Belief propagation in imprecise Markov trees

Filip Hermans, Gert de Cooman & Erik Quaeghebeur
SYSTeMS Research Group, EESA Department, Faculty of Engineering, Ghent University
Technologiepark-Zwijnaarde 914, 9052 Zwijnaarde
{Filip.Hermans,Gert.deCooman,Erik.Quaeghebeur}@UGent.be

Basic notions and notation

We consider a rooted and directed discrete tree with finite width and depth:

− we call \( T \) the set of its nodes \( s \)
− we denote the set of parents of node \( s \) by \( P(s) \); if \( P(s) \) is non-empty then \( P(s) = \{ p \} \) with \( p \) the mother of \( s \)
− we denote the set of children of \( s \) by \( C(s) \), and the set of its siblings by \( S(s) \)
− we write \( s \prec r \) if node \( s \) precedes node \( r \)
− \( D \) denotes the set of descendants of \( s \), and \( A(s) = \{ t \in T \mid t \subset s \} \) its set of ancestors; \( s \subset T \) means that \( s \subset T \) and \( s \neq s \)

We also use \( s \sim D(s) \cup \{ s \} \) and \( \pi \sim A(s) \cup \{ s \} \)

With each node \( s \) of the tree, there is associated a variable \( X_s \), assuming values in a finite non-empty set \( \mathcal{X}_s \).

Local uncertainty models

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

• a conditional lower expectation \( \mathcal{G}_s(X_s \mid \pi) \) on \( \mathcal{R}^{\mathcal{X}_s} \); for every possible mother-value \( \pi \), we have a lower expectation \( \mathcal{G}_s(X_s \mid \pi) \)

• an unconditional lower expectation \( \mathcal{G}_s \) on \( \mathcal{R}^{\mathcal{X}_s} \)

We use the common generic notation \( \mathcal{G}_s(X_s \mid \pi) \) for all these local belief models.

Lower expectations

Instead of specifying a single (precise) probability model, an expert expresses his beliefs by giving bounds on this belief model. For example: “the probability of \( A \) is higher than that of \( B \). This leads to a convex closed set of precise probability models, a so-called credal set \( \mathcal{P} \).

Specifying a convex set \( \mathcal{P} \) of probability mass functions \( \rho \) on finite set \( \mathcal{X}_s \) is equivalent to specifying lower and upper expectations, defined for any \( y \in \mathcal{R}^{\mathcal{X}_s} \) by

\[
\mathcal{L}(\rho) = \min \left\{ \sum_{x \in \mathcal{X}_s} \mathcal{G}_s(x \mid \pi) \rho(x) \mid \rho \in \mathcal{P} \right\}
\]

\[
\mathcal{U}(\rho) = \max \left\{ \sum_{x \in \mathcal{X}_s} \mathcal{G}_s(x \mid \pi) \rho(x) \mid \rho \in \mathcal{P} \right\}
\]

Observe the conjugacy relationship:

\[
\mathcal{G}_s(x \mid \pi) = \mathcal{U}(\rho) - \mathcal{L}(\rho)
\]

The real functional \( \mathcal{L} \) is bounded, non-negatively homogeneous and super-additive. There is a one-to-one relationship between credal sets and lower expectation functionals.

Interpretation of the graphical model

Epistemic irrelevance \( T \) is irrelevant to \( X \) whenever the belief model (lower expectation \( \mathcal{G} \)) about \( X \) does not change when we learn something about \( T \):

\[
(\forall y \in \mathcal{R}^{\mathcal{X}} \forall \pi \in \mathcal{P})(\mathcal{L}(y)) = \mathcal{L}(y).
\]

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

Interpretation of the graphical structure

Consider some node \( s \), its mother \( w(s) \) and the set \( \pi = T \backslash \{ w(s) \} \) of the non-parent non-descendants of \( s \). Then conditional on the mother variable \( X_{w(s)} \), the non-parent non-descendant variables \( X_s \) are assumed to be epistemically irrelevant to the variables \( X_{w(s)} \) associated with \( s \) and its descendants.

This means that for all \( s \in T \), for all \( \pi \subset T \) and for all \( \pi \in \mathcal{P} \),

\[
(\forall y \in \mathcal{R}^{\mathcal{X}_s})(\mathcal{G}_{\pi}(y)) = \mathcal{G}_{\pi}(y).
\]

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 \( \mathcal{G}_s(X_s) \), we can construct the most conservative joint lower expectation \( \mathcal{G} \) for all variables in the tree in a recursive fashion, from leaves to root.

Belief updating

We treat the imprecise Markov tree as an expert system, i.e. we are interested in making inferences about the value of the variable \( X \) in some target node \( s \), when we know the values \( x_s \) of the variables \( X_t \) in a set \( \mathcal{E} \subseteq T \) \( \{ s \} \) of evidence nodes.

Assuming that \( \mathcal{G}_s(X_{\mathcal{E}}) > 0 \), we can do this by conditioning the joint \( \mathcal{P} \) on the available evidence \( \{ X_{\mathcal{E}} = x_{\mathcal{E}} \} \).

\[
\mathcal{G}_s(x_s) = \max \{ \mu \in \mathcal{R} \mid \mathcal{G}_s(x_s \mid \mathcal{E}) \geq \mu \}.
\]

We set \( s \) to be the greatest element of the chain \( E \cup \{ t \} \), i.e. the instantiated node closest to \( s \), and let \( s_i \) be its successor in the chain \( t \).

\[
\mathcal{G}_s(x_s) = \max \left\{ \mathcal{G}_s(x_s \mid \pi) + \sum_{s_i \in \pi} \mathcal{G}_s(x_s \mid \{ s_i \}) \right\}.
\]

The messages are defined as

\[
\mathcal{G}_s(x_s) = \max \{ \mu \in \mathcal{R} \mid \mathcal{G}_s(x_s \mid \mathcal{E}) \geq \mu \}.
\]

where

\[
q^E_s \geq q^P \quad \text{and} \quad q^P \geq \sum_{s_i \in \pi} \mathcal{G}_s(x_s \mid \{ s_i \}).
\]

A simple example involving dilation

Consider the following imprecise Markov chain:

Suppose that \( X = \{ a, b \} \), \( \mathcal{G}_s \) is a linear model \( q \), with mass function \( \alpha \) and that \( \mathcal{G}_s(X_1) \) is a linear model \( Q(X_1) \) with conditional mass function \( q(X_1 \mid X_1 - 1) \).

We make no restrictions on the local model \( \mathcal{G}_s(X_1) \).

After applying the belief updating that

\[
\begin{align*}
\mathcal{G}_s(X_1 \mid X_1 - 1) & = \mathcal{G}_s(x_1 \mid x_1 - 1) = \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
& = \mathcal{G}_s(x_1 \mid x_1 - 1) + q(x_1) \\
& = \mathcal{G}_s(x_1 \mid x_1 - 1) + \mathcal{G}_s(x_1) \\
\end{align*}
\]

When \( \alpha = \beta \), which happens for instance if the local model for \( X_1 \) is precise, then we see that, with obvious notations,

\[
\begin{align*}
\mathcal{G}_s(X_1 \mid X_1 - 1) & = \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
& = \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
& = \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
\end{align*}
\]

and therefore \( X_1 \) indeed separates \( X_1 \) from \( X_2 \). But in general, letting \( \alpha = q(x_1 \mid x_1 - 1) \) and \( \beta = q(x_1) \), we get

\[
\mathcal{G}_s(X_1 \mid X_1 - 1) = \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
= \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
= \sum_{x_1} q(x_1 \mid x_1 - 1) q(x_1) \\
\]

As soon as \( \alpha < \beta \), \( X_1 \) no longer separates \( X_1 \) from \( X_2 \), and we witness dilation because of the additional observation of \( X_1 \).

Online character recognition by imprecise HMMs

The first two charts of Dante’s Divina Commedia were fed to a HMM with length 2. Mimicking an OCR-device, the output (observation nodes) were 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% (7252/7743)
Accuracy of imprecise interpolations 64.97% (243/374)
Diversity 98.15% (7380/7743)
Set accuracy 93.58% (390/374)
Single accuracy 93.45% (7012/7749)
Indeterminate output size 3.97 over 21