MODELING NON STATIONARY HIDDEN SEMI-MARKOV CHAINS WITH TRIPLET MARKOV CHAINS AND THEORY OF EVIDENCE

Wojciech Pieczynski

INT/GET, Département CITI, CNRS UMR 5157
9, rue Charles Fourier, 91000 Evry, France

ABSTRACT

Hidden Markov chains, enabling one to recover the hidden process even for very large size, are widely used in various problems. On the one hand, it has been recently established that when the hidden chain is not stationary, the use of the theory of evidence is equivalent to consider a triplet Markov chain and can improve the efficiency of unsupervised segmentation. On the other hand, hidden semi-Markov chains can also be considered as particular triplet Markov chains. The aim of this paper is to use these two points simultaneously. Considering a non stationary hidden semi-Markov chain, we show that it is possible to consider two auxiliary random chains in such a way that unsupervised segmentation of non stationary hidden semi-Markov chains is workable.

1. INTRODUCTION

Let $Z = (X, Y)$, with $X = (X_1, ..., X_n)$, $Y = (Y_1, ..., Y_n)$ be two random chains, where each X_i takes its values in $\Omega = \{\omega_1, ..., \omega_k\}$ and each Y_i takes its values in R. We will say that $Z = (X, Y)$ is a classical hidden Markov chain with independent noise (HMC-IN) if its distribution reads $p(z) = p(x_1)p(x_2|x_1)...p(x_n|x_{n-1})p(y_1|x_1)...p(y_n|x_n)$. Then the hidden chain X is a Markov one, and we add "independent noise" because the random variables Y_1, …, Y_n are independent conditionally on X. This condition can be relaxed and there exist hidden Markov models in which X is Markovian and the random variables Y_1, …, Y_n are not independent conditionally on X; such a model will be called hidden Markov chains (HMC).

HMC-IN are widely used because unknown realisations of X can be estimated from observed realisations of Y by different Bayesian methods, even for very large values of n. Furthermore, different model parameters estimation methods, like the "expectation-maximization" (EM) algorithm or the "iterative conditional estimation" (ICE) one, are available, which enables unsupervised estimation of X from Y.

Classically, HMC-IN have been extended in two directions:

(i) In HMC-IN the hidden chain X is a Markov one, and thus the sojourn duration distribution in each state is exponential. In hidden semi-Markov chains with independent noise (HSMC-IN), which form an extension of HMC-IN, this distribution is of any kind. HSMC-IN are useful in many situations, as images sequence analysis [5], speech processing [6], or still tracking problems [15], among others;

(ii) more recently, HMC-IN have been extended to "pairwise Markov chains" (PMC [9]), in which one directly assumes the Markovianity of $Z = (X, Y)$ and in which X is no longer necessarily a Markov chain, and to "triplet Markov chains" (TMC [10, 13]), in which one introduces a third auxiliary random chain $U = (U_1, ..., U_n)$ and assumes the Markovianity of the triplet $T = (X, U, Y)$. When the random variables $U_1, ..., U_n$ take their values in a discrete finite space, both PMC and TMC still enable to estimate X from Y by Bayesian methods. Let us mention that TMC can be also used when the three chains X, U, and Y are continuous; in this case Bayesian segmentation methods are replaced by Kalman, or particle, filtering techniques [1, 4].

Otherwise, we have two following recent results:

(iii) let us return to classical HMC-IN, with unknown parameters and with a non stationary process X. It is possible to estimate the model parameters by EM or ICE; however, such estimation methods assume the stationarity of X and would thus necessarily give wrong results, which in turn can imply poor restoration results of $X = x$. It then has been shown that replacing the non stationary priors by "evidential" stationary ones ("evidential" refers to the theory of evidence [2, 14]) enables to improve the final segmentation results [7, 8]. Moreover, introducing such "evidential" priors is identical, from the mathematical viewpoint, to consider a particular TMC [7, 8];

(iv) very recently it has been showed that HSMC-IN can be seen as particular TMC [11, 13].

The aim of this paper is to use points (iii) and (iv) above simultaneously. Considering a non stationary hidden semi-Markov chain, we introduce a TMC in which the auxiliary chain has two components. The first one models the HSMC-IN, as in [11], and the second one enables us to take into account, via evidential priors, its non stationarity, as in [7, 8]. We detail the use of such a model in Byayesian segmentation and briefly describe both EM and ICE parameter estimation methods.

2. HIDDEN SEMI MARKOV CHAIN AS A TRIPLET MARKOV CAHINNTRODUCTION

Since [11] is under submission and the problem is just mentioned in [13], let us briefly specify, via an example, why each HSMC-IN is a particular TMC. Let (X,Y) be an HSMC-IN with the distribution given by :

(a) the distribution of X_1 on Ω, denoted by $p(x_1)$;

(b) the transitions $(p(x_i|x_{i-1}))_{i\geq 2}$ verifying $p(x_i|x_{i-1})=0$ for $x_i=x_{i-1}$;

(c) q a probability distribution on the set of natural numbers N, and

(d) distributions $p(y_i|x_i)$ on R.

Otherwise, let us consider a general TMC $T=(X,U,Y)$, with each U_i taking its values in $N^*=N-\{0\}$. The general form if its distribution is then defined by $p(t_1)=p(x_1,u_1,y_1)$ and the transitions $p(t_{n+1}|t_n)=p(x_{n+1},u_{n+1},y_{n+1}|x_n,u_n,y_n)$. These transitions can be detailed in different ways; in the following, we will consider them as being written

$$p(x_{n+1},u_{n+1},y_{n+1}|x_n,u_n,y_n)= p(x_{n+1}|x_n,u_n,y_n)\times p(u_{n+1}|x_n,u_n,y_n,x_{n+1})\times p(y_{n+1}|x_n,u_n,y_n,x_{n+1},u_{n+1}) \quad (1)$$

Then HSMC-IN defined by (a)-(d) can be seen as a TMC defined by $p(t_1)=p(x_1,u_1,y_1)$ and (1), where (d is the Dirac's mass):

$$p(x_{n+1}|x_n,u_n,y_n)=p(x_{n+1}|x_n,u_n)=d(x_n) \text{ if } u_n>1, \text{ and } p(x_{n+1}|x_n) \text{ if } u_n=1; \quad (2);$$

$$p(u_{n+1}|x_{n+1},x_n,u_n,y_n)=p(u_{n+1}|u_n,x_n)=d(u_n-1) \text{ if } u_n>1, \text{ and } q(u_{n+1}) \text{ if } u_n=1; \quad (3);$$

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

Finally, we have a particular TMC where (1) reduces to

$$p(x_{n+1},u_{n+1},y_{n+1}|x_n,u_n,y_n)= p(x_{n+1}|x_n,u_n)\ p(u_{n+1}|x_n,u_n)\ p(y_{n+1}|x_{n+1}) \quad (5)$$

with $p(x_{n+1}|x_n,u_n)$, $p(u_{n+1}|x_n,u_n)$, and $p(y_{n+1}|x_{n+1})$ given by (2), (3), and (4), respectively. Then the "backward" probabilities $\beta_i(x_i,u_i)=p(y_{i+1},...y_n|x_i,u_i)$, needed to different useful computations, are recursively calculated by

$$\beta_n(x_n,u_n)=1, \text{ and } \beta_i(x_i,u_i)= \sum_{u_{i+1},x_{i+1}}\beta_{i+1}(x_{i+1},u_{i+1})p(x_{i+1}|x_i,u_i)p(u_{i+1}|x_i,u_i)p(y_{i+1}|x_{i+1}) \quad (6)$$

for $1\leq i\leq n-1$

Of course, the sum in (6) is particular because of (2) and (3).
In order to simplify notations, let us put $V=(X,U)$. Thus $v_i=(x_i,u_i)$ for each $1\leq i\leq n$. In particular, (6) is written

$$\beta_n(v_n)=1, \text{ and } \beta_i(v_i)=\sum_{v_{i+1}}\beta_{i+1}(v_{i+1})p(v_{i+1}|v_i)p(y_{i+1}|v_{i+1}) \quad (7)$$

for $1\leq i\leq n-1$

(7) Looks like the formulas of a very classical hidden Markov chain (V,Y); however, let us remark that the distributions $p(y_{i+1}|v_{i+1})=p(y_{i+1}|x_{i+1},u_{i+1})$ verify

$$p(y_{i+1}|x_{i+1},u_{i+1})=p(y_{i+1}|x_{i+1}) \quad (8)$$

which means that a same noise distribution can remain valid for different classes v_{i+1}, which is not usual in classical models.
However, once it has been noticed that (V,Y) is a hidden Markov chain, the "forward" probabilities $\alpha_i(v_i)=p(v_i,y_1,...,y_i)$ can be computed recursively by

$$\alpha_i(v_1)=p(v_1,y_1), \text{ and } a_{i+1}(v_{+i})=\sum_{v_i}\alpha_i(v_i)p(v_{i+1}|v_i)p(y_{i+1}|v_{i+1}) \quad (9)$$

for $1+1\leq i\leq n$.

Finally, backward and forward probabilities can be classically used to calculate:

$$p(v_{i+1}|v_i, y) = \frac{\beta_{i+1}(v_{i+1})}{\beta_i(v_i)} \tag{10}$$

$$p(v_i|y) = \frac{\alpha_i(v_i)\beta_i(v_i)}{\sum_{v'_i} \alpha_i(v'_i)\beta_i(v'_i)} \tag{11}$$

$$p(v_i, v_{i+1}|y) = \frac{\alpha_i(v_i)p(v_{i+1}|v_i)p(y_{i+1}|v_{i+1})\beta_{i+1}(v_{i+1})}{\sum_{v'_i, v'_{i+1}} \alpha_i(v_i)p(v_{i+1}|v_i)p(y_{i+1}|v_{i+1})\beta_{i+1}(v_{i+1})} \tag{12}$$

Therefore we have a first TMC $T = (X, U, Y)$ which models the fact that (X, Y) is an HSMC-IN.

Remark 1

Let us briefly mention that, as described in [11], each equation among (2)-(4) can be extended and thus the present viewpoint representing HSMC-IN as particular TMC enables one to propose numerous generalizations of HSMC-IN. For example, let us replace in equation (4) $p(y_{n+1}|x_{n+1})$ by $p(y_{n+1}|u_n, x_{n+1})$. Such an extension means that the distribution of the noise at $n+1$ also depends on the residual sojourn duration u_n. This can be understood intuitively; in fact, when this duration is large we are far away from the boundary among two different classes, and when $u_n = 1$, we know that the class has just changed ($x_{n+1} \neq x_n$). Thus replacing $p(y_{n+1}|x_{n+1})$ by $p(y_{n+1}|u_n, x_{n+1})$ enables one to model the fact that on boundaries the distribution of the noise can be different from its distribution "inside" of a given class. Then the modelling of the non stationarity of X proposed in this paper can still be extended to these different generalizations of the HSMC-IN model.

3. HIDDEN EVIDENTIAL MARKOV CHAIN

Let us consider the HMC-IN $T = (V, Y)$ above and let us consider that this HMC-IN is not stationary. In other words, the distribution $p(v_i, y_i, v_{i+1}, y_{i+1})$ depends on $1 \le i \le n-1$. Furthermore, let us assume that this non stationarity is due to the non-stationarity of V, which means that $p(y_i|v_i)$ does not depend on $1 \le i \le n$. Such a situation has been studied in [8] by the use of Dempster-Shafer theory of evidence. By replacing the non stationary prior distribution of V by evidential priors as explained in [7, 8], we necessarily replace the non stationary distribution of $V = (X, U)$ by some particular evidential priors. As we are going to specify, such a replacement amounts to introducing in the HMC-IN (V, Y) a third random chain $V' = (V'_1, ..., V'_n)$, where each V'_i takes its values in a finite set $\Delta' = \{\lambda_1, ..., \lambda_m\}$, and leads to the consideration of the TMC (V, V', Y). Finally a non-stationary HMC-IN (V, Y) will be replaced by a stationary TMC (V, V', Y).

Let us specify with some more details the use of the theory of evidence. Each V_i taking its values in $\Delta = \{\delta_1 ..., \delta_r\}$, let us denote by $[\mathrm{P}(\Delta)]$ the power set of Δ. We will consider the so-called "evidential Markov chain" (EMC), denoted by m, which verifies:

(i) m is defined on $[\mathrm{P}(\Delta)]^n$ and takes its values in $[0,1]$;

(ii) $\sum_{A \in [\mathrm{P}(\Delta)]^n} m(A) = 1$ and $m(A_1 \times ... \times A_n) = 0$ if one at least among A_1, .., A_n is $\emptyset$;

(iii) m is of the "Markovian" form : $m(A_1, A_2, ..., A_n) = m(A_1)m(A_2|A_1)...m(A_n|A_{n-1})$.

We see how an EMC extends a Markov chain; in fact, when $m(A_1 \times ... \times A_n) = 0$ if one at least among A_1, .., A_n is not a singleton, m is a Markov chain on Δ^n.

The approach proposed in [8] is based on the two following points:

1. The posterior distribution $p(v|y)$ of an HMC-IN, which is needed to Bayesian restoration, can be seen as the DS fusion of the prior Markov distribution $p(v) = p(v_1)p(v_2|v_1)...p(v_n|v_{n-1})$ of V with the probability $q^y(v_1, ..., v_n) \propto p(y_1|v_1)...p(y_n|v_n)$ (where "$\propto$" means "proportional to") defined on Δ^n by $Y = y$;

2. When the distribution $p(v)$ of the Markov chain V is incompletely known, it can be replaced by an EMC obtained from $p(v)$, whose aim is to model the lack of precise knowledge of $p(v)$. It can be fused with q^y using DS fusion. The result of the latter fusion is a probability distribution on Δ^n and, although it is not necessarily a Markov distribution, it can be used to perform Bayesian restorations. Indeed, the latter feasibility is due to the fact that the fused distribution is a triplet Markov chain [8]. To be more precise, $p(v)$ is replaced by an EMC m such that $m(A_1 \times ... \times A_n) \neq 0$ only on $(A_1 \times ... \times A_n)$ such that each A_i is in $\Delta' = \{\{\delta_1\}..., \{\delta_r\}, \Delta\}$. Thus a homogeneous Markov chain on Δ^n is defined by r^2 parameters, and a

homogeneous EMC on $\{\{\delta_1\},...,\{\delta_r\},\Delta\}$ is defined by $(r+1)^2$ parameters, which means that EMC considered here is not so much more complex than the corresponding Markov chain.

Finally, let us specify how the DS fusion (DS fusion) allows one to calculate different quantities of interest.

Let us consider the DS fusion of the probability $q^y(v_1,...,v_n) \propto p(y_1|v_1)...p(y_n|v_n)$ with $m(A_1 \times ... \times A_n)$ defined above. We have:

$$(q^y \oplus m)(v_1,...,v_n) \propto \sum_{(v_1,...,v_n)\in(v'_1,...,v'_n)} q^y(v_1,...,v_n)m(v'_1,...,v'_n), \qquad (13)$$

where each v_i is in $\Delta = \{\delta_1...,\delta_r\}$, each v'_i is in $\Delta' = \{\{\delta_1\},...,\{\delta_r\},\Delta\}$, and $(v_1,...,v_n) \in (v'_1,...,v'_n)$ means that $v_1 \in v'_1$, .., $v_n \in v'$. Let us insist on the fact that in (13) $(v_1,...,v_n)$ and the sum is taken over $(v'_1,...,v'_n)$ such that $(v_1,...,v_n) \in (v'_1,...,v'_n)$.

Important is then to notice that $(v_1,...,v_n)m(v'_1,...,v'_n)$ in (13) defines a Markov chain, and the DS fusion is the calculation of some marginal distribution of this Markov chain.

To be more precise, let us consider $n-1$ functions $f_2^{y_1,y_2}$, $f_3^{y_3}$,..., $f_n^{y_n}$ defined on $\Delta^2 \times (\Delta')^2$ by:

$$f_2^{y_1,y_2}(v_1,v_2,v'_1,v'_2) = 1_{[v_1\in v'_1, v_2\in v'_2]}\, p(y_1|v_1)p(y_2|v_2)m(v'_1,v'_2) \qquad (14)$$

and

$$f_{i+1}^{y_{i+1}}(v_i,v_{i+1},v'_i,v'_{i+1}\ y_{i+1}) = 1_{[v_i\in v'_i, v_{i+1}\in v'_{i+1}]}\, p(y_{i+1}|v_{i+1})m(v'_{i+1}|v'_i) \qquad (15)$$

for $2 \le i \le n-1$.

On the one hand, the product $f_2^{y_1,y_2}(v_1,v_2,v'_1,v'_2) \times f_3^{y_3}(v_2,v_3,v'_2,v'_3) \times ... \times f_n^{y_3}(v_{n-1},v_n,v'_{n-1},v'_n)$ defines (for fixed y) a Markov chain on $\Delta^n \times (\Delta')^n$ with calculable $p(v_{i+1},v'_{i+1}|v_i,v'_i,y)$, $p(v_i,v'_i,v_{i+1},v'_{i+1}|y)$ and $p(v_i,v'_i|y)$ (see Lemma below). On the other hand, we see that this product is proportional to $q^y(v_1,...,v_n)m(v'_1,...,v'_n)$ in (13). These two results show that the result of the DS fusion $(q^y \oplus m)(v_1,...,v_n)$ in (13) is the marginal distribution (the distribution of $(V_1,...,V_n)$) of the distribution of the Markov chain $((V_1,V'_1),...,(V_n,V'_n))$ defined by the functions $f_2^{y_1,y_2}$, $f_3^{y_3}$,..., $f_n^{y_n}$.

Finally, we can state the following result:

Proposition 1

Let $(V,Y) = (V_1,Y_1,...,V_n,Y_n)$ be a classical HMC-IN, each V_i taking its values in $\Delta = \{\delta_1...,\delta_r\}$ and each Y_i taking its values in R. Therefore, $p(v) = p(v_1)p(v_2|v_1)...p(v_n|v_{n-1})$, and $p(y|v) = p(y_1|v_1)...p(y_n|v_n)$.

Let $q^y(v_1,...,v_n) \propto p(y_1|v_1)...p(y_n|v_n)$ be the probability defined on Δ^n by $Y=y$, and let m be an EMC on $\Delta' = \{\{\delta_1\},...,\{\delta_r\},\Delta\}$ extending the prior distribution $p(v) = p(v_1)p(v_2|v_1)...p(v_n|v_{n-1})$.

Then the probability $q^y \oplus m$ given by (13), which extends the posterior probability $p(v|y)$ and which is not necessarily of Markovian form, is a marginal probability of a finite Markov chain. As a consequence, $q^y \oplus m$ can be used to perform different restorations like MPM.

This result has been successfully applied to the case of non-stationary hidden $V = (V_1,...,V_n)$, see [8].

4. HIDDEN EVIDENTIAL SEMI-MARKOV CHAIN

Let us return to the situation of the previous section, with the hidden semi-Markov chain $X = (X_1,...,X_n)$, $Y = (Y_1,...,Y_n)$, where each X_i takes its values in $\Omega = \{\omega_1,...,\omega_k\}$ and each Y_i takes its values in R. As specified above, the HSMC-IN (X,Y) can also be seen as a TMC (X,U,Y), which also is an HMC-IN (V,Y), with $V = (X,U)$. Let us consider the case of non-stationary semi-Markov chain $X = (X_1,...,X_n)$, which means that $V = (X,U)$ also is non stationary. Thus this non-stationarity of $V = (X,U)$ can be managed as specified in Proposition 1. Knowing that $p(v|y)$ is workable as being marginal distribution of the Markov chain $p(v,v'|y)$, we see that $p(x|y)$ is workable as being a marginal distribution of $p(v|y) = p(x,u|y)$, and thus a marginal distribution of the Markov chain $p(x,u,v'|y)$. As an example, let us detail how the posterior marginal distributions $p(x_i|y)$, enabling one to perform the MPM segmentation, are computable.

Proposition 2

Let X be a non-stationary semi-Markov chain, (X,Y) a non-stationary HSMC-IN, (X,U,Y) the corresponding non-stationary TMC, and (X,U,V',Y) a stationary Markov chain obtained from $(X,U,Y)=(V,Y)$ as specified above.

Then $p(x_i|y)$ is computable in two steps:

(i) $p(x_i,u_i|y)=p(v_i,v'_i|y)=\sum_{v'_i}p(v_i,v'_i|y)$, and

(ii) $p(x_i|y)=\sum_{u_i}p(x_i,u_i|y)$.

Finally, we arrived at a TMC $(X,U,V',Y)=(X,W,Y)$, where the third random process $W=(U,V')$ model the fact that X is semi-Markov on the one hand, and the fact that X is non-stationary, on the other hand.

Remark 2

We pointed out in Remark 1 above how the model $T=(X,U,V',Y)$ can be extended by considering different extensions offered by (2)-(5). However, all over the paper we kept the notations "HMC-IN" and HSMC-IN" to clearly specify that in all models considered the random variables Y_1, .., Y_n remain independent conditionally on X, which means that the "noise" is of a quite simple – and undoubtedly too simple in numerous situations - form. However, extending this simple case to correlated noise is quite straightforward in the Gaussian noise case. When the noise is correlated and not necessarily Gaussian, it is then possible to use the theory of Copulas, as proposed in the simple HMC-IN case in [3].

Lemma

Let $H=(H_1,...,H_n)$ be a random chain, each H_i taking its values in a finite set.

Then H is a Markov chain if and only if there exist positive functions s_1, .., s_{n-1} such that $p(h_1,...,h_n)\propto s_1(h_1,h_2)...s_{n-1}(h_{n-1},h_n)$.

If H is a Markov chain, the transitions and the marginal distributions are given by

$$p(h_{i+1}|h_i)=q_{i+1}(h_i,h_{i+1})\beta_{i+1}(h_{i+1})/\beta_i(h_i),$$

$$p(h_i)=\alpha_i(h_i)\beta_i(h_i)/\sum_{h_i'}\alpha_i(h_i')\beta_i(h_i'),$$

with

$\alpha_1(h_1)=1$, $\alpha_{i+1}(h_{i+1})=\sum_{h_{i+1}}q_{i+1}(h_i,h_{i+1})\alpha_i(h_i)$, and

$\beta_n(h_n)=1$, $\beta_i(h_i)=\sum_{h_{i+1}}q_{i+1}(h_i,h_{i+1})\beta_{i+1}(h_{i+1})$.

5. PARAMETER ESTIMATION

Let us briefly mention how the parameter estimation can be performed by two general methods "Expectation-Maximization" (EM), and "Iterative Conditional Estimation" (ICE). To simplify things, let us assume that the variables U_i take their values in a finite set. Moreover, we assume that the TMC $T=(V,V',Y)$ is "Gaussian" in that $p(y|v,v')$ are Gaussian. Assuming T stationary in that neither $p(v_i,v'_i,v_{i+1},v'_{i+1})$ nor $p(y_i|v_i,v'_i)=p(y_i|x_i)$ depend on i, we have to estimate the finite distribution $p(v_i,v'_i,v_{i+1},v'_{i+1})$ on $\Delta^2\times(\Delta')^2$, and, for k classes, k means and k variances of the k Gaussian distributions $p(y_i|x_i)$ on R. Let us put $\Delta^2\times(\Delta')^2=\{1,...,M\}$, $\tau_j=p[(v_i,v'_i,v_{i+1},v'_{i+1})=j]$ and $\tau=(\tau_1,...,\tau_M)$. Otherwise, let $\mu=(\mu_1,...,\mu_k)$ be the means and $\sigma^2=(\sigma_1^2,...,\sigma_k^2)$ the variances. Then the parameters to be estimated are $\theta=(\tau,\mu,\sigma^2)$. Both EM and ICE methods are iterative: taking an initial value θ^0, the problem is to compute θ^{m+1} from the observation $Y=y=(y_1,...,y_n)$ and θ^m.

EM method runs as follows:

(i) use θ^m to calculate $\tau_j^{i,m}=p[(v_i,v'_i,v_{i+1},v'_{i+1})=j|y,\theta^m]$ (using the Lemma) and $p_q^{i,m}=p[x_i=\omega_q|y,\theta^m]$ (deduced from $\tau_j^{i,m}$);

(ii) put $\tau_j^{m+1}=\dfrac{\tau_j^{1,m}+..\tau_j^{n-1,m}}{n-1}$ for $1\le j\le M$,

$\mu_q^{m+1}=\dfrac{y_1p_q^{1,m}+...+y_np_q^{n,m}}{p_q^{1,m}+...+p_q^{n,m}}$, and

$\sigma_q^{2,m+1}=\dfrac{(y_1-\mu_q^{m+1})^2p_q^{1,m}+...+(y_n-\mu_q^{m+1})^2p_q^{n,m}}{p_q^{1,m}+...+p_q^{n,m}}$ for $1\le q\le k$,

which gives θ^{m+1}.

In ICE τ_j^{m+1} is obtained as in EM and, to obtain μ_q^{m+1} and $\sigma_q^{2,m+1}$, one simulates a realisations $x^1=(x_1^1,...,x_n^1)$, .., $x^a=(x_1^a,...,x_n^a)$ of $X=(X_1,...,X_n)$ according to $p(x|y,\theta^m)$ (recall that the distribution of $(V,V')=(X,U,V')$ conditional on Y is a Markov distribution with computable transitions $p(v_{i+1}, v'_{i+1} | v_i, v'_i, y, \theta^m)$, and thus simulating $(V, V') = (X, U, V')$ gives $X = x$). For $1 \le l \le a$ and $1 \le q \le k$, let $x^{l,q} = (x_1^{l,q}, ..., x_{n'}^{l,q})$ be the sub-sample of $x^l = (x_1^l, ..., x_n^l)$ such that $x_1^{l,q} = \omega_q$, .., $x_{n'}^{l,q} = \omega_q$. This sub-sample is then used to estimate, by the classical estimators, the mean $\mu_q^{m+1,l}$, and the variance $\sigma_q^{2,m+1,l}$. Then $\mu_q^{m+1} = \dfrac{\mu_q^{m+1,1} + ... + \mu_q^{m+1,a}}{a}$, and $\sigma_q^{2,m+1} = \dfrac{\sigma_q^{2,m+1,1} + ..\sigma_q^{2,m+1,a}}{a}$. In practice, one often takes $a = 1$.

6. CONCLUSION AND PERSPECTIVES

We dealt in this paper with unsupervised segmentation of the hidden non stationary semi-Markov chains. The main tool used was the triplet Markov chain model, which has been obtained by introduction of two auxiliary chains. The first auxiliary chain modelled the semi-Markovianity, and the second one modelled the non stationarity.

As perspectives, we can mention different possibilities of further extensions. In particular, more complex noises can be introduced via the Copula theory, as described in the simple hidden Markov chains case in [3]. Otherwise, the mono sensor case considered in this paper (observations in R) can be extended to the multisensor one [12].

7. REFERENCES

[1] B. Ait-el-Fquih and F. Desbouvries, Kalman filtering for triplet Markov chains : applications and extensions, Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP 05), March 19-23, Philadelphia, USA, 2005.

[2] A. Appriou, Probabilités et incertitude en fusion de données multisenseurs, Revue Scientifique et Technique de la Défense, No. 11, pp. 27-40, 1991.

[3] N. Brunel and W. Pieczynski, Unsupervised signal restoration using Copulas and Pairwise Markov chains, IEEE Workshop on Statistical Signal Processing (SSP 2003), Saint Louis, Missouri, September 28-October 1, 2003.

[4] F. Desbouvries and W. Pieczynski, Particle Filtering in Pairwise and Triplet Markov Chains, Proceedings of the IEEE – EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003), Grado-Gorizia, Italy, June 8-11, 2003.

[5] S. Faisan, L. Thoraval, J.-P. Armspach, F. Heitz, Hidden Semi-Markov Event Sequence Models: Application to Brain Functional MRI Sequence Analysis, IEEE Int. Conf. Image Processing, ICIP 2002, Rochester, USA, Sept. 2002.

[6] Y. Guedon and C. Cocozza-Thivent, Use of the Derin' s algorithm in hidden semi-Markov models for automatic speech recognition, International Conference on Acoustics, Speech, and Signal Processing, pp. 282-285, 1989.

[7] P. Lanchantin et W. Pieczynski, Chaînes et arbres de Markov évidentiels avec applications à la segmentation des processus non stationnaires, *Traitement du Signal*, Vol. 22, No. 1, 2005.

[8] P. Lanchantin and W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, *IEEE Trans. on Signal Processing*, Vol. 53, No. 8, 2005.

[9] W. Pieczynski, Pairwise Markov chains, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 5, pp. 634-639, 2003.

[10] W. Pieczynski, C. Hulard, and T. Veit, Triplet Markov chains in hidden signal restoration, SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece, 2002.

[11] W. Pieczynski, Chaînes Chaînes semi-markoviennes cachées et chaînes de Markov triplet, soumis à *Comptes Rendus de l'Académie des Sciences – Mathématique*, décembre 2004 *(in French)*.

[12] W. Pieczynski, Fusion de Dempster-Shafer dans les chaînes triplet partiellement de Markov, *Comptes Rendus de l'Académie des Sciences – Mathématique*, Série I, Vol. 339, Issue 11, pp. 797-802, 2004.

[13] W. Pieczynski and F. Desbouvries, On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France, May 2005.

[14] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, Princeton, 1976.

[15] S.-Z. Yu and H. Kobayashi, A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking, *Signal Processing*, Vol. 83, No. 2, pp. 235-250, 2003.