Abstract

In order to use this paper, we developed an evolutionary algorithm that, given a WIKIPEDIA graph, finds the best patterns between two target entities. Given a training set of WIKIPEDIA nodes, our algorithm learns a new classification model, which can be used to predict target entities and new properties.

In our demo paper, we present a high-level overview of our graph pattern learner and show its application to enable fusion of WIKIPEDIA datasets.

Introduction:

Task:
• Simulate human associations
• For a given source node predict target nodes as humans would

Approach:
• Training Data:
• Node Pairs (associated)
• SPARQL Endpoint
• Graph Pattern Learner
• Learns SPARQL Queries
• In ensemble can predict training target for corresponding training source
• Apply trained model to user inputs

Graph Pattern Learner:
• Evolutionary Algorithm
• Individuals: SPARQL BGPs
• Fitness: “Good Patterns”
• Mutations
• Mating: Combining Triples (randomly)
• InC: Random length paths
• Coverage by restarting training runs with remaining pairs
• Clustering of learned patterns via training pair fulfillment (ASK queries)
• Fusion training on target candidate vectors w.r.t learned patterns

Evaluation:

<table>
<thead>
<tr>
<th>Method</th>
<th>MP@25</th>
<th>MP@50</th>
<th>MP@75</th>
<th>NDCG@25</th>
<th>NDCG@50</th>
<th>NDCG@75</th>
<th>MAP</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle</td>
<td>65.4%</td>
<td>75.6%</td>
<td>85.8%</td>
<td>21.9%</td>
<td>27.4%</td>
<td>33.0%</td>
<td>10.1%</td>
<td>85.8%</td>
</tr>
<tr>
<td>REFSol</td>
<td>59.6%</td>
<td>72.6%</td>
<td>77.3%</td>
<td>16.0%</td>
<td>20.4%</td>
<td>21.9%</td>
<td>8.4%</td>
<td>76.9%</td>
</tr>
<tr>
<td>DBpedia</td>
<td>49.4%</td>
<td>62.5%</td>
<td>67.7%</td>
<td>12.6%</td>
<td>15.9%</td>
<td>16.6%</td>
<td>6.8%</td>
<td>69.8%</td>
</tr>
<tr>
<td>NLB-VLK &amp; PL</td>
<td>32.0%</td>
<td>43.5%</td>
<td>47.1%</td>
<td>6.8%</td>
<td>9.1%</td>
<td>9.6%</td>
<td>3.3%</td>
<td>49.2%</td>
</tr>
</tbody>
</table>

Graph Pattern Learner Mutations:

Learned Patterns form a Feature Space

• Training pairs and patterns form a boolean vector-space w.r.t. SPARQL/ASK queries
• We can use this to cluster similar patterns

Evaluation:

<table>
<thead>
<tr>
<th>Method</th>
<th>MP@25</th>
<th>MP@50</th>
<th>MP@75</th>
<th>NDCG@25</th>
<th>NDCG@50</th>
<th>NDCG@75</th>
<th>MAP</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oracle</td>
<td>65.4%</td>
<td>75.6%</td>
<td>85.8%</td>
<td>21.9%</td>
<td>27.4%</td>
<td>33.0%</td>
<td>10.1%</td>
<td>85.8%</td>
</tr>
<tr>
<td>REFSol</td>
<td>59.6%</td>
<td>72.6%</td>
<td>77.3%</td>
<td>16.0%</td>
<td>20.4%</td>
<td>21.9%</td>
<td>8.4%</td>
<td>76.9%</td>
</tr>
<tr>
<td>DBpedia</td>
<td>49.4%</td>
<td>62.5%</td>
<td>67.7%</td>
<td>12.6%</td>
<td>15.9%</td>
<td>16.6%</td>
<td>6.8%</td>
<td>69.8%</td>
</tr>
<tr>
<td>NLB-VLK &amp; PL</td>
<td>32.0%</td>
<td>43.5%</td>
<td>47.1%</td>
<td>6.8%</td>
<td>9.1%</td>
<td>9.6%</td>
<td>3.3%</td>
<td>49.2%</td>
</tr>
</tbody>
</table>

Examples of Learned Graph Patterns in the Trained Model:

More Info & Contact:
w3id.org/associations
joern.hees@kit.edu