Complex Query Answering on Neural Knowledge Graphs with Rich Semantics

Yangqiu Song
Department of CSE, HKUST
Slides Credit: Jiaxin Bai and Zihao Wang
Knowledge Graphs

• Large-scale knowledge graphs about entities and their attributes (property-of) and relations (thousands of different predicates)

• Developed since Google released its knowledge graph in 2012
  • Millions of entities and concepts
  • Billions of relationships

Google Knowledge Graph (2012)
570 million entities and 18 billion facts
Why is it still Important?

• Large language models (LLMs) tend to better memorize head (popular, more frequent) knowledge

<table>
<thead>
<tr>
<th>Prompt</th>
<th># parsed</th>
<th># correct</th>
<th>Acc (%)</th>
<th>Hit@5 (%)</th>
<th># parsed</th>
<th># correct</th>
<th>Acc (%)</th>
<th>Hit@5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>0</td>
<td>0</td>
<td>0.00</td>
<td>7.95</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>JP</td>
<td>46</td>
<td>26</td>
<td>29.55</td>
<td>61.36</td>
<td>50</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>MJP</td>
<td>85</td>
<td>37</td>
<td>42.04</td>
<td>79.55</td>
<td>97</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1: Email address recovery results on sampled emails from the Enron Email Dataset.
Entity/Facts as Memories

Figure credit: Verga et al., 2021
What is Missing?

- Retrieval Augmented Generation (RAG)

Triple may not be enough
Q: What is the effect of the Fed raising the interest rate?

The Federal Reserve's main monetary policy instrument is its Federal funds rate target. By adjusting this target, the Fed affects a wide range of market interest rates and in turn indirectly affects stock prices, wealth and currency exchange rates. …

Interbank borrowing is essentially a way for banks to quickly raise money. … Raising the federal funds rate will dissuade banks from taking out such inter-bank loans, which in turn will make cash that much harder to procure.

The effect of the Fed raising interest rates can lead to higher borrowing costs for consumers and businesses, which can slow down economic growth. It can also make it more expensive for people to take out loans, such as mortgages and car loans. However, it can also help to control inflation and stabilize the economy in the long term.
Q: What is the effect of the Fed raising the interest rate?

The effect of the Fed raising interest rates can lead to higher borrowing costs for consumers and businesses, which can slow down economic growth. It can also make it more expensive for people to take out loans, such as mortgages and car loans. However, it can also help to control inflation and stabilize the economy in the long term.
Retrieval Augmented Generation (Cont’d)

<table>
<thead>
<tr>
<th>Texts</th>
<th>Embeddings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Match</td>
<td>Exact match</td>
</tr>
<tr>
<td></td>
<td>Semantic match</td>
</tr>
<tr>
<td>Search space</td>
<td>Sparse vectors</td>
</tr>
<tr>
<td></td>
<td>Dense vectors</td>
</tr>
<tr>
<td>In-context learning</td>
<td>Symbolic</td>
</tr>
<tr>
<td></td>
<td>Neural</td>
</tr>
</tbody>
</table>
What is the effect of the Fed raising the interest rate?

What is the effect of the Fed raising the interest rate while China dropping the rate?

What is the effect of the Fed raising the interest rate while China dropping the rate and OPEC+ reduces oil production?

What is the effect of the Fed raising the interest rate while China dropping the rate and OPEC+ reduces oil production, and the consequence of this effect?

What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?
What is the effect of the Fed raising the interest rate to **more than 6%** while China dropping the rate to **less than 2%** and OPEC+ reduces oil production by **1 million barrels a day**, and the consequence of this effect?
What is the effect of the Fed raising the interest rate to more than 6% while China dropping the rate to less than 2% and OPEC+ reduces oil production by 1 million barrels a day, and the consequence of this effect?
More Generally

• Traditional NoSQL or Graph databases
• Semantic match of strings

Can we support complex queries such as join, intersection, counting, etc., on top of that?
Neural Graph Databases

• Graph structure + vector storage
  • Leveraging the power of LLMs for textual data

• Query executor to support complex queries
  • Query encoder
  • Inductive method to be robust with insertion, deletion, and modification
  • Fuzzy semantic search
  • Generalizable to incomplete data

Figure taken from: https://towardsdatascience.com/neural-graph-databases-cc35c9e1d04f
Complex Queries on Neuralized Knowledge Graphs

• A working example: Tree-Formed Queries (TFQ):
  • Tree-form query family contains the queries that can be converted into the computational tree

Natural Language: Find non-American directors whose movie won Golden Globes or Oscar?

Logical Formula: \[ q = V_? \exists V_1. (\text{Won}(V_1, \text{GoldenGlobes}) \lor \text{Won}(V_1, \text{Oscar})) \land \neg \text{BornIn}(V_?, \text{America}) \land \text{Direct}(V_?, V_1) \]

Set Operator Tree: \( \text{DirectorOf}(\text{WinnerOf}(\text{GoldenGlobes}) \cup \text{WinnerOf}(\text{Oscar})) \cap \text{BornIn}(\text{America})^C \)

Example from: Zihao Wang, Weizhi Fei, Hang Yin, Yangqiu Song, Ginny Y Wong, and Simon See. Wasserstein-Fisher-Rao Embedding: Logical Query Embeddings with Local Comparison and Global Transport In Findings of ACL 2023
The Design Space of Neural TFQ Answering

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity set</td>
<td>$\mathcal{E}$</td>
<td>The entity set in KG</td>
</tr>
<tr>
<td>Relation set</td>
<td>$\mathcal{R}$</td>
<td>The relation set in KG</td>
</tr>
<tr>
<td>Set embedding space</td>
<td>$\mathcal{X}$</td>
<td>Embedding space</td>
</tr>
<tr>
<td>Set embedding lookup</td>
<td>$E_X: \mathcal{E} \mapsto \mathcal{X}$</td>
<td>Singleton set embedding</td>
</tr>
<tr>
<td>Entity embedding space</td>
<td>$\mathcal{Y}$</td>
<td>Embedding space</td>
</tr>
<tr>
<td>Entity embedding lookup</td>
<td>$E_Y: \mathcal{E} \mapsto \mathcal{Y}$</td>
<td>Entity embedding</td>
</tr>
<tr>
<td>Set intersection</td>
<td>$I: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$</td>
<td>Binary or N-ary</td>
</tr>
<tr>
<td>Set union</td>
<td>$U: \mathcal{X} \times \cdots \times \mathcal{X} \mapsto \mathcal{X}$</td>
<td>Binary or N-ary</td>
</tr>
<tr>
<td>Set complement</td>
<td>$C: \mathcal{X} \mapsto \mathcal{X}$</td>
<td>Replaceable with set difference</td>
</tr>
<tr>
<td>Set projection</td>
<td>$P: \mathcal{X} \times \mathcal{R} \mapsto \mathcal{X}$</td>
<td>One-hop link prediction</td>
</tr>
<tr>
<td>Scoring function</td>
<td>$s: \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$</td>
<td>How much an entity is in a set</td>
</tr>
</tbody>
</table>

Converting to computational tree makes it possible to model set operations with neural networks.

### The Design Space of Neural TFQ Answering

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definition</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity set</td>
<td>$\mathcal{E}$</td>
<td>Known notation</td>
</tr>
<tr>
<td>Relation set</td>
<td>$\mathcal{R}$</td>
<td>Known notation</td>
</tr>
<tr>
<td>Set embedding space</td>
<td>$\mathcal{X}$</td>
<td></td>
</tr>
<tr>
<td>Set embedding lookup</td>
<td>$\mathbb{E}_\mathcal{X}: \mathcal{E} \rightarrow \mathcal{X}$</td>
<td>Simplified</td>
</tr>
<tr>
<td>Entity embedding space</td>
<td>$\mathcal{Y}$</td>
<td>[Entity embedding: Slot 2]</td>
</tr>
<tr>
<td>Entity embedding lookup</td>
<td>$\mathbb{E}_\mathcal{Y}: \mathcal{E} \rightarrow \mathcal{Y}$</td>
<td>Simplified</td>
</tr>
<tr>
<td>Set intersection</td>
<td>$I: \mathcal{X} \times \cdots \times \mathcal{X} \rightarrow \mathcal{X}$</td>
<td>[Slot 3]</td>
</tr>
<tr>
<td>Set union</td>
<td>$U: \mathcal{X} \times \cdots \times \mathcal{X} \rightarrow \mathcal{X}$</td>
<td>[Slot 4]</td>
</tr>
<tr>
<td>Set complement</td>
<td>$C: \mathcal{X} \rightarrow \mathcal{X}$</td>
<td>[Slot 5]</td>
</tr>
<tr>
<td>Set projection</td>
<td>$P: \mathcal{X} \times \mathcal{R} \rightarrow \mathcal{X}$</td>
<td>[Slot 6]</td>
</tr>
<tr>
<td>Scoring function</td>
<td>$s: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$</td>
<td>[Slot 7]</td>
</tr>
</tbody>
</table>

Each method will be introduced by filling 7 slots.
Embedding Space and Set Representations

\[ q = V \land \exists V: Win(TuringAward, V) \land \neg Citizen(Canada, V) \land Graduate(V, V) \]

The multi-hop logical operations make the query answers diversified. The answers embeddings are set(s) scattered in the embedding space.
What’s Still Missing to Support RAG?

What is the effect of the Fed raising the interest rate to \textit{more than 6%} while China dropping the rate to \textit{less than 2%} and OPEC+ reduces oil production by \textit{1 million barrels a day}, and the consequence of this effect?

\[ p(y_t | y_{t-1}, \ldots, y_1, q, c_1, \ldots, c_M) \]
This Talk

• Neural KG CQA on Entities and Numerical Values

• Neural KG CQA on Eventuality Knowledge Graphs
## Numerical Complex Query Answering

<table>
<thead>
<tr>
<th>Category</th>
<th>Complex Queries</th>
<th>Interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical CQA</td>
<td>$q_2 = V_? \cdot \exists X_1, X_2: \text{Win}(V_?, \text{TuringAward}) \land \text{GreaterThan}(1927, X_2) \land \text{BornIn}(V_?, X_2)$</td>
<td>Find the Turing award winners that is born before the year of 1927.</td>
</tr>
<tr>
<td>Numerical CQA</td>
<td>$q_3 = V_? \cdot \exists X_1, X_2: \text{LocatedIn}(V_?, \text{UnitedStates}) \land \text{HasLatitude}(V_?, X_1) \land \text{GreaterThan}(X_1, X_2) \land \text{HasLatitude}(\text{Beijing}, X_2)$</td>
<td>Find the states in US that have a higher latitudes than Beijing.</td>
</tr>
<tr>
<td>Numerical CQA</td>
<td>$q_4 = V_? \cdot \exists X_1, X_2, X_3: \text{LocatedIn}(V_?, \text{UnitedStates}) \land \text{HasPopulation}(V_?, X_1) \land \text{SmallerThan}(X_1, X_2) \land \text{TimesByTwo}(X_2, X_3) \land \text{HasPopulation}(\text{California}, X_3)$</td>
<td>Find the states in US that have a twice smaller population than California?</td>
</tr>
</tbody>
</table>
Number Reasoning Network

Find the cities that have a higher latitudes than Japanese cities.

\[ q = V \cdot \exists V_1, X_1, X_2: \text{HasLatitude}(V_?, X_2) \land \text{GreaterThan}(X_2, X_1) \land \text{HasLatitude}(V_1, X_1) \land \text{LocatedIn}(V_1, \text{Japan}) \]

(1) Relational Projection
(2) Attribute Projection
(3) Numerical Projection
(4) Reverse Attribute Projection

Jiaxin Bai, Chen Luo, Zheng Li, Qingyu Yin, Bing Yin, Yangqiu Song: Knowledge Graph Reasoning over Entities and Numerical Values. KDD 2023: 57-68
Number Reasoning Network

(1) Relational Projection (rp):
Query Embedding $\rightarrow$ Entity Set

(2) Attribute Projection (ap):
Query/Set Embedding $\rightarrow$ Value Distribution

(3) Numerical Projection (np):
Value Distribution $\rightarrow$ Value Distribution

(4) Reverse Attribute Projection (rap):
Value Distribution $\rightarrow$ Query Embedding
Number Reasoning Network

(1) Relational Projection: 
Adopted from the backbones: GQE, Query2Box, Query2Particles.

(2) Attribute Projection

(3) Numerical Projection

(4) Reverse Attribute Projection

(2) (3) (4) Other Projections: Gated Transitions

\[ p_i = W_p q_i + b_p \]  
Linear projection

\[ z_i = \sigma (W_z e_a + U_z p_i + b_z) \]

\[ r_i = \sigma (W_r e_a + U_r p_i + b_r) \]

\[ t_i = \varphi (W_h e_a + U_h (r_i \odot p_i) + b_h) \]  
MLP

\[ \theta_{i+1} = (1 - z_i) \odot p_i + z_i \odot t_i \]  
Gate selection
Number Reasoning Network

Entity embeddings:
Adopted from the backbones:
GQE, Query2Box, Query2Particles.

Input number embeddings
- DICE
- Sinusoidal

\[
\psi(v)_d = \begin{cases} 
\sin^{d-1}(\alpha) \cos(\alpha) & \text{if } d \equiv 0 \,(\mod\,2) \\
\sin^D(\alpha) & \text{if } d \equiv 1 \,(\mod\,2) 
\end{cases}
\]

Number Reasoning Network

• Logic Operators on Entities, adopted from the backbones:
  • GQE, Query2Box, Query2Particles.
• Logic Operator on Value Distribution:
  • Intersection and Union: DeepSet
    \[ a_i = \text{Attn}(W_q \theta_i^T, W_k \theta_i^T, W_v \theta_i^T) \]
    \[ \theta_{i+1} = \text{MLP}(a_i) \]
  • Relational Projection (rp)
  • Attribute Projection (ap)
  • Numerical Projection (np)
  • Reverse Attribute Projection (rap)
Number Embeddings and Learning Objective

Use maximize a posteriori probability (MAP) estimation to derive an objective function for type-aware reasoning.

\[ \hat{\theta}_I(v, t) = \arg \max_{\theta_I} f(\theta_I | v, t) \]
\[ = \arg \max_{\theta_I} f(\psi(v) | \theta_I, t) g(\theta_I | t) \]
\[ = \arg \min_{\theta_I} -\log f(\psi(v) | \theta_I) - \log g(\theta_I | t) \]

(Bayes’ Rule; Remove the denominator: a constant in argmax)

(Conditional Independent of \( v \) and \( t \) on \( \theta_I \))

\( v \) is the positive answer value.
\( t \) is the type of this value like date, length, size etc.
\( \theta_I \) is the distribution parameters in the last step.
\( \psi(v) \) is number embeddings

Set the parameteration as:

\[ f(\psi(v) | \theta_I) = p_{\theta_I}(\psi(v)) \]
\[ g(\theta_I | t) = \phi_t(\theta_I) \]

\[ L_A = \frac{1}{M} \sum_{j=1}^{M} (- \log p_{\theta_I(j)}(\psi(v(j))) - \log \phi_t(j)(\theta_I(j))) \]

\begin{align*}
\text{happenedOnDate} & \quad \text{createdOnDate} & \quad \text{date_of_death} \\
\text{latitude} & \quad \text{longitude} & \quad \text{date_of_birth} \\
\text{person.height_mt} & \quad \text{film_release_date} & \quad \text{org.date_founded} \\
\text{location.date_founded} & \quad \text{date_of_death} & \quad \text{date_of_birth} \\
\end{align*}

Jiaxin Bai, Chen Luo, Zheng Li, Qingyu Yin, Bing Yin, Yangqiu Song: Knowledge Graph Reasoning over Entities and Numerical Values. KDD 2023: 57-68
Number Embeddings and Learning Objective

End-to-end training by Joint optimization of two losses:

$$L_A = \frac{1}{M} \sum_{j=1}^{M} \left( - \log p_{\theta_{I}(j)} \left( \psi(v_{j}) \right) \right) - \log \phi_{t_{I}(j)}(\theta_{I}(j))$$

The likelihood of the value $v_{j}$ sampled from distribution of $\theta_{I}(j)$.

The likelihood of the distribution parameter $\theta_{I}(j)$ is of type $t_{I}(j)$.

$$L_E = -\frac{1}{N} \sum_{j=1}^{N} \log p(q_{I}(j), v_{j})$$

The likelihood of the entity $v_{j}$ is the answer of the query encoding $q_{I}(j)$.

$j$ means the $j$-th sample, and $I$ means the last step of distribution parameter encoding.
Sampling Data

Numerical Values Types on Different KGs:

General Query Types:

Number related query types:
# Data Statistics

<table>
<thead>
<tr>
<th>Graphs</th>
<th>Data Split</th>
<th>#Nodes</th>
<th>#Rel.</th>
<th># Attr.</th>
<th>#Rel. Edges</th>
<th>#Attr. Edges</th>
<th>#Rev. Attr. Edges</th>
<th>#Num. Edges</th>
<th>#Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>FB15K</td>
<td>Training</td>
<td>25,106</td>
<td>1,345</td>
<td>15</td>
<td>947,540</td>
<td>20,248</td>
<td>20,248</td>
<td>27,020</td>
<td>1,015,056</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>26,108</td>
<td>1,345</td>
<td>15</td>
<td>1,065,982</td>
<td>22,779</td>
<td>22,779</td>
<td>27,376</td>
<td>1,138,916</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>27,144</td>
<td>1,345</td>
<td>15</td>
<td>1,184,426</td>
<td>25,311</td>
<td>25,311</td>
<td>27,389</td>
<td>1,262,437</td>
</tr>
<tr>
<td>DB15K</td>
<td>Training</td>
<td>31,980</td>
<td>279</td>
<td>30</td>
<td>145,262</td>
<td>33,131</td>
<td>33,131</td>
<td>25,495</td>
<td>237,019</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>34,191</td>
<td>279</td>
<td>30</td>
<td>161,978</td>
<td>37,269</td>
<td>37,269</td>
<td>25,596</td>
<td>262,112</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>36,358</td>
<td>279</td>
<td>30</td>
<td>178,394</td>
<td>41,411</td>
<td>41,411</td>
<td>25,680</td>
<td>286,896</td>
</tr>
<tr>
<td>YAGO15K</td>
<td>Training</td>
<td>32,112</td>
<td>32</td>
<td>7</td>
<td>196,616</td>
<td>21,732</td>
<td>21,732</td>
<td>26,616</td>
<td>266,696</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>33,078</td>
<td>32</td>
<td>7</td>
<td>221,194</td>
<td>22,748</td>
<td>22,748</td>
<td>26,627</td>
<td>293,317</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>33,610</td>
<td>32</td>
<td>7</td>
<td>245,772</td>
<td>23,520</td>
<td>23,520</td>
<td>26,631</td>
<td>319,443</td>
</tr>
<tr>
<td>Graphs</td>
<td>Data Split</td>
<td>1p</td>
<td>2p</td>
<td>2i</td>
<td>3i</td>
<td>pi</td>
<td>ip</td>
<td>2u</td>
<td>up</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>FB15K</td>
<td>Training</td>
<td>304,633</td>
<td>138,192</td>
<td>226,729</td>
<td>288,874</td>
<td>260,057</td>
<td>233,834</td>
<td>284,301</td>
<td>284,931</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>8,271</td>
<td>15,860</td>
<td>23,359</td>
<td>28,836</td>
<td>25,081</td>
<td>22,930</td>
<td>29,187</td>
<td>29,210</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>7,969</td>
<td>15,431</td>
<td>23,346</td>
<td>28,865</td>
<td>24,810</td>
<td>22,232</td>
<td>29,212</td>
<td>29,274</td>
</tr>
<tr>
<td>DB15K</td>
<td>Training</td>
<td>124,851</td>
<td>99,698</td>
<td>140,427</td>
<td>190,413</td>
<td>171,353</td>
<td>163,687</td>
<td>190,364</td>
<td>194,244</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>3,529</td>
<td>10,388</td>
<td>9,792</td>
<td>13,817</td>
<td>14,594</td>
<td>16,651</td>
<td>19,512</td>
<td>19,792</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>3,387</td>
<td>10,047</td>
<td>9,914</td>
<td>14,603</td>
<td>14,642</td>
<td>15,897</td>
<td>19,504</td>
<td>19,773</td>
</tr>
<tr>
<td>YAGO15K</td>
<td>Training</td>
<td>84,014</td>
<td>76,238</td>
<td>136,282</td>
<td>183,850</td>
<td>162,712</td>
<td>145,994</td>
<td>183,963</td>
<td>183,459</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>2,833</td>
<td>7,986</td>
<td>10,757</td>
<td>16,884</td>
<td>13,485</td>
<td>13,899</td>
<td>18,444</td>
<td>19,105</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>2,713</td>
<td>7,949</td>
<td>10,935</td>
<td>17,171</td>
<td>13,481</td>
<td>13,526</td>
<td>18,433</td>
<td>18,997</td>
</tr>
</tbody>
</table>
## Main Results on Three KGs

<table>
<thead>
<tr>
<th>Query Encoding</th>
<th>Attribute</th>
<th>Hit@1</th>
<th>Hit@3</th>
<th>Hit@10</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GQE</strong></td>
<td>Baseline</td>
<td>10.33</td>
<td>18.19</td>
<td>27.91</td>
<td>16.29</td>
</tr>
<tr>
<td></td>
<td>NRN + DICE</td>
<td>11.03</td>
<td>19.18</td>
<td>29.01</td>
<td>17.15</td>
</tr>
<tr>
<td></td>
<td>NRN + Sinusoidal</td>
<td><strong>11.14</strong></td>
<td><strong>19.39</strong></td>
<td><strong>29.23</strong></td>
<td><strong>17.31</strong></td>
</tr>
<tr>
<td><strong>Q2P</strong></td>
<td>Baseline</td>
<td>10.22</td>
<td>17.35</td>
<td>26.61</td>
<td>15.81</td>
</tr>
<tr>
<td></td>
<td>NRN + DICE</td>
<td>11.86</td>
<td>19.70</td>
<td>29.46</td>
<td>17.84</td>
</tr>
<tr>
<td></td>
<td>NRN + Sinusoidal</td>
<td><strong>12.25</strong></td>
<td><strong>20.16</strong></td>
<td><strong>29.96</strong></td>
<td><strong>18.28</strong></td>
</tr>
<tr>
<td><strong>Q2B</strong></td>
<td>Baseline</td>
<td>11.81</td>
<td>20.93</td>
<td>31.19</td>
<td>18.41</td>
</tr>
<tr>
<td></td>
<td>NRN + DICE</td>
<td>12.52</td>
<td>22.09</td>
<td>32.34</td>
<td>19.34</td>
</tr>
<tr>
<td></td>
<td>NRN + Sinusoidal</td>
<td><strong>12.75</strong></td>
<td><strong>22.22</strong></td>
<td><strong>32.46</strong></td>
<td><strong>19.51</strong></td>
</tr>
</tbody>
</table>
This Talk

• Neural KG CQA on Entities and Numerical Values

• Neural KG CQA on Eventuality Knowledge Graphs
Complex query on eventuality graphs are **different** from the entity-relation graph

**Whether and when** the eventualities occur are important

<table>
<thead>
<tr>
<th>Queries</th>
<th>Type</th>
<th>Interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1 = V \cdot \exists V: \text{Interact}(V, V) \wedge \text{Assoc}(V, Alzheimer) \wedge \text{Assoc}(V, MadCow)$</td>
<td>Entity</td>
<td>Find the substances that interact with the proteins associated with Alzheimer’s and Mad cow disease.</td>
</tr>
<tr>
<td>$q_2 = V \cdot \text{Precedence}(\text{Food is bad, PersonX add soy sauce}) \wedge \text{Reason}(\text{Food is bad}, V)$</td>
<td>Eventuality</td>
<td>Food is bad before PersonX add soy sauce. What is the reason for food being bad?</td>
</tr>
<tr>
<td>$q_3 = V \cdot \text{Precedence}(V, \text{PersonX go home}) \wedge \text{ChosenAlternative}(\text{PersonX go home, PersonX buy an umbrella})$</td>
<td>Eventuality</td>
<td>Instead of buying an umbrella, PersonX go home. What happened before PersonX go home?</td>
</tr>
</tbody>
</table>
(Activities, States, Events, and their Relations)

Principle 1: Comparing semantic meanings by fixing grammar (Katz and Fodor, 1963)
Principle 2: The need of language inference based on ‘partial information’ (Wilks, 1975)

https://github.com/HKUST-KnowComp/ASER
Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, Cane Wing-Ki Leung: ASER: A Large-scale Eventuality Knowledge Graph. WWW 2020: 201-211
Conceptualization and Normalization

Conceptualized ASER

- PersonX gives PersonY Red-Meat
- PersonX order Meat
- PersonX be hungry
- PersonX be thirsty
- PersonX eat dish
- PersonX be full

Eventualities:
- He orders meat (Pr = 0.1)
- He orders beef (Pr = 0.2)
- He orders chicken (Pr = 0.1)

Relations:
- Conjunction (0.05)
- Synchronous (7.5)
- Succession (0.042)
- Precedence (0.042)
- Precedence (0.057)
- Succession (0.5)

See the full paper for more details.

https://github.com/HKUST-KnowComp/ASER
Hongming Zhang, Xin Liu, Haojie Pan, Haowen Ke, Jiefu Ou, Tianqing Fang, Yangqiu Song: ASER: Towards large-scale commonsense knowledge acquisition via higher-order selectional preference over eventualities. Artif. Intell. 309: 103740 (2022)
Discourse Relations and Implicit Constraints

- PersonX did not eat anything because PersonX was full

Reason($PersonX$ did not eat anything, $PersonX$ was full)

Occurrence Constraint

\[ \eta(PersonX \text{ did not eat anything}) \land \eta(PersonX \text{ was full}) \land \eta(PersonX \text{ did not eat anything}) \leftarrow \eta(PersonX \text{ was full}) \]

Temporal Constraints

\[ \tau(PersonX \text{ did not eat anything}) > \tau(PersonX \text{ was full}) \]

\[ \eta(A) = 1 \text{ if and only if it occurs} \]

\[ \tau(A) > \tau(B) : A \text{ happens after } B \]
Discourse Relations and Implicit Constraints

- Food is bad before PersonX add soy sauce

Precedence\((\text{Food is bad, PersonX adds soy sauce})\)

<table>
<thead>
<tr>
<th>Occurrence Constraint</th>
<th>(\eta(\text{Food is bad}) \land \eta(\text{PersonX adds soy sauce}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Constraints</td>
<td>(\tau(\text{Food is bad}) &lt; \tau(\text{PersonX adds soy sauce}))</td>
</tr>
</tbody>
</table>

\(\tau(A) < \tau(B)\) : A happens before B
\(\eta(A) = 1\) if and only if it occurs
Discourse Relations and Implicit Constraints

• Instead of buying an umbrella, PersonX go home

\[ \eta(A) = 1 \text{ if and only if it occurs} \]
## Logical Constraints behind Discourse Relations

<table>
<thead>
<tr>
<th>Discourse Relations</th>
<th>Semantics</th>
<th>Implicit Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedence(A, B)</td>
<td>A occurs before B.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Succession(A, B)</td>
<td>A occurs after B happens.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Synchronous(A, B)</td>
<td>A occurs at the same time as B.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Reason(A, B)</td>
<td>A occurs because B.</td>
<td>$\eta(A) \land \eta(B) \land (\eta(A) \leftarrow \eta(B))$</td>
</tr>
<tr>
<td>Result(A, B)</td>
<td>A occurs as a result B.</td>
<td>$\eta(A) \land \eta(B) \land (\eta(A) \rightarrow \eta(B))$</td>
</tr>
<tr>
<td>Condition(A, B)</td>
<td>If B occurs, A.</td>
<td>$\eta(A) \rightarrow \eta(B)$</td>
</tr>
<tr>
<td>Concession(A, B)</td>
<td>B occurs, although A.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Contrast(A, B)</td>
<td>B occurs, but A.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Conjunction(A, B)</td>
<td>A and B both occur.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Instantiation(A, B)</td>
<td>B is a more detailed description of A.</td>
<td>$\eta(A) \land \eta(B)$</td>
</tr>
<tr>
<td>Restatement(A, B)</td>
<td>A restates the semantics of B.</td>
<td>$\eta(A) \leftrightarrow \eta(B)$</td>
</tr>
<tr>
<td>Alternative(A, B)</td>
<td>A and B are alternative situations.</td>
<td>$\eta(A) \lor \eta(B)$</td>
</tr>
<tr>
<td>ChosenAlternative(A, B)</td>
<td>Instead of B occurs, A.</td>
<td>$\eta(A) \land \neg \eta(B)$</td>
</tr>
<tr>
<td>Exception(A, B)</td>
<td>A, except B.</td>
<td>$\neg \eta(A) \land \eta(B) \land (\neg \eta(B) \rightarrow \eta(A))$</td>
</tr>
</tbody>
</table>

$\tau(A) \prec \tau(B)$ : A happens before B, $\eta(A) = 1$ if and only if it occurs
Implicit Temporal Logical Constraints

Consider the differences between the following questions:

Question 1: What is the reason of food being bad?
PersonX added soy sauce is a plausible answer.

Question 2: Food is ready bad before PersonX adds soy sauce. What is the reason of food being bad?
Adding soy sauce is not a plausible answer. Because adding soy sauce happens after food being bad, Because of this, it cannot be the reason of it. (Causality implies temporal relation)

Food is bad before PersonX add soy sauce. What is the reason for food being bad?

Query: $q = V_\tau \cdot \text{Precedence}(\text{Food is bad}, \text{PersonX adds soy sauce}) \land \text{Reason}(\text{Food is bad}, V_\tau)$

Constraints:

$\eta(\text{Food is bad}) \land \eta(\text{PersonX adds soy sauce}) \land \tau(\text{Food is bad}) < \tau(\text{PersonX adds soy sauce}) \land \eta(\text{Food is bad}) \land \eta(V_\tau) \land (\eta(V_\tau) \rightarrow \eta(\text{Food is bad})) \land \tau(\text{Food is bad}) > \tau(V_\tau)$

Occurrence:

$\eta(\text{Food is bad}) \land \eta(\text{PersonX adds soy sauce}) \land \eta(\text{Food is bad}) \land \eta(V_\tau) \land (\eta(V_\tau) \rightarrow \eta(\text{Food is bad}))$

Temporal:

$\tau(\text{Food is bad}) < \tau(\text{PersonX adds soy sauce}) \land \tau(\text{Food is bad}) > \tau(V_\tau)$
Consider the differences between the following questions:

**Question 1:** What happened before I go home?
Buying umbrella is a plausible answer.

**Question 2:** Instead of buying an umbrella, I go home. What happened before I go home?
Buying umbrella is not a plausible answer. Because the instead-of relation implicitly suggests it did not happen.

Instead of buying an umbrella, PersonX go home. What happened before PersonX go home?

**Query:**

\[ q = V_? \cdot \text{Precedence}(V_?, \text{PersonX go home}) \land \text{ChosenAlternative}((\text{PersonX go home, PersonX buy an umbrella}) \]

**Constraints:**

\[ \tau(V_?) < \tau(\text{PersonX go home}) \land \eta(V_?) \land \eta(\text{PersonX go home}) \land \eta(\text{PersonX go home}) \land \neg \eta(\text{PersonX buy an umbrella}) \]

**Occurrence:**

\[ \eta(V_?) \land \eta(\text{PersonX go home}) \land \neg \eta(\text{PersonX buy an umbrella}) \]

**Temporal:**

\[ \tau(V_?) < \tau(\text{PersonX go home}) \]
Logical Query with Implicit Constraints

Question: Food is bad before PersonY adds soy sauce. Instead of adding vinegar, PersonY adds ketchup. PersonX complains after V. PersonX leaves the restaurant after V. The reason V is V’?. What is V’?

Query on Graph: \( q = V \). \( \exists V: \text{Succession(PersonX complains, V)} \land \text{Succession(PersonX leaves restaurant, V)} \land \text{Reason(V, V')} \land \text{Precedence(Food is bad, PersonY adds soy sauce)} \land \text{ChosenAlternative(PersonY adds ketchup, PersonY adds vinegar)} \)

Without context and its constraints: 4 answers

With Implicit Constraints: Only 2 answers
(1) $s_{i,m} = <q_i, c_h^{(m)}> $
Computes the relevance of query embedding to the head of the memory key at position $m$.

(2) $v_i = \sum_{m=1}^{M} s_{i,m} (c_r^{(m)} + c_t^{(m)})$
Computes the aggregated memory values across $M$ memory cells with the importance weighted by relevance scores.

(3) $q_i = q_i + MLP(v_i)$
Computes the query embedding with memory values with the help of a MLP layer.
### Complex Eventuality Queries on ASER

<table>
<thead>
<tr>
<th>Data Split</th>
<th>Types</th>
<th>Query with Occurrence Constraints</th>
<th>Query with Temporal Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#Queries</td>
<td>#Answers</td>
</tr>
<tr>
<td>Train Queries</td>
<td>6</td>
<td>58,797</td>
<td>4.74</td>
</tr>
<tr>
<td>Validation Queries</td>
<td>15</td>
<td>22,320</td>
<td>7.20</td>
</tr>
<tr>
<td>Test Queries</td>
<td>15</td>
<td>24,466</td>
<td>7.93</td>
</tr>
</tbody>
</table>

- The tables shows the types and number queries;
- The number of answers on ASER;
- The number of logically contradictory answers.
The MEQE Combined with Various QE methods

<table>
<thead>
<tr>
<th>Models</th>
<th>Occurrence Constraints</th>
<th>Temporal Constraints</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hit@1</td>
<td>Hit@3</td>
<td>MRR</td>
</tr>
<tr>
<td>+ MEQE</td>
<td>10.20</td>
<td>15.54</td>
<td>14.31</td>
</tr>
<tr>
<td>+ MEQE</td>
<td>15.15</td>
<td>20.67</td>
<td>19.38</td>
</tr>
<tr>
<td>Nerual MLP</td>
<td>13.03</td>
<td>19.21</td>
<td>17.75</td>
</tr>
<tr>
<td>FuzzQE</td>
<td>11.68</td>
<td>18.64</td>
<td>17.07</td>
</tr>
<tr>
<td>+ MEQE</td>
<td>14.76</td>
<td>21.12</td>
<td>19.45</td>
</tr>
</tbody>
</table>
Conclusions

• Complex query answering on neural knowledge graphs/bases has great potential to support retrieval augmented generation (RAG)

• Two things are missing
  • Number and attribute understanding
  • Discourse relation modeling for logical queries

• We implemented models that can handle numbers and discourse relations
Future Work

• Incorporating complex query answering into retrieval augmented generation

• Exploring more strategies for complex query answering with rich semantics to handle Open World Problems
The Challenge of the Open World Problem

How to break the data incompleteness

Unobserved KG

Query

Observed KG

Graph Traversal

\( A_u \)

\( A_o \)
Neural strategy: End-to-end Training

Unobserved KG

Observed KG

Graph Traversal

Neural Models

Query

Inference

Training

$A_u$

$A_o$
Symbolic Strategy: Completion and Search

Unobserved KG \rightarrow \text{Graph Traversal} \rightarrow A_u

KG Completion/Population \rightarrow \text{Query} \rightarrow \text{Graph Traversal} \rightarrow A_o

Observed KG \rightarrow \text{Graph Traversal}
Other Works on CQA

Benchmarking EFO-1 (Existential First-Order Queries with Single Free Variable)

EFO-1 queries with cycles

EFO-K more than one variables

Data Models

Query encoder with OT

Learning in the inference step as a GNN (one-hop logical inference based MPNN)

Sequence encoder of queries

Zihao Wang, Yangqiu Song, Ginny Y. Wong, and Simon See. Logical Message Passing Networks with One-hop Inference on Atomic Formulas. In The Eleventh International Conference on Learning Representations, ICLR 2023
Zihao Wang, Weizhi Fei, Hang Yin, Yangqiu Song, Ginny Y Wong, and Simon See. Wasserstein-Fisher-Rao Embedding: Logical Query Embeddings with Local Comparison and Global Transport In Findings of ACL 2023
Hang Yin, Zihao Wang, and Yangqiu Song. Benchmarking the Combinatorial Generalizability of Complex Query Answering on Knowledge Graphs. In NeurIPS Datasets and Benchmarks Track, 2021
Thank you for your attention 😊