Neural Language Models as Commonsense Representation Engines

Antoine Bosselut
Language Models

- **The New York Times**
  - *Finally, a Machine That Can Finish Your Sentence*
  - Completing someone else's thoughts
  - But new systems are starting to mimic natural language

- **The New Yorker**
  - *The Next Word*
  - Will predictive text take us?

- **The New York Times**
  - *A Breakthrough for A.I.*
  - Technology: Passing an 8th-Grade Spelling Test

- **Vox**
  - *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.

---

- **Switch Transformer**
  - ~1 T
  - (Jan 2021)

- **GPT3**
  - 175B
  - (July 2020)
Limitations of Symbolic CSKGs

- Insufficient Coverage
- Limited expressivity
- No Contextualization
Kai knew that things were getting out of control and managed to keep his temper in check.

Limitations of Symbolic CSKGs
Limitations of Symbolic CSKGs

- Situations rarely found \textit{as-is} in commonsense knowledge graphs

\begin{itemize}
  \item (X goes to the mall, Effect on X, buys clothes)
  \item (X goes the mall, Perception of X, rich)
  \item (X gives Y some money, Reaction of Y, grateful)
\end{itemize}

\textbf{ATOMIC}

(Sap et al., 2019)
Limitations of Symbolic CSKGs

- Situations rarely found as-is in commonsense knowledge graphs
- Connecting to knowledge graphs can yield incorrect nodes

Kai knew that things were getting out of control and managed to keep his temper in check.

X keeps X’s temper

X keeps ___ under control

X keeps X’s ___ in check
Limitations of Symbolic CSKGs

- Situations rarely found **as-is** in commonsense knowledge graphs
- Connecting to knowledge graphs can yield **incorrect** nodes

Kai knew that things were getting out of control and managed to keep his temper in check
Limitations of Symbolic CSKGs

- Situations rarely found as-is in commonsense knowledge graphs
- Connecting to knowledge graphs can yield incorrect nodes
- Suitable nodes are often uncontextualized

Kai knew that things were getting out of control and managed to keep his temper in check.
Limitations of Symbolic CSKGs

- Situations rarely found as-is in commonsense knowledge graphs
- Connecting to knowledge graphs can yield incorrect nodes
- Suitable nodes are often uncontextualized

Kai knew that things were getting out of control and managed to keep his temper in check.
Challenge

How do we provide machines with large-scale commonsense knowledge?
Constructing Knowledge Graphs

Observe world → Write commonsense knowledge facts → Store facts in knowledge graph

(person, CapableOf, buy)

(Miller, 1995) (Singh et al., 2002) (Lenat, 1995) (Sap et al., 2019)
Commonsense knowledge is immeasurably vast, making it impossible to manually enumerate.
Extracting Knowledge Graphs from Text

Gather Textual Corpus

Automatically extract knowledge

Store in knowledge graph

Underspecified Reporting Bias

John went to the grocery store to buy some steaks. He was going to prepare dinner for his daughter’s birthday. She was turning 5 and would be starting elementary school soon.

(Banko et al., 2007)
(Zhang et al., 2020)

(Speer et al., 2017)
(Tandon et al., 2019)
Learning Relations from Existing KGs

Gather training set of knowledge tuples

Learn relationships among entities

Predict new relationships

Store in knowledge graph

(person, CapableOf, buy)

(Socher et al., 2013)
(Bordes et al., 2013)
(Riedel et al., 2013)
(Toutanova et al., 2015)
(Yang et al., 2015)
(Trouillon et al., 2016)
(Nguyen et al., 2016)
(Dettmers et al., 2018)
Commonsense KGs are Different

<table>
<thead>
<tr>
<th>Knowledge Graph</th>
<th># Entities</th>
<th># Edges</th>
<th>Average Fan-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConceptNet - 100k</td>
<td>78088</td>
<td>100000</td>
<td>1.25</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>256570</td>
<td>610536</td>
<td>2.25</td>
</tr>
<tr>
<td>FB15k-237</td>
<td>14505</td>
<td>272115</td>
<td>16.98</td>
</tr>
</tbody>
</table>

Knowledge base completion assumes explicit connectivity

Malaviya et al., AAAI 2020
## Commonsense KGs are Different

<table>
<thead>
<tr>
<th>Knowledge Graph</th>
<th># Entities</th>
<th># Edges</th>
<th>Average Fan-in</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConceptNet - 100k</td>
<td>78088</td>
<td>100000</td>
<td>1.25</td>
</tr>
<tr>
<td>ATOMIC</td>
<td>256570</td>
<td>610536</td>
<td>2.25</td>
</tr>
<tr>
<td>FB15k-237</td>
<td>14505</td>
<td>272115</td>
<td>16.98</td>
</tr>
</tbody>
</table>

Knowledge base completion assumes explicit connectivity.

Malaviya et al., AAAI 2020
Challenges

• Commonsense knowledge is **immeasurably vast**, making it impossible to manually enumerate.

• Commonsense knowledge is often implicit or underspecified, and therefore cannot be directly extracted from text.

• Commonsense knowledge resources are quite sparse, making them difficult to extend by only learning from examples.

**How else can we learn commonsense knowledge at scale?**
Deep Language Models

Input Vector

Stacked Transformer Blocks

Predict Next Word

Allen sailed across oceans in bought a boat <END>

\[
\begin{align*}
& h_0 \\
& h_1 \\
& h_2 \\
& h_3 \\
& h_{T-3} \\
& h_{T-2} \\
& h_{T-1} \\
& h_T \end{align*}
\]

(Radford et al., 2018, 2019, many others)
Learning in Language Models

Text Corpus

Transformer Language Model

Used to Learn

(Radford et al., 2018, 2019, many others)
Do language models have commonsense knowledge?
Knowledge in Language Models

Sentence: I wanted to learn to sail, so I bought a...

Predictions:
- 14.2% boat
- 5.4% sail
- 2.6% new
- 2.0% small
- 1.4% canoe

Sentence: I wanted to learn to drive, so I bought a...

Predictions:
- 7.5% new
- 7.6% car
- 1.7% Honda
- 1.7% BMW
- 1.3% Ford

Sentence: I wanted to learn to read, so I bought a...

Predictions:
- 17.2% book
- 15.2% copy
- 3.4% Kindle
- 2.4% new
- 1.7% few

Sentence: I wanted to learn to fly, so I bought a...

Predictions:
- 5.5% plane
- 3.8% new
- 1.6% small
- 1.6% Boeing
- 1.5% jet

https://demo.allennlp.org/next-token-lm
Knowledge Prompting

(Dante, <born_in>, ?)

map relation to one or more natural language prompts

“Dante was born in [MASK].”

Neural LM Memory Access

Florence

e.g. ELMo/BERT

Petroni et al., EMNLP 2019
Instance:
Christmas was a special holiday to Eric but not Adam since ____ was a Jew.

Question Generation:
Christmas was a special holiday to Eric but not Adam since ____ was a Jew.

Answer Generation:

Clarification Question:
What is the definition of Christmas?

Clarification:
The definition of Christmas is the celebration of the birth of Christ. The purpose of Christmas is to celebrate the birth of Christ. The definition of Christmas is a Christian Holiday.
### Prompt Sensitivity

<table>
<thead>
<tr>
<th>Prompts</th>
<th>$y_{\text{man}}$</th>
<th>$y_{\text{mine}}$</th>
<th>$y_{\text{para}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>manual</td>
<td>Intel -1.06</td>
<td>Microsoft -1.77</td>
<td>Microsoft -2.23</td>
</tr>
<tr>
<td>mined</td>
<td>Microsoft -2.21</td>
<td>They -2.43</td>
<td>Intel -2.30</td>
</tr>
<tr>
<td>paraphrased</td>
<td>IBM -2.76</td>
<td>It -2.80</td>
<td>default -2.96</td>
</tr>
<tr>
<td></td>
<td>Google -3.40</td>
<td>Sega -3.01</td>
<td>Apple -3.44</td>
</tr>
<tr>
<td></td>
<td>Nokia -3.58</td>
<td>Sony -3.19</td>
<td>Google -3.45</td>
</tr>
</tbody>
</table>

**Jiang et al., TACL 2020**
### Prompt Sensitivity

**Prompts**

<table>
<thead>
<tr>
<th>manual</th>
<th>DirectX is developed by $y_{\text{man}}$</th>
<th>mined</th>
<th>DirectX is released by $y_{\text{mine}}$</th>
<th>paraphrased</th>
<th>DirectX is created by $y_{\text{para}}$</th>
</tr>
</thead>
</table>

**Top 5 predictions and log probabilities**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Company</th>
<th>$y_{\text{man}}$</th>
<th>Company</th>
<th>$y_{\text{mine}}$</th>
<th>Company</th>
<th>$y_{\text{para}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Intel</td>
<td>-1.06</td>
<td>Microsoft</td>
<td>-1.77</td>
<td>Microsoft</td>
<td>-2.23</td>
</tr>
<tr>
<td>2</td>
<td>Microsoft</td>
<td>-2.21</td>
<td>They</td>
<td>-2.43</td>
<td>Intel</td>
<td>-2.30</td>
</tr>
<tr>
<td>3</td>
<td>IBM</td>
<td>-2.76</td>
<td>It</td>
<td>-2.80</td>
<td>default</td>
<td>-2.96</td>
</tr>
<tr>
<td>4</td>
<td>Google</td>
<td>-3.40</td>
<td>Sega</td>
<td>-3.01</td>
<td>Apple</td>
<td>-3.44</td>
</tr>
<tr>
<td>5</td>
<td>Nokia</td>
<td>-3.58</td>
<td>Sony</td>
<td>-3.19</td>
<td>Google</td>
<td>-3.45</td>
</tr>
</tbody>
</table>

**Candidate Sentence $S_i$**

<table>
<thead>
<tr>
<th>Sentence</th>
<th>log $p(S_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>“musician can playing musical instrument”</td>
<td>−5.7</td>
</tr>
<tr>
<td>“musician can be play musical instrument”</td>
<td>−4.9</td>
</tr>
<tr>
<td>“musician often play musical instrument”</td>
<td>−5.5</td>
</tr>
<tr>
<td>“a musician can play a musical instrument”</td>
<td><strong>−2.9</strong></td>
</tr>
</tbody>
</table>

*Jiang et al., TACL 2020*

*Feldman et al., EMNLP 2019*
Context Sensitivity

DirectX is created by

<table>
<thead>
<tr>
<th>Company</th>
<th>Log Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft</td>
<td>-1.77</td>
</tr>
<tr>
<td>Intel</td>
<td>-2.43</td>
</tr>
<tr>
<td>Sega</td>
<td>-3.01</td>
</tr>
<tr>
<td>Sony</td>
<td>-3.19</td>
</tr>
<tr>
<td>Google</td>
<td>-3.40</td>
</tr>
</tbody>
</table>

Prompts
- “musician can playing musical instrument”
- “musician can be play musical instrument”
- “musician often play musical instrument”
- “a musician can play a musical instrument”

Candidate Sentence $S_i$ | $\log p(S_i)$
--- | ---
“musician can playing musical instrument” | $-5.7$
“musician can be play musical instrument” | $-4.9$
“musician often play musical instrument” | $-5.5$
“a musician can play a musical instrument” | $-2.9$

Weir et al., CogSci 2020

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Model Predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ___ has fur.</td>
<td>dog, cat, fox, ...</td>
</tr>
<tr>
<td>A ___ has fur, is big, and has claws.</td>
<td>cat, bear, lion, ...</td>
</tr>
<tr>
<td>A ___ has fur, is big, has claws, has teeth, is an animal, eats, is brown, and lives in woods.</td>
<td>bear, wolf, cat, ...</td>
</tr>
</tbody>
</table>

Jiang et al., TACL 2020

Feldman et al., EMNLP 2019

Weir et al., CogSci 2020
Do language models have commonsense?

• Distinction between *encoding* commonsense knowledge and *expressing* commonsense knowledge
Do language models have commonsense?

- Distinction between **encoding** commonsense knowledge and **expressing** commonsense knowledge

- Probing with prompts measures whether LMs can **express** commonsense knowledge and the results are **mixed**
Do language models encode commonsense knowledge?
From Unstructured to Structured Knowledge

Transformer Language Model

Used to Communicate?
Transformer Language Models

Allen sailed across oceans in boat

(Radford et al., 2018, 2019, many others)
Transformer Language Models

- Trained to generate the next word given a set of preceding words
Transformer Language Models

- Trained to generate the next word given a set of preceding words
- Follow-up tokens can be generated using generated tokens as input

Language Model

Allen sailed across oceans ... He bought a boat
Transformer Language Models

- Trained to generate the next word given a set of preceding words
- Follow-up tokens can be generated using generated tokens as input
Structure of Knowledge Tuple

- **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**

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

**Language Model**

- **person**
- **sails**
- **across**
- **oceans**
- **<requires>**
- **buy**
- **a**
- **boat**

Bosselut et al., ACL 2019
Learning Structure of Knowledge

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

Bosselut et al., ACL 2019
Generate commonsense knowledge for any input concept
Generate commonsense knowledge for any input concept

COMmonsEnse Transformers

Bosselut et al., ACL 2019
Antoine gives a tutorial on commonsense knowledge.

Before, X needed to be a teacher.

X perceived as smart.

Others will want to ask questions.

Others will want to gain knowledge.

Now, none want to gain knowledge.
Audience listens to tutorial

- classroom
- motivated by knowledge
- starts with listen to teacher
- ends with applaud
- causes they learn
Why does this work?
Transfer Learning from Language

mango → is a → fruit

Bosselut et al., ACL 2019
Transfer Learning from Language

mango \rightarrow \text{is a} \rightarrow \text{fruit}

mango \rightarrow \text{used for} \rightarrow \text{salsa}

Bosselut et al., ACL 2019

ConceptNet
Transfer Learning from Language

mango → is a fruit

mango → used for salsa

mango → Same Model, Not Pretrained on language

mango → ?

Bosselut et al., ACL 2019
Transfer Learning from Language

mango → is a fruit

mango → used for salsa

mango → Same Model, Not Pretrained on language → is a spice

Bosselut et al., ACL 2019
Can’t language models just do this?
Do Language Models know this?

Sentence:

mango is a

Predictions:

2.1% great
1.9% very
1.2% new
1.0% good
1.0% small
← Undo

https://demo.allennlp.org/next-token-lm
Do Language Models know this?

Sentence:

mango is a

Predictions:

2.1% great
1.9% very
1.2% new
1.0% good
1.0% small
← Undo

4.2% good
4.0% very
2.5% great
2.4% delicious
1.8% sweet
← Undo

a mango is a

https://demo.allennlp.org/next-token-lm
Do Language Models know this?

Sentence:

mango is a

Predictions:
- 2.1% great
- 1.9% very
- 1.2% new
- 1.0% good
- 1.0% small
- ← Undo

Sentence:
a mango is a

Predictions:
- 4.2% good
- 4.0% very
- 2.5% great
- 2.4% delicious
- 1.8% sweet
- ← Undo

Sentence:

A mango is a

Predictions:
- 4.2% fruit
- 3.5% very
- 2.5% sweet
- 2.2% good
- 1.5% delicious
- ← Undo
Do Masked Language Models know this?

Sentence:

mango is a [MASK].

Mask 1 Predictions:
69.7% .
9.3% ;
1.7% !
0.8% vegetable
0.7% ?

Sentence:

mango is a [MASK].

Mask 1 Predictions:
7.6% staple
7.6% vegetable
4.6% plant
3.5% tree
3.5% fruit

Sentence:

A mango is a [MASK].

Mask 1 Predictions:
16.0% banana
12.1% fruit
5.9% plant
5.5% vegetable
2.5% candy

https://demo.allennlp.org/masked-lm
Let’s talk about the elephant in the room.
What about GPT-3?
Does GPT-3 have commonsense knowledge?

Q: What is your favorite animal?  
A: My favorite animal is a dog.

Q: Why?  
A: Because dogs are loyal and friendly.

Q: How many eyes does a giraffe have?  
A: A giraffe has two eyes.

Q: Why don't animals have three legs?  
A: Animals don't have three legs because they would fall over.
COMET (BART): 435x smaller model (~400M params), informed by ATOMIC\textsuperscript{20}

GPT-3 (Few Shot): 175B parameters!! pre-trained with a ton of web text (~500B tokens)

Hwang et al., AAAI 2021
Commonsense Transformers

- Learn implicit knowledge at scale from language models and web-scale text

Pre-trained Language Model
Commonsense Transformers

- Learn implicit knowledge at scale from language models and web-scale text
- Learn explicit structure of knowledge from symbolic knowledge graphs

Pre-trained Language Model + Seed Knowledge Graph Training
Commonsense Transformers

- Learn implicit knowledge at scale from language models and web-scale text
- Learn explicit structure of knowledge from symbolic knowledge graphs
- Resulting knowledge model generalizes structure to other concepts
What are the implications of representing commonsense knowledge in this manner?
transformer-style architecture — input format is natural language

- event can be fully parsed

Kai knew that things were getting out of control and managed to keep his temper in check
Commonsense Knowledge for any Situation

- transformer-style architecture — input format is natural language
  - event can be fully parsed
  - knowledge generated *dynamically* from neural knowledge model

Kai knew that things were getting out of control and managed to keep his temper in check

Kai wants to avoid trouble
Kai intends to be calm
Kai stays calm
Kai is viewed as cautious
Kai knew that things were getting out of control and managed to keep his temper in check.

Reasoning with Knowledge Graphs
Commonsense Knowledge for any Situation

- transformer-style architecture — input format is natural language
  - event can be fully parsed
  - knowledge generated *dynamically* from *neural* knowledge model

Kai knew that things were getting out of control and managed to keep his temper in check

Kai wants to avoid trouble
Kai intends to be calm
Kai stays calm
Kai is viewed as cautious
Kai knew that things were getting out of control and managed to keep his temper in check.

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

Bosselut et al., AAAI 2021
Kai knew that things were getting out of control and managed to keep his temper in check.

- Kai wants to avoid trouble.
- Kai intends to be calm.
- Kai stays calm.
- Kai is viewed as cautious.

Dynamic Construction of Knowledge Graphs

Root Node

Kai knew that things were getting out of control and managed to keep his temper in check.

Kai wants to avoid trouble.

Kai intends to be calm.

Kai stays calm.

Kai is viewed as cautious.

Bosselut et al., AAAI 2021
Kai wants to avoid trouble
Kai intends to be calm
Kai stays calm
Kai is viewed as cautious

Dynamic Construction of Knowledge Graphs

Generated Commonsense Inference Nodes

Kai knew that things were getting out of control and managed to keep his temper in check
Kai knew that things were getting out of control and managed to keep his temper in check.
Kai knew that things were getting out of control and managed to keep his temper in check.

Kai wants to avoid trouble.

Kai intends to be calm.

Kai stays calm.

Kai is viewed as cautious.

$\ell = 1$
Kai knew that things were getting out of control and managed to keep his temper in check.

Kai wants to avoid trouble.

Kai intends to be calm.

Kai stays calm.

Kai is viewed as cautious.

Kai then avoids trouble.

Kai wants to be safe.

Kai feels relieved.

Others want to avoid trouble.

\[ \ell = 1 \]

\[ \ell = 2 \]
Kai knew that things were getting out of control and managed to keep his temper in check. Kai wants to avoid trouble. Kai intends to be calm. Kai stays calm. Kai is viewed as cautious.
Kai knew that things were getting out of control and managed to keep his temper in check. Kai wants to avoid trouble. 

Kai intends to be calm. 

Kai stays calm. 

Kai is viewed as cautious. 

\( \ell = 1 \)

\( \ell = L \)

relieved

scared

anxious

Bosselut et al., AAAI 2021
What about the relations? Those are fixed.
Fixed Relation Sets

Number of Relations

<table>
<thead>
<tr>
<th>Set</th>
<th>Number of Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOMIC</td>
<td>9</td>
</tr>
<tr>
<td>ConceptNet</td>
<td>34</td>
</tr>
<tr>
<td>ATOMIC 2020</td>
<td>23</td>
</tr>
<tr>
<td>TransOMCS</td>
<td>20</td>
</tr>
</tbody>
</table>
Few-shot Knowledge Models

Using prompts induces rapid knowledge model adaptation in T5!

PersonX goes to the mall <xl Intent> to buy clothes

PersonX goes to the mall because they wanted to buy clothes

Da et al., 2021
Few-shot Knowledge Models

Human Judgment - Plausibility

- COMET (T5) - full: 84.6
- COMET (T5) - few shot: 78.6
- GPT3: 73.0

Just ~100 examples!
5 examples / relation
Can we model more complex commonsense knowledge?
Path Knowledge Models

PersonX saw a fight was breaking out <xWant> wants to avoid trouble <xEffect> leaves

Wang et al., EMNLP Findings 2020
PersonX saw a fight was breaking out. The Language Model wants to avoid trouble, which leads to the tail entity leaving. Wang et al., EMNLP Findings 2020.
Path Knowledge Models

Multi-step Knowledge Model

Path entities

<xWant> wants to avoid trouble <xEffect> leaves

Head entity
PersonX saw a fight was breaking out

Tail entity

Wang et al., EMNLP Findings 2020
Visual Commonsense Knowledge Models

Park et al., ICCV 2020

Person 1

- Realize the ship is sinking.
- Start moving against the water.
- Get caught in a rush of water.

Before Person 1 needed to ...

- Because Person 1 wanted to ...
- After Person 1 will most likely ...

Person 2

- Try to help [Person 2].
- Save himself from drowning.
- Wait for help to arrive.

Before Person 2 needed to ...

- Because Person 2 wanted to ...
- After Person 2 will most likely ...

- Notice water washing in.
- Swim towards the statute.
- Sense his own death.
- Be washed away.
- Scream for help.

- Get to the top of the deck.
- Swim to safety.
- Sink in the water.

- Gasp for air.
- Scream for help.
- Swim towards the statute.
- Notice water washing in.
VisualCOMET

Language Model

Tail entity

Tail entity
gasp
for
air
<END>

ROI Feature
ROI Feature

Visual Context

Head entity

<Person2> is holding onto a ...

Gasp
for
air

<relation>

Park et al., ICCV 2020
VisualCOMET

Multimodal Knowledge Model

Visual Context

Head entity

<relation>

Tail entity

gasp

for

air

<EOLD>

Park et al., ICCV 2020
Okay, what’s the catch?
Limitations of Knowledge Models

- **Base Self-supervised Model**
  - biases, history

- **Seed Knowledge Graph**
  - bias, language, relations, schema
Limitations of Knowledge Models

- Base Self-supervised Model
- biases, history

- Seed Knowledge Graph
- bias, language, relations, schema

- Generation Algorithm
- diversity, mode collapse
Sarcasm generation
Chakrabarty et al. 2020 @ ACL 2020

Therapy Chatbot
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
Where else can commonsense knowledge access improve our systems?
Kai knew that things were getting out of control and managed to keep his temper in check.

- X keeps ___ under control
- X sweats
- X avoids a fight
- X keeps X’s temper
- X wants to show strength
- X wants to avoid trouble
- X wants to show strength
- X keeps X’s ___ in check

**Link to static Knowledge Graph**

**Generate dynamic graph with COMET**

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

**contextual knowledge**

**no linking**

**bad links**

**context-free knowledge**

**Bosselut et al., AAAI 2021**
Language Models as CSKBs?

• Large-scale language models encode a lot of commonsense knowledge implicitly, but it’s not not directly accessible

• We can develop methods to extract it, but they need to be adaptable, robust and efficient

• Need to rethink how we design commonsense knowledge graphs if transfer from language models is a new use case
Building Commonsense Knowledge Bases

Quality of KB is important!
Careful validation is critical so that LMs can learn from precise examples.

More varieties of relations!
Represent a wide range of commonsense relationships (Use prompts for fast adaptation)

Focus on textually less explicit examples
These are less likely to be known by LMs, thus more impactful in knowledge transfer

Hwang et al., AAAI 2021
References & Additional Reading


