IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 18, Issue 2, Ver. II (Mar-Apr. 2016), PP 37-41
www.iosrjournals.org

Role of Fuzzy Set in Students' Performance Prediction

Mrs. Jyoti Upadhyay[1,] Dr. Pratima Gautam[2]
[1]*Ph.D. (Research Scholar) AISECT University, Bhopal, M.P.*
[2]*Dean (IT Dept.) AISECT University, Bhopal, M.P*

Abstract: *We can use educational data mining to predict student' performance on the basis of different attribute. In this paper, the classification task is used to predict the result of students. Decision tree (DT) learning is one of the most famous practical methods for analytical study. We use DT as prediction method in this paper. In this paper we propose a model using fuzzy set to predict more accurate result.Fuzzy logic brings in an improvement of analysis aspects due to the elasticity of fuzzy sets formalism. Therefore, we proposed a decision tree on fuzzy set data, which combines ID3 with fuzzy theory.The results are compared to some other popular classification algorithms.*

Key words: *Decision tree, Fuzzy set, Data mining*

I. Introduction

1.1 Decision tree: In data mining Decision trees (DT) are among the most popular prediction technique. Although DT's are better known in their role as classifiers, they also have prominent applications in regression, clustering and feature selection.A decision tree presents possible outcomes of a decision through graph. We use specialized software to draw graph, which is rapid miner. DT is useful for focusing discussion when a group must make a decision.

Ross Quinlan developed ID3 (Iterative Dichotomiser 3) in 1986. ID3 algorithm based on top-down search approach on the given data sets. To select the attribute that is most useful for classifying a given sets, use information gain. To find an optimal way to classify a learning set we need some function, which provides the most balanced splitting [1]. A data set contains attributes, we have to select most suitable attribute for the root node of DT. We are using Entropy and information gain. Entropy is the index used to measure degree of impurity [2].

The Entropy is calculated as follows: Entropy=$\sum_j -p \log_2 p\, j$

Splitting criteria used for splitting of nodes of the tree is Information gain. To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S, is defined as:

Gain (S , A) = Entropy (S) $-\sum_{v \epsilon value\ (A)} |sv| / |s|$Entropy ($s_v$)

For the best split point we have to calculate gain for each attribute and repeat the process until we didn't get label attribute.

1.2 fuzzy Set: The selection of the best classification algorithm for a given dataset is a very widespread problem. In this sense it requires to make several methodological choices. Among them, in this research it focuses on the decision tree algorithms from classification methods, which is used to assess the classification performance and to find the best algorithm in obtaining qualitative student data. Fuzzy set theory is also known as possibility theory, was proposed by Lotfi Zadeh in 1965. A researcher seems it as an alternative to traditional theory. Most important, fuzzy set theory allows us to deal with inexact facts. For example, if we says that a student who has 80% is eligible for any course but less % is not considerable then what happen for 79.8%? This can be thought of as an extension of traditional crisp sets, in which each element must either be in or not in a seti.e. Fuzzy set.Fuzzy sets are defined on a non-fuzzy universe of discourse, which is an ordinary sets.Fuzzy logic set can model the normal attributes as linguistic variables (such as good, poor, avg, tall, high) into the inductive generation of the tree structure. Fuzzy decision trees (FDTs) are better suitable to deal with the uncertainty, fuzziness [3].A fuzzy sets F of a universe of discourse U is characterized by a membership function μF (x) which assigns to every elementx ∈ U ,a membership degree μF (x)∈[0,1]. Anelement x∈U is said to be in a fuzzy sets F if and only if μ (x)>0 and to be a full member if and A only if μF (x) =1[4]. Membership functions can either be chosen by the user arbitrarily or based on the user's experience.Typically, a fuzzy subset A can be represented as,

A = {μA (x1) / x1},{μA (x2) / x2 },...,{μA (xn) / xn }

Where the separating symbol / is used to associate the membership value with its coordinate on the horizontal axis. Fuzzy sets and fuzzy logic allow providing a symbolic framework for knowledge clarity. Fuzzy decision trees differ from traditional crisp decision trees in following manner, they use splitting criteria based on fuzzy restrictions and the fuzzy sets representing the defined data.

Fuzzy decision tree required to develop the following things attribute value partitioning methods, branching attribute selection method, branching test method to decide which degree data follows down branches of a node, and leaf node labeling methods to determine classes for which leaf nodes stand[5]. A data set with some condition attributes and one decision attributes can be presented in the form of knowledge representation system J = (U , C ∪D) , whereU ={u1,u2....,us}is the set of data samples, C = {c1 , c2 ...., cn } is the set of condition attributes.D = {d } is the one-elemental set with the decision attribute or class label attribute. Suppose this class label attribute has m distinct values and defining m distinct classes , di (for i=l, ..,m).

For a given subset Sj , information gain s expressed as

$I(s_{1j},\ldots s_{mj}) = -\sum m\, p_{ij} \log_2 p_{ij}$

So information gain of attribute ci is given by

$Gain(ci) = I(s1\,j, \ldots s_{mj}) - E(ci)$

The attribute withhighest information gain is the most informative attribute of the given data set.

II. Methodology

For our work we are using Rapid Miner tool. Which is the most powerful, easy to use graphical user interface for the design of analytic processes[6]. In this paper, we propose a computational model based on fuzzy-rough decision trees to learn the most significant factor for student performance.Our methodology uses fuzzy rough sets to discard irrelevant features on the basis of their dependency. Our work methodology shows in fig.1.We collect data from govt. schools of Chhattisgarh for the session 2014-15.Using Rapidminer we obtained different confidence and support values. Model Evaluation shows accuracy of model

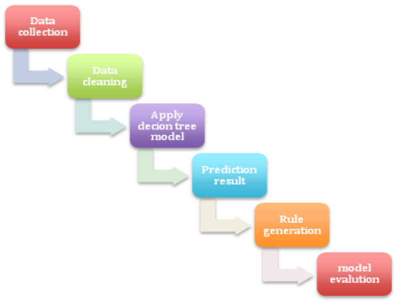

Figure 1: work methodology

III. Proposed work

To implement above methodology we will follow some steps, which are:

3.1 Data Set: For Hypothesis formation we select Educational Environment.For an educational environment it is important to know which factor can affect result of students so that they can emphasis on particular factor and improve successful result and get less drop out ratio. For mining select relevant data shows in table 1.

Attribute	Values
Prev_year %	Result of 12^{th} class
Cast	SC,ST,OBC,GEN
Attendance %	Attendance % in present class
Location	RURAL,URBAN,SEMI-URBAN
Recent Result(label class)	Pass,Fail

We convert data set into Fuzzy Set.

mae of students	Cast	previous %	Attandance	Location	recent result
ımeet Singh	Gen.	48%	78	Urban	fail
veta Prasad	ST	62%	80	Semi-urban	fail
eelam Singh	Gen.	55%	90	Urban	pass
Madhur Kumar	Gen.	40%	60	Rural	
aran Arora	Gen.	61%	80	Urban	pass
kanksha Singh	Gen.	52%	56	Urban	fail
aurav Gayali	Gen.	59%	89	Semi-urban	pass
ıainy Joseph	Gen.	69%	90	Urban	
amal Narayan	SC	61%	91	Rural	fail

Figure 2:Data Set

Fuzzy Set:

Class of object can define Fuzzy set, there is no noisy margins for object [7].A fuzzy set formed by combination of linguistic variable.

Nmae of students	Cast	Location	recent result	Fprevious %	Fattendance
Sumeet Singh	Gen.	Urban	fail	theird	good
Sweta Prasad	ST	Semi-urban	fail	second	good
Neelam Singh	Gen.	Urban	pass	second	good
B.Madhur Kumar	Gen.	Rural		theird	avg
Karan Arora	Gen.	Urban	pass	second	good
Akanksha Singh	Gen.	Urban	fail	second	avg
Gaurav Gayali	Gen.	Semi-urban	pass	second	good
Shainy Joseph	Gen.	Urban		first	good
Kamal Narayan	SC	Rural	fail	second	good

Figure 3: Fuzzy Data Set

3.2 Data Preparation done: The success of data mining techniques depends highly on an appropriate pre-processing of the data. Pre-processing includes data selection, data normalization and transformation. Selection Data selection is critical for the result of a data mining process. Although a relation between a certain attribute and the desired result is not obvious, the attribute has to be considered as well because some information may be hidden in the data.

3.3 Unsuitable Size: Remove inconsistent or remove noisy data and apply treatment about incomplete and erroneous data. Apply Data transformation into Modified Data (after it data transforming into a new format).We apply filter operator of Rapid miner, which takes an Example Set as input and returns a new Example Set including only those examples that satisfy the specified condition. Several predefined conditions are provided; users can select any of them. Users can also define their own conditions to filter examples. This operator may reduce the number of examples in an Example Set but it has no effect on the number of attributes. The select Attributes operator is used to select required attributes.

Figure 4: graphical representation of attributes

3.4 Experiment: We collected qualitative data for experiment and 10-fold cross validation applied. After applying model we evaluate it and generate confusion matrix.

IV. Result and Discussion:

Data set consist 56 students.On the basis of decision tree algorithm fuzzy_attn has the highest information gain, then the decision tree generated as shown in figure:5. We extract Classification rule with the help of decision tree which shows location and cast also play an important role in students' performance.

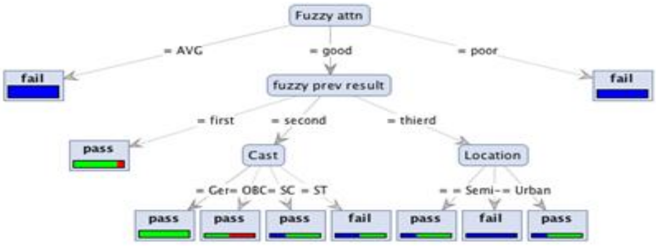

Figure 5:Decision tree

We use the condition attribute Fuzzy_attn, Fuzzy_prev_result, cast and location. Result as class label.
Fuzzy attn = AVG: fail {fail=17, pass=0, Pass=0}
Fuzzy attn = good
fuzzy prev result = first: pass {fail=0, pass=7, Pass=1}
fuzzy prev result = second
Cast = Gen.: pass {fail=0, pass=7, Pass=0}
Cast = OBC: pass {fail=0, pass=1, Pass=1}
Cast = SC: pass {fail=1, pass=2, Pass=0}
Cast = ST: fail {fail=1, pass=1, Pass=0}
fuzzy prev result = thierd
Location = Rural: pass {fail=1, pass=2, Pass=0}
Location = Semi-urban: fail {fail=2, pass=0, Pass=0}
Location = Urban: pass {fail=1, pass=2, Pass=0}
Fuzzy attn = poor: fail {fail=9, pass=0, Pass=0}

Row No.	recent result	prediction(recent result)	confidence(fail)	confidence(pass
.	fail	fail	0.500	0.500
!	fail	fail	0.500	0.500
;	pass	pass	0	1
!	pass	pass	0	1
;	fail	fail	1	0
;	pass	pass	0	1
'	fail	pass	0.333	0.667
;	pass	fail	0.500	0.500
)	pass	pass	0	0.833
.0	pass	pass	0	1
.1	fail	fail	1	0
.2	fail	fail	1	0
.3	fail	fail	1	0
.4	pass	pass	0.333	0.667
.5	fail	fail	1	0
.6	pass	fail	0.500	0.500
.7	fail	fail	1	0
.8	fail	fail	1	0
.9	fail	fail	1	0

Figure 6: Predicted Result

We can interpret from the rule for confidence values are that the students'' performance will be poor if their attendance is poor. Also their location will have an impact on their result.

V. Model Evaluation

Rapid Miner is the most powerful, easy to use and intuitive graphical user interface for the design of analytic processes [8].Classification performance is referred as the characteristics and how successful the models formulated using DT and Fuzzy rough set decision tree learning algorithms are in accurately classifying data points from the testing dataset and/or the independent one-class dataset.The confusion matrix shows the accuracy of the DT for the given data sets. The proposed model was able to classify 82% of the input instances correctly. The results show clearly that the proposed method performs well compared to other similar methods in the literature.

PerformanceVector:
accuracy: 81.50% +/- 10.97% (mikro: 81.63%)

ConfusionMatrix:

True:	fail	pass	Pass
fail:	24	2	1
pass:	5	16	1

kappa: 0.636 +/- 0.207 (mikro: 0.639)

ConfusionMatrix:

True:	fail	pass	Pass
fail:	24	2	1
pass:	5	16	1

Performance Vector for non Fuzzy Set
accuracy: 76.00% +/- 14.97% (mikro: 75.51%)

ConfusionMatrix:

True:	fail	pass	Pass
fail:	25	6	0
pass:	4	12	2

kappa: 0.502 +/- 0.327 (mikro: 0.501)

ConfusionMatrix:

True:	fail	pass	Pass
fail:	25	6	0
pass:	4	12	2

VI. Conclusion

Our work is highly concerned with fuzzy set and decision tree .we propose that numeric data can be represented by fuzzy values. When we apply decision tree model on fuzzy set then it produce more accurate result as we we can see through confusion matrix. Our analysis shows that students' performance can be affected by many factors. In our next paper we will present more attributes with weight.

References

[1]. T.Miranda Lakshmi , A.Martin ,others,An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data,I.J.Modern Education and Computer Science, 5, 18-27, 2013

[2]. Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby,Data Mining: A prediction for Student's Performance Using Classification Method,World Journal of Computer Application and Technology,vol2,43-47, 2014

[3]. Y.-l. Chen, T. Wang, B.-s. Wang, and Z.-j. Li, "A survey of fuzzy decision tree classifier," Fuzzy Information and Engineering, vol. 1, pp. 149–159, June 2009.

[4]. FuzzyLogical Toolbox: zmf. Matlab Help. The MathWorks, Inc. 1984-2004.

[5]. TIEN-CHIN WANG,HSIEN-DA LEE,Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy

[6]. https://rapidminer.com/

[7]. Nasution, H., Design Methodology of Fuzzy Logic Control, UTM, 2002.

[8]. Han J, Rodriguze J, Beheshti M. Diabetes data analysis and prediction model discovery using rapid miner. 2nd International Conference on Future Generation Communication and Networking, IEEE; 2008