IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012
ISSN (Online): 1694-0814
www.IJCSI.org

Fuzzy ID3 Decision Tree Approach for Network Reliability Estimation

Ak. Ashakumar Singh[1], Momtaz Thingujam[2]

[1]Department of Computer Science, Thoubal College, Manipur University,
Manipur-795138, India.

[2] Department of Computer Science, Manipur University, Imphal,
Manipur-795001 India.

Abstract

The Computer communication networks have been rapidly increasing recently to share expensive hardware and software resources, and provide access to main system from distant locations. The reliability and the cost of these systems are important considerations that are largely determined by placement of the nodes and the links between nodes. Network Reliability Estimation is a process of predicting the efforts and cost in terms of money, schedule and staff for any network system. In this study, we investigate the use of Fuzzy ID3 decision tree for network reliability estimation; it is designed by integrating the principles of ID3 decision tree and the fuzzy set-theoretic concepts, enabling the model to handle uncertain and imprecise data when describing the reliabilities of the networks, which can improve greatly the accuracy of obtained estimates. MMRE is used as measures of prediction accuracy for this study. The results are compared with those produced by the crisp version of the ID3 decision tree.

Keywords: *Network reliability, network design, Decision Tree, Fuzzy ID3, Fuzzy Entropy.*

1. Introduction

The design of reliable communication networks is a significant problem in the telecommunications industry. Improving the accuracy of the reliability estimation models available to project managers would facilitate more effective control of time and budgets during network design. An important stage of network design is to find the best layout of components to minimize cost while meeting a performance criterion such as transmission delay, throughput or reliability [1]. Generally, a large scale network has a multilevel and hierarchical structure consisting of a backbone network and several local access networks [2]. This paper is focused on large scale backbone communication network design where the relevant reliability metric is all terminal network reliability [3].

In this paper, we are concerned with network cost estimation models based on fuzzy decision trees especially Fuzzy Interactive Dichotomizer 3.

There are three major advantages when using estimation by decision trees (DT). First, decision trees approach may be considered as "white boxes", it is simple to understand and easy to explain its process to the users, contrary to other learning methods. Second, it allows the learning from previous situations and outcomes. The learning criterion is very important for cost estimation models because network design technology is supposed to be continuously evolving. Third, it may be used to feature subset selection to avoid the problem of cost driver selection in network cost estimation model.

On the other hand, fuzzy logic has been used in network cost estimation. It's based on fuzzy set theory, which was introduced by Zadeh in 1965 [4]. Attempts have been made to rehabilitate some of the existing models in order to handle uncertainties and imprecision problems. The aim of this study is to evaluate and to discuss the use of fuzzy decision trees, especially the fuzzy ID3 algorithm in designing DT for network reliability estimation. Instead of crisp DT, fuzzy DT may allow to exploit complementary advantages of fuzzy logic theory which is the ability to deal with inexact and uncertain information when describing the network designing.

The paper is organized as follows: In section 2, we present the fuzzy ID3 decision tree for network reliability estimation. The description of datasets used to perform the empirical studies and the evaluation criteria adopted to measure the predictive accuracy of the designed models are given in section 3. Section 4 focuses on the experimental design. In Section V, we present and discuss the obtained results when the fuzzy ID3 is used to estimate the network designing. A comparison of the estimation results produced by means of the fuzzy ID3 Models and the crisp ID3 model are also provided in section 5. A conclusion and an overview of future work conclude the paper in the last section 6.

2.Fuzzy ID3 for Network Reliability Estimation

Based on the Concept Learning System algorithm, Quinlan proposed a decision tree called the Interactive Dichotomizer 3 (ID3). The ID3 technique is based on information theory and attempts to minimize the expected number of comparisons.

The fuzzy ID3 is based on a fuzzy implementation of the ID3 algorithm [5,6]. It's formed of one root node, which is the tree top, or starting point, and a series of other nodes. Terminal nodes are leaves (effort). Each node corresponds to a split on the values of one input variable (cost drivers). This variable is chosen in order to reach a maximum of homogeneity amongst the examples that belong to the node, relatively to the output variable.

IJCSI
www.IJCSI.org

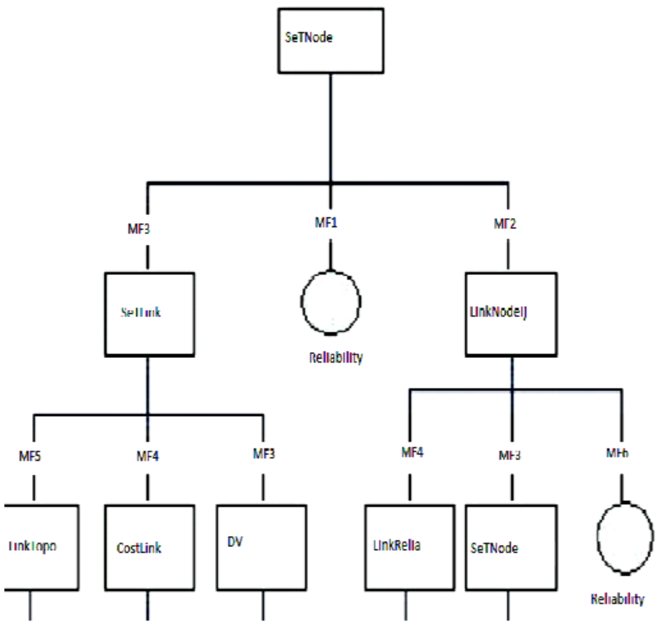


Fig. 1 An example of fuzzy ID3 decision tree for network reliability estimation

.

The major characteristic of fuzzy ID3 is that an example belongs to a node to a certain degree. The proportion of P_k^n of examples with k classification at node n is calculated using the membership degrees as follows:

$$P_n^k = \frac{\sum_{i=1}^{N} u_k(y_i) \wedge u_n(x_i)}{\sum_{c=1}^{k}\sum_{i=1}^{N} u_c(y_i) \wedge u_n(x_i)} \qquad (1)$$

Where K represents the classes and N is the number of examples in the subset. $u_k(y_i)$ is the membership degree on the attribute i that belongs to the class k and $u_n(x_i)$ is the membership degree of the attribute i at node n.

Λ represents the conjunction operator. T-norm, which generalizes intersection in the domain of fuzzy sets, is usually used for fuzzy conjunction. The most popular T- norms are minimum and product. The fuzzy entropy uses the membership degree of examples at a particular node and contributes to enhance the discriminative power of an attribute, is computed as:

$$H_n = -\sum_k p_k^n * \log(p_k^n) \qquad (2)$$

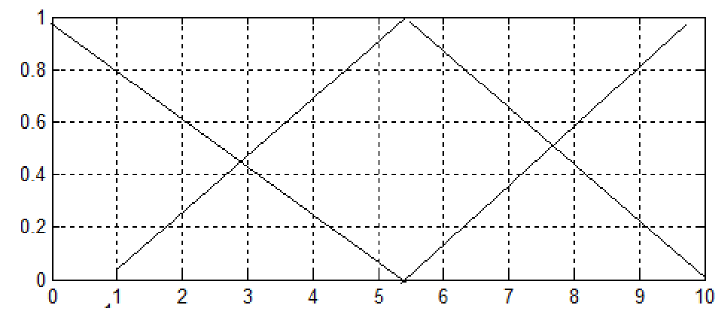
The growth of the fuzzy ID3 is realized by expanding a node of tree characterized by the highest information gain. The information gain is calculated as follows:

$$G_n^j = H_n - \sum_{l=1}^{M} w_l H_l \qquad (3)$$

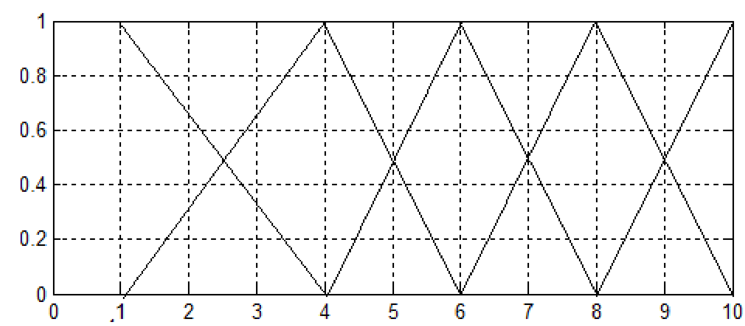
Where H_n is the entropy in the node n. H_l is the entropy of the node that belongs to the fuzzy set L of the j variable. W_l is the fuzzy relative weight.

The node *n* is split into as many sub-nodes as there are attributes. The algorithm terminated when all attributes are used for splits, or when all examples at a node have the same classification.

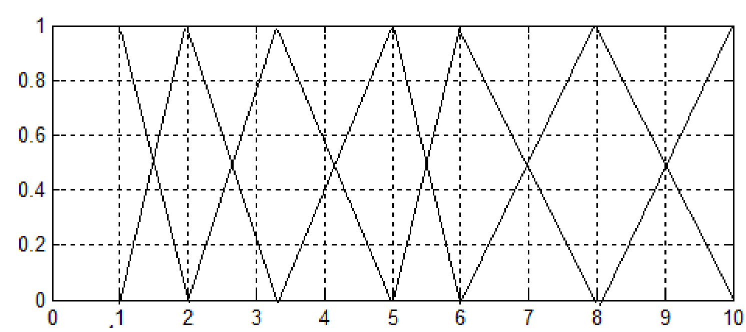
The fuzzification of the network cost drivers converts crisp cost drivers into membership degrees to the different fuzzy sets of the partition. Many algorithms can be found in the specialized literature for generating partitions from data, we choose the hierarchical fuzzy partitioning [7]. It corresponds to an ascending procedure. At each step, for each given variable, two fuzzy sets are merged. This method combines two different clustering techniques, hierarchical clustering and fuzzy clustering techniques. The triangular membership functions are used to represent the fuzzy sets because of its simplicity, easy comprehension and computational efficiency. Figure 2 illustrates the membership function associated to the fuzzy sets of the SeTNode attribute.



(a) Membership function of 3 fuzzy sets defined for the SeTNode cost driver



(b) Membership function of 5 fuzzy sets defined for the SeTNode cost driver



(c) Membership function of 7 fuzzy sets defined for the SeTNode cost driver

Fig. 2 Membership function associated to the fuzzy of the SeTNode attribute.

The fuzzy decision tree is interpreted by rules, each path of the branches from root to leaf can be converted into a rule with condition part represents the attributes on the passing branches from the root to the leaf and condition part represents the class at the leaf of the form: IF (condition 1 and condition 2….. and condition n) THEN C, where the conditions are extracted from the nodes and C is the leaf.

3. Data Description and Evaluation Criteria

3.1 Data Description

This section describes the dataset used to perform this empirical study and the evaluation criteria adopted to measure the estimates accuracy of the designed network cost estimation model based on fuzzy ID3 method.

The dataset is described using 9 numerical attributes as given in table 1.

Table 1: Attributes for network cost estimation

Attributes	Description
SeTNode	Set of nodes (terminal)
SeTLink	Set of links (edge, arcs)
LinkNodeij	Link between nodes i and j
LinkRelia	Link Reliability
DV	Decision variables
LinkTopo	Link Topology
TerRelia	All terminal Reliability
ReliaReq	Network Reliability Requirement
CostLink	Cost of Link

3.2 Evaluation criteria

We employ the following criteria to measure the accuracy of the estimates generated by the fuzzy ID3. A common criterion for the evaluation of reliability estimation models is the magnitude of relative error (MRE), which is defined as

$$MRE = \left| \frac{Reliability_{actual} - Reliability_{estimated}}{Reliability_{actual}} \right| \quad (4)$$

Where $Reliability_{actual}$ is the actual reliability of the network in the dataset and $Reliability_{estimated}$ is the estimated Reliability that was obtained using a model or a technique.

The MRE value are calculated for each attribute in the dataset, while mean magnitude of relative error (MMRE) computes the average over N items

$$MMRE = \frac{1}{N}\sum_{I=1}^{N} \left| \frac{Reliability_{actual,i} - Reliability_{estimated,j}}{Reliability_{actual,i}} \right| \times 100 \quad (5)$$

The acceptable target values for MMRE are MMRE $\leq$ 25. This indicates that on the average, the accuracy of the established estimation model would be less than 25%.

4. Experiment Design

This section describes the experiment design of the fuzzy ID3 decision tree on the given datasets.

The use of fuzzy ID3 to estimate network design effort requires the determination of the parameters, namely the number of input variables, the maximum number of fuzzy sets for each input variable, the significant level value and the conjunction operator. The last two parameters play an essential role in the generation of Fuzzy Decision trees. It greatly affects the calculation of fuzzy entropy and classification results of Fuzzy Decision trees.

The number of input variables is the number of the attributes describing in the used dataset. Therefore, when applying fuzzy ID3 to the given dataset, the number of input variables is equal to 9. Concerning the maximum number of fuzzy sets is the maximum partition size for each variable, is fixed to 7 for all experiments.

In the present paper we are interested in studying the impact of the fuzzy conjunction operators (t-norms) and the significant level parameter (β) on the accuracy of fuzzy ID3. The significant level is the membership degree for an example to be considered as belonging to the node.

For each dataset, two models of fuzzy ID3 were generated. The first Fuzzy ID3 effort estimation model uses the product entropy conjunction operator to measure the fuzzy entropy (t-norm=product), and the second model uses the minimum entropy conjunction operator to calculate the fuzzy entropy t-norm=min). These conjunction operators are the two commonly used t-norm operators because of their well behavior and their computational simplicity [8]

The minimum entropy conjunction operator is defined as:

$$u_k(y_i) \wedge u_n(x_i) = \min[u_k(y_i), u_n(x_i)] \quad (6)$$

Concerning, the product entropy conjunction operator is given as:

$$u_k(y_i) \wedge u_n(x_i) = u_k(y_i) * u_n(x_i) \quad (7)$$

For each model, a series of experiments is conducted with the fuzzy ID3 algorithm each time using a different value of the significant level parameter (β). The significant level is varied within the interval [0, 1].

5. Results

This section presents and discusses the results obtained when applying the fuzzy ID3 to the given datasets. The calculations were made using MatLab software. We conducted several experiments using different configurations of fuzzy ID3. For these experiments, a holdout validation on the entire datasets was performed. Datasets were randomly split into two groups: training set and test set.

Two models of fuzzy ID3 were designed. The first Fuzzy ID3 effort estimation model (Model 1) uses the formula of the conjunction operator given in Eq. (7) to compute fuzzy entropy, and the second model (Model 2) uses the formula of the conjunction operator given in Eq. (6). For each model, different configurations have been obtained by varying the significant level (β). The aim is to determine which configuration improves the estimates.

We have trained and tested the two models using the given dataset. The results for the different configurations have been compared. Figure 3 show the accuracy of the two fuzzy ID3 models, measured in terms of MMRE.

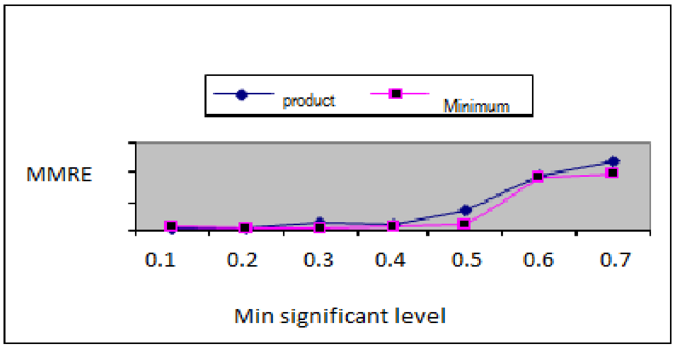


Fig. 3: Relationship between the accuracy of fuzzy ID3(MMRE), the used conjunction operator and the SL value.

Figure 3 compares the accuracy of the two models, in terms of MMRE, when varying the significant level. We note that the fuzzy ID3 model using the product conjunction operator generates a lower MMRE that the other model using the minimum conjunction operator for significant level value less than 0.2.

For example, for β=0.1 the model 1 generates a lower prediction error (MMRE=2.45) than the model 2 (MMRE=5.31). By against, model 2 generates a lower MMRE than the model 1 for significant level value greater than or equal to 0.2. For example, for β=0.5 the model 2 generate a lower prediction error (MMRE=9.09) than the model 1 (MMRE=34.48).

Table 2 summarizes the results obtained using different configurations of fuzzy ID3 for the given dataset. It shows the variation of the accuracy according to the significant level value and to the used conjunction operator.

Table 2: MMRE results of different fuzzy ID3 configurations for the given dataset

Significant level (β)	Accuracy of Fuzzy ID3	
	t-norm=Product	t-norm=Minimum
	MMRE	MMRE
0.1	2.45	5.31
0.2	4.09	1.82
0.3	11.7	3.87
0.4	8.49	5.82
0.5	34.48	9.09
0.6	93.08	90
0.7	119.3	97.41
0.8	210	111.83
0.9	176.99	176.99

Table 3: Result of the different models used on the given dataset

Performance criteria	Crisp ID3	Model 1	Model 2
MMRE	24	2.45	1.82

The experimental results show that the fuzzy ID3 models show better estimation accuracy than the crisp ID3 model in terms of MMRE.

6. Conclusions

In the paper, we have empirically studied two fuzzy ID3 models for network reliability estimation. Each one used a different formula to compute the fuzzy entropy. These fuzzy ID3 models were trained and tested using the given datasets. The results show that the use of an optimal significant level value and an adequate conjunction operator for computing the fuzzy entropy improves greatly the estimates generated by fuzzy ID3 model. The comparison with the crisp version of ID3 decision tree shows encouraging results.

References

[1] R.H.Jan, F.J.Hwang, and S.T.Chen, “Topological optimization of a communication network subject to a reliability constraint”, *IEEE Transactions on Reliability*, Vol.42(1993), 63-70

[2] R.R.Boorstyn and H.Frank, “Large scale network topological optimization”, IEEE Transactions on communication, Com-25(1977), 29-7

[3] C.J.Colbourn, *The Combinatories of Network Reliability*, Oxford University Press(1987).

[4] L.A.Zadeh, “Fuzzy sets”, Information and Control, Vol. 8, 1965,338-353.

[5] M.Umano, H.Okamoto, I.Hatono, H.Tamura, F.Kawachi, S.Umedzu, J.Kinoshita, “Fuzzy Decision Trees by Fuzzy ID3 algorithm and its Applications to Diagnosis System”, In

proceedings of the third IEEE Conference on Fuzzy Systems, vol. 3, Orlando, 1994, 2113-2118.

[6] R.Weber, Fuzzy ID3: a class of methods for automatic knowledge acquisition, proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, July, 17-22, 1992, 265-268

[7] S.Guillaume and B.Charnomordic, "Generating an interpretable family of Fuzzy partitions", IEEE Transactions on Fuzzy Systems, 12(3), June 2004. 324-335.

[8] E.H.Mamdani, "Applications of Fuzzy Set Theory to Control Systems: A survey", in Fuzzy Automata and Decision Processes, M.M.Gupta, G.N.Saridis and B.R.Gaines, eds., North-Holland, New York, 1977, 1-13.

Biography:

Ak. Ashakumar Singh graduated in Mathematics from Manipur University, Imphal and passed MCA in the year 2000 from the same varsity. He was awarded Ph.D. in the area of Computer Science from the Dept. of Mathematics of the same varsity in the year 2008. Then produced eight M.Phil scholars in Computer Science and now supervising three scholars leading to Ph.D. in *Computer* Science. The area of research is on Soft computing and related applications of computer science.

Ms. Momtaz Thingujam had received her B.Tech degree (C.Sc.) from IET , Lucknow U.P(INDIA) and M.Tech degree from AAIDU, Allahabad U.P (INDIA) respectively.
She is currently an Assistant Professor (Senior Scale) in Computer Science Department, Manipur University. Previously she was a Programmer in Department of Science and Technology, Manipur and Computer Center, Manipur University. Her primary research interests are Natural Language Processing and Computer Network. She also worked as a Co-Investigator for a project (Development of Manipuri language Technology) sponsored by Department of Information Technology, New Delhi.