Signal Processing 128 (2016) 8–17

Contents lists available at ScienceDirect

Signal Processing

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

Unsupervised classification using hidden Markov chain with unknown noise copulas and margins

Stéphane Derrode [a,*], Wojciech Pieczynski [b]

[a] École Centrale de Lyon, LIRIS, CNRS UMR 5205, Écully, France
[b] SAMOVAR, Telecom Sudparis, CNRS, Université Paris-Saclay, 9, rue Charles Fourier, Evry, France

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form
7 March 2016
Accepted 17 March 2016
Available online 24 March 2016

Keywords:
Hidden Markov models
Dependent noise
Model selection
Iterative conditional estimation
Copulas
Unsupervised classification
Pearson's system of distributions

ABSTRACT

We consider the problem of unsupervised classification of hidden Markov models (HMC) with dependent noise. Time is discrete, the hidden process takes its values in a finite set of classes, while the observed process is continuous. We adopt an extended HMC model in which the rich possibilities of different kinds of dependence in the noise are modelled via copulas. A general model identification algorithm, in which different noise margins and copulas corresponding to different classes are selected in given families and estimated in an automated way, from the sole observed process, is proposed. The interest of the whole procedure is shown via experiments on simulated data and on a real SAR image.

1. Introduction

The paper deals with the problem of unsupervised estimation of a hidden discrete process $\boldsymbol{X}_1^N = (X_1, \dots, X_N)$ from an observed continuous one $\boldsymbol{Y}_1^N = (Y_1, \dots, Y_N)$. Hidden Markov models (HMMs) are very widely used to deal with the problem. Indeed, they allow recursive computations of different quantities used in optimal Bayesian processing in linear time. There are many papers following the pioneering ones [1,2], dealing with various application areas. Let us mention some recent general papers or books about general setting [3–5], signal and image processing [3], economy and finance [6,7], or biology [8,9]. Besides, copulas [10,11] are also of interest in numerous situations, due to their ability of modelling dependent non-Gaussian data [12–15]. Their use goes increasing in different areas. Mainly applied in economy and finance [16–21], they are becoming increasingly used in other fields, such as in signal or image processing [22–25] or in ecology [26–28].

However, despite their great benefit when used separately, there is very little research and applications that combines them. First papers on the subject date from about ten years: copulas use has been introduced at temporal level in hidden Markov chains with dependent noise (HMC-DN) in [29], at vectorial level in hidden Markov chains in [30], and in hidden Markov trees in [31]. Some applications using vectorial-level copulas have been proposed in the context of hidden Markov chains [32], hidden Markov trees [33], hidden Markov fields [34,35], or general Bayesian networks [36]. They were showed to be especially useful in multi-sensor image processing where sensors are dependent and not Gaussian [34,35]. Temporal-level copulas remain, for their part, very little used. This is certainly due to the fact that the observations in HMMs are usually assumed to be independent conditionally on the hidden data, and thus there is no dependency to model. However, taking into account the noise dependence is of interest, and using the right copulas can have strong influence on the efficiency of Bayesian processing methods in HMMs with correlated noise [37].

Our paper deals with the problem of unsupervised classification of hidden Markov chains with copulas used at temporal level. The novelty of the work is to propose a general method allowing one to search the best copulas in a finite set of admissible copulas, as well as the best margins in a finite set of admissible margins. In addition, the admissible sets of copulas and margins can vary with the hidden discrete data. This allows one to select, from the only observed data, the best model in a quite rich set of possible models. Therefore we simultaneously extend, first, the method presented in [37] where the copulas where searched while the forms of margins were assumed known and, second, the method presented in [38,39] where the margins were searched while assuming independence.

Let us notice that the presented results can be almost directly applied to more complex models than the HMC-DNs considered.

* Corresponding author.

http://dx.doi.org/10.1016/j.sigpro.2016.03.008
0165-1684/

Indeed, when parameter estimation is concerned, dealing with "pairwise Markov models" (PMMs) [40,41] or even "triplet Markov models" (TMMs), which includes non stationary PMMs [42], hidden semi-Markov models [43], or still hidden bivariate Markov models [44], is a quite similar problem [42,43].

The organization of the paper is the following. In next Section we recall the basics about HMM and how a dependent noise can be modelled using a copula representation. The general model identification method we propose is then specified in Section three. Section four is devoted to recall the classic computations in HMM-DN for different quantities of interest. Fifth section contains some systematic experiments and the segmentation result of a real SAR image. The last Section draws conclusions and proposes a few perspectives.

2. HMM with dependent noise and copulas

Let us consider two random sequences $\boldsymbol{X}_1^N = (X_1, \ldots, X_N)$ and $\boldsymbol{Y}_1^N = (Y_1, \ldots, Y_N)$, taking their values in $\Omega = \{1, \ldots, K\}$ and $\mathbb{R}$ respectively. $\boldsymbol{X}_1^N$ is hidden, while $\boldsymbol{Y}_1^N$ is observed, and the problem is to estimate $\boldsymbol{X}_1^N$ from $\boldsymbol{Y}_1^N$. Optimal Bayesian methods can be used for the classic hidden Markov models (HMMs), whose distribution is defined with

$$p\left(\boldsymbol{x}_1^N, \boldsymbol{y}_1^N\right) = p(x_1)p(y_1|x_1)p(x_2|x_1)p(y_2|x_2)\ldots p(x_N|x_{N-1})p(y_N|x_N). \tag{1}$$

HMMs can also be defined as verifying two hypotheses:

$$\boldsymbol{X}_1^N \text{ is Markov;} \tag{2}$$

$$p(\boldsymbol{y}_1^N|\boldsymbol{x}_1^N) = \prod_{n=1}^{N} p(y_n|x_n). \tag{3}$$

Let us notice that (3) means that the random variables $Y_1, \ldots, Y_N$ are independent conditionally on $\boldsymbol{X}_1^N$; for this reason we will call the classic HMM (2) and (3) "HMM with independent noise" (HMM-IN).

It is possible to consider more general models in which both processes $(\boldsymbol{X}_1^N, \boldsymbol{Y}_1^N)$ and $\boldsymbol{X}_1^N$ are Markov and in which the same Bayesian processing as in HMM-IN remains possible. The distribution of such models is written

$$p(x_{n+1}, y_{n+1}|x_n, y_n) = p(x_{n+1}|x_n)p(y_{n+1}|x_n, y_n, x_{n+1}). \tag{4}$$

In these kind of models, called HMM with dependent noise (HMM-DN) $Y_1,\ldots,Y_N$ are (possibly) dependent conditionally on $\boldsymbol{X}_1^N$. Thus an HMM-IN is an HMM-DN for which $p(y_{n+1}|x_n, y_n, x_{n+1}) = p(y_{n+1}|x_{n+1})$.

Remark 2.1. It has been shown in [41,40] that the Markovianity of $\boldsymbol{X}_1^N$ is not even required, and the following model called "pairwise Markov model" (PMM):

$$p\left(\boldsymbol{x}_1^N, \boldsymbol{y}_1^N\right) = p\left(x_1, y_1\right) \sum_{n=1}^{N-1} p\left(x_{n+1}, y_{n+1}|x_n, y_n\right) \tag{5}$$

allows the same processing than HMM-DNs.

In this paper we will deal with the stationary reversible case, which means that $p(x_n, y_n, x_{n+1}, y_{n+1})$ does not depend on $n = 1, \ldots, N-1$, and the distributions $p(x_{n+1}, y_{n+1}|x_n, y_n)$ and $p(x_n, y_n|x_{n+1}, y_{n+1})$ are equal. In that case, an HMM-DN is a particular case of PMM for which we have

$$p(y_{n+1}|x_{n+1}, x_n) = p(y_{n+1}|x_{n+1}), \tag{6}$$

for all $n \in [1, N-1]$, see [41]. Thus in the model considered in this paper we have simultaneously (4) and (6). Let us notice that (6) does not imply that $p(y_{n+1}|x_n, y_n, x_{n+1})$ can be reduced to a simpler expression: the distribution of Y_{n+1} conditional on X_n, Y_n, X_{n+1} can depend on the three variables.

The distribution of such a stationary reversible HMM-DN $\left(\boldsymbol{X}_1^N, \boldsymbol{Y}_1^N\right)$ is defined by

$$p(x_1, y_1, x_2, y_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2). \tag{7}$$

The aim of this paper is to consider $p(y_1, y_2|x_1, x_2)$ in (7) under very general form and to propose a way for its estimation, together with $p(x_1, x_2)$, from the observed sequence $\boldsymbol{Y}_1^N$. More precisely, for given (x_1, x_2), $p(y_1, y_2|x_1, x_2)$ is defined by

- two margins $p(y_1|x_1, x_2) = p(y_1|x_1) = f_{x_1}^l(y_1)$ and $p(y_2|x_1, x_2) = p(y_2|x_2) = f_{x_2}^r(y_2)$, according to (6) ($l$ and r stand for 'left' and 'right' to distinguish between the left and right variables, see below);
- a copula C with pdf $c\left(F_{x_1}^l(y_1), F_{x_2}^r(y_2)|x_1, x_2\right) = c_{x_1,x_2}\left(F_{x_1}^l(y_1), F_{x_2}^r(y_2)\right)$, where F is the cumulative distribution function (cdf) corresponding to f.

We recall that a copula C is defined as a cumulative distribution function on $[0, 1]^2$ such that the corresponding marginal cumulative functions are identity, which also means that the corresponding marginal distributions on $[0, 1]$ are uniform distributions, see e.g. [10]. Let $h(y_1, y_2)$ be a probability distribution on $\mathbb{R}^2$, which will be assumed continuous in this paper. Let $H(y_1, y_2)$ be the corresponding cumulative function, $h^l(y_1)$ and $h^r(y_2)$ the corresponding marginal densities, and $H^l(y_1)$, $H^r(y_2)$ the associated cumulative functions. According to Sklar's theorem [11] there exists an unique copula C such that

$$H(y_1, y_2) = C\left(H^l(y_1), H^r(y_2)\right). \tag{8}$$

Setting $c(u, v) = \dfrac{\partial C(u, v)}{\partial u \partial v}$ and deriving (8) with respect to y_1, y_2 gives

$$h(y_1, y_2) = h^l(y_1)h^r(y_2)c(H^l(y_1), H^r(y_2)). \tag{9}$$

Thus any continuous probability distribution $h(y_1, y_2)$ is given by a triplet h^l, h^r, and a probability distribution c on $[0, 1]^2$ with uniform margins. Conversely, such a triplet defines a probability distribution on $[0, 1]^2$ with (9). Such a representation of $h(y_1, y_2)$ is of interest as every distribution among h^l, h^r, c can be modified independently from the two others. For example, a Gaussian copula $h(y_1, y_2)$ is given by Gaussian margins h^l, h^r, and a Gaussian copula c. Replacing in (8) c with another non Gaussian copula c' we obtain a non Gaussian distribution $H'(y_1, y_2)$ with Gaussian margins. We can also keep the Gaussian copula c and replace the Gaussian margins by any other ones. This offers a very rich set of possibilities easy to handle with.

We will assume that for each $(x_1, x_2) \in \Omega^2$, each $f_{x_1}^l$ and each $f_{x_2}^r$ belongs to a parametric set of distributions, which themselves belongs to a finite family of parametric sets of distributions. For example, imagine that f_1^l can be Gaussian or Gamma, f_1^r can be Beta, Gamma or Rayleigh, f_2^l can be Beta or Gamma, f_2^r can be exponential and so on for $x_1 = 3, \ldots, K$, $x_2 = 3, \ldots, K$. Thus, for each $(x_1 = i, x_2 = j)$, we have to find what is the general form of the distributions f_i^l and f_j^r, and we have to find the parameters, which precisely define the distribution of the determined shape. Similarly, for each $(x_1 = i, x_2 = j)$ we have to find general form of copula $c_{i,j}$ and estimate the parameters, which set the copula in the set of copulas having the same form. For example, $c_{1,1}$ can be Gumbel or Gaussian, $c_{1,2}$ can be Gaussian or Clayton, $c_{2,1}$ can be Student,