



ELSEVIER

Contents lists available at ScienceDirect

## Neurocomputing

journal homepage: [www.elsevier.com/locate/neucom](http://www.elsevier.com/locate/neucom)

## Hidden Markov models with set-valued parameters

Denis Deratani Mauá<sup>a,\*</sup>, Alessandro Antonucci<sup>b</sup>, Cassio Polpo de Campos<sup>c</sup><sup>a</sup> Universidade de São Paulo, São Paulo, Brazil<sup>b</sup> Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Manno, Switzerland<sup>c</sup> Queen's University Belfast, Belfast, UK

## ARTICLE INFO

## Article history:

Received 30 March 2015

Received in revised form

11 July 2015

Accepted 4 August 2015

Available online 5 November 2015

## Keywords:

Hidden Markov models

Imprecise probability

Sensitivity analysis

Credal networks

Robust statistics

## ABSTRACT

Hidden Markov models (HMMs) are widely used probabilistic models of sequential data. As with other probabilistic models, they require the specification of local conditional probability distributions, whose assessment can be too difficult and error-prone, especially when data are scarce or costly to acquire. The imprecise HMM (iHMM) generalizes HMMs by allowing the quantification to be done by sets of, instead of single, probability distributions. iHMMs have the ability to suspend judgment when there is not enough statistical evidence, and can serve as a sensitivity analysis tool for standard non-stationary HMMs. In this paper, we consider iHMMs under the strong independence interpretation, for which we develop efficient inference algorithms to address standard HMM usage such as the computation of likelihoods and most probable explanations, as well as performing filtering and predictive inference. Experiments with real data show that iHMMs produce more reliable inferences without compromising the computational efficiency.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Hidden Markov models (HMMs) are popular probabilistic descriptions of paired sequences of states and observations [1], with applications in speech [1] and text processing [2], activity recognition [3] and computational biology [4], to name but a few. An HMM assumes that the states have been generated by a first-order Markov Chain process, while each observation is generated based only on the paired state. The specification of an HMM comprises an *initial state probability distribution*, which specifies the probability that the process originates in a given state, a *state transition probability distribution*, which specifies the probability that the process will transit from a given state to another, and a *symbol emission probability distribution*, which specifies the probability of observing a symbol conditional on a state.

In many domains, the transitions between consecutive hidden states and the relation between a hidden variable and the corresponding observation are affected by severe uncertainty. This is the case, for instance, when data are scarce [5], observations are missing not-at-random [6], and information is conflicting. In such cases, the use of probability distributions to represent uncertainty might be inadequate and lead to overly confident inferences [5,7,8].

Credal sets [9] are closed and convex sets of probability distributions that allow for a more general representation of uncertainty,

including the situations just described. For instance, complete ignorance about a variable is represented as the credal set of all probability distributions on that variable, instead of the more common representation as a uniform probability distribution. The *imprecise (Multinomial) Dirichlet model* (IDM) learns credal sets from categorical data in a situation of near ignorance, providing a more reliable (although less informative) model of the underlying distribution than the more common Multinomial-Dirichlet model [10].

This paper presents efficient algorithms for inference with *imprecise hidden Markov models* (iHMMs), which allow the specification of a time- and state-discrete HMM with initial state, state transition and symbol emission credal sets in lieu of probability distributions. iHMMs provide a sound way to handle severe uncertainty, with two direct benefits. First, they allow us to suspend judgment when there is insufficient statistical evidence to support a decision [11]. Second, they provide an efficient tool for performing sensitivity analysis [12] in standard non-stationary HMMs, allowing parameters to vary jointly, and in time.

In the rest of the paper, we review the related work (Section 2) and the basics of HMMs (Section 3) and iHMMs (Section 4); we then describe algorithms to deal with common uses such as comparing models according to the data likelihood (Section 5), predicting the current/next state given past observations (Section 6) and finding the most likely hidden state sequence for a given sequence of observations (Section 7). Experiments with speech and action recognition, text completion, and part-of-speech tagging (Section 8) provide evidence that iHMMs are indeed capable of making reliable decisions and evaluating the sensitivity of HMMs to

\* Corresponding author.

E-mail addresses: [denis.maua@usp.br](mailto:denis.maua@usp.br) (D.D. Mauá), [alessandro@idsia.ch](mailto:alessandro@idsia.ch) (A. Antonucci), [c.decampos@qub.ac.uk](mailto:c.decampos@qub.ac.uk) (C.P. de Campos).

the learning sample size. Conclusions and future work are described in Section 9.

## 2. Related work

Bayesian networks are probabilistic models where conditional independences are represented by a graph whose nodes are identified with random variables [13]. HMMs are part of a special class of Bayesian networks (viz. tree-shaped Bayesian networks), one for which efficient inference algorithms are available. As with HMMs, Bayesian networks require uncertainty to be represented by conditional probability distributions. Credal networks [14] extend Bayesian networks to allow uncertainty to be modeled as credal sets. The iHMMs we discuss here are special cases of tree-shaped credal networks.

Drawing inferences from credal networks is a notoriously hard problem. Posterior inference is NP-hard already in tree-shaped credal networks, even when variables take on at most three values [15]; it is NP-hard also in polytree-shaped networks when there is no evidence, even if we allow (provably good) approximate results [16]. A few tractable cases appear in the literature. Fagioli and Zaffalon [17] developed a polynomial-time algorithm for polytree-shaped credal networks with binary variables. Zaffalon and Fagioli [18] described a method to compute unconditional posterior bounds in tree-shaped networks. Mauá et al. [19] proved the existence of a fully polynomial-time approximation scheme for networks of bounded treewidth and bounded variable cardinality.<sup>1</sup>

There has also been intense work on fast approximate algorithms (with no accuracy guarantees). The GL2U algorithm implements a message passing scheme similar to loopy belief propagation in Bayesian networks, that computes upper and lower probabilities in polynomial-time [20]. Other researchers have proposed the use of greedy heuristics [21,22]. Recently, Antonucci et al. [23] developed an approximate method based on linear programming relaxations that was shown to outperform other approximate methods for marginal inference.

The algorithmic techniques discussed in the previous paragraphs deal with the interpretation of imprecision in the parameters known as strong independence. Strong independence, which we adopt in this work, assumes the existence of an ideal probability distribution which we cannot characterize for lack of resources. Epistemic irrelevance (or its symmetrical counterpart epistemic independence) makes no such claim, and allows for the possibility that there might not be any single probability distribution capable of representing our (uncertain) knowledge. De Cooman et al. [24] presented an efficient algorithm for single-query marginal inferences in tree-shaped credal networks under epistemic irrelevance. Their algorithm can be used to efficiently perform filtering (i.e., estimating the marginal probability of the future state given a sequence of observations). Recently, it was shown that filtering on iHMMs provides the same results whether one adopts strong independence or epistemic irrelevance [25,15]. Hence, filtering is also polynomial-time computable under strong independence. We develop later an alternative algorithm for filtering in iHMMs under strong independence. Our algorithm follows more closely the syntax of HMMs and strong independence, and it is arguably easier to understand and implement for a non-expert in imprecise probability models.

De Bock and De Cooman [26] designed an algorithm that computes the maximal joint state sequences of an iHMM under epistemic irrelevance in time polynomial in the input and linear in the

number of maximal sequences. A state sequence is maximal if there is no other state sequence with greater probability under any distribution induced by the model. De Boom et al. [27] devised an analogous algorithm for iHMMs under strong independence with similar time complexity. Finding maximal state sequences is a conservative generalization of the most likely state sequence inference in HMMs. As the number of maximal sequences can be exponential in the number of state variables, their algorithm does not qualify as efficient (i.e., polynomial-time) inference in a strict sense (unless we are satisfied with selecting an arbitrary bounded subset of maximal sequences). In Section 7, we present polynomial-time algorithms for computing unconditional maximin and maximax state sequences; these can be seen as another possible generalization of the most likely state sequence inference in HMMs.

Yet another generalization of the most likely state sequence inference is the computation of E-admissible state sequences. A state sequence is E-admissible if it is a most likely state sequence for at least one distribution induced by an iHMM. Very recently, De Bock et al. [28] developed an algorithm that efficiently decides whether the set of E-admissible state sequences has cardinality strictly greater than one in bounded treewidth models. They showed how this algorithm can be used to measure the sensitivity of MAP inferences to perturbations in the parameters.

The use of credal sets in modeling sequential data is not new. Kozine and Utkin [29] investigated Markov chains with interval-valued transition probabilities. De Cooman et al. [30] used credal sets for sensitivity analysis in Markov chains. Škulj [31,32] defined imprecise Markov chains, and analyzed some basic asymptotic behaviors such as regularity and ergodicity. Crossman et al. [33] studied imprecise Markov chains with absorbing states. Antonucci et al. [34] investigated the use of iHMMs under epistemic irrelevance for tracking tasks. Benavoli et al. [8] defined an iHMM over continuous variables aimed at robust filtering. An imprecise version of the Baum–Welch procedure [1], used to estimate the parameters of an HMM when the state sequence is not observable, was developed by Antonucci et al. [35], and tested on an activity recognition task. Van Camp and De Cooman [36] extended the learning of iHMMs from data to the case of epistemic irrelevance. In [37], the authors designed a method for comparing two iHMMs according to their asymptotic data likelihood, and also applied it on an activity recognition task.

## 3. Hidden Markov models

A Hidden Markov model (HMM) describes a stochastic process over a sequence of *state variables*  $Q_1, \dots, Q_T$  and *manifest variables*  $O_1, \dots, O_T$ . Each state variable  $Q_t$ ,  $t = 1, \dots, T$ , takes values in a finite set  $\mathcal{Q} = \{1, \dots, N\}$ ; each manifest variable  $O_t$  takes value in a finite set  $\mathcal{O} = \{1, \dots, M\}$ .<sup>2</sup> We denote an arbitrary value of state variable  $Q_t$  by  $q_t$ ,  $i$  or  $j$ , and similarly for  $O_t$ . The parameter  $t$  that indexes either family of variables is called *time (step)*; we use a temporal metaphor and refer to the relative indexes of variables by using terms such as past, future and present in the usual way. The stochastic process satisfies the following properties:

- P1: A state variable  $Q_t$  is stochastically independent of all the variables in the past given its immediate predecessor state variable  $Q_{t-1}$ , that is,  $\mathbb{P}(Q_t = q_t | Q_{1:t-1} = q_{1:t-1}, O_{1:t-1} = o_{1:t-1}) = \mathbb{P}(Q_t = q_t | Q_{t-1} = q_{t-1})$ , where the

<sup>1</sup> A fully polynomial-time approximation scheme is a family of algorithms parameterized by a rational  $\epsilon > 0$ , each returning, in time polynomial in the input and in  $1/\epsilon$ , a solution whose value is within a multiplicative factor of  $(1 + \epsilon)$  of the optimum.

<sup>2</sup> The constraint that all state variables or all manifest variables share the same sample space is introduced for simplicity of notation and because they are commonly observed in applications; they can be easily relaxed to generic discrete variables without invalidating any of the results developed in this paper. A further generalization, with some limitations, to the case of continuous manifest variables is also discussed in Section 8.

Download English Version:

<https://daneshyari.com/en/article/408632>

Download Persian Version:

<https://daneshyari.com/article/408632>

[Daneshyari.com](https://daneshyari.com)