CONTRIBUTION OF BELIEF FUNCTIONS TO HIDDEN MARKOV MODELS WITH AN APPLICATION TO FAULT DIAGNOSIS

Emmanuel Ramasso

FEMTO-ST Institute, UMR CNRS 6174 - UFC / ENSMM / UTBM,
Automatic Control and Micro-Mechatronic Systems Department, Besançon, France, 25000
emmanuel.ramasso@femto-st.fr

ABSTRACT

Evidence-theoretic propagations of temporal belief functions are proposed to deal with possibly dependent observations and for partially supervised learning of HMM. Solutions are formulated in Transferable Belief Model framework and experiments concern a diagnosis problem.

Keywords: Evidential Hidden Markov Models, State Sequence Recognition, Partially Supervised Learning, System's Health Assessment.

1. INTRODUCTION

Temporal data modelling and analysis are important in many fields [1, 2] and one common solution is to use *state sequence* representation, learning and inference algorithms. A state generally represents a stationarity or a functioning mode in the data and describes the dynamical system at a given time while transitions represent the system dynamics. In practical applications, states are hidden and only features are observable. An additional modelling step is thus required to relate numerical features to states. In order to manage imperfections, the modelling should cope with uncertainty using [3] probability theory, possibility theory or evidence theory, the latter being more general.

Hidden Markov Model (HMM) [1] is a famous probabilistic and non-deterministic finite state machine for state sequence learning and inference under the concept of hidden states managing both discrete and continuous *features*. In HMM, the modelling method allows to generate a conditional probability (or a likelihood), denoted $b_j(\mathbf{O}_t)$ for the j-th state, where $\mathbf{O}_t = [O_1 \ O_2 \ \dots \ O_F]$ is the set of F features. Three problems concerning HMM can be tackled. What if information on states are represented by possibilities or belief functions or in several formalisms? How to express a lack of knowledge without choosing any artificial priors and is it then possible to take it into account in probabilitic HMM learning? What if observations are not independent (conditionally on states)?

In order to tackle these problems, one solution is to marry HMM mechanims with other representations of uncertainty [3]. Mohamad and Gader [4] introduced the *Generalized HMM* in which the generalization is narrowed down to *possibility measures* and thereby their framework is not able to manage belief functions. One advantage of their framework is the possibility to *manage dependent observations by using fuzzy operators* but the authors used the product thereby assuming statistical independence. Pieczynski *et al.* [5, 6] pioneered in mixing belief functions with Markov chains leading to promising results. The main idea was to alleviate the problem of prior modelling in Markov chains [7] using Dempster's rule of combination (generalizing Bayesian inference [8]). However, either the prior or the belief state is evidential but not both, thereby underlying probability assumptions are present. Rombaut *et al.* [9] proposed a generalization of a Petri Net to belief functions based on the Generalized Bayesian Theorem (GBT) [8]. However, it is not robust to noise because links between states at successive instants are given by an evolving and sparse transition matrix depending on sensors measures. Moreover, no classification criterion was proposed. The fourth paper [10] proposed a deterministic state machine called *Belief Scheduler* based on the *Temporal Evidential Filter* [11]. An original inference criterion based on conflict is proposed for classification of sequences. The problem is the sensitivity of some threshold settings.

Compared to previous work, evidential forward-backward-Viterbi propagations are extended to belief functions with underlying probability assumptions. The extension is based on special operators developed in the Transferable Belief Model (TBM) [12] in order to consider possibly dependent observations. The problem of partially supervised learning of probabilistic HMM is also tackled using belief functions.

2. PROBABILISTIC HMM IN A NUTSHELL

States are said hidden: only F observations in $\mathbf{O}_t$ are available at instants $t \in \{1..T\}$ (T is the sequence length). The number of states is N and $s_i^t \in \Omega_t$ is the i-th state at t where $\Omega_t = \{s_1^t, s_2^t, \dots s_N^t\}$ is the frame of discernment. At each t, the likelihood of state i conditional to the current observation is denoted $b_i(\mathbf{O}_t) = P(\mathbf{O}_t|s_i^t)$ and is provided by a modelling technique such as a mixture of Gaussians [1]. The transition from a state s_i^{t-1} to a state s_j^t is made with a probability $a_{ij} = P(s_j^t|s_i^{t-1})$ and the transition matrix is $\mathbb{A} = [a_{ij}], i \in \{1 \dots N\}, j \in \{1 \dots N\}$ with size $N \times N$ and $\sum_j a_{ij} = 1$. Transitions between states start from an initial distribution ($t = 1$) denoted $\pi_i, i \in \{1 \dots N\}$ and probabilities on states are then updated along time according to transitions and observations. An HMM $\lambda = \{\mathbb{A}, \mathbb{B}, \Pi\}$ is thus characterized by three elements: the transition matrix $\mathbb{A}$, the observation models $\mathbb{B}$ (generating likelihoods $b_i(\mathbf{O}_t)$) and the initial distribution Π.

Estimating the likelihood of an observation sequence given the model is computationally feasible using the *forward* propagation that evaluates the *forward* variable $\alpha_t(s_j^t) = P(\mathbf{O}_1, \mathbf{O}_2 \dots \mathbf{O}_t, s_j^t|\lambda) = \sum_{\forall s_i \in \Omega_{t-1}} \alpha_{t-1}(s_i) \cdot a_{ij} \cdot b_j(\mathbf{O}_t)$. For a classification problem, we need to compute the log-likelihood of observations for each HMM model λ and then to choose the best model by: $\lambda^* = \text{argmax}_{\lambda}\, \mathcal{L}_p$ with $\log \mathcal{L}_p = -\frac{1}{T} \sum_t \log \left(\sum_{s_j \in \Omega_t} \alpha_t(s_j) \right)$. The *backward* propagation allows to compute the *backward* variable $\beta_t(s_i) = P(O_T, O_{T-1} \dots\, O_{t+1}|s_i, \lambda) = \sum_{\forall s_j \in \Omega_{t+1}} \cdot a_{ij} \cdot b_j(\mathbf{O}_{t+1}) \cdot \beta_{t+1}(s_j)$. Two particular combinations of these variables are used in HMM learning [1]. The first is the smooth estimate of probabilities $\gamma_t(s_j^t) = P(s_j^t|\mathbf{O}_{1:T}, \lambda)$ with $\gamma_t(s_j) = \alpha_t(s_j) \cdot \beta_t(s_j)$, that is particularly used to both determine the most probable state at a given instant and models parameters. The second combination is $\xi_t(s_i^t, s_j^{t+1}) = P(s_i^t, s_j^{t+1}|\, \mathbf{O}_{1:T}, \lambda)$ with $\xi_t(s_i^t, s_j^{t+1}) \propto \alpha_t(s_i) \cdot a_{ij} \cdot b_j(\mathbf{O}_{t+1}) \cdot \beta_{t+1}(s_j)$ and allows to compute the expected probabilities of transition by: $a_{ij} \propto \sum_{t=1}^{T-1} \xi_t(s_i^t, s_j^{t+1})$. Lastly, the *Viterbi* algorithm determines the best sequence of hidden states. It is actually a *forward* propagation where the *sum-product* operator is replaced by a *max-product*.

3. TRANSFERABLE BELIEF MODEL (TBM)

The belief of an agent on subsets of the frame of discernment Ω_t can be represented by a belief mass assignment (BBA) $m^{\Omega_t} : 2^{\Omega_t} \to [0, 1]$, $A \to m^{\Omega_t}(A)$. The belief mass satisfies $\sum_{A \subseteq \Omega_t} m^{\Omega_t}(A) = 1$ and can be assigned to singleton state ($|A| = 1$) or to *subsets* ($|A| > 1$), *without assumption concerning additivity* that is one important difference with probabilities. Conditional BBA can be used to represent state of knowledge, e.g. $m^{\Omega_t|\Omega_{t-1}}[S_i](\cdot)$ is a BBA defined on Ω_t conditionally to subset $S_i \subseteq \Omega_{t-1}$. The mass $m^{\Omega_t}(\emptyset)$ is the mass of conflict and dividing each belief mass by $1 - m^{\Omega_t}(\emptyset)$ while forcing $m^{\Omega_t}(\emptyset) = 0$ is called Dempster *normalization*.

Several functions, which are in *one-to-one correspondance* [12], can be computed from a BBA which are then used to simplify the computation of combination rules. In the sequel, the function w^{Ω_t} (representing the weights of the canonical conjunctive decomposition (WCD) [14]) will be used. They are defined $\forall B \subseteq \Omega_t$ by:

$$w^{\Omega_t}(B) = \prod_{B \subseteq C} q^{\Omega_t}(C)^{(-1)^{|C|-|B|+1}} \tag{1}$$

where the commonality q is $q^{\Omega_t}(B) = \sum_{C \supseteq B} m^{\Omega_t}(C)$. We will also use the credibility $bel^{\Omega_t}(B) = \sum_{C \subseteq B} m^{\Omega_t}(C)$ and the plausibility $pl^{\Omega_t}(B) = \sum_{C \cap B \neq \emptyset} m^{\Omega_t}(C)$ with:

$$pl^{\Omega_t}(\Omega_t) = 1 - m^{\Omega_t}(\emptyset) \tag{2}$$

The combinations of BBAs can then be achieved by four main rules [14]: the conjunctive combination (CRC, $\textcircled{\cap}$), the disjunctive combination (DRC), the "cautious" CRC (CCRC, $\textcircled{\wedge}$), and the "bold" DRC. The two last ones are used when sources of belief are not independent. The CCRC is:

$$w^{\Omega_t}_{1 \circledwedge 2}(B) = w^{\Omega_t}_1(B) \circledast w^{\Omega_t}_2(B) \tag{3}$$

where $\circledast$ is a "generalized" cautious rule (GCR) based on positive t-norms/t-conorms (see [14]-§5.1 for details). This offers the possibility to use an infinity number of t-norm/t-conorm operators ($\top$ and $\perp$) defined in possibility theory. For example [14] the Frank t-norm can be used:

$$x \top_s y = \begin{cases} x \wedge y & \text{if } s = 0 \\ x \cdot y & \text{if } s = 1 \\ \log_s \left(1 + \dfrac{(s^x - 1) \cdot (s^y - 1)}{s - 1} \right) & \text{otherwise} \end{cases} \tag{4}$$

where $s \in [0, 1]$ and $\wedge$ is the *minimum*. If $s = 0$ (resp. $s = 1$), the CCRC (resp. the CRC) is obtained. If the CRC is used and the obtained belief mass is normalized then we obtain *Dempster rule* [15].

The conditioning process is a special case of conjunctive combination using the CRC between a first belief mass (to be conditioned) with a second one (said categorical) where the latter has the particularity to be not nil only for one element $A \subset \Omega_t$ ("A" is the condition).

Decision making in TBM is made by choosing the best hypothesis from the pignistic probability distribution [12] obtained $\forall s_k \in \Omega_t$ by:

$$\mathbf{BetP}\{m^{\Omega_t}\}(s_k) = \frac{1}{1 - m^{\Omega_t}(\emptyset)} \sum_{B \subseteq \Omega_t, s_k \in B} \frac{m^{\Omega_t}(B)}{|B|} \tag{5}$$

4. PARTIALLY-HIDDEN MARKOV MODELS

Before tackling the problem of evidential propagations with dependent observations, let consider an intermediate situation where one uses probabilistic HMM when data are annotated by belief functions. Let $\mathbf{O_t}$ be an instance of the learning set. One can have information about the belief of this observation to belong to one of the K states, i.e.

each observation in the learning set is annotated by a belief function $m_o^{\Omega_t}$. This can be useful because one can learn a HMM in a supervised / unsupervised / partially supervised manner. Indeed, in case of *supervised learning*, the belief function of each data in the learning set is categorical (e.g. for all instances, the whole mass is assigned to one singleton) thus $m_o^{\Omega_t}(B) = 1$ for $B \subseteq \Omega_t$ and $\forall B \neq C \subseteq \Omega, m_o^{\Omega_t}(C) = 0$. In *unsupervised learning*, the belief function of each data is vacuous, i.e. $m_o^{\Omega_t}(\Omega_t) = 1$ and *semi-supervised learning* is a mixing of both previous cases. Finally, in *partially-supervised learning*, belief functions are "general" thus $m_o^{\Omega_t}(B) \in [0,1[, \forall B \subseteq \Omega_t$. So the problem is: how to take into account partial labels (fourth case) in probabilistic HMM learning? For that, we rely on [16] concerning EM-based parameters learning using partial labels which has shown promising results for static data.

Let consider the forward variable that has a pivotal role in HMM learning: $\alpha_t(s_j) = P(\mathbf{O}_1, \mathbf{O}_2 \ldots \mathbf{O}_t, s_j|\lambda) = \sum_{\forall s_i \in \Omega_{t-1}} \alpha_{t-1}(s_i) \cdot a_{ij} \cdot b_j(\mathbf{O}_t)$. In case a partial knowledge on states at t is available in the form of a belief function on states $m_o^{\Omega_t}$, it can be taken into account by the conjunctive combination with α_t. Since the probabilistic forward variable has only positive values for singletons, the conjunctive combination leads to:

$$\left(\alpha_t \circledcirc m_o^{\Omega_t}\right)(s_i) = \alpha_t(s_i) \cdot pl_o^{\Omega_t}(s_i) \tag{6}$$

and thus the forward variable can be rewritten as:

$$\alpha_t(s_j) = \sum_{\forall s_i \in \Omega_{t-1}} \alpha_{t-1}(s_i) \cdot a_{ij} \cdot b_j(\mathbf{O}_t) \cdot pl_o^{\Omega_t}(s_j) \tag{7}$$

A similar result can be obtained for the backward variable. Given a learning set annotated by belief functions, only plausibility on singletons need to be stored and then used in the forward and backward recursions for HMM learning[1].

5. THE CLASSIFICATION PROBLEM

Evidential forward algorithm - The evidential forward propagation is used to assess the *forward* variable at instant t and requires 1) the BBA $m_\alpha^{\Omega_{t-1}}[\lambda]$ of the *forward* variable at the previous instant $t-1$ (conditional to the parameters that governs the corresponding EvHMM), 2) the conditional WCD $w_a^{\Omega_t}[S_i^{t-1}]$ defined on Ω_t conditionally to subsets of preceding states ($S_i^{t-1} \subseteq \Omega_{t-1}$) and representing transitions between subsets of states at $t-1$ and t, and 3) the WCD $w_b^{\Omega_t}[\mathbf{O}_t]$ obtained from observations at t.

Proposition 1 *The computation of the evidential forward variable at t satisfies a "prediction – update" mechanism:*

$$w_\alpha^{\Omega_t}[\lambda](S_j^t) = w_b^{\Omega_t}[\mathbf{O}_\mathbf{t}](S_j^t) \circledast \Big(\sum_{S_i^{t-1} \subseteq \Omega_{t-1}} m_\alpha^{\Omega_{t-1}}[\lambda](S_i^{t-1}) \cdot w_a^{\Omega_t}[S_i^{t-1}](S_j^t) \Big) \tag{8}$$

GCR rules $\circledast$ endow the evidential forward propagation with a panel of operators that could be chosen according to the application and the data [14]. It has been demonstrated in [14] that these operators are suited for dependent data. Moreover, another advantage is that the prior on states at $t=1$ can be "vacuous" and this state of knowledge is easily represented by $w_\alpha^{\Omega_1}[\lambda](S_i^1) = 1, \forall S_i^1 \subseteq \Omega_1$ meaning total ignorance. Besides, this evidential version generalizes the probabilistic one when transitions, observations and prior are all Bayesian (i.e. the belief masses are only on singleton hypotheses) and when one uses Dempster rule [15]. The complexity of the forward variable is in $T \times |\Omega_t| \times 2^{|\Omega_t|}$.

A conflict may appear at each instant between prediction and observation meaning both quantities are contradictory. This *conflict* must be cancelled out because it is absorptive by a conjunctive rule. Actually, the required normalization procedure is the same as in probabilistic HMM [1]. Unlike probabilistic HMM where the normalisation process consists in redistributing uniformly $1 - \sum_j \alpha_t(j)$ to each state at t, belief function framework provides complex and sound redistribution rules [17]. The parallel between both formalisms is more obvious when seeing the (seldom emphasized) relation between conflict and plausibility (Eq. 2) that represents a bridge between HMM and EvHMM. As shown in many applications (object association [18], image processing [19], change detection and state sequence recognition [11, 10]), conflict is a useful information that is used below for classification of state sequences.

Classification in EvHMM - Given an observation sequence $\mathbf{O}_{1:T}$ (length T) and a set of N EvHMM $\lambda_{1\ldots N}$, how to choose the EvHMM that best fits observations?

Proposition 2 *The lower is the conflict throughout the whole observation sequence, the better is the EvHMM λ for explaining these observations. The whole conflict and the best EvHMM are given by:*

$$\mathcal{L}_c(\lambda_n) = \frac{1}{T} \sum_{t=1}^{T} \log(1 - m_\alpha^{\Omega_t}[\lambda_n](\emptyset)) \tag{9a}$$

$$\lambda^* = \underset{n}{\operatorname{argmax}}\ \mathcal{L}_c(\lambda_n) \tag{9b}$$

This criterion reduces to the probabilistic one when using Dempster rule and Bayesian BBAs. Conclusively, the evidential forward algorithm can be used for online filtering of belief functions on states and for the classification of observation sequences in five steps: 1) the initial BBA ($t=1$) on states can be vacuous, 2) Eq. 8 is applied from $t=2$ to T, 3) the conflict at t is stored and then redistributed, 4) the classification criterion (Eq. 9) is finally computed.

6. TIME-INSTANT STATE RECOGNITION AND TRANSITION ESTIMATION

Evidential backward algorithm - The evidential backward algorithm is similar to the forward one and requires the BBA

[1]Thanks to Pr. Denœux (HEUDIASYC lab, thierry.denoeux@utc.fr) for discussions on this subject.

$m_\beta^{\Omega_{t+1}}[\lambda]$ of the backward variable defined on subsets of states at $t+1$, the set of conditional plausibilities $pl_a^{\Omega_{t+1}}[S_i^t]$ representing transitions from subsets of states at t and $t+1$ and the observed BBA $m_b^{\Omega_{t+1}}[\mathbf{O_{t+1}}]$.

Proposition 3 *The belief on states at t in the backward recursion is obtained by:*

$$w_\beta^{\Omega_t}(S_i^t) = \sum_{S_j^{t+1} \subseteq \Omega_{t+1}} m_{\beta \circledast b}^{\Omega_{t+1}}(S_j^{t+1}) \cdot w_a^{\Omega_t}[S_j^{t+1}](S_i^t) \quad (10)$$

where $m_{\beta \circledast b}^{\Omega_{t+1}}$ is the conjunctive combination of BBAs at Ω_{t+1}, i.e. $m_\beta^{\Omega_{t+1}}$ and $m_b^{\Omega_{t+1}}$. The posterior conditional WCDs $w_a^{\Omega_t}[S_j^{t+1}]$ are obtained from plausibilities by calculating the conditional commonalities using the GBT [8]:

$$q_a^{\Omega_t|\Omega_{t+1}}[S_j^{t+1}](S_i^t) = \prod_{s_i^t \in S_i^t} pl_a^{\Omega_{t+1}|\Omega_t}[s_i^t](S_j^{t+1}) \quad (11)$$

then used in Eq. 1 to obtain WCDs. Unknown prior at T is modelled by a vacuous BBA ($w_\beta^{\Omega_T}(S_i^T) = 1, \forall S_i^T \subseteq \Omega_T$). Eq. 10 reduces to the probabilistic case if all BBAs are Bayesian and if Dempster rule is used.

Instant state recognition - As in probabilistic HMM, the conjunctive combination of both forward and backward variables (called γ-variable) is performed when observations are available as a whole.

Proposition 4 *The γ-variable is given by:*

$$w_\gamma^{\Omega_t}(S_j^t) = w_\alpha^{\Omega_t}(S_j^t) \circledast w_\beta^{\Omega_t}(S_j^t) \quad (12)$$

One has to use the same rule $\circledast$ as in the forward and backward variables. The γ-variable can then be exploited for two functionalities. The first one is *offline smoothing* of belief functions on states. The second one is the *detection of the best state* s_t^* at a given instant and for a given EvHMM. This state is found out by maximizing the pignistic probability (Eq. 5) based on $m_\gamma^{\Omega_t}, \forall t \in \{1 \dots T\}$, i.e. $s_t^* = \text{argmax}_{\ s_i \in \Omega_t} \mathbf{BetP}\{m_\gamma^{\Omega_t}\}(s_i)$. As in the probabilistic case [1], this technique is not always well-suited for state sequence recognition because is too "local". Instead, an evidential version of the Viterbi algorithm is proposed (§ 7).

Transition estimation - One can mimic the computation of the so-called ξ-variable [1] that is used in probabilistic HMM for transition estimation in an Expectation-Maximization (EM) learning process. A proposition has been formulaed in [13] for independent observations. However, by doing so, one can face a serious problem: loosing the interest of belief functions. Indeed, EM is an iterative procedure while (evidential) transition estimation is based on conjunctive combinations. Therefore, due to the repeted conjunctive combinations at each iteration, the transition matrix is expected to gradually tend to a probabilistic one. Therefore, we rather propose below an estimation formula of transitions that is independent of transitions themselves. It consists in computing the expected belief mass of making a transition from one state to another. For that, it is required to combine conjunctively two observed BBAs defined at two successive instants and then taking the mean of these belief masses as proposed below.

Proposition 5 *An estimation of the expected transition from subsets to subsets from observations is given by:*

$$m_{\hat{a}}^{\Omega_t \times \Omega_{t+1}} \propto \sum_{t=1}^{T-1} \left(m_b^{\Omega_t \uparrow \Omega_t \times \Omega_{t+1}} \circledast m_b^{\Omega_{t+1} \uparrow \Omega_t \times \Omega_{t+1}} \right) (13)$$

up to a constant $\frac{1}{T-1}$ and where $m_b^{\Omega_t \uparrow \Omega_t \times \Omega_{t+1}}$ is the vacuous extension [12] of the belief mass $m_b^{\Omega_t}[\mathbf{O}_t]$ on the cartesian product defined by: $m_b^{\Omega_t \uparrow \Omega_t \times \Omega_{t+1}}(B) = m_b^{\Omega_t}(C)$ if $C \times \Omega_{t+1} = B$ and 0 otherwise. If one has prior information on transitions, it can be combined conjunctively with observations in Eq. 13.

The backward recursion requires Eq. 11 to be computed and for that one needs plausibilities conditional to *singleton states*. They can be obtained from the joint belief mass distribution (Eq. 13) by conditioning it by each singleton state $s_i^t \in \Omega_t$ followed by a marginalization onto Ω_{t+1}, formally $m_{\hat{a}}^{\Omega_t \times \Omega_{t+1}}[s_j^{t+1}]^{\downarrow \Omega_t}$ (the two processes are explained in [8, 12]). This solution satisfies the constraint of independence of conditional beliefs (transition) to apply the GBT. Due to the *duality GBT-DRC* [8], Eq. 11 computes *implicitly* the beliefs conditional to unions of states by the disjunctive rule. Transitions expressed as a joint belief mass can also be used directly in both forward-backward recursions and this requires to compute the conjunctive combination on a joint space followed by a marginalization onto Ω_t to obtain the evidential variables at t. Eq. 13 has been used to estimate the parameters of the Temporal Evidential Filter [11].

7. STATE SEQUENCE FROM OBSERVATIONS

For a given model, the goal is to determine the best state sequence $s_*^1, s_*^2, \dots s_*^T$ given the sequence of observations $\mathbf{O}_{1:T}$. In probabilistic HMM, the powerful Viterbi algorithm, based on dynamic programming, achieves this goal by selecting the sequence with maximum likelihood without computing the N^T sequences.

An optimal evidential Viterbi decoding - If we consider sequences of singleton states $s^1, s^2, \dots s^T$, then one can apply exactly the same algorithm as in probabilistic HMM except that transitions and observations are represented by the plausibilities on (and conditional to) singletons computed from the related belief masses. The Viterbi metric is a thus propagated plausibility initialized to 1 at the beginning of the sequence reflecting missing prior. The complexity is then the same as in the probabilistic case.

A sub-optimal evidential Viterbi decoding - The previous algorithm, even if optimal in the sense of maximum plausibility / likelihood, reduces belief functions to single values due to the conditioning and propagation processes.

In [13] was proposed an algorithm that draws benefits from belief function modelling and that is able to *postpone* the decision concerning the best predecessors. For that, at each step, it updates a propagated metric that is a belief function $w_\delta^{\Omega_t}$. This algorithm is a Viterbi-*like* decoder since decisions are not directly propagated. In the sequel, conditioning on EvHMM parameters $[\lambda]$ are omitted for simplicity.

Proposition 6 *When observations are not independent, the Viterbi-like metric at t is defined* $\forall S_j^t \subseteq \Omega_t$ *by:*

$$w_\delta^{\Omega_t}(S_j^t) = w_b^{\Omega_t}[\mathbf{O}_t](S_j^t) \circledast \sum_{S^{t-1}\subseteq\Omega_{t-1}} w_a^{\Omega_t|\Omega_{t-1}}[S^{t-1}\cap\mathcal{A}^{t-1}](S_j^t)\cdot m_\delta^{\Omega_{t-1}}(S^{t-1}) \quad (14)$$

where $\mathcal{A}^{t-1} = \bigcup_{s_j^t\in\Omega_t}\psi'(s_j^t)$ is the union of predecessors of states at t. The main idea behind conditioning on the set of predecessors $\mathcal{A}^{t-1}$ is to make a one-step forward propagation of the belief masses of the most important subsets of $m_\delta^{\Omega_{t-1}}$. Set $\mathcal{A}^{t-1}$ is obtained by first searching which is the predecessor state at $t-1$ of each state at t and this is performed in three steps. First, since the predecessor of a state s_j^t is unknown, it is needed to condition $w_\delta^{\Omega_{t-1}}$ on all possible candidates $s_i^{t-1}\in\Omega_{t-1}$. Thus, a WCD $w_{\delta,i}^{\Omega_t}$ is computed by applying Eq. 14 and replacing $w_a^{\Omega_t|\Omega_{t-1}}[S^{t-1}\cap\mathcal{A}^{t-1}]$ by $w_a^{\Omega_t|\Omega_{t-1}}[S^{t-1}\cap s_i^{t-1}]$. The second step aims at making a decision concerning predecessors for each state s_j^t based on $m_{\delta,i}^{\Omega_t}$ using the decision criterion (Eq. 5):

$$\mathcal{P}_t[s_i^{t-1}](s_j^t) = \mathbf{BetP}\{m_{\delta,i}^{\Omega_t}\}(s_j^t) \quad (15)$$

Conditioning on s_i^{t-1} in the first step is equivalent to a conjunctive combination with a categorical BBA (§ 3) and the generated conflict $m_{\delta,i}^{\Omega_t}(\emptyset)$ quantifies how irrelevant is the hypothesis "*the predecessor of* s_j^t *is* s_i^{t-1}". When the two first steps are done for all previous states $s_i^{t-1}\in\Omega_{t-1}$, the most probable predecessor of $s_j^t\in\Omega_t$ is found by:

$$\psi_t'(s_j^t) = \underset{s_i^{t-1}\in\Omega_{t-1}}{\operatorname{argmax}}\left[\left(1-m_{\delta,i}^{\Omega_t}(\emptyset)\right)\cdot\mathcal{P}_t[s_i^{t-1}](s_j^t)\right] \quad (16)$$

where $\psi_t'(s_j^t)$ stores the best predecessor of s_j^t. Note that these three first steps are strictly equivalent to the probabilistic mechanism. Then, values of ψ_t' are used to generate $\mathcal{A}^{t-1}$ (Eq. 14). They are also used to compute a propagated metric defined at each t by:

$$\mathcal{Q}_t(s_*^t,\lambda) = \mathcal{Q}_{t-1}(\psi_t'(s_j^t),\lambda)\cdot pl_\delta^{\Omega_t}(s_j^t) \quad (17)$$

At $t=1$, $\mathcal{Q}_t(s_*^t,\lambda)=1$ (reflecting missing prior) and at $t=T$, the best state $s_*^T = \operatorname{argmax}_{s_j^T\in\Omega_T}\mathcal{Q}_T(s_j^T,\lambda)$ is used for state backtracking by $s_*^t = \psi_{t+1}'(s_*^{t+1})$. The value of the metric $\mathcal{Q}_T(s_*^T,\lambda)$ at T quantifies the plausibility of the state sequence given the EvHMM and can be used for state sequence classification.

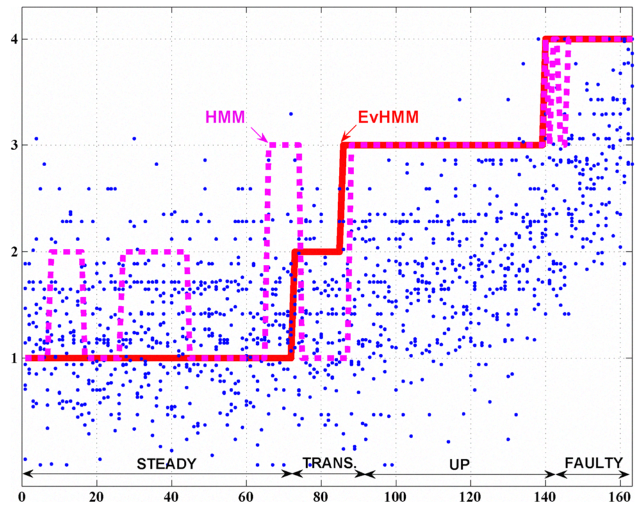

Fig. 1. EvHMM and HMM Viterbi decodings. Modes: steady (1), transition (2), up (3), faulty (4). Blue dots: observations with values normalized in $[0,4]$ for visualization.

8. ILLUSTRATION ON FAULT DIAGNOSIS

The evidential Viterbi-like decoder is illustrated on challenge data of diagnostic and prognostic of machine faults from the Int. Conf. on Prognostics and Health Management [21] (available at `https://dashlink.arc.nasa.gov/member/goebel/`). Data set consists of multiple multivariate time series with sensor noise (blue dots in Fig. 1). Each time serie is from a different engine of the same fleet and each engine starts with different degrees of initial wear and manufacturing variation unknown to the user and considered normal. The engine is operating normally at the start and develops a fault at some point. The fault grows in magnitude until system failure (we have used the file *train_FD001.txt* for training and *test_FD001.txt* for testing).

Given a new observation sequence measured on an engine, the goal is to diagnose its current mode and to determine whether the system is faulty. A fault occurs when a sequence of four modes is detected: steady → transition → up → faulty. One detector is built for each mode with an EM run on the training set using mixture of Gaussians with three components. A four states EvHMM is built where the evidential transition matrix is estimated as proposed in this paper and using the training set. The chosen GCR operator $\circledast$ is the Frank t-norm (Eq. 4) with parameter $s=0.9$. To exploit the EvHMM, we first use the GBT to transform the likelihoods generated by each detector into belief function distributions [8, 20]. These beliefs are then used in the Viterbi decoder. To compare EvHMM with HMM, we transform the evidential transition matrix into a pignistic one and use the likelihoods directly.

An illustration of the detection by both systems is given in Fig. 1 (HMM in light magenta and EvHMM in bold red).

	Steady	Transition	Up	Faulty
Steady	**82**	48	7	0
Transition	17	**42**	18	0
up	1	10	**57**	4
Faulty	0	0	18	**96**

	Steady	Transition	Up	Faulty
Steady	**87**	36	2	0
Transition	13	**59**	25	0
up	0	5	**56**	2
Faulty	0	0	17	**98**

Table 1. Confusion matrices for HMM (top, global accuracy: 68%) and EvHMM (bottom, global accuracy: 75%), values are percents.

The differences in detections come mainly from the lack of data for HMM learning while EvHMM are less sensitive. This lack of sensitivity on the size of the dataset, peculiar to belief functions [16], is important in fault diagnosis to decrease the number of tests on machines to build datasets. Moreover, the evidential Viterbi is better in part because it postpones the decision until the last instant thanks to the conditioned forward propagation. Table 1 presents confusion matrices of detections by HMM and EvHMM where EvHMM provides a classification rate of 75% against 68% for HMM. Confusion matrices reflects also an obvious mixing between "transitions" with both "steady" and "up" phases. One solution should be to consider "transition" as belonging to both "steady" and "up" phases while keeping only three modes. Sharing the same training set will not be a problem using GCR rules.

9. CONCLUSION

Extensions of HMM propagations to belief functions are presented. Contributions of belief functions to HMM are: 1) Possibility to process temporal belief functions, 2) Opportunity to mix several uncertainty formalisms into a flexible one, 3) Partially supervised learning of HMM, 4) Processing possibly-dependent observations by parametrized operators [14], 5) Availability of a variety of combination operators and normalization processes. Higher computational cost is expected but solutions exist to reduce it.

A thorough exploration of EvHMM is required for further applications. Moreover, generalizing the approach developed here to other graphical models as well as the development of a criterion for learning in EvHMM taking into account both models and transitions together are under study.

10. REFERENCES

[1] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, pp. 257–285, 1989.

[2] K. P. Murphy, *Dynamic Bayesian Networks: Representation, inference and learning*, Ph.D. thesis, Berkeley (CSD), 2002.

[3] G.J. Klir and M.J. Wierman, *Uncertainty-based information. Elements of generalized information theory*, Studies in fuzzyness and soft computing. Physica-Verlag, 1999.

[4] M.A. Mohamed and P. Gader, "Generalized HMM - part I: Theoretical frameworks," *IEEE Tr. on Fuzzy Systems*, vol. 8, pp. 67–81, 2000.

[5] L. Fouque, A. Appriou, and W. Pieczynski, "Evidential markovian model for data fusion and unsupervised image classification," in *ICIF'00*, vol. TuB4, pp. 25–31.

[6] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *IJAR*, vol. 45, pp. 1–16, 2007.

[7] H. Soubaras, "An Evidential Measure of Risk in Evidential Markov Chains ," in *ECSQARU'09*, LNCS Vol. 5590.

[8] Ph. Smets, "Beliefs functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem," *IJAR*, vol. 9, pp. 1–35, 1993.

[9] M. Rombaut, I. Jarkass, and T. Denoeux, "State recognition in discrete dynamical systems using Petri nets and Evidence theory," in *ESCQARU'99*, pp. 352–361.

[10] E. Ramasso, C. Panagiotakis, M. Rombaut, D. Pellerin, and G. Tziritas, "Human shape-motion analysis in athletics videos for coarse to fine action/activity recognition using Transferable Belief Model," *ELCVIA*, vol. 4, pp. 32–50, 2009.

[11] E. Ramasso, M. Rombaut, and D. Pellerin, "State filtering and change detection using TBM conflict," *IEEE Tr. on CSVT*, vol. 17, pp. 944–949, 2007.

[12] Ph. Smets and R. Kennes, "The Transferable Belief Model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[13] E. Ramasso, M. Rombaut, and D. Pellerin, "Forward-backward-viterbi procedures in TBM for state sequence analysis using belief functions," in *ECSQARU'07*, pp. 405–417.

[14] T. Denoeux, "Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence," *Artificial Intelligence*, vol. 172, pp. 234–264, 2008.

[15] A.P. Dempster, "A generalization of Bayesian inference," *Journal of the RSS*, vol. 30, pp. 205–247, 1968.

[16] E. Côme, L. Oukhellou, T. Denoeux, and P. Aknin, "Mixture model estimation with soft labels," in *SMPS'08*.

[17] Ph. Smets, "Analyzing the combination of conflicting belief functions," *Info. Fusion*, vol. 8, pp. 387–412, 2005.

[18] B. Ristic and Ph. Smets, "The TBM global distance measure for the association of uncertain combat ID declarations," *Info. Fusion J.*, vol. 7, pp. 276–284, 2006.

[19] A.S. Capelle, O. Colot, and C. Fernandez-Maloigne, "Evidential segmentation scheme of multi-echo MR images for the detection of brain tumors using neighborhood information," *Info. Fusion*, vol. 5, pp. 203–216, 2004.

[20] F. Delmotte and Ph. Smets, "Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model," *IEEE Tr. on SMC*, vol. 34, 2004.

[21] A. Saxena, K. Goebel, D. Simon and N. Eklund "Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation," in *IEEE PHM'08*, Denver CO.