Unsupervised segmentation of random discrete data using triplet Markov chains

Mohamed Ben Mabrouk and Wojciech Pieczynski

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

Abstract. The hidden Markov chain (HMC) model is a couple of random sequences (X, Y), in which X is an unobservable Markov chain, and Y is its observable noisy version. Classically, HMC enables one to use different Bayesian restoration techniques. HMC model, which is powerful in hidden data restoration and widely used, has recently been extended to “triplet Markov chain” (TMC) model, which is obtained by adding a third chain U and considering the Markovianity of the triplet $T = (X, U, Y)$. When U is not too complex, X can still be recovered from Y. The aim of the paper is to propose a new particular TMC model, which is not more complex that the classical HMC one, and which can be more efficient in unsupervised segmentation.
Keywords: hidden Markov chains, triplet Markov chains, unsupervised segmentation, image segmentation, iterative conditional estimation.

1 Introduction

Let $X = (X_i)_{1 \le i \le i}$ and $Y = (Y_{n\,i})_{1 \le i \le i}$ two stochastic processes, where X is hidden and Y is observable. In the whole paper, each X_i takes its values in a finite set of classes $\Omega = \{\omega_1, \ldots, \omega_K\}$ and each Y_i takes its values in R. The problem of estimating X from Y, which occurs in numerous applications, can be solved with Bayesian methods once one has chosen some accurate distribution $p(x, y)$ for $Z = (X, Y)$. The hidden Markov chain (HMC) model is the simplest and most well known model. Its applications covers numerous fields, see [Koski, 2001], [Ephraim and Merhav, 2002], [Cappé *et al.*, 2005]. The HMC has been extended to triplet Markov chains model (TMC [Pieczynski *et al.*, 2002]), which is obtained by adding a third chain U and considering the Markovianity of the triplet $T = (X, U, Y)$. TMC are strictly more general than HMC [Pieczynski 2007] and present a very rich family of models ([Ait-el-Fquih

and Desbouvries, 2006], [Lanchantin and Pieczynski, 2004], [Pieczynski and Desbouvries, 2005], among others).
The aim of this paper is to present an original triplet Markov chain and to show that this new model can be of interest, with respect to the classical HMC, in unsupervised segmentation of hidden discrete signal. Its main originality, which opens new perspectives, is to take the auxiliary chain U continuous on the one hand, and to consider that the hidden states distributions are independent conditionally on U, on the other hand. The latter hypothesis, which has never been considered until now at our knowledge, makes possible Bayesian segmentation without any model approximations.

2 Classical hidden Markov chain and new triplet one

Let $X = (X_n)_{1\le n\le N}$ and $Y = (Y_n)_{1\le n\le N}$ be two stochastic processes. X is hidden and each X_n takes its values in a finite set of classes $\Omega = \{\omega_1, ..., \omega_K\}$, while Y is observed and each Y_n takes its values in the set of real numbers R. The problem is to estimate $X = x$ from $Y = y$. As we will use the following Bayesian Maximum Posterior Mode (MPM) segmentation method $\hat{s}_{MPM}(y_1, ..., y_n) = (\hat{x}_1, ..., \hat{x}_n)$, with $\hat{x}_n = \arg\max_{x_n\in\Omega} p(x_n|y)$, in the whole paper, let us concentrate on the fact that the posterior marginal distributions $p(x_n|y)$ can be calculated. In the classical hidden Markov chain (HMC) the distribution of (X,Y) is given by $p(x, y) = p(x_1)p(y_1|x_1)p(x_2|x_1)p(y_2|x_2)...p(y_n|x_n)$; then the classical "Forward" probabilities $\alpha(x_n) = p(x_n, y_1, ..., y_n)$, and the "Backward" ones $\beta(x_n) = p(y_{n+1}, ..., y_N|x_n)$, can be calculated recursively by $\alpha(x_1) = p(x_1, y_1)$; $\alpha(x_{n+1}) = \sum_{x_n\in\Omega} \alpha(x_n)p(x_{n+1}, y_{n+1}|x_n, y_n)$ and $\beta(x_N) = 1$; $\beta(x_n) = \sum_{x_{n+1}\in\Omega} \beta(x_{n+1})p(x_{n+1}, y_{n+1}|x_n, y_n)$. The marginal posterior distributions of the hidden state are given by $p(x_n|y) \propto \alpha(x_n)\beta(x_n)$, and the transition of the posterior Markov distribution $p(x|y)$ are calculated by $p(x_{n+1}|x_n, y_1, ..., y_N) \propto p(x_{n+1}, y_{n+1}|x_n, y_n)\beta(x_{n+1})$.

Considering a triplet Markov chain (TMC) consists of introducing a third stochastic process $U = (U_i)_{1 \le i \le n}$ and assuming that $T = (X, U, Y)$ is a Markov chain. As stated in Introduction, when each U_i takes its values in a finite set $\Lambda = \{\lambda_1, ..., \lambda_m\}$, different TMC T, with discrete U, have been successfully applied in different situations [Pieczynski and Desbouvries, 2005]. Here we propose a different model, in which each U_i takes its values in R. Therefore each $T_i = (X_i, U_i, Y_i)$ takes its values in $\Omega \times R^2$ and its distribution is defined by $p(t_1)$ and the transitions $p(t_{i+1} | t_i)$. In the model we propose, called "New" TMC (NTMC) in the following, these probability distributions are:

$$p(x_1, u_1, y_1) = p(u_1) p(x_1 | u_1) p(y_1 | u_1) ,$$
$$p(x_{i+1}, u_{i+1}, y_{i+1} | x_i, u_i, y_i) = p(u_{i+1} | u_i) p(x_{i+1} | u_{i+1}) p(y_{i+1} | u_{i+1}) \qquad (1)$$

Equivalently, we can say that $T = (X, U, Y)$ is a NTMC if it verifies: (i) U is a Markov chain, and (ii) $p(x, y | u) = \prod_{i=1}^{n} p(x_i | u_i) p(y_i | u_i)$.

We notice that in NTMC X_i and Y_i are independent conditionally on U_i.

The difference between the classical HMC-IN and the NTMC is also seen from their dependence graphs, presented in Fig. 1.

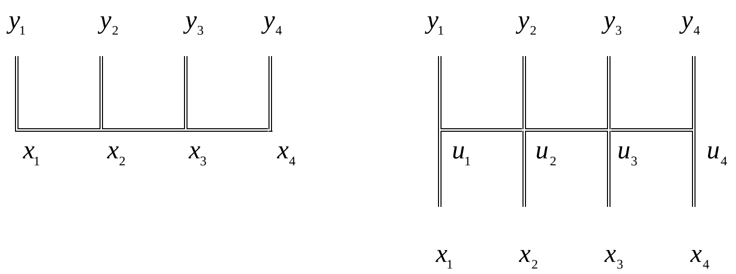


Fig. 1. Dependence graphs for HMC (left) and NTMC (right).

The problem is the same as above: estimate $X = x$ from $Y = y$. We need $p(x_i | y)$ and thus we are going to show that these posterior marginal distributions are calculable in NTMC. We can state the following result:

Proposition

Let $T = (X, U, Y)$ be a NTMC verifying (1). If $(U, Y) = (U_i, Y_i)_{1 \le i \le n}$ is Gaussian, then:
(i) the Gaussian distribution $p(u_i | y)$ is computable with the number of operations linear in n;
(ii) $p(x_i | y) = \int_R p(x_i | u_i) p(u_i | y) du_i$.

Proof. As $(U, Y) = ((U_i, Y_i))_{1 \le i \le n}$ is a classical Gaussian hidden Markov chain, the fact that the Gaussian distribution $p(u_i | y)$ is computable with the number of operations linear in n is classical result ([Rauch, Tung and Striebel, 1965], [Ephraim and Merhav, 2002]. To show $p(x_i | y) = \int_R p(x_i | u_i) p(u_i | y) du_i$, we use the fact that Y and X are independent conditionally on U, which means that $p(x, u | y) = p(x | u) p(u | y)$. The last equality implies that $p(x_i, u_i | y) = p(x_i | u_i) p(u_i | y)$, which leads to (ii) and ends the proof.

Let us remark that in spite of its simplicity, the NTMC is very different from the classical HMC. In fact, it is possible to show, using the results presented in [Pieczynski, 2007], that in NTMC neither X, nor even (X, Y), is necessarily a Markov chain.

3. Learning the NTMC

Let us consider a stationary NTMC $T = (X, U, Y)$, which means that $p(t_i, t_{i+1})$ does not depend on $i = 1$, …, $n-1$. We learn such a NTMC with the general "Iterative Conditional Estimation" (ICE) method [Fjørtoft *et al.*, 2003]. ICE is based on the following principle. Let $\theta = (\theta_1, ..., \theta_m)$ be the vector of all real parameters defining $p(t)$ and let $\hat{\theta}(t)$ an estimator of θ defined from the complete data $t = (x, u, y)$. ICE is an iterative method consisting on:
(i) initialize θ^0;

(ii) compute $\theta_i^{q+1} = E[\hat{\theta}_i(X,U,Y)|Y = y,\theta^q)$ for the components θ_i for which this computation is workable;

(iii) for other components θ_i, simulate (x_1^q,u_1^q), …, (x_l^q,u_l^q) according to $p(x,u|y,\theta^q)$ and put $\theta_i^{q+1} = \dfrac{\hat{\theta}(x_1^q,u_1^q,y)+...+\hat{\theta}(x_l^q,u_{l1}^q,y)}{j}$.

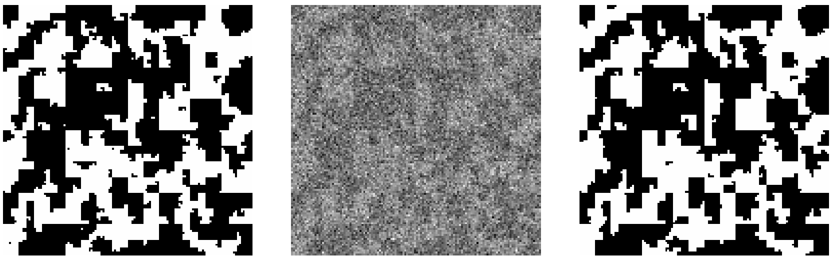
We see that ICE is applicable under very slight two hypotheses: existence of an estimator $\hat{\theta}(t)$ from the complete data, and the ability of simulating (X,U) according to $p(x,u|y)$. The first hypothesis is not really a constraint because if we are not able to estimate θ from complete data (x,u,y), there is no point in searching an estimator from incomplete ones y. The second hypothesis is also verified in the NTMC; in fact, we have $p(x,u|y) = p(x|u)p(u|y)$ and both $p(u|y)$, $p(x|u)$ can be sampled.

4. Experiments

Let us consider the following NTMC $T = (X,U,Y)$. Each X_i takes its values in $\Omega = \{\omega_1,\omega_2\}$, and each U_i and Y_i are in R. The distribution of $T = (X,U,Y)$ is then defined by $p(t_1,t_2) = p(u_1,u_2,y_1,y_2)p(x_1|u_1)p(x_2|u_2)$, where $p(u_1,u_2,y_1,y_2)$ is a Gaussian distribution on R^4 verifying $p(u_1,u_2,y_1,y_2) =$ $= p(u_1,u_2)p(y_1|u_1)p(y_2|u_2)$.

In experiments below we will take $p(u_1,u_2)$ Gaussian with both means equal to zero, and correlation equal to r. Then $p(y_1|u_1)$ (which is equal to $p(y_2|u_2)$) is defined by its mean m, its variance σ^2, and the correlation c between U_1 and Y_1. Otherwise, we will assume that the proportions of two classes are equal, and we will take $p(x_1 = \omega_1|u_1) = \dfrac{\exp[u_1]}{1+\exp[u_1]}$, $p(x_1 = \omega_2|u_1) = \dfrac{1}{1+\exp[u_1]}$. Thus we have four parameters $\theta = (r,c,m,\sigma^2)$. We see how NTMC is different from the classical stationary HMC, in which $p(x_1,x_2,y_1,y_2) = p(x_1,x_2)p(y_1|x_1)p(y_2|x_2)$ is defined by $p(x_1,x_2)$, which is a probability on $\Omega^2 = \{\omega_1,\omega_2\}^2$, and two

Gaussian densities $p(y_1|x_1=\omega_1)$, $p(y_1|x_1=\omega_2)$ on R. Thus in HMC we have one real parameter for $p(x_1,x_2)$ (given that the proportions of the two classes are equal) and four parameters (two means and two variances) for $p(y_1|x_1)$.
Finally, we have four parameters for the NTMC considered and five parameters for the HMC one.
We present below three series of experiments. In all of them, realizations of X and Y are represented as images, classically obtained by the Hilbert-Peano transformation of a mono dimensional set of indices to a bi-dimensional set of pixels [Fjørtoft *et al.*, 2003].
In the first series, we simulate a stationary HMC, where $p(x_1=\omega_1, x_2=\omega_1) = = p(x_1=\omega_2, x_2=\omega_2) = 0.49$, $p(y_1|x_1=\omega_1)$ is Gaussian $N(0,1)$, and $p(y_1|x_1=\omega_2)$ is Gaussian $N(1,1)$. NTMC based segmentation gives the error ratio of 3.80%, while the HMC based segmentation gives the error ratio of 3.07%. This experiment, and other similar ones we performed, indicates that the NTMC is robust with respect to HMC data. The realizations $X=x$, $=y$, and the NTMC based segmentation results are presented in Fig. 1.



$X=x$ $\qquad$ $Y=y$ $\qquad$ NTMC based $\hat{x}=MPM(y)$

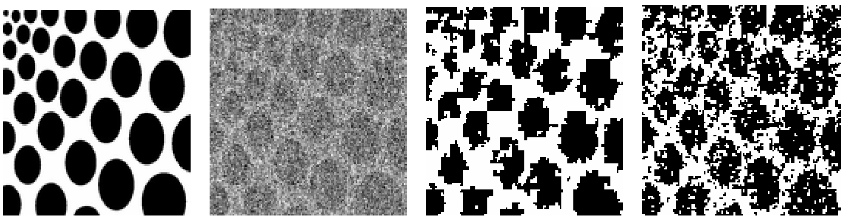
Fig. 1. Class image $X=x$, observed image $Y=y$, and NTMC based segmentation (error ratio 3.80%). HMC based segmentation provides the error ratio of 3.07%

In the second experiment, we consider the same Markov chain realization $X=x$, which is now corrupted with correlated noise. The noise is obtained by the following spatial moving average: taking (U_n) independent and Gaussian $N(0,1)$, we consider $V_n = \alpha U_n + \beta \sum_{j\in A(n)} U_j$, where $A(n)$ is neighbourhood of pixel n. For a given form of neighbourhood, α and β are chosen in such a way that the marginal distributions $p(y_1 | x_1 = \omega_1)$, $p(y_1 | x_1 = \omega_2)$ are the same as above: $N(0,1)$ and $N(1,1)$, respectively. According to the results presented in Fig. 2, we see that NTMC is more robust with respect of the noise correlation.

$Y = y$ NTMC (8.95%) HMC (19.09%)

Fig. 2. NTC and HMC unsupervised segmentation s with error rates Correlated noise obtained with $\alpha = 0.85$, $\beta = 0.18$ (correlation of 0.45)

Finally, in the third experiment we consider a hand-written image corrupted by the same correlated noise. Therefore the data are neither HMC not NTMC ones and we see, according to Fig. 3, that NTMC works better than the classical HMC.



$X = x$ $Y = y$ NTMC (14.49%) HMC (20.12%)

Fig. 3. NTC and HMC unsupervised segmentation s with error rates.

5. Conclusion

We proposed in this paper a new triplet Markov chain (NTMC) allowing one to deal with the unsupervised statistical data segmentation. Although not more complex, the new model is very different from the classical hidden Markov chain (X, Y). Iin fact, in NTMC neither the searched X nor the observable Y is, in general, a Markov chain. The number of parameters to be estimated and the computer time needed are similar in both HMC and NTMC, while the latter better resist to the correlation of the noise.

References

[Ait-el-Fquih and Desbouvries, 2006] B. Ait-el-Fquih and F. Desbouvries, Kalman filtering in triplet Markov chains, *IEEE Trans. on Signal Processing,* Vol. 54, No. 8, pp. 2957-2963, 2006.

[Cappé *et al.*, 2005] O. Cappé, E. Moulines, T. Ryden, *Inference in hidden Markov models,* Springer, Series in Statistics, 2005.

[Ephraim and Merhav, 2002] Y. Ephraim and N. Merhav, Hidden Markov processes, *IEEE Trans. on Information Theory,* Vol. 48, No. 6, pp. 1518-1569, 2002.

[Fjørtoft *et al.*, 2003] R. Fjørtoft, Y. Delignon, W. Pieczynski, M. Sigelle, and F. Tupin, Unsupervised segmentation of radar images using hidden Markov chains and hidden Markov random fields, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 41, No. 3, pp. 675-686, 2003.

[Koski, 2001] T. Koski, Hidden Markov models for bioinformatics, Kluwer Academic Publishers, 2001.

[Lanchantin and Pieczynski, 2004] P. Lanchantin and W. Pieczynski, Unsupervised non stationary image segmentation using triplet Markov chains, Advanced Concepts for Intelligent Vision Systems (ACIVS 04), Aug. 31-Sept. 3, Brussels, Belgium, 2004.

[Pieczynski *et al.*, 2002] 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.

[Pieczynski, 2007] W. Pieczynski, Multisensor triplet Markov chains and theory of evidence, *International Journal of Approximate Reasoning*, accepted, June 2006.

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

[Rauch, Tung and Striebel, 1965] H. E. Rauch, F. Tung, and C. T. Striebel, Maximum likelihood estimates of linear dynamic systems, *AIAA J.*, Vol. 3, No. 8, pp. 1445–1450, August 1965.