Multi-Hop Knowledge Graph Reasoning with Reward Shaping

Authors: Xi Victoria Lin, Richard Socher, Caiming Xiong
Presented by: Arka Sadhu

April 1, 2019
Given a knowledge graph (KG), we want to build a QA system.

Q: Collaborators of Tom Hanks?
A: Steven Spielberg, Phyllida Lloyd.

Here we focus only on Structured QA. So the query would be (Tom Hanks, collaborators, ?).
1. Knowledge graph Embedding
   - Encode nodes into an embedding space, score the triplet (source, relation, target) [ConvE]
   - Highly Efficient and Accurate
   - Not interpretable: Can’t ask why the triplet is correct?

2. Path Based Multi-hop Reasoning
   - Sequential reasoning over nodes and edges [MINERVA]
   - Use RL to learn the decision making process
   - Highly interpretable
   - Significant performance drop: Mean Reciprocal Rank (MRR) drops from 95.7% (ConvE) to 82.5% (Minerva) on UMLS benchmark dataset
The problem is formulated as a Markov Decision Process (MDP): States($S$), Actions($A$), Rewards($R$), Transition($T$)

Start at source entity $e_s$ (Tom Hanks) with the query relation $r_q$ (collaborators). Suppose at time step $t$ we are at entity $e_t$. Define the states $S$ of the MDP as $(e_t, (e_s, r_q))$

Each state has a self-loop and we process for at-most $T$ steps so every state is essentially a terminal state as well.

If the terminal state is a correct answer give a reward 1 else 0.

$$R_b(s_T) = 1\{(e_s, r_q, e_T) \in G\}$$

Can be trained via Policy Gradients using REINFORCE.
**Policy function**

\[ \pi_{\Theta}(a_t | s_t) \]

**Design choice:**
include search history in the state representation

**Figure:** Schematic Diagram of Minerva.
Every entity and relation is mapped to a dense embedding vector.

At time step $t$ we take action $a_t = (r_{t+1}, e_{t+1})$ which is represented by the concatenation of the two vectors.

Search history consists of all previously taken actions modeled by LSTM.

The policy network is given as

$$
\pi_\theta(a_t|s_t) = \text{SoftMax}(A_t \times W_2 \text{ReLU}(W_1[e_t; h_t; r_q]))
$$
Problem with Path-Based QA

1. False Negative Target: Since KGs are incomplete, even if a target node is visited it would get 0 reward as it is not a training example.

2. False Positive Path: An agent may arrive at a correct answer via spurious paths and may over-fit to these.

Figure: An example Knowledge Graph

- California
- United States
- Steven Spielberg
- The Post
- Tom Hanks
- Meryl Streep
- Phyllida Lloyd
- born_in
- locate_in
- live_in
- produced_in
- director
- cast_in
- cast_in
- collaborator?
**False-negative & nearly-correct entities ➞ true-negatives**

Figure: Incorrect reward distribution during training
1. **Reward Shaping for False Negative Target:** Use soft-rewards computed from pre-trained embedding models ($f$) for the targets. The new reward:

$$ R(s_T) = R_b(s_T) + (1 - R_b(s_T))f(e_s, r_q, e_T) $$

2. **Action dropout for False Positive Paths:** Mask outgoing path at random so that agent explores more diverse paths and doesn’t over-fit to spurious paths. The new policy:

$$ \tilde{\pi}_\theta(a_t|s_t) \propto (\pi_\theta(a_t|s_t) \cdot m + \epsilon) $$

$$ m_i \sim Bernoulli(1 - \alpha), i = 1 \ldots |A_t| $$
Figure: Reward Shaping using pretrained ConvE
Action Dropout

Randomly offset the **sampling probabilities** with rate $\alpha$ and renormalize

$$\tilde{\pi}_\Theta(a_t | s_t) \propto \pi_\Theta(a_t | s_t) \cdot m + \epsilon$$

$$m_i \sim \text{Bernoulli}(1 - \alpha), i = 1, \ldots, N$$

**Figure:** Masking paths to allow more exploration.
### Evaluation

<table>
<thead>
<tr>
<th>Name</th>
<th># Ent.</th>
<th># Rel.</th>
<th># Fact</th>
<th># Degree Avg</th>
<th># Degree Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinship</td>
<td>104</td>
<td>25</td>
<td>8,544</td>
<td>85.15</td>
<td>82</td>
</tr>
<tr>
<td>UMLS</td>
<td>135</td>
<td>46</td>
<td>5,216</td>
<td>38.63</td>
<td>28</td>
</tr>
<tr>
<td>FB15k-237</td>
<td>14,505</td>
<td>237</td>
<td>272,115</td>
<td>19.74</td>
<td>14</td>
</tr>
<tr>
<td>WN18RR</td>
<td>40,945</td>
<td>11</td>
<td>86,835</td>
<td>2.19</td>
<td>2</td>
</tr>
<tr>
<td>NELL-995</td>
<td>75,492</td>
<td>200</td>
<td>154,213</td>
<td>4.07</td>
<td>1</td>
</tr>
</tbody>
</table>

**Evaluation Protocol:** MRR (Mean Reciprocal Rank)

**Figure:** KG benchmarks.
Model Ablation

Figure: Model Ablation using Dev Sets MRR x100

Esp. helpful on dense graphs

Figure: Model Ablation using Dev Sets MRR x100
Action dropout helps in every case suggesting thorough exploration of the path space is important across datasets.

Reward Shaping helps in Kinship and UMLS, but doesn’t give much performance boost on the other three. Likely reason is ConvE is itself not very good at these datasets and might be adding noise to the system.
SOTA Comparison

Figure: Test Set MRR x 100 compared to SOTA approaches
### SOTA Comparison

**Figure:** Query answering performance compared to state-of-the-art embedding based approaches (top part) and multi-hop reasoning approaches (bottom part). The @1, @10 and MRR metrics were multiplied by 100.

<table>
<thead>
<tr>
<th>Model</th>
<th>UMLS</th>
<th>Kinship</th>
<th>FB15k-237</th>
<th>WN18RR</th>
<th>NELL-995</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@1</td>
<td>@10</td>
<td>MRR</td>
<td>@1</td>
<td>@10</td>
</tr>
<tr>
<td>DistMult (Yang et al., 2014)</td>
<td>82.1</td>
<td>96.7</td>
<td>86.8</td>
<td>48.7</td>
<td>90.4</td>
</tr>
<tr>
<td>ComplEx (Trouillon et al., 2016)</td>
<td>89.0</td>
<td>99.2</td>
<td>93.4</td>
<td>81.8</td>
<td>98.1</td>
</tr>
<tr>
<td>ConvE (Dettmers et al., 2018)</td>
<td><strong>93.2</strong></td>
<td><strong>99.4</strong></td>
<td><strong>95.7</strong></td>
<td>79.7</td>
<td><strong>98.1</strong></td>
</tr>
<tr>
<td>NeuralLP (Yang et al., 2017)</td>
<td>64.3</td>
<td>96.2</td>
<td>77.8</td>
<td>47.5</td>
<td>91.2</td>
</tr>
<tr>
<td>NTP-λ (Rocktäschel et. al. 2017)</td>
<td>84.3</td>
<td><strong>100</strong></td>
<td>91.2</td>
<td>75.9</td>
<td>87.8</td>
</tr>
<tr>
<td>MINERVA (Das et al., 2018)</td>
<td>72.8</td>
<td>96.8</td>
<td>82.5</td>
<td>60.5</td>
<td>92.4</td>
</tr>
<tr>
<td>Ours(ComplEx)</td>
<td>88.7</td>
<td>98.5</td>
<td>92.9</td>
<td><strong>81.1</strong></td>
<td><strong>98.2</strong></td>
</tr>
<tr>
<td>Ours(ConvE)</td>
<td><strong>90.2</strong></td>
<td><strong>99.2</strong></td>
<td><strong>94.0</strong></td>
<td>78.9</td>
<td><strong>98.2</strong></td>
</tr>
</tbody>
</table>
Some missing entries are due to high computational cost of path-based models.

The proposed method achieves significantly higher performance than previous path-based approaches and similar performance to embedding based approach.

The relationship between reward-shaping module performance and the resulting model is not consistent in the case of WN18RR where Ours(ComplEx) performs better than Ours(ConvE) despite ComplEx performing worse than ConvE.

The proposed method can be seen as an interpretable counterpart of the embedding approach without any significant loss in performance.
Interpretable Results

Laura Carmichael \(\xrightarrow{\text{profession}}\) ?
- Richard Bonneville
  - people/person/profession
  - actor/actress
  - \(7.819\times 10^{-3}\)
- Penelope Wilton
  - people/person/profession
  - comedienne
  - \(1.395\times 10^{-4}\)
- Laura Carmichael
  - /award_winner
  - SAG Award
    - /award_nomination/award
  - /people/person/nationality
    - United Kingdom
      - /people/person/nationality
- Elizabeth McGovern
  - /event_participant
- Don Cheadle
  - /people/person/profession
  - writer
  - \(9.466\times 10^{-5}\)
- Ricky Gervais
  - /people/person/profession
  - film maker
  - \(5.122\times 10^{-5}\)
- Nicholas Cage
  - /people/person/profession
  - TV producer
  - \(4.713\times 10^{-5}\)

FB15k-237 (Toutanova and Chen 2016)
Conclusion

1. Proposes Reward Shaping and Action Dropout to address false negative targets and false positive paths.
2. Similar performance to embedding based approaches and is highly interpretable.
The figures are taken from the author’s presentation at EMNLP’18
http://victorialin.net/pubs/MultiHopKG_emnlp2018_talk.pdf