



# Big Data Analytics for Uncertain Data Using Different Classification Techniques: A Survey

Devashri Kulkarni<sup>1</sup>, Vedika Joshi<sup>2</sup>, Suvarna Machpalle<sup>3</sup>, Vaishnavi Alurkar<sup>4</sup>

B. E Student, Department of Information Technology, VPKBIET Baramati, Maharashtra, India.

B. E Student, Department of Information Technology, VPKBIET Baramati, Maharashtra, India.

B. E Student, Department of Information Technology, VPKBIET Baramati, Maharashtra, India.

B. E Student, Department of Information Technology, VPKBIET Baramati, Maharashtra, India.

**ABSTRACT:** Today lot of uncertain data is present in various areas. There is need of data classification and analysis. To solve this problem, we introduced big data analytics for uncertain data classification. In this system we will present an approach that will be easy to implement and guarantees to achieve accuracy and execution time. The decision tree and fuzzy decision tree algorithm is used for uncertain data classification. It contains classification and partitioning by using map-reduce. Uncertain Datasets are used as input for classification technique and they generate classified data. Decision trees are widely used to classification and integrating decision trees with fuzzy set theory to deal with uncertainty of big data analytics. Fuzzy classification consists three main steps fuzzification, defuzzification and inference system.

**KEYWORDS:** Decision tree, fuzzy Decision tree, fuzzy logic, classification.

## I. INTRODUCTION

A decision tree is an algorithm used for supervised learning problems such as classification or regression. Classification is the process of organizing data into categories for its most effective and efficient use. It assigns categories to collection of data in order to make it more accurate and effective. Decision tree builds classification model in the form of tree structure and breaks down a dataset into smaller subset and final result is decision nodes. It represents classification or decision. Fuzzy decision tree algorithm is used to classify the uncertain data. It uses fuzzy logic with fuzzy set theory to overcome uncertainties and improve accuracy. Fuzzy logic is many-valued form in which truth values of variable may be in 0 and 1. It uses two techniques first is fuzzification which map mathematical input values into fuzzy membership function and second is de-fuzzification which map fuzzy output into crisp output and then used for decision.

## II. RELATED WORK

In the past years, some decision tree learning algorithms have been proposed for managing big data by adopting the map-reduce paradigm on the Hadoop [7]- [8]. The different algorithms used for generating decision trees and by applying fuzzy logic, try to make it more accurate and efficient. To the best of our knowledge, there do not exist works that have discussed FDT for cloud computing environments taking into account accuracy, complexity and scalability [1]. Machine learning in financial application differs from other domains in how the quality of a model is accessed where in most applications "accuracy of prediction" is often the only metric used. [3]. It also works on scaled dominance approach which is newly adopted to calculate the rule weight that takes the fuzzy frequency of class into account. [3]. It has been introduced in [9]- [10].



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 11, November 2018

## III. METHODOLOGY AND ALGORITHM

Fuzzy stands for uncertain data and fuzzy classification is the process of grouping element into fuzzy set. fuzzy set means unbalanced data elements of that fuzzy set having degree of membership. fuzzy classification corresponds to a membership function  $\mu$  that indicates whether an individual is member of class [3]. Data classification techniques used to find information entropy then fuzzification and defuzzification. It also involves the fuzzy inference system to done the task of map-reduce. Fuzzy decision tree is directed acyclic graph, where each internal node denotes a test on an attribute, each branch represents the outcome of the test and each leaf node holds one or more class labels. [1]

### Algorithm:

**Algorithm:** of FDT learning process [1]

**Require:** training set TR, set X of attribute, splitting method split Met, stopping method stop Met.

```
1. Procedure FDTLearning(in:TR,X,Split Met,Stop Met)
2. root ← create a new node
3. tree ← TREEGROWING(root,TR,X,Split Met,Stop Met)
4. return tree
5. end procedure
6. procedure TreeGrowing(in:node,S,X,Split Met,Stop Met)
7. if STOPMET(node) then
8. node ← mark node as leaf
9. else
10. splits ← SELECTATTRIBUTE(X,S,Splits Met)
11. for each splitz in splits do
12. Sz ← get the set of instances from S determined by splitz
13. childz ← create one node by using splitz and Sz
14. node ← connect the node with TREE GROWING(childz,Sz, Xz, Split Met,Stop Met)
15. end for
16. end if
17. return node
18. end procedure
```

## IV. ARCHITECHTURE

In this section, we introduce architecture for classification process of uncertain data. In that two main steps are defined first is FDT learning and second is Fuzzy partitioning. FDT learning consist approach of learning algorithm and partitioning is a methodology for generating fuzzy sets to represent the underlying data. It follows the entropy based mechanism .fuzzy discritizer is used to generate strong fuzzy partitions.It also introduces membership functionalong with techniques used to process input data that are fuzzification and defuzzification and inference system used to map the input to an output by using fuzzy logic.

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 11, November 2018

System diagram: [6]

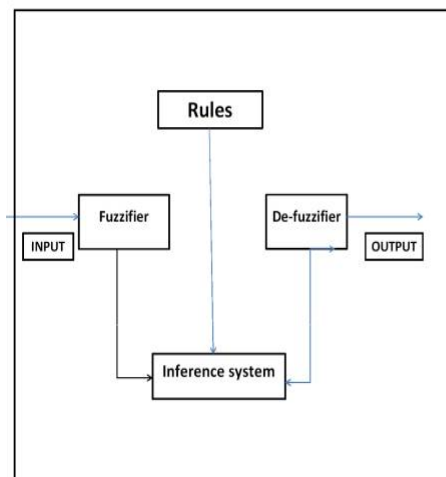


fig no.1.system design[6]

## V. PROCEDURE TO BUILD FUZZY SET

The proposed First consider the feature set suppose  $f=\{1.....F\}$  then define fuzzy membership function " $\mu_{\tilde{A}}(x)$ ". fuzzy sets are value of this membership function. It uses triangular membership function to build fuzzy set. if a, b, c are points then a and c are endpoints and b is core.

Process is start with first fuzzify the input values into fuzzy membership functions. then execute all applicable rules in the rulebase to compute the fuzzy output functions. Defuzzify the fuzzy output function to get "crisp" output value.

## VI. CONCLUSION

In summary, for classification of uncertain data the decision tree and fuzzy decision tree are used in effective way. They are based on Hadoop map reduce functions. fuzzy logic is applied on input data and input is mapped into inference system and then mapped membership function are into crisp output. It provides more accurate and effective data.

## REFERENCES

1. On Distributed Fuzzy Decision Trees for Big Data. Armando Segatori, Francesco Marcelloni , Member IEEE, and Witold Pedrycz , Fellow, IEEE
2. T. Goetz, The Decision Tree: Taking Control of Your Health in the New Era of Personalized Medicine. Emmaus, PA, USA: Rodale, 2010.
3. Multi-Objective Evolutionary Optimization of Type-2 Fuzzy Rule-based Systems for Financial Data Classification .Michela Antonelli, Dario Bernardo, Hani Hagrass, Fellow, IEEE, Francesco Marcelloni, Member, IEEE
4. Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy Department of Information Managementfortune Institute of Technology Kaohsiung,Taiwan
5. A complete fuzzy decision tree technique Cristina Olaru\*, Louis WehenkelUniversity of Liege, Department of Electrical Engineering and Computer Science, Monte\_ore Institute, B28,B-4000, Liege, Belgium Received 5 February 2002; received in revised form 29 January 2003; accepted 18 February 2003.
6. Fuzzy Decision Tree of Risks Assessment Generated From Risk Response Hafeth I. Najil, Rouwaida Hussein Ali2.1,2 Civil Engineering Department, University of Diyala, Province, Iraq.
7. R. Wang, Y.-L. He, C.-Y. Chow, F.-F. Ou, and J. Zhang, "Learning ELMTtree from big data based on uncertainty reduction," *Fuzzy Sets Syst.*, vol. 258, pp. 79–100, 2015.



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 11, November 2018

8. W. Dai and W. Ji, "A MapReduce implementation of C4.5 decision tree algorithm," *Int. J. Database Theory Appl.*, vol. 7, no. 1, pp. 49–60, 2014.
9. Triguero, D. Peralta, J. Bacardit, S. García, and F. Herrera, "MRPR: A MapReduce solution for prototype reduction in big data classification," *Neurocomputing*, vol. 150, pp. 331–345, 2015.
10. Bechini, F. Marcelloni, and A. Segatori, "A MapReduce solution for associative classification of big data," *Inf. Sci.*, vol. 332, pp. 33–55, 2016.