Searching Patterns for Relation Extraction over the Web: Rediscovering the Pattern-Relation Duality

Yuan Fang1,2 Kevin Chen-Chuan Chang1,2
1 University of Illinois at Urbana-Champaign, United States
2 Advanced Digital Sciences Center, Singapore

Key Finding: Pattern-Relation Duality (PR Duality)

Original Intuition [Brin’98]

Tuples and patterns for a desired relation R can be qualified by the metrics of precision and recall, both of which are propagated between matching patterns and tuples—essentially such propagations correspond to random walks on a graph of interrelated patterns and tuples: recall is a forward walk from R to tuples and patterns, and precision is a backward walk from tuples and patterns to R.

How to Qualify Patterns and Tuples?

- Probabilistic Inferences between tuples and patterns
- Through contexts (acting as bridges)

1) QuestP: Quest Backward for Precision Inference
   \[ P(t) = \frac{\sum_{p \in \mathcal{P}(t)} P(p)}{\sum_{p \in \mathcal{P}(t)} P(p)} \]
   if \( t \in \mathcal{T}_0 \);
   otherwise.

2) QuestF: Quest Forward for Recall Inference
   \[ R(t) = \frac{\sum_{p \in \mathcal{P}(t)} R(p)}{\sum_{p \in \mathcal{P}(t)} R(p)} \]

- The inferences can be interpreted as random walks

How to Propagate the Metrics?

- Extracting three target relations on the Web
- Baselines: QXtract and Snowball (Q&S)
- Three different schemes of PRDualRank:
  - Dual-Ext (scoring tuples with extraction patterns only)
  - Dual-Sch (scoring tuples with search patterns only)
  - Dual-Combine (average of the above two)

Experiments: Our Results

- Our Results

The dual problem is: How to rank patterns? How to rank tuples using the patterns?

Motivation: Use syntactic patterns to extract tuples.

- Need to interrelate them for mutual reinforcement
- Patterns and tuples co-occur in text fragments
- Co-occurring \( p \) and \( t \) form a context \( c = (t, p) \)
- A context is a particular “interpretation”
  - Whether the pair \( (t, p) \) is relevant or irrelevant
- Contexts thus interrelate tuples and patterns into a Context Graph
  - Which is an affinity graph of semantic relationships

Example Context Graph \( G = (T, C, P) \)

<table>
<thead>
<tr>
<th>Tuples T</th>
<th>Context C</th>
<th>Patterns P</th>
</tr>
</thead>
<tbody>
<tr>
<td>I_1</td>
<td>C_1</td>
<td>P_1</td>
</tr>
<tr>
<td>I_2</td>
<td>C_2</td>
<td>P_2</td>
</tr>
<tr>
<td>I_3</td>
<td>C_3</td>
<td>P_3</td>
</tr>
<tr>
<td>I_4</td>
<td>C_4</td>
<td>P_4</td>
</tr>
</tbody>
</table>

- \( t_1 = \) (Beijing, China)
- \( t_2 = \) (Shanghai, China)
- \( p_1 = \) #city is ... capital of #country
- \( p_2 = \) #city is an ancient city in #country
- \( t_1 \) and \( p_1 \) co-occur as \( c_1 \) in text fragment 
  - “Beijing is an ancient city in China,…” for 4 times.
- \( t_1 \) and \( p_2 \) co-occur as \( c_2 \) in text fragment 
  - “Beijing, a big city in China,…” for 5 times.

How to Interrelate Patterns and Tuples?

- Need to interrelate them for mutual reinforcement
- Patterns and tuples co-occur in text fragments
- Co-occurring \( p \) and \( t \) form a context \( c = (t, p) \)

How to Qualify Patterns and Tuples?

- Probabilistic Inferences between tuples and patterns
- Deterministic precision and recall: 
  \[ P(p) = \frac{|C_R \cap I_p|}{|I_p|} \]
  \[ R(p) = \frac{|C_R \cap I_p|}{|C_R|} \]

Which is an affinity graph of semantic relationships