State sequence prediction in imprecise hidden Markov models

Jasper De Bock and Gert de Cooman

{Jasper.DeBock,Gert.deCooman}@UGent.be

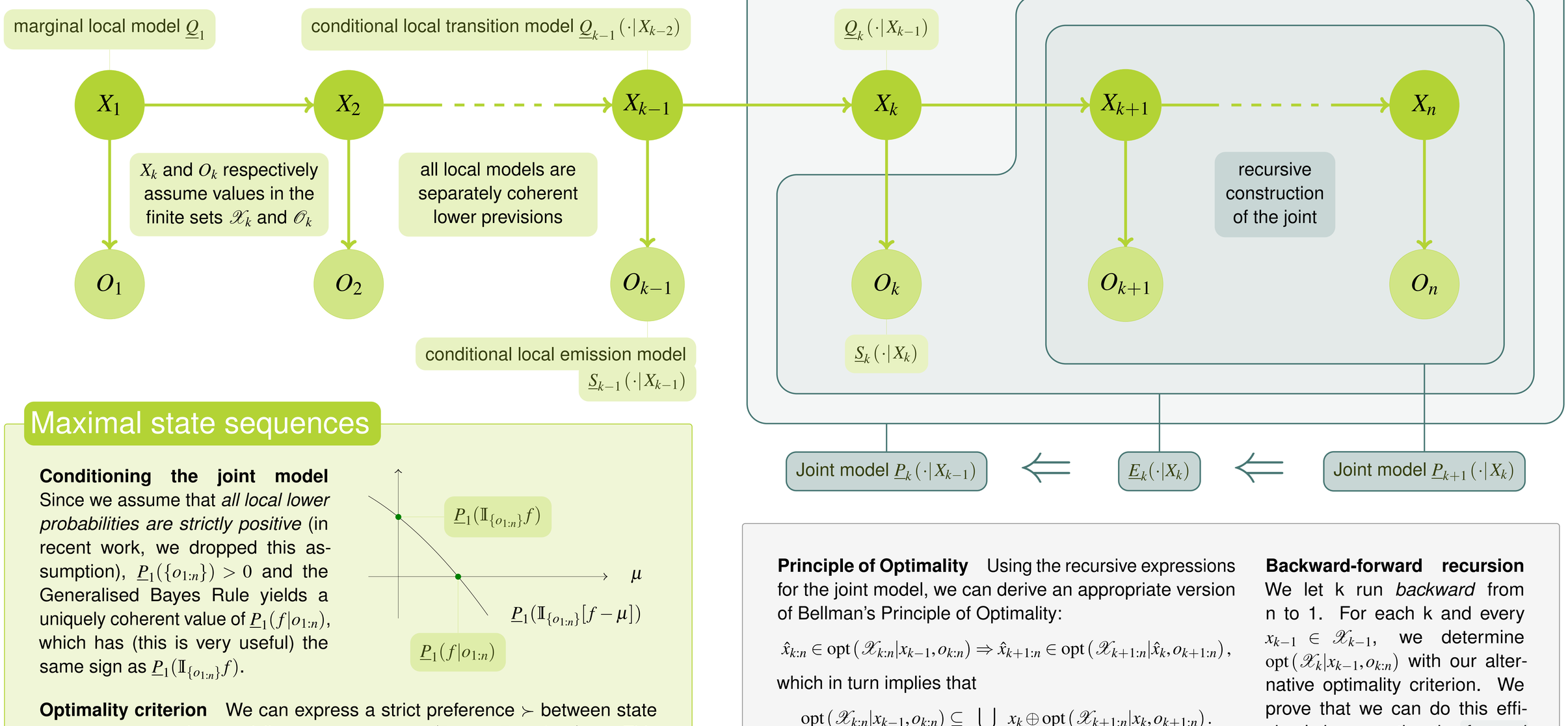
SYSTeMS, Ghent University, Belgium

Interpretation of the graphical model

Interpretation of the graphical structure Our imprecise hidden Markov model (iHMM) represents the following irrelevance assessments: *conditional on its mother variable, the non-parent non-descendants of any variable in the tree are epistemically irrelevant to this variable and its descendants.*

Epistemic irrelevance Y *is irrelevant to* X whenever the belief model (lower prevision $\underline{P}$) about X does not change when we learn something about Y:

$$(\forall g \in \mathscr{L}(\mathscr{X}))(\forall y \in \mathscr{Y})\ \underline{P}(g) = \underline{P}(g|y).$$



Independent natural extension

$\underline{E}_k(\cdot|X_k) := \underline{S}_k(\cdot|X_k) \otimes \underline{P}_{k+1}(\cdot|X_k)$

factorising! $\Rightarrow$ very handy recursive expressions!

Marginal extension

$\underline{P}_k(\cdot|X_{k-1}) := \underline{Q}_k(\underline{E}_k(\cdot|X_k)|X_{k-1})$

Maximal state sequences

Conditioning the joint model Since we assume that *all local lower probabilities are strictly positive* (in recent work, we dropped this assumption), $\underline{P}_1(\{o_{1:n}\}) > 0$ and the Generalised Bayes Rule yields a uniquely coherent value of $\underline{P}_1(f|o_{1:n})$, which has (this is very useful) the same sign as $\underline{P}_1(\mathbb{I}_{\{o_{1:n}\}}f)$.

Optimality criterion We can express a strict preference $\succ$ between state sequences $\hat{x}_{1:n}$ and $x_{1:n}$ as follows: $\hat{x}_{1:n} \succ x_{1:n} \Leftrightarrow \underline{P}_1(\mathbb{I}_{\{\hat{x}_{1:n}\}} - \mathbb{I}_{\{x_{1:n}\}}|o_{1:n}) > 0$. This induces a strict partial order $\succ$ on the set of state sequences $\mathscr{X}_{1:n}$, and we consider a sequence $\hat{x}_{1:n}$ to be optimal when it is undominated, or maximal, in this strict partial order:

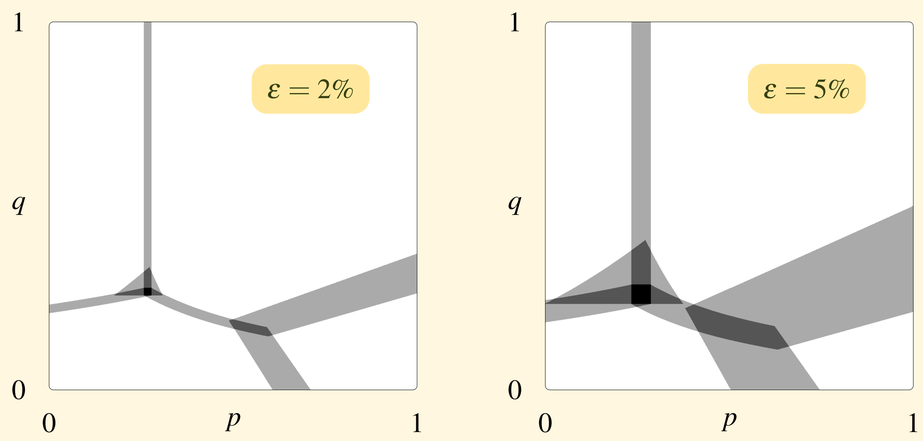
$$\begin{aligned}\hat{x}_{1:n} \in \operatorname{opt}(\mathscr{X}_{1:n}|o_{1:n}) &\Leftrightarrow (\forall x_{1:n} \in \mathscr{X}_{1:n})\ x_{1:n} \not\succ \hat{x}_{1:n} \\ &\Leftrightarrow (\forall x_{1:n} \in \mathscr{X}_{1:n})\ \underline{P}_1(\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}|o_{1:n}) \leq 0 \\ &\Leftrightarrow (\forall x_{1:n} \in \mathscr{X}_{1:n})\ \underline{P}_1(\mathbb{I}_{\{o_{1:n}\}}[\mathbb{I}_{\{x_{1:n}\}} - \mathbb{I}_{\{\hat{x}_{1:n}\}}]) \leq 0.\end{aligned}$$

In an analogous manner we define the optimal subsequences:

$$\hat{x}_{k:n} \in \operatorname{opt}(\mathscr{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow (\forall x_{k:n} \in \mathscr{X}_{k:n})\ \underline{P}_k(\mathbb{I}_{\{o_{k:n}\}}[\mathbb{I}_{\{x_{k:n}\}} - \mathbb{I}_{\{\hat{x}_{k:n}\}}]|x_{k-1}) \leq 0.$$

A first experiment

Motivation and description No algorithm, however cleverly designed, will be able to find all maximal sequences efficiently if there are too many. Because this number of maximal sequences is so important, we study its behaviour in more detail. We consider a binary, stationary iHMM with precise emission models. The imprecise marginal and transition models are generated by mixing precise models with a vacuous one, using a mixture coefficient ε. For a particular observation sequence of length three, we plot the number of maximal sequences as a function of the transition probabilities p and q. As this number grows from 1 to 4 the areas go from white to black.



Results We see that there are large regions of transition probability space where the number of maximal elements remains fairly small. The plots also display quite interesting behaviour. If we let the imprecision grow, by using higher ε, the areas with multiple maximal sequences become larger. They are expanded versions of the lines of indifference that occur in the precise case.

Principle of Optimality Using the recursive expressions for the joint model, we can derive an appropriate version of Bellman's Principle of Optimality:

$$\hat{x}_{k:n} \in \operatorname{opt}(\mathscr{X}_{k:n}|x_{k-1}, o_{k:n}) \Rightarrow \hat{x}_{k+1:n} \in \operatorname{opt}(\mathscr{X}_{k+1:n}|\hat{x}_k, o_{k+1:n}),$$

which in turn implies that

$$\operatorname{opt}(\mathscr{X}_{k:n}|x_{k-1}, o_{k:n}) \subseteq \bigcup_{x_k \in \mathscr{X}_k} x_k \oplus \operatorname{opt}(\mathscr{X}_{k+1:n}|x_k, o_{k+1:n}).$$

An alternative criterion If we limit ourselves to the possible sequences selected by the Principle of Optimality, we can reformulate the optimality criterion as follows:

$$\hat{x}_{k:n} \in \operatorname{opt}(\mathscr{X}_{k:n}|x_{k-1}, o_{k:n}) \Leftrightarrow \alpha_k^{\operatorname{opt}}(\hat{x}_k|x_{k-1}) \leq \alpha_k(\hat{x}_{k:n}).$$

repeat for i from k to n until no:

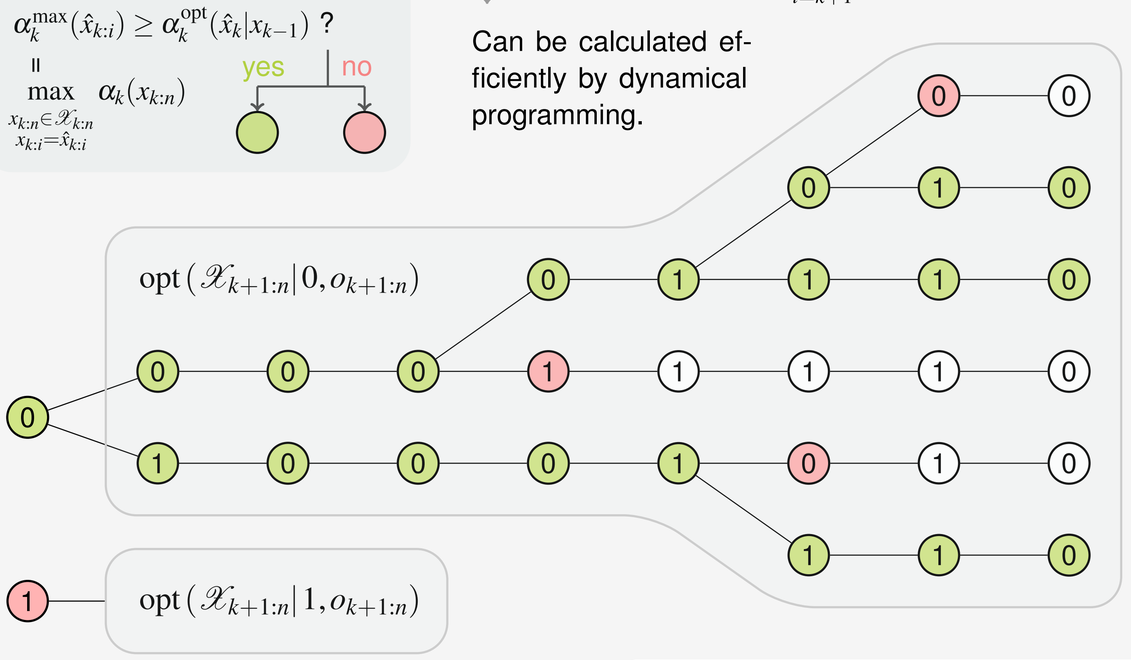
$$\alpha_k^{\max}(\hat{x}_{k:i}) \geq \alpha_k^{\operatorname{opt}}(\hat{x}_k|x_{k-1})\ ?$$

$$\| \quad \max_{\substack{x_{k:n} \in \mathscr{X}_{k:n} \\ x_{k:i} = \hat{x}_{k:i}}} \alpha_k(x_{k:n})$$

yes / no

$$\alpha_k(\hat{x}_{k:n}) = \overline{S}_k(\{o_k\}|\hat{x}_k) \prod_{i=k+1}^{n} \overline{S}_i(\{o_i\}|\hat{x}_i)\overline{Q}_i(\{\hat{x}_i\}|\hat{x}_{i-1})$$

Can be calculated efficiently by dynamical programming.



The EstiHMM algorithm

Computational complexity

Theoretical analysis We prove that the computational complexity is at worst quadratic in the length of the Markov chain, cubic in the number of states, and roughly speaking linear in the number of maximal sequences (each backward step in the backward-forward loop has a linear complexity in the number of maximal elements at that stage).

Empirical confirmation In order to demonstrate that our algorithm is indeed quite efficient, we let it determine the maximal sequences for a random output sequence of length 100. The iHMM we use to determine the maximal sequences is generated by mixing precise local models with a vacuous one, using a mixture coefficient ε. For $\varepsilon = 2\%$, there are 5 maximal sequences and it takes 0.2 seconds to calculate them. If we let ε grow to 5%, we get 764 maximal sequences and these can be determined in 32 seconds. This demonstrates that the complexity is indeed linear in the number of solutions and that the algorithm can efficiently calculate the maximal sequences even for long output sequences.

Backward-forward recursion We let k run *backward* from n to 1. For each k and every $x_{k-1} \in \mathscr{X}_{k-1}$, we determine $\operatorname{opt}(\mathscr{X}_k|x_{k-1}, o_{k:n})$ with our alternative optimality criterion. We prove that we can do this efficiently by executing the *forward running procedure* demonstrated in the figure below. The sequences of green nodes are the elements of $\operatorname{opt}(\mathscr{X}_k|x_{k-1}, o_{k:n})$. In this binary example, $\mathscr{X} = \{0, 1\}$.