Knowledge Augmentation for Reasoning in Language

Part of the ACL 2023 Tutorial
“Complex Reasoning in Natural Language”

(Bill) Yuchen Lin
https://yuchenlin.xyz
Knowledge Augmentation for Reasoning

A question.

Blablabla......

Now I can reason better!

Retrieve

Knowledge Bases

Relevant Knowledge

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion
Knowledge Sources

Structured Knowledge

- WIKIDATA
- DBpedia
- ConceptNet
- ATOMIC
- ASCENT++

Un/Semi-structured Knowledge

- WIKIPEDIA
- Wiktionary
- PubMed
- wikiHow
- arXiv
- GenericsKB

Parametric Knowledge

- LAMA
- COMET
- OpenAI GPTs

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion
Structured Knowledge

Symbolic Structures of Knowledge

Neural Language Models (e.g., BERT [9])

How can we incorporate structured knowledge into neural language models?
Multiple-Choice QA
Question: xxxxx?
Candidates: A) a_1, B) a_2, C) a_3

KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning

KagNet [10]

Features
- LM encoder to encode the text of question + candidate
- Graph encoder to encode knowledge subgraphs
- Late fusion of Text Embeddings + Graph Embeddings
- Interpretable

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion
Multi-hop graph encoder ➡ Scalable

Text embeddings as virtual nodes in graph encoding

Early fusion of graphs and text encoders
Un/Semi-structured Knowledge

KGs can be limited:
1. Incompleteness
2. Only for binary relations

Task: Open-Ended QA

Question: xxxxx?
(A target knowledge corpus.)

Complex questions need multiple reasoning steps.
Who voices the dog in the TV show Family Guy?
What can help alleviate global warming?
What will separate iron filings from sand?
Dense Passage Retrieval (DPR) [14]
Multi-hop Dense Retrieval (MDR) [15]

Q: *What was the nickname of Judy Lewis's father?*

**P1: Judy Lewis** (born Judith Young; November 6, 1935 – November 25, 2011) was an American actress, writer, producer, and therapist. She was the secret biological daughter of actor Clark Gable and actress Loretta Young.

**P2: William Clark Gable** (February 1, 1901 – November 16, 1960) was an American film actor, often referred to as "The King of Hollywood". He had roles in more than 60 motion pictures in a wide variety of genres during a career that lasted 37 years...
Differentiable Fact-Following Operations (DrFact) [17]

A corpus of common-sense facts, e.g., GenericsKB, $\mathcal{F}$

$f_i \in \mathcal{F}$

A fact is a sentence of generic commonsense knowledge

$c_j \in \mathcal{V}$

A concept is a noun or noun-chunk that are mentioned in $\mathcal{F}$

$f_1 = \text{carbon dioxide is the major greenhouse gas contributing to global warming.}$

$f_2 = \text{trees remove carbon dioxide from the atmosphere through photosynthesis.}$

$f_3 = \text{the atmosphere contains oxygen, carbon dioxide, and water.}$

A knowledge corpus as a hypergraph.

A weighted set of facts $\rightarrow$ A sparse vector.

<table>
<thead>
<tr>
<th>Intro</th>
<th>Structured Knowledge</th>
<th>Un/Semi-structured Knowledge</th>
<th>Parametric Knowledge</th>
<th>Conclusion</th>
</tr>
</thead>
</table>
Differentiable Fact-Following Operations (DrFact) [17]

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion

q: What can help alleviate global warming?

Carbon dioxide is the major greenhouse gas contributing to global warming.

Fact-Following

initial facts

the next round?

Trees remove carbon dioxide from the atmosphere through photosynthesis.
Differentiable Fact-Following Operations (DrFact) [17]

\[ F_t = \text{Fact-Follow}(F_{t-1}, q_t) \]

Self-Following

Input: Previous-Hop

Output: Next-Hop

\( F_{t-1} \)

Sparse Fact-to-Fact Matrix

\(|\mathcal{F}| \times |\mathcal{F}|\)

Sparse Retrieval as Filtering

\( g(h_{t-1}, q_t) = z_t \)

Sparse-Dense Mixing

Dense Retrieval

MIPS

\( h_{t-1} \)

Weighted Aggregation of Fact Embeddings in \( F_{t-1} \)

\( q_t \)

Question Embedding at Hop-\( t \)
## A quick comparison

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Structure</td>
<td>A set of documents</td>
<td>A set of documents</td>
<td>Mention-Entity Bipartite Graph</td>
<td>Concept-Fact Hypergraph</td>
</tr>
<tr>
<td>Multi-hop Reasoning Formulation</td>
<td>-</td>
<td>- / Multiple-Round</td>
<td>Entity-Following</td>
<td>Fact-Following</td>
</tr>
<tr>
<td>Index for Dense Retrieval</td>
<td>-</td>
<td>Passage Embeddings</td>
<td>Mention Embedding</td>
<td>Fact Embeddings</td>
</tr>
<tr>
<td>Sparse Retrieval Method</td>
<td>TF-IDF based Index+ BM25 Ranking Func.</td>
<td>-</td>
<td>Entity-Mention Cooccurrence</td>
<td>Fact-to-Fact Matrix</td>
</tr>
<tr>
<td>Multi-Hop Questions</td>
<td>-</td>
<td>- / Single model</td>
<td>Aggregating Multiple Models</td>
<td>A single model w/ Self-Following</td>
</tr>
<tr>
<td>Intermediate Supervision</td>
<td>-</td>
<td>-</td>
<td>N/A</td>
<td>Distant Supervision</td>
</tr>
</tbody>
</table>

- **Intro** | **Structured Knowledge** | **Un/Semi-structured Knowledge** | **Parametric Knowledge** | **Conclusion**

13
Textual Knowledge

📚 Structured Knowledge

📖 Un/Semi-structured Knowledge

We are very accurate, easy to modify, trustworthy, and verifiable; However, we are incomplete & hard to query!

Parametric Knowledge

We are super generalizable and easy to query! But, we often hallucinate and make mistakes.

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion
• We can be very expensive.
• You have no idea about our data/weights.

Symbolic Knowledge Distillation [19]

<table>
<thead>
<tr>
<th>ATOMIC \textsuperscript{10X}</th>
<th>X starts running</th>
<th>xEffect so, X</th>
<th>gets in shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>X and Y engage in an argument</td>
<td>xWant so, X wants</td>
<td>to avoid Y</td>
<td></td>
</tr>
<tr>
<td>X learns to type fast</td>
<td>xNeed X needed</td>
<td>to have taken typing lessons</td>
<td></td>
</tr>
<tr>
<td>X steals his grandfather’s sword</td>
<td>xEffect so, X</td>
<td>is punished by his grandfather</td>
<td></td>
</tr>
<tr>
<td>X takes up new employment</td>
<td>xIntent because X wants</td>
<td>to be self sufficient</td>
<td></td>
</tr>
</tbody>
</table>

Intro | Structured Knowledge | Un/Semi-structured Knowledge | Parametric Knowledge | Conclusion
Input: The word children means <mask> or more kids.

Knowledge: The word child means one kid. The word children is the plural form of the word child.

Prediction: two
Reinforced Knowledge Introspector (Rainier) [21]

### Task: NumerSense

**Prompt**
-Generate some numerical facts about objects. Examples:
  - Input: *penguins have* `<mask>` *wings.*
  - Knowledge: *Birds have two wings. Penguin is a kind of bird.*
  - Input: *a typical human being has* `<mask>` *limbs.*
  - Knowledge: *Human has two arms and two legs.*
  - Input: *How many legs does an easel have?*
  - Knowledge: 

**Generated Knowledge Prompting (GKP)**

- A tripod is a kind of easel.
- A human is a kind of easel.
Conclusion

Structured Knowledge
- WIKIDATA
- DBpedia
- ConceptNet
- Ai2 ATOMICS

Un/Semi-structured Knowledge
- WIKIPEDIA
- PubMed
- Semantics Scholar
- wikiHow
- arXiv
- Ai2 GenericsKB

Instance-based Knowledge
- Datasets
- PromptSource
- FLAN Collection

Parametric Knowledge
- LAMA
- Ai2 COMET
- ChatGPT & GPT-4

Other Tools
- KagNet
- QA-GNN
- MHGRN
- GreaseLM
- DPR
- MDP
- DrKIT
- DrFact
- SKD
- GKP
- Rainier

Reference is on the website: https://yuchenlin.xyz/
Future Directions

1. How can we merge all these different sources of knowledge into a unified knowledge model?

2. How can we deliver a more faithful and interpretable reasoning models with low cost?

3. How do we collect and ground knowledge of real world and social interactions between humans?