Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Approximate Reasoning
39 (2005) 235–255

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Upper entropy of credal sets. Applications to credal classification

Joaquín Abellán, Serafín Moral *

Dpto. Ciencias de la Computación, Universidad de Granada, Granada 18071, Spain

Received 1 February 2004; accepted 1 October 2004
Available online 25 November 2004

Abstract

We present an application of the measure of entropy for credal sets: as a branching criterion for constructing classification trees based on imprecise probabilities which are determined with the imprecise Dirichlet model. We also justify the use of upper entropy as a global uncertainty measure for credal sets and present a deduction of this measure. We have carried out several experiments in which credal classification trees are built taking a global uncertainty measure as a basis. The results show how the introduced methodology improves the performance of traditional methods (Naive Bayes and C4.5), by providing a much lower error rate.

Keywords: Imprecise probabilities; Uncertainty; Upper entropy; Imprecision; Non-specificity; Classification; Classification trees; Credal sets

1. Introduction

Classification is an important problem in the area of machine learning in which traditional probability theory has been extensively used. Basically, we have an incoming set of observations, called the training set, and, generally, we want to

* Corresponding author. Tel.: +34 9 58 242819; fax: +34 9 58 243317.
E-mail addresses: jabemu@teleline.es (J. Abellán), smc@decsai.ugr.es (S. Moral).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2004.10.001

obtain a model to assign a value of the class variables to any new observation. The set of observations used to assess the quality of this model is also called the test set. Classification has notable applications in medicine, recognition of hand-written characters, astronomy, banking, etc. The learned classifier can be represented as a Bayesian network, a neural network, a classification tree, etc. These methods normally use the theory of probability to estimate the parameters with a stopping criterion to limit the complexity of the classifier and to avoid overfitting.

In some previous papers [4–6], we have introduced a new procedure to build *classification trees* based on the use of *imprecise probabilities*. Classification trees have their origin in Quinlan's ID3 algorithm [24], and a basic reference is the book by Breiman et al. [8]. In this paper, we also apply decision trees for classification, but as in [32], the *imprecise Dirichlet model* [29] is used to learn the model and to decide among the possible classes.

In classical probabilistic approaches, *information gain* [24] is used to build the tree, but then other procedures must subsequently be used to prune it, since information gain tends to build structures which are too complex. We have shown that if imprecise probabilities are used and the information gain is computed by measuring the total amount of uncertainty of the associated *credal set* (a closed and convex set of probability distributions), then the problem of overfitting disappears and results improve.

In [1–3], we studied how to measure the uncertainty of a credal set by generalizing the measures used in the *theory of evidence*, [11,26]. We considered two main sources of uncertainty: *entropy* and *non-specificity*. We proved that the proposed functions satisfy the most basic desiderata of these types of measures [2,14,20].

We previously proved that by using a global uncertainty measure which is the result of adding an entropy measure and a non-specificity measure, classification results are better than those obtained by the C4.5 classification method, based on Quinlan's ID3 algorithm. In this paper, we have carried out some experiments in which the upper entropy of the probability distributions of a credal set is used to measure its uncertainty, and we find that the results obtained are even better.

We consider two methods of building classification trees. In the first method, [4], we start with an empty tree and in each step, the variable which produces the largest decrease in the entropy of the class variable is selected for branching. The second method quantifies the uncertainty of each individual variable in each node in the same way, but also considers the results of adding two variables at the same time. In this way, we aim to discover relationships involving more than two variables that were not seen when investigating the relationships of a single variable with the class variable.

In traditional probability, adding a new branch always produces a decrease in entropy. It is necessary to include an additional criterion so as not to create models which are too complex and therefore overfit the data. With credal sets, adding a branch can produce smaller entropy but, at the same time, it will always give rise to greater non-specificity. Under these conditions, we follow the same procedure as in probability theory, but measuring the total uncertainty of adding a branch.