EstiHMM: an efficient algorithm for state sequence prediction in imprecise hidden Markov models

Jasper De Bock

Supervisor: prof. dr. ir. Gert De Cooman

Abstract— **We develop an efficient algorithm that calculates the maximal state sequences in an imprecise hidden Markov model by means of coherent lower previsions. Initial results show that this algorithm is able to robustify the inferences made by a classical precise model.**

Keywords— **Imprecise hidden Markov model, coherent lower prevision, epistemic irrelevance, maximal state sequence, Viterbi algorithm.**

I. Introduction

We present an efficient exact algorithm for estimating state sequences from outputs (or observations) in imprecise hidden Markov models (iHMM), where both the uncertainty linking one state to the next, and that linking a state to its output, are represented using coherent lower previsions. The notion of independence we associate with the credal network representing the iHMM is that of epistemic irrelevance. We consider as best estimates for state sequences the (Walley–Sen) maximal sequences [1] for the posterior joint state model (conditioned on the observed output sequence), associated with a gain function that is the indicator of the state sequence.

Our algorithm corresponds to (and generalises) finding the state sequence with the highest posterior probability in HMMs with precise transition and output probabilities, for wich the existing Viterbi-algorithm [2] provides an efficient sollution. However, for imprecise-probabilistic local models, such as coherent lower previsions, we know of no algorithm in the literature for which the computational complexity comes even close to that of Viterbi. We solve this problem by developping an algorithm which has a computational complexity that is at worst quadratic in the length of the Markov chain, cubic in the number of states, and essentially linear in the number of maximal state sequences.

II. Coherent lower previsions

The theory of coherent lower previsions was the most important mathematical tool to develop this algorithm. We refer to [3] for an in-depth study of this theory.

Coherent lower previsions are a way of working with imprecise probabilities. Whereas classical probability theory assumes that a model can be represented by a single probability mass function, the theory of imprecise probabilities works with a sets of possible probability mass functions and therefore allows for imprecision to be modelled. For people who have never heard of this notion, looking at it as a way of robustifying the classical theory is perhaps the most useful interpretation.

Consider a set $\mathscr{M}$ of probability mass functions on $\mathscr{X}$. With each $p \in \mathscr{M}$, we can associate a linear prevision P_p, defined on the set $\mathscr{L}(\mathscr{X})$ of all real-valued maps on $\mathscr{X}$. For every $f \in \mathscr{L}(\mathscr{X})$, $P_p(f)$ is the expected value of f, according to the probability mass function p. We then define the lower prevision $\underline{P}_{\mathscr{M}}$ that corresponds with $\mathscr{M}$ as follows:

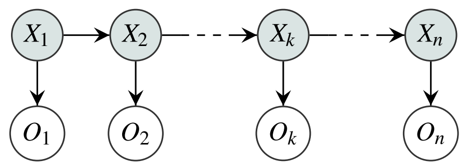
$$\underline{P}_{\mathscr{M}}(f) = \min\left\{P_p(f) \colon p \in \mathscr{M}\right\}.$$

If $\mathscr{M}$ is closed and convex (which makes it a credal set), it can be shown [3] that $\underline{P}_{\mathscr{M}}$ satisfies a set of interesting mathematical properties and we call $\underline{P}_{\mathscr{M}}$ a coherent lower prevision. Futhermore, it is proven in [3] that such coherent lower previsions are completely equivalent with their corresponding credal set. This implies that we can use the theory of coherent lower previsions as a tool for reasoning with closed convex sets of probability mass functions. From now on, we will not indicate the credal set $\mathscr{M}$ anymore and we will simply talk about coherent lower previsions $\underline{P}$, but one should keep in mind that there is always a credal set that corresponds with such a coherent lower prevision.

In a similar way, we can also define conditional lower previsions, which are an extension of the classical conditional expectation functional. More information about these conditional lower previsions can be found in [3].

III. Imprecise hidden Markov Models

An imprecise hidden Markov model can be depicted using the following probabilistic graphical model:



The *state variables* $X_1, \ldots, X_n$ assume values in the respective finite sets $\mathscr{X}_1, \ldots, \mathscr{X}_n$, and the *output variables* $O_1, \ldots, O_n$ assume values in the respective finite sets $\mathscr{O}_1, \ldots, \mathscr{O}_n$. We denote generic values of X_k by x_k or $\hat{x}_k$ and generic values of O_k by o_k. If we talk about complete sequences of variables, we will use the index $1:n$ instead of k. For example, $O_{1:n}$ is the *output sequence*, which takes values $o_{1:n}$ in the set $\mathscr{O}_{1:n}$.

The local uncertainty models associated with the nodes of the network are (conditional) coherent lower previsions, and the independence notion used to interpret the graphical structure is that of epistemic irrelevance [3], [4].

We show how we can use the ideas in [5] (independent natural extension and marginal extension) to construct a most conservative joint model $\underline{P}_1$ from the imprecise local transition and emission models, and derive a number of interesting and useful formulas from that construction.

IV. The problem to be solved

In a hidden Markov model, the states are not directly observable, but the outputs are, and the general aim is to use the outputs to estimate the states. We concentrate on the following problem: *Suppose we have observed the output sequence $o_{1:n}$, estimate the state sequence $x_{1:n}$.*

The first step in our approach consists in updating (or conditioning) the joint model $\underline{P}_1$ on the observed outputs $O_{1:n} = o_{1:n}$. We do this by using the so-called regular extension [3]:

$$\underline{P}_1(f|o_{1:n}) = \max\left\{\mu \in \mathbb{R}\colon \underline{P}_1(\mathbb{I}_{\{o_{1:n}\}}[f-\mu]) \geq 0\right\},$$

for all $f \in \mathscr{L}(\mathscr{X}_{1:n})$.

The next step consists in using the posterior model $\underline{P}_1(\cdot|o_{1:n})$ to find best estimates for the state sequence $x_{1:n}$. On the Bayesian approach, this is usually done by solving a decision-making, or optimisation, problem: we associate a gain function $\mathbb{I}_{\{x_{1:n}\}}$ with every candidate state sequence $x_{1:n}$, and select as best estimates those state sequences $\hat{x}_{1:n}$ that maximise the expected gain, resulting in state sequences with maximal posterior probability.

Here we generalise this decision-making approach towards working with imprecise probability models. The criterion we use to decide which estimates are optimal for the given gain functions is that of (Walley–Sen) *maximality* [1], [3].

We prove (for the specific case of our interpretation of an iHMM) that the collection $\operatorname{opt}(\mathscr{X}_{1:n}|o_{1:n})$ of all optimal state sequences $\hat{x}_{1:n}$ can be defined as follows:

$$\begin{aligned}&\hat{x}_{1:n} \in \operatorname{opt}(\mathscr{X}_{1:n}|o_{1:n})\\ &\qquad\Leftrightarrow (\forall x_{1:n} \in \mathscr{X}_{1:n})\underline{P}_1(\mathbb{I}_{\{o_{1:n}\}}[\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}]) \leq 0.\end{aligned}$$

In summary then, we develop an efficient algorithm for finding the set of maximal estimates $\operatorname{opt}(\mathscr{X}_{1:n}|o_{1:n})$.

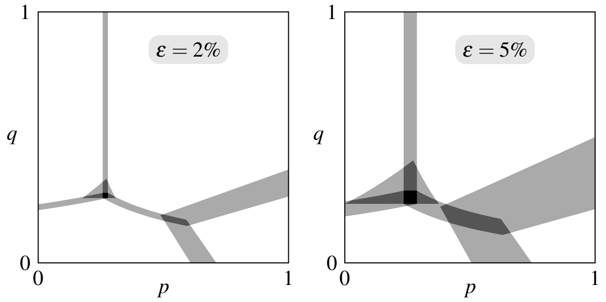
V. The EstiHMM-algorithm

Finding all maximal state sequences in $\operatorname{opt}(\mathscr{X}_{1:n}|o_{1:n})$ seems a daunting task at first: it has a search space that grows exponentially in the length of the Markov chain. However, we have been able to use the basic formulas mentioned in the section on iHMMs to derive an appropriate version of Bellman's [6] Principle of Optimality, which allows for an exponential reduction of the search space. By using a number of additional tricks, we are then able to devise an algorithm for finding all maximal state sequences that is essentially linear in the number of such maximal sequences, quadratic in the length of the chain, and cubic in the number of states. We have given this algorithm the name EstiHMM: Estimation in imprecise Hidden Markov Models.

We perceive the complexity of the EstiHMM algorithm to be comparable to that of the Viterbi algorithm, especially after realising that the latter makes the simplifying step of resolving ties more or less arbitrarily in order to produce only a single optimal state sequence. This is something we will not allow our algorithm to do, as this would completely remove the advantage of robustness our algorithm offers.

VI. Some experiments

Using the EstiHMM algorithm, we have calculated the number of maximal sequences corresponding to a precise model that has been contaminated to make it imprecise. The degree of imprecision is expressed by the parameter ε. In the figures below, the number of maximal sequences is depicted as a function of the precise transition probabilities p and q for two different degrees of imprecision. Darker areas correspond to more maximal sequences. We observe that the areas with more than one maximal sequence enlarge as the imprecision grows and show that they are expanded versions of the lines of indifference that occur in the precise case.



VII. An application

As a first and simple application, we use the EstiHMM algorithm to try and correct mistakes in words that have been processed by Optical Character Recognition software (OCR). We compare our results with those of the Viterbi algorithm and show that our algorithm offers a more robust solution.

For example, the word CHE was wrongfully read as CNE by the OCR software. Using a precise model, the Viterbi algorithm could not correct this mistake, as it suggested that the original correct word was ONE. The EstiHMM algorithm on the other hand, using an imprecise model, concluded that it was not sure enough to offer a single solution and suggested CBE, CHE, CNE, CZE and ONE as possible solutions, thereby including the correct one.

VIII. Conclusions

Interpreting the graphical structure of an imprecise hidden Markov model as a credal network under epistemic irrelevance, leads to an efficient algorithm for finding the maximal state sequences for a given output sequence. Preliminary experiments and a first simple application show that this algorithm extends the well-known Viterbi algorithm in such a way that it allows for imprecision and that it can be used to robustify the inferences made by a classical precise model.

Acknowledgments

The author would like to acknowledge the huge amount of time, help and knowledge that Gert de Cooman, Erik Quaeghebeur and Filip Hermans have offered him during his research.

References

[1] Matthias C. M. Troffaes, "Decision making under uncertainty using imprecise probabilities," *International Journal of Approximate Reasoning*, vol. 45, no. 1, pp. 17–29, January 2007.

[2] Andrew J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[3] Peter Walley, *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.

[4] Inés Couso, Serafín Moral, and Peter Walley, "A survey of concepts of independence for imprecise probabilities," *Risk Decision and Policy*, vol. 5, pp. 165–181, 2000.

[5] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon, "Epistemic irrelevance in credal nets: the case of imprecise Markov trees," *International Journal of Approximate Reasoning*, vol. 51, pp. 1029–1052, 2010.

[6] Richard Bellman, *Dynamic Programming*, Princeton University Press, Princeton, 1957.