

## Phasic Triplet Markov Chains

Mohamed El Yazid Boudaren, Emmanuel Monfrini,  
Wojciech Pieczynski, and Amar Aïssani

**Abstract**—Hidden Markov chains have been shown to be inadequate for data modeling under some complex conditions. In this work, we address the problem of statistical modeling of phenomena involving two heterogeneous system states. Such phenomena may arise in biology or communications, among other fields. Namely, we consider that a sequence of meaningful words is to be searched within a whole observation that also contains arbitrary one-by-one symbols. Moreover, a word may be interrupted at some site to be carried on later. Applying plain hidden Markov chains to such data, while ignoring their specificity, yields unsatisfactory results. The Phasic triplet Markov chain, proposed in this paper, overcomes this difficulty by means of an auxiliary underlying process in accordance with the triplet Markov chains theory. Related Bayesian restoration techniques and parameters estimation procedures according to the new model are then described. Finally, to assess the performance of the proposed model against the conventional hidden Markov chain model, experiments are conducted on synthetic and real data.

**Index Terms**—Bayesian restoration, biology and genetics, hidden Markov chains, Markov processes, maximal posterior mode, maximum a posteriori, triplet Markov chains, Viterbi algorithm

### 1 INTRODUCTION

HIDDEN Markov chains (HMCs) have been for long used for data modeling in a wide range of applications fields, particularly in signal and image processing. In fact, HMCs have been extensively applied in image segmentation [1], land change detection [2], handwritten word recognition [3], speech recognition [4], document image analysis, tumor classification [5], or even acoustic musical signal recognition [6]. Other potential applications include communications [7], bioinformatics [8] or genome structure recognition [9]. Let us also mention some general books [8], [10] and pioneering papers [11], [12], [13] where the HMC theory is expansively described.

When the unobserved signal of interest can be modeled via a finite Markov chain, and when the noise is not too complex, HMCs are quite robust and allow to provide satisfactory results. However, they can be inefficient for some complex data such as nonstationary ones [14], [15].

HMCs have then been generalized in many directions to enlarge their modeling capabilities while keeping their associated techniques workable. In particular, pairwise Markov chains (PMCs) [16] can improve HMCs results [17] (also see [18] for on-line demos and source codes). Let  $X$  be the hidden chain and  $Y$  the observed one. The originality of PMCs with respect to the classic HMCs is that  $X$  is not necessarily Markovian but, since the couple  $(X, Y)$  is of Markov form,  $X$  is Markovian conditionally on  $Y$ , which makes feasible the classical Bayesian processing.

Further extensions of PMCs, firstly proposed in [19], consist in adding a third finite random sequence  $U$  and considering that the

triplet  $(U, X, Y)$  is a Markov chain. Then, such a “triplet” Markov chain (TMC) can still be used to search the hidden  $X$ , and the computing complexity is comparable to that of conventional HMCs. The family of TMCs is very rich and the added  $U$  can be used to model different situations. In [14], [15], it models different stationarities of the distribution of  $(X, Y)$ , and thus  $U$  models the switches in parameters defining the distribution of  $(X, Y)$ . In [20] authors use a TMC to model the additional information contained in an atlas. Another possibility is to use  $U$  to model the semi-Markovianity of  $X$ , and thus the classic hidden semi-Markov model appears as a particular TMC [21]. Similarly, the Gaussian mixture model (GMM) [22] can also be considered as a particular TMC. Using Dempster-Shafer fusion [23] in Markovian context leads to another class of TMCs with interesting possibilities of integrating different partial pieces of information in conventional HMCs [24], [25], [26], [27], [28], [29]. In addition, different uses of  $U$  can be dealt with simultaneously as, for example, in the case of nonstationary hidden semi-Markov model [26]. TMCs have also been used for continuous hidden sequences in Kalman filtering [30], in prediction [31], or still optimal fast filtering in a particular class of switching systems [32]. Finally, let us mention that hidden Markov fields have also been extended to triplet Markov fields [33], and have been successfully applied to complex structure data classification [34], in SAR images processing [35], [36], [37], [38] or biometry [39].

In this paper we introduce a particular TMC to deal with the following problem. Let  $X = (X_1, \dots, X_N)$  be a random sequence where each  $X_n$  takes its values in the state set  $\Omega = \{\omega_1, \omega_2\}$ ; and let  $Y = (Y_1, \dots, Y_N)$  be an observable discrete sequence, where each  $Y_n$  takes its values in a finite symbol set  $\Upsilon = \{v_1, \dots, v_M\}$ . Let us assume that the only observations of interest, which will be called word symbols, are those corresponding to  $X_n = \omega_2$ , whereas those corresponding to  $X_n = \omega_1$ , are “solitary symbols” and do not contain any useful information. The sequence of word symbols is a sequence of words, each being of length  $K$ . One of the difficulties when searching words is that they may be interrupted at any moment  $n$  (which means that  $x_n = \omega_1$ ) to be resumed at the next site  $n'$  such that  $x_{n'} = \omega_2$  (see Fig. 1). Therefore, for  $X_n = \omega_2$ , it is not immediate to see to which place in a word  $Y_n = y_n$  corresponds, and to answer this question one must know the past  $X_1 = x_1, X_2 = x_2, X_{n-1} = x_{n-1}$  which makes the use of conventional Markov models and even high-order ones like [40] unsuitable. For this purpose, we define the phase of a symbol as the length of the sub-word to be fulfilled at the next occurrence of  $\omega_2$ . Hence, a symbol is of phase zero if the beginning of a new word is expected (see Fig. 1). The idea we propose is then to summarize the useful past information in an auxiliary discrete sequence  $U$ , and to assume that the triplet  $T = (U, X, Y)$  is a Markov chain. As we are going to see, the words sequence of interest can then be searched by classic Bayesian methods in an unsupervised way. Let us specify that such problems may occur in communications. Consider a transmitter who sends through a channel, using some kind of modulation, a sequence of fixed-size reserved words to a receiver with the following characteristic: The transmission may be interrupted at any moment to be resumed afterward. In the meantime, random data are received. The problem is to reconstitute the meaningful phrase of interest, typically a command or a password. The interruption may be due to a waver access to the channel by the transmitter. This latter may also deliberately includes random nonsense data to mislead sniffing attempts. An analogous situation occurs in some biological data like DNA where a sequence of nucleotide triplets (words of length 3), encoding a protein, is to be searched within the whole sequence.

The remainder of the paper is organized as follows: Section 2 gives an overview about HMCs, PMCs and TMCs formalisms. Section 3 describes the proposed Phasic triplet Markov chain (PTMC) and its

- M.E.Y. Boudaren is with Ecole Militaire Polytechnique, BP 17, Bordj El Bahri, Algiers 16111, Algeria. E-mail: boudaren@gmail.com.
- E. Monfrini and W. Pieczynski are with the Department of CITI, Institut Mines-Télécom, Télécom SudParis, 9 rue Charles Fourier, 91011 Evry, France. E-mail: {emmanuel.monfrini, wojciech.pieczynski}@it-sudparis.eu.
- A. Aïssani is with Université des Sciences et de la Technologie Houari Boumediene (USTHB), BP 32, El Alia, Algiers 16111, Algeria. E-mail: amraissani@yahoo.fr.

Manuscript received 16 Apr. 2013; revised 24 Jan. 2014; accepted 19 May 2014. Date of publication 2 June 2014; date of current version 9 Oct. 2014.

Recommended for acceptance by F. Fleuret.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.

Digital Object Identifier no. 10.1109/TPAMI.2014.2327974

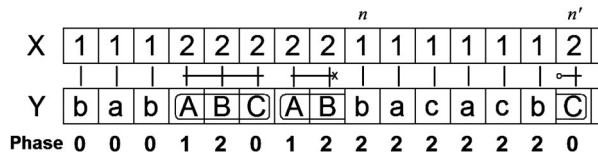


Fig. 1. Portions of sequences  $X$  and  $Y$  for  $\Omega = \{1, 2\}$ ,  $\Upsilon = \{a, b, c\}$  and  $K = 3$ . Word symbols are capitalized for illustrative purpose.

corresponding theory. Experimental results are provided and discussed in Section 4. Concluding remarks and future improvements are given in Section 5.

## 2 HIDDEN, PAIRWISE AND TRIPLET MARKOV CHAINS

Hidden Markov chains have shown their limitation to handle some data specificities. For this purpose, these models have been generalized to PMCs and TMCs which are of strictly increasing generality. The aim of this section is to briefly describe HMC, PMC and TMC models.

### 2.1 Hidden Markov Chains

Let us consider two processes: a hidden discrete process  $X = (X_1, \dots, X_N)$  where each  $X_n$  takes its values in the state set  $\Omega = \{\omega_1, \dots, \omega_S\}$ ; and an observable discrete process  $Y = (Y_1, \dots, Y_N)$  where each  $Y_n$  takes its values in a finite symbol set  $\Upsilon = \{v_1, \dots, v_M\}$ .

The pairwise process  $Z = (X, Y)$  is said to be a HMC if the hidden process of interest  $X$  is a Markov chain. Moreover, when the conventional noise independence assumptions hold, the joint distribution of  $Z$  may be expressed as

$$p(x, y) = p(x_1)p(y_1 | x_1) \prod_{n=2}^N p(x_n | x_{n-1})p(y_n | x_n) \quad (1)$$

This model is fully defined through the parameters set  $\Theta = (\Pi, A, B)$  where  $\Pi$  is a distribution on  $\Omega$  such as  $\pi_i = p(x_1 = \omega_i)$ ,  $A$  is a transition matrix defined on  $\Omega^2$  as  $a_{i,j} = p(x_n = \omega_j | x_{n-1} = \omega_i)$  and  $B$  is a distribution defined on  $\Omega \times \Upsilon$  as  $b_i(m) = p(y_n = v_m | x_n = \omega_i)$ . Such a model is called “independent-noise hidden Markov chain” (IN-HMC). As commonly done in the literature, we will use “HMC” to refer to IN-HMC.

In accordance with HMC formalism, and when the model parameters  $\Theta$  are known, the hidden process of interest  $x$  may be estimated from  $y$  by means of some low-time-consuming Bayesian techniques such as marginal posterior mode (MPM) [11] or maximum *a posteriori* (MAP) [41]. The reader may refer to [13] where both techniques are described. Furthermore, when the model parameters are unknown, these latter may be estimated in an unsupervised manner by means of several iterative algorithms such as expectation-maximization algorithm (EM) [11], [42], its stochastic version (SEM) [43] or iterative conditional estimation (ICE) [17], [44].

### 2.2 Pairwise Markov Chains

Let  $X = (X_1, \dots, X_N)$  and  $Y = (Y_1, \dots, Y_N)$  be two stochastic sequences as in the previous sub-section. The pairwise process  $Z = (X, Y)$  is said to be a “Pairwise Markov chain” if its joint distribution is written

$$p(z) = p(z_1) \prod_{n=2}^N p(z_n | z_{n-1}) \quad (2)$$

which simply means that  $Z = (X, Y)$  is a Markov chain. The transition probability can then be expressed as  $p(z_n | z_{n-1}) = p(x_n | x_{n-1}, y_{n-1})p(y_n | x_n, x_{n-1}, y_{n-1})$ . One can say that a PMC is a

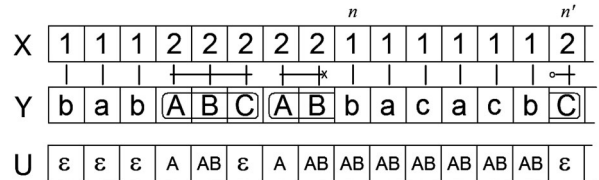


Fig. 2. Introduction of the auxiliary process  $U$  to take the phase shift into account.

classic HMC where  $p(x_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})$  and  $p(y_n | x_n, x_{n-1}, y_{n-1}) = p(y_n | x_n)$  for each  $n = 2, \dots, N$  (see [15] for the proof in general context). This means that the noise distribution can be more complex in PMC and thus, assuming that  $X$  Markovian is equivalent to lay down a constraint on the noise distribution.

In spite of this generality of PMCs over HMCs, all Bayesian processing remain possible, with comparable computation complexity. Moreover, the greater generality of PMCs can result in significantly greater efficiency in unsupervised data segmentation [17].

### 2.3 Triplet Markov Chains

Let  $X = (X_1, \dots, X_N)$  and  $Y = (Y_1, \dots, Y_N)$  be two stochastic sequences as above, with  $X$  hidden and  $Y$  observable. Considering a triplet Markov chain to search  $X$  from  $Y$  consists of introducing a third process  $U = (U_1, \dots, U_N)$ , where each  $U_n$  takes its values in a finite set  $\Sigma = \{\sigma_1, \dots, \sigma_L\}$ , and assuming that the triplet process  $T = (U, X, Y)$  is Markovian. In the case of discrete finite  $X$  considered in this paper, TMCs are not very different from the PMCs from purely mathematic viewpoint. In fact, as both  $X$  and  $U$  are discrete finite, one can say setting  $V = (U, X)$ , that  $T = (U, X, Y)$  is a TMC if and only if  $(V, Y)$  is a PMC. Thus, different Bayesian methods can still be used to search  $V$  from  $Y$  and, having found  $V = (U, X)$  one has found  $X$ . In addition,  $U$  can also be estimated, which can be of interest. In spite of this mathematical equivalency between PMCs and TMCs, the latter turns out to be quite richer. Of course, it is strictly more general as in a Markov chain  $T = (U, X, Y)$ , the couple  $Z = (X, Y)$  is not necessarily Markovian [15].

## 3 PHASIC TRIPLET MARKOV CHAINS

### 3.1 Model Definition

Let us consider the situation described in section 1 and let  $P = (P_1, \dots, P_N)$  be the phase process associated with processes  $X$  and  $Y$  where  $p_n$  is the phase of symbol  $y_n$ .

Let us introduce another process  $U = (U_1, \dots, U_N)$  where each  $U_n$  takes its values in the set  $\Gamma = \bigcup_{k=0}^{K-1} \Upsilon^k$  such as  $|u_n| = p_n$  where  $|w|$  is the length of the word  $w$ .  $\Gamma$  is then the set of words whose length ranges from 0 to  $K - 1$  symbols where the empty word denoted  $\varepsilon$  is the word of length 0. The process  $U$  stores the information about the last word portion to be completed at the next occurrence of state  $\omega_2$ . Hence, the realization of  $U_n$  is updated at each occurrence of  $\omega_2$ . The empty word  $\varepsilon$  corresponds to the situation where we have to start a new word from the beginning (Fig. 2).

Considering  $T_n = (U_n, X_n, Y_n)$ , the process  $T = (T_1, \dots, T_N)$  is called a “triplet process” associated with  $U, X$  and  $Y$ . Realizations of such processes will be denoted by lowercase letters, and for sake of simplicity  $p(t_n)$  will stand for  $p(T_n = t_n)$ . We will also denote by  $LF(w)$  the set of left factors of a word  $w \in \Lambda = \bigcup_{k=0}^K \Upsilon^k$  defined as  $LF(w) = \{z \in \Lambda | \exists w' \in \Lambda : w = z \cdot w'\}$  where “ $\cdot$ ” is the concatenation operator.

Let  $f$  and  $g$  be two emission probabilities defined on  $\Upsilon$  and  $\Upsilon^K$  respectively as  $f_{y_n} = p(y_n | x_n = \omega_1)$ , and  $g_{y_{n..n+K}} = p(y_{n..n+K} | x_{n..n+K} = (\omega_2, \dots, \omega_2), p_n = 1)$ . Assuming that  $X$  is Markovian, let

$\Pi$  be a distribution on  $\Omega$  such as  $p(x_1 = \omega_i) = \pi_i$  and let  $A$  be a matrix defined on  $\Omega^2$  as  $p(x_n = \omega_j | x_{n-1} = \omega_i) = a_{i,j}$ .

The process  $T = (U, X, Y)$  is called a Phasic triplet Markov chain (PTMC) of order  $K$  (or a  $K$ -PTMC) if it is a triplet Markov chain. Hence, its distribution is expressed as  $p(t) = p(t_1)p(t_2 | t_1) \dots p(t_N | t_{N-1})$ . Since we deal with a homogeneous TMC,  $p(t)$  is fully defined though the initial probabilities  $p(t_1) = p(x_1)p(u_1, y_1 | x_1)$  and the transition ones  $p(t_n | t_{n-1}) = p(x_n | x_{n-1})p(u_n, y_n | x_n, u_{n-1})$  where

$$p(u_1, y_1 | x_1) = \begin{cases} \mathbb{1}_{u_1 = \varepsilon} f_{y_1} & \text{if } x_1 = \omega_1 \\ \mathbb{1}_{u_1 = y_1} \frac{\sum_{w \in LF^{-1}(y_1)} g_w}{\sum_{w \in \Upsilon^K} g_w} & \text{if } x_1 = \omega_2. \end{cases} \quad (3)$$

and

$$p(u_n, y_n | x_n, u_{n-1}) = \begin{cases} \mathbb{1}_{u_n = u_{n-1}} f_{y_n} & \text{if } x_n = \omega_1 \\ \mathbb{1}_{u_n = u_{n-1} \cdot y_n} \frac{\sum_{w \in LF^{-1}(u_{n-1} \cdot y_n)} g_w}{\sum_{w \in LF^{-1}(u_{n-1})} g_w} & \text{if } x_n = \omega_2, u_{n-1} \notin \Upsilon^{K-1} \\ \mathbb{1}_{u_n = \varepsilon} \frac{g_{u_{n-1} \cdot y_n}}{\sum_{w \in LF^{-1}(u_{n-1})} g_w} & \text{if } x_n = \omega_2, u_{n-1} \in \Upsilon^{K-1}. \end{cases} \quad (4)$$

The conventional HMC is a first-order PTMC (for which  $K=1$ ). The auxiliary process realizations  $u_n$  are then all set to  $\varepsilon$  and one finds again the usual initial and transition probabilities of the HMC. Notice that the pairwise process  $Z = (X, Y)$  associated to the PTMC can be considered as a more general hidden Markov chain since the hidden process  $X$  is assumed to be Markovian here.

### 3.2 Simulation of a PTMC

The simulations of  $X$ ,  $Y$  and  $U$  are performed recursively according to the following:

- 1) Simulation of  $T_1$  is conducted as follows:
  - a) Realization  $x_1$  is sampled by drawing from the initial prior distribution  $\Pi$ ;
  - b) Realization  $y_1$  is sampled by drawing from the probabilities
$$p(y_1 | x_1) = \begin{cases} f_{y_1} & \text{if } x_1 = \omega_1 \\ \frac{\sum_{w \in LF^{-1}(y_1)} g_w}{\sum_{w \in \Upsilon^K} g_w} & \text{if } x_1 = \omega_2 \end{cases} \quad (5)$$
- c) Realization  $u_1$  is deduced from  $x_1$  and  $y_1$  as follows:  $u_1 = \varepsilon$  if  $x_1 = \omega_1$  and  $u_1 = y_1$  otherwise.
- 2) Simulation of  $T_{n+1}$  is achieved as follows:
  - a) Realization  $x_{n+1}$  is sampled by drawing from the transition matrix  $A$ ;
  - b) Realization  $y_{n+1}$  is sampled by drawing from the probabilities

$$p(y_{n+1} | u_n, x_{n+1}) = \begin{cases} f_{y_{n+1}} & \text{if } x_{n+1} = \omega_1 \\ \frac{\sum_{w \in LF^{-1}(u_n \cdot y_{n+1})} g_w}{\sum_{w \in LF^{-1}(u_n)} g_w} & \text{if } x_{n+1} = \omega_2. \end{cases} \quad (6)$$

- c) Realization  $u_{n+1}$  is deduced from  $u_n$ ,  $x_{n+1}$  and  $y_{n+1}$  as follows:

$$u_{n+1} = \begin{cases} u_n & \text{if } x_{n+1} = \omega_1 \\ u_n \cdot y_{n+1} & \text{if } x_{n+1} = \omega_2, u_n \notin \Upsilon^{K-1} \\ \varepsilon & \text{if } x_{n+1} = \omega_2, u_n \in \Upsilon^{K-1} \end{cases} \quad (7)$$

**Remark 1.** The realization of process  $U$  being deducible from realizations of  $X$  and  $Y$ , the simulation of a PTMC may be instead achieved by involving only these latter two processes.

### 3.3 Bayesian Restoration of a PTMC

In this section, we tackle the problem of Bayesian restoration of data governed by a PTMC. For this purpose, we assume the model parameters to be known and we focus on the MPM and MAP estimators. The Bayesian MPM restoration of the hidden process is performed according to the following estimator for each  $n = 1, 2, \dots, N$ :

$$[\hat{x} = \hat{s}_{MPM}(y)] \iff [\hat{x}_n = \arg \max_{\omega \in \Omega} p(x_n = \omega | y)] \quad (8)$$

The MPM estimation of (8) can be achieved thanks to the probability functions forward and backward expressed as  $\alpha_n(x_n, u_n) = p(y_1, \dots, y_n, x_n, u_n)$  and  $\beta_n(x_n, u_n) = p(y_{n+1}, \dots, y_N | x_n, u_n)$  respectively, and that can be computed using the following recursion:

$$\alpha_1(x_1, u_1) = p(t_1),$$

$$\alpha_n(x_n, u_n) = \sum_{x_{n-1}, u_{n-1}} \alpha_{n-1}(x_{n-1}, u_{n-1}) p(t_n | t_{n-1}); \quad (9)$$

and

$$\beta_N(x_N, u_N) = 1,$$

$$\beta_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1}) p(t_{n+1} | t_n). \quad (10)$$

Accordingly, the posterior probability is

$$p(x_n, u_n | y) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n), \quad (11)$$

and the posterior marginal probability required to perform the estimation of (8) can be derived according to

$$p(x_n = \omega | y) = \sum_{u_n} p(x_n = \omega, u_n | y). \quad (12)$$

**Remark 2.** It has been established in [27] that the number of elementary operations to evaluate marginal distributions  $p(x_n | y)$  in a TMC increases proportionally to data size  $N$ . Hence, this remains valid for PTMC.

Let us focus on the MAP estimation which is given by the following estimator:

$$[\hat{x} = \hat{s}_{MAP}(y)] \iff [\hat{x} = \arg \max_{x \in \Omega^N} p(x | y)] \quad (13)$$

**Remark 3.** Let  $T = (U, X, Y)$  be a PTMC and let  $V = (X, U)$ .

We have the following:

$$[\hat{x} = \hat{s}_{MAP}(y)] \iff [\hat{v} = \arg \max_{v \in [\Omega \times \Gamma]^N} p(t)].$$

We describe now a Viterbi-like algorithm to perform the MAP restoration. For this purpose, let us consider the quantity  $\delta_n(v_n) = \max_{v_1, \dots, v_{n-1} \in \Omega \times \Gamma} p(t_1, \dots, t_n)$  that can be computed using recursive formulation analogous to the forward probability function in the following manner:

$$\begin{aligned} \delta_1(v_1) &= p(v_1, y_1), \psi_1(v_1) = 0, \\ \delta_{n+1}(v_{n+1}) &= \max_{v_n \in \Omega \times \Gamma} \{\delta_n(v_n) p(t_{n+1} | t_n)\}, \\ \psi_{n+1}(v_{n+1}) &= \arg \max_{v_n \in \Omega \times \Gamma} \{\delta_n(v_n) p(t_{n+1} | t_n)\}, \end{aligned} \quad (14)$$

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

$$\hat{v}_N = \arg \max_{v_n \in \Omega \times \Gamma} \delta_N(v_n), \hat{v}_n = \psi_{n+1}(v_{n+1}). \quad (15)$$

### 3.4 Parameters Estimation from Complete Data

In this sub-section, we show how to estimate the model parameters from complete data. For this purpose, let  $Z = (Z^1, \dots, Z^R)$  be a set of  $R$  complete sequences where  $Z^r = (Z_r^1, \dots, Z_r^{N_r})$  and  $Z^r = (X^r, Y^r)$ . For a given value of order  $K$ , the model parameters  $\Theta = (\Pi, A, f, g)$  can be computed according to the following formulae:

$$\pi_i = \frac{\sum_{r=1}^R \mathbb{1}_{x_1^r = \omega_i}}{R}, \quad (16)$$

$$a_{i,j} = \frac{\sum_{r=1}^R \sum_{n=1}^{N_r-1} \mathbb{1}_{(x_n^r, x_{n+1}^r) = (\omega_i, \omega_j)}}{\sum_{r=1}^R \sum_{n=1}^{N_r-1} \mathbb{1}_{x_n^r = \omega_i}}, \quad (17)$$

$$f_{v_m} = \frac{\sum_{r=1}^R \sum_{n=1}^{N_r} \mathbb{1}_{(x_n^r, y_n^r) = (\omega_1, v_m)}}{\sum_{r=1}^R \sum_{n=1}^{N_r} \mathbb{1}_{x_n^r = \omega_1}}, \quad (18)$$

$$g_w = \frac{\sum_{r=1}^R \sum_{n=1}^{N_r} \mathbb{1}_{(x_n^r, u_n^r, w) = (\omega_2, \varepsilon, u_{n-1}^r, y_n^r)}}{\sum_{r=1}^R \sum_{n=1}^{N_r} \mathbb{1}_{(x_n^r, u_n^r) = (\omega_2, \varepsilon)}}. \quad (19)$$

### 3.5 Parameters Estimation from Incomplete Data

The model parameters can also be estimated in an unsupervised manner from a realization of  $Y = (Y_1, \dots, Y_N)$  provided that the model order  $K$  is beforehand known. Thereafter, we describe an EM-like algorithm to estimate the set of model parameters  $\Theta = (\Pi, A, f, g)$ . For this purpose, let  $\chi_n(v_n) = p(v_n | y)$  and  $\xi_n(v_n, v_{n+1}) = p(v_{n+1} | v_n, y)$ , that can be expressed through the recursive probability functions  $\alpha_n$  and  $\beta_n$  as follows:

$$\chi_n(v_n) \propto \alpha_n(v_n) \beta_n(v_n), \quad (20)$$

$$\xi_n(v_n, v_{n+1}) \propto \alpha_n(v_n) p(t_{n+1} | t_n) \beta_{n+1}(v_{n+1}). \quad (21)$$

The EM-like procedure runs as follows:

- 1) Consider an initial value of the parameters set  $\Theta^0 = (\Pi^0, A^0, f^0, g^0)$ .

- 2) For each iteration  $q$ , calculate  $\Theta^{q+1}$  from  $\Theta^q$  and  $y$  in two steps:

- a) E Step: Compute  $\alpha_n^q(v_n)$  and  $\beta_n^q(v_n)$ , and then  $\chi_n^q(v_n)$  and  $\xi_n^q(v_n, v_{n+1})$ ;
- b) M Step: Calculate  $\Theta^{q+1}$  as follows:

$$\pi_i^{q+1} = \sum_{u_n \in \Gamma} \chi_1^q(x_n = \omega_i, u_n), \quad (22)$$

$$a_{i,j}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{(u_n, u_{n+1}) \in \Gamma^2} \xi_n^q(x_n = \omega_i, u_n, x_{n+1} = \omega_j, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{u_n \in \Gamma} \chi_n^q(x_n = \omega_i, u_n)}, \quad (23)$$

$$f_{v_m}^{q+1} = \frac{\sum_{n=1}^N \sum_{u_n \in \Gamma} \mathbb{1}_{y_n = v_m} \chi_n^q(x_n = \omega_1, u_n)}{\sum_{n=1}^N \sum_{u_n \in \Gamma} \chi_n^q(x_n = \omega_1, u_n)}, \quad (24)$$

$$g_w^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n \in \Omega \times \Gamma} \mathbb{1}_{w = u_n, y_{n+1} = \varepsilon} \xi_n^q(v_n, x_{n+1} = \omega_2, u_{n+1} = \varepsilon)}{\sum_{n=1}^N \chi_n^q(x_n = \omega_2, u_n = \varepsilon)}. \quad (25)$$

## 4 EXPERIMENTAL STUDY

In this section, we assess the performance of the proposed PTMC against the HMC and PMC models in the task of data segmentation. For this purpose, we present two series of experiments. The first set of experiments, in which we deal with DNA sequences, aims to highlight the prevalence of PTMCs over HMCs and PMCs when real parameters are available. On the other hand, the aim of the second set of experiments, which are concerned with sampled data, is basically to check the validity of the EM parameters estimation procedure.

The segmentation performance may be expressed in terms of mislabeled symbols ratio, precision, recall, or F-measure among other criteria [45], [46]. However, a phase error may have an impact on word symbols even if these latter are assigned to their correct class ( $\omega_2$ ), which strongly affects the data significance while the misclassification ratio may remain high. For this purpose, it would be more relevant to consider word symbols (belonging to  $\omega_2$ ) and having different phase labels as belonging to distinct classes. Accordingly, we will adopt the ‘‘Phasic error ratio’’ that provides the proportion of symbols assigned to a wrong class or assigned correctly as a word symbol but with a wrong phase label. Similarly, we will consider a  $(K+1) \times (K+1)$  Phasic confusion matrix, based on which, performance metrics will be computed.

For this study, we will adopt the commonly used F-measure metric (also known as  $F_1$ -measure) which is the harmonic mean of precision  $p$  and recall  $r$ .

$$F = \frac{2pr}{p+r} \quad (26)$$

The F-measure ranges between 0 and 1 and the higher is its value, the better is the classification quality. The global F-measure is computed here by averaging, over  $K+1$  classes, the F-measure values computed locally on each class  $i$ .

$$F_i = \frac{2p_i r_i}{p_i + r_i}, F = \frac{\sum_{i=1}^{K+1} F_i}{K+1}. \quad (27)$$

### 4.1 Statistical Segmentation of DNA Coding Sequences into Exons/Introns

Deoxyribonucleic acid (DNA) is a molecule that encodes the genetic information of all known living organisms. Such information are encoded as a sequence of nucleotides (guanine, adenine, thymine, and cytosine denoted  $G, A, T$  and  $C$  respectively). Many



works have viewed the biological sequences like DNA as strings on alphabets of four nucleotides  $A, C, G$  and  $T$  [47]. Stochastic grammars such as HMCs can be then useful for various problems for biological sequence analyses [48]. Indeed, biological sequences can be modeled as the output of a process that evolves through a sequence of discrete hidden states. Hence, the applicability of HMCs to computational biology has gained recognition during the last two decades [48]. In [49], an HMM has been applied to *E. Coli* to identify gene structure. HMCs have also been generalized to allow one hidden state to generate more than one symbol [50] which provides a solution to represent genes with their various functional features. In [51], a Bayesian basecalling solution to the hidden Markov model of electropherograms is proposed. An MCMC algorithm is developed to estimate the sequence of bases in the presence of unknown system parameters. For the same sequencing problem, Nelson et al. propose an HMC based approach applied on rice DNA sequences [52]. High-order Markov chains have also been widely considered for DNA sequences modeling [51], [53]. The reader may also refer to [54] where the use of HMCs in biological sequences modeling is reviewed. Finally, let us mention that some systems devoted to DNA sequences segmentation such as VEIL [54] and Genie [55] are based on HMCs.

Let us highlight that the aim of this study is not to propose a formalism that surpasses the existing stochastic approaches devoted to the DNA sequences modeling, but rather to establish that the proposed PTMC outperforms the conventional HMC by taking into account such data particularity. Of course, a more elaborated approach based on PTMC should focus on DNA biological features [48].

For human beings, among other species, only a small fraction of the global genome sequence encodes protein. The coding DNA sequence (CDS) is that portion of a gene's DNA, composed of "exons" (sequence of nucleotides triplets that code for protein), and "introns" (non-meaningful nucleotides located within that portion). An encoding word may be interrupted at a certain site (beginning of an intron) to be resumed at the next occurrence of exon. Hence, DNA sequences are typical examples of data where the application of conventional HMCs suffers from phase shift problem. We will see that the proposed PTMC is well-suited to model this situation.

In this set of experiments, we tackle the problem of human CDS segmentation into introns and exons. For this purpose, we describe how PTMC is used to model CDS and then we evaluate the model ability to perform the task of segmenting a CDS into Introns and Exons. Let  $Y = (Y_1, \dots, Y_N)$  be a CDS with each  $Y_n \in \{A, C, G, T\}$  and let  $X = (X_1, \dots, X_N)$  with  $X_n \in \{I, E\}$  where  $I$  stands for intron and  $E$  for exon. Assuming that  $Z = (X, Y)$  is a three-PTMC denoted  $\lambda_\Theta$  where  $\Theta = (\Pi, A, f, g)$  is a known set of parameters, the aim is to estimate  $x$  from  $y$  using the restoration techniques described in Section 3.3.

We consider 50 different human DNA sequences chosen randomly from the ExonIntron Database (EID) [56], and from which CDSs are extracted. Since each gene has its own properties, we devote a different PTMC (and hence different parameters  $\Theta$ ) to each CDS. For each model, parameters  $\Theta$  are estimated based on the complete data  $z$  as described in Section 3.4. Segmentation is then conducted using MPM and MAP techniques based on the model corresponding to the CDS under consideration. Similarly, the same experiment is conducted through HMCs and PMCs in a classical way. Segmentation results in terms of Error ratio (Err), Phasic error ratio (Ph-Err), F-measure (F) and Phasic F-measure (Ph-F) are reported in Table 1 (Err and Ph-Err are expressed in percent).

The segmentation results confirm the dominance of the proposed PTMC over both HMC and PMC. The difference in terms of performance can be even striking. Notice that MAP restoration

TABLE 1  
Results of CDS Segmentation into Introns/Exons

	HMC		PMC		3-PTMC	
	MPM	MAP	MPM	MAP	MPM	MAP
Err	22.2	23.7	16	16.4	8.8	7.7
Ph-Err	34.9	32.8	35.2	31.1	23.8	8
F	0.69	0.75	0.73	0.71	<b>0.92</b>	<b>0.92</b>
Ph-F	0.46	0.45	0.47	0.47	0.57	<b>0.92</b>

provides the most suitable performance since the aim here is to determine the most probable state sequence whereas the MPM restoration suffers from the phase shift phenomenon due to the data specificity. More precisely, MPM technique being purely local with no consideration for data global structure, the estimated hidden sequence is not necessarily a valid path. On the other hand, MAP technique holistically estimates the most valid path from all the possible ones. According to the defined Phasic F-measure, which is the most significant performance metric here, the proposed PTMC is much more efficient than HMC and PMC.

## 4.2 Statistical Segmentation of Simulated PTMCs

In this set of experiments, we check the PTMC model validity and its associated Bayesian restoration techniques on data sampled according to a PTMC model  $\lambda_\Theta$  where  $\Theta = (\Pi, A, f, g)$  as described in Section 3.2. The restoration results are also compared to those provided by HMC and PMC models. The same experiment is conducted for two different values of the model order  $K = 2$  (Experiment 1) and  $K = 3$  (Experiments 2 and 3). For all experiments  $T = (T_1, \dots, T_N)$  with  $N = 5000$ ,  $\Omega = \{\omega_1, \omega_2\}$  and  $\Upsilon = \{a, b, c\}$ . The data are sampled according to

$$\Pi = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, A = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}.$$

*Experiment 1.*  $f$  has a uniform distribution over  $\Upsilon$ , whereas  $g$  has a uniform distribution over  $\{ab, bc, ca\}$ .

*Experiment 2.*  $f$  has a uniform distribution over  $\Upsilon$ , whereas  $g$  has a uniform distribution over  $\{aab, bac, cab\}$ .

*Experiment 3.*  $f$  has a uniform distribution over  $\Upsilon$ , whereas  $g$  has the following form:  $g(w) = 19/81$  for  $w \in \{aab, abc, acb\}$  and  $g(w) = 1/81$  for  $w$  elsewhere.

For all data sets, restoration of  $x$  from  $y$  is performed according to MPM and MAP techniques using real parameters  $\Theta$  and estimated parameters  $\hat{\Theta}$  (from  $y$  using EM- procedure as described in Section 3.5). For unsupervised parameters estimation, parameters are initialized using random values for  $\Pi$  and  $A$ , and uniform values for  $f$  and  $g$  over their domains. The EM- procedure is run for 15 iterations. Finally, as we deal with sampled data, each experiment is achieved 100 times and average values of performance metrics are summarized in Table 2.

Overall, the results obtained show that the proposed PTMC outperforms the conventional HMC and PMC. The performance of the HMC is bad. In experiment 1, for instance, even when real parameters are available, HMC has no way to discriminate between the two classes since they have the same marginal symbol emission distributions, and hence, symbols are randomly assigned to one of the two classes. On the other hand, PMCs performance is better, particularly when real parameters are available. Finally, the PTMC performs well even in unsupervised context. Notice that in terms of Phasic metrics, MAP yields particularly better results than MPM due to the data nature. Indeed MAP supplies the more likely states sequence that corresponds to the symbols sequence. Such holistic estimation respects the symbols phase property.

TABLE 2  
Restoration Results of Sampled Data

		Experiment 1				Experiment 2				Experiment 3			
		Err	Ph-Err	F	Ph-F	Err	Ph-Err	F	Ph-F	Err	Ph-Err	F	Ph-F
HMC	$\tau_{\theta} - MPM$	44.9	29	0.38	0.25	14.4	42	0.85	0.47	29.8	50.6	0.69	0.32
	$\tau_{\theta} - MAP$	45.7	33.8	0.33	0.24	26.9	44.5	0.71	0.33	38.1	43.7	0.55	0.23
	$\tau_{\theta}^{\wedge} - MPM$	45.1	28.6	0.41	0.28	22.8	40.8	0.74	0.33	36	39.2	0.64	0.35
	$\tau_{\theta}^{\wedge} - MAP$	45.4	34.1	0.36	0.23	33.6	32.5	0.65	0.33	44.1	29	0.39	0.14
	$\tau_{\theta} - MPM$	4.7	32.3	0.95	0.63	8.2	38	0.92	0.50	14.4	39.4	0.85	0.42
PMC	$\tau_{\theta} - MAP$	6.1	30	0.94	0.65	12.2	41.8	0.87	0.51	17.8	40.3	0.81	0.44
	$\tau_{\theta}^{\wedge}$	6	30.7	0.94	0.63	27.2	48.8	0.73	0.39	24	43.1	0.73	0.32
	$\tau_{\theta}^{\wedge} - MPM$	6.9	31.4	0.93	0.62	27.5	47.4	0.73	0.33	24.8	41.7	0.70	0.32
	$\tau_{\theta}^{\wedge} - MAP$	<b>1.7</b>	22	<b>0.99</b>	0.72	<b>1.1</b>	37.6	<b>0.99</b>	0.46	<b>5.4</b>	35.2	<b>0.94</b>	0.54
	$\tau_{\theta} - MPM$	1.9	<b>1.9</b>	0.98	<b>0.98</b>	<b>1.1</b>	<b>1.1</b>	<b>0.99</b>	<b>0.99</b>	6.3	<b>6.3</b>	<b>0.94</b>	<b>0.93</b>
PTMC	$\tau_{\theta} - MAP$	<b>1.7</b>	27.7	0.98	0.65	1.4	35	<b>0.99</b>	0.50	6.1	30.5	<b>0.94</b>	0.54
	$\tau_{\theta}^{\wedge} - MPM$	2	<b>2</b>	0.98	<b>0.98</b>	1.4	<b>1.4</b>	<b>0.99</b>	<b>0.99</b>	<b>6</b>	<b>6</b>	<b>0.94</b>	<b>0.93</b>
	$\tau_{\theta}^{\wedge} - MAP$												

Furthermore, in all experiments, the EM-like procedure of PTMC furnishes good quality estimated parameters and performance metrics in the unsupervised context are comparable to those provided by real parameters whereas the EM-procedure of the HMC and PMC turns out to be inefficient due to the mismatch between the model structure and the data nature.

## 5 CONCLUSION

In this paper, we proposed a new TMC model that takes into account the Phasic nature of some complex data. Unlike memory-less conventional HMCs, the proposed PTMC employs an auxiliary process in order to keep in memory some earlier data, useful to the different estimation tasks of interest. We showed how model parameters can be estimated from complete and incomplete data. We also provided related Bayesian restoration procedures. The model dominance over the usual HMC and PMC models has been checked through experiments conducted on sampled data and DNA sequences.

As future improvements, we intend to extend the PTMC model to consider the situation where words have different lengths. Another interesting direction would be to further investigate the use of the proposed model in the field of DNA sequences alignment and digital data transmission through a communications channel.

## ACKNOWLEDGMENTS

Funding for this research was provided in part by Ecole Militaire Polytechnique, Algeria. Authors would like to thank Marc Uro who helped to use EID database, and Frdric Lehmann for the fruitful discussions about PTMC applicability in communications. Authors would also like to thank reviewers for their constructive comments.

## REFERENCES

- [1] W. Qian and D. Titterton, "On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions," *J. Appl. Stat.*, vol. 16, no. 2, pp. 267–281, 1989.
- [2] C. Carincotte, S. Derrode, and S. Bourenane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 432–441, Feb. 2006.
- [3] M.-Y. Chen, A. Kundu, and J. Zhou, "Off-line handwritten word recognition using a hidden Markov model type stochastic network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 481–496, May. 1994.
- [4] F. Jelinek, *Statistical Methods for Speech Recognition*. Cambridge, MA, USA: MIT press, 1997.
- [5] M. Hatt, F. Lamare, N. Boussian, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest, and D. Visvikis, "Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET," *Phys. Med. Biol.*, vol. 52, no. 12, pp. 3467–3491, 2007.

- [6] C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 360–370, Apr. 1999.
- [7] G. K. Kaleb and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2406–2413, Jul. 1994.
- [8] T. Koski, *Hidden Markov Models for Bioinformatics*, vol. 2. New York, NY, USA: Springer, 2001.
- [9] G. A. Churchill, "Hidden Markov chains and the analysis of genome structure," *Comput. Chem.*, vol. 16, no. 2, pp. 107–115, 1992.
- [10] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, vol. 6. New York, NY, USA: Springer, 2005.
- [11] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *The Ann. Math. Stat.*, vol. 41, pp. 164–171, 1970.
- [12] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.
- [13] 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.
- [14] 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.
- [15] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise," *Signal Process.*, vol. 91, no. 2, pp. 163–175, 2011.
- [16] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 634–639, May 2003.
- [17] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.
- [18] S. Derrode. (2011). On-line demos for recent extensions of HMC model. [Online]. Available: <http://www.fresnel.fr/perso/hmcest/>
- [19] W. Pieczynski, C. Hulard, and T. Veit, "Triplet Markov chains in hidden signal restoration," in *Proc. SPIE's Int. Symp. Remote Sens.*, 2003, pp. 58–68.
- [20] S. Bricq, C. Collet, and J.-P. Armspach, "Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas," in *Proc. 3rd IEEE Int. Symp. Biomed. Imaging: Nano Macro*, 2006, pp. 386–389.
- [21] J. Lapuyade-Lahorgue and W. Pieczynski, "Unsupervised segmentation of hidden semi-Markov non stationary chains," in *Proc. AIP Conf. Proc.*, 2006, vol. 872, pp. 347–354.
- [22] T. M. Nguyen and Q. J. Wu, "Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 42, no. 1, pp. 193–202, Feb. 2012.
- [23] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton University Press, 1976.
- [24] M. E. Y. Boudaren, E. Monfrini, W. Pieczynski, and A. Aissani, "Dempster-Shafer fusion of multisensor signals in nonstationary Markovian context," *EURASIP J. Adv. Sig. Process.*, vol. 2012, p. 134, 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] 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.
- [27] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *Int. J. Approximate Reas.*, vol. 45, no. 1, pp. 1–16, 2007.
- [28] E. Ramasso, "Contribution of belief functions to hidden Markov models with an application to fault diagnosis," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, 2009, pp. 1–6.
- [29] H. Soubaras, "On evidential Markov chains," in *Foundations of Reasoning Under Uncertainty*. New York, NY, USA: Springer, 2010, pp. 247–264.

- [30] B. Ait-El-Fquih and F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2957–2963, Aug. 2006.
- [31] N. Bardel and F. Desbouvries, "Exact Bayesian prediction in a class of Markov-switching models," *Methodol. Comput. Appl. Probability*, vol. 14, no. 1, pp. 125–134, 2012.
- [32] W. Pieczynski, "Exact filtering in conditionally Markov switching hidden linear models," *Comptes Rendus Math.*, vol. 349, no. 9, pp. 587–590, 2011.
- [33] D. Benboudjema and W. Pieczynski, "Unsupervised statistical segmentation of nonstationary images using triplet Markov fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1367–1378, Aug. 2007.
- [34] J. Blanchet and F. Forbes, "Triplet Markov fields for the classification of complex structure data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1055–1067, Jun. 2008.
- [35] L. Gan, Y. Wu, M. Liu, P. Zhang, H. Ji, and F. Wang, "Triplet Markov fields with edge location for fast unsupervised multi-class segmentation of synthetic aperture radar images," *IET Image Process.*, vol. 6, no. 7, pp. 831–838, 2012.
- [36] F. Wang, Y. Wu, Q. Zhang, P. Zhang, M. Li, and Y. Lu, "Unsupervised change detection on SAR images using triplet Markov field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 697–701, Jul. 2013.
- [37] P. Zhang, M. Li, Y. Wu, L. Gan, M. Liu, F. Wang, and G. Liu, "Unsupervised multi-class segmentation of SAR images using fuzzy triplet Markov fields model," *Pattern Recognit.*, vol. 45, no. 11, pp. 4018–4033, 2012.
- [38] P. Zhang, M. Li, Y. Wu, M. Liu, F. Wang, and L. Gan, "SAR image multi-class segmentation using a multiscale TMF model in wavelet domain," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 6, pp. 1099–1103, Nov. 2012.
- [39] D. Benboudjema, N. Othman, B. Dorizzi, and W. Pieczynski, "Challenging eye segmentation using triplet Markov spatial models," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Process.*, 2013, pp. 1927–1931.
- [40] S. Sarawagi and W. W. Cohen, "Semi-Markov conditional random fields for information extraction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, vol. 17, pp. 1185–1192.
- [41] G. D. Forney Jr, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [42] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382, Hoboken, NJ, USA: Wiley, 2007.
- [43] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: An experimental study in the mixture case," *J. Stat. Comput. Simul.*, vol. 55, no. 4, pp. 287–314, 1996.
- [44] 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.
- [45] C. J. van Rijsbergen, *Information Retrieval*. London, U.K.: Butterworth, 1979.
- [46] S. Wang and X. Yao, "Relationships between diversity of classification ensembles and single-class performance measures," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 1, pp. 206–219, Jan. 2013.
- [47] D. B. Searls, "The language of genes," *Nature*, vol. 420, no. 6912, pp. 211–217, 2002.
- [48] Y. Sakakibara, "Grammatical inference in bioinformatics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 7, pp. 1051–1062, Jul. 2005.
- [49] A. Krogh, I. S. Mian, and D. Haussler, "A hidden Markov model that finds genes in *E. coli* DNA," *Nucleic Acids Res.*, vol. 22, no. 22, pp. 4768–4778, 1994.
- [50] G. D. Stormo and D. Haussler, "Optimally parsing a sequence into different classes based on multiple types of evidence," in *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 1994, vol. 2, pp. 369–375.
- [51] K.-C. Liang, X. Wang, and D. Anastassiou, "Bayesian basecalling for DNA sequence analysis using hidden Markov models," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 4, no. 3, pp. 430–440, Jul. 2007.
- [52] R. Nelson, S. Foo, and M. Weatherspoon, "Using hidden Markov modeling in DNA sequencing," in *Proc. 40th Southeastern Symp. Syst. Theory*, 2008, pp. 215–217.
- [53] A. E. Raftery, "A model for high-order Markov chains," *J. Royal Stat. Soc., Ser. B (Methodol.)*, vol. 47, no. 3, pp. 528–539, 1985.
- [54] J. Henderson, S. Salzberg, and K. H. Fasman, "Finding genes in DNA with a hidden Markov model," *J. Comput. Biol.*, vol. 4, no. 2, pp. 127–141, 1997.
- [55] D. K. D. Haussler and M. G. R. F. H. Eeckman, "A generalized hidden Markov model for the recognition of human genes in DNA," in *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 1996, pp. 134–142.
- [56] S. Saxonov, I. Daizadeh, A. Fedorov, and W. Gilbert, "EID: The Exon-Intron databasean exhaustive database of protein-coding intron-containing genes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 185–190, 2000.