Expressivity Analysis for PL-Languages

Manfred Jaeger
Aalborg University

Kristian Kersing, Luc De Raedt
Freiburg University
The Problem
“Alphabet soup” (L.Getoor): Prism, SLP, RBN, PRM, BLP, MLN, Blog, …

Questions:
- Where are these languages similar?
- Where are these languages different?
- What are the particular strengths/weaknesses of language XYZ?

First issue to investigate:
- What is the expressive power of the different languages?

Later:
- What is the complexity of inference?
- What is the complexity of learning?
Elements of a Solution

- Goal: establish general framework with re-usable components for expressivity analysis
- Find common semantic ground
- Consider translations of (syntactic) models and embeddings of their semantics.
- A language \( L' \) is at least as expressive as a language \( L \), if each \( L \)-model \( M \) can be translated into an \( L' \)-model \( M' \), so that the semantics of \( M' \) “contains” the semantics of \( M \).

![Diagram](attachment:image.png)
Common Semantic Ground: Multi-valued Herbrand Interpretations

PL-languages define distributions for random variables that can be written as ground atoms:

\begin{align*}
  \text{blood\_pressure}(\text{tom}) & \quad \text{sister}(\text{susan}, \text{tom}) & \quad \text{genotype}(\text{mother}(\text{paul})) \\
  \text{blood\_pressure}(\text{susan}) & \quad \text{sister}(\text{susan}, \text{paul}) & \quad \text{genotype}(\text{father}(\text{paul})) \\
  \cdots & \quad \cdots & \quad \cdots
\end{align*}

With each relation symbol is associated a (finite) state space:

\begin{align*}
  \text{states(blood\_pressure)} &= \{\text{high, normal, low}\} \\
  \text{states(sister)} &= \{\text{true, false}\} \\
  \text{states(genotype)} &= \{\text{AA, Aa, aa}\}
\end{align*}

*Herbrand Interpretation*: assignment of a truth value to all ground atoms constructible from a vocabulary $S$ of relation, function, and constant symbols.

*Multi-valued Herbrand Interpretation*: assignment of a state to all ground atoms constructible from a vocabulary $S$ of relation, function, and constant symbols.

*PL-model*: defines a probability distribution over all Multi-valued Herbrand Interpretations for a given vocabulary $S$. 
Any PL-model can be represented by an ordinary Bayesian network. Are PL-languages just shorthand notations for large Bayesian networks?
Modularity of Representations

The power and usefulness of PL-languages derives from the fact that they split the specification of a complex model into a generic (intensional) and a domain-specific (extensional) part:

*General Genetic Linkage Model*

*Input Pedigree* (can be represented as a Bayesian network)
A (preliminary) analysis of several languages:

<table>
<thead>
<tr>
<th></th>
<th>Intensional</th>
<th>Extensional</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBN</td>
<td>rbn</td>
<td>Input Structure</td>
</tr>
<tr>
<td>PRM</td>
<td>prm</td>
<td>Skeleton Structure</td>
</tr>
<tr>
<td>BLP</td>
<td>intensional part</td>
<td>extensional part</td>
</tr>
<tr>
<td>MLN</td>
<td>mln</td>
<td>constants</td>
</tr>
<tr>
<td>Prism</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ground atoms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>without</td>
</tr>
<tr>
<td></td>
<td></td>
<td>msw’s in SLD tree</td>
</tr>
</tbody>
</table>

Updated plan:

![Diagram of embedding process]
Embeddings

$P$: probability distributions over $MVHI(S)$
$P'$: probability distributions over $MVHI(S')$

An embedding of $P$ in $P'$ is a mapping

$$h : MVHI(S) \rightarrow 2^{MVHI(S')}$$

such that for all $w, w' \in MVHI(S)$:

$$P(w) = P'(h(w)) \text{ and } h(w) \cap h(w') = \emptyset$$

Write $P \preceq P'$ if there is such an embedding.
If $P \preceq P'$, then every probabilistic query about $P$ can be answered from the model $P'$ (one can consider weaker forms of embeddings, so that only restricted types of queries for $P$ are supported by $P'$).
Putting Everything Together…

Language $L'$ is at least as expressive as $L$, $L \preceq L'$, if

$$\exists t_{int} \forall M_{int} \exists t_{ext} \forall M_{ext} \quad P(M_{int}, M_{ext}) \preceq P(t_{int}(M_{int}), t_{ext}(M_{ext}))$$

Example Result

$$MLN \preceq RBN$$  \hspace{1cm} (precisely:$MLN \preceq_c RBN$)