International Journal of Approximate Reasoning 56 (2015) 249–263

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Robust classification of multivariate time series by imprecise hidden Markov models

Alessandro Antonucci [a,*], Rocco De Rosa [b], Alessandro Giusti [a], Fabio Cuzzolin [c]

[a] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Manno, Switzerland
[b] Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano, Milano, Italy
[c] Department of Computing, Oxford Brookes University Wheatley Campus, Wheatley, Oxford, United Kingdom

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form 4 July 2014
Accepted 9 July 2014
Available online 14 July 2014

Keywords:
Multivariate time series
Credal classification
Hidden Markov models
Gaussian mixtures
Imprecise probability
Bhattacharyya distance

ABSTRACT

A novel technique to classify time series with imprecise hidden Markov models is presented. The learning of these models is achieved by coupling the EM algorithm with the imprecise Dirichlet model. In the stationarity limit, each model corresponds to an imprecise mixture of Gaussian densities, this reducing the problem to the classification of static, imprecise-probabilistic, information. Two classifiers, one based on the expected value of the mixture, the other on the Bhattacharyya distance between pairs of mixtures, are developed. The computation of the bounds of these descriptors with respect to the imprecise quantification of the parameters is reduced to, respectively, linear and quadratic optimization tasks, and hence efficiently solved. Classification is performed by extending the k-nearest neighbors approach to interval-valued data. The classifiers are credal, meaning that multiple class labels can be returned in the output. Experiments on benchmark datasets for computer vision show that these methods achieve the required robustness whilst outperforming other precise and imprecise methods.

1. Introduction

The theory of *imprecise probability* [1] extends the Bayesian theory of subjective probability to cope with sets of distributions. This potentially provides more robust and realistic models of uncertainty. These ideas have been applied to classification and a number of classifiers based on imprecise probabilities are already available. Most of these approaches are based on graphical models, whose parameters are imprecisely quantified with a set of priors by means of the *imprecise Dirichlet model* [2]. The first attempt in this direction is the *naive credal classifier* [3], which generalizes the naive Bayes classifier to imprecisely specified probabilities. Each prior in the imprecise Dirichlet model defines a precise classifier. When two precise classifiers of this kind assign a different class label to the same instance, the imprecise classifier returns both labels, and the instance is said to be *prior-dependent*. Conversely, when a single label is returned, this is independent of the prior. Classifiers of this kind, possibly returning multiple class labels in the output, are called *credal*.[1] The separation between

* Corresponding author.
E-mail addresses: alessandro@idsia.ch (A. Antonucci), rocco.derosa@unimi.it (R. De Rosa), alessandrog@idsia.ch (A. Giusti), fabio.cuzzolin@brookes.ac.uk (F. Cuzzolin).

[1] Besides the naive, other examples of credal classifiers proposed in the literature include models with more complex topologies [4] and imprecise averages of precise models [5]. Alternative quantification techniques not based on the imprecise Dirichlet model have been proposed for general topologies in [6].

http://dx.doi.org/10.1016/j.ijar.2014.07.005
0888-613X/

prior-dependent and other instances induced by a credal classifier typically corresponds to a distinction between hard and easy-to-classify instances, with the accuracy of a precise classifier significantly lower on the prior-dependent instances rather than on the prior-independent ones. In this sense, credal classifiers are suitable as preprocessing tools, assigning the right class label to prior-independent instances and partially suspending the judgement otherwise.

Despite the relatively large number of credal classifiers proposed in the literature, no credal models specifically intended to classify temporal data have been developed so far.[2] This is cumbersome since, on the other side, dynamical models such as Markov chains and *hidden Markov models* (HMMs) have been already extended to imprecise probabilities to model non-stationary dynamic processes [8,9]. As a matter of fact, HMMs in their precise formulation have been often applied to classification of time series (e.g., [10]), while no similar attempts have been made in the imprecise case. This can be partially explained by the lack of algorithms to learn imprecise-probabilistic models from incomplete data (e.g., because referred to hidden variables) and, more marginally, by the lack of suitable inference algorithms.

It therefore seems natural to merge these two lines of research and develop credal classifiers for time series based on imprecise HMMs. To achieve that, we first show how to learn an imprecise HMM from a discrete-time sequence. The technique, already tested in previous works [7,11] combines the imprecise Dirichlet model with the popular EM algorithm, generally used to learn precise HMMs. After this step, each sequence is associated with an imprecise HMM. In the limit of infinitely long models, HMMs might converge to a condition of *stationarity*, even in the imprecise case [12,13]. A major claim of this paper is that, in this limit, the model becomes considerably simpler without losing valuable information for classification purposes.

In the stationarity limit, the imprecise HMM becomes an *imprecise mixture* (i.e., with multiple specification of the weights) of Gaussian densities over the observable variable (i.e., the joint observation of the features). Two novel algorithms are proposed to perform classification with these models. The first, called IHMM-E, evaluates the mixture expected value, which becomes a static attribute for a standard classification setup. The second, called IHMM-B, uses the Bhattacharyya distance between two mixtures as a descriptor of the dissimilarity level between sequences. Being associated with imprecise-probabilistic models, those descriptors cannot be precisely evaluated and only their lower and upper bounds with respect to the constraints on the parameters can be evaluated. This is done efficiently by solving a linear (for IHMM-E) and a quadratic (for IHMM-B) optimization task.

After this step, IHMM-E summarizes the sequence as an interval-valued observation in the feature space. To classify this kind of information, a generalization of the *k-nearest neighbors* algorithm to support multivariate interval data is developed. The same approach can be used to process the interval-valued (univariate) distances between sequences returned by IHMM-B. Both algorithms are credal classifiers for time series, possibly assigning more than a single class label to a sequence. Performances are tested on some of the most important computer vision benchmarks. The methods we propose achieve the required robustness in the evaluation of the class labels to be assigned to a sequence and outperform alternative imprecise methods with respect to state-of-the-art metrics [14] to compare performances of credal and traditional classifiers.

The performance is also good when compared with *dynamic time warping*, the state-of-the-art approach to the classification of time series. The reason is the high dimensionality of the computer vision data: dynamic time warping is less effective when coping with multivariate data [15], while the methods in this paper are almost unaffected by the dimensionality of the features.

The paper is organized as follows. In Section 2, we introduce the basic ideas in the special case of precise HMMs obtained from univariate data. Then, in Section 3, we define imprecise HMMs and discuss the learning of these models from multivariate data. The new algorithms IHMM-E and IHMM-B are detailed in Sections 4 and 5. A summary of the two methods together with a discussion about their computational complexity and the performance evaluation are in Section 6. Experiments and conclusive remarks are in Sections 7 and 8.

2. Time series classification

Let us introduce the key features of our approach and the necessary formalism in the precise univariate case. Variables $O_1, O_2, \ldots, O_T$ denote the observations of a particular phenomenon at T different (discrete) times. These are assumed to be *observable*, i.e., their actual (real) values are available and denoted by $o_1, o_2, \ldots, o_T$. If the observations are all sampled from the same distribution, say $P(O)$, the empirical mean converges to its theoretical value (strong law of large numbers):

$$\lim_{T\to+\infty} \frac{\sum_{i=1}^{T} o_i}{T} = \int_{-\infty}^{+\infty} o \cdot P(o) \cdot \mathrm{d}o. \tag{1}$$

Under the stationarity assumption, the empirical mean is therefore a sensible descriptor of the sequence. More generally, observations at different times can be sampled from different distributions (i.e., the process can be non-stationary). Such a situation can be modeled by pairing O_t with an auxiliary discrete variable X_t, for each $t = 1, \ldots, T$. The values of variables

[2] The only exception is a previous work of the authors [7]. The algorithms in the present paper are a natural evolution of those approaches. Numerical tests showing much better performances of the new methods are reported in Section 7.