Expert Systems with Applications xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Example-dependent cost-sensitive decision trees

Alejandro Correa Bahnsen *, Djamila Aouada, Björn Ottersten

*Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, Luxembourg*

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Cost-sensitive learning
Cost-sensitive classifier
Credit scoring
Fraud detection
Direct marketing
Decision trees

ABSTRACT

Several real-world classification problems are example-dependent cost-sensitive in nature, where the costs due to misclassification vary between examples. However, standard classification methods do not take these costs into account, and assume a constant cost of misclassification errors. State-of-the-art example-dependent cost-sensitive techniques only introduce the cost to the algorithm, either before or after training, therefore, leaving opportunities to investigate the potential impact of algorithms that take into account the real financial example-dependent costs during an algorithm training. In this paper, we propose an example-dependent cost-sensitive decision tree algorithm, by incorporating the different example-dependent costs into a new cost-based impurity measure and a new cost-based pruning criteria. Then, using three different databases, from three real-world applications: credit card fraud detection, credit scoring and direct marketing, we evaluate the proposed method. The results show that the proposed algorithm is the best performing method for all databases. Furthermore, when compared against a standard decision tree, our method builds significantly smaller trees in only a fifth of the time, while having a superior performance measured by cost savings, leading to a method that not only has more business-oriented results, but also a method that creates simpler models that are easier to analyze.



## 1. Introduction

Classification, in the context of machine learning, deals with the problem of predicting the class of a set of examples given their features. Traditionally, classification methods aim at minimizing the misclassification of examples, in which an example is misclassified if the predicted class is different from the true class. Such a traditional framework assumes that all misclassification errors carry the same cost. This is not the case in many real-world applications. Methods that use different misclassification costs are known as cost-sensitive classifiers. Typical cost-sensitive approaches assume a constant cost for each type of error, in the sense that, the cost depends on the class and is the same among examples (Elkan, 2001; Kim, Choi, Kim, & Suh, 2012). Although, this class-dependent approach is not realistic in many real-world applications, for example in credit card fraud detection, failing to detect a fraudulent transaction may have an economical impact from a few to thousands of Euros, depending on the particular transaction and card holder (Sahin, Bulkan, & Duman, 2013). In churn modeling, a model is used for predicting which customers are more likely to abandon a service provider. In this context, failing to identify a profitable or unprofitable churner have a significant different financial impact (Glady, Baesens, & Croux, 2009). Similarly, in direct marketing, wrongly predicting that a customer will not accept an offer when in fact he will, has a different impact than the other way around (Zadrozny, Langford, & Abe, 2003). Also in credit scoring, where declining good customers has a non constant impact since not all customers generate the same profit (Verbraken, Bravo, Weber, & Baesens, 2014). Lastly, in the case of intrusion detection, classifying a benign connection as malicious have a different cost than when a malicious connection is accepted (Ma, Saul, Savage, & Voelker, 2011).

Methods that use different misclassification costs are known as cost-sensitive classifiers. In particular we are interested in methods that are example-dependent cost-sensitive, in the sense that the costs vary among examples and not only among classes (Elkan, 2001). However, the literature on example-dependent cost-sensitive methods is limited, mostly because there is a lack of publicly available datasets that fit the problem (Aodha & Brostow, 2013). Example-dependent cost-sensitive classification methods can be grouped according to the step where the costs are introduced into the system. Either the costs are introduced prior to the training of the algorithm, after the training or during training (Wang, 2013). In Fig. 1, the different algorithms are grouped according to the stage in a classification system where they are used.

* Corresponding author.
*E-mail addresses:* alejandro.correa@uni.lu (A. Correa Bahnsen), djamila.aouada@uni.lu (D. Aouada), bjorn.ottersten@uni.lu (B. Ottersten).

http://dx.doi.org/10.1016/j.eswa.2015.04.042
0957-4174/

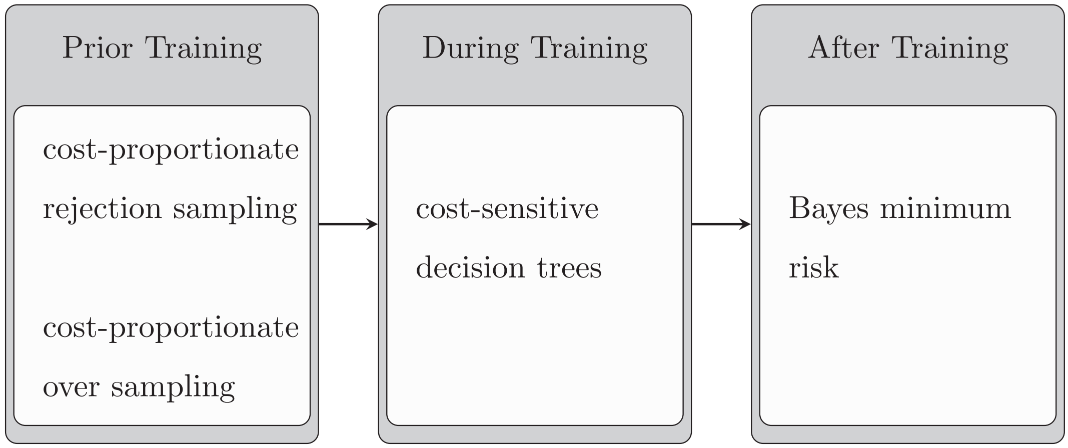


**Fig. 1.** Different example-dependent cost-sensitive algorithms grouped according to the stage in a classification system where they are used.

The first set of methods that were proposed to deal with cost-sensitivity consist in re-weighting the training examples based on their costs, either by cost-proportionate rejection-sampling (Zadrozny et al., 2003), or cost-proportionate over-sampling (Elkan, 2001). The rejection-sampling approach consists in selecting a random subset by randomly selecting examples from a training set, and accepting each example with probability equal to the normalized misclassification cost of the example. On the other hand, the over-sampling method consists in creating a new set, by making $n$ copies of each example, where $n$ is related to the normalized misclassification cost of the example. Recently, we proposed a direct cost approach to make the classification decision based on the expected costs. This method is called Bayes minimum risk (*BMR*), and has been successfully applied to detect credit card fraud (Correa Bahnsen, Stojanovic, Aouada, & Ottersten, 2013; Correa Bahnsen, Stojanovic, Aouada, & Ottersten, 2014). The method consists in quantifying tradeoffs between various decisions using probabilities and the costs that accompany such decisions.

Nevertheless, these methods still use a cost-insensitive algorithm, and either by modifying the training set or the output probabilities convert it into a cost-sensitive classifier. Therefore, leaving opportunities to investigate the potential impact of algorithms that take into account the real financial example-dependent costs during the training of an algorithm.

The last way to introduce the costs into the algorithms is by modifying the methods. The main objective of doing this, is to make the algorithm take into account the example-dependent costs during the training phase, instead of relying on a pre-processing or post-processing method to make classifiers cost-sensitive. In particular this has been done for decision trees (Draper, Brodley, & Utgoff, 1994; Kretowski & Grześ, 2006; Ling, Yang, Wang, & Zhang, 2004; Li, Li, & Yao, 2005; Ting, 2002; Vadera, 2010). In general, the methods introduce the misclassification costs into the construction of a decision trees by modifying the impurity measure, and weight it with respect of the costs (Lomax & Vadera, 2013). However, in all cases, approaches that have been proposed only deal with the problem when the cost depends on the class and not on the example.

In this paper we formalize a new measure in order to define when a problem is cost-insensitive, class-dependent cost-sensitive or example-dependent cost-sensitive. Moreover, we go beyond the aforementioned state-of-the-art methods, and propose a decision tree algorithm that includes the example-dependent costs. Our approach is based first on a new example-dependent cost-sensitive impurity measure, and secondly on a new pruning improvement measure which also depends on the cost of each example.

We evaluate the proposed example-dependent cost-sensitive decision tree using three different databases. In particular, a credit card fraud detection, a credit scoring and a direct marketing databases. The results show that the proposed method outperforms state-of-the-art example-dependent cost-sensitive methods. Furthermore, when compared against a standard decision tree, our method builds significantly smaller trees in only a fifth of the time. Furthermore, the source code used for the experiments is publicly available as part of the *CostSensitiveClassification*[1] library.

By taking into account the real financial costs of the different real-world applications, our proposed example-dependent cost-sensitive decision tree is a better choice for these and many other applications. This is because, our algorithm is focusing on solving the actual business problems, and not proxies as standard classification models do. We foresee that our approach should open the door to developing more business focused algorithms, and that ultimately, the use of the actual financial costs during training will become a common practice.

The remainder of the paper is organized as follows. In Section 2, we explain the background behind example-dependent cost-sensitive classification and we define a new formal definition of cost-sensitive classification problems. In Section 3, we make an extensive review of current decision tree methods, including by the different impurity measures, growth methods, and pruning techniques. In Section 4, we propose a new example-dependent cost-sensitive decision tree. The experimental setup and the different datasets are described in Section 5. Subsequently, the proposed algorithm is evaluated on the different datasets. Finally, conclusions of the paper are given in Section 7.

## 2. Cost-sensitive cost characteristic and evaluation measure

In this section we give the background behind example-dependent cost-sensitive classification. First we present the cost matrix, followed by a formal definition of cost-sensitive problems. Afterwards, we present an evaluation measure based on cost. Finally, we describe the most important state-of-the-art methods, namely: Cost-proportionate sampling and Bayes minimum risk.

### 2.1. Binary classification cost characteristic

In classification problems with two classes $y_i \in \{0, 1\}$, the objective is to learn or predict to which class $c_i \in \{0, 1\}$ a given example $i$

[1] https://github.com/albahnsen/CostSensitiveClassification