International Journal of Civil Engineering and Technology (IJCIET)
Volume 10, Issue 03, March 2019, pp. 713-718. Article ID: IJCIET_10_03_068
Available online at http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=10&IType=03
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

ASSORTMENT OF DATA WITH BINARY TREE USING SUPERVISED LEARNING

Boshra F. Zopon Al_Bayaty
Computer Science Department, College of Science, University of Mustansiriyah, Baghda, Iraq

Murooj Khalid Ibraheem
College of Arts, University of Mustansiriyah, Baghdad Iraq

ABSTRACT

Classification can be addressed by supervised & unsupervised approaches. Decision trees like binary tree follows supervised approach for these executions. There is necessity to classify the data and provide the consolidated update and delete. This paper talks about an empirical associated functionality. This work is completed by following standard norms of Software Engineering. Agile scum methodology helped to ensure exact implementation in Software Development Life Cycle (SDLC) as stated in Software Requirement specification (SRS).

Keywords: Supervised learning, assorted tree for binary search, Agile SCRUM model.

Cite this Article: Boshra F. Zopon Al_Bayaty, and Murooj Khalid Ibraheem, Assortment of Data with Binary Tree Using Supervised Learning, *International Journal of Civil Engineering and Technology*, 10(3), 2019, pp. 713-718.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=10&IType=03

1. INTRODUCTION

Software Engineering methodologies Artificial Intelligence are few of the core competencies required for the hassle-free implementation of the Assortment of data with binary tree using supervised learning. Also, Machine learning concepts, Algorithms helps in bridging the gap between the A.I. models and their actual implementation while dealing with the data. In machine learning Machines are trained to take their own decisions by observation of data, identifying the pattern and learning through this will help in decision making newly considered data that is the item to be classified or predicted.

- Supervised Learning:In this Artificial Neural Network system is trained with respect to target to extract the process or algorithm; later on, the same algorithm is applied for the classification or regression of the subject.

Training Input (known)-Process (**To be known**)-Output (Target-Known)
Testing Input (known)-Process(**known**)-Output (**Known at the Instance**)

UnSupervised Learning: - Unsupervised model does not require any prior knowledge as there is no learning involved. Rather it uses cluster or likewise approaches and identify the patterns in the provided data through the vigilant analysis of the data.

Pattern Identification Input (known)-Process (**known - analysis**)-Output (Conclusions about patterns)
Testing Input (known)-Process (**known- clustering**)-Output (**Known at the Instance**)

- Reinforcement Learning: - Reinforcement learning deals with the approach where the agents are trained to adopt with the environment can take the decisions accordingly.
- Deep Learning: - Most often the conclusions are made by overlooking the patterns or training the system with limited number of layers. This may happen because of the lack of reserved time for the process or the severity of the training required for the said situations. Complexity or depth of the learning process is enhanced to firmly identify the exact patterns or to train the system with more accuracy.

2. LITERATURE SURVEY

Generic purpose of the training may be classification or regression.

Classification: - System is categorized into several categories and the sample is identified the class from which it belongs to.

Regression: - It is the process of tracing the path to which the forthcoming values are supposed to be adhered to.

There are some very important classic algorithms which guidelines the research in supervised learning approaches. There number of important algorithm available but only those algorithms are listed which might be helpful in heading towards Assortment of data with binary tree using supervised learning.

Naïve Bayes Algorithm

Naïve Bayes probability considers the negation while calculating the probability, it is also known as the posterior probability. Given the prior probability along with the proof then the posterior probability can be calculated. This can be stated in terms of P(A) and p(B) as below-

P(A/B) = (P(A)*P(B/A))/p(B)

In easier language it can be stated as

Posterior= (Prior*Likelihood)/Evidences

2.1. K - Nearest Neighborhood Algorithm (K.N.N.)-

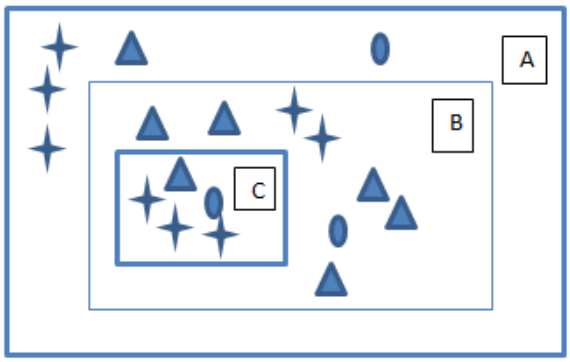

Figure 1 the K - Nearest Neighborhood Algorithm

KNN Algorithm works on the phenomenon of calculation of Euclidean distance between the Object to be classified and Actual members of the class. As mentioned in the diagram mentioned above, consider there are two categories star and triangle and oval are the object to be classified. A, B, C represents different values of K

A. K =16, In this case target object oval belongs to class **star** as there are maximum representatives equal to 8
B. B – K=11, In this case target object oval belongs to class **triangle** as there are maximum representatives equal to 6
C. C – k=4, In this case target object oval belongs to class **star** as there are maximum representatives equal to 3.

1. Support Vector Machine (S.V.M.) – Support Vector Machine is one of the most important approach delivering comparatively highest accuracy. It works in 3D representation as well, as it can categorize nonlinear and non-plainer members with the help of a functionality called kernel. Genric equation of SVM is y=wx+b
2. Where, w is weight, b is bias, x is input, and y is output for #D system there are vectors.
3. Logistic Regression – Logistic regression classifies linearly inseparable data with log function. This regression function can be represented with identity equation as below.
4. y=(1+e^-x)/ (1+e^x)
5. Problem statement for the experiment is Classification of the data with moderate level of accuracy while performing the function like insert, delete, and update.

3. SOFTWARE REQUIREMENTS

To Starts with implementation some requirements important needed for the experiment are as mentioned below:

Table 1 Important software requirements needed for implementation

Sr. No.	Name of Item	Specifications
1.	Data Set	Maximum nodes till 1000
2.	Sampling	Random Sampling Method
3.	Nature of Data	Afresh
4.	Front End	C#
5.	Back End	Files
6.	Input	Command with value. Ex. Insert 445
7.	Process	Search, Insert, Delete, Update

4. EXPERIMENTAL SET UP AND IMPLEMENTATION

Binary tree is constructed with c# to classify the data and perform search, insert, delete and update functionalities.

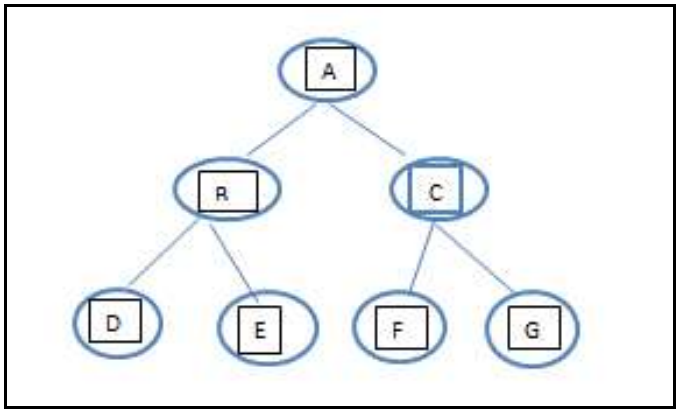

Figure 1 the Binary Tree

Each node represents unique code, which helps in identifying the point of time to perform the binary tree operation.

Insertion If the value at node is smaller than value at root then addition takes place at left otherwise addition is done at right side.

Deletion If the node to be deleted has the child node remove that node. If node has left child, it will take place. If the node to be deleted has the child node remove that node. If node has right child, it will take place.

Updating Updating process is the combination of insertion followed by deletion process as mentioned above.

5. RESULT SOFTWARE TESTING AND VALIDATION

Software testing is an important phase of any Software Development Life Cycle. Experiment performed here can be tested with the help of several parameters mentioned below. Software testing acts as a guideline for developing the software. It helps in checking the performance of the software and helps to identify the adherence to the Software Requirement Specification. Once the Software Testing and validation phase is cleared one can say that software developed follows all the norms mentioned in the SRS with required quality checks.

There are specific validation mechanisms used, in case we are classifying the data then precision, recall and F-Measure is important. The time complexity analysis and the graphical

representation is added advantage. There is absolutely no restriction on the number of nodes which can be added to the tree. Each time based on the query tree is reconstructed at run time. This binary tree assorting mechanism is better in terms of efficiency and time complexity. Output representation in terms of graph gives easy interpretation of current structure of tree.

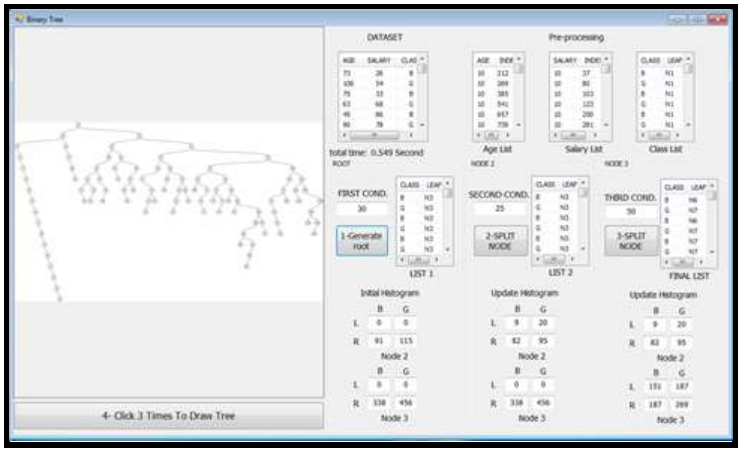

Figure 2 Output representation in terms of graph of current structure of tree.

As mentioned in the screenshot the after providing the data set one can perform the process of insertion, deletion and updating. While doing so the option L indicates left node and R indicates right node. The outcome of the graph can be clearly noted by looking at the reconstructed tree.

6. CONCLUSION

Accession time and representation of data is improved significantly by using the assorted binary tree approach. The data is in terms input which is also referred as a target object is specified at respective location in the graph. Definitely there is scope for improvement by applying this technique to the decision-making techniques that are required in day today life like purchasing decision on e-commerce sites.

ACKNOWLEDGMENTS

The authors would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) Baghdad – Iraq for its support in the present work.

REFERENCES

[1] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Implementation SVM to Solve Multiple Meaning of Word Problem, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 9, November 2014. www.ijiset.com ISSN 2348 – 7968.

[2] Classification Algorithms for Data Mining: A Survey Raj Kumar Department of Computer Science and Engineering Jind Institute of Engg. & Technolog, Jind, Haryana, India Dr. Rajesh Verma Department of Computer Science and Engineering Kurukshetra Institute of Technology & Management, Kurukshetra, Haryana, India.

[3] International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN: 2456-3307 478 Development of a Rule Based Classification System to Identify a Suitable Classifier for a

Particular Dataset Subham Datta*, Gautam, Tapas Saha Department of Computer Science, Pondicherry University, Puducherry, India

[4] 2-14- M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera. 2012. "A review on e4sembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 42, no. 4, pp. 463– 484, (Jul. 2012.).

[5] Boshra F. Zopon AL_Bayaty, Shashank Joshi, Empirical Implementation Decision Tree Classifier to WSD Problem, International Conference on Emerging Trends Science and Cutting-Edge Technology (ICETSCET), YMCA, 28, Sep 2014, International Journal of Advanced Technology in Engineering and Science www.ijates.com,Volume No.02, Special Issue No. 01, September 2014 ISSN (online): 2348 – 7550.

[6] Ms. A J. Chamatkar et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 4, Issue 5(Version 3), May 2014, pp.38-41, Importance of Data Mining with Different Types of Data Applications and Challenging Areas Ms. Aruna J. Chamatkar* , Dr. P.K. Butey** * (Department of Electronics & Computer Science, RTM Nagpur University, Nagpur) ** (Department of Computer Science, RTM Nagpur University ,Nagpur-32).

[7] Boshra F. Zopon AL_Bayaty, Dr. Shashank Joshi, "Empirical Comparative Study to Supervised Approaches for WSD Problem: Survey", International conference IEEE Canada, IHTC, Ottawa, 31 May- 4th June 2015. (IEEE Explore paper)

[8] http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7238056&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel7%2F7196573%2F7238033%2F07238056.pdf%3Farnumber%3D7238056.

[9] Briand, L. C. 2008. Novel Applications of Machine Learning in Software Testing. 2008 The Eighth International Conference on Quality Software, 3–10. https://doi.org/10.1109/QSIC.2008.29

[10] Zhang, L., Tan, J., Han, D., & Zhu, H. 2017. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discovery Today, 0(0), 1–6. https://doi.org/10.1016/j.drudis.2017.08.010

[11] A. M. Law and W. D. Kelton. 2000. Simulation Modeling and Analysis. Boston, MA, USA: McGraw Hill.

[12] H. He and E. A. Garcia. 2009. "Learning from imbalanced data," IEEE Trans. Knowledge. Data Eng., vol. 21, no. 9, pp. 1263–1284, (Sep. 2009).

[13] Tan,Steinbach, Kumar Introduction to Data Mining. 18/4/2004.

[14] Quinlan JR (1993) C4.5: Programs for machine learning. Morgan Kaufmann Publishers, San Mateo

[15] Hunt EB, Marin J, Stone PJ (1966) Experiments in induction. Academic Press, New York

[16] Quinlan JR (1979) Discovering rules by induction from large collections of examples. In: Michie D (ed)

[17] S. M. Weiss and C. A. Kulikowski.Computer Systems that Learn: Classi

[18] cation and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufman, 1991.

[19] Zheng, Z. (2000). Constructing X-of-N Attributes for Decision Tree Learning. Machine Learning 40: 35–75.

[20] *Machine Learning*, vol. 1, no. 1. ID3 is based off the Concept Learning System (CLS) algorithm.