

## Constructing Cost Sensitive Decision Trees Based on Multi-Objective Optimization

---

**Hao Tang<sup>1</sup>**

*Guangdong University of Technology, Guangdong, 201503, China*

*E-mail: 122260851@ qq. com*

We propose a multi-objective optimization based on the cost sensitive decision tree building method. The misclassification cost, test cost, waiting time cost and information gain rate as four optimization goals by using the method of linear weighting are adopted to transfer the multi-objective optimization problem into a single objective optimization problem, as the splitting attribute selection criterion; and then we put forward the specific strategy of building the minimum cost decision tree and a hybrid testing decision tree method; finally, use our algorithm and two other algorithms in two real datasets to build and test the decision trees. The experimental results show that our method of decision tree features less cost, more efficient and stronger generalization ability. The method is especially useful in terms of medical diagnostic.

*CENet2015  
12-13 September 2015  
Shanghai, China*

---

<sup>1</sup>Speaker

## 1. Introduction

The data mining to face three major challenges includes the mining method, the mining and the mining object constraint. This classification problem covers the three major challenges in mining method. In general, the early effect of classification based on accuracy rate was a standard; but in recent years, experts believe that different  $q$  classification error caused by the loss is not the same; therefore, the reducing amount of misclassification cost shall be taken into account as the measure scale.

The decision tree is one of the widely used classification methods. The accuracy rate of the classification results is mainly concerned as to the algorithm of most decision tree; however, it can not fully meet the needs because it only has the the classification accuracy in practice, such as the cost of algorithm test and misclassification are focusing on the factors, which should be considered in application. Literature respectively defines the misclassification cost and the cost of testing, while the required time in obtaining the attribute value in process is called waiting-time cost[1]. In actual classification as comprehensive consideration various cost factors contributed to more price sensitive classification problem, the solution of this problem usually adopts two kinds of typical strategies: firstly, many original prices will change into a new comprehensive cost; secondly, each kind of price is treated as a single object by using the multi-objective optimization technique optimization. Now, the first strategy is to primarily use the construction cost sensitive decision trees[2]. With the method of weighted will cost different decision tree under the unified standard to measure [3], or use the weighted method to make a different price to be merged with another new cost [4]. Taking the strategy can achieve more eclectic results in all kinds of costs, but not the same price. Usually, it is more difficult with the unified standard to carry out measurement, for example, the misclassification cost, the test cost and the waiting time cost; therefore, Literature put forward a method to solve many cost sensitive classification problems based on optimized technology of multi-objectives, the cost of which is defined as a goal of optimization[2]. Among the candidates for building decision tree, the minimum cost decision tree will make choice therefrom; however, construct the minimal cost of the decision tree. This method will increase the complexity of time and space.

In this study, the misclassification cost, the test cost, the waiting time cost and the information gain rate construction optimization problem are used to be the rule of the attribute selection criteria in the process of constructing a decision tree, so as to construct a minimum cost decision tree. Then, the specific broussonetia papyrifera strategies and the test strategy have been put forward for the missing values in the data.

## 2. Construction of the Decision Tree

Assume that the training data and the test data contain the missing values, including the following three steps when we construct the decision tree learning process: (1) choose the breakdown property; (2) establish the cost sensitive decision trees; (3) establish the decision tree test.

### 2.1 Cost Definition

Literature summarized nine main costs, which cover misclassification errors, cost of tests, cost of teacher, cost of Intervention, cost of unwanted achievements, cost of computation, cost of cases, human-computer interaction cost and cost of instability[1]. In this paper, we mainly consider the test cost, the waiting time cost and the cost of misclassification cost .

#### 2.1.1 Misclassification Cost (MC)

The cost of misclassification errors (MC): the punishment cost resulted from wrong classification. As to  $i$  as a class of class label , the predicted category is  $j$ ; thus the

misclassification cost is  $MC_{i,j}$ , where  $MC_{i,j} \in [0, +\infty)$ . In broussonetia papyrifera process, select attribute  $A$  as the splitting attribute, thus the misclassification cost is  $MC(A)$ , where  $MC \in [0, +\infty)$ ;

### 2.1.2 Test Cost (TC)

Cost of tests(TC): document provides six types of testing, including the batch test strategies; however, with introduction of the concept of discount price, TC may have some changes in constructing the decision tree. When the misclassification cost is far greater than the cost of testing, this attribute value test should not be performed in the forecast [5]. With the value of the attribute  $A$ , the test cost is  $TC(A)$  and the actual values are given by expert;

### 2.1.3 Waiting Time Cost (WC)

The waiting time cost: the cost caused by getting the attribute value and waiting for the test results[6]. On the one hand, associated with the actual length of the waiting time, we can use the waiting time to measure the waiting time cost. If a property test of  $A$  waits for the results of another Attribute  $B$  before we can carry out, the waiting time cost of Attribute  $A$  is related to the waiting time cost of Attribute  $B$ , we call Attribute  $A$  as a priori attribute for Attribute  $B$ ; on the other hand, waitfor the test cost with the individual differences; therefore, as to the test Attribute  $A$ , the waiting time cost can be defined as  $WC(A) = WC_T(A) \oplus WC_T(B) \oplus WC_S(A)$ , where  $WC_T(A)$  is the actual waiting time of testing Attribute  $A$ ;  $WC_T(B)$  is the actual waiting time of a prior attribute  $B$  for Attribute  $A$ .  $WC_S(A)$  is that difference factors of the test object, the resource caused by the waiting time cost.  $WC_S(A)$  is determined by the domain experts; in this sense, the  $WC_T(A)$ ,  $WC_T(B)$  and  $WC_S(A)$  metrics are not the same, So we can't simply add up, and they must be unified on metrics. In our definition, use the  $\oplus$  connection  $WC_T(A)$ ,  $WC_T(B)$  and  $WC_S(A)$ . The said  $WC(A)$  is decided by the three parties.

## 2.2 Splitting Attribute Selection

Given a training set of  $S = \{x_n, y_n\}$ ,  $n = 1, \dots, N$ , where  $N$  is the number of samples and the input vector  $x_n$  belongs to a range of  $X \subseteq R^N$ ;  $y_n$  belongs to the category labels set  $Y = \{1, 2, \dots, K\}$ . Each instance of  $\{x_n, y_n\}$  is independent from each other in an unknown distribution  $D: X \times Y$  [7]. Accordingly, given a attribute test cost set  $TC = \{TC(A) | A \in T\}$ , where the waiting time cost for property is the set of  $WC = \{WC(A) | A \in T\}$ , where,  $T$  is the collection of all properties.

The cost sensitive decision tree of the splitting attribute selection task is to use the training set. The test set and the cost of waiting time cost are to find a splitting Attribute  $A$  so that as to any  $A^* \in T'$ ,  $T'$  is the set of all to be split attribute, then

$$\begin{cases} MC(A) \leq MC(A^*) \\ TC(A) \leq TC(A^*) \\ WC(A) \leq WC(A^*) \\ GainRatio(A, S) \geq GainRatio(A^*, S) \end{cases} \quad (2.1)$$

where,  $GainRatio(A, S)$  is the information gain rate of Attribute  $A$ ; thus the splitting attribute selection will constitute a multi-objective optimization problem for the cost sensitive decision tree. After standardization of various price and information gain rate, using the linear weighted sum method to multi-objective optimization problem into single objective optimization problem, then

$$\underset{A \in T'}{\text{Minimize}} F(A) = \alpha_1 MC'(A) + \alpha_2 TC'(A) + \alpha_3 WC'(A) + \alpha_4 (1 - GainRatio'(A, S)) \quad (2.2)$$

where,  $\sum_{i=1}^4 \alpha_i = 1$ [8]. In the actual application, different applications for the cost sensitive

decision tree construction will be different, so we change the value  $\alpha_i$ , namely the change in weight of the type (2.2), to adapt to different application requirements.

### 2.3 Create the Cost Sensitive Decision Trees

According to Formula (2.2), our algorithm chooses the plitting attribute, which can make the  $F$  minimum and generate a node. Similarly in  $C4.5$ , our algorithm is based on the attribute of local optimal choice and is not back, so we are not likely to create the global optimal decision tree in the end; however, take the local optimal method can greatly improve the efficiency of the algorithm achievements.

Firstly, according to Formula (2.2), we choose to make  $F$  minimum attribute as the current node. If there are two or more properties making the minimum at the same time, we further choose attributes according to the following strategy:

- 1) the property has smaller classification error rate of cost before it is not be standardized
- 2) the property has lower testing cost before it is not be standardized;
- 3) the property with less waiting time cost before it is not be standardized;

Secondly, as to the training data with missing values. experimental verification of the various methods handled the missing values to construct the cost sensitive decision trees. In the training data, we will use the method of internal nodes to handle missing values to construct the cost sensitive decision trees.

Thirdly, what kind of conditions is adopted to stop production. Similar to  $C4.5$ , In the following two conditions, when either of them is satisfied, we stop building.

- (a) in all instances of a certain node class, the label is the same;
- (b) To split the collection  $T'$  of attributes is an empty set.

### 2.4 Create the Cost Sensitive Decision Trees

When the construction of a decision tree is completed, we will test the decision tree by using the test data to verify the decision tree whether it meets the requirements, that is, whether the misclassification cost, test cost and waiting time cost can meet certain standards and the needs of practical application or not. As to the test data of missing values, T.K.Jan etc presented four kinds of test strategies and summed up the four kinds of test strategies[7]. S.C. Zhang summed up the current testing strategy of cost sensitive decision trees, covering the main order test and batch testing, and then proposed a mixed strategy of these two kinds of test strategy[5]. This article takes this hybrid test strategy with details shown as follows:

Firstly, according to Formula (2.3), each attribute validity (utility) is calculated.

$$Utility = \frac{MC}{TC + WC} \quad (2.3)$$

Secondly, with Formula (2.4), the bulk properties of validity is calculated. The bulk properties should satisfy the following two points:

- (a) Because these properties as the bulk properties are calibrated by experts in the field and the public part of the test cost between these attributes, named  $CC$  (Common Cost) here;
- (b) all of these bulk properties to spend type (Tangible Cost) of the total cost of resources can not exceed the test case.

Validity of the mass properties calculation formula is shown as follows:

$$Utility = \frac{nMC}{T\_C} \quad (2.4)$$

where  $T\_C = \sum_{i=1}^n TC_i - (n-1)CC + \max_{i=1, \dots, n} \{WC_i\}$ ,  $n$  refers to a number of attributes in the batch attribute.

Finally, the attribute to the validity of the greatest is calculated by the above two steps of test first. When the validity is the largest single attribute, the test strategy can be seen as a smooth test strategy; on the other hand, when the validity is the largest bulk property, it can be regarded as the batch test strategy.

### 3. Experimental Results and Analysis

In order to verify the validity of the algorithm, we chose two very broad applications and the attribute values to discrete data sets from the UCI machine learning repository, in which, the basic information of the data set is shown in Table 1. Each data set we divide it into two parts: the training set (60%) and test set (40%). As the car data sets don't have missing values and the mushroom has 2480 missing values (only 1.39%), we constructed by missing completely at random (MCAR) mechanism respectively the proportion accounted for 10%, 20%, 30%, 40%, 50% and 60% value of attribute missing in the two data sets. As to each attribute, the cost of testing was randomly assigned between 1 and 100 and the waiting time cost were randomly assigned between [0,50], with the misclassification cost shown in Table 2 and table 3. In particular, the misclassification cost is a relative value and this standard is not same as the test cost and the waiting time cost.

Data set	The number of instances	The number of attributes	Category
Car	1728	6	(unacceptable/acceptable/good/v-good)1210/384/69/65
Mushroom	8124	22	(edible/ poisonous)4208/3916

**Table 1:** Experimental Data Set

predicted category original category	edible	poisonous
	edible	0
poisonous	800	0

**Table 2:** Misclassification of Cost of Mushroom Data Set

predicted category original category	unacceptabl e	acceptable	good	v-good
unacceptable	0	100	200	300
acceptable	200	0	100	200
good	400	200	0	100
v-good	600	400	200	0

**Table 3:** Misclassification of Cost of Car Data set

We used three different splitting attribute criteria to construct a cost sensitive decision trees under different missing rates, respectively based on the information gain ratio criterion (M1),the criterion of minimum total cost (M2) and US-based multi-objective optimization criterion (M3). Among them, in our method, the weights of various costs are the same, that is, 0.25; and then the lack of training data is worth of testing the decision tree, M1 is used to C4.5 method for dealing with the missing values, M2 is used to the methods for dealing with missing values, M3 is using the hybrid testing method in this paper. The experimental results are shown in Fig. 1 and Fig. 2.

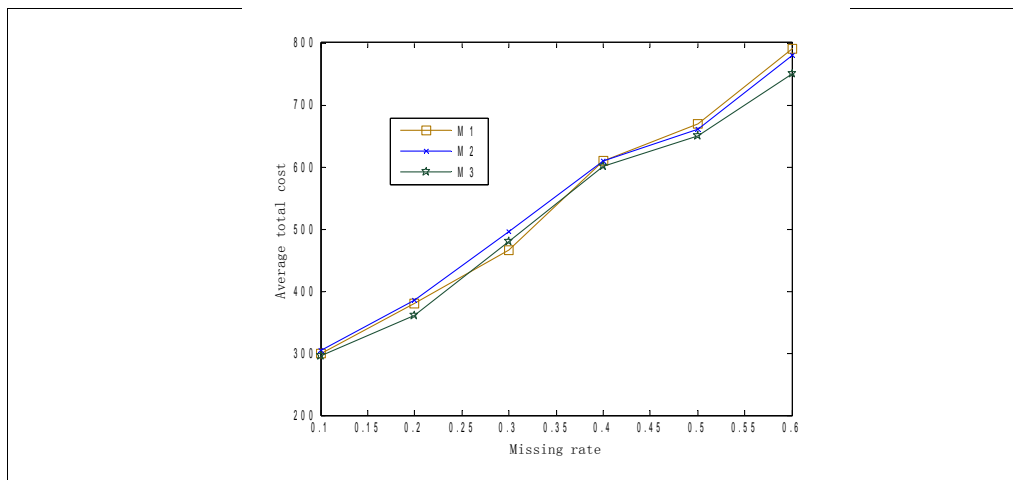


Figure 1 : Car Data Set of Three Kinds of Algorithm Performance under Different Loss Rate

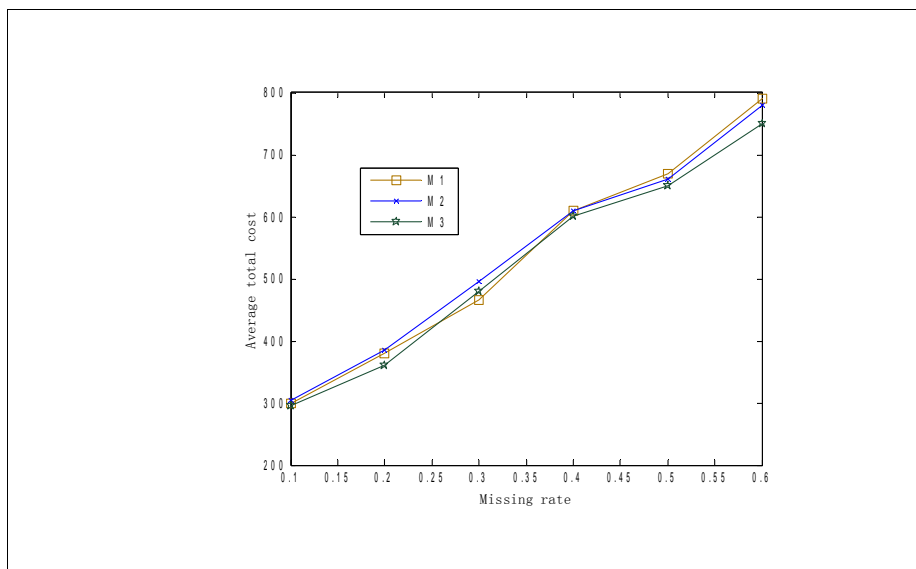


Figure 2: Mushroom Data Set of Three Kinds of Algorithm Performance under Different Loss Rate

POS (CENet2015) 056

Among them, the mean values of misclassification cost, test cost and waiting time cost are the average total cost. From Fig. 1 and Fig. 2, with the increased loss rate, the average total cost also increases because the constructed decision tree performance decreases while the training data test cost and the waiting time cost will be high with higher loss rate; however, when compared with the other two algorithms, our algorithm performs better and more efficiently.

#### 4. Conclusion

In this paper, We put the misclassification cost, test cost, waiting time cost and information gain rate structure multi-objective optimization model, then using the linear weighted sum method, the multi-objective optimization problem into a single objective optimization problem are applied to solve the model, as the attribute selection criterion, and then build a decision tree, the application of a hybrid test of the decision tree method, The experimental results show that, our proposed algorithm is of high efficiency and strong applicability; however, we note that cost sensitive decision tree is of the most importance to reduce the misclassification cost; therefore, in the future work, we put various cost change into multi-objective optimization problems of mathematical models, the misclassification cost is firstly considered in solving the optimal value.

At the same time, as to the excessive pursuit of low misclassification cost, it is likely to be of high misclassification rate for the cost and the practical application of the classification error rate is too high and it will influence the decision tree; therefore, in the subsequent work, in the construction of decision tree, it is necessary to take into account the classification error rate.

#### References

- [1] T. P. *Types of cost in inductive concept learning*. In: Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning, Stanford University, California. 15–11 (2000).
- [2] G. Z. Ma. *Research on Optimization of Multi-cost Sensitive BackPropogation Neural Network*. Doctoral Dissertation of Huazhong University of Science and Technology, Huazhong University of Science and Technology, 2009 (In Chinese).
- [3] Q. Yang, C. Ling, X. Chai. *Test-cost sensitive classification on data with missing values*. IEEE Transactions on Knowledge and Data Engineering. 18[5]:626-638(2006).
- [4] J. V. Davis, J. Ha, C. J. Rossbach, H. E. Ramadn, E. Witchel. *Cost-sensitive decision tree learning for forensic classification*. Proc of 17th European Conference on Machine Learning. Berlin Germany, Springer Berlin Heidelberg, Sep 18-22, 622-629(2006).
- [5] S. C. Zhang, X. F. Zhu, J. L. Zhang and C. Q. Zhang. *Cost-Time sensitive decision tree with missing values*. Knowledge Science, Engineering and Management, Springer Berlin Heidelberg, Berlin Germany, pp.447-459(2007).
- [6] S. Zhang, Z. Qin, C. Ling., S. Sheng. "Missing is Useful": *Missing Values in Cost-sensitive Decision Trees*. IEEE Transactions on Knowledge and Data Engineering 17(12), 1689–1693(2005).
- [7] T. K. Jan, D. W. Wang, C. H. Lin, H. T. Lin. *A simple methodology for soft cost-sensitive classification*. KDD '12 Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, New York, pp.141-149.(2012).
- [8] R. T. Marler, J. S. Arora. *Survey of multi-objective optimization methods for engineering*. Structural and Multidisciplinary Optimization. 2004(4):369-395