ISSN 2320 – 2602
Volume 3, No.8, August 2014
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2014/ijacst06382014.pdf

# Optimized Fuzzy Decision Tree for Structured Continuous-Label Classification

**K.Jeyalakshmi[1], J.Jenifer[2]**

[1]Assistant Professor, Pg and Research Department of Computer Science, Hindustan college of Arts and Science, Bharathiyar University,Coimbatore, India , Email: jay_lakme@yahoo.com

[2]Research Scholar, Pg and Research Department of Computer Science, Hindusthan college of Arts and Science,BharathiyarUniversity,Coimbatore,India,Email:jeniferbtw@gmail.com.

## ABSTRACT

Mainly understandable decision trees have been intended for perfect symbolic data. Conventional crisp decision trees (DT) are extensively used for classification purpose. However, there are still many issues particularly when we used the numerical (continuous valued) attributes. Structured continuous-label classification is one type of classification in which the label is continuous in the data. Although, the main aim is to classify data into classes that are a set of predefined ranges and can be ordered in a hierarchy. Some of those difficulties can be addressed using fuzzy decision trees (FDT). Over the years, additional methodologies have been investigated and proposed to deal with continuous or multi-valued data, and with missing or noisy features. In recent times, with the growing popularity of fuzzy representation, a few researchers independently have proposed to utilize fuzzy representation in decision trees to deal with similar situations. Fuzzy concept connects the gap between symbolic and non symbolic data by using linking qualitative linguistic terms with quantitative data. In our proposed work, a new method of optimized fuzzy decision trees is presented. This method proposed a novel decision tree technique for handling continuous valued attributes with user defined membership. The efficient results of our proposed fuzzy decision trees are compared at the end in the experimentation. From the experimentation, we are conclude that the proposed fuzzy decision trees is well effective than the existing system interms of accuracy rate as well as reduce the error performance of the system.

**Keywords:** Continuous-label classification, discretization method, fuzzy decision trees algorithm, information gain and decision tree (DT).

## 1. INTRODUCTION

Classification is one of the most common tasks in data mining to solve a wide range of real problems, such as credit scoring [1], bankruptcy prediction [2], [3], medical diagnosis [4], pattern recognition [5], and multimedia classification [6], [7]. It is recognized as a powerful way for companies to develop effective knowledge-based decision models to gain competitive advantages [8]. The current classification methods include decision trees (DTs), neural networks, logistic regression, and nearest neighbor. DTs are widely used for classification because of their ability to manage noisy data [9] and return well-organized results that are easily interpreted, computationally efficient, and robust. Several algorithms, such as ID3 [10] and C4.5 [11], have been devised for the construction of DTs and are applied in several areas, including risk management [12], customer relationship management [13], text categorization [14], [15], medical diagnosis [16], and credit rating of loan applicants [17]–[19].

In traditional DT algorithms, a label (target variable) of a tuple is either categorical or Boolean; that is, the algorithms operate under the assumption that the labels are nominal and flat. However, several practical situations involve more complex classification scenarios, in which the label to be predicted can occur in a variety of types, for example, continuous variable, hierarchically related variable, or both. To bridge this gap in the research, Hu *et al.* [20] proposed a method called the continuous label classifier (CLC) algorithm, which added a novel dynamic partition mechanism to partition continuous labels during the construction of a DT. For hierarchy-related problems with class labels, Chen *et al.* [21] proposed a hierarchical class label classifier (HLC) algorithm that can construct a DT from data with hierarchical class labels. They alsodemonstrated the unsuitability of traditional methods for hierarchicalclass labels and showed the differences resulting from their use. The HLC [21] and CLC [20] algorithms address several crucial problems associated with the induction of DT from data with either continuous or hierarchical labels. However, both algorithms have the following substantial limitations: 1) The HLC algorithm can be used only for class labels and is unable to manage continuous labels, and 2) for the CLC algorithm, if the partitioned ranges are excessively fragmentary because of extensive variation in continuous-label distribution, they are not useful in prediction accuracy or an understanding of the induced models. This condition makes us difficult to use these rules to form business strategies.

In addition, current algorithms cannot directly manage a situation in which the labels are continuous in the data and can be naturally organized as a hierarchical structure. This type of problem is encountered in several domains, including the insurance industry [22], functional genomics [23], supply chains [22], risk management [24], and object recognition [25]. In functional genomics, biologists create a set of possible gene functions that are subsequently organized hierarchically. A hierarchical structure is necessary because each gene may have multiple functions. It is crucial to obtain an interpretable model to understand the interactions between various genes. In each of these cases, the labels are continuous in the data and can be naturally organized as a hierarchical structure of class labels, which defines an abstraction over class labels.

## 2. RELATED WORKS

K. Riesen and H. Bunke [7], in pattern recognition and related areas, graph based representations provide a versatile alternative to the broadly used feature vectors. This work is motivated by the concept of representing graphs by using dissimilarities and enlarges the preliminary work to the more common setting of Lipschitz embeddings. The embedding framework that will be employed in this paper is based on Lipschitz embedding in conjunction with graph edit distance. Qiang Yang, Jie Yin, Charles Ling, and Rong Pan [13], in this work present novel algorithm for post processing decision trees to obtain actions that are associated with attribute-value changes, so as to maximize the profit-based objective functions. These techniques can determine cost effective actions to convert customers from undesirable classes to desirable classes. The method here take incorporates data mining and decision making strongly by preparing the decision making issues directly on top of the data mining outcomes in a post processing step.

M. Doumpos, C. Zopounidis, and Vassiliki [5], a methodology to construct additive models using the SVM framework. The most important advantage of additive models is that they keep hold of the simplicity and clearness/interpretability of linear models, combined with the nonlinear behavior of more difficult classifiers. This can be more significant in numerous contexts where pattern classification methods need not only to be accurate although also transparent and reasonable. This methodology is based on the fact that any real function can be approximated by a piecewise linear function with an arbitrarily large number of linear segments. Expect for the classification performance of the models, the computational effort needed to build and use the models is also a crucial issue.

Z. Barutcuoglu, R. E. Schapire and O. G. Troyanskaya [23], multi-label hierarchical classification in the machine learning phase is defined as the common setting where an example can be a member of any number of classes in a hierarchy. This proposed technique of correcting inconsistent predictions in a multi-label class hierarchy is used to enhanece performance for the majority of functions. In addition to, this system removing inconsistencies, Bayesian approach also completely transforms unbounded real-valued classifier results into marginal probabilities and offers good calibration. Michal R. Chmielewski and Jerzy W. Grzymala-Busse [26], a novel technique of transforming any local discretization approach into a global one is proposed in this work. In this work propose a method which is used hierarchical cluster analysis to discretize attributes. When clustering cannot be completed any further, appropriate postprocessing using a class-entropy evaluate is carry out to combine neighboring intervals. This approach of discretization is called the *cluster analysis method*. The discretization approaches offered here can be classified as either *local* or *global*.

## 3. PROPOSED MECHANISM

This representation, based on fuzzy sets and used in approximate reasoning, is especially applicable to bridging the conceptual gap between subjective/ambiguous features and quantitative data. As a consequence of the smoothness of gradual fuzzy sets and approximate reasoning approaches used, fuzzy representation is also sufficient for dealing with inaccurate and noisy data. Fuzzy rules, according to the fuzzy sets, make use of those qualities of fuzzy representation in an understandable structure of rule bases. In our proposed work, a novel technique according to the partitioning the continuous-valued attributes is proposed. Our proposed approach could be used in the majority algorithms for constructing Decision Tree without wipe out their original properties.

Our proposed algorithm is very similar to ID3. On the other hand, while ID3 selects the test attribute according to the information gain which is calculated by the probability of ordinary data, our algorithm does it by the probability of membership values for data. In the FDT described, the membership function for attribute values is user defined.

We use the algorithm, which is summarized in the following:

1. Generates the root node

2. Tests for leaf node

3. Finds a test attribute

   a. Divides the data according to this attribute

   b. Generates new nodes for fuzzy subsets

4. Makes recursion of the process for the new nodes from point 2.

In our technique, only point 3a is altered as follows: in the beginning, we must describe the cut points. With the intention of select the cut points, first, the attribute values are sorted in an ascending order. Then, we have some possible cut points between data with different classes. We utilize the following acronyms: the information of data (D), the class information entropy E (attribute, D) after discretization with attribute, and the information gain G (attribute, D).

Our proposed algorithm fluctuates from the conventional ID3 algorithm in the following ways.

- There is a membership grade j, $(0 < i < 1)$ given for all input examples.
- The algorithm not only builds a leaf node if all data be a member of the same class excluding also in the following cases:
  - If the proportion of a data set of a class CK is greater than or equal to a given threshold (pre pruning of subsequent nodes because "nearly all" data belong to the same class),
  - If the number of elements in a data set is less than a given threshold (pre pruning because of “numerical tininess” of the set) or o If there are no more attributes for classification (in ID3 there is a null class for this leaf node).
- More than one class name may be allotted to one leaf node (the valid improvement of FDT).
- The fuzzy sets of all attributes are user defined. Each attribute is considered as a linguistic variable. (In our opinion, this is not necessary. The membership function can be calculated from the boundary points of the interval using the algorithm)

## 4. EXPERIMENTAL RESULTS

We analyze and compare the performance offered by existing decision tree C4.5, Hierarchical continuous-label classifier (HCC) technique and proposed fuzzy decision tree (fuzzy HCC) method. The performance is evaluated by the parameters such as accuracy and error rate. Based on the comparison and the results from the experiment show the proposed approach works better than the existing system.

### Accuracy

Accuracy can be calculated from formula given as follows

$$\text{Accuracy} = \frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

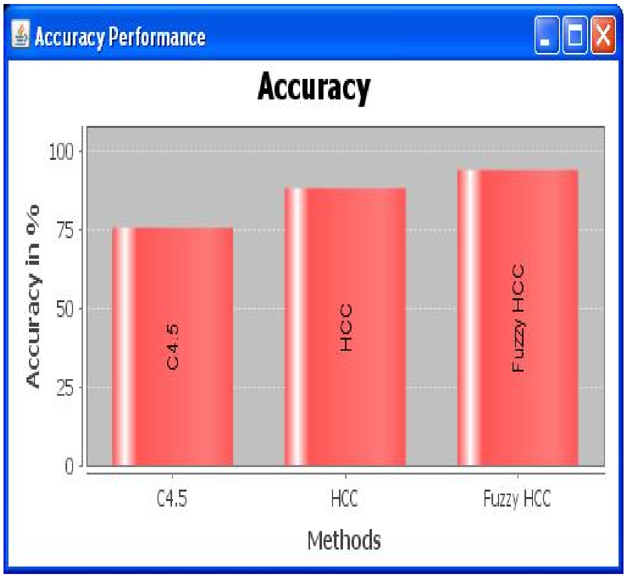


Fig.1. Accuracy comparison

This graph shows the accuracy rate of existing decision tree C4.5, Hierarchical continuous-label classifier (HCC) technique and proposed fuzzy decision tree (fuzzy HCC) method based on two parameters of accuracy and methods such as existing and proposed system. In this graph, x axis will be the methods (existing and proposed system) and y axis will be accuracy in %. From the graph we can see that, accuracy of the system is reduced somewhat in existing system than the proposed system. From this graph we can say that the accuracy of proposed system is increased which will be the best one.

**Error rate**

Error rate can be calculated from formula given as follows

$$\text{Error rate} = \frac{\text{False positive} + \text{False negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

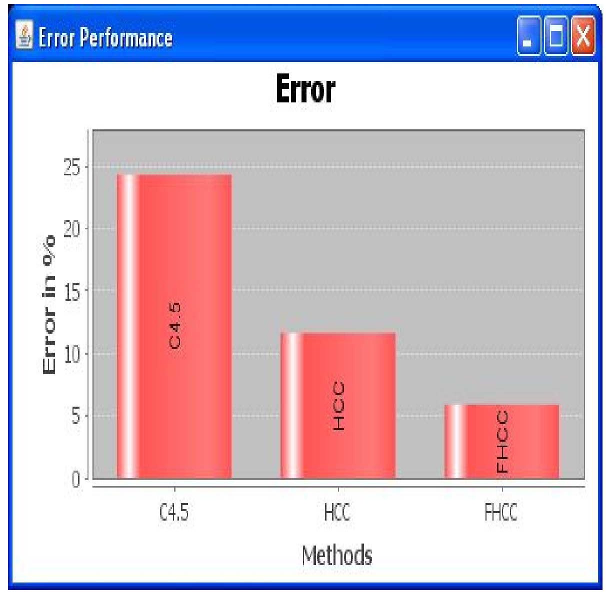


Fig.2. Error rate comparison

This graph shows the error rate of existing decision tree C4.5, Hierarchical continuous-label classifier (HCC) technique and proposed fuzzy decision tree (fuzzy HCC) method based on two parameters of error rate and methods such as existing and proposed system. In this graph, x axis will be the methods (existing and proposed system) and y axis will be error rate in %. From the graph we can see that, error rate of the system is increased somewhat in existing system than the proposed system. From this graph we can say that the error rate of proposed system is reduced which will be the best one.

## 5. CONCLUSION

Decision tree (DT) classifiers are generally used for data with categorical or Boolean class labels. To the best of our knowledge, this is the first classification algorithm for learning DT classifiers from data with hierarchical continuous labels. In the existing system, proposes a novel classification algorithm for learning DT classifiers from data with structured continuous labels. But in this system, many issues are there too see. These drawbacks are addressed in our proposed system. We are proposing the effective and optimized fuzzy decision tree approach. Our proposed work, proposes a efficient approach that is according to partitioning the continuous-valued attributes. The suggested technique could be used in most algorithms for building DT (e.g. Fuzzy ID3) without obliterating their original properties. The experimental results showed that our proposed dexcision tree is superior to existing decision tree because of its efficient prediction accuracy and specificity. In addition, the overall performance of our system is superior to that of existing decision tree method, also in the time complexity.

**REFERENCES**

[1] G. Paleologo, A. Elisseeff, and G. Antonini, "Subagging for credit scoring models," *Eur. J. Oper. Res.*, vol. 201, no. 2, pp. 490–499, Mar. 2010.

[2] D. L. Olson, D. Delen, and Y. Meng, "Comparative analysis of data mining methods for bankruptcy prediction," *Decision Support Syst.*, vol. 52, no. 2, pp. 464–473, Jan. 2012.

[3] Y.-C. Hu and C.-J. Chen, "A PROMETHEE-based classification method using concordance and discordance relations and its application to bankruptcy prediction," *Inf. Sci.*, vol. 181, no. 22, pp. 4959–4968, Nov. 2011.

[4] B. Legrand, C. Che Sau, O. Sim-Heng, N. Soek-Ying, and N. Palanisamy, "Automated identification of chromosome segments involved in translocations by combining spectral karyotyping and banding analysis," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 38, no. 6, pp. 1374– 1384, Nov. 2008.

[5] M. Doumpos, C. Zopounidis, and V. Golfinopoulou, "Additive support vector machines for pattern classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 3, pp. 540–550, Jun. 2007.

[6] H.Kaiqi, T. Dacheng,Y.Yuan, L.Xuelong, and T. Tieniu, "Biologically inspired features for scene classification in video surveillance," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 1, pp. 307–313, Feb. 2011.

[7] K. Riesen and H. Bunke, "Graph classification by means of Lipschitz embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1472–1483, Dec. 2009.

[8] S.-H. Liao, W.-J. Chang, C.-C. Wu, and J. M. Katrichis, "A survey of market orientation research (1995–2008)," *Ind. Market. Manage.*, vol. 40, no. 2, pp. 301–310, Feb. 2011.

[9] J. R. Cano, F. Herrera, and M. Lozano, "Evolutionary stratified training set selection for extracting classification rules with trade off precision interpretability," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 90–108, Jan. 2007.

[10] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[11] J. R. Quinlan, *C4.5: Programs for Machine Learning.* San Mateo, CA: Morgan Kaufmann, 1993.

[12] Y. O. Taner, "Model selection in omnivariate decision trees using structural risk minimization," *Inf. Sci.*, vol. 181, no. 23, pp. 5214–5226, Dec. 2011.

[13] Y. Qiang, Y. Jie, C. Ling, and P. Rong, "Extracting actionable knowledge from decision trees," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 43–56, Jan. 2007.

[14] T. Songbo, "An improved centroid classifier for text categorization," *Expert Syst. Appl.*, vol. 35, no. 1/2, pp. 279–285, Jul. 2008.

[15] W. Shang, H. Huang, H. Zhu, Y. Lin, Y. Qu, and Z.Wang, "A novel feature selection algorithm for text categorization," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 1–5, Jul. 2007.

[16] C.-Y. Fan, P.-C. Chang, J.-J. Lin, and J. C. Hsieh, "A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 632–644, Jan. 2011.

[17] A. P. Sinha and H. Zhao, "Incorporating domain knowledge into data mining classifiers: An application in indirect lending," *Decision Support Syst.*, vol. 46, no. 1, pp. 287–299, Dec. 2008.

[18] B. W. Yap and I. R. Ibrahim, "Using data mining predictive models to classify credit card applicants," in *Proc. 6th Int. Conf. Adv. IMS*, pp. 394–398, 2010.

[19] Z. Zenglian and C. Min, "Research of credit risk of commercial bank's personal loan based on CHAID decision tree," in *Proc. 2nd Int. Conf. AIMSEC*, pp. 7160–7164, 2011.

[20] H. W. Hu, Y. L. Chen, and K. Tang, "A dynamic discretization approach for constructing decision trees with a continuous label," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 11, pp. 1505–1514, Nov. 2009.

[21] Y. L. Chen, H. W. Hu, and K. Tang, "Constructing a decision tree from data with hierarchical class labels," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 4838–4847, Apr. 2009.

[22] J. B. Ayers, *Handbook of Supply Chain Management,* 2nd ed. New York: Auerbach, 2006.

[23] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Bioinformatics*, vol. 22, no. 7, pp. 830–836, Apr. 2006.

[24] I. Bose and R. K.Mahapatra, "Business data mining—A machine learning perspective," *Inf. Manage.*, vol. 39, no. 3, pp. 211–225, Dec. 2001.

[25] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla, "Estimating 3D hand pose using hierarchical multi-label classification," *Image Vis. Comput.*, vol. 25, no. 12, pp. 1885–1894, Dec. 2007.

[26] M. R. Chmielewski and J. W. Grzymala-Busse, "Global discretization of continuous attributes as preprocessing for machine learning," in *Proc. 3rd Int. Workshop Rough Sets Soft Comput.*, pp. 294–301, 1994.