Differentiable Learning of Logical Rules for Knowledge Base Reasoning

Fan Yang, Zhilin Yang, William W. Cohen
Carnegie Mellon University
{fanyang1,zhiliny,wcohen}@cs.cmu.edu

Motivation
First-order logical rules are useful for knowledge base reasoning.
- Interpretable
- Transferrable to unseen entities.

Learning probabilistic logical rules is difficult -- it requires learning
- the discrete structure, i.e. the particular set of rules to include, and
- the continuous parameters, i.e. confidence associated with each rule.

Our Approach
An end-to-end differentiable framework --
Neural Logic Programming (Neural LP).

Learning -- Objective function
Learn weighted chain-like logical rules to reason over the knowledge base.
\[
\alpha \text{ query } (Y, X) \leftarrow R_n (Y, Z_n) \land \cdots \land R_1 (Z_1, X)
\]

Using TensorLog operators, the objective is:
\[
\max_{\{\alpha_i, \beta_i\}} \sum_{(y, x)} \text{score}(y \mid x) = \max_{\{\alpha_i, \beta_i\}} \sum_{(y, x)} v_y^T \left( \sum_i (\alpha_i \Pi_{\beta_i} M_i v_x) \right)
\]
- \(l\) indexes over all possible rules
- \(\alpha_i\) is the confidence of the rule
- \(\beta_i\) is an ordered list of all relations in the rule

Learning -- Recurrent formulation
An equivalent but recurrent formulation to allow end-to-end differentiable optimization.

TensorLog operators
E = set of entities. R = set of binary relations.
- For each entity \(i\), define \(v_i\) in \(\{0, 1\}^{\mid E\mid}\).
- For each relation \(R\), define \(M_{R_i}\) in \(\{0, 1\}^{\mid E\mid \times \mid E\mid}\) where the \((i, j)\) entry is 1 if and only if \(R(i, j)\).

Key idea
A neural controller that learns to compose TensorLog operators.

Experiments
Statistical relational learning

<table>
<thead>
<tr>
<th></th>
<th>ISG</th>
<th>Neural LP</th>
</tr>
</thead>
<tbody>
<tr>
<td>T=2</td>
<td>43.5</td>
<td>92.0</td>
</tr>
<tr>
<td>T=3</td>
<td>43.3</td>
<td>90.2</td>
</tr>
</tbody>
</table>

WikiMovies with natural language queries.

<table>
<thead>
<tr>
<th>Knowledge base</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>directed_by</td>
<td>Who is the writer of the film Blade Runner?</td>
</tr>
<tr>
<td>written_by</td>
<td>What year was the movie Blade Runner released?</td>
</tr>
</tbody>
</table>

Key-Value Memory Network

Neural LP

Knowledge base completion
- Example of learned rules

<table>
<thead>
<tr>
<th></th>
<th>WN18</th>
<th>FB15K</th>
<th>FB15KSelected</th>
</tr>
</thead>
<tbody>
<tr>
<td>TransE</td>
<td>0.01</td>
<td>0.48</td>
<td>0.53</td>
</tr>
<tr>
<td>Neural LP</td>
<td>94.49</td>
<td>73.28</td>
<td>27.97</td>
</tr>
</tbody>
</table>

- Inductive and transductive settings