



Equivalence between conditional random fields and partially Markov models

Elie Azeraf, Wojciech Pieczynski

► To cite this version:

Elie Azeraf, Wojciech Pieczynski. Equivalence between conditional random fields and partially Markov models. 2023. <hal-04280748v2>

HAL Id: hal-04280748

<https://hal.science/hal-04280748v2>

Preprint submitted on 15 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Equivalence between conditional random fields and partially Markov models

E. Azeraf and W. Pieczynski

Abstract—Hidden Markov models (HMM) have been used successfully in various applications for around sixty years. Moreover, conditional random fields (CRFs) are considered an alternative to HMMs and appear in the literature as different and somewhat competing models. It has recently been shown that basic linear chain CRFs (LC-CRFs), considered different from HMMs, are in fact equivalent to them in the sense that for each LC-CRF there exists an HMM, which can be specified, whose a posteriori distribution is identical to the given LC-CRF. In this article, we extend this result to general CRFs. We consider partially Markovian models (PMM), which are sequentially defined models extending HMMs, and we show that they are equivalent to CRFs. Equivalence means that the posterior distributions of the PMMs are CRFs and, conversely, for each given CRF there exists a PMM – which we specify – whose posterior distribution is the given CRF. The first assertion is well known and the proof of the second constitutes the main contribution of this article. The advantage of presenting a CRF in its PMM form is that the latter is sequential, whereas its original former is not. This property is essential in certain sequential processing, such as filtering or forecasting.

Index Terms— Conditional random fields, Filtering, Hidden Markov models, Pairwise Markov chains, Partially Markov models.

----- ◆ -----

1 INTRODUCTION

LET $Z_{1:N} = (Z_1, \dots, Z_N)$ be a stochastic sequence, with $Z_n = (X_n, Y_n)$. Random variables X_1, \dots, X_N take their values in a finite set Λ , while Y_1, \dots, Y_N take their values either in a discrete or continuous set Ω . Probabilistic model is a distribution – or a family of distributions – which will be denoted with $p(z_{1:N})$, or $p(x_{1:N}, y_{1:N})$. More generally, for $1 \leq m < n \leq N$, we will note $z_{m:n} = (z_m, \dots, z_n)$, and similarly for $x_{m:n}, y_{m:n}$. Conditional distributions will be denoted with $p(\cdot | \cdot)$. For example, we have $p(x_n | y_{1:N}) = P[X_n = x_n | Y_{1:N} = y_{1:N}]$, $p(x_{n+1} | x_{1:n}) = P[X_{n+1} = x_{n+1} | X_{1:n} = x_{1:n}]$, and so on.

The aim of the paper is to show the equivalence between sequential models called “Partially Markov models” (PMMs, [39]) and non-sequential models called “Conditional random fields” (CRFs, [21], [47]). The interest is that sequential PMMs are easier to manipulate than non-sequential CRFs; in particular, they allow new observations to be taken into account more easily. This contribution shows that apparently less general PMMs have actually the same modeling power as CRFs. PMMs are defined by the distribution $q(x_{1:N}, y_{1:N})$ verifying (1.1), while CRFs considered in this paper are defined by the distribution $p(x_{1:N} | y_{1:N})$ verifying (1.2).

$$q(x_{1:N}, y_{1:N}) = \quad (1.1)$$

$$q(x_1)q(y_1|x_1) \prod_{n=1}^{N-1} q(x_{n+1}, y_{n+1}|x_n, y_{1:n}).$$

$$p(x_{1:N} | y_{1:N}) = \quad (1.2)$$

$$p(x_1 | y_{1:N}) \prod_{n=1}^{N-1} p(x_{n+1} | x_n, y_{1:n}),$$

In the whole paper, we adopt the following definitions:

Definition 1.1

(i) a distribution $q(x_{1:N}, y_{1:N})$ is “equivalent” to a distribution $p(x_{1:N} | y_{1:N})$ if and only if (iff)

$$q(x_{1:N} | y_{1:N}) = p(x_{1:N} | y_{1:N}); \quad (1.3)$$

(ii) two distributions $r(x_{1:N}, y_{1:N})$, $s(x_{1:N}, y_{1:N})$ are equivalent iff

$$r(x_{1:N} | y_{1:N}) = s(x_{1:N} | y_{1:N}). \quad (1.4)$$

Let us consider the following popular CRFs distribution:

$$p(x_{1:N} | y_{1:N}) = ce^{-[\sum_{n=1}^{N-1} \varphi(x_n, x_{n+1}, y_{1:N}) + \sum_{n=1}^N \phi(x_n, y_{1:N})]}, \quad (1.5)$$

with c constant and φ, ϕ real valued functions. Actually, such distributions are particular cases of (1.2); indeed, they are (1.2) with the conditions $p(x_1 | y_{1:N}) > 0$ for $p(x_1 | y_{1:N})$, and $p(x_{n+1} | x_n, y_{1:n}) > 0$ for every $n = 1, \dots, N - 1$.

It is easy to see that the posterior distribution $q(x_{1:N} | y_{1:N})$ of a PMM (1.1) is a CRF (1.2); the converse proposition is trickier. The problem we deal with is to answer the following questions. Having a CRF $p(x_{1:N} | y_{1:N})$, does there exist a PMM $q(x_{1:N}, y_{1:N})$ such

• E. Azeraf is with Watson Department, IBM France. E-mail elie.azeraf@gmail.com
 • W. Pieczynski is with Samovar Telecom Sudparis Institut Polytechnique de Paris, Palaiseau, France. E-mail: wojciech.pieczynski@telecom-sudparis.eu

that $q(x_{1:N}|y_{1:N}) = p(x_{1:N}|y_{1:N})$? If so, can probabilities $q(x_1)q(y_1|x_1)$, $q(x_{n+1}, y_{n+1}|x_n, y_{1:n})$ in (1.1) be calculated from the probabilities $p(x_1|y_{1:N})$, $p(x_{n+1}|x_n, y_{1:N})$ in (1.2)? Can we specify all the solutions of the problem? We answer these questions in the theorem in the next section. Let us remark that one immediate practical consequence is that if we wish to estimate of $X_{1:N}$ from $Y_{1:N}$ by a Bayesian classifier, we can use either CRFs or any equivalent PMMs. Indeed, Bayesian estimators only depend on the *a posteriori* distribution $p(x_{1:N}|y_{1:N})$.

CRFs found huge deal of applications, and continue to be widely used as attested by hundreds journal papers appearing each year. Examples include image segmentation [11], [4], [23], named entity recognition [32], or still part-of-speech tagging [15]. Let us also cite [29] as review article and [19], [47] as books, among others. PMMs have been little studied so far; we can cite [22] where they were used to estimate non-stationary data hidden with long memory noise. This opens up new possibilities, as it means that sequential PMMs can be used in place of non-sequential CRFs without losing CRFs modelling power.

The equivalence between CRFs and PMMs makes it possible to show the equivalence between "hidden" CRFs (HCRFs, [41], [42]) and "triplet" PMMs (called triplet partially Markov chains in [39]). HCRFs consist of adding a latent variable $U_{1:N}$, and considering the triplet $(X_{1:N}, U_{1:N}, Y_{1:N})$. By setting $V_{1:N} = (X_{1:N}, U_{1:N})$, HCRF is given assuming that $(V_{1:N}, Y_{1:N})$ is a CRF. HCRFs are more general than CRFs because in HCRF $(X_{1:N}, U_{1:N}, Y_{1:N})$ the couple $(X_{1:N}, U_{1:N})$ is Markov (conditionally on $Y_{1:N}$), and $X_{1:N}$ may be Markovian or may not be Markovian. HCRFs are applicable in all situations CRFs are, and there is an abundant literature showing their interest. In particular, they have been applied to telephone classification [17], [45], detection of cell division [36], human action recognition [51], [52], human facial expression recognition [44], gesture recognition [48], fault diagnosis [49], CET image segmentation [50], and many others. A similar idea has been proposed for PMMs in [39], leading to "triplet" PMMs (TPMMs). TPMMs have been applied to non-stationary Markov chains hidden with long memory noise in [22]. According to the theorem, we can state that HCRFs and TPMMs are equivalent. This opens up enormous possibilities of applications of TPMMs in situations where HCRFs have performed well and where the sequential nature of TPMMs is of interest

Once the equivalence between PMMs and CRFs has been established, we propose in section III the following contribution, which is related to particular PMMs called

"pairwise Markov chains" (PMCs, [38]). The distribution of PMCs is given by

$$q(x_{1:N}, y_{1:N}) = q(x_1)q(y_1|x_1) \prod_{n=1}^{N-1} q(x_{n+1}, y_{n+1}|x_n, y_n), \quad (1.6)$$

which simply means that the stochastic sequence $Z_{1:N}$ of pairs $Z_n = (X_n, Y_n)$ is Markovian. PMCs extend the classic hidden Markov models (HMMs), which are still widely used [5], [43], [9], [13], among others. The distribution of HMMs is written as

$$q(x_{1:N}, y_{1:N}) = q(x_1)q(y_1|x_1) \prod_{n=1}^{N-1} q(x_{n+1}|x_n)q(y_{n+1}|x_{n+1}). \quad (1.7)$$

As

$$q(x_{n+1}, y_{n+1}|x_n, y_n) = q(x_{n+1}|x_n, y_n)q(y_{n+1}|x_n, y_n, x_{n+1}), \quad (1.8)$$

(1.6) and (1.7) mean that HMMs are PMCs in which $q(x_{n+1}|x_n, y_n) = q(x_{n+1}|x_n)$ and $q(y_{n+1}|x_n, y_n, x_{n+1}) = q(y_{n+1}|x_{n+1})$. This shows that PMCs are clearly more general than HMMs and, indeed, the interest of the former over the latter is highlighted in various articles. In particular, they have been applied to image segmentation [1], [2], [16], [38], sound classification [24], fuzzy segmentation [24], or multiple target tracking [30], [31], [37]. An extension to apply Bayesian variational Bayesian methods is proposed in [34], and some theoretical studies can be found in [20], [25], [26], [27]. Since PMCs are applicable under the same conditions as HMMs while being more general, it is interesting to ask which family of CRFs is equivalent to PMCs. Our contribution consists in characterizing the family of CRFs equivalent to PMCs. More precisely, we show that a CRF $p(x_{1:N}|y_{1:N})$ is equivalent to a PMC $q(x_{1:N}, y_{1:N})$ iff $p(x_{n+1}|x_n, y_{1:N}) = p(x_{n+1}|x_n, y_{n+1:N})$ for every $n = 1, \dots, N-1$. Subsequently, we give some examples showing that some classic CRFs are equivalent to PMCs. In particular, applying this result to the simple CRF whose distribution is defined with

$$p(x_{1:N}|y_{1:N}) = \kappa(y_{1:N}) \exp[\sum_{n=1}^{N-1} V_n(x_n, x_{n+1}) + \sum_{n=1}^N U_n(x_n, y_n)], \quad (1.9)$$

where $\kappa(y_{1:N})$ is the normalizing constant, we find again the fact that it is equivalent to classic HMM (1.8), which has been recently established in a direct way in [3]. Subsequently, another case studied shows that the classic Minimum Entropy Markov Models (MEMMs [33]) are equivalent to particular PMCs.

Dependence graphs of HMMs, PMCs, and PMMs are presented in Figure.

This article is organized as follows. In the next section, we show the equivalence, in the sense of definition 1.1, between PMMs (1.1) and CRFs (1.2). In Section III, we specify the CRFs equivalent to PMCs. We also specify the family of CRFs equivalent to triplet Markov chains (TMCs [40]). We present conclusions and perspectives in Section IV.

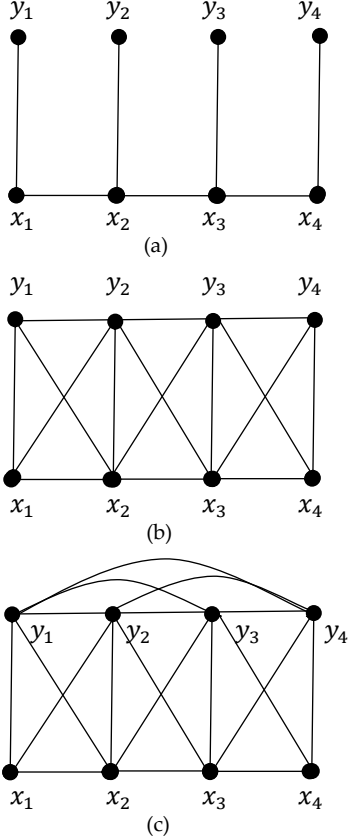


Figure. Dependence graphs. (a): hidden Markov model (HMM); (b) pairwise Markov chain (PMC); (c) partially Markov model (PMM).

2. EQUIVALENCE BETWEEN PMMS AND CRFS

In this paragraph, we show that for each CRF (1.2) there exists an equivalent PMM (1.1). More precisely, we specify how to compute all distributions $q(x_1)$, $q(y_1|y_1)$, $q(x_2, y_2|x_1, y_1)$, $q(x_3, y_3|x_1, y_{1:2})$, ..., $q(x_N, y_N|x_{N-1}, y_{1:N-1})$, defining all equivalent PMMs (1.1), from $p(x_1|y_{1:N})$, $p(x_2|x_1, y_{1:N})$, ..., $p(x_N|x_{N-1}, y_{1:N})$ defining the CRF (1.2) under consideration.

We will use the following Lemma 1.

Lemma 1. Let $W_{1:N} = (W_1, \dots, W_N)$ be random sequence, taking its values in a finite set Δ . Then

(i) $W_{1:N}$ is a Markov chain iff there exist $N - 1$ functions $\varphi_1, \dots, \varphi_{N-1}$ from Δ^2 to \mathbb{R}^+ such that

$$p(w_1, \dots, w_N) = K \varphi_1(w_1, w_2) \dots \varphi_{N-1}(w_{N-1}, w_N), \quad (2.1)$$

with K constant;

(ii) for HMM defined with $\varphi_1, \dots, \varphi_{N-1}$ verifying (2.1), $p(w_1)$, and $p(w_{n+1}|w_n)$ for $n = 2, \dots, N - 1$, are given with

$$p(w_1) = \frac{\beta_1(w_1)}{\sum_{w_1} \beta_1(w_1)};$$

$$p(w_{n+1}|w_n) = \frac{\varphi_n(w_n, w_{n+1}) \beta_{n+1}(w_{n+1})}{\beta_n(w_n)}, \quad (2.2)$$

where $\beta_1(w_1), \dots, \beta_N(w_N)$ verify the backward recursion

$$\beta_N(w_N) = 1,$$

$$\beta_n(w_n) = \sum_{w_{n+1}} \varphi_n(w_n, w_{n+1}) \beta_{n+1}(w_{n+1}). \quad (2.3)$$

Proof. 1. Let $W_{1:N}$ be Markov: $p(w_1, \dots, w_N) = p(w_1)p(w_2|w_1)p(w_3|w_2) \dots p(w_N|w_{N-1})$. Then (2.1) is verified by $\varphi_1(w_1, w_2) = p(w_1)p(w_2|w_1)$, $\varphi_2(w_2, w_3) = p(w_3|w_2)$, ..., $\varphi_{N-1}(w_{N-1}, w_N) = p(w_N|w_{N-1})$.

2. Conversely, let $p(w_1, \dots, w_N)$ verifies (2.1). This implies that for each $n = 1, \dots, N - 1$ we have

$$p(w_{n+1}|w_1, \dots, w_n) = \frac{p(w_1, \dots, w_n, w_{n+1})}{p(w_1, \dots, w_n)} = \quad (2.4)$$

$$\frac{\sum_{(w_{n+2}, \dots, w_N)} \varphi_1(w_1, w_2) \dots \varphi_n(w_n, w_{n+1}) \varphi_{n+1}(w_{n+1}, w_{n+2}) \dots \varphi_{N-1}(w_{N-1}, w_N)}{\sum_{(w_{n+1}, \dots, w_N)} \varphi_1(w_1, w_2) \dots \varphi_n(w_n, w_{n+1}) \varphi_{n+1}(w_{n+1}, w_{n+2}) \dots \varphi_{N-1}(w_{N-1}, w_N)} =$$

$$\frac{\varphi_n(w_n, w_{n+1}) \sum_{(w_{n+2}, \dots, w_N)} \varphi_{n+1}(w_{n+1}, w_{n+2}) \dots \varphi_{N-1}(w_{N-1}, w_N)}{\sum_{(w_{n+1}, w_{n+2}, \dots, w_N)} \varphi_n(w_n, w_{n+1}) \varphi_{n+1}(w_{n+1}, w_{n+2}) \dots \varphi_{N-1}(w_{N-1}, w_N)} =$$

$$p(w_{n+1}|w_n),$$

which shows that $p(w_1, \dots, w_N)$ is Markov.

Besides, let us set, for $n = 1, \dots, N - 1$:

$\beta_n(w_n) = \sum_{(w_{n+1}, \dots, w_N)} \varphi_n(w_n, w_{n+1}) \dots \varphi_{N-1}(w_{N-1}, w_N)$
On the one hand, we see that $\beta_n(w_n) = \sum_{w_{n+1}} \varphi_n(w_n, w_{n+1}) \beta_{n+1}(w_{n+1})$. On the other hand, according to (2.4) we have $p(w_{n+1}|w_n) = \frac{\varphi_n(w_n, w_{n+1}) \beta_{n+1}(w_{n+1})}{\beta_n(w_n)}$. As $p(w_1) = \frac{\beta_1(w_1)}{\sum_{w_1} \beta_1(w_1)}$, (2.2) and (2.3) are verified, which ends the proof. \square

Proposition 2.1 Let $X_{1:N} = (X_1, \dots, X_N)$, $Y_{1:N} = (Y_1, \dots, Y_N)$ be two stochastic sequences taking their values in finite sets Λ, Ω , respectively.

Let $q(x_{1:N}, y_{1:N})$ be a PMM of form (1.1). Then $q(x_{1:N}|y_{1:N})$ is a of form (1.2), with

$$q(x_{n+1}|x_n, y_{1:N}) = \frac{q(x_{n+1}, y_{n+1}|x_n, y_{1:n}) \beta_{n+1}(x_{n+1})}{\beta_n(x_n)}, \quad (2.5)$$

where $\beta_n(x_n)$ are defined with

$$\beta_n(x_n) = q(y_{n+1:N}|x_n, y_{1:n}), \quad (2.6)$$

and can be computed recursively with

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

$$\beta_n(x_n) = \sum_{x_{n+1}} q(x_{n+1}, y_{n+1}|x_n, y_{1:n}) \beta_{n+1}(x_{n+1}) \quad (2.7)$$

Proof. The proof is immediate applying Lemma 1 to $X_{1:N} = W_{1:N}$, $\varphi_1(x_1, x_2) = q(x_1, y_1)q(x_2, y_2|x_1, y_{1:2})$, $\varphi_2(x_2, x_3) = q(x_3, y_3|x_2, y_{1:2})$, ..., $\varphi_{N-1}(x_{N-1}, x_N) = q(x_N, y_N|x_{N-1}, y_{1:N})$ \square

The theorem below specify the reverse proposition, which is more difficult to prove and is the core contribution of the paper.

For a given PMM $q(x_{1:N}, y_{1:N})$, let us consider the "backward" probabilities

$$\beta_N(x_N) = 1, \text{ and} \quad (2.8)$$

$$\beta_n(x_n) = q(y_{n+1:N}|x_n, y_{1:n}) \text{ for } n = N-1, \dots, 1.$$

Similarly to what is true in classic HMMs, they verify (2.7), rewritten as

$$\beta_N(x_N) = 1, \text{ and} \quad (2.9)$$

$$\beta_n(x_n) = \quad (2.10)$$

$$\sum_{x_{n+1}} q(y_{n+1}|x_n, y_{1:n})q(x_{n+1}|x_n, y_{1:n+1})\beta_{n+1}(x_{n+1}).$$

We specify the converse proposition in Theorem below:

Theorem. Let

$$p(x_{1:N}|y_{1:N}) = \quad (2.11)$$

$$p(x_1|y_{1:N})p(x_2|x_1, y_{1:N}) \dots p(x_N|x_{N-1}, y_{1:N}).$$

be a CRF. Then the PMM

$$q(x_{1:N}, y_{1:N}) = \quad (2.12)$$

$$q(y_1)q(x_1|y_1) \prod_{n=1}^{N-1} q(y_{n+1}|x_n, y_{1:n})q(x_{n+1}|x_n, y_{1:n+1})$$

is equivalent to $p(x_{1:N}|y_{1:N})$ if and only if

(i) sequence

$$q(y_1), q(y_2|x_1, y_1), \dots, q(y_N|x_{N-1}, y_{1:N-1}) \quad (2.13)$$

is arbitrary;

(ii) for given sequence (2.13),

$$q(x_1|y_1), q(x_2|x_1, y_{1:2}), \dots, q(x_N|x_{N-1}, y_{1:N})$$

is defined from $p(x_{1:N}|y_{1:N})$ and (2.13) with

the following backward recursion:

$$q(x_N|x_{N-1}, y_{1:N}) = p(x_N|x_{N-1}, y_{1:N}); \quad (2.14)$$

for $n = N-1, \dots, 2$:

$$q(x_n|x_{n-1}, y_{1:n}) = \left[\sum_{y_{n+1:N}} \frac{\beta_n(x_n)}{p(x_n|x_{n-1}, y_{1:n})} \right]^{-1}, \quad (2.15)$$

where $\beta_n(x_n)$ is computed from $q(y_{n+1}|x_n, y_{1:n})$, $q(x_{n+1}|x_n, y_{1:n+1})$, and $\beta_{n+1}(x_{n+1})$ with (2.7);

$$q(x_1|y_1) = \left[\sum_{y_{2:N}} \frac{q(y_{2:N}|x_1, y_1)}{p(x_1|y_{1:N})} \right]^{-1}. \quad (2.16)$$

Proof. Let $p(x_{1:N}|y_{1:N})$ be a given CRF (2.11), and $q(y_1)$, $q(y_2|x_1, y_1)$, ..., $q(y_N|x_{N-1}, y_{1:N-1})$ arbitrary sequence.

We search a PMM $q(x_{1:N}, y_{1:N})$ (2.12) verifying $q(x_{1:N}|y_{1:N}) = p(x_{1:N}|y_{1:N})$. This equality is equivalent to

$$\begin{aligned} q(x_N|x_{N-1}, y_{1:N}) &= p(x_N|x_{N-1}, y_{1:N}), \\ q(x_{N-1}|x_{N-2}, y_{1:N}) &= p(x_{N-1}|x_{N-2}, y_{1:N}), \\ &\dots, \\ q(x_2|x_1, y_{1:N}) &= p(x_2|x_1, y_{1:N}), \\ q(x_1|y_{1:N}) &= p(x_1|y_{1:N}). \end{aligned} \quad (2.17)$$

Let's number the equations (2.17) as follows: equation (1) for the one in the first line, equation (2) for the one in the second line, ..., equation (N-1) for the last equation. To show that all the equations are true with $q(x_N|x_{N-1}, y_{1:N})$, ..., $q(x_1|y_{1:N})$ verifying (2.15), we show that the first equation in (2.17) is true, and we show that if the first $N-n$ equations (for $1 \leq n \leq N-1$) are true with $q(x_N|x_{N-1}, y_{1:N})$, ..., $q(x_1|y_{1:N})$ verifying (2.15), then the $(N-n-1)$ th equation also is.

As

$$\begin{aligned} q(x_N, y_N|x_{N-1}, y_{1:N-1}) &= \\ q(y_N|x_{N-1}, y_{1:N-1})q(x_N|x_{N-1}, y_{1:N}), \end{aligned}$$

we see that the first equality in (2.17) is verified for $q(x_N|x_{N-1}, y_{1:N}) = p(x_N|x_{N-1}, y_{1:N})$ and for any $q(y_N|x_{N-1}, y_{1:N-1})$.

Let us assume that $q(x_N|x_{N-1}, y_{1:N})$, ..., $q(x_{n+1}|x_n, y_{1:n+1})$ verify (2.16), and verify (2.17) for $N-n$ first equations. Let us search all $q(x_n, y_n|x_{n-1}, y_{1:n-1})$ such that (2.16) is verified for $(N-n-1)$ th equation $q(x_n|x_{n-1}, y_{1:N}) = p(x_n|x_{n-1}, y_{1:N})$.

We have

$$\begin{aligned} q(x_n|x_{n-1}, y_{1:N}) &= q(x_n|x_{n-1}, y_{1:n}, y_{n+1:N}) = \\ \frac{q(x_n, y_{n+1:N}|x_{n-1}, y_{1:n})}{q(y_{n+1:N}|x_{n-1}, y_{1:n})} &= \quad (2.18) \\ \frac{q(x_n|x_{n-1}, y_{1:n})q(y_{n+1:N}|x_{n-1}, x_n, y_{1:n})}{q(y_{n+1:N}|x_{n-1}, y_{1:n})}. \end{aligned}$$

As q is a PMM, we can write

$$q(y_{n+1:N}|x_{n-1}, x_n, y_{1:n}) = q(y_{n+1:N}|x_n, y_{1:n}), \quad (2.19)$$

so that finally we have

$$q(x_n|x_{n-1}, y_{1:N}) = \frac{q(x_n|x_{n-1}, y_{1:n})q(y_{n+1:N}|x_n, y_{1:n})}{q(y_{n+1:N}|x_{n-1}, y_{1:n})} \quad (2.20)$$

Then the question is: knowing that $q(y_{n+1:N}|x_n, y_{1:n})$ in (2.20) is given (recall that it is equal to $\beta_n(x_n)$, given with recursion (2.7) from $q(x_N|x_{N-1}, y_{1:N})$, ...,

$q(x_{n+1}|x_n, y_{1:n+1})$ and $q(y_N|x_{N-1}, y_{1:N-1})$, ..., $q(y_{n+1}|x_n, y_{1:n})$, the latter sequence being arbitrary but fixed in the proof), can we find $q(x_n|x_{n-1}, y_{1:n})$ such that (2.20) holds with $q(x_n|x_{n-1}, y_{1:n}) = p(x_n|x_{n-1}, y_{1:n})$? Consider $r(\cdot)$ distributions on Λ conditional on $(x_{n-1}, y_{1:n})$, and let us set $v = x_n$, $w = y_{n+1:N}$. The question above is equivalent to: can we find $r(v)$ such that

$$r(v|w) = \frac{r(v)r(w|v)}{r(w)}? \quad (2.21)$$

In other words, we have $r(v|w)$ and $r(w|v)$, and we search $r(v)$ verifying (2.21). According to Lemma 2 below, the solution is

$$r(v) = \left[\sum_w \frac{r(w|v)}{r(v|w)} \right]^{-1}. \quad (2.22)$$

Since $q(y_{n+1:N}|x_n, y_{1:n}) = \beta_n(x_n)$, (2.22) is equivalent to (2.15). Finally, (2.16) is obtained applying Lemma 2 to $r(\cdot)$ distributions on Λ conditional on y_1 , and setting $v = x_1$, $w = y_{2:N}$, which completes the proof. \square

Lemma 2. Let V, W be two discrete finite random variables taking their values in Λ . Let $r(w|v)$, $r(v|w)$ be two conditional distributions. Then $r(v)$ is given by (2.22).

Proof. Since $r(v, w) = r(w|v)q(v) = r(v|w)p(w)$, we have $r(w) = \frac{r(w|v)}{r(v|w)}r(v)$, and hence $\sum_w r(w) = r(v) \sum_w \frac{r(w|v)}{r(v|w)} = 1$, which gives (2.22) and completes the proof. \square

Remark 2.1 CRFs and PMMs are equivalent; however, they are parameterized differently and can therefore give different results in real applications. Indeed, in real applications, one has to estimate the parameters in a preliminary step. Since the parameters are different in CRFs and PMMs, the estimators used are different and may be more or less suitable for the intended application. Thus, even in real-life situations where the sequential nature of PMMs is not of importance, there is no reason to prefer one family of models to the other. Therefore, even in situations where the sequential nature of PMMs is not important and where CRFs - or HCRFs, see remark 2.2 below - give excellent results, it is still worth applying PMMs - or TPMMs - for comparison.

Remark 2.2 As mentioned in Introduction, equivalence between CRFs and PMMs implies equivalence between "hidden" CRFs (HCRFs, [41], [42]) and "triplet" PMMs (TPMMs, [39]). Indeed, HCRFs (or TPMMs) $(X_{1:N}, U_{1:N}, Y_{1:N})$ also are CRFs (or PMMs) $(V_{1:N}, Y_{1:N})$, with $V_{1:N} = (X_{1:N}, U_{1:N})$.

3. CRFs EQUIVALENT TO PMCS

Consider the "pairwise Markov chains" given by (1.7). As PMCs are an extension of HMMs, they are applicable in the same situations and are likely to improve the results provided by HMMs. As we shall see, some CRFs presented as alternative to HMMs are equivalent to PMCs. This is the case for Maximum Entropy Markov Models (MEMMs) [33].

Proposition 3.1 below, which we can consider as a corollary of the theorem, characterizes the family of CRFs equivalent to the family of PMCs.

Proposition 3.1 A CRF (1.2) is equivalent to a PMC (1.6) if and only if for $n = 1, \dots, N-1$,

$$p(x_{n+1}|x_n, y_{1:N}) = p(x_{n+1}|x_n, y_{n+1:N}). \quad (3.1)$$

Proof. 1. Let $q(x_{1:N}, y_{1:N})$ be a PMC verifying $q(x_{1:N}|y_{1:N}) = p(x_{1:N}|y_{1:N})$. Thus for $n = 1, \dots, N-1$, $q(x_{n+1}|x_n, y_{1:N}) = p(x_{n+1}|x_n, y_{1:N})$. Setting $\beta_n(x_n) = q(y_{n+1:N}|x_n, y_{1:n})$ for $n = 1, \dots, N-1$, and $\beta_N(x_N) = 1$, we have, according to the classic properties of PMCs,

$$\beta_n(x_n) = \sum_{x_{n+1}} q(y_{n+1}|x_n, y_n) q(x_{n+1}|x_n, y_{n+1:N}) \beta_{n+1}(x_{n+1}), \text{ and}$$

$$q(x_{n+1}|x_n, y_{1:N}) = \frac{q(y_{n+1:N}|x_n, y_n) q(x_{n+1}|x_n, y_{n+1:N}) \beta_{n+1}(x_{n+1})}{\beta_n(x_n)} =$$

$$q(x_{n+1}|x_n, y_{n+1:N}).$$

Thus $p(x_{n+1}|x_n, y_{1:N}) = p(x_{n+1}|x_n, y_{n+1:N})$, which is (3.1);

2. Let $p(x_{1:N}|y_{1:N})$ be a CRF (1.2) verifying (3.1). We have $q(x_N|x_{N-1}, y_{1:N}) = p(x_N|x_{N-1}, y_{1:N})$.

Conversely, let us show that $q(x_{n+1}|x_n, y_{1:n+1}) = q(x_{n+1}|x_n, y_{n+1})$ implies $q(x_n|x_{n-1}, y_{1:n}) = q(x_n|x_{n-1}, y_n)$. We look for a PMC in which $q(y_{n+1:N}|x_n, y_{1:n}) = q(y_{n+1:N}|x_n, y_n)$. Then (2.15) gives

$$q(x_n|x_{n-1}, y_{1:n}) = \left[\sum_{y_{n+1:N}} \frac{q(y_{n+1:N}|x_n, y_{1:n})}{p(x_n|x_{n-1}, y_{1:n})} \right]^{-1} = \left[\sum_{y_{n+1:N}} \frac{q(y_{n+1:N}|x_n, y_n)}{p(x_n|x_{n-1}, y_{n:N})} \right]^{-1} = \frac{p(x_n|x_{n-1}, y_{n:N})}{\sum_{y_{n+1:N}} q(y_{n+1:N}|x_n, y_n)} = p(x_n|x_{n-1}, y_{n:N}),$$

which completes the proof. \square

Example 3.1 Let $p(x_{1:N}|y_{1:N})$ be a CRF defined with $p(x_1|y_{1:N}), p(x_2|x_1, y_{1:N}), \dots, p(x_N|x_{N-1}, y_{1:N})$. What is the simplest equivalent PMM $q(x_{1:N}, y_{1:N})$? Since $q(y_1), q(y_2|x_1, y_1), \dots, q(y_N|x_{N-1}, y_{1:N-1})$ are arbitrary, let us take them all equal to the constant $c = \frac{1}{\#\Lambda}$. We note that in such a PMM Y_1, \dots, Y_N are independent and equi-distributed, which may seem somewhat unusual. However, this has no bearing on the problem of finding

$X_{1:N}$ from $Y_{1:N}$ with Bayesian classifiers; indeed, they are independent from $p(x_{1:N}, y_{1:N})$. We have

$$q(x_{1:N}, y_{1:N}) = q(x_1, y_1) \prod_{n=1}^{N-1} q(x_{n+1}, y_{n+1} | x_n, y_{1:n}) = c^N q(x_1 | y_1) \prod_{n=1}^{N-1} q(x_{n+1} | x_n, y_{1:n+1}) \quad (3.2)$$

According to the Theorem, $q(x_{n+1} | x_n, y_{1:n+1})$ are computable by the following backward recursion:

$$\begin{aligned} - \beta_N(x_N) &= 1, q(x_N | x_{N-1}, y_{1:N}) = p(x_N | x_{N-1}, y_{1:N}); \\ - \beta_n(x_n) &= c \sum_{x_{n+1}} q(x_{n+1} | x_n, y_{1:n}) \beta_{n+1}(x_{n+1}) = c^{N-n}; \\ - q(x_n | x_{n-1}, y_{1:n}) &= c^{N-n} \left[\sum_{y_{n+1:N}} \frac{1}{p(x_n | x_{n-1}, y_{1:N})} \right]^{-1}; \\ - q(x_1 | y_1) &= \left[\sum_{y_{2:N}} \frac{q(y_{2:N} | x_1, y_1)}{p(x_1 | y_{1:N})} \right]^{-1}. \end{aligned}$$

Knowing that the classifiers only depend on $p(x_{1:N} | y_{1:N})$, this shows that any “generative” classifier defined from a PMM can be calculated in a discriminative way, by replacing in the calculations $q(y_2 | x_1, y_1), \dots, q(y_N | x_{N-1}, y_{1:N-1})$ by $c = \frac{1}{\#\Lambda}$, which extends to PMMs the results related to Maximum a posteriori (MAP) and Maximum posterior mode (MPM) classifiers presented in the framework of HMMs in [3].

Example 3.2 In classic HMM $q(x_{1:N}, y_{1:N})$ we have $q(x_{n+1}, y_{n+1} | x_n, y_n) = q(x_{n+1} | x_n) q(y_{n+1} | x_{n+1})$. Then

$$q(y_{n+1} | x_n, y_n) = \sum_{x_{n+1}} q(x_{n+1} | x_n) q(y_{n+1} | x_{n+1}); \quad (3.3)$$

$$q(x_{n+1} | x_n, y_n, y_{n+1}) = \frac{q(x_{n+1} | x_n) q(y_{n+1} | x_{n+1})}{\sum_{x_{n+1}} q(x_{n+1} | x_n) q(y_{n+1} | x_{n+1})}. \quad (3.4)$$

We remark that $q(x_{n+1} | x_n, y_n, y_{n+1}) = q(x_{n+1} | x_n, y_{n+1})$. As in example 3.1, replacing $q(y_{n+1} | x_n, y_n)$ with $c = \frac{1}{\#\Lambda}$ will not change $q(x_{1:N} | y_{1:N})$, so PMM $r(x_{1:N}, y_{1:N})$ defined with

$$r(x_{1:N}, y_{1:N}) = c^N q(x_1 | y_1) \prod_{n=1}^{N-1} q(x_{n+1} | x_n, y_{n+1}) \quad (3.5)$$

is equivalent to the HMM $q(x_{1:N}, y_{1:N})$. As above, we arrive at a model $r(x_{1:N}, y_{1:N})$, which may seem odd, in which Y_1, \dots, Y_N are independent and uniformly distributed on Λ , with $r(x_{n+1} | x_n, y_n, y_{n+1})$ transitions of a classic HMM. Of course, such a PMM is not an HMM.

Example 3.3 Let us consider the Maximum Entropy Markov Models (MEMMs) $p(x_{1:N} | y_{1:N})$ introduced in [33], which can be seen as ancestor of CRFs. They verify

$$p(x_{1:N} | y_{1:N}) = p(x_1 | y_1) p(x_2 | x_1, y_2) \dots p(x_N | x_{N-1}, y_N) \quad (3.6)$$

MEMMs have been proposed as alternative to HMMs. MEMMs are neither extensions of HMMS nor particular HMMs. Indeed, in some respects the CRF given with HMM (simply its posterior distribution $p(x_{1:N} | y_{1:N})$) is more general than MEMM: for example, in MEMM $p(x_n | y_{1:N}) = p(x_n | y_{1:n})$, whereas in the HMM $p(x_n | y_{1:N})$ depends on all components of $y_{1:N}$. In other respects the

CRF given with HMM is less general than the MEMM. For example, in MEMM $p(x_{n+1} | x_n, y_{n+1})$ are of any form, while in HMM, according to (1.4), they are proportional to the product $p(x_{n+1} | x_n) p(y_{n+1} | x_{n+1})$, which is therefore a particular form.

Moreover, we can notice that MEMM is a CRF verifying $p(x_{n+1} | x_n, y_{1:N}) = p(x_{n+1} | x_n, y_{n+1})$. Therefore, according to Proposition 2, MEMM is equivalent to a PMC $q(x_{1:N}, y_{1:N})$ given with transitions $q(x_{n+1}, y_{n+1} | x_n, y_n) = q(y_{n+1} | x_n, y_n) q(x_{n+1} | x_n, y_n, y_{n+1})$, where $q(y_{n+1} | x_n, y_n)$ are arbitrary. Taking them constant as in example 1, we have

$$\begin{aligned} q(x_n | x_{n-1}, y_{1:n}) &= q(x_n | x_{n-1}, y_{n-1}, y_n) = \\ &= \left[\sum_{y_{n+1:N}} \frac{q(y_{n+1:N} | x_n, y_{1:n})}{p(x_n | x_{n-1}, y_{1:N})} \right]^{-1} = \left[\sum_{y_{n+1:N}} \frac{q(y_{n+1:N} | x_n, y_{1:n})}{p(x_n | x_{n-1}, y_n)} \right]^{-1} = \\ &= \left[\frac{1}{p(x_n | x_{n-1}, y_n)} \right]^{-1} = p(x_n | x_{n-1}, y_n), \end{aligned}$$

and so MEMM is equivalent to PMC given with transitions $q(x_{n+1}, y_{n+1} | x_n, y_n) = c p(x_{n+1} | x_n, y_{n+1})$.

Remark 3.1 The advantage of PMM over the CRF form can be seen in the filtering problem. Let $p(x_{1:n} | y_{1:n})$ be a CRF, and $q(x_{1:n}, y_{1:n})$ an equivalent PMM. The problem is to estimate X_n from $Y_{1:n}$, which requires the calculation $p(x_n | y_{1:n})$. In filtering, it is interesting to calculate $p(x_1 | y_1), p(x_2 | y_{1:2}), \dots, p(x_n | y_{1:n}), p(x_{n+1} | y_{1:n+1}), \dots$ as quickly as possible. Let us assume that $p(x_n | y_{1:n}) = q(x_n | y_{1:n})$ are given, each with its own parametrization. When using the equivalent CRF form, there is no direct link between $p(x_n | y_{1:n})$ and $p(x_{n+1} | y_{1:n+1})$, so $p(x_n | y_{1:n})$ cannot be used directly and $p(x_{n+1} | y_{1:n+1})$ is calculated as the marginal distribution of the Markov chain $p(x_1 | y_{1:n+1}), p(x_2 | y_{1:n+1}), \dots, p(x_{n+1} | y_{1:n+1})$. This requires n sums over Λ , so its complexity is $n(\#\Lambda)$, which can be a problem when n increases. Using the PMM form, we have

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

so that $p(x_{n+1}, y_{1:n+1})$ can be computed from $p(x_n, y_{1:n})$ with complexity $\#\Lambda$, and hence $p(x_{n+1} | y_{1:n+1})$ is computable from $p(x_n | y_{1:n})$ with complexity $2(\#\Lambda)$, independent from n .

Remark 3.2 Consider a HCRF $(X_{1:N}, U_{1:N}, Y_{1:N})$ such that the CRF $(V_{1:N}, Y_{1:N})$ - where $V_{1:N} = (X_{1:N}, U_{1:N})$ - verifies (3.1). It is then equivalent to PMCs $(V_{1:N}, Y_{1:N}) = (X_{1:N}, U_{1:N}, Y_{1:N})$, models known as “triplet Markov chains” (TMCs) [40]. Here the situation is the opposite to those mentioned in examples 2.1 and 2.2: authors have applied TMCs in different situations, while there was no obvious

reason for CRF specialists to be interested in their equivalent CRFs verifying (3.1). More precisely, TMCs have found applications in the modelling of non-stationary time series of normalized vegetation indices [6], the segmentation of 3D magnetic resonance brain imaging [8], the modelling of consumer loan repayment behavior [10], the activity classification [28], or non-stationary sequences or images segmentation [7], [14], or [16] where the third process is continuous. Let us also cite recent TMCs using variational Bayesian inference [34], [35]. As TMCs are particular TPMMs, they are equivalent to a sub-family of CRFs. This means that in situations where the sequential nature of TMCs is not required, the results obtained with a TMC are likely to be improved with a more general CRF.

4. CONCLUSIONS AND PERSPECTIVES

We studied the relationships between CRFs and different HMM extensions. The main contribution is the proof of equivalence, within the meaning of definition 1.1, between CRFs [21] and PMMs [39]. More precisely, for each CRF (1.2), we have given all PMMs (1.1) whose posterior distribution is the given CRF.

In summary, we have the following four equivalences:

- (i) Simple CRFs (1.10) are equivalent to HMMs (1.8);
- (ii) CRFs satisfying (3.1) are equivalent to pairwise Markov chains (1.7);
- (iii) General CRFs (1.2) are equivalent to PMMs (1.1);
- (iv) Hidden CRFs (HCRFs [41], [42]) are equivalent to Triplet PMMs (TPMMs [39]).

Point (i) was recently established in [3] (see also [18], where a similar result is established under stronger assumptions), points (ii) to (iv) are contributions of the article.

The potential application interest of these results is that the HMM, PMC, PMM and TPMM are sequential – within the meaning of remark 3.1 –, while the CRFs are not. This means that in situations where sequential processing is attractive, it is possible to use HMM, PMC, PMM or TPMM instead of equivalent CRFs without losing modeling power. In cases where the sequential nature of treatments is not essential, no family is *a priori* more interesting than the other; indeed, both families have identical modeling power. However, because their parameters are different, the parameter estimates are different and, therefore, the final yields may be different. In such situations, it would be interesting to test both models, because even if the effectiveness of a model from one family is satisfactory, it can be improved by a model from the other family.

We also noticed that likely in HMMs studied in [3], in PMMs Bayesian generative classifiers can be calculated in a discriminative manner.

Funding: this research was partly funded by the French Government Agency Association Nationale Recherche technologie (ANRT).

5. REFERENCES

- [1] M. Ameer, N. Idrissi, and C. Daoui, “Color images segmentation using pairwise Markov chain”, 6th International Conference on Multimedia Computing and Systems (ICMCS), Rabat, Morocco, May 10-12, 2018.
- [2] A.K. Atiampo and G.L. Loum, “Unsupervised image segmentation with pairwise Markov chains based on nonparametric estimation of copula using orthogonal polynomials”, *International Journal of Image and Graphics*, vol. 16, no. 4, 2016.
- [3] E. Azeraf, E. Monfrini, and W. Pieczynski, “Equivalence between LC-CRF and HMM, and discriminative computing of HMM-based MPM and MAP”, *Algorithms*, 16, 173, 2023.
- [4] I. Batatia, “A deep learning method with CRF for instance segmentation of metal-organic frameworks in scanning electron microscopy images”, European Signal Processing Conference, pp. 625, 2021.
- [5] L.E. Baum and T. Petrie, “Statistical inference for probabilistic functions of finite state Markov chains”, *Annals of Mathematical Statistics*, 37, pp. 1554–1563, 1966.
- [6] A. Ben Abbes, I.R. Farah, and V. Barra, “A non-stationary NDVI time series modelling using triplet Markov chain”, *International Journal of Information and Decision Sciences*, vol. 11, no. 2, pp. 163–179, 2019.
- [7] M.Y. Boudaren, E. Monfrini, W. Pieczynski, and A. Aissani, “Phasic triplet Markov chains”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2310–2316, November 2014.
- [8] S. Brick, C. Collet, and J.-P. Armspach, “Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas”, 3rd IEEE ISBI, Arlington, United States, April 06-09, 2006.
- [9] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models; Springer Series in Statistics*, Springer: Berlin/Heidelberg, Germany, 2005.
- [10] S. Chen, and X. Jiang, “Modeling repayment behavior of consumer loan in portfolio across business cycle: A triplet Markov model approach”, *Complexity*, 2020.
- [11] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 40, no 4, pp. 834–848, 2018.
- [12] J. H. Cristancho Cuervo, J.F. Delgado Saa, and L.A. Ripoll Solano, “Analysis of instantaneous brain interactions contribution to a motor imagery classification task”, *Frontiers in Computational Neuroscience*, 16, 2022.
- [13] P. Dymarski, *Hidden Markov models*, IntechOpen, 2011.
- [14] C. Fernandes and W. Pieczynski, “Non-stationary data segmentation with hidden evidential semi-Markov chains”,

- International Journal of Approximate Reasoning*, 109025, November 2023.
- [15] S. Gahbiche-Braham, H. Bonneau-Maynard, T. Lavergne, and F. Yvon, "Joint segmentation and POS tagging for Arabic using a CRF-based classifier", 8th International Conference on Language Resources and Evaluation, pp. 2107-2113, 2012.
 - [16] H. Gangloff, K. Morales, and Y. Petetin, "Deep parameterizations of pairwise and triplet Markov models for unsupervised classification of sequential data", *Computational Statistics and Data Analysis*, 180, 2023.
 - [17] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt, "Hidden conditional random fields for phone classification", 9th European Conference on Speech Communication and Technology, 1117-1120, 2005.
 - [18] G. Heigold, H. Ney, P. Lehnen, T. Gass, and R. Schluter, "Equivalence of Generative and Log-Linear Models", *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1138-1148, 2010.
 - [19] R. Klinger, *Conditional Random Fields for Named Entity Recognition*. Shaker, 2011.
 - [20] C. Kuljus and J. Lember, "Pairwise Markov models and hybrid segmentation approach", *Methodology and Computing in Applied Probability*, vol. 25, no. 37, 2023.
 - [21] J. Lafferty, A. McCallum, and F.C. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data", *Proceedings of the International Conference on Machine Learning*, 2001.
 - [22] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of triplet Markov chains hidden with long-memory noise", *Signal Processing*, vol. 88, no. 5, pp. 1134-1151, 2008.
 - [23] J. Lapuyade-Lahorgue, and S. Ruan, "Segmentation of multicorrelated images with copula models and conditionally random fields", *Journal of Medical Imaging*, vol. 9, no. 1, 2022.
 - [24] S. Le Cam, F. Salzenstein, and C. Collet, "Fuzzy pairwise Markov chain to segment correlated noisy data", *Signal Processing*, vol. 88, no. 10, pp. 2526-2541, 2008.
 - [25] J. Lember and J. Sova, "Exponential forgetting of smoothing distributions for pairwise Markov models", *Electronic Journal of Probability*, 2021.
 - [26] J. Lember and J. Sova, "Regenerativity of Viterbi process for pairwise Markov models", *Journal of Theoretical Probability*, vol. 34, no. 1, pp. 1-33, 2021.
 - [27] J. Lember, "Local Viterbi property in decoding", *Information and Inference*, vol. 12, no. 3, pp. 1-37, 2023.
 - [28] H. Li, S. Derrode, and W. Pieczynski, "An adaptive and on-line IMU-based locomotion activity classification method using a triplet Markov model", *Neurocomputing*, 362, pp. 94-105, 2019.
 - [29] D.Y. Liliana, and C. Basaruddin, "A Review on conditional random Fields as a sequential classifier in machine learning", *International Conference on Electrical Engineering and Computer Science*, pp.143-148, 2017.
 - [30] J. Liu, C. Wang, W. Wang, and Z. Li, "Particle probability hypothesis density filter based on pairwise Markov chains", *Algorithms*, vol. 12, no. 2, 2019.
 - [31] R. Mahler, "The pairwise-Markov Bernoulli filter", *IEEE Access*, 8, 168229-168245, 2020.
 - [32] A. McCallum and W. Li, "Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons, 7th Conference on Natural Language Learning", pp. 188-191, 2003.
 - [33] A. McCallum, D. Freitag, and F. Pereira, "Maximum Entropy Markov Models for information extraction and segmentation", *Seventeenth International Conference on Machine Learning*, 17, pp. 591-598, 2000.
 - [34] K. Morales, and Y. Petetin, "Variational bayesian inference for pairwise Markov models", *IEEE Workshop on Statistical Signal Processing Proceedings*, pp. 251-255, 2021.
 - [35] K. Morales, and Y. Petetin, "A probabilistic semi-supervised approach with triplet Markov chains", *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2023.
 - [36] L.-S. Ong, X. Zhang, B. Kundukad, J. Dauwels, P. Doyle, and H. H. Asada, "Detecting cell division of *Pseudomonas aeruginosa* bacteria from bright-field microscopy images with hidden conditional random fields", *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, p. 3985, 2016.
 - [37] Y. Petetin and F. Desbouvries, "Bayesian multi-object filtering for pairwise Markov chains", *IEEE Trans. on Signal Processing*, vol. 61, no. 18, pp. 4481-4490, 2013.
 - [38] W. Pieczynski, "Pairwise Markov chains", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 634-639, 2003.
 - [39] W. Pieczynski, "Triplet partially Markov chains and tree", 2nd International Symposium on Image/Video Communications over fixed and mobile networks, Brest, France, July, 7-9, 2004.
 - [40] W. Pieczynski, C. Hulard, and T.Veit, "Triplet Markov chains in hidden signal restoration", *Proceedings of the International Society for Optical Engineering*, 4885 pp. 58-68, 2002.
 - [41] A. Quattoni, M. Collins, and T. Darrell, "Conditional random fields for object recognition", *Neural Information Processing Systems (NIPS 2004)*, pp. 1097-1104, 2004.
 - [42] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1848-1853, 2007.
 - [43] L. Rabiner and B. Juang, "An introduction to hidden Markov models", *IEEE ASSP Magazine*, 3, pp. 4-16, 1986.
 - [44] M.H. Siddiqi, R. Ali, A.M. Khan, Y. Park, and S. Lee, "Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields", *IEEE Trans. on Image Processing*, vol. 24, no. 4, pp. 1386-1398, 2015.
 - [45] Y. Sung, and D. Jurafsky, "Hidden conditional random fields for phone recognition", 2009 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2009, pp. 107-112, 2009.
 - [46] C. Sutton, A. McCallum, and K. Rohanimanesh, "Dynamic conditional random fields: factorized probabilistic models for labeling and segmenting sequence data", *Journal of Machine Learning Research*, vol. 8, pp. 693-723, 2007.
 - [47] C. Sutton, A. McCallum, *An Introduction to Conditional Random Fields*, Now Publishers, 2012.
 - [48] B.W. Sy, A. Quattoni, L.-P. Morency, D. Demirdjian, and T. Darrell, "Hidden conditional random fields for gesture recognition", *IEEE*

- Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1521, 2006.
- [49] P. Tang, and T.W.S. Chow, "Wireless sensor-networks conditions monitoring and fault diagnosis using neighborhood hidden conditional random field", *IEEE Trans. on Industrial Informatics*, vol. 12, no. 3, pp. 933-940, 2016.
 - [50] R. Trullo, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen, "Segmentation of organs at risk in thoracic CT images using a SharpMask architecture and conditional random fields", *International Symposium on Biomedical Imaging*, pp. 10, 2017.
 - [51] M. Vrigkas, E. Mastora, C. Nikou, and I.A. Kakadiaris, "Robust incremental hidden conditional random fields for human action recognition", *Lecture Notes in Computer Science*, vol. 11241, pp. 126-136, 2018.
 - [52] Y. Wang and G. Mori, "Hidden part models for human action recognition: probabilistic versus max margin", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1310-1323, 2011.



Elie Azeraf obtained a master's degree in statistics from Télécom SudParis, Evry, France, and ENSAE Paris, Palaiseau, France, in 2018. He then prepared a PhD from Télécom SudParis, Institut Polytechnique de Paris, in collaboration with IBM France's Watson department, which he defended in 2022. He is currently Senior Data Scientist at Capgemini Engineering, in Aix-en-Provence, France. His research interests include machine learning, hidden Markov chains, and Neural networks applied to NLP.



Wojciech Pieczynski received the Doctorat d'Etat in mathematical statistics from the University Paris VI, France, in 1986. A professor at Telecom Sudparis since 1990, he has been head of the department CITI since 2004. He is the author or co-author of more than eighty journal papers, and he has been supervisor of twenty-five PhD students. His current research focuses on Markov models and Dempster-Shafer theory of evidence, with applications to tracking and forecasting.