Short Paper

Unified Representation of Sets of Heterogeneous Markov Transition Matrices

Mohamed El Yazid Boudaren and Wojciech Pieczynski

***Abstract*—Markov chains are very efficient models and have been extensively applied in a wide range of fields covering queuing theory, signal processing, performance evaluation, time series, and finance. For discrete finite first-order Markov chains, which are among the most used models of this family, the transition matrix can be seen as the model parameter, since it encompasses the set of probabilities governing the system state. Estimating such a matrix is, however, not an easy task due to possible opposing expert reports or variability of conditions under which the estimation process is carried out. In this paper, we propose an original approach to infer a consensus transition matrix, defined in accordance with the theory of evidence, from a family of data samples or transition matrices. To validate our method, experiments are conducted on nonstationary label images and daily rainfall data. The obtained results confirm the interest of the proposed evidential modeling with respect to the standard Bayesian one.**

***Index Terms*—Hidden Markov chains, Markov chains, model selection, theory of evidence.**

I. Introduction

Let $X = (X_1, ..., X_N)$ be a stochastic process where each X_n takes its values in a discrete finite state set $\Omega = \{\omega_1, ..., \omega_K\}$. According to the first-order Markov chain model, the joint distribution of X is given by

$$p(x) = p(x_1) \prod_{n=2}^{N} p(x_n | x_{n-1}). \quad (1)$$

When the probabilities $p(x_n|x_{n-1})$ are independent on n, they are given by a unique transition matrix $A = [a_{ij}]_{i,j=1}^{K}$, where $a_{ij} = p(x_n = \omega_j | x_{n-1} = \omega_i)$, whereas $p(x_1)$ is given by a vector Π, where $\pi_i = p(x_1 = \omega_i)$. Hence, the estimation of the transition matrix A is crucial to accurately monitor the system under consideration. For this purpose, many statistical approaches may be used [1]–[3]. However, such an estimation is not always an easy task. In particular, we can mention the following difficulties.

1) The transition probabilities may be fluctuating, and any estimation under such conditions may lead to a provisional transition matrix that could be unsuitable for future use.
2) There may be no consensus between experts about the meaning of states themselves. Thus, expertise reports may be different even if the same data are considered.
3) A set of transition matrices may be provided by independent experts. In such cases, where no data are available, it is crucial to derive a consensus transition matrix.

Manuscript received December 24, 2014; revised March 27, 2015 and May 14, 2015; accepted July 16, 2015. Date of publication July 24, 2015; date of current version March 29, 2016.

M. E. Y. Boudaren is with Ecole Militaire Polytechnique, Algiers 16111, Algeria (e-mail: boudaren@gmail.com).

W. Pieczynski is with the Department of CITI, Institut Mines-Télécom, 91011 Evry, France (e-mail: wojciech.pieczynski@telecom-sudparis.eu).

Digital Object Identifier 10.1109/TFUZZ.2015.2460740

In this paper, we propose to use the theory of evidence (TE) [4]–[9] to take into account these different sources of possible lack of precision. Markov chains, hidden or not, are widely used, and the TE also is. However, in spite of the recognized efficiency of these two tools, there are only few research studies that deal with them simultaneously. Let us summarize some results linked with such works. Ramasso and Denoeux use belief functions to take into account partial knowledge about hidden states of a hidden Markov model [10]. Hidden evidential Markov chains (HEMCs) are used in [11] and [12] to deal with possible nonstationarities of the hidden chain, with application to unsupervised image segmentation. In [13], authors use HEMCs for power quality disturbance classification. Decision problems are also dealt with using HEMCs in [14]. Other applications of HEMCs include particle filtering [15], prognostics [16], and fault diagnosis [17]. Extensions of such models to a multisensor case with possibly correlated noise are proposed in [18]. Let us also briefly mention the use of TE in the hidden Markov field context, where Bayesian processing is somewhat different from the one used in the Markov chain context, with applications to image processing [19]–[22]. However, compared with the volume of publications on HMCs and TE separately, these different results appear as quite marginal, and thus, there are wide perspectives for further developments.

The idea presented in this paper is to remedy to different problems mentioned above by extending the frame of discernment $\Omega = \{\omega_1, ..., \omega_K\}$ of the "hard" Markov chain to $P(\Omega) = \{\emptyset, \{\omega_1\}, ..., \Omega\}$ to consider compound hypotheses, as in the HEMC, typically to take into account the uncertainty and the unreliability associated with such hypotheses. Hence, the corresponding transition matrix A^* is defined on $P(\Omega)^2$, and at each site n, the system is modeled through a random variable U_n in $P(\Omega)$. Furthermore, if the system state, denoted X_n, is accessible (or observable) at some date n, we have necessarily $X_n \in U_n$. The conventional Markov chain can then be perceived as a particular case, where $U_n = X_n$ for each $1 \leq n \leq N$. To summarize, one can consider that observing X_n at date n does not fully describe the system state at that date. It is to say that X_n can be seen as a noisy version of U_n. For this purpose, we propose a genuine approach based on hidden Markov chains, according to which the exclusive states X_n are considered independent of each other conditional on $U = u$, whereas each X_n is emitted by its associated U_n in accordance with the well-known Pignistic probability transform [5].

The remainder of this paper is organized as follows. Section II recalls some concepts of theory of evidence and introduces "evidential Markov chains" (EMCs). Section III describes the proposed approach to infer a consensus transition matrix from a family of data samples or transition matrices. Experimental results, obtained on label images and daily rainfall data, are provided in Section IV. Section V ends the paper.

II. Markov Chains and Theory of Evidence

Let $\Omega = \{\omega_1, ..., \omega_K\}$ be a finite state set, called in the TE language "frame of discernment," and let $P(\Omega) = \{A_1, ..., A_q\}$ be its power set, with $q = 2^K$. A function M from $P(\Omega)$ to $[0, 1]$ is called a "basic belief assignment" (*bba*) if $M(\emptyset) = 0$ and $\sum_{A \in P(\Omega)} M(A) = 1$. A *bba* M defines then a "plausibility" function pl from $P(\Omega)$ to $[0, 1]$ given by $Pl(A) = \sum_{A \cap B \neq \emptyset} M(B)$ and a "credibility" function Cr from $P(\Omega)$ to $[0, 1]$ given by $Cr(A) = \sum_{B \subset A} M(B)$. Given a *bba* M, the associated Pl and Cr are related through $Pl(A) + Cr(A^c) = 1$ so that each of them defines the other. In addition, Pl and Cr can be defined by some axioms, and each of them defines then a unique *bba* M. Furthermore, a probability function p can be seen as a particular *bba*, which is null outside singletons. We can see that in such a case, Pl is equal to Cr, and both are a classic probability. We will say that a *bba* is "Bayesian" or "probabilistic" when, being null outside singletons, it defines a probability, and we will say that it is "evidential" otherwise. We see how TE extends the classic probabilities defined on a finite set. Moreover, when two *bba*s M_1 and M_2 represent two pieces of evidence, we can combine, or fuse, them using the so-called Dempster–Shafer fusion (DS fusion), which gives $M = M_1 \oplus M_2$ defined by

$$M(A) = (M_1 \oplus M_2)(A) = \frac{1}{H} \sum_{B_1 \cap B_2 = A \neq \emptyset} M_1(B_1) M_2(B_2) \tag{2}$$

where H is a normalizing constant given by

$$H = \sum_{(B_1, B_2) \in P(\Omega)^2 | B_1 \cap B_2 \neq \emptyset} M_1(B_1) M_2(B_2).$$

One can then see that when either M_1 or M_2 is Bayesian (with $H \neq 0$), the fusion result M is also Bayesian.

To illustrate the interest of extending the Bayesian modeling to the evidential one, let us consider the following example. Let $\Omega = \{\omega_1, \omega_2\}$, and let $P(\Omega) = \{\emptyset, \omega_1, \omega_2, \Omega\}$ be its associated powerset. Let us consider a family of probabilities $(p_\theta)_{\theta \in \Theta}$ defined on $\Omega = \{\omega_1, \omega_2\}$, and let us consider the following "lower" probability $\tilde{p}(\omega_k) = \inf_{\theta \in \Theta} p_\theta(\omega_k)$. Let M be a mass function defined by $M(\{\omega_1\}) = \tilde{p}(\omega_1)$, $M(\{\omega_2\}) = \tilde{p}(\omega_2)$, and $M(\{\omega_1, \omega_2\}) = 1 - \tilde{p}(\omega_1) - \tilde{p}(\omega_2)$. The latter quantity illustrates the fluctuation of the accurate varying probability p. Hence, adopting the *bba* M instead makes it possible to handle the variability of p while adopting a nonvarying value of M. The interest of such modeling has been shown theoretically in the independent nonstationary data classification context in [11].

To establish a link between TE and the aim of this paper, let us recall the EMC model. A *bba* M defined on $P(\Omega^N)$ is said to be an EMC if it is null outside $[P(\Omega)]^N$ and if it can be written as

$$M(u_1, u_2, .., u_N) = M(u_1) M(u_2 | u_1) ... M(u_N | u_{N-1}). \tag{3}$$

Setting $a^*_{ij} = M(u_j | u_i)$, $A^* = [a^*_{ij}]$ is the transition mass defined on $P(\Omega)^2$, which will be called in the remainder of this paper "evidential transition matrix" in contrast with the Bayesian transition matrix A, defined on Ω^2. The EMC given by (3) generalizes the Markov chain of (1) and allows one to take into account the possible fluctuations of the transition probabilities. This result has been successfully used within the triplet Markov chain formalism to extend the well-known hidden Markov chain model [23] in many directions allowing, in particular, nonstationary data modeling [11], [24], [25], data fusion [18], or both simultaneously [26].

III. Unification of Markov Transition Matrices

In this section, we propose an original approach based on hidden Markov chains and TE to unify a family of heterogeneous Markov transition matrices, or data samples. First, we discuss the motivation behind our evidential modeling. Then, we describe the proposed approach and its related techniques. Finally, we recall the Bayesian information criterion (BIC) that will be used later for performance evaluation purposes.

A. Model Motivation

Let us consider problem 1 discussed in Section I. Let us consider a nonstationary sequence $X = (X_1, ..., X_N)$, or alternatively, a family of heterogeneous subsequences $(X^s)_{s=1}^S$ (for a given S) that could be merged into one sequence X, each X_n taking its values in $\Omega = \{\omega_1, ..., \omega_K\}$. Let us assume X Markovian, with the distribution $p(x_1, ..., x_N)$ given by an initial distribution and transition matrices $A_2, ..., A_N$, possibly varying with n. Estimating the transition matrix from the complete data $X = x$, while ignoring its fluctuations, may possibly yield a nonrepresentative Bayesian transition matrix A. On the other hand, conducting such estimation locally on samples $(X^s)_{s=1}^S$ will possibly lead to S different transition matrices $(A_s)_{s=1}^S$. The problem then is to propose a stationary model, which would be easier to estimate, and which would approximate the unknown distribution $p(x_1, ..., x_N)$ more accurately than any homogeneous Markov distribution. To illustrate this situation, let us consider the following problem of daily rainfall modeling. Let $\Omega = \{\omega_1, \omega_2\}$, where ω_1 and ω_2 denote "Wet" and "Dry," respectively. Of course, the transitional probabilities between the system states depend on the year season and may even vary over the years. Hence, estimating the associated transition matrix A on different samples of data $(X^s)_{s=1}^S$, possibly taken from different year months, will probably result in S different transition matrices $(A_s)_{s=1}^S$. Let us now consider $X = (X_1, ..., X_n)$ a sequence corresponding to daily rainfall over a year period, from day 1 to n, and let us assume that we are interested in estimating X_{n+1} knowing $(X_1, ..., X_n) = (x_1, ..., x_n)$. From the viewpoint of Markov chain model of (1), the distribution of X_{n+1} can be directly derived from the transition matrix A through $p(X_{n+1} = \omega_j | X_n = \omega_i) = a_{ij}$. On the other hand, according to our evidential model, the situation where $(X_{n+1}, U_{n+1}) = (\omega_1, \{\omega_1\})$ can be different from $(X_{n+1}, U_{n+1}) = (\omega_1, \{\omega_1, \omega_2\})$, in both of which $X_{n+1} = \omega_1$, since the system behavior is governed by the transition matrix

A^* giving $p(U_{n+1} = \gamma_j | U_n = \gamma_i) = a^*_{ij}$. In other words, the fluctuation of the transitional probabilities with respect to Ω is modeled through the transitional probabilities on $P(\Omega)$ involving the compound state $\{\omega_1, \omega_2\}$, which can be perceived as an inadequacy measurement of the standard Markov chain model.

To consider problem 2, let us assume that we have two experts. At date n, we may have $X_n = \omega_1$ according to expert 1, while expert 2 considers that $X_n = \omega_2$. To overcome this lack of consensus about state meaning, one can set $U_n = \{\omega_1, \omega_2\}$. In the daily rainfall problem introduced above, for instance, assume that assigning a day to "Dry" or "Wet" depends on a threshold value (defined by an expert) of the precipitation amount [27] (see Section IV-B).

Let us now consider problem 3, where only a finite family of Bayesian transition matrices $(A_s)_{s=1}^S$ are available. To "combine" these matrices (possibly different, due to the reasons described in the two problems above), we propose first to sample S realizations $(X^s)_{s=1}^S$, where $X^s = (X_1^s, ..., X_N^s)$ for a given N using $(A_s)_{s=1}^S$ as specified in (1). This gives a family of data samples that can be merged into one data sequence X. One finds again the situation of problem 1.

B. Model Definition

We now propose a general evidential modeling allowing to handle the problems discussed above. Let $(X_1, ..., X_N) = (x_1, ..., x_N)$, where each X_n takes its values in Ω. The main idea is to consider X as a noisy version of a hidden stationary evidential chain $U = (U_1, ..., U_N)$ taking its values in $P(\Omega) = \{\gamma_1, ..., \gamma_q\}$ with $q = 2^N$. The pairwise process $\lambda = (U, X)$ is then a hidden Markov chain [23], where U is defined in the evidential domain $P(\Omega)$ associated with the state set Ω. Thus, we consider a Markov *bba* $M(u_1, ..., u_N)$ given by (3) and $p(x_1|u_1), ..., p(x_N|u_N)$ that we have to define. Then, we have

$$p(u, x) = M(u_1)p(x_1|u_1) \prod_{n=2}^{N} M(u_n|u_{n-1})p(x_n|u_n). \quad (4)$$

According to the "Pignistic probability transform" [5], we set

$$p(X_n = \omega_i | U_n = \gamma_v) = \frac{1_{\omega_i \in \gamma_v}}{|\gamma_v|}. \quad (5)$$

The distributions $p(x_1|u_1), ..., p(x_N|u_N)$ being defined, the model is defined through $M(u_1), M(u_2|u_1), ..., M(u_N|u_{N-1})$. As the *bba* M is stationary, the model is then fully defined through $M(u_1)$ and $M(u_2|u_1)$. Let $\theta = (\varpi, A^*)$ be the set of parameters of λ, where $\varpi_i = M(u_1 = \gamma_i)$ and $a^*_{ij} = M(u_2 = \gamma_j | u_1 = \gamma_i)$.

C. Estimation and Processing Techniques

We propose now to consider the three classic problems related to the proposed evidential model λ:

a) evaluation of the likelihood $p(x)$;
b) estimation of the realization of the hidden process U;
c) estimation of the model parameters θ.

Before showing how to solve problems a–c, let us define the probability functions $\alpha_n(u_n) = p(x_1, ..., x_n, u_n)$, $\beta_n(u_n) = p(x_{n+1}, ..., x_N | u_n)$, $\xi_n(u_n) = p(u_n|x)$, and $\psi_n(u_n, u_{n+1}) = p(u_n, u_{n+1}|x)$. We have then the following computation rules, which are similar to the ones used in HMCs:

$$\alpha_1(u_1) = M(u_1)p(x_1|u_1);$$

$$\alpha_n(u_n) = \sum_{u_{n-1}} \alpha_{n-1}(u_{n-1}) M(u_n|u_{n-1}) p(x_n|u_n) \quad (6)$$

$$\beta_N(u_N) = 1;$$

$$\beta_n(u_n) = \sum_{u_{n+1}} \beta_{n+1}(u_{n+1}) M(u_{n+1}|u_n) p(x_{n+1}|u_{n+1}) \quad (7)$$

$$\xi_n(u_n) \propto \alpha_n(u_n)\beta_n(u_n) \quad (8)$$

$$\psi_n(u_n, u_{n+1}) \propto \alpha_n(u_n) M(u_{n+1}|u_n) p(x_{n+1}|u_{n+1}) \beta_{n+1}(u_{n+1}). \quad (9)$$

Solution of problem a: The marginal distribution $p(x)$ can be computed from $(x_1, ..., x_N)$ through

$$p(x) = \sum_{u \in P(\Omega)^N} p(u, x) = \sum_{u_N \in P(\Omega)} \alpha_N(u_N). \quad (10)$$

Remark 1. Let us notice that the distribution $p(x)$ can be seen as the DS fusion of $m(u_1, ..., u_N)$ with the very simple distribution $q(x_1, ..., x_N) = q^1(x_1) \ldots q^1(x_N)$, where q^1 is the uniform distribution on Ω: $p = m \oplus q$.

Solution of problem b: The estimation of the realization of the underlying process U can be of interest. For this purpose, one can use some Bayesian techniques such as marginal posterior mode (MPM) [28] or maximum a priori (MAP) [29].

The MPM estimator, which can be computed through (8), is given by the following formula:

$$[\hat{u} = \hat{u}_{\text{MPM}}(x)] \Longleftrightarrow \left[\hat{u}_n = \arg\max_{u \in P(\Omega)} p(u_n|x)\right]. \quad (11)$$

On the other hand, the MAP estimator is given by the following formula:

$$[\hat{u} = \hat{u}_{\text{MAP}}(x)] \Longleftrightarrow \left[\hat{u} = \arg\max_{u \in P(\Omega)^N} p(u|x)\right]. \quad (12)$$

To compute such estimator, let us consider the quantity

$$\delta_n(u_n) = \max_{(u_1, .., u_{n-1}) \in [P(\Omega)]^{n-1}} p(u_1, x_1, .., u_n, x_n)$$

that can be computed in the following recursive manner:

$$\delta_1(u_1) = M(u_1)p(x_1|u_1);$$

$$\varphi_1(u_1) = \emptyset;$$

$$\delta_{n+1}(u_{n+1}) = \max_{u_n \in P(\Omega)} \{\delta_n(u_n) M(u_{n+1}|u_n) p(x_{n+1}|u_{n+1})\};$$

$$\varphi_{n+1}(u_{n+1}) = \arg\max_{u_n \in P(\Omega)} \{\delta_n(u_n) M(u_{n+1}|u_n) p(x_{n+1}|u_{n+1})\} \quad (13)$$

where $\varphi_n(u_n)$ is the predecessor of u_n giving $\delta_n(u_n)$. The optimal path is then derived as follows:

$$\hat{u}_N = \arg\max_{u_n \in P(\Omega)} \delta_N(u_n);$$
$$\hat{u}_n = \varphi_{n+1}(u_{n+1}) \text{ for } n = N-1, ..., 1. \quad (14)$$

Solution of problem c: The model parameters θ can be estimated via one of the well-known mixture estimation algorithms like expectation-maximization (EM) [28], [30], stochastic EM [31], or iterative conditional estimation [32].

In this paper, we propose to use the EM algorithm, which runs as follows:

1) Find an initial value $\theta^{(0)}$ of the parameters.
2) Compute $\theta^{(q+1)}$ from $\theta^{(q)}$ and x as follows:
   a) *Step E:* Use (6)–(9) to compute $\xi_n(u_n)$ and $\psi_n(u_n, u_{n+1})$ using the current parameters $\theta^{(q)}$.
   b) *Step M:* Compute $\theta^{(q+1)}$ as follows:

$$\varpi_i = \frac{1}{N}\sum_{n=1}^{N} \xi_n(\gamma_i) \quad (15)$$

$$a_{ij}^* = \frac{\sum_{n=1}^{N-1} \psi_n(\gamma_i, \gamma_j)}{\sum_{n=1}^{N-1} \xi_n(\gamma_i)} \quad (16)$$

until an end criterion is reached.

Notice that during the execution of EM, the distributions $p(x_1|u_1), ..., p(x_N|u_N)$ are forced to remain of the form given by (5).

D. Bayesian Information Criterion

One of the major issues in model selection is to increase the model fitting while keeping the model dimension, i.e., the number of its parameters, reasonable. For this purpose, some criteria like the BIC [33] or Akaike information criterion (AIC) [34] introduce a penalty term proportional to the number of parameters.

Let $X = x$ be a data sample. Let λ be the candidate model, where θ_λ denotes its parameter set, with p being its number of free parameters. Let N be the dimension of the data. Let $L = p(x|\lambda)$ be the marginal likelihood of the observed data given the model λ, and let $\hat{L}$ be the maximized value of the likelihood function $p(x|\hat{\theta}_\lambda)$ of the model λ. Hence, the BIC is given by

$$\text{BIC}(x|\hat{\theta}_\lambda) = -2\ln(\hat{L}) + p\ \ln(N). \quad (17)$$

Hence, when a family of models $\Lambda = \{\lambda_1, ..., \lambda_D\}$ is available, the best model is the one given by

$$\lambda_d = \min_{1 \le d \le D} \text{BIC}(x|\hat{\theta}_{\lambda_d}). \quad (18)$$

On the other hand, the AIC is given by

$$\text{AIC}(x|\hat{\theta}_\lambda) = -2\ln(\hat{L}) + 2p. \quad (19)$$

Accordingly, the BIC tends to favor models with small N (for $N \ge 8$, $\ln(N) \ge 2$). Since we deal with bigger values of N in this paper, we will use BIC for comparison sake between the proposed model and the Markov chain model to favor this latter.

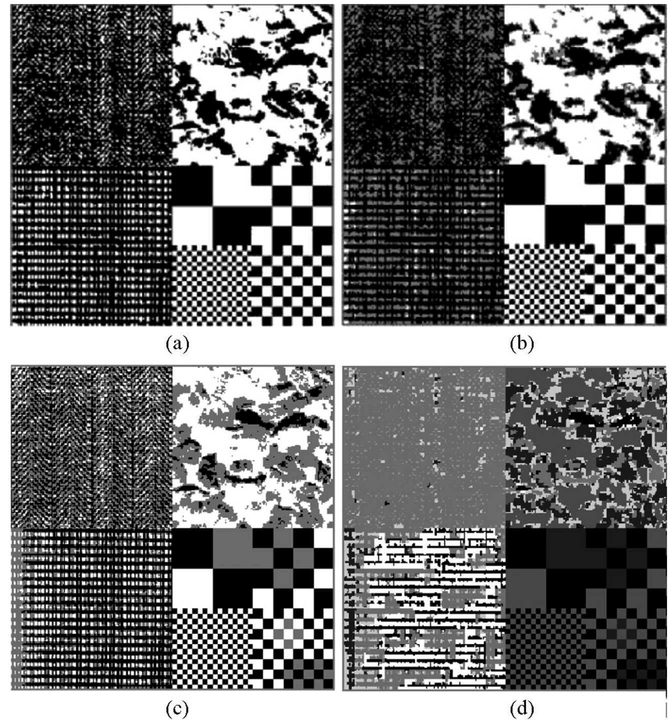

Fig. 1. Estimation of the underlying process U for the "Texture" image. (a) Texture image, $K = 2$. (b) MPM-based estimate of U, where ω_1, ω_2 and $\{\omega_1, \omega_2\}$ are depicted in black, white, and gray, respectively. (c) Texture image, $K = 3$. (d) MPM-based estimate of U, where $\omega_1, \omega_2, \omega_3, ..., \{\omega_1, \omega_2, \omega_3\}$ are depicted gradually from black to white.

IV. Experiments

The aim of this section is to show, through experiments, the effectiveness of the proposed approach to derive a consensus evidential Markov transition matrix and to show, according to BIC metric, the interest of such modeling with respect to the classical Bayesian transition matrix. For this purpose, we consider two series of experiments. In the first one, we deal with an image modeling problem, where images are converted to and from 1-D sequences via the Hilbert–Peano scan as done in [11]. In the second series, we consider the problem of consensus transition matrix estimation from daily rainfall data.

A. Inference of Evidential Transition Matrix From Data Samples

Let us consider the "Texture" image of size 256×256 [see Fig. 1(a)]. We have a realization of X with $\Omega = \{\omega_1, \omega_2\}$, where ω_1 and ω_2 correspond to black pixels and white ones, respectively. The image is then divided into four square subimages X^1 (upper left), X^2 (upper right), X^3 (down right), and X^4 (down left) of the same size 128×128 [see Fig. 1(a)]. Such subimages can be seen as four different realizations of a same system. Having $X = x$, $X_1 = x_1$, $X_2 = x_2$, $X_3 = x_3$, and $X_4 = x_4$, how could one model the corresponding system?

We first consider the standard Bayesian modeling. Hence, assuming X Markovian and stationary, we estimate the transition matrices A, A_1, A_2, A_3, and A_4 based on sequences X, X^1,

TABLE I
COMPARISON BETWEEN TRANSITION MATRICES PERFORMANCES BASED ON BIC METRIC, TEXTURED IMAGE, $K = 2$

	A_1	A_2	A_3	A_4	A	A^*	A^{**}
X^1	**15 074**	21 627	32 394	19 143	18 726	15 979	16 135
X^2	29 152	11 491	15 143	18 809	12 449	**11 221**	11 507
X^3	22 977	5749	**3331**	13 628	6880	5129	5097
X^4	16 442	19 343	28 249	**15 539**	17 055	16 568	16 568
X	83 558	58 132	79 031	63 628	55 024	**48 406**	48 789

TABLE II
COMPARISON BETWEEN TRANSITION MATRICES PERFORMANCES BASED ON BIC METRIC, TEXTURED IMAGE, $K = 3$

	A_1	A_2	A_3	A_4	A	A^*	A^{**}
X^1	**29 509**	43 442	102 312	33 094	34 981	31 283	30 614
X^2	39 352	**16 400**	36 956	26 938	18 851	16 785	17 763
X^3	34 869	12 135	**3452**	18 435	11 463	5413	6695
X^4	32 781	40 119	74 766	**28 352**	30 550	30 544	31 142
X	136 343	111 945	217 301	106 640	95 685	**82 616**	84 775

X^2, X^3, and X^4, respectively. We obtain

$$A = \begin{pmatrix} 0.877 & 0.123 \\ 0.195 & 0.805 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} 0.848 & 0.152 \\ 0.704 & 0.296 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.841 & 0.159 \\ 0.089 & 0.911 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 0.979 & 0.021 \\ 0.021 & 0.979 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0.858 & 0.142 \\ 0.471 & 0.529 \end{pmatrix}.$$

As we can see, the transition matrices obtained are different from each other. This situation corresponds actually to problem 1 discussed in Section I. The whole image can then be perceived as a varying system, for which a representative transition matrix is to be estimated, for possible future use.

To apply our approach, let us consider an underlying process U taking its values in $P(\Omega)$. Assuming (U, X) a hidden Markov chain and setting (5), the evidential transition matrix is estimated in an unsupervised way from $X = x$ using the EM algorithm as described in Section III. We obtain

$$A^* = \begin{pmatrix} 0.865 & 0.003 & 0.132 \\ 0.005 & 0.961 & 0.034 \\ 0.268 & 0.037 & 0.695 \end{pmatrix}.$$

The performance evaluation results of all the transition matrices are provided in Table I. The evidential transition matrix A^* is more appropriate than the Bayesian ones to model the whole data sequence X. Indeed, for A, $BIC = 48\,406$, whereas for A^*, $BIC = 55\,024$. On the other hand, A^* is also more suitable than A to locally model subimages.

In addition, it is possible to quantify the usefulness of the proposed evidential modeling over the Bayesian one by considering the joint mass function m^* associated with A^*, where $m^*_{ij} = M(U_n = \gamma_i, U_{n+1} = \gamma_j)$ and which can be estimated using a formula similar to (16) with no denominator:

$$m^* = \begin{pmatrix} 0.427 & 0.001 & 0.065 \\ 0.001 & 0.253 & 0.009 \\ 0.065 & 0.009 & 0.169 \end{pmatrix}.$$

In fact, the sum of quantities allocated to m^*_{ij}, where γ_i or γ_j are nonsingletons is quite high (0.317), which shows how the evidential modeling can overcome the lack of precision of the Bayesian modeling.

The MPM-based estimate of the evidential process U is shown in Fig 1(b). The presence of the underlying class $\{\omega_1, \omega_2\}$ (depicted in gray) refers in some manner to the lack of precision of the Bayesian modeling. Indeed, for each site n, the uncertainty is then associated with sites where $U_n \neq X_n$.

Finally, let us assume that the four Bayesian transition matrices A_1, A_2, A_3, and A_4 are provided by four independent experts, regardless of the data sequence $X = x$ corresponding to the image. The problem is to derive a representative transition matrix of the family $(A_s)_{s=1}^4$. For this purpose, we apply the sampling algorithm of Section III-A. Setting $N = 15\,000$, we obtain

$$A^{**} = \begin{pmatrix} 0.851 & 0.034 & 0.115 \\ 0.058 & 0.939 & 0.003 \\ 0.302 & 0.004 & 0.694 \end{pmatrix}$$

which is close to A^* estimated on the genuine image realization. The performance of A^{**} has also been assessed via BIC metric, and the results obtained are reported in Table I.

As we can see, the performance of A^{**} estimated by sampling ($BIC = 48\,789$) is close to the one provided by A^* ($BIC = 48\,406$). In all cases, A^{**} is more suitable than the Bayesian A. This shows again the greater generality of the proposed evidential modeling over the standard Bayesian one.

To consider a more complicated image modeling problem, the same experiment is conducted on a three-class version of the same "Texture" image [see Fig. 1(c)], where ω_1, ω_2, and ω_3 correspond to black, gray and white pixels, respectively. The performance evaluation of the associated transition matrices is provided in Table II (also see Fig. 1(d) for the MPM-estimate of U). The results obtained confirm the interest of the evidential modeling with respect to the classic Markov chain.

B. *Inference of Consensus Transition Matrix From Rainfall Data*

Let us consider the problem of daily rainfall modeling discussed in Section III-A. Let $X = (X_1, ..., X_N)$, where each X_n takes its values in $\Omega = \{\omega_1, \omega_2\}$, where ω_1 and ω_2 denote "Wet" and "Dry," respectively. Assigning a day to one of the two classes depends on a beforehand established threshold. In [27], for instance, three threshold values have been considered, and accordingly, three transition matrices have been estimated per month over a period of 60 years (1941–2000). In this experiment, the same set of matrices has been used to sample a sequence of data $X = (X_1, ..., X_N)$ with $N = 21\,915$ (corresponding

TABLE III
COMPARISON BETWEEN TRANSITION MATRICES PERFORMANCE BASED ON BIC METRIC, DAILY RAINFALL DATA

threshold	0.1 mm	1 mm	3 mm
A_1	**26 965**	27 051	26 376
A_2	27 292	26 747	25 445
A_3	28 574	27 198	**25 040**
A^*	27 010	**26 570**	25 256

to 60 years). Hence, we have three different interpretations of data: $X^1 = (x_1^1, ..., x_N^1)$, $X^2 = (x_1^2, ..., x_N^2)$, and $X^3 = (x_1^3, ..., x_N^3)$. The aim is then to check that for each data sample (i.e., interpretation), the evidential transition matrix is better suited than the two Bayesian ones associated with the two other data samples.

Hence, for each sequence X^i, we have estimated the Bayesian transition matrix. The obtained matrices are

$$A_1 = \begin{pmatrix} 0.690 & 0.310 \\ 0.293 & 0.707 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} 0.617 & 0.383 \\ 0.253 & 0.746 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0.529 & 0.471 \\ 0.203 & 0.797 \end{pmatrix}.$$

Since the meaning attached to the data sample is different, the transition matrices also are. The problem is then to estimate the consensus evidential transition matrix. Applying our approach, we obtain the evidential transition matrix

$$A^* = \begin{pmatrix} 0.643 & 0.102 & 0.255 \\ 0.042 & 0.737 & 0.221 \\ 0.146 & 0.233 & 0.621 \end{pmatrix}.$$

The performance evaluation results of the transition matrices are given in Table III. As we can see, the BIC values corresponding to the evidential transition matrix are very close to the ones based on the Bayesian transition matrices estimated locally on each data sample. For X^2, the performance of A^* is even higher than the A_2 one. This is due to the fact that data samples are nonstationary regardless of the interpretation related to the precipitation threshold value, given the fluctuations of transitional probabilities depending on the year month. This proves the interest of the proposed modeling and the inference approach.

V. CONCLUSION

In this paper, we have proposed an original approach, based on EMCs, to infer a consensus transition matrix from a set of data samples or, a set of Markov transition matrices. The proposed approach has been validated through experiments on label images and daily rainfall data. The results obtained confirm, on one hand, the interest of the evidential modeling with respect to the conventional Bayesian one and, on the other hand, the validity of the proposed approach to infer the consensus transition matrix. An interesting future direction would be to consider the same extension for Markov random fields and general Bayesian networks.

REFERENCES

[1] T. W. Anderson and L. A. Goodman, "Statistical inference about Markov chains," *Ann. Math. Statist.*, vol. 28, pp. 89–110, 1957.

[2] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, vol. 356. Princeton, NJ, USA: van Nostrand, 1960.

[3] K. L. Chung, *Markov Chains*. New York, NY, USA: Springer, 1967.

[4] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 1976.

[5] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.

[6] C. Romer and A. Kandel, "Applicability analysis of fuzzy inference by means of generalized Dempster-Shafer theory," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 4, pp. 448–453, Nov. 1995.

[7] F. Janez and A. Appriou, "Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods," *Int. Approx. Reas.*, vol. 18, no. 1, pp. 1–19, 1998.

[8] C. Lucas and B. N. Araabi, "Generalization of the Dempster-Shafer theory: A fuzzy-valued measure," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 3, pp. 255–270, Jun. 1999.

[9] R. R. Yager, "Set measure directed multi-source information fusion," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1031–1039, Dec. 2011.

[10] E. Ramasso and T. Denoeux, "Making use of partial knowledge about hidden states in HMMs: An approach based on belief functions," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 2, pp. 395–405, Apr. 2014.

[11] P. Lanchantin and W. Pieczynski, "Unsupervised restoration of hidden nonstationary Markov chains using evidential priors," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3091–3098, Aug. 2005.

[12] L. Fouque, A. Appriou, and W. Pieczynski, "An evidential Markovian model for data fusion and unsupervised image classification," in *Proc. 3rd Int. Conf. Inf. Fusion*, 2000, pp. TUB4/25–TUB4/32.

[13] H. Dehghani, B. Vahidi, R. Naghizadeh, and S. Hosseinian, "Power quality disturbance classification using a statistical and wavelet-based hidden Markov model with Dempster–Shafer algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 47, pp. 368–377, 2013.

[14] H. Soubaras, "On evidential Markov chains," in *Foundations of Reasoning Under Uncertainty*. New York, NY, USA: Springer, 2010, pp. 247–264.

[15] T. Reineking, "Particle filtering in the Dempster–Shafer theory," *Int. J. Approx. Reason.*, vol. 52, no. 8, pp. 1124–1135, 2011.

[16] E. Ramasso and R. Gouriveau, "Prognostics in switching systems: Evidential Markovian classification of real-time neuro-fuzzy predictions," in *Proc. Prognostics Health Manag. Conf.*, 2010, pp. 1–10.

[17] E. Ramasso, "Contribution of belief functions to hidden Markov models with an application to fault diagnosis." in *Proc. IEEE Int. Workshop Mach. Learning Signal Process.*, 2009, pp. 1–6.

[18] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *Int. J. Approx. Reason.*, vol. 45, no. 1, pp. 1–16, 2007.

[19] S. Foucher, M. Germain, J.-M. Boucher, and G. B. Benie, "Multisource classification using ICM and Dempster-Shafer theory," *IEEE Trans. Instrum. Measur.*, vol. 51, no. 2, pp. 277–281, Apr. 2002.

[20] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence," *Image Vision Comput.*, vol. 24, no. 1, pp. 61–69, 2006.

[21] S. Le Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover," *Pattern Recog.*, vol. 31, no. 11, pp. 1811–1823, 1998.

[22] F. Tupin, I. Bloch, and H. Maître, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sensing.*, vol. 37, no. 3, pp. 1327–1343, May 1999.

[23] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[24] M. E. Y. Boudaren, E. Monfrini, and W. Pieczynski, "Unsupervised segmentation of random discrete data hidden with switching noise distributions," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 619–622, Oct. 2012.

[25] M. E. Y. Boudaren, E. Monfrini, and W. Pieczynski, "Unsupervised segmentation of nonstationary pairwise Markov chains using evidential priors," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 2243–2247.

[26] M. E. Y. Boudaren, E. Monfrini, W. Pieczynski, and A. Aïssani, "Dempster-Shafer fusion of multisensor signals in nonstationary Markovian context," *EURASIP J. Adv. Signal Process.*, vol. 2012, p. 134, 2012.

[27] D. U. J. Sonnadara and D. R. Jayewardene, "A Markov chain probability model to describe wet and dry patterns of weather at Colombo," *Theor. Appl. Climatol.*, vol. 119, pp. 333–340, 2015.

[28] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, pp. 164–171, 1970.

[29] G. D. Forney, Jr, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[30] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382. New York, NY, USA: Wiley, 2007.

[31] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: An experimental study in the mixture case," *J. Statist. Comput. Simul.*, vol. 55, no. 4, pp. 287–314, 1996.

[32] J. P. Delmas, "An equivalence of the EM and ICE algorithm for exponential family," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2613–2615, Oct. 1997.

[33] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[34] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control.*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.