Optimized Fuzzy Decision Tree Data Mining for Engineering Applications

Liam Evans and Niels Lohse

Faculty of Engineering, University of Nottingham, Nottingham,
England, NG7 2RD, United Kingdom
{epxle1,niels.lohse}@nottingham.ac.uk

Abstract. Manufacturing organizations are striving to remain competitive in an era of increased competition and every-changing conditions. Manufacturing technology selection is a key factor in the growth of an organization and a fundamental challenge is effectively managing the computation of data to support future decision-making. Classification is a data mining technique used to predict group membership for data instances. Popular methods include decision trees and neural networks. This paper investigates a unique fuzzy reasoning method suited to engineering applications using fuzzy decision trees.

The paper focuses on the inference stages of fuzzy decision trees to support decision-engineering tasks. The relaxation of crisp decision tree boundaries through fuzzy principles increases the importance of the degree of confidence exhibited by the inference mechanism. Industrial philosophies have a strong influence on decision practices and such strategic views must be considered. The paper is organized as follows: introduction to the research area, literature review, proposed inference mechanism and numerical example. The research is concluded and future work discussed.

Keywords: Fuzzy Decision Tree (FDT), Classification and Prediction, Knowledge Management, Manufacturing Technology Selection, Intelligent Decision-Making.

1 Introduction

Decision-making in the manufacturing sector is a complex and imperative practice that requires accurate judgment and precise classification. It consists of the wide evaluation of alternatives options against an intolerable set of conflicting criteria. The rapid development of available technologies and complexity manufacturing technologies offer has made the task of technology selection difficult. Rao [1] notes how manufacturing technologies have continually gone through gradual and sometimes revolutionary changes. Fast changing technologies on the product front cautioned the need for an equally fast response from the manufacturing industries. To meet the challenges, manufacturing industrials have to select appropriate strategies, product designs, processes, work piece and tool materials, machinery and equipment, *etc*. The selection decisions are complex, as decision-making is more challenging today.

P. Perner (Ed.): ICDM 2011, LNAI 6870, pp. 228–239, 2011.

Classification problems have aroused interest of many researchers in recent years. In general, a classification problem is to assign certain membership classes to objects (events, phenomena), described by a set of attributes. In practice, classification algorithms involve obtaining some data on input and putting appropriate classes on output, mostly assuming a given object attribute and class set [2]. Fuzzy Decision Trees (FDTs) are a form of induced decision trees that combine the theory of fuzzy to soften sharp decision boundaries, which are inherent in traditional decision tree algorithms. A fuzzy region represents each node in the decision tree and the firing to some degree of each node forms the inference technique to produce the final classification.

FDTs are an effective data-mining technique that support classification based on historical data through a case repository of previous decisions. Knowledge acquisition is regarded as the bottleneck of expert system development in the artificial intelligence field. Knowledge is difficult to capture and express, it is also extensive and costly to conduct. Human experts may be able to master their respective task, but unable to communicate such activities into an intelligent system. Capturing knowledge through historical cases is potentially a suitable and easier to conduct task. Initial studies suggest that previous evaluated technologies stored in the form of cases can enable quick classifications based on new project requirements by adopting the FDT technique.

Each node in a FDT is represented by a fuzzy set, itself defined by a fuzzy membership function. An unclassified example, based on the input of fuzzy requirements, pass through the tree and result in all branches firing to some degree. It is common for membership grades throughout the tree to be combined using pre-selected inference techniques to produce an overall classification. Shortcomings of existing techniques are the lack of consideration for the value of attributes that can account for changes in the expected classification. In addition, summing the respective values is not appropriate for FDTs. When performing the selection process for engineering domains, certain factors are deemed essential for the validity of choice and should have a bearing on the outcome, which in turn relates the organizational strategy and vision.

This paper presents a discussion of inference techniques to support fuzzy decision tree classification. The paper notes on the unique factors of engineering applications and draws on key challenges essential to the reasoning algorithm. A unique inference mechanism is proposed in section three and section four provides a numerical example for further clarity. Finally, the research is concluded and future research discussed in section five.

2 Literature Review

In our daily life we always face situations where we have to make decisions. We often use our past experience to decide on current events, where experience can be thought of as experimental data. Some applications are very complex such that it is very hard for us to deduce good decision models based on our experience. Furthermore, experimental approaches to decide on new cases may be too expensive and time-consuming. Machine learning represents an efficient and automated approach to construct decision models from previously collected data (known cases) and apply the constructed models to unknown, similar cases to make a decision [3]. Machine learning has received extensive interest from researchers in classification and prediction problems. Many have been successfully applied and bring improvements compared with existing decision support practices.

Anand and Buchner [4] define data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets. It is an extremely useful theoretical application and broad ranges of techniques applied to problems exist. The aim is to identify useful patterns within a dataset to predict suitable outcomes for decision makers. Decision trees are one of the most popular machine-learning techniques [5]; they are praised for their ability to represent the decision support information in a human comprehensible form [5, 6]. However, they are recognized as a highly unstable classifier with respect to small changes in training data [3, 7]. Decision tree rule induction is a method to construct a set of rules that classify objects from knowledge of a training set of examples, whose classes are previously known. The process of classification can be defined as the task of discovering rules or patterns from a set of data. The objectives of any classification task is to at least equal and essentially exceed a human decision maker in a consistent and practical manner [8].

A fundamental problem associated with decision trees to support classification is the sharp boundaries that exist in separating the attributes within the tree. The partitioning is strict and small changes can lead to different classifications being sought. To overcome some of the deficiencies of crisp decision trees, the relaxation of these boundaries can be achieved through the creation of fuzzy regions at each node. Unknown cases travel through all paths with a certain degree of confidence, instead of maintaining one definite path. The degree of confidence exhibited by a specific attribute value is determined by a fuzzy membership grade.

Janikow [5] best summaries and describes the four steps of a fuzzy decision tree induction mechanism:

1. Data fuzzification.
2. Building a fuzzy decision tree.
3. Converting the fuzzy decision tree into a set of fuzzy rules.
4. Applying the fuzzy rules to make classification and/or prediction (inference).

Data fuzzification is applied to numerical data. The purpose is to reduce the information overload in the decision support process. Fuzzy membership functions are selected to represent the partitioning attributes and are crucial to the performance of fuzzy decision trees. The tree building procedure recursively partitions the training dataset based on the value of a select splitting feature. Several information measures exist in the literature with the purpose of identifying influential branching features. A node in the tree is considered a leaf node, when all the objects at the node belong to the same class, the number of objects in the node is less than a certain threshold, the ratio between objects membership in different classes is greater than a given threshold, or no more features are available.

The building procedure will often initiate with the most informative attribute beginning the initial splitting and continuing with the second most instructive, *etc*. This continues till all objects are classified within the data set. If the dataset were extensive, unique classifications would create a large tree. Pruning can examine the performance of a particular branch within the tree to decide whether or not to stop the growth down that specific branch, reducing the size of the tree. The final stage of an induction mechanism is the inference procedure.

Inference has long been a method for reasoning and thus deducing an outcome from a set of facts. The technique involves combining the mathematical information generated from firing a number of IF-THEN rules from a knowledge base. The knowledge base consists of a series of fuzzy IF-THEN rules extracted from each path of the tree. Keeley [8] discusses the technique in four stages: (i) Combining the information of the antecedent of a particular rule, (ii) Applying the resultant value to the consequence of that particular rule, (iii) Combing the resultants from all rules, (iv) interpreting the outcome.

The rules in fuzzy decision trees are fuzzy rather than crisp, and therefore have antecedents, consequences, or both. The chosen fuzzy inference paradigm is applied to combine the information generated from firing the rules, and produce a fuzzy set of fuzzy value outcome [8]. Typical decision tree IF-THEN rules produce a singleton as the outcome; however, fuzzy models usually produce a fuzzy region.

The latest study of fuzzy inference reasoning mechanisms suitable for decision tree rules require a singleton output as discussed by Abu-halaweh [3]. As the test object falls down the numerous paths and through each attribute within the tree, a level of certainty can be concluded at each partitioning point. The first method discussed by Abu-halaweh [3] corresponds to labeling the leaf node with the class that has the greatest membership value, whilst the second labels the leaf node with all class names along with their membership values

In the first method, as the object propagates down the fuzzy decision tree, its membership value in all of the decision leaf nodes is calculated. Then the object is assigned the class label of the leaf node that has the greatest membership value. In other words, it is assigned the same label of the fuzzy rule with the maximum firing strength (max-min). In the second method, it will reach each leaf node with some certainty or membership value. However, since the leaf nodes are labeled with all class names and their membership values, the class proportion in the leaf node multiplies the certainties. Then the certainties of each class are summed, and the test object is assigned the class label with the greatest certainty [3].

In terms of manufacturing technology selection, data mining has been identified as a potential key factor that can support manufacturing decision-making practices. Harding et al [9] recognized that knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors, and to compete efficiently and effectively to the best of its ability. Data mining for manufacturing began in the 1990s [10-12] and it has gradually progressed by receiving attention from the production community. Data mining is now used in many different areas of manufacturing to extract knowledge for use in predictive maintenance, fault detection, design, production, quality assurance, scheduling, and decision support systems [9]

Shortcomings of the two noted reasoning mechanisms are firstly that objects can be classified as unsuitable solutions based on a single membership value within a path that has received the highest fuzzy membership grade. An unknown object may then not adhere to requirements and be incorrectly classified. Secondly, the attribute splitting points within a tree often signify different levels of importance related to activities such as corporate vision and strategy. It is possible that low importance factors placed high within the decision tree can classify decisions that do not meet the appropriate requirements. Finally, the summation of each membership value, independent of weighting by proportion, can classify solutions that contain longer paths to received higher scores and therefore be recommended for selection. It is likely that longer paths with smaller values will be classified compared to shorter paths that have higher values.

To overcome these shortcomings, section three of this paper describes a new analytical methodology that aims to generate a dependable and flexible fuzzy reasoning mechanism suitable for engineering applications where the impact of criteria weighting is of paramount importance.

3 Proposed Inference Mechanism

In a fuzzy rule-based classification system, two main components can be recognized: 1) the Knowledge Base (KB), composed of a Rule Base (RB) and a Data Base (DB), which is specific for a given classification problem, and 2) a fuzzy rule-based reasoning mechanism. The classification system coherently combines both components that start with a set of correctly classified examples (historical case examples). The aim is to assign class labels to new examples with minimum error and acceptable similarity. This process is described in Figure 1 and the detailed structure of the components is discussed in the following subsections.

A fuzzy reasoning method is an inference procedure that derives a set of conclusions from a fuzzy rule set and a case example. The method combines the information of the rules fired with the pattern to be classified for an unknown case. This model is described in the following.

Knowledge Base

a) **Extracted rules** from the fuzzy decision tree form an IF-THEN rule base and a coterie of fuzzy sets. Each rule is extracted and varies in length depending upon the purity of the decision tree. They are formed as:

"IF a set of conditions are satisfied, AND a different set of conditions are satisfied, THEN a set of consequences is deduced".

$$\mathbf{R_{k1\ldots kn}:\ IF\ } x_{1\ldots n} \mathbf{\ is\ } Ax_{1\ldots n} \mathbf{\ AND\ } x_{n} \mathbf{\ is\ } Ax_{n} \mathbf{\ THEN\ \ solution\ \ is\ } L_{j1\ldots jn}$$

Where:

$R_{k1\ldots kn}$	is a rule with a unique case identified number
$x_{1\ldots n}$	is an attribute in the tree
$Ax_{1\ldots n}$	is the rating of the attribute $x_{1\ldots n}$
$L_{j1\ldots jn}$	is the case solution (end leaf)

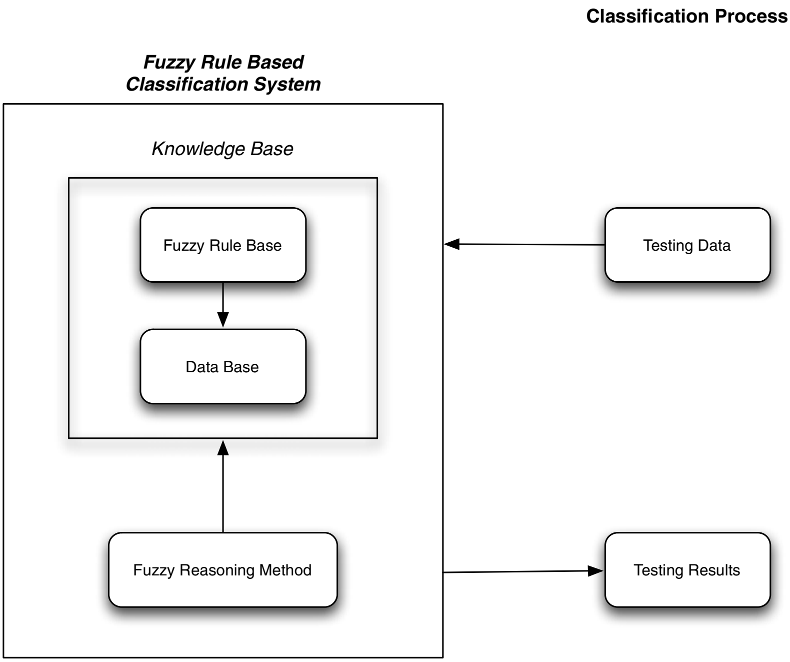

Fig. 1. Design of a Fuzzy Rule-based Classification System

Fuzzy Reasoning Method

b) **Weighting of parameters** enables a level of quantitative property to be assigned to each splitting attribute within the tree. In engineering applications, different attributes have alternative levels of importance that can affect the expected outcome. An appropriate process of identifying quantitative scoring for alternative parameters is to use the pair-wise comparison technique. Each attribute is considered and decision makers express their preference between two mutually distinct alternatives. For example, if the alternatives are Attribute1 and Attribute2, the following are the possible pairwise comparisons.

Attribute1 is preferred over Attribute2:	"Att.1 > Att.2"
Attribute2 is preferred over Attribute1:	"Att.2 > Att.1"
Preference is indifferent between both alternatives:	"Att.1 = Att.2"

To calculate the final scoring, a normalized quantitative property is determined for each of the alternatives within the comparison table. Each attribute weight is calculated and expressed $wx_{n.}$ as a percentage.

c) **Probability** is an important technique for decision analysis where the level of certainty can play a role in classification. The probability is shown at each attribute within the decision tree and identifies the amount of objects that lie below that particular attribute. Probability can provide an insight into the strength of a solution appearing within a rule when a class object is repeated on a number of occasions. The advantage of incorporating probability into the reasoning algorithm is the ability to consider a single path on a number of occasions that may contain more than one identical final object.

The probability of an object relating to an attribute is shown as:

$$P_x = \textit{probability} \text{ of } L_{j1\ldots jn} \text{ within } x_{1\ldots n} \text{ for } Ax_{1\ldots n}$$

d) **New object classification** allows the requirements of an unknown case to be classified. Starting at the root node, the tested object is defined in fuzzy terms and expressed as an optimal position within a fuzzy membership set. The input is expressed for each criteria rating and the output fuzzy membership value is allocated for each of the fuzzy functions within the membership set. The output fuzzy membership value is expressed as F_{MV} for each attribute partition.

For each attribute partition, the selected position along the fuzzy membership set forms a numerical output value as shown for each function in Figure 2. The input value of 3.5 concludes a score of 0.8 and 0.2 for the linguistic terms ‘ManyConstraints’ and ‘PossibleConstraints’.

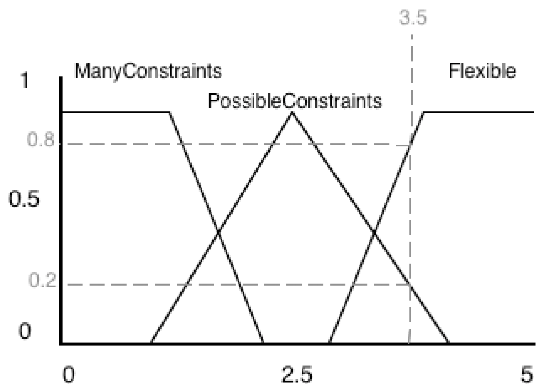

Fig. 2. Fuzzy Membership Function Example

e) **The weighted fuzzy membership value (wMV)** is calculated at each attribute within the decision tree to conclude interim scores for use in the final calculation of a rule. The weight fuzzy membership value identifies the probability of the final leaf object appearing in a particular attribute. An attribute information set is shown as:

Attribute x_n		
Category	(%)	n
Alternative L_1	50	25
Alternative L_2	20	10
Alternative L_n	20	10
Probability (P_x)	x	x
		$wx_n = x$

The final weighted attribute expressed as:

$$wMV = P_x \text{ x } wx_n \text{ x } F_{MV}.$$

f) End leaf calculation for rule class object. For each rule, as the object propagates down the tree, the *wMV* is noted at each of the decision tree nodes. The number of noted *wMV* depends entirely on the length of the rule and the average is calculated by summing each *wMV* and dividing it by the number of *wMV*. The average is expressed as:

$$\text{Score } R_{\text{kn}} = \frac{1}{n}\sum_{i=1}^{n} wMV_n$$

Where:

n = number of *wMV* within the rule.

By calculating the average score reflects that the length of a rule may vary. Long and short rules are deemed equal and each leaf node is given an equal opportunity for identifying a similarity score. The consideration of weighting each attribute replicates human reasoning where particular attributes are deemed more or less important, and affects the final result dependent on the level of importance.

g) Summary. The final phase is to summaries the results of the calculated scores for each rule. Firstly, for rules that contain the same object class name (i.e. the solutions are identical), the maximum rule score will represent that object. The object classification that received the highest score is deemed to be the most appropriate and a suitable solution based on the new project input requirements. Therefore, the classification is the highest scoring solution.

To conclude the proposed inference mechanism, the seven stages aim to provide a methodical approach that is considerate to the value of alternative attributes and the form in which the fuzzy decision tree generates the rule base for engineering applications. This paper will now present a brief numerical example to illustrate the approach.

4 Numerical Example

In this section, we present a numerical example to demonstrate the applicability of the approach within its intended domain. Using a fictitious dataset, a case repository of twenty cases containing four alternative class objects was applied. The dataset was fuzzfied and contained seven attributes. Each case contained a unique identification number and linguistic term to represent the performance of the class label within the case. The case repository is shown in Table 1.

Table 1. Case Repository

	Technical					Financial	Strategy	
	TL	PT	SL	CM	SC	PC	MO	
Case Number	**Technological Longevity**	**Process Time**	**Skill Level / Training**	**Change Management**	**Supply Chain Management**	**Project Cost**	**Long Term Manufacturing Objectives**	**Solution**
002	High	Average	Semi/Training	Acceptable	Improved Chain	Average	Inline w/Obj'	Fixed Tooling
003	Low	Low	Unskilled	Acceptable	Acceptable	Low Cost	Non-related	Fixed Tooling
004	High	Very Low	Semi/Training	Good	Improved Chain	Rel Low Cost	Inline w/Obj'	Laser Scanner
006	Very Low	Very Low	Unskilled	Good	Acceptable	Low Cost	Non-related	Fixed Tooling
007	High	Very Low	Semi/Training	Good	Improved Chain	Average	Inline w/Obj'	Photogrammetry
008	Medium	Low	Skilled	Acceptable	Potential Issues	Average	Partially	Robot
010	Low	Low	Skilled	Unmanageable	Potential Issues	High Cost	Partially	Robot
011	Medium	Low	Semi/Training	Acceptable	Acceptable	Rel Low Cost	Partially	Fixed Tooling
012	High	Low	Semi/Training	Good	Acceptable	Low Cost	Inline w/Obj'	Laser Scanner
013	Very High	High	Skilled	Acceptable	Improved Chain	Low Cost	Inline w/Obj'	Laser Scanner
014	Very Low	Average	Unskilled	Acceptable	Acceptable	Low Cost	Inline w/Obj'	Photogrammetry
015	Very Low	High	Semi/Training	Acceptable	Acceptable	Low Cost	Non-related	Photogrammetry
016	Very Low	Very High	Skilled	Good	Improved Chain	Low Cost	Inline w/Obj'	Robot
017	High	Low	Skilled	Good	Potential Issues	High Cost	Partially	Fixed Tooling
018	Medium	Low	Unskilled	Good	Potential Issues	Average	Inline w/Obj'	Fixed Tooling
019	Low	Very Low	Semi/Training	Acceptable	Acceptable	Low Cost	Non-related	Fixed Tooling
020	High	Average	Skilled	Unmanageable	Acceptable	Low Cost	Inline w/Obj'	Laser Scanner
021	Low	Low	Semi/Training	Acceptable	Improved Chain	Low Cost	Inline w/Obj'	Laser Scanner
022	High	Very Low	Skilled	Good	Acceptable	High Cost	Non-related	Robot
023	Medium	Low	Skilled	Acceptable	Improved Chain	Low Cost	Inline w/Obj'	Robot

From the dataset, we form a knowledge representation system for $J = (U, C \cup D)$ where: $U = \{1,\ldots.20\}$, C = {TL, PT, SL, CM, SC, PC, MO}, D = {Fixed_Tooling, Laser_Scanner, Photogrammetry, Robot}. Using the fuzzy decision tree building procedure proposed by Wang and Lee [13], the following information gain scores were concluded for each of the attributes:

Gain (SL) = 0.6344, Gain (TL) = 0.4958, Gain (PT) = 0.4336, Gain (CM) = 0.1174, Gain (SC) = 0.2693, Gain (PC) = 0.3379, and Gain (MO) = 0.3757.

Since skill level received the highest information gain among the seven attributes, it is selected as the initial partitioning of the tree and placed at the top. Upon initial splitting of the tree, it became apparent that skill level does not uniquely classify each alternative; the tree is not pure. We therefore select the second highest attribute to

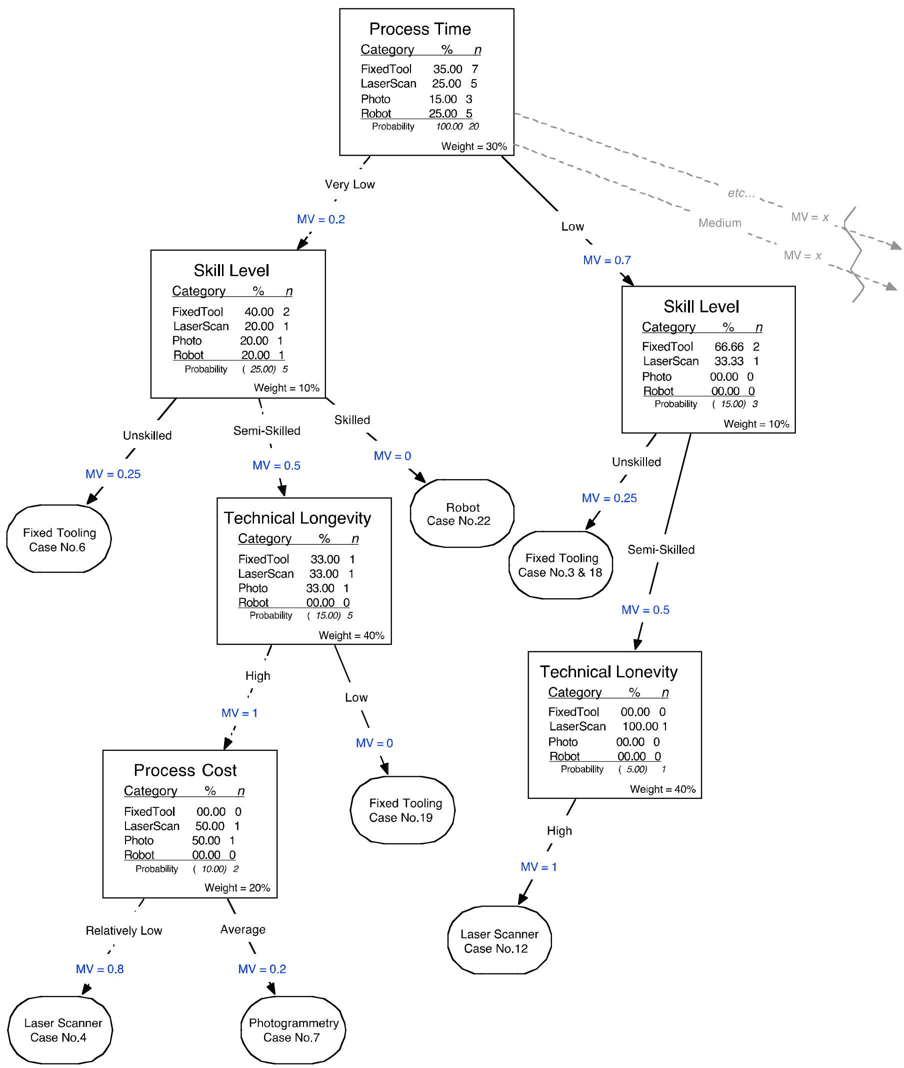

Fig. 3. Partial Fuzzy Decision Tree

split the tree further until it is fully classified. We present a partial section of the fuzzy decision tree in Figure 3 to demonstrate the fuzzy splitting of data. In addition, the figure represents the output fuzzy membership values for each of the attributes for the new project classification. These have been selected as optimal positions within each fuzzy set in order to classify the new example.

In order to calculate the resultant score for each of the end nodes, we follow the equation in step five of the methodology. For example, we will demonstrate using the dashed rule for 'Laser Scanner, Case No.4'.

R$_{k4}$: IF *ProcessTime* is *VeryLow* **AND** *SkillLevel* is *SemiSkilled* **AND** *TechnicalLongevity* is High **AND** *ProcessCost* is *RelativelyLow* **THEN** Solution is *LaserScanner*

$$Rule_{LaserScanner\ Case\ No.4} =$$

$$\frac{(100\% \times 30\% \times 0.2) + (25\% \times 10\% \times 0.5) + (40\% \times 15\% \times 1) + (20\% \times 10\% \times 0.8)}{4} = 0.03715$$

If we follow the same algorithm for the eight different class objects within the tree, we can conclude that Fixed Tooling Case No.3 & 18 received the highest similarity score and therefore is the classification object. The case is then represented as a new case in the repository and stored for future use. As the decision maker wishes to determine the appropriate classification result among the objects within the repository, the highest scoring technology is deemed appropriate and a ranking of the solutions is not shown.

5 Conclusion

In this paper, we have proposed a fuzzy reasoning method for fuzzy decision tree inference of engineering applications. The approach considers each rule within a tree independent of the length determined by the tree builder algorithm. Although most fuzzy reasoning using IF-THEN rules determine a fuzzy region as the output, fuzzy decision tree rules consist of fuzzy partitioning at each attribute and not for the end nodes. Therefore the rules are multiple input, single output equations that output a numerical score. The output rule receiving the highest score is deemed the most suitable classification.

Existing publications tend to lack consideration for the value of attributes, which relate directly to an organization and have an influence on the outcome. The well-publicized max-min method identifies the weakest membership function in a rule and uses that value to represent the object class. Identifying the lowest score is not ideal because a rule may be well represented by other attributes. The methodology proposed in this paper combines the importance of different attribute values by determining a normalized level of importance through the pair-wise comparison technique. The average fuzzy membership value of each rule is calculated to act as the final object class to consider stronger and weak similarity scores. To conclude, the proposed model is deemed as more effective compared with existing algorithms and well suited to applications where levels of importance can change over time to allow the decision-maker to input different requirements.

The work described in this paper is part of a research project that is investigating how fuzzy decision trees can support manufacturing technology selection within the engineering domain. Future work will investigate the effectiveness of the proposed approach in a corporate environment for comparison with existing practices.

Acknowledgments. The authors gratefully acknowledge the support provided by the Innovative Manufacturing Research Centre (IMRC) at the University of Nottingham, funded by the Engineering and Physical Science Research Council (EPSRC) under Grant No. EP/E001904/1 for the research reported in this paper.

References

1. Rao, V.: Decision making in the manufacturing environment using an improved PROMETHEE method. International Journal of Production Research 48(16), 4665–4682 (2010)
2. Chang, P.C., Fan, C.Y., Dzan, W.Y.: A CBR-based fuzzy decision tree approach for database classification. Expert Systems with Applications 37(1), 214–225 (2010)
3. Abu-halaweh, N.M.: Integrating Information Theory Measures and a Novel Rule-Set-Reduction Technique to Improve Fuzzy Decision Tree Induction Algorithms, Deptartment of Computer Science. Georgia State University, Atlanta (2009)
4. Anand, S.S., Buchner, A.G.: Decision support using data mining. Financial Times Management, London (1998)
5. Janikow, C.: Fuzzy decision trees: issues and methods. IEEE Transactions on Systems, Man, and Cybernetics, Part B 28(1), 1 (1998)
6. Quinlan, J.: Decision trees and decision-making. IEEE Transactions on Systems, Man and Cybernetics 20(2), 339–346 (1990)
7. Olaru, C., Wehenkel, L.: A complete fuzzy decision tree technique. Fuzzy Sets and Systems 138(2), 221–254 (2003)
8. Crockett, K.A.: Fuzzy Rule Induction from Data Domains, Department of Computing and Mathematics, p. 219. Manchester Metropolitan University, Manchester (1998)
9. Harding, J., Shahbaz, M., Kusiak, A.: Data mining in manufacturing: a review. Journal of Manufacturing Science and Engineering 128, 969 (2006)
10. Lee, M.: The knowledge-based factory. Artificial Intelligence in Engineering 8(2), 109–125 (1993)
11. Irani, K.B., et al.: Applying machine learning to semiconductor manufacturing. IEEE Expert Intelligent Systems and their Applications, 41–47 (1993)
12. Piatetsky-Shapiro, G.: The data-mining industry coming of age. IEEE Intelligent Systems and their Applications 14(6), 32–34 (1999)
13. Wang, T.C., Lee, H.D.: Constructing a fuzzy decision tree by integrating fuzzy sets and entropy. WSEAS Transactions on Information Science and Applications 3(8), 1547–1552 (2006)