Constraint-Based Inference in Probabilistic Logic Programs

Arun Nampally, Timothy Zhang, C. R. Ramakrishnan
Department of Computer Science, Stony Brook University,
Stony Brook, NY 11794
{anampally,thzhang,cram}@cs.stonybrook.edu

May 2, 2018

A Proofs

Proposition (Closure properties)**.** *OSDDs are closed under conjunction and disjunction operations.*

Proof. Let $\psi = (s,k,Y)[\gamma_i : \psi_i]$ and $\psi' = (s',k',Y')[\gamma'_j : \psi'_j]$ be two OSDDs.

Let $\oplus$ denote either $\wedge$ or $\vee$, then by the definition of $\psi \oplus \psi'$ ordering is preserved. Depending on the ordering of the OSDDs, $\psi \oplus \psi'$ has three cases. If $(s,k) \prec (s',k')$ (resp. $(s',k') \prec (s,k)$ then $\psi \oplus \psi'$ is constructed by leaving the root and edge lables intact at (s,k,Y) (resp. (s',k',Y')). In this case urgency, mutual exclusion, and completeness are all preserved since ψ (resp. ψ') is an OSDD and the root and its edge labels are unchanged.

If $(s,k) = (s',k')$ urgency is preserved since $\forall i \forall j\ \gamma_i \wedge \gamma'_j$ are the constructed edges of $\psi \oplus \psi'$ and individually these γ_i and γ'_j satisfied urgency. If we take two distinct edge constraints $\gamma_i \wedge \gamma'_j$ and $\gamma_k \wedge \gamma_l$ it is the case that $[\![\gamma_i \wedge \gamma'_j \wedge \gamma_k \wedge \gamma'_l]\!] = \emptyset$ since either $i \neq k$ or $j \neq l$ and both $[\![\gamma_i \wedge \gamma_k]\!] = \emptyset$ and $[\![\gamma'_j \wedge \gamma'_l]\!] = \emptyset$. Let σ be the grounding substitution of $\cup_{i,j} Vars(\gamma_i \wedge \gamma'_j) \setminus \{Y\}$ that is compatible with constraint formula labeling the path to the node (s,k,Y). To prove completeness, we note that $\cup_j [\![\gamma_i\sigma \wedge \gamma'_j\sigma]\!]_Y = [\![\gamma_i\sigma]\!]_Y$. Therefore, $\cup_{i,j}[\![(\gamma_i \wedge \gamma'_j)\sigma]\!] = type(Y)$. □

Proposition. *Let $\psi = (s,k,Y)[\gamma_i : \psi_i]$ and $\psi' = (s',k',Y')[\gamma'_j : \psi'_j]$ be two OSDDs, then*

$$\mathscr{G}(\psi \oplus \psi') = \mathscr{G}(\psi) \oplus \mathscr{G}(\psi').$$

Proof. When $(s,k) \prec (s',k')$, then $\mathscr{G}(\psi) \oplus \mathscr{G}(\psi') = (s,k,Y)[\alpha_r : \mathscr{G}(\psi_r[\alpha_r/Y]) \oplus \mathscr{G}(\psi')]$. But $\mathscr{G}(\psi \oplus \psi') = \mathscr{G}((s,k,Y)[\gamma_i : \psi_i \oplus \psi']) = (s,k,Y)[\alpha_r : \mathscr{G}(\psi_r \oplus \psi'[\alpha_r/Y])]$.

Thus, we consider the case where $(s,k) = (s',k')$. Both ground explanation graphs have the same root, therefore the ground explanations in $\mathscr{G}(\psi) \oplus \mathscr{G}(\psi')$ are obtained by combining subtrees connected which have the same edge label. Given grounding substitution σ on $\cup_{i,j} Vars(\gamma_i \wedge \gamma'_j) \setminus \{Y\}$ that is compatible with the constraint formula labeling the path from root to the node under consideration, if some value $\alpha \in type(Y)$ is such that it satisfies $\gamma_i\sigma$ and $\gamma'_j\sigma$ for specific i, j, then in $\psi \oplus \psi'$, $\alpha \in [\![(\gamma_i \wedge \gamma_j)\sigma]\!]_Y$, therefore the same subtrees are combined. $\square$

Proposition (Condition for Measurability)**.** *A satisfiable constraint formula is measurable w.r.t all of its variables if and only if it saturated.*

Proof. First we prove that saturation is a sufficient condition for measurability.

The proof is by induction on the number of variables in γ. When $|Vars(\gamma)| = 1$ the proposition holds since the only satisfiable constraint formulas with a single variable are $\{X = c\}$ for some $c \in Dom(X)$ or formulas of the form $\{X \neq c_1, X \neq c_2, \ldots, X \neq c_m\}$ for some distinct set of values $\{c_1, \ldots, c_m\} \subset Dom(X)$. Clearly the formulas are measurable w.r.t X.

Assume that the proposition holds for saturated constraint formulas with n variables. Now consider a satisfiable constraint formula γ with $n+1$ variables which is saturated. Let $X \in Vars(\gamma)$. Consider the graph obtained by removing X and all edges incident on X from the constraint graph of γ. It represents a saturated constraint formula γ' with n variables. This is because for any three variables A, B, C distinct from X, if $A = B, B = C$ then A, C are connected by an "$=$" edge. Similarly, if $A = B, B \neq C$, then A, C are connected by an "$\neq$" edge. Further for any variable A other than X, if Z is the set of variables connected to A by "$\neq$" edges, then there exists edges between each pair of these nodes. This is due to the definition of saturation which is satisfied by γ.

But, by inductive hypothesis γ' is measurable w.r.t each of its variables. Now consider computing the measure of X in γ. If X is connected to any node Y with an "$=$" edge, then measure of X is 1. If X is not connected to any node with an "$=$" edge, then it is either disconnected from other nodes or connected to them by only "$\neq$" edges. In either case m_X is computed by subtracting the number of nodes connected to X by "$\neq$" edges from the domain.

To prove that saturation is a necessary condition we use proof by contradiction. Assume there exists a measurable constraint formula γ which is not saturated. Then there exists a variable $X \in Vars(\gamma)$ and a set $\mathscr{Z}$ which is the set of nodes connected to the node for X by "$\neq$" edge and for some pair of elements $A, B \in \mathscr{Z}$, there is no edge between them. Since we take closure of "$=$" edges, we can assume that $\gamma \not\models A = B$. So there must exist two substitutions σ, σ' where $A = B$ and $A \neq B$ respectively. The number of solutions of X under these two substitutions is clearly different, which is a contradiction. $\square$