Forward-Backward-Viterbi procedures in the Transferable Belief Model for state sequence analysis using belief functions

Emmanuel Ramasso, Michèle Rombaut, Denis Pellerin

HAL Id: hal-00348577

https://hal.archives-ouvertes.fr/hal-00348577

Submitted on 19 Dec 2008

Forward-Backward-Viterbi procedures in the Transferable Belief Model for state sequence analysis using belief functions

Emmanuel Ramasso, Michèle Rombaut, and Denis Pellerin

GIPSA-lab, Images and Signal Department,
46 av. Félix Viallet, 38031 Grenoble, France.
Emails: `e_ramasso @ yahoo.fr`,
{`rombaut,pellerin`} `@ gipsa-lab.inpg.fr`

Abstract. The Transferable Belief Model (TBM) relies on belief functions and enables one to represent and combine a variety of knowledge from certain up to ignorance as well as conflict inherent to imperfect data. A lot of applications have used this flexible framework however, in the context of temporal data analysis of belief functions, a few work have been proposed. Temporal aspect of data is essential for many applications such as surveillance (monitoring) and Human-Computer Interfaces. We propose algorithms based on the mecanisms of Hidden Markov Models usually used for state sequence analysis in probability theory. The proposed algorithms are the "credal forward", "credal backward" and "credal Viterbi" procedures which allow to filter temporal belief functions and to assess state sequences in the TBM framework. Illustration of performance is provided on a human motion analysis problem.

1 Introduction

Analysis of state sequence is important in many fields such as Signal Processing and Computer Vision [1–3]. State sequence analysis is generally performed in Bayesian framework using Hidden Markov models (HMM) [1, 2] where probabilities are used to handle uncertainty on states. In HMM, one can only observe some features related to states but not the states directly in part because of noise. Given a sequence of noise observations, HMM machinery is able to retrieve the best sequence of states using a Viterbi decoding [1] relying on a forward propagation scheme. The latter is used for state filtering (online) whereas smoothing (offline) is performed using a backward procedure. Particular combinations of forward–backward procedures allow to estimate HMM parameters such as the state transition matrix [1]. Usual HMM can only be applied on probabilities.

Transferable Belief Model (TBM) [4] can model more complex cases of uncertainty than probabilities. It is a general model used to represent and combine a variety of knowledge. In particular, doubt and conflict are explicitly emphasized. Doubt smartly represents ignorance (useful to initialize HMM and to represent state transition) and conflict emphasizes the contradiction within a fusion process (can be exploited for state sequence analysis [5]).

The TBM has seldom been used for temporal data and state sequence analysis in a noise context. In this paper a method is proposed for this purpose and that we call credal HMM (CrHMM). The CrHMM combines the genericity of TBM and mecanisms of HMM. The idea to generalize HMM to evidence theory was initiated by Pieczynski et al. (see [6] for a recent work) but the generalization is based on Dempster's rule of combination (with normalization) and assumes that either the prior or (exclusively) the observation is evidential (and generally a *simple* BBA obtained by discounting). Therefore, the combination yields a Bayesian belief function. The formalism proposed in this paper handles general belief functions (as understood in TBM), is strongly based on conflict and relies on Smets' work concerning Evidential Network [7–9]. Moreover, a first *credal Viterbi* algorithm is proposed.

Credal forward-backward algorithms are expressed by commonalities in order to mimick their counterpart in HMM. Then, particular combinations of credal forward-backward algorithms are presented for CrHMM learning. Lastly, a credal Viterbi decoding algorithm is proposed to retrieve the best sequence of states when knowledge on observations and priors is modelled by belief functions. Both the credal forward and the credal Viterbi decoding generate one criterion (based on conflict information) used for inference in the context of competing CrHMMs. Illustrations of algorithms capabilities concern human motion analysis in videos.

Section 2 presents HMM machinery. Section 3 describes some TBM's tools used in this paper. Section 4 is devoted to the credal version of the forward algorithm. Section 5 focuses on the extension of the backward algorithm and on some variables useful for learning credal HMM. Section 6 presents the credal Viterbi decoding. Lastly, illustrations are provided.

2 Hidden Markov models: basics

In this section, main elements of HMM are recalled. For the remainder of this paper, we assume the reader is familiar with basics in HMM. The reader can read the well known tutorial of Rabiner [1] for details. We will refer to it several times in this paper. All conditions of independence are assumed to be satisfied in both the probabilistic [1] and credal cases [7, 10]. Fig. 1 depicts forward and backward processes explained hereafter. This figure will be widely used for the generalization in TBM framework.

Fig. 1. Computation (from [1]) of forward (α_t) and backward (β_t) variables using past and observations likelihood (b_{s_i}). Circles for states and arrows for propagation direction.
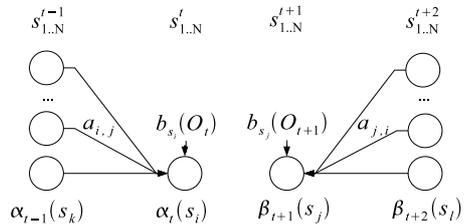
An HMM is a stochastic state machine for the recognition of state sequence from observations. A sequence is supposed to be composed of N states s_i^t at time

$t \in \{1 \dots T\}$ and at any time $s_i^t \in \Omega_t$ where Ω_t is called frame of discernment (FoD) defined by $\Omega_t = \{s_1^t, s_2^t, \dots s_N^t\}$. The time t will be explicitly denoted as superscript of states when required, for instance s_i^t is the state s_i at t. At a given time t, states are said *hidden* and only observations denoted $\mathbf{O}_t \in \Re^F$ are effectively measured, one at each time t. Typically, observations $\mathbf{O}_t$ represent a set of F features values. Models of observations are used to infer the likelihood l of states given observations, i.e. [1] $b_{s_i}(\mathbf{O}_t) = l(s_i^t|\mathbf{O}_t) = P(\mathbf{O}_t|s_i^t)$. These likelihoods are generally provided by a mixture of Gaussians (MoG) for each state [1]. Then, inference on Ω_t is performed by Bayes' theorem providing the posterior probability of state s_i given observations and priors. Sequence analysis by an HMM requires a *transition matrix* $A = [a_{ij}] = P(s_j^t|s_i^{t-1})$ (Markovian assumption), a *prior* π (π_i for state s_i) and *observation models* (MoG). These three elements represent a particular HMM λ. In case distinct sequences are to be recognized, one particular HMM λ is necessary for each sequence. An HMM is used for several tasks [2]: online filtering (forward variable α), offline smoothing (γ-variable obtained from backward β and α), learning (ξ-variable obtained from β, α and γ) and evaluation (Viterbi decoding δ) of states sequences.

3 TBM background

In this section are recalled some basics of TBM used to derive the CrHMM.

3.1 Basic concepts

An agent's belief is expressed on a finite FoD Ω_t and is represented by a basic belief assignment (BBA) m^{Ω_t} from 2^{Ω_t} to $[0, 1]$ with respect to $\sum_{S_i \subseteq \Omega_t} m^{\Omega_t}(S_i) = 1$. For the sake of simplicity, braces around sets will be forgotten, $\{s_1^t, s_3^t\} = \{s_1^t\} \cup \{s_3^t\} \equiv s_1^t \cup s_3^t$. Belief functions are non-additive measures and this allows to explicitly model doubt between states. This is a fundamental difference with probability theory. In the sequel, a singleton is denoted with small cap (e.g. $s_i^t \in \Omega_t$) whereas big cap is used for union of singletons (e.g. $S_i^t \subseteq \Omega_t$).

A BBA m^{Ω_t} can be converted into plausibility pl^{Ω_t} and commonality q^{Ω_t} functions. These functions represent belief in different ways, emphasize some particularities and allow to decrease some computational aspects. We denote $f^{\Omega_t} \in \{q^{\Omega_t}, pl^{\Omega_t}, m^{\Omega_t}\}$ these functions which are in one-to-one correspondance. The relations that we will use are defined $\forall S_i, S_k \subseteq \Omega_t$ by:

$$pl^{\Omega_t}(S_i) = \sum_{S_i \cap S_k \neq \emptyset} m^{\Omega_t}(S_k) \quad (1) \quad m^{\Omega_t}(S_i) = \sum_{S_k \supseteq S_i} (-1)^{|S_k|-|S_i|} q^{\Omega_t}(S_k) (2)$$

$$m^{\Omega_t}(S_i) = \sum_{S_k \supseteq S_i} (-1)^{|S_k|-|S_i|} q^{\Omega_t}(S_k) (3) \quad q^{\Omega_t}(S_i) = \sum_{S_k \subseteq S_i} (-1)^{|S_k|+1} pl^{\Omega_{t+1}}(S_k) (4)$$

3.2 Combination, extension and marginalization

Given two distinct [10] BBAs, $m_1^{\Omega_t}$ and $m_2^{\Omega_t}$ defined on the same FoD Ω_t, the *conjunctive rule of combination* (CRC) is defined $\forall S_i, S_k, S_l \subseteq \Omega_t$ by:

$$(m_1^{\Omega_t} \, \textcircled{\scriptsize $\cap$} \, m_2^{\Omega_t})(S_l) = \sum_{S_i \cap S_k = S_l} m_1^{\Omega_t}(S_i) \cdot m_2^{\Omega_t}(S_k) \tag{5}$$

and equivalently $(q_1^{\Omega_t} \, \textcircled{\scriptsize $\cap$} \, q_2^{\Omega_t})(S_l) = q_1^{\Omega_t}(S_l) \cdot q_2^{\Omega_t}(S_l)$ (commonalities simplifies computation). The *disjunctive rule of combination* (DRC) is [11, 7]:

$$(m_1^{\Omega_t} \, \textcircled{\scriptsize $\cup$} \, m_2^{\Omega_t})(S_l) = \sum_{S_i \cup S_k = S_l} m_1^{\Omega_t}(S_i) \cdot m_2^{\Omega_t}(S_k) \tag{6}$$

In case FoDs are different, e.g. one wants to combine m^{Ω_t} and $m^{\Omega_{t+1}}$, then BBAs must be *extended* on the common FoD $\Omega_t \times \Omega_{t+1}$ before combination using the so-called vacuous extension [7] denoted "$\uparrow$". E.g. m^{Ω_t} is redefined as:

$$m^{\Omega_t \uparrow \Omega_t \times \Omega_{t+1}}(S) = \begin{cases} m^{\Omega_t}(S_i) \text{ if } S = S_i \times \Omega_{t+1} \\ \qquad\qquad \text{and } S_i \subseteq \Omega_t \\ 0 \qquad\quad \text{otherwise} \end{cases} \tag{7}$$

After combination by the CRC or by the DRC, the result $m^{\Omega_t \times \Omega_{t+1}}$ can be marginalized onto Ω_{t+1} or Ω_t. Assuming the marginal is computed on Ω_{t+1}, the marginalization operator, denoted "$\downarrow$", is defined $\forall S_j \subseteq \Omega_{t+1}$ by:

$$m^{\Omega_t \times \Omega_{t+1} \downarrow \Omega_{t+1}}(S_j) = \sum_{\substack{S \subseteq \Omega_t \times \Omega_{t+1} \\ S^{\downarrow \Omega_{t+1}} = S_j}} m^{\Omega_t \times \Omega_{t+1}}(S) \tag{8}$$

with "$S^{\downarrow \Omega_{t+1}} = S_j$" means S_j is the projection of S on Ω_{t+1}.

3.3 Conditioning and ballooning extension

In a state sequence, conditional beliefs describes how true can be the states in $S_j \subseteq \Omega_{t+1}$ given the previous states in $S_i \subseteq \Omega_t$. Given two BBAs: $m_i^{\Omega_t}$ defined by $m_i^{\Omega_t}(S_i) = 1$ and $m^{\Omega_t \times \Omega_{t+1}}$, the conditional belief $m^{\Omega_{t+1}}[S_i^t]$ is defined by:

$$m^{\Omega_{t+1}}[S_i^t] = \left(m_i^{\Omega_t \uparrow \Omega_t \times \Omega_{t+1}} \, \textcircled{\scriptsize $\cap$} \, m^{\Omega_t \times \Omega_{t+1}} \right)^{\downarrow \Omega_{t+1}} \tag{9}$$

Conversely, if a conditional BBA $m^{\Omega_{t+1}}[s_i^t]$ is provided $\forall s_i \in \Omega_t$, it is possible to cancel the conditioning revision using the *ballooning extension* [7]. Let us denote $v_j = ((s_i \times \Omega_{t+1}) \cap S)^{\downarrow \Omega_{t+1}})$ (where $s_i \times \Omega_{t+1}$ is called cylindrical extension of state (singleton) $s_i \in \Omega_t$ on $\Omega_t \times \Omega_{t+1}$), then Smets proves [7]:

$$q^{\Omega_t \times \Omega_{t+1}}(S) = \prod_{s_i \in \Omega_t} q^{\Omega_{t+1}}[s_i](v_j) \tag{10}$$

Ballooning extension and vacuous extension [7] are used in this paper in order to compute a belief on a joint space $\Omega_t \times \Omega_{t+1}$.

3.4 The Generalized Bayesian Theorem

Bayes' theorem was extended in TBM framework by Smets [7] and is called the Generalized Bayesian Theorem (GBT). The GBT alleviates the problem of priors since belief functions allow to represent total ignorance.

Definition - Given a vacuous prior $(m^{\Omega_t}(\Omega) = 1)$ and the set of conditional beliefs $f^{\Omega_{t+1}}[s_i]$, the posterior $f^{\Omega_{t+1}}(S_j)$, $\forall S_j \in \Omega_{t+1}$ is:

$$f^{\Omega_{t+1}}(S_j) = \sum_{S_i \subseteq \Omega_t} f^{\Omega_{t+1}}[S_i](S_j) \cdot m^{\Omega_t}(S_i) \tag{11}$$

Since the conditional belief is initially given conditionally to each singleton state $s_i \in \Omega_t$, the belief defined conditionally to a subset S_i is obtained by the DRC (Eq. (6)) assuming distinctness [7]:

$$f^{\Omega_{t+1}}[S_i] = \bigcup_{s_i \in S_i} f^{\Omega_{t+1}}[s_i], \;\; \forall S_i \subseteq \Omega_t \tag{12}$$

Eq. (11)-(12) are the core of Evidential Networks [7, 8, 10].

Motivation: from likelihood to belief function - Assume a set of features $\mathbf{O}_t$ taking values in $\Re^F$. As emphasized in [12], often the conditional belief over $\Re^F$ given s_i is represented by a probability function. In this case $pl^{\Re^F}[s_i](\mathbf{O}_t) = P(\mathbf{O}_t|s_i) = l(s_i|\mathbf{O}_t)$, so the vector of plausibilities equals the vector of likelihoods of s_i given $\mathbf{O}_t$ [7, 12]. Given the likelihoods $l(s_i|\mathbf{O}_t)$ for each $s_i \in \Omega_t$, then for $\mathbf{O}_t \in \Re^F$ and for each $S \subseteq \Omega_t$, Smets [7] proves:

$$q_b^{\Omega_t}[\mathbf{O}_t](S) = \prod_{s_i \in S} l(s_i|\mathbf{O}_t) \tag{13}$$

where $q_b^{\Omega_t}[\mathbf{O}_t]$ is the posterior commonality conditionally to $\mathbf{O}_t$ and defined on Ω_t. It is the counterpart of the probability $b_{s_i}(\mathbf{O}_t)$ [1] but now the commonality is defined for union of states.

3.5 Decision-making

After combination of multiple sources of belief, a resulting BBA m^{Ω_t} is obtained. Decision-making under uncertainty and imprecision based on belief functions must be made either on the pignistic probability [7] or on plausibilities [12], both assign a value to each element (singleton) $s_i \in \Omega_t$. In the sequel, the pignistic probability is used and it is defined by:

$$\mathbf{BetP}\{m^{\Omega_t}\}(s_i) = \sum_{S_k \subseteq \Omega_t} \frac{|s_i \cap S_k|}{|S_k|} \frac{m^{\Omega_t}(S_k)}{1 - m^{\Omega_t}(\emptyset)} \tag{14}$$

where $m^{\Omega_t}(\emptyset)$ is the conflict value and $1 - m^{\Omega_t}(\emptyset)$ is a normalizing coefficient.

4 Filtering belief functions: credal forward algorithm

In HMM, the forward algorithm allows to filter (online) probabilities that evolves along time. The forward algorithm relies on the forward variable generally denoted $\alpha(t)$ [1].

A similar *credal forward algorithm* can be obtained. Commonalities are used in order to mimic their counterpart based on probabilities. The algorithm is a simple propagation scheme that follows left part of Fig. 1. The credal forward algorithm consists of three steps:

1. **Initialization :** $\forall S_i \subseteq \Omega_1$, apply Eq. (15).
2. **Induction:** $1 \le t \le T-1$, $S_j \subseteq \Omega_{t+1}$, apply Eq. (16).
3. **Termination:** apply Eq. (17).

We denote $f_\alpha^{\Omega_t} \in \{q_\alpha^{\Omega_t}, pl_\alpha^{\Omega_t}, m_\alpha^{\Omega_t}\}$ the credal version of the forward variable. Subscripts α, a and b are used to mimick, respectively, the forward variable, transition probabilities and observations likelihood as defined for usual HMM [1].

4.1 Initialization

One asset of using belief functions is the possibility to explicitly model a vacuous prior for $t = 1$ (no prior):

$$q_\alpha^{\Omega_{t=1}}(S_i) = 1, \quad \forall S_i \subseteq \Omega_t \tag{15}$$

or equivalently $m_\alpha^{\Omega_{t=1}}(\Omega) = 1$. This lack of information is well handled and smartly represented using belief functions whereas probabilities require priors. One can also use more complex initialization such as consonant belief functions.

4.2 Induction

Induction can be retrieved from Smets' work on Evidential Network [7, 8]. It relies on the credal forward variable. Given:

1. $m_\alpha^{\Omega_t}$ the BBA of the forward variable of states at t,
2. $q_a^{\Omega_{t+1}}[S_i]$, $\forall S_i \subseteq \Omega$ the set of conditional commonality distribution which links states and set of states from t to $t+1$ (obtained from Eq. (11) and (12)),
3. $q_b^{\Omega_{t+1}}[\mathbf{O}_t]$ the observations, obtained e.g. from a mixture of Gaussians and GBT (Eq. (13)) at $t+1$.

The credal forward variable is a commonality that combines past information, transition and current observations (left part of Fig. 1) by:

$$q_\alpha^{\Omega_{t+1}}(S_j) = \Big(\sum_{S_i \subseteq \Omega_t} m_\alpha^{\Omega_t}(S_i) \cdot q_a^{\Omega_{t+1}}[S_i](S_j) \Big) \cdot q_b^{\Omega_{t+1}}[\mathbf{O}_t](S_j) \tag{16}$$

defined $\forall S_j \subseteq \Omega_{t+1}$. This equation has a close form compared to the probability-based forward pass but this one works on belief functions and on sets of states. The first part is the application of the GBT with priors ($m_\alpha^{\Omega_t}(S_i)$) and conditional commonalities ($q_a^{\Omega_{t+1}}[S_i](S_j)$). The second part represents observations ($q_b^{\Omega_{t+1}}[\mathbf{O}_t](S_j)$) conjunctively combined using the CRC.

4.3 Termination step

In HMM, the termination step of the forward algorithm [1] is $\sum_{s_i \in \Omega_T} \alpha_T(s_i) = P(\mathbf{O}_{1:T}|\lambda)$ and represents the state sequence probability. However, we can not apply the same termination step with the credal version because belief on focal

sets of $m_\alpha^{\Omega_t}$ always sum to 1 for any t. Actually, the BBA obtained at T does not reflect the whole sequence but only the belief on states at T. Instead, we propose to exploit conflict for state sequence analysis. The proposed criteria has a similar role to the log-likelihood used for MAP classification. Given several HMM $\lambda_1 \ldots \lambda_k \ldots \lambda_K$, the best model λ^* explaining observations on $[1, T]$ minimizes conflict along the sequence of observations $\mathbf{O}_{1:T}$:

$$L_c^1(\lambda) = \frac{1}{T} \sum_{t=1}^{T} \log(m_\alpha^{\Omega_t}[\lambda](\emptyset)) \ (17) \qquad\qquad \lambda^* = \underset{k}{\operatorname{argmin}} \ L_c^1(\lambda_k) \qquad (18)$$

During induction, we keep track of this conflict and then normalize the BBA. Eq. (17) is similar to the one proposed and justified by Ristic and Smets [13] where it is used as a distance measure between objets for association purpose. Using commonalities, the computation in the credal forward algorithm is close to the probability-based version [1] but the former is greedy (see [14, 7] for computational solutions).

5 Smoothing and towards learning

State sequence can be filtered offline, this is called smoothing [2]. Smoothing is used for learning HMM parameters. Rabiner [1] presents three variables used for smoothing/learning: the backward (β) variable (as in right part of Fig. 1), the γ-variable and the ξ-variable. The first one is generally combined with the forward variable for offline smoothing. The ξ-variable is exploited for learning the transition matrix of HMM. Learning the transition matrix includes an iterated procedure that we will not be developed for the CrHMM. We only propose the equivalent expressions of the three variables.

5.1 The credal backward induction

Likewise to the credal forward variable, the credal backward variable is computed using belief propagation and following the right part of Fig. 1. The credal backward induction is defined $\forall S_i \subseteq \Omega_t$ by:

$$q_\beta^{\Omega_t}(S_i) = \sum_{S_j \subseteq \Omega_{t+1}} \left(\left(m_\beta^{\Omega_{t+1}} \copyright m_b^{\Omega_{t+1}}[\mathbf{O}_t] \right)(S_j) \cdot q_a^{\Omega_t}[S_j](S_i) \right) \qquad (19)$$

where $\left(m_\beta^{\Omega_{t+1}} \copyright m_b^{\Omega_{t+1}}[\mathbf{O}_t] \right)(S_j)$ is the value of the BBA on set S_j resulting from the combination by the CRC of both $m_\beta^{\Omega_{t+1}}$ and $m_b^{\Omega_{t+1}}[\mathbf{O}_t]$. Since only $q_a^{\Omega_{t+1}}[S_i](S_j)$ is known (i.e. the conditional belief of proposition S_j at $t + 1$ given S_i at t), it is required to derive $q_a^{\Omega_{t+1}}[S_j](S_i)$ from it. For that, we use the Generalized Likelihood Principle [7] which postulates that $pl_a^{\Omega_{t+1}}[S_i](S_j) = pl_a^{\Omega_t}[S_j](S_i)$, where $pl_a^{\Omega_{t+1}}[S_i](S_j)$ is the conditional plausibility (Eq. (1)) of being in state S_j at $t + 1$ given S_i at t and which is known. From plausibility $pl_a^{\Omega_t}[S_j]$, commonality $q_a^{\Omega_t}[S_j]$ is derived by Eq. (4). The backward variable is initialized with a vacuous prior.

5.2 The credal γ variable

The joint observation of both the credal forward and backward variables can be obtained by the CRC:

$$q_\gamma^{\Omega_t} = q_\alpha^{\Omega_t} \textcircled{$\cap$} q_\beta^{\Omega_t} \qquad (20)$$

obtained $\forall S_i \subseteq \Omega_t$. The credal variable $q_\gamma^{\Omega_t}$ might be useful for learning HMM parameters in the credal case. Another role of $q_\gamma^{\Omega_t}$ is to assess the best state s_t^* at t of the current sequence by $s_t^* = \text{argmax }_{s_i \in \Omega_t} \mathbf{BetP}\{m_\gamma^{\Omega_t}\}(s_i)$ for $1 \le t \le T$ and where $\mathbf{BetP}$ is defined by Eq. (14).

5.3 The credal ξ variable

Since the probability-based ξ variable is used to learn the transition matrix [1], it would be interesting to define it in TBM framework for further learning. In case learning might be done online or offline, we propose hereafter two versions of the credal ξ variable: one to estimate it on-the-fly and denoted $f_{\xi_{\text{on}}}^{\Omega_{t-1} \times \Omega_t}$ and one for the off-line case denoted $f_{\xi_{\text{off}}}^{\Omega_t \times \Omega_{t+1}}$, both defined on the product space of two successive time slices. The joint space allows to explicitly model the link (transition) between each couple of states likewise to the transition matrix but the credal version explicitly models doubt between states.

On-line estimation In the online case, only the available information up to time t are combined. Thus it is based on the credal forward algorithm, observations (at t), priors (if available, at $t-1$) and conditional beliefs (t given $t-1$). We define it for $S_i \subseteq \Omega_{t-1}$, $S_j \subseteq \Omega_t$ and $S \subseteq \Omega_t \times \Omega_{t-1}$ as:

$$q_{\xi_{\text{on}}}^{\Omega_{t-1} \times \Omega_t}(S) = q_\alpha^{\Omega_{t-1}\uparrow\Omega_{t-1}\times\Omega_t}(S) \times q_a^{\Omega_{t-1}\times\Omega_t}(S) \times q_b^{\Omega_t}[\mathbf{O}_t]^{\uparrow\Omega_{t-1}\times\Omega_t}(S) \quad (21)$$

and where $q_a^{\Omega_{t-1} \times \Omega_t}(S)$ is computed by Eq. (10). Moreover, the marginalization of $q_{\xi_{\text{on}}}^{\Omega_{t-1} \times \Omega_t}$ onto Ω_t results in the forward variable, i.e. $q_\alpha^{\Omega_t} = q_{\xi_{\text{on}}}^{\Omega_{t-1} \times \Omega_t \downarrow \Omega_t}$.

Off-line estimation For the offline case, both backward (up to $t+1$) and forward (up to t) propagations are combined. The link between t and $t+1$ is made by the conditonal belief. We define it $\forall S \subseteq \Omega_t \times \Omega_{t+1}$ by:

$$\begin{aligned} q_{\xi_{\text{off}}}^{\Omega_t \times \Omega_{t+1}}(S) = {} & q_\alpha^{\Omega_t\uparrow\Omega_t\times\Omega_{t+1}}(S) \times q_\beta^{\Omega_{t+1}\uparrow\Omega_t\times\Omega_{t+1}}(S) \\ & \times q_b^{\Omega_{t+1}}[\mathbf{O}_t]^{\uparrow\Omega_t\times\Omega_{t+1}}(S) \times q_a^{\Omega_t\times\Omega_{t+1}}(S) \end{aligned} \qquad (22)$$

There is one commonality $q_{\xi_{\text{off}}}^{\Omega_t \times \Omega_{t+1}}$ at each time. The joint form of the conditional belief is obtained by applying Eq. (10) on $q_a^{\Omega_{t+1}}[S_j](S_i)$ as explained Section 5.1 for the backward variable. Note that marginalizing $q_{\xi_{\text{off}}}^{\Omega_t \times \Omega_{t+1}}$ onto Ω_{t+1} yields the credal γ variable, i.e. $q_\gamma^{\Omega_{t+1}} = q_{\xi_{\text{off}}}^{\Omega_t \times \Omega_{t+1} \downarrow \Omega_{t+1}}$.

6 Credal Viterbi procedure

The goal is to determine which state $s_i \in \Omega_{t-1}$ at $t-1$ (*current candidate*) accounts for the ended-state sequence s_j at time t (*current hypothesis*).

State-of-knowledge between t and $t-1$ is given by Eq. (21). Therefore, likewise to the probabilistic case, the credal Viterbi relies on the forward pass. Since it is required to test each hypothesis $s_j \subseteq \Omega_t$, the BBA obtained from Eq. (21) and (3) is conditioned on the current hypothesis s_j. Moreover, it is necessary to determine which state s_i is the best candidate therefore, the conditioning result is marginalized onto the space Ω_{t-1} where the current candidate s_i is defined. Hence, the BBA used for decision is formally defined by:

$$m_{\text{dec}}^{\Omega_{t-1}}[s_j] = m_{\xi_{\text{on}}}^{\Omega_{t-1} \times \Omega_t}[s_j]^{\downarrow \Omega_{t-1}} \tag{23}$$

Conditioning is equivalent to a conjunctive combination with a categorical belief mass of the form $m_j^{\Omega_{t-1} \times \Omega_t}(s_j) = 1$ (Section 3.3). Therefore, a conflict may appear and quantifies the *incoherence of being in state s_j at t*. This conflict is used in the decision process and it is the core of the credal Viterbi procedure. Let us denote $\mathbf{C}_t(s_j)$ the value of the coherence at time t:

$$\mathbf{C}_t(s_j) = 1 - m_{\text{dec}}^{\Omega_{t-1}}[s_j](\emptyset) \tag{24}$$

Decision must be made on the pignistic probability distribution (Eq. (14)) defined over Ω_{t-1} from Eq. (23). This space, Ω_{t-1}, is characterized by a set of coherences $\mathbf{C}_{t-1}(s_i)$ (s_i is used because time index is $t-1$) computed at the previous time slice from Eq. (24). Therefore, coherences are taken into account in the decision process performed at time t on space Ω_{t-1} by weighting the pignistic probabilities defined over Ω_{t-1}:

$$P_t[s_j](s_i) = \mathbf{C}_{t-1}(s_i) \cdot \mathbf{BetP}\{m_{\text{dec}}^{\Omega_{t-1}}[s_j]\}(s_i) \tag{25}$$

It is required to normalize the coherences in order to obtain a probability distribution. The BBA $m_{\text{dec}}^{\Omega_{t-1}}[s_j]$ is obtained at time t from knowledge at both t and $t-1$ (it is projected onto Ω_{t-1}) whereas coherences concern time index $t-1$ (not t) obtained at the previous time. From the obtained (weighted) probability distribution (Eq. (25)), the decision (choosing the best candidate s_i) is made.

Lastly, we compute another log-contradiction criteria $L_c^2(\lambda)$ using the results of the credal Viterbi procedure (see step 4b)). Indeed, when the path is recovered, we take into account the coherence along the sequence. The basic idea is that the lower the conflict along the path, the better the sequence corresponds to the model λ. We define it in the following algorithm:

1. **Initialization :** $\forall s_i \in \Omega_1$, $\mathbf{C}_1(s_i) = 1$, $L_c^2(\lambda) = 0$ and $\psi^{\Omega_1}(s_i) = 0$
2. **Recursion:** $2 \leq t \leq T-1$, $\forall s_j \in \Omega_t$, $\forall s_i \in \Omega_{t-1}$
   a) Compute $m_{\text{dec}}^{\Omega_{t-1}}[s_j]$ with Eq. (23)
   b) $P_t[s_j](s_i) = \mathbf{C}_{t-1}(s_i) \cdot \mathbf{BetP}\{m_{\text{dec}}^{\Omega_{t-1}}[s_j]\}(s_i)$
   c) $\psi_t(s_j) = \text{argmax}_{s_i \in \Omega_{t-1}} [P_t[s_j](s_i)]$
   d) Compute $\mathbf{C}_t(s_j)$ with Eq. (24)

3. **Termination:** $s_T^* = \text{argmax}\,_{\forall s_j \in \Omega_T}[\max_{\forall s_i \in \Omega_{T-1}} P_T[s_j](s_i)]$
4. **Path backtracking:**
   a) $s_t^* = \psi_{t+1}(s_{t+1}^*)$
   b) $L_c^2(\lambda) \leftarrow L_c^2(\lambda) + \log(1 - \mathbf{C}_t(s_t^*))$

where "$\leftarrow$" is an assignment in order to update the second proposed log-contradiction criteria. The variable $\mathbf{C}_t$ is computed at t for all $s_j \subseteq \Omega_t$ (line d)) and is used for the next step.

Transition and ambiguity detection - The coherence measure is used for states transition detection. When the decision is made (step 3)), one can compel the probability (step 2b)) to be greater than a given threshold σ in order to trust the decision (e.g. $\sigma = 1/N$ with N the number of states).

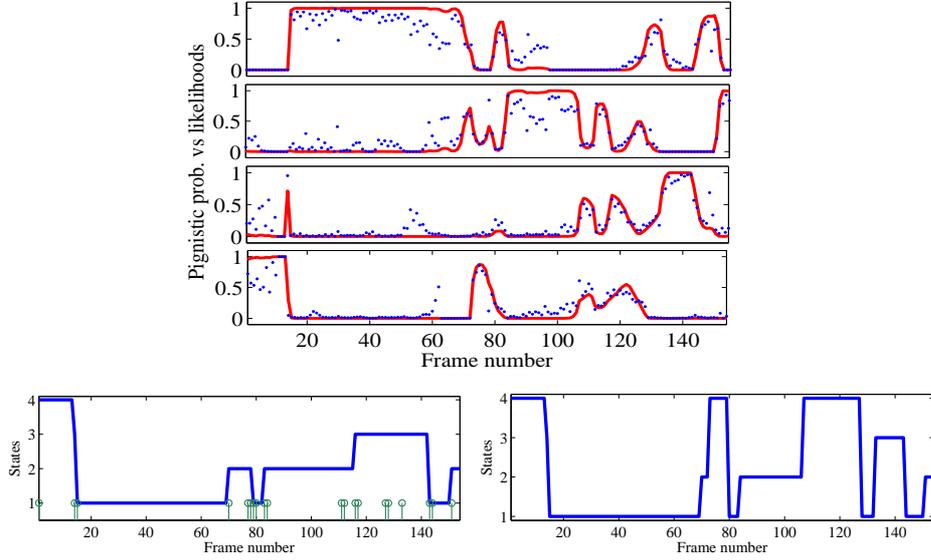## 7 Illustration: human motion analysis

Fig. 2. Results: **(top)** Gaussian mixture outputs (blue dots) and smoothing by credal γ variable (red, bold). **(bottom-left)** Credal Viterbi decoding in CrHMM. **(bottom-right)** Viterbi decoding of HMM. Ground truth: running on $[12, 65]$, jumping on $[66, 103]$, falling on $[104, 154]$ (there is no standing-up action in this sequence).

Our aim is to *illustrate* some of the proposed algorithms : smoothing, evaluation and classification *a posteriori*. For that, we use data of previous communications [15, 16] that concern the recognition of three athletics jumps (activity): long jumps, pole vaults and high jumps. For this paper, each activity is modelled by one HMM and one CrHMM. Each model of activity is made of four

states (actions) with $\Omega_t = \{running, jumping, falling, standing\}$. Given a set of features extracted at each video frame, a likelihood on each action is computed using a mixture of gaussians. Likelihoods at each video frame are transformed into belief on actions using Eq. (13). HMM parameters are learned using Baum-Welch algorithm [1] on the half of the database. Learning CrHMM is still in its infancy: actually we proposed some equations in previous sections that need to be embedded in an iterated procedure such as CrEM [17] in order to compute it. The credal transition matrix is thus computed from the probabilistic one by transforming each conditional probability into conditional consonant BBA [18] that is discounted contextually [19] to spread beliefs on subsets (thus we obtain a "filled" matrix).

Smoothing - We use the credal γ variable (Eq. (20)) that combines both forward-backward algorithms to filter (offline) a noise BBA. Fig. 2-top shows the smoothing results with likelihoods (blue dots) vs. pignistic probabilities (red line) after application of the credal γ variable.

Inference - We apply the credal Viterbi algorithm on the previous set of likelihoods in order to decode the best sequences of states. The results of both the credal (Fig. 2-bottom-left) and probabilistic (Fig. 2-bottom-right) Viterbi decoding are shown. The proposed credal Viterbi scheme is able to detect the transitions (which size varies) as well as ambiguous instants (where the largest pignistic is low). These information are represented by green stems on the figure (with $\sigma = 1/N = 0.25$). The credal Viterbi yields better decoding here.

Classification - In order to assess the recognition criteria integrated in the Viterbi decoding (L_c^2), we propose a classification problem. We use 26 videos (about 3000 frames) of pole vaults analyzed by the three models (long jump (λ_1), high jump (λ_2) and pole vault (λ_3) models). We compare the criteria L_c^2 ($\lambda^* = \text{argmin}_{k=\{1,2,3\}} L_c^2(\lambda_k)$) to the usual log-likelihoods (L_l, $\lambda^* = \text{argmax}_{k=\{1,2,3\}} L_l(\lambda_k)$). We assess two things:

1) The classification rate: we compute it by dividing the number of correctly classified videos by 26. We obtain 81% for the CrHMM and 58% for the usual HMM. The classification rate is better for the CrHMM mainly because of doubt that is used to model knowledge about states. When a wrong decision is made in a HMM, it is propagated. In the credal Viterbi (CrHMM), the decision is not propagated since we keep the belief mass for the next instant.

2) The mean square of the difference between the two best normalized criteria values (by ranking the values for the criteria L_l and L_c^2): it reflects the "discriminative power" during classification (ideally one criteria equals 1 and the two others are null). We obtain 7% for the CrHMM and 1% for the usual HMM. This means that the CrHMM is much more discriminative than the HMM on this dataset therefore, the decision is more reliable.

8 Conclusion

We have presented a general methodology to use HMM machinery in Transferable Belief Model (TBM) framework which is useful when knowledge is rep-

resented by belief functions instead of probabilities. The proposed modelling exploits doubt and conflict emphasized in the TBM. Experiments have illustrated some concepts and demonstrated encouraging results for video-oriented applications. Future work will be focused on learning and comparison with fuzzy HMM.

References

1. L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77:257–285, 1989.
2. K. P. Murphy. *Dynamic Bayesian Networks: Representation, inference and learning*. PhD thesis, UC Berkeley (CSD), 2002.
3. M. Shah. Understanding human behavior from motion imagery. *Machine Vision and Applications*, 14:210–214, 2003.
4. P. Smets and R. Kennes. The Transferable Belief Model. *Artificial Intelligence*, 66(2):191–234, 1994.
5. E. Ramasso, D. Pellerin, and M. Rombaut. Belief Scheduling for the recognition of human action sequence. In *Int. Conf. on Inf. Fusion*, Florence, Italia, 2006.
6. W. Pieczynski. Multisensor triplet Markov chains and theory of evidence. *IJAR*, 45:1–16, 2007.
7. P. Smets. Beliefs functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. *IJAR*, 9:1–35, 1993.
8. H. Xu and P. Smets. Evidential reasoning with conditional belief functions. *Uncertainty in Artificial Intelligence*, 94:598–606, 1994.
9. B. Ben Yaghlane, P. Smets, and K. Mellouli. Directed evidential networks with conditionnal belief functions. In *ECSQARU*, pages 291–305, 2003.
10. B. Yaghlane, P. Smets, and K. Mellouli. Belief Function Independence: II. The Conditional Case. *IJAR*, 31:31–75.
11. D. Dubois and H. Prade. A set-theoretic view of belief functions – Logical operations and approximation by fuzzy sets. *I.J. of Gen. Sys.*, 12(3):193–226, 1986.
12. F. Delmotte and Ph. Smets. Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model. *IEEE TSMC*, 34(4):457–471, 2004.
13. B. Ristic and P. Smets. The TBM global distance measure for the association of uncertain combat ID declarations. *Information Fusion*, 7:276–284, 2006.
14. T. Denoeux and A.B. Yaghlane. Approximating the combination of belief functions using the fast Moebius transform in a coarsened frame. *IJAR*, 37:77–101, 2002.
15. E. Ramasso, C. Panagiotakis, M. Rombaut, and D. Pellerin. Human action recognition in videos based on the Transferable Belief Model - Application to athletics jumps. *Pattern Analysis and Applications*, 2007. Accepted for publication.
16. C. Panagiotakis, E. Ramasso, G. Tziritas, M. Rombaut, and D. Pellerin. Shape-motion based athlete tracking for multilevel action recognition. In 4th Int. Conf. on Articulated Motion and Deformable Object, Spain, pages 385–394, 2006.
17. P. Vannoorenberghe and Ph. Smets. Partially supervised learning by a Credal EM approach. In *ECSQARU*, volume 3571, pages 956–967, 2005.
18. D. Dubois, H. Prade, and P. Smets. New semantics for quantitative possibility theory. In *ECSQARU*, volume 2143, pages 410–421, 2001.
19. D. Mercier, B. Quost, and T. Denoeux. Refined modeling of sensor reliability in the belief function framework using contextual discounting. *Information Fusion*. To appear in 2007.