Graph Learning for Natural Language Understanding

Antoine Bosselut
What is NLP?

• Automated understanding of natural language input

• Coherent generation of natural language output
NLP Applications

Machine Translation:

Question Answering:

Personal Assistants:

Specialized Applications:

Legal Documents
Health Records
Business Intelligence
Customer Research
Modern NLP

A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test

How I’m using AI to write my next novel

Meet GPT-3. It Has Learned to Code (and Blog and Argue).
The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.

https://www.stateof.ai/
Trustworthy NLP systems?

I ate my avocado for lunch in the cafeteria.

J'ai mangé mon avocat pour le déjeuner à la cafétéria.

The girl spilt orange juice on herself and started crying.

"Why are you crying?" her dad asked.
"Because my clothes are wet", replied the girl.
"And why are they wet?"

"Because I fell in the swimming pool."
"And why did you fall in the swimming pool?"

"Because I couldn't see the water", the girl replied.

The moral of the story is:
Always wear a blindfold when you go swimming.

https://twitter.com/VeredShwartz/status/1370178821916266501/photo/1
Trustworthy NLP systems?

This is your machine learning system?

Yup! You pour the data into this big pile of linear algebra, then collect the answers on the other side.

What if the answers are wrong?

Just stir the pile until they start looking right.

I ate my avocado for lunch in the canteen.

I ate my lawyer for lunch in the cafeteria.
Trustworthy NLP systems

Want predictions that are:

• grounded to real world dynamics
• made for the right reasons

Need to provide models with:

• domain knowledge
• reasoning abilities
• common sense

https://xkcd.com/1838/
Outline

Language Models as KG Representations

BRS MCC - ACL 2019
HBLDSBC - AAAI 2021
JBBC - NAACL 2021
DLLCB - AKBC 2021

Language Models with Graph Reasoning

BLC - AAAI 2021
YRBL - NAACL 2021
ALBCM - EMNLP 2021
ZBYRML - In Prep
Language Models as Knowledge Engines

https://demo.allennlp.org/next-token-lm
<table>
<thead>
<tr>
<th>Sentence:</th>
<th>Predictions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>I wanted to learn to sail, so I bought</td>
<td><strong>Prediction</strong></td>
</tr>
<tr>
<td></td>
<td>I wanted to learn to dance, so I bought a <strong>dress. I had never ...</strong></td>
</tr>
<tr>
<td></td>
<td>I wanted to learn to dance, so I bought a <strong>$10,000 house ...</strong></td>
</tr>
<tr>
<td></td>
<td>I wanted to learn to dance, so I bought a <strong>censored</strong></td>
</tr>
<tr>
<td></td>
<td>I wanted to learn to dance, so I bought a <strong>US$10 pair of ...</strong></td>
</tr>
<tr>
<td></td>
<td>I wanted to learn to dance, so I bought a <strong>little kit and took part ...</strong></td>
</tr>
</tbody>
</table>

https://demo.allennlp.org/next-token-lm
Problem

- **Problem:** Language models encode commonsense knowledge *implicitly* from pretraining, but its representation *is not robust or reliable*
Can we use knowledge graphs to teach language models to more effectively represent knowledge?
Learning Structure of Knowledge

head entity

person sails across oceans

relation

<requires>

tail entity

buy a boat

(entity to generate)
Learning Structure of Knowledge

Given a **head entity** and a **relation**, learn to generate the **tail entity**

\[
\mathcal{L} = - \sum \log P(\text{tail words} | \text{head words}, \text{relation})
\]

_Bosselut, Rashkin, Sap, Malaviya, Celikyilmaz, and Choi (ACL 2019)_
Learning Structure of Knowledge

Language Model → Knowledge Model: 
Hypothesizes knowledge in the structure of the examples used for training

Bosselut, Rashkin, Sap, Malaviya, Celikyilmaz, and Choi (ACL 2019)
Generate knowledge graph
knowledge for any input concept
**Solution**

- **Problem:** Language models encode commonsense knowledge *implicitly* from pretraining, but its representation is *not robust or reliable*.

- **Solution:** Use knowledge graphs to *explicitly teach commonsense relationships*, allowing LMs to generalise relationships to a broader set of implicitly encoded concepts.
Kai knew that things were getting out of control and managed to keep his temper in check.

Static vs. Dynamic Knowledge

Kai intended to be calm
Kai is viewed as cautious
Kai wants to avoid trouble
Kai stays calm

Link to static Knowledge Graph

Generate dynamic graph with GraphLM

context-free knowledge

X keeps ___ under control
X keeps X's temper
X sweats
X avoids a fight
X wants to show strength
X keeps X's ___ in check

contextual knowledge

no linking

Bosselut et al., AAAI 2021
Applications

Sarcasm generation
Chakrabarty et al. 2020 @ ACL 2020

Therapy Chatbots
Kearns et al. 2020 @ CHI EA 2020

Personalized Dialogue
Majumder et al. 2020 @ EMNLP 2020

Simile generation
Chakrabarty et al. 2020 @ EMNLP 2020

Text-Based Games
Dambekodi et al. 2020 @ arXiv:2012.02757

Automated Storytelling
Ammanabrolu et al. 2021 @ AAAI 2021

COMET
Bosselut et al. 2019 @ ACL 2019
Outline

Language Models as Graph Representations

BRSMCC - ACL 2019
HBLDSBC - AAAI 2021
JBBC - NAACL 2021
DLLCB - AKBC 2021

Language Models with Graph Reasoning

BLC - AAAI 2021
YRBLL - NAACL 2021
ALBCM - EMNLP 2021
ZBYRML - In Prep
Neural Reasoning

If it is not used for hair, a round brush is an example of what?
A. hair brush  B. bathroom  C. art supplies*  D. shower

What does the LM think about round brushes?

- Is used to
  - paint a picture
  - brush hair with
  - brush the hair
  - paint a portrait
If it is not used for hair, a round brush is an example of what?

A. hair brush  B. bathroom  C. art supplies*  D. shower

What does the LM think about round brushes?

- Is used to
  - paint a picture
  - brush hair with
  - brush the hair
  - paint a portrait
Deep learning models exploit biases (Bolukbasi et al., 2016), annotation artifacts (Gururangan et al., 2018), surface patterns (Li & Gauthier, 2017), etc.

They do not learn viable reasoning capabilities

“All the impressive achievements of deep learning amount to just curve fitting”

(Pearl, 2018)
If it is not used for hair, a round brush is an example of what?

A. hair brush  B. bathroom  C. art supplies*  D. shower
How can we coordinate the reasoning implied by language with explicit knowledge from graphs?
If it is **not** used for hair, a round brush is an example of what?

- A. hair brush
- B. bathroom
- C. art supplies*
- D. shower
If it is not used for hair, a round brush is an example of what?

A. hair brush  B. bathroom  C. art supplies*  D. shower
If it is not used for hair, a round brush is an example of what?

A. hair brush  B. bathroom  C. art supplies*  D. shower

**Joint graph** representation that brings language and knowledge together in a shared semantic space.
If it is **not** used for hair, a **round brush** is an example of what?

A. hair brush  
B. bathroom  
C. **art supplies***  
D. shower

**Joint graph** representation that brings language and knowledge together in a shared semantic space.
If it is not used for hair, a round brush is an example of what?
A. hair brush  B. bathroom  C. art supplies*  D. shower
If it is not used for hair, a round brush is an example of what?
A. hair brush  B. bathroom  C. art supplies*  D. shower

Joint reasoning over language and KG through multiple layers of GNN forward pass.
If it is **not** used for **hair**, a **round brush** is an example of what?

A. hair brush  
B. bathroom  
C. art supplies*  
D. shower
Solution

• **Problem:** Machine learning models *shortcut reasoning* by learning to exploit easier statistical patterns (e.g., biases)

• **Solution:** *Jointly model* language and knowledge in a *shared graph structure*, enabling *interaction* between modalities
Experimental Setup

**CommonsenseQA:** reasoning with commonsense knowledge

**OpenBookQA:** reasoning with elementary science knowledge

What do people typically do while playing guitar?

- (A) cry
- (B) hear sounds
- **(C) singing**
- (D) arthritis
- (E) making music

Which of these would let the most heat travel through?

- (A) a new pair of jeans
- **(B) a steel spoon in a cafeteria**
- (C) a cotton candy at a store
- (D) a calvin klein cotton hat
Experimental Results

**CommonsenseQA**

- RoBERTa (Liu, 2019)
- RelNet (Santoro, 2017)
- MHGRN (Feng, 2020)
- QA-GNN (Ours)

**OpenbookQA**

- AristoRoBERTa (Liu, 2019)
- RelNet (Santoro, 2017)
- MHGRN (Feng, 2020)
- QA-GNN (Ours)
Structured Reasoning - Negation

Original Question

If it is not used for hair, a round brush is an example of what?
A. hair brush  B. art supplies*

GNN 1st Layer  GNN Final Layer  Our GNN Prediction

The attention weight by Z over hair decreases and increases on round brush, The attention weight by round brush over painting increases.
Structured Reasoning - Negation

Original Question
If it is **not** used for hair, a **round brush** is an example of what?
A. hair brush  B. art supplies*  

Negation Removed
If it is used for **hair**, a **round brush** is an example of what?  
A. hair brush  B. art supplies

GNN 1st Layer

- **Z**
- **hair**
- **round brush**
- **painting**
- **art supply**

GNN Final Layer

- **Z**
- **hair**
- **round brush**
- **painting**
- **art supply**

A. hair brush (#1)
B. art supplies (#2)

Our GNN Prediction

A. hair brush (#2)
B. art supplies (#1)

The attention weight by **Z** over **hair** decreases and increases on **round brush**, The attention weight by **round brush** over **painting** increases.
If it is not used for hair, a round brush is an example of what?

A. hair brush  B. art supplies

*Entity Changed (hair → art)*

If it is not used for art, a round brush is an example of what?

A. hair brush  B. art supplies

The attention weight by Z on art becomes low in the final layer of the GNN. Round brush attends to hair brush now.
Impact

• **Problem:** Machine learning and symbolic methods *not* individually robust enough for language-based reasoning

• **Solution:** Jointly model language and knowledge in a *shared graph structure*, enabling *interaction* between modalities

• **Impact:** Knowledge graph becomes a *scaffold for neural reasoning* and natural language understanding
Outline

Language Models as Graph Representations

Language Models with Graph Reasoning

BRSMCC - ACL 2019
HBLDSBC - AAAI 2021
JBBC - NAACL 2021
DLLCB - AKBC 2021

BLC - AAAI 2021
YRBL - NAACL 2021
ALBCM - EMNLP 2021
ZBYRML - In Prep
Final Thoughts

• **Representation:** Learning **knowledge relationships** from KG examples enables pretrained language models to **generalise those relationships** to implicitly encoded knowledge learned during pretraining.

• **Reasoning:** **Interfacing language models with graph structure** allows language models to **learn scaffolds** for robust neural reasoning.
References


• Additional work in progress!

atcbosselut.github.io
Collaborators