A solution for the learning problem in Evidential (Partially) Hidden Markov Models based on Conditional Belief Functions and EM

Emmanuel Ramasso

FEMTO-ST Institute, UMR CNRS 6174 - UBFC / ENSMM / UTBM,
Applied Mechanics / Automatic Control & Micro-Mechatronic Systems Dep.,
25000, Besançon, France
emmanuel.ramasso@femto-st.fr

Abstract. Evidential Hidden Markov Models (EvHMM) is a particular Evidential Temporal Graphical Model that aims at statistically representing the kynetics of a system by means of an Evidential Markov Chain and an observation model. Observation models are made of mixture of densities to represent the inherent variability of sensor measurements, whereas uncertainty on the latent structure, that is generally only partially known due to lack of knowledge, is managed by Dempster-Shafer's theory of belief functions. This paper is dedicated to the presentation of an Expectation-Maximization procedure to learn parameters in EvHMM. Results demonstrate the high potential of this method illustrated on complex datasets originating from turbofan engines where the aim is to provide early warnings of disfunction.

Keywords: Evidential Temporal Graphical Model, Evidential latent variable, Markov Chain, Belief Functions, Parameter learning

1 Introduction

The statistical representation of multi-dimensional time-series originating from a dynamical system consists in finding a concise and meaningful mathematical model that can be easily interpreted and used to undertand the behavior of the system. Those models can then be used to enhance data-driven phenomenological physics model with better prediction capabilities in in-service applications. However, sources of uncertainty are numerous in real-world applications which accounts for systems' oversizing to ensure people safety and equipments' availability. Uncertainty quantification thus plays a critical role during both systems' design (upstream) and in-service monitoring (downstream).

Dempster-Shafer's theory of belief functions is a mathematical framework that allows to represent, quantify and propagate uncertainties. Its application in mechanical engineering has however been limited due to a lack of tools to handle temporal data. The additional temporal dimension compared to static data makes both the inference and learning problems more difficult than with probability theory.
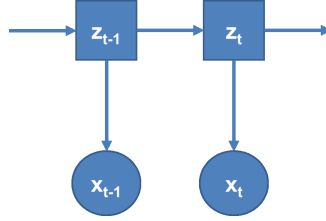
Fig. 1. Graphical representation of a hidden Markov model: $\mathbf{z}$ is discrete and hidden, $\mathbf{x}$ is continuous and observed, t is the time index.

This paper is focused on a simple – yet not tackled – problem that is the estimation of parameters in Hidden Markov Models when uncertainty is no more represented by probabilities but by belief functions. This model has been initially called Evidential Hidden Markov Models (EvHMM) and represents a particular statistical discrete-latent Markov model which is depicted in Fig. 1. This model assumes that the system is driven by a doubly stochastic process: A Markov chain on discrete hidden (not observed) variables called states, and an observation model that statistically represents the distribution of sensor measurements (recorded on the system) defined conditionally to the states.

Inference mechanisms in EvHMM has been proposed by the author in [15, 20] (not recalled here) to estimate the belief functions over hidden variables given both an observation model and data. Those mechanisms provide exact belief functions and enable one to compute the equivalent of a likelihood of a given model for some sequences of observations, as well as to estimate sequences of hidden states. Those procedures can then be used to explore relevant regions in the feature or parameter space.

Only the learning problem is considered subsequently, with some assumptions required to make the problem tractable [17].

2   Parameter learning in Evidential discrete-latent models

2.1   The criterion

The quality of a model such as depicted in Fig. 1 can be quantified by minimizing the *amount of conflict* between the model and the data. Time-series are denoted as $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \ldots \mathbf{x}_T]$ with length T in D dimensions with $\mathbf{x}_t = (x_1, \ldots, x_D)'$ called feature vector. The latent states are represented by discrete random variables $z_1, z_2 \ldots z_t$ taking values in a finite set $\Omega_z = \{s_1, s_2, \ldots, s_K\}, s_i \cap s_j = \emptyset$.

Finding the parameters θ^* in a latent variable model, in which uncertainty is managed by belief functions, and made of one observed variable $\mathbf{x}_t$ and one discrete hidden variable $\mathbf{z}_t$ with $t = 1 \ldots T$, can be turned into the maximization of the potential support assigned to the subset $(\mathbf{x}_1, \Omega_z), (\mathbf{x}_2, \Omega_z) \ldots (\mathbf{x}_T, \Omega_z)$ after observing all data vectors:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \; pl^{\mathbb{R}^T \times \Omega_z^T} \left((\mathbf{x}_1, \Omega_z), (\mathbf{x}_2, \Omega_z) \ldots (\mathbf{x}_T, \Omega_z) \mid \theta \right) \qquad (1)$$

$\mathbb{R}$ is the domain of $\mathbf{x}_t$. For short, this criterion is rewritten as argmax $_\theta\, pl(\mathbf{X}, \boldsymbol{\Omega_z} \mid \theta)$. By definition of the plausibility function, the criterion can be computed by summing belief masses assigned to all configurations of the hidden variables $\mathbf{S}$:

$$pl(\mathbf{X}, \boldsymbol{\Omega_z} \mid \theta) = \sum_{\mathbf{S} \not\ni \emptyset} m(\mathbf{X}, \mathbf{S} \mid \theta) \tag{2}$$

Direct maximization of this criterion is untractable but by making use of the latent structure, it is possible to formulate the problem differently in order to use an EM procedure (Expectation-Maximization) [6]. In EM, the E-step can indeed be used to estimate the distribution over latent variables given both the data and the current values of the parameters $\theta^{(q)}$ (at iteration q) while the M-step can be used to find the parameters $\theta^{(q+1)}$ that allows to maximize an "auxiliary" function such that the criterion (that is not directly maximized) does not decrease. For that, we can rewrite the criterion as

$$pl(\mathbf{X}, \boldsymbol{\Omega_z} \mid \theta) = \sum_{\mathbf{S} \not\ni \emptyset} R(\mathbf{S}) \frac{m(\mathbf{X}, \mathbf{S} \mid \theta)}{R(\mathbf{S})} \tag{3}$$

where $S \subseteq \Omega_z$ and R is a distribution such that $\sum_A R(A) = 1$ that allows Jensen's inequality to be applied [8, Eq. 5]:

$$\log pl(\mathbf{X}, \boldsymbol{\Omega_z} \mid \theta) \geq \mathcal{Q}_{m,m}(\theta^{(q)}, \theta) - H_{m,m}(\theta^{(q)}, \theta^{(q)}) \tag{4a}$$

$$\mathcal{Q}_{m,m}(\theta^{(q)}, \theta) = \sum_{\mathbf{S} \not\ni \emptyset} R(\mathbf{S}, \theta^{(q)}) \log m(\mathbf{X}, \mathbf{S} \mid \theta) \tag{4b}$$

$$H_{m,m}(\theta^{(q)}, \theta^{(q)}) = \sum_{\mathbf{S} \not\ni \emptyset} R(\mathbf{S}, \theta^{(q)}) \log R(\mathbf{S}, \theta^{(q)}) \tag{4c}$$

$$s.t. \sum_{\mathbf{S}} R(\mathbf{S}, \theta^{(q)}) = 1 \tag{4d}$$

where $H_{m,m}$ depends only on previous estimates $\theta^{(q)}$ and allows to underline that when the function in the logarithm ideally evolves towards the target R (which can change at each iteration) then $\mathcal{Q}_{m,m} - \mathcal{H}_{m,m} \to 0$. Since $R(\mathbf{S})$ must sum up to 1, it follows that a rational choice for R is a BBA denoted as m_γ subsequently and $\mathcal{Q}_{m,m}$ is an expectation taken with respect to m_γ.

2.2   E/M-steps

An EM-like procedure can thus be applied. At iteration q, the E-step aims at maximizing the expectation 4b given fixed parameters $\theta^{(q)}$. We can cancel its derivative with repect to R using appropriate Lagrangian multipliers (integrating the aforementionned constraint on R) to get the maximizer $m_\gamma^{(q)}$:

$$\text{E-step:} \ \Rightarrow m_\gamma^{(q)} = \frac{m(\mathbf{X}, \mathbf{S} \mid \theta^{(q)})}{\sum_{\mathbf{S}' \not\ni \emptyset} m(\mathbf{X}, \mathbf{S}' \mid \theta^{(q)})} \equiv m(\mathbf{S} \mid \mathbf{X}, \theta^{(q)}) \tag{5}$$

$m_\gamma^{(q)}(\cdot \mid \mathbf{X})$ is the posterior BBA on states given observations. The posterior is then used in the M-step to find the best estimate $\theta^{(q+1)}$ for the next iteration so that it maximizes the expectation under $m_\gamma^{(q)}$:

$$\text{M-step: } \Rightarrow \theta^{(q+1)} = \underset{\theta}{\operatorname{argmax}} \ \mathbb{E}_{m_\gamma^{(q)}} [\log m(\mathbf{X}, \mathbf{S} \mid \theta)] \tag{6}$$

The algorithm iterates likewise to standard EM until the relative increase of the support $pl(\mathbf{X}, \boldsymbol{\Omega_z})$ between two consecutive iterations remains below a threshold.

Property 1. Since R is a BBA, then Jensen's inequality holds so that this algorithm is guaranteed to converge.

The proof follows the same line of reasoning as in standard EM.

Conjecture 1. Similarly to the auxiliary function in EM [1, Theorem 2.1], the maximization of the lower bound $\mathcal{Q}_{m,m}$ does not decrease the total support.

This conjecture was implicitly assumed in [22] for the Credal EM algorithm applied to Gaussian Mixture Model.

Remark 1. According to the model considered, it can be practically feasible to check whether the conjecture holds or not. It is the case for EvHMM [17] by using the evidential forward propagation [15].

2.3   Incorporating evidential prior to adjust the posterior BBA

The target BBA $m_\gamma^{(q)}$ computed in the E-step is of paramount interest to reestimate the parameters. In cases of model's misspecification (choice of $\mathbb{A}$ for instance) or biases induced by the data collection process, this BBA may eventually lead to wrong parameter estimates.

One solution was proposed in [22]. It considers that the prior knowledge on hidden variables are encoded by a set of T belief functions. For temporal data, the prior may be defined on Ω_z with a BBA $m_{\text{prior}(t)}^{\Omega_z}$. Note that it is likely to encounter situations where the prior can be defined on $\Omega_z \times \Omega_z$. If nothing is known about the hidden variables, then $\forall t, m_{\text{prior}(t)}^{\Omega_z}(\Omega_z) = 1$. Those priors can then be incorporated into the computation of the mathematical expectation (Eq. 4a) by Dempster's rule $\textcircled{\cap}$ as proposed in [22]:

$$m_\gamma^{(q)} \leftarrow m_\gamma^{(q)} \textcircled{\cap} m_{\text{prior}(t)} \tag{7}$$

The second solution relies on the Theory of Weighted Distributions (TWD) [10] that allows to incorporate prior knowledge on expectations computed in EM [3].

3   Learning parameters in EvHMM

3.1   What is an EvHMM?

An EvHMM is a particular evidential discrete-latent model enhanced by a Markov chain [15] in which the states can be partially observable with some degree of uncertainties. It is defined by two main sets of parameters:

- Transition matrix $\mathbb{A}$: An entry a_{ij} represents the belief mass of observing subset S_j at time t given that the system was in subset S_i at $t-1$.
- Observation model: Allows to generate the belief mass on subset S_j at t given observation $\mathbf{x}_t$. Observations are supposed to follow a multivariate Gaussian Mixture Model (GMM) for each state, characterized by parameters $\Phi = \{\boldsymbol{\mu}, \mathbf{c}, \boldsymbol{\Sigma}\}$ representing the means, covariances and mixing weights.

The symbol $\theta = \{\mathbb{A}, \boldsymbol{\mu}, \mathbf{c}, \boldsymbol{\Sigma}\}$ represents the set of parameters of an EvHMM.

In this model, the joint BBA (Eq. 6) located in the logarithm of the criterion can not be expressed using only products which makes the estimation of parameters untractable.

Assumption 1 *It is possible to decouple the estimation of the transitions parameters in the Markov chain from the parameters in the observation model.*

This decoupling appears *naturally* in standard HMM due to factorisation [2, Chap. 13]. The criterion can thus be rewritten as $\mathcal{Q}_{m,m} = \mathcal{Q}_{m,m}^a + \mathcal{Q}_{m,m}^b$ where $\mathcal{Q}_{m,m}^a$ is related to the transitions while $\mathcal{Q}_{m,m}^b$ to the observation model.

3.2 M-step for the Markov chain

Suppose that the transition matrix is made of BBAs $m_a^{\Omega_z}(\cdot \mid S_{t-1}), S_{t-1} \subseteq \Omega_z$. A sequence $\mathbf{S} = (S_1, S_2, \ldots S_t \ldots S_T), S_t \subseteq \Omega_z$ starting at S_1 requires to considering that S_1 is true at $t = 1$, S_2 is true at $t = 2$ and so on.

Proposition 1. *The total support assigned to a sequence* $\mathbf{S} = (S_1, S_2, \ldots S_t \ldots S_T), S_t \subseteq \Omega_z$ *can be quantified by the plausibility on* $\Omega_z^T = \Omega_z \times \Omega_z \times \ldots \Omega_z$ *(T times) after conditioning on the sequence. Given a vacuous BBA on initial states, the total support is given by:*

$$pl^{\Omega_z^T}(\mathbf{S}) = \prod_{t=2}^{T} pl_a^{\Omega_z}(S_t \mid S_{t-1}) \tag{8}$$

It defines an Evidential Markov Chain (EMC).

Note that the solution is different if the prior is not vacuous. The solution is also different from the result proposed in [11, Def. 4.1]:

Definition 1 (proposed in [11]). *An EMC has been defined as*

$$m^{\Omega_z^T}(\mathbf{S}) = \prod_{t=2}^{T} m_a^{\Omega_z}(S_t \mid S_{t-1}) \tag{9}$$

This definition is of practical interest in the sequel since estimating the transition given plausibilities (Eq. 8), although exact, would lead to incoherences due to

the presence of BBA in the suggested EM procedure (Section 2). The proposed criterion has thus the following form:

$$\mathcal{Q}_{m,m}^a(\mathbb{A}^{(q)}, \mathbb{A}) = \sum_{t=2}^{T} \sum_{S_j \subseteq \Omega_z} \sum_{S_i \subseteq \Omega_z} m_{\xi(t,t-1)}^{\Omega_z \times \Omega_z}(S_i, S_j \mid \mathbb{A}^{(q)}) \log m_a^{\Omega_z}(S_j \mid S_i, \mathbb{A})$$

(10)

where $m_{\xi(t,t-1)}$ is a BBA defined on two consecutive time slices that represents the probability mass of observing two given subsets. The maximization of $\mathcal{Q}_{m,m}^a$ with respect to m_a at iteration (q) requires to take the derivative of $\mathcal{Q}_{m,m}^a$ and using appropriate Lagrangian multipliers (ensuring that $\sum_B m_a^{\Omega_z}(B \mid S_{t-1}) = 1, \forall S_{t-1} \subseteq \Omega_z$ yielding:

$$m_a^{(q+1)}(S_{j,t} \mid S_{i,t-1}) = \frac{\sum_{t=2}^{T} m_{\xi(t,t-1)}^{\Omega_z \times \Omega_z}(S_i, S_j \mid \mathbb{A}^{(q)})}{\sum_{t=2}^{T} \sum_{\emptyset \neq S_l \subseteq \Omega_z} m_{\xi(t,t-1)}^{\Omega_z \times \Omega_z}(S_i, S_l \mid \mathbb{A}^{(q)})}$$

(11)

By assuming that the BBAs defined conditionally to subsets are computed by the DRC based only on BBAs defined conditionally to singletons, it follows that Eq. 11 allows to estimate $|\Omega_z| \times 2^{|\Omega_z|}$ parameters.

3.3   M-step for the observation model

In [22], the authors suggested an approach (EM-like) to estimate the parameters in a GMM using belief functions to represent uncertainty on mixing (discrete latent) variables. The criterion relies on both BBA and plausibilities generating inconsistencies for reestimation formulas. We can thus aim at maximizing an approximation of the support similarly to the Markov chain given by:

$$Q_b^{m,m}(\theta^{(q)}, \theta) = \sum_{t=1}^{T} \sum_{S \subseteq \Omega_z} m_{\gamma,t}^{\Omega_z}(S \mid \mathbf{X}, \mathbb{A}^{(q)}, \Phi^{(q)}) \log m_b^{\Omega_z}(S \mid \mathbf{x}_t, \Phi)$$

(12)

where it is important to remark that m_γ is made dependent not only on the current parameters of the observation model ($\Phi^{(q)}$) but also on the EMC ($\mathbb{A}^{(q)}$). Indeed, the Evidential Forward-Backward algorithm proposed in [15] can compute this quantity, which is related to Eq. 11 by a marginal operation likewise to standard HMM [12, Eq. 38].

The Generalized Bayesian Theorem (GBT) [21] allows to deduce the BBA $m_b^{\Omega_z}(S \mid \mathbf{x}_t, \Phi)$ given plausibilities conditional to singleton $pl^{\Omega_z}(\mathbf{x}_t \mid S_t, \Phi)$ [5]:

$$m_b^{\Omega_z}(S \mid \mathbf{x}_t, \Phi) = \prod_{s_k \in S} pl^{\mathbb{R}}(\mathbf{x}_t \mid s_k, \Phi) \prod_{s_k \notin S} \left(1 - pl^{\mathbb{R}}(\mathbf{x}_t \mid s_k, \Phi)\right)$$

(13)

where $pl^{\mathbb{R}}(\mathbf{x}_t \mid s_k, \theta), \forall s_k \in \Omega_z$ is given by a GMM [12, Sect. 4A]. Making use of Eq. 13, the criterion $\mathcal{Q}_{m,m}^b$ can be rewritten as:

$$\mathcal{Q}_{m,m}^b(\theta^{(q)}, \theta) = \sum_{t=1}^{T} \sum_{s_l \in \Omega_z} \left\{ pl_{\gamma,t}^{\Omega_z}(s_l \mid \theta^{(q)}) \log pl(\mathbf{x}_t \mid s_l, \theta) \right.$$
$$\left. + bel_{\gamma,t}(\overline{s_l} \mid \theta^{(q)}) \log \left(1 - pl(\mathbf{x}_t \mid s_l, \theta)\right) \right\}$$

(14)

where $bel_{\gamma,t}$ is the belief function.

Assumption 2 *The contribution of "$bel_{\gamma,t}(\overline{s_l} \mid \theta^{(q)}) \log \left(1 - pl(\mathbf{x}_t \mid s_l, \theta)\right)$" is negligible compared to $pl_{\gamma,t}^{\Omega_z}(s_l \mid \theta^{(q)}) \log pl(\mathbf{x}_t \mid s_l, \theta)$.*

This assumption does not narrow the expression down to a probabilistic formulation because the weight $pl_{\gamma,t}^{\Omega_z}(s_l \mid \theta^{(q)})$ makes use of the information held by all subsets that contain s_l.

For illustration purpose, we consider one Gaussian component for each singleton state. The criterion can thus be approximated as:

$$\mathcal{Q}_{m,m}^b(\theta^{(q)}, \theta) \approx \sum_{t=1}^{T} \sum_{s_l \in \Omega_z} pl_{\gamma,t}^{\Omega_z}(s_l \mid \theta^{(q)}) \log pl(\mathbf{x}_t \mid s_l, \theta) \tag{15}$$

The means $\mu_k, k = 1 \ldots K$ for the next iteration are obtained by

$$\mu_j^{(q+1)} = \frac{\sum_t pl_{\gamma,t}^{(q)}(s_j)\, \mathbf{x}_t}{\sum_t pl_{\gamma,t}^{(q)}(s_j)} \tag{16}$$

and the covariances by

$$\Sigma_j^{(q+1)} = \frac{\sum_t pl_{\gamma,t}^{(q)}(s_j) \cdot (\mathbf{x}_t - \mu_j)\,(\mathbf{x}_t - \mu_j)'}{\sum_t pl_{\gamma,t}^{(q)}(s_j)} \tag{17}$$

Due to component annealing observed in practice [9], the mixture weights were not considered.

3.4   E-step

$m_{\gamma,t}$ represents the knowledge on subsets of states after observing $\mathbf{X}$ which is obtained by the evidential forward-backward algorithm [15]. This algorithm can be written using commonality functions which allows point-wise multiplication and therefore with limited complexity.

4   Results

4.1   Turbofan engine datasets

The turbofan datasets were generated using the CMAPSS simulation environment that represents an engine model of the 90,000 lb thrust class [7, 19]. The authors used a number of editable input parameters to specify operational profile, closed-loop controllers, environmental conditions (various altitudes and temperatures). Some efficiency parameters were modified to simulate various degradations in different sections of the engine system. Selected fault injection parameters were varied to simulate continuous degradation trends. The datasets generated possess unique characteristics that make them very useful and suitable

for developing classification and prognostics algorithms [16]: Multi-dimensional response from a complex non-linear system, high levels of noise, effects of faults and operational conditions, and plenty of units simulated with high variability.

In the present paper, the 100 training instances of dataset #1 are considered to illustrate the EvHMM on a complex system. These instances were generated by considering one operating condition and one fault mode. The data were collected from various parts of the system to record effects of different degradation mechanisms on 21 sensor measurements. The time-series thus represent different degradation behaviors in multiple units. From sensor measurements in each instance of the training dataset #1, a health indicator is built as proposed in [13]. The health indicators (HI) are depicted in Figure 2 where we can observe high variability in terms of noise and degradation level. For classification purpose, a ground truth of the state sequences corresponding to each time-series was proposed in [18] which is used subsequently, and available at `https://fr.mathworks.com/matlabcentral/ fileexchange/54808-segmentation-of-cmapss-trajectories-into-states`. The comparison between the ground truth and the estimations provided by both HMM and EvHMM is made by the Adjusted Rand Index (ARI) [23] that tends to 1 if the sequence estimated and the ground truth are equal.
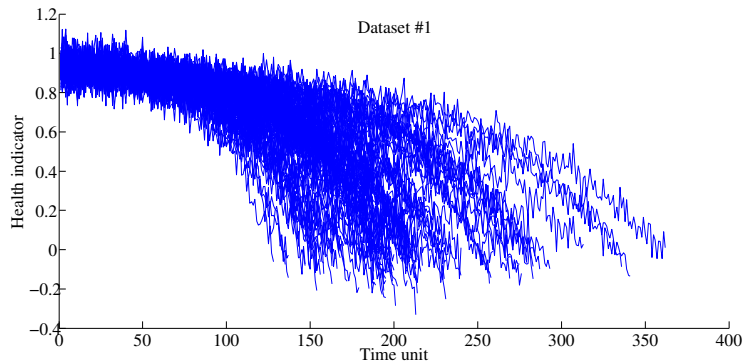
**Fig. 2.** Instances in training dataset #1.

4.2   Influence of a possibly wrong ground truth

The ground truth may be corrupted by errors due, for instance, to the lack of expertise on the degradation, to the noise on the HI or to some parameter tuning. To evaluate the influence of labeling errors, we proceed as proposed in [4, 14] where at each time step t of a training instance, an error probability q_t is drawn randomly from a beta distribution with mean ρ and standard deviation 0.2. Then, with probability q_t, the state y_t is replaced by a completely random value $\tilde{y}_t$ with a uniform distribution over possible states. We thus obtain noisy

labels corresponding to a crisp random labeling. Note that $\rho = 1$ corresponds to the unsupervised case (no labels). The use of prior in HMM was proposed in [14] with available code at `https://fr.mathworks.com/matlabcentral/fileexchange/55172-continuous-partially-hidden-markov-models-with-uncertain-noisy-labels`.

Training and testing sets were then generated to evaluate the EvHMM. 20 instances and the corresponding labels were randomly selected for training, the remaining 80 are kept for testing (without labels), and this process is repeated 8 times. For each run, the labels were corrupted 10 times with the random process explained previously. For each value of ρ, 6400 results were thus obtained which are represented by box plots in Figure 3.
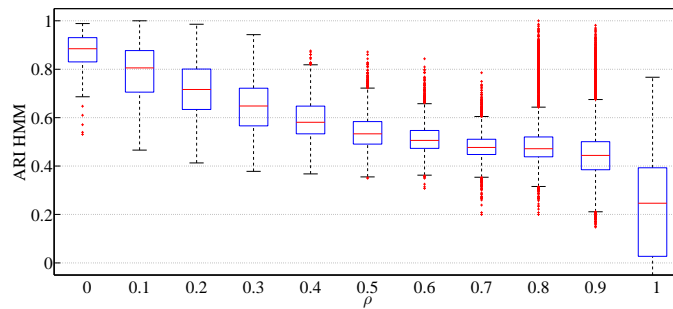
(a) HMM with noisy labels
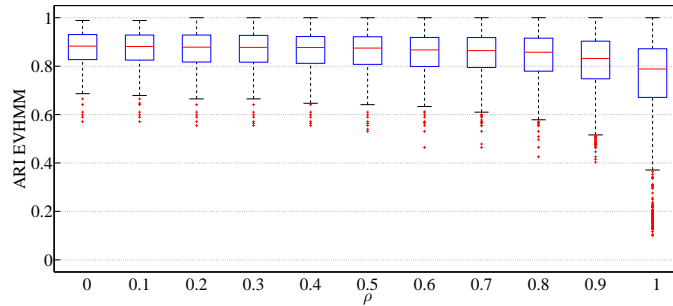
(b) EvHMM with noisy labels

Fig. 3. Performance (ARI) of EvHMM and HMM for the classification of hidden states with respect to the quantity and quality of prior about states (controlled by ρ).

Figure 3(a) and 3(b) depict the evolution of the performance of both HMM and EvHMM when *noisy* labels are considered. It can be observed that the behavior of the EvHMM is highly different from standard HMM. The former indeed appears more robust to label switching compared to the latter, with a

stable performance around 90% until high level of noise ($\rho = 0.8$), whereas the performance of the HMM highly decreases with report to the noise level ρ. For noisy labels, the EvHMM always outperformed the HMM, except for the supervised case with $\rho = 1$ yielding similar results. Note that the low performance of the HMM in this case was also underlined in [14]. The 75-th percentiles of the EvHMM's performance are almost always close to 1 while the boxes for the HMM follows the decreasing trend of the median with report to the noise level. Therefore, with an appropriate initialization (made similarly for both models), the EvHMM may lead to perfect recognition for this dataset.

5   Conclusion and on-going work

EvHMM (Evidential Hidden Markov Model) is a new method for time-series modelling based on Dempster-Shafer's theory of belief functions. The main difference with standard Hidden Markov Models is the consideration of random disjunctive sets in the Markov chain. The use of belief functions to quantify and propagate uncertainty and imprecision on subsets of random latent variables allows to represent the gradual evolutions of a state variables of a system which is monitored through sensors.

Some preliminary results are presented on complex datasets on which the proposed EvHMM depicts high performance compared to standard HMM in presence of noisy labels.

In [17], it is shown that the likelihood (termed as a plausibility) can be exactly computed so that the conjecture and approximations proposed in this learning procedure to make the solution tractable can be easily checked in real applications.

The application of the proposed procedure to various evidential latent models is considered for future work.

Codes will be available on `https://fr.mathworks.com/matlabcentral/profile/authors/7468430-emmanuel-ramasso`.

References

1. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. The annals of mathematical statistics pp. 164–171 (1970)
2. Bishop, C.: Pattern Recognition and Machine Learning. Springer (2006)

3. Cano, P., Ramasso, E.: Ascertainment-adjusted parameter estimation approach to improve robustness against misspecification of health monitoring methods. Mechanical Systems and Signal Processing (2016), submitted, revision 2

4. Côme, E., Oukhellou, L., Denoeux, T., Aknin, P.: Learning from partially supervised data using mixture models and belief functions. Pattern recognition 42(3), 334–348 (2009)

5. Delmotte, F., Smets, P.: Target identification based on the Transferable Belief Model interpretation of Dempster-Shafer model. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 34(4), 457–471 (2004)

6. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39(1), 1–38 (1977)

7. Frederick, D., DeCastro, J., Litt, J.: User's guide for the commercial modular aero-propulsion system simulation (C-MAPSS). Tech. rep., National Aeronautics and Space Administration (NASA), Glenn Research Center, Cleveland, Ohio 44135, USA (2007)

8. Jensen, J.L.W.V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Mathematica 30(1), 175–193 (1906)

9. Naim, I., Gildea, D.: Convergence of the EM algorithm for gaussian mixtures with unbalanced mixing coefficients. In: Langford, J., Pineau, J. (eds.) Proceedings of the 29th International Conference on Machine Learning (ICML-12). pp. 1655–1662. ACM, New York, NY, USA (2012)

10. Patil, G.: Weighted distributions, vol. 4. John Wiley & Sons, Ltd, Chichester (2002), pp. 2369-2377

11. Pieczynski, W.: Multisensor triplet Markov chains and theory of evidence. International Journal of Approximate Reasoning 45(1), 1–16 (2007)

12. Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77, 257–285 (1989)

13. Ramasso, E.: Investigating computational geometry for failure prognostics. Int. Journal on Prognostics and Health Management 5(5), 1–18 (2014)

14. Ramasso, E., Denoeux, T.: Making use of partial knowledge about hidden states in HMMs: an approach based on belief functions. Fuzzy Systems, IEEE Transactions on 22(2), 395–405 (2014)

15. Ramasso, E., Rombaut, M., Pellerin, D.: Forward-backward-viterbi procedures in TBM for state sequence analysis using belief functions. In: ECSQARU. pp. 405–417 (2007)

16. Ramasso, E., Saxena, A.: Performance benchmarking and analysis of prognostic methods for CMAPSS datasets. International Journal on Prognostics and Health Management 5(2), 1–15 (2014)

17. Ramasso, E.: Inference and learning in evidential discrete-latent markov models. IEEE Trans. on Fuzzy Systems (2016), submitted, ver. 2 (Jan. 30, 2016)

18. Ramasso, E.: Segmentation of CMAPSS health indicators into discrete states for sequence-based classification and prediction purposes. Tech. Rep. 6839, FEMTO-ST institute (January 2016)

19. Saxena, A., Goebel, K., Simon, D., Eklund, N.: Damage propagation modeling for aircraft engine run-to-failure simulation. In: IEEE Prognostics and Health Management (2008)

20. Serir, L., Ramasso, E., Zerhouni, N.: Time-sliced temporal evidential networks: The case of evidential HMM with application to dynamical system analysis. In: Prognostics and Health Management (PHM), 2011 IEEE Conference on. pp. 1–10 (June 2011)

21. Smets, P.: Beliefs functions: The disjunctive rule of combination and the generalized Bayesian theorem. IJAR 9, 1–35 (1993)
22. Vannoorenberghe, P., Smets, P.: Partially supervised learning by a credal EM approach. In: Godo, L. (ed.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science, vol. 3571, pp. 956–967. Springer Berlin Heidelberg (2005)
23. Vinh, N., Epps, J., Bailey, J.: Information theoretic measures for clustering comparison: Is a correction for chance necessary? In: Proc. of the 26th Annual Int. Conf. on Machine Learning. pp. 1073–1080 (2009)