determine the weight parameter, how to adjust the relative importance of the two types of errors. Fourth, in order to reduce the error outside the sample, the K-fold cross-link test was used. Finally, as a search strategy for a subset, a dual PSO (Particle Swarm Optimization) system was used with a (Mutation Binary PSO) MBPSO engine. The main contribution and technical innovations in this article are the following:

(i) Prepare a test of Kolmogorov-Smimov's hypothesis about characteristics related to the length of capital, and had values of p less than 0.001.

(ii) The method used to select functions based on wrappers, which allows to achieve high classification accuracy, as well as to select important functions.

(iii) The decision tree C4.5 is used as a wrapper classifier and the PSO is used with mutation (MBPSO) as a wrapper search strategy.

(iv) The results of the selection function with MBPSO were found to be better than (Genetic algorithm) GA, (Iterated Local Search) ILS, (Ant Colony Optimization) ACO, (Restarted simulated annealing) RSA, (Binary PSO) BPSO and PSO results.

(v) It is provided that the functions related to capital are related to efficiency.

(vi) The wrapper has proved to be more efficient than filters in terms of efficient classification.

S. Rasoul Safavian, David Landgrebe [18] presented a paper that provides an overview of the current (Decision tree classifier) DTC design methods and various existing problems. Some of the goals were as follows:

(i) To combine disparate problems in the decision tree classifiers closely together and, possibly, to motivate some new ideas;

(ii) Provide a more unified set of decision tree classifiers;

(iii) And to warm "random" users of these methods from possible "traps" of each method. Recognizing the difficulties associated with the development of optimal DTC, many heuristic methods have been proposed in the literature. Finally, it is pointed out that the productivity of DTC depends on how well the tree is designed; special attention is paid to this stage. In particular, the time spent in development can be fully justified when the number of samples to be classified is significant. A heuristic procedure that includes a time component in the structure of a tree structure is proposed in [19].

II.III. Steepest Ascent Hill Climbing

One of the mutants of the heuristic algorithm is the steepest ascent hill climbing. The simplest technique to look for is hill climbing, which is also called greedy search or steep ascent. Steep ascent rises and moves the current node to the child with maximum accuracy, the end if no child is improved compared with the current node [20].

In the steepest ascent, all the successors of nodes are examined and compared to their significance, and then the best of them is taken as a node of a successor. This leads to a complete local search and identification of the best possible successor of the node at any time. The steepest ascent on the rise always requires the presence of the best nodes as successor nodes for the continuation of the procedure.

Literature review of some papers that used Steepest Ascent Hill Climbing is given below:

In the task of selecting a subset of characteristics, learning algorithms experience the problem of selecting the appropriate subset of functions to focus, that ignores everything else. Ron Kohavi and George H. John [20] have presented a paper examining the strengths and weaknesses of the complex approach and showing a number of advanced designs. Bypass approach compared with induction without selecting a subset of objects and Relief-filter approach to the selection of a subset of objects. It is noted that for the wrapper approach, the space of the search, operators, and evaluation function is required. For the estimation function, the cross-validation method is used as a precision estimate based on the results of [21]. Optimal functions are dependent on the particular prejudices and heuristic learning algorithms, and therefore the approach of the wrapper, of course, corresponds to this definition. A common search space to the operators of addition and deletion is used as a basis for comparison of two engines: hill climbing and the best-first search. It seems that the best-first search with the composite operators is to be a strong performer and improves the Naïve-Bayes, C4.5 and ID3, regarding accuracy and clarity, the number of functions. The comparisons include two distinct families of asynchronous algorithms: Naïve-Bayes and decision trees. In some cases, the results of these changes have been very different. Artificial data sets were mostly noiseless, while real data contained noise.

Jayanta Basak [22] presented a paper presenting a new classification system called interactive adaptive decision trees known as OADT, a tree-like network, such as a decision tree, capable of learning online, for example, neural networks. The new objective measure obtained for supervised learning with OADT. OADT is useful in the case of data flow and is constrained to the memory position where data sets cannot be stored explicitly. The OADT structure is similar to the HME structure (a hierarchical mixture of experts) [23, 24]. However, OADT, leaf nodes is the solution rather than the root node, which differs from the HME structure. The model uses gradient descent training, similar to the class of gradient descent algorithms used in primary neural networks. Gradient descent algorithms can be included more efficiently; with the exception of gradient descent, in OADT, to improve performance. With a set of real data, it should be noted that OADT works better than HME. However, this requires further study to determine whether OADT gives higher estimates than HME as a whole.

Yueh-Min Huang, et al. [25] presented a document that analyzes many classification algorithms and is used to predict the creditworthiness of the bank's customers based on current account information. The goal was to determine the range of the credit score, which can be implemented by risk managers. This study suggests two basic methods for the purification of training materials, which contain asymmetric distribution and conflicting information. One technique includes random choice in instances of training data established in accordance with the values of attributes of the class, for making the same data distribution for each class in the training sets. Another way is to create a combinatorial classifier with a combination of the PRISM device and ID3 class. Suggested methods can obtain constant high coefficient of TP prediction for each class. In addition, this study also provides additional information on the financial sector to identify the most vulnerable factors for particular classes. Results show that the classification accuracy without cleaning the data, for all classes is about 80%; but after cleaning, the percentage of training data sets increases by 64% for each class. Finally, an error of about 15% is allowed.

Parikshit Gopalan, et al. [26] illustrated an algorithm for the uniform distribution at the entrances. Conceptually presented article is parallel to recent work in the direction of the agnostic study of half-spaces [27]; algorithmically it is much more complicated. The results show that for uniform distribution without queries, agnostically studying sparse polynomials reduced to learning parities with random noise, for example, to the problem of noisy parity. Algorithm 2 (Gradient Descent using KM) does not solve this problem; it is possible to construct a function f, where algorithm 2 gives a hypothesis with an accuracy of ½, whereas there exists a formula DNF (Disjunctive normal form) of polynomial size with a correlation of at least 1/ poly (n) with f.

III. CONCLUSION and Future Scope

In this paper, we have shown that three types of methods can be used to solve the problems identified in the designing of the decision tree. The problem of scalability can be solved with the help of RainForest framework and the second problem of finding the optimal solution can be solved using heuristics and steepest ascent hill climbing algorithms. These methods can be combined together to overcome other problems as well. Alterations can be done to RainForest framework and it can be used with steepest ascent hill climbing algorithm to overcome the problem of excessive memory usage while finding the best optimal solution.

REFERENCES

[1] Anuradha, G. Gupta, " A self explanatory review of decision tree classifiers", In the Proceedings of the 2014 IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, India, pp.1-7, 2014.

[2] R. Garg, V. Mongia, "Dimensionality Reduction and Comparison of Classification Models for Breast Cancer Prognosis", International Journal of Computer Sciences and Engineering, Vol.6, Issue.1, pp.308-312, 2018.

[3] M. Fernandes, "Data Mining: A Comparative Study of its Various Techniques and its Process", International Journal of Scientific Research in Computer Sciences and Engineering, Vol.5, Issue.1, pp.19-23, 2017.

[4] N. Ghuse, P. Pawar, A. Potgantwar, "An Improved Approach For Fraud Detection in Heath Insurance Using Data Mining Techniques", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.5, pp.13, 2017.

[5] D. Singh, N. Choudhary, J. Samota, "Analysis of Data Mining Classification with Decision tree Technique", Global Journal of Computer Science and Technology, Software & Data Engineering, Vol.13, Issue.13, pp.1-6, 2013.

[6] J. Gehrke, R. Ramakrishnan, V. Ganti, "RainForest- A framework for Fast Decision Tree Construction of Large Datasets", In the Proceedings of 24th VLDB conference, New York, USA, Vol.98, pp.416-427, 1998.

[7] Y. Yang, W. Chan, "Taiga: Performance Optimization of the C4.5 Decision Tree Construction Algorithm", Tsinghua Science and Technology, Vol.21, Issue.4, pp.415-425, 2016.

[8] H. Wang, C. Zaniolo, "CMP: A Fast Decision Tree Classifier Using Multivariate Predictions", In the Proceedings of the 16th International Conference on Data Engineering (Cat. No. 00CB37073), San Diego, CA, USA, pp.449, 2000.

[9] D. Thakur, N. Markandaiah, D. S. Raj, "Re optimization of ID3 and C4.5 decision tree", In the Proceedings of the 2010 IEEE International Conference on Computer and Communication Technology (ICCCT), Allahabad, Uttar Pradesh, India, pp.448-450, 2010.

[10] L. Fu, "Novel Efficient Classifiers Based on Data Cube", International Journal of Data Warehousing and Mining (IJDWM), Vol.1, Issue.3, pp.15-27, 2005.

[11] S. B. Roy, H. Wang, G. Das, U. Nambiar, M. Mohania, "Minimum-effort driven dynamic faceted search in structured databases", In the Proceedings of the 17th ACM conference on International and knowledge management, Napa Valley, California, USA, pp.13-22, 2008.

[12] A. Newell, H. A. Simon, "Computer science as empirical inquiry: symbols and search", ACM Turing Award Lecture, New York, USA,2007.

[13] X.-Z. Wang, D. S. Yung, E. C. C. Tsang, " A Comparative study on heuristic algorithms for generating fuzzy decision trees", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol.31, Issue.2, pp.215-226, 2001.

[14] P. D. Turney, "Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm", Journal of artificial intelligence research, Vol.2, pp.369-409, 1994.

[15] J. J. Grefenstette, "Optimization of Control Parameters for Genetic Algorithms", IEEE Transactions on System, Man, and Cybernetics, Vol.16, Issue.1, pp.122-128, 1986.

[16] J. R. Quinlan, "C4.5: Programs for machine learning", Elsevier, California, USA, 2014.

[17] Y. Zhang, S. Wang, P. Phillips, G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection", Knowledge-Based Systems, Vol.64, pp.22-31, 2014.

[18] S. R. Safavian, D. Landgrebe, "A survey of decision tree classifier methodology", IEEE Transactions on Systems, Man, and Cybernetics, Vol.21, Issue.3, pp.660-674, 1991.

[19] P. H. Swain, H. Hauska, "The decision tree classifier: Design and potential", IEEE Transactions on Geoscience Electronics, Vol.15, Issue.3, pp.142-147, 1977.

[20] R. Kohavi, G. H. John, "Wrappers for feature subset selection", Artificial Intelligence, Vol.97, Issue.1-2, pp.273-324, 1997.

[21] R. Kohavi, " A study of cross-validation and bootstrap for accuracy estimation and model selection", International Joint Conference on Artificial Intelligence (IJCAI), Vol.14, Issue.2, pp.1137-1145, 1995.

[22] J. Basak, "Online Adaptive decision trees", Neural Computation, Vol.16, Issue.9, pp.1959-1981, 2004.

[23] M. I. Jordan, R. A. Jacobs, 'Hierarchical mixtures of experts and the EM algorithm", In the Proceedings of 1993 International Conference on Neural Networks (IJCNN-93), Nagoya, Japan, pp.1339-1344, 1993.

[24] M. I. Jordan, R. A. Jacobs, 'Hierarchical mixtures of experts and the EM algorithm", Neural Computation, Vol.6, Issue.2, pp.181-214, 1994.

[25] Y.-M. Huang, C.-M. Hung, H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem", Vol.7, Issue.4, pp.720-747, 2006.

[26] P. Gopalan, A. T. Kalai, A. R. Klivans, "Agnostically learning decision trees", In the Proceedings of the fortieth annual ACM symposium on Theory of computing, Victoria, British Columbia, Canada, pp.527-536, 2008.

[27] A. T. Kalai, A. R. Klivans, Y. Mansour, R. A. Servedio, "Agnostically learning halfspaces", In the Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05), Pittsburgh, PA, USA, 2005.

Authors Profile

Ms. Nikita Sandhu pursed Bachelor of Engineering from Sushila Devi Bansal College of Technology, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Indore in 2016. She is currently pursuing Master of Technology from Department of Computer Science and Engineering, Guru Jambheshwar University of Science & Technology, Hisar since 2016.

Mr. Sunil Kumar pursed Bachelor of Technology from Guru Jambheshwar University of Science & Technology, Hisar in 2009 and Master of Technology from Department of Computer Science and Application, Kurukshetra University, Kurukshetra in 2011. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Department of Computer Science and Engineering, Guru Jambheshwar University of Science & Technology, Hisar since 2014.

