Epistemic irrelevance in credal networks: the case of imprecise Markov trees

Gert de Cooman, Filip Hermans, Alessandro Antonucci, Marco Zaffalon

SYSTeMS, Ghent University, Belgium & IDSIA, Switzerland

{Gert.deCooman,Filip.Hermans}@UGent.be & {alessandro,zaffalon}@idsia.ch

UNIVERSITEIT GENT

Basic notions and notations

We consider a rooted and directed discrete *tree* with finite width and depth. With each node s of the tree, there is associated a variable X_s assuming values in a finite non-empty set $\mathscr{X}_s$.

We now add a *local uncertainty model* to each of the nodes:

- a separately coherent conditional lower prevision $\underline{Q}_s(\cdot|X_{\mathrm{pa}(s)})$ on $\mathscr{L}(\mathscr{X}_s)$: for each possible parent value $X_{\mathrm{pa}(s)} = x_{\mathrm{pa}(s)}$, we have a lower prevision $\underline{Q}_s(\cdot|x_{\mathrm{pa}(s)})$.
- a coherent unconditional lower prevision $\underline{Q}_\square$ on $\mathscr{L}(\mathscr{X}_\square)$.

A joint lower prevision will be denoted by $\underline{P}$ instead of $\underline{Q}$. The set of all nodes following s with s included is denoted $\downarrow s$. $\mathrm{pa}(s)$, $\mathrm{ch}(s)$, $\mathrm{sib}(s)$ are respectively the parent, the children and the siblings of node s.

Interpretation of the graphical model

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:

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

Irrelevance is not symmetrical and does not imply d-separation in trees.

Interpretation of the graphical structure Consider any node s, its (single) parent $\mathrm{pa}(s)$ and the set $\bar{s}$ of the non-parent non-descendants of s. Then *conditional on the parent variable $X_{\mathrm{pa}(s)}$, the non-parent non-descendant variables $X_{\bar{s}}$ are assumed to be epistemically irrelevant to the variables $X_{\downarrow s}$ associated with s and its descendants.*

This means that for all $s \in T$, for all $S \subseteq \bar{s}$ and for all $z_{S\cup\mathrm{pa}(s)} \in \mathscr{X}_{S\cup\mathrm{pa}(s)}$:

$$\underline{P}_s(\cdot|z_{\mathrm{pa}(s)}) = \underline{P}_s(\cdot|z_{S\cup\mathrm{pa}(s)}).$$

This makes the tree an *imprecise Markov tree* (IMT).

Recursive construction of the joint Using the interpretation of the graphical structure, and the local belief models $\underline{Q}_s(\cdot|X_{\mathrm{pa}(s)})$, we can construct the most conservative joint lower prevision $\underline{P}$ for all variables in the tree in a recursive fashion, from leaves to root.

Message from node before target node

$$\underline{\pi}_9^\mu := \underline{P}_9(\phi_9^\mu|X_{\mathrm{pa}(9)}) = \underline{Q}_9\left(\psi_9^\mu|X_8\right),$$

where

$$\psi_9^\mu(x) = \begin{cases} \underline{\pi}_{10}^\mu(x) \prod_{s\in\mathrm{sib}(10)} \underline{\pi}_s(x) & \text{if } \underline{\pi}_{10}^\mu(x) \geq 0, \\ \underline{\pi}_{10}^\mu(x) \prod_{s\in\mathrm{sib}(10)} \overline{\pi}_s(x) & \text{if } \underline{\pi}_{10}^\mu(x) < 0. \end{cases}$$

Message from the target node

$$\underline{\pi}_t^\mu := \underline{P}_t(\phi_t^\mu|X_{\mathrm{pa}(t)}) = \underline{Q}_t\left(\psi_t^\mu|X_{10}\right),$$

where

$$\psi_t^\mu(x) = \begin{cases} (g(x)-\mu) \prod_{c\in\mathrm{ch}(t)} \underline{\pi}_c(x) & \text{if } g(x) \geq \mu, \\ (g(x)-\mu) \prod_{c\in\mathrm{ch}(t)} \overline{\pi}_c(x) & \text{if } g(x) < \mu. \end{cases}$$

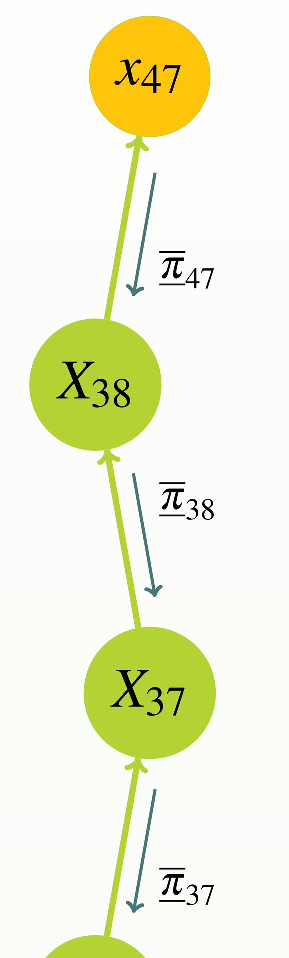
Message from an unobserved node not preceding t

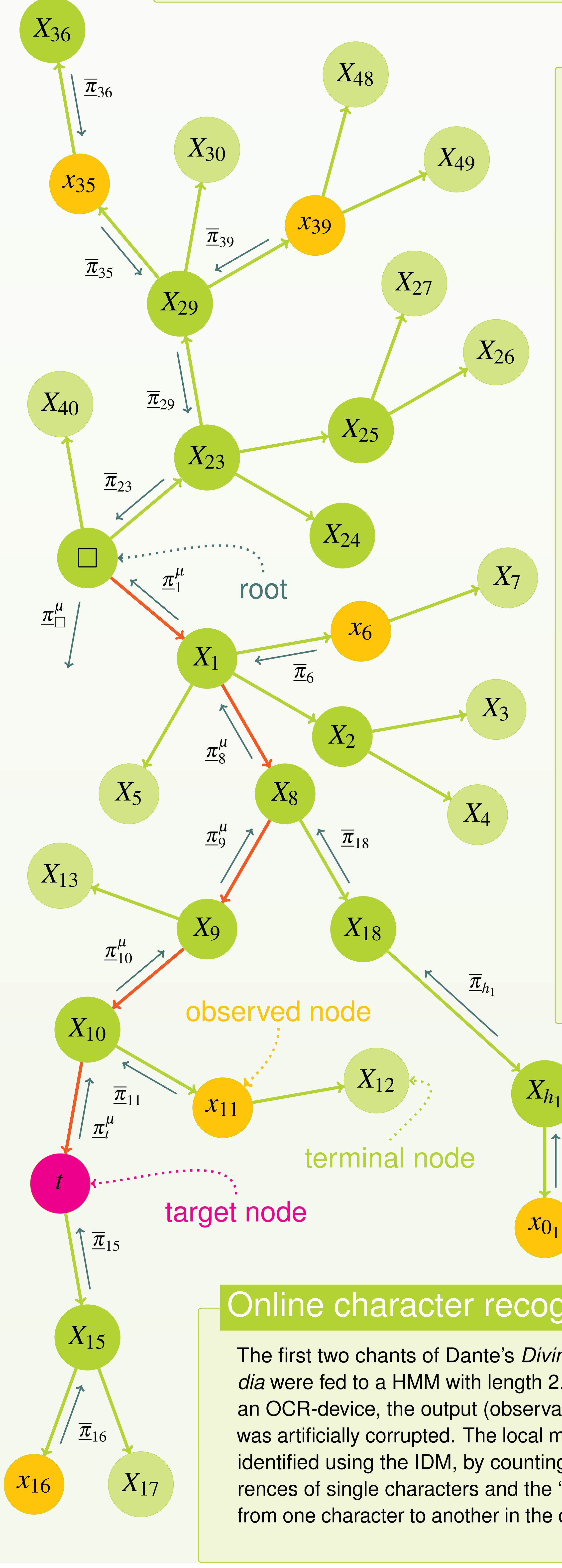
$$\underline{\pi}_{15} := \underline{P}_{15}(\phi_{15}^\mu|X_{\mathrm{pa}(15)}) = \underline{Q}_{15}\Big(\prod_{c\in\mathrm{ch}(15)} \underline{\pi}_c\ |X_t\Big).$$

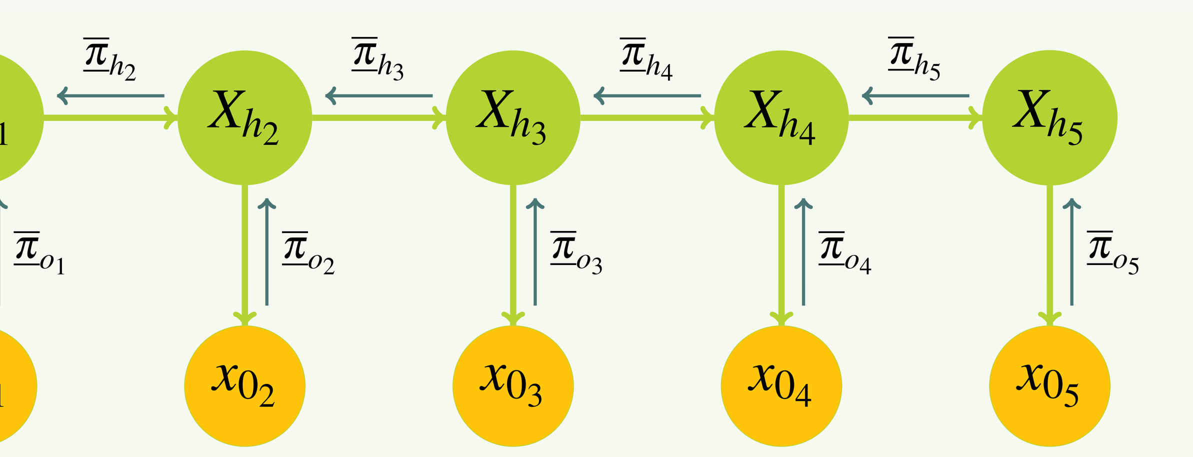
Message from an evidence node not preceding t

$$\underline{\pi}_{16} := \underline{P}_{16}(\phi_{16}^\mu|X_{\mathrm{pa}(16)}) = \underline{Q}_{16}(I_{\{x_{16}\}}\prod_{c\in\mathrm{ch}(16)}\underline{\pi}_c(x_{16})\,|X_{15}),$$

$$= \underline{Q}_{16}(\{x_{16}\}|X_{15})\prod_{c\in\mathrm{ch}(16)}\underline{\pi}_c(x_{15}).$$





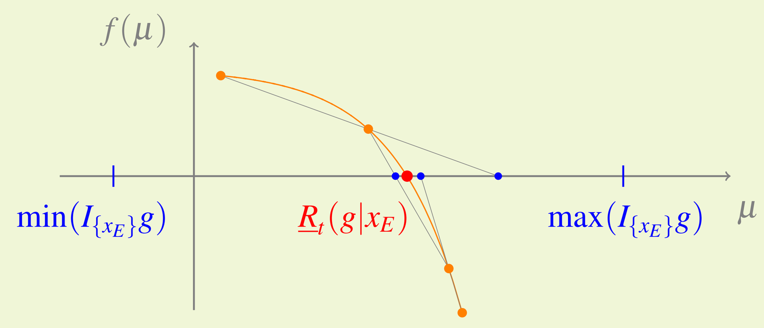


The imprecise Markov tree as an expert system

We are interested in making inferences about the value of the variable X_t in some target node t, when we know the values x_E of the variables X_E in a set E of evidence nodes. Assuming that $\overline{P}(\{x_E\}) > 0$, we can do this by conditioning the joint $\underline{P}$ on the available evidence '$X_E = x_E$':

$$\underline{R}_t(g|x_E) = \max\{\mu \in \mathbb{R} \colon \underline{P}(I_{\{x_E\}}[g-\mu]) \geq 0\}.$$

If we are able to calculate the joint $\underline{P}(I_{\{x_E\}}[g-\mu])$, then we can compute $\underline{R}_t(g|x_E)$ using a bracketing algorithm.



It can be proven that $f(\mu) := \underline{P}(I_{\{x_E\}}[g-\mu])$ is continuous, concave and descending. This speeds up the root-finding algorithm drastically.

For any given node s, define the local gamble g_s^μ

$$g_s^\mu := \begin{cases} I_{x_s} & \text{if } s \in E, \\ g-\mu & \text{if } s = t, \\ 1 & \text{else.} \end{cases}$$

which means that $\underline{P}(I_{\{x_E\}}[g-\mu]) = \underline{P}(\prod_{s\in T} g_s^\mu)$.

If we define ϕ_s^μ by $\phi_s^\mu := \prod_{c\in\downarrow s} g_c^\mu$ then

$$g_\square = I_{\{x_E\}}[g-\mu],$$
$$\phi_s^\mu = g_s^\mu \prod_{c\in\mathrm{ch}(s)} \phi_c^\mu.$$

Now we are able to define the messages $\underline{\pi}_s^\mu$

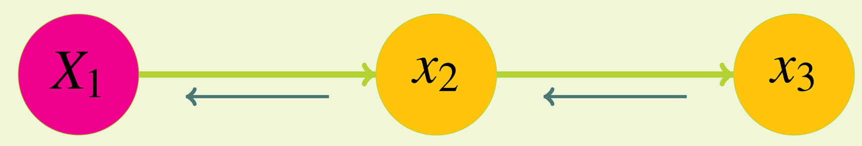
$$\underline{\pi}_s^\mu = \underline{P}(\phi_s^\mu|\mathrm{pa}(s)) \quad \text{and} \quad \overline{\pi}_s^\mu = \overline{P}(\phi_s^\mu|\mathrm{pa}(s)).$$

Thus, treathing the imprecise Markov chain as an expert system means looking for the μ that makes $\underline{\pi}_\square^\mu = \underline{P}(I_{\{x_E\}}[g-\mu])$ equal to zero.

A simple example involving dilation

Consider the following imprecise Markov chain:

$\mathscr{X}_1 = \{a,b\}$	$\mathscr{X}_2 = \{x_2,\ldots\}$	$\mathscr{X}_3 = \{x_3,\ldots\}$
$q(a) := \underline{Q}(a)$	$q(x_2\vert a) := \underline{Q}(\{x_2\}\vert a)$	$\underline{q} := \underline{Q}(\{x_3\}\vert x_2)$
$q(b) := \underline{Q}(b)$	$q(x_2\vert b) := \underline{Q}(\{x_2\}\vert b)$	$\overline{q} := \overline{Q}(\{x_3\}\vert x_2)$



$\underline{\pi}_2 = q(x_2|\cdot)\underline{q}$ $\qquad$ $\underline{\pi}_3 = \underline{Q}(x_3|\cdot)$

If the gamble of interest g is equal to $I_{\{a\}}$ we find after applying the algorithm

$$\underline{\pi}_1^\mu = q(a)\psi_1^\mu(a) + q(b)\psi_1^\mu(b)$$
$$= q(a)(1-\mu)q(x_2|a)\underline{q} - q(b)\mu q(x_2|b)\overline{q}$$

and therefore

$$\underline{r} := \underline{R}_1(\{a\}|x_{\{2,3\}}) = \frac{q(a)q(x_2|a)\underline{q}}{q(a)q(x_2|a)\underline{q} + q(b)q(x_2|b)\overline{q}}$$

$$\overline{r} := \overline{R}_1(\{a\}|x_{\{2,3\}}) = \frac{q(a)q(x_2|a)\overline{q}}{q(a)q(x_2|a)\overline{q} + q(b)q(x_2|b)\underline{q}}.$$

When $\underline{q} = \overline{q}$, which happens for instance if the local model for X_3 is precise, then we see that, with obvious notations,

$$\overline{r} = \underline{r} = \frac{q(a)q(x_2|a)}{q(a)q(x_2|a) + q(b)q(x_2|b)} =: p(a|x_2)$$

and therefore X_2 indeed separates X_3 from X_1. But in general, letting $\alpha := q(a)q(x_2|a)$ and $\beta := q(b)q(x_2|b)$, we get

$$\overline{r} - p(a|x_2) = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\overline{q}+\beta\underline{q}}$$

$$p(a|x_2) - \underline{r} = \frac{\alpha\beta}{\alpha+\beta}\frac{\overline{q}-\underline{q}}{\alpha\underline{q}+\beta\overline{q}}.$$

As soon as $\overline{q} > \underline{q}$, X_2 no longer separates X_3 from X_1, and we witness *dilation* because of the additional observation of X_3!

Online character recognition by imprecise HMMs

The first two chants of Dante's *Divina Commedia* were fed to a HMM with length 2. Mimicking an OCR-device, the output (observation nodes) was artificially corrupted. The local models were identified using the IDM, by counting the occurrences of single characters and the "transitions" from one character to another in the original text.

Accuracy	93.96%	(7275/7743)
Accuracy (if imprecise indeterminate)	64.97%	(243/374)
Determinacy	95.17%	(7369/7743)
Set-accuracy	93.58%	(350/374)
Single accuracy	95.43%	(7032/7369)
Indeterminate output size	2.97	over 21