Event and Commonsense
Event-centric Natural Language Understanding (Part IV)

Hongming Zhang
HKUST/UPenn

Feb 2021
AAAII Tutorials
Event-centric Natural Language Understanding
Commonsense is crucial for NLU

Example: John stepped in a puddle and had to go home to change.
Outline

- Understanding Commonsense from the Angle of Events

- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling

- Schema-level Event Knowledge Acquisition

- Conclusion
Commonsense Knowledge

- Modern Definition of Commonsense Knowledge (Liu & Singh, 2004)
  - “While to the average person the term ‘commonsense’ is regarded as synonymous with ‘good judgement’”
  - “the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people.”
  - “Commonsense is about preference and not always true”
    - If you forget someone’s birthday, they may be unhappy with you.
    - But if your friends understand that you are busy, he will not be angry.

Unlike factual knowledge, they are not inevitably true.

Commonsense is about preference.
What kinds of preference?

- Semantic meaning in our language can be described as “a finite set of mental primitives and a finite set of mental combination.” (Jackendoff, 1990)
- The primitive units of semantic meanings include
  - Thing (or entity)
    - cat
  - State
    - The cat is cute.
    - The cat is smiling.
  - Event
    - The cat is running.

States describe things.
Events describe the changing of states.

We want to understand humans’ preferences about things, states, and events.

How to represent the preference?

- The lower bound of a semantic theory (Katz and Fodor, 1963)
  - Linguistic description – grammar = semantics
  - Understanding language needs both “the speaker’s knowledge of his language and his knowledge about world” (Katz and Fodor, 1963)

Selectional Preference

- Selectional Preference (Resnik, 1993)
  - A relaxation of selectional restrictions (Katz and Fodor, 1963) and is often used as syntactic features (Chomsky, 1965).
  - Applied to IsA hierarchy in WordNet and verb-object relations.
  - With this formulation, we can easily use the frequency/plausibility scores of different combinations to reflect humans’ preference.

- Examples:
  - ("Cat" -IsA- "Animal") > ("Cat" -IsA- "Plant")
  - ("eat" -dobj- "food") > ("eat" -dobj- "rock")
Higher-order Selectional Preference

- **First-order**
  - dobj: (“eat”->dobj->“food”) > (“eat”->dobj->“house”)
  - Nsubj: (“sing”->nsubj->“singer”) > (“sing”->nsubj->“house”)
  - ...

- **Second-order (Zhang et al., 2019)**
  - Nsubj-amod / dobj-amod
  - (“eat”->nsubj->“[SUB]”->amod->“hungry”) > (“eat”->dobj->“[OBJ]”->amod->“hungry”)

- **Higher-order**
  - (“I eat dinner”->Causes->“I am full”) > (“I eat dinner”->Causes->“I am hungry”)

Commonsense can be represented by the higher-order selectional preference over eventualities.

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL 2019.
Transferability from event knowledge to Commonsense

Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. IJCAI 2020.
Transferability from event knowledge to Commonsense

1. stand
2. think
3. die
4. learn
5. make mistake
6. lie
7. typically have
8. create society
9. have cell
10. create life

“human" CapableOf

1. be friendly
2. be happy
3. pain
4. marriage
5. be quaint
6. be unhappy
7. be allergic
8. be desperate
9. be apart
10. be silly

“love” Causes

Event-centric KG

Human-defined commonsense
Outline

- Understanding Commonsense from the Angle of Events

- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling

- Schema-level Event Knowledge Acquisition

- Conclusion
## Event-centric KBs

<table>
<thead>
<tr>
<th>Event-centric KBs</th>
<th># Events</th>
<th># Event relation</th>
<th># Relation Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrameNet (Baker et al., 1998)</td>
<td>27,691</td>
<td>1,709</td>
<td>7</td>
</tr>
<tr>
<td>ACE (Aguilar et al., 2014)</td>
<td>3,290</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PropBank (Palmer et al., 2005)</td>
<td>112,917</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NomBank (Meyers et al., 2004)</td>
<td>114,576</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TimeBank (Pustejovsky et al., 2003)</td>
<td>7,571</td>
<td>8,242</td>
<td>1</td>
</tr>
<tr>
<td>ConceptNet (Liu and Singh, 2004)</td>
<td>74,989</td>
<td>116,097</td>
<td>4</td>
</tr>
<tr>
<td>Event2Mind (Smith et al., 2018)</td>
<td>24,716</td>
<td>57,097</td>
<td>3</td>
</tr>
<tr>
<td>ProPora (Dalvi et al., 2018)</td>
<td>2,406</td>
<td>16,269</td>
<td>1</td>
</tr>
<tr>
<td>ATOMIC (Sap et al., 2019)</td>
<td>309,515</td>
<td>877,108</td>
<td>9</td>
</tr>
<tr>
<td>ATOMIC 2020* (Hwang et al., 2020)</td>
<td>-</td>
<td>165,164</td>
<td>4</td>
</tr>
</tbody>
</table>

**Pro:** High quality  
**Con:** Expensive; Small Scale; Limited relation types

*For ATOMIC 2020, we only count the unique edges and ignore the edges it inherits from other KBs.*
Outline

- Understanding Commonsense from the Angle of Events

- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling

- Schema-level Event Knowledge Acquisition

- Conclusion
Knowlywood (Tandon et al., 2015)

- KG Format
  - Node: Verb + Object
  - Edge: Temporal Relation

- Resource
  - 560 movie scripts

- Extraction Methodology

---

Example

“Knock door”->“open up entrance”->“enter office”

Quantity

<table>
<thead>
<tr>
<th>Source</th>
<th>#Input Scripts</th>
<th>#Scenes</th>
<th>#Unique Activities</th>
<th>Parent</th>
<th>Participant</th>
<th>Prev</th>
<th>Next</th>
<th>Loc.</th>
<th>Time</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie scripts</td>
<td>560</td>
<td>148,296</td>
<td>244,789</td>
<td>0.87</td>
<td>0.86</td>
<td>0.78</td>
<td>0.85</td>
<td>0.79</td>
<td>0.79</td>
<td>0.84</td>
</tr>
<tr>
<td>TV series</td>
<td>290</td>
<td>886,724</td>
<td>565,394</td>
<td>0.89</td>
<td>0.85</td>
<td>0.81</td>
<td>0.84</td>
<td>0.82</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>Sitcoms</td>
<td>179</td>
<td>286,266</td>
<td>200,550</td>
<td>0.88</td>
<td>0.85</td>
<td>0.81</td>
<td>0.87</td>
<td>0.81</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>Novels</td>
<td>103</td>
<td>383,795</td>
<td>137,365</td>
<td>0.84</td>
<td>0.84</td>
<td>0.78</td>
<td>0.88</td>
<td>0.85</td>
<td>0.72</td>
<td>0.84</td>
</tr>
<tr>
<td>Crowdsrc.</td>
<td>25</td>
<td>3,701</td>
<td>9,575</td>
<td>0.82</td>
<td>0.91</td>
<td>0.91</td>
<td>0.87</td>
<td>0.74</td>
<td>0.40</td>
<td>0.86</td>
</tr>
<tr>
<td>Knowlywood</td>
<td>1,157</td>
<td>1,708,782</td>
<td>964,758</td>
<td>0.87</td>
<td>0.86</td>
<td>0.84</td>
<td>0.85</td>
<td>0.78</td>
<td>0.84</td>
<td>0.85±0.01</td>
</tr>
<tr>
<td>ConceptNet 5</td>
<td>-</td>
<td>-</td>
<td>4,757</td>
<td>0.15</td>
<td>0.81</td>
<td>0.92</td>
<td>0.91</td>
<td>0.33</td>
<td>N/A</td>
<td>0.46±0.02</td>
</tr>
</tbody>
</table>
ASER (Zhang et al., 2020)

- **KG Format**
  - A Hybrid graph
  - Node: Eventualities in the format of dependency graphs
  - Edge: All discourse relations

- **Resource**
  - 11B token textual corpora (i.e., Yelp, NYT, Wikipedia, Reddit, Subtitles, E-books)

- **Extraction**

Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. ASER: A Large-scale Eventuality Knowledge Graph. WWW 2020.
ASER Example

A hybrid graph of
- Each eventuality is a hyper-edge of words
- Heterogeneous edges among eventualities

194 million eventualities, 64 million edges
ASER Quantity and Quality (Edge)
Comparison with Other event KGs

PS: In ConceptNet 5.0, more edges are added, but only the core part, which is inherited from ConceptNet 1.0 (Liu & Singh, 2004), is related to commonsense knowledge.
Outline

- Understanding Commonsense from the Angle of Events

- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling

- Schema-level Event Knowledge Acquisition

- Conclusion
Knowledge Discovery from Pre-trained LMs

- Language Model
  - Examples: GPT-1/2/3

- COMET (Bosselut et al., 2019):
  - Commonsense Transformers for Automatic Knowledge Graph Construction

Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: Commonsense Transformers for Automatic Knowledge Graph Construction. ACL 2019.
Event Temporal Commonsense

- TacoLM (Zhou et al., 2020)
  - A general time-aware language model that distinguishes temporal properties in fine-grained contexts.

![Graph showing temporal distinctions]

I moved my chair  I moved my piano  I moved to a different city
Event Temporal Commonsense

**Step 1: Information Extraction**

- Use high-precision patterns to acquire temporal information
  - Unsupervised automatic extraction
- Overcomes reporting biases with a large amount of natural text

**Step 2: Joint Language Model Pre-training**

- Multiple temporal dimensions
  - Duration $\sim 1 / \text{Frequency}$
    - "I brush my teeth every morning"
    - Duration of "brushing teeth" < morning
  - Further generalization to combat reporting biases

**Goal:** build a general time-aware LM with minimal supervision

**Output:** TacoLM - a time-aware general BERT
Event Temporal Commonsense

Original sentence:
I played basketball for 2 hours.

SRL Parse:
- Arg-0: Verb
- Arg-1: Arg-Tmp

Pattern Matching:
- for 2 hours: matches Duration pattern

Information Extraction:
- Event: I played basketball
- Dimension: Duration, Hours

Baseline Model: Pre-trained BERT-base
Main objective: mask some tokens and recover them
How we mask:
- With some probability, mask temporal value while keeping others
- Otherwise, mask a certain portion of E1...En while keeping temporal value unchanged

Max (P(Event|Dim,Val) + P(Val|Event,Dim)); Preserving original LM capability

Joint training with language model
Outline

- Understanding Commonsense from the Angle of Events

- Instance-level Event Knowledge Acquisition
  - Human Annotation
  - Automatic Event Knowledge Extraction
  - Language Modeling

- Schema-level Event Knowledge Acquisition

- Conclusion
Event Graph Schema Induction (Li et al., 2020)

- History repeats itself: Instance graphs (a) and (b) refer to very different event instances, but they both illustrate a same scenario.
- Select salient and coherent paths based on Path Language Model, and merge them into graph schemas.

Path Language Model

- Path Language Model is trained on two tasks
  - Autoregressive Language Model Loss: capturing the frequency and coherence of a single path
  - Neighbor Path Classification Loss: capturing co-occurrence of two paths

![Diagram of Path Language Model](image)

1 = neighbor
0 = not neighbor

Word

<table>
<thead>
<tr>
<th>Word</th>
<th>Position</th>
<th>Segment</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>1 2 3 4 5 6</td>
<td>0 0 0 0 0</td>
<td>1 2 1 2 1 0</td>
</tr>
<tr>
<td>attacker</td>
<td>7 8 9 10 11 12</td>
<td>1 1 1 1 1 1</td>
<td>1 2 1 2 1 0</td>
</tr>
<tr>
<td>GPE</td>
<td>13 14</td>
<td>1 1 1 1 1 1</td>
<td>0</td>
</tr>
<tr>
<td>agent</td>
<td>Transport</td>
<td>[SEP]</td>
<td>Artic</td>
</tr>
</tbody>
</table>

Path A:
Attack - attacker - GPE - agent - Transport

Path B:
Attack - instrument - WEA - artifact - GPE - agent - Transport
A salient schema can serve as a skeleton to recover instance graphs.

- We use each graph schema to match back to each ground-truth instance graph and evaluate their intersection in terms of Precision and Recall.

Precision = \[ \frac{\sum_{s \in S} \sum_{g \in G} |g \cap s|}{\sum_{s \in S} |s|} \]

Recall = \[ \frac{\sum_{s \in S} \sum_{g \in G} |g \cap s|}{\sum_{g \in G} |g|} \]
Use the state-of-the-art IE system OneIE (Lin et al, 2020) to decode converts each input document into an IE graph.

Each path in the graph schema is encoded as a single global feature for scoring candidate IE graphs.

OneIE promotes candidate IE graphs containing paths matching schema graphs.

### Dataset

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Entity</th>
<th>Event Trigger Identification</th>
<th>Event Trigger Classification</th>
<th>Event Argument Identification</th>
<th>Event Argument Classification</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>90.3</td>
<td>75.8</td>
<td>72.7</td>
<td>57.8</td>
<td>55.5</td>
<td>44.7</td>
</tr>
<tr>
<td>+PathLM</td>
<td>90.2</td>
<td>76.0</td>
<td>73.4</td>
<td>59.0</td>
<td>56.6</td>
<td>60.9</td>
</tr>
</tbody>
</table>
Outline

■ Understanding Commonsense from the Angle of Events

■ Instance-level Event Knowledge Acquisition
  □ Human Annotation
  □ Automatic Event Knowledge Extraction
  □ Language Modeling

■ Schema-level Event Knowledge Acquisition

■ Conclusion
Key takeaways

- There is a transferability from event knowledge to commonsense knowledge
- Compared with commonsense, acquiring event knowledge is cheaper and more scalable.
- All existing acquisition systems have advantages and limitations.

<table>
<thead>
<tr>
<th></th>
<th>Quality</th>
<th>Scale</th>
<th>Relation Coverage</th>
<th>Explainability</th>
<th>Robustness</th>
<th>Downstream Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Annotation</td>
<td>High</td>
<td>Small</td>
<td>Middle</td>
<td>High</td>
<td>High</td>
<td>Difficult</td>
</tr>
<tr>
<td>Automatic Event Knowledge Extraction</td>
<td>Middle</td>
<td>Large</td>
<td>High</td>
<td>High</td>
<td>Middle</td>
<td>Difficult</td>
</tr>
<tr>
<td>Language Model</td>
<td>Middle</td>
<td>Large</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Easy</td>
</tr>
</tbody>
</table>
Key References

- **ConceptNet**: Hugo Liu and Push Singh, ConceptNet - a practical commonsense reasoning tool-kit, BTTJ, 2004
- **ATOMIC**: Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, Yejin Choi, ATOMIC: An Atlas of Machine Commonsense for If-Then Reasoning. AAAI 2019
- **COMET**: Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. COMET: commonsense transformers for automatic knowledge graph construction. ACL 2019.
- **ASER**: Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. ASER: A Large-scale Eventuality Knowledge Graph. WWW 2020.
- **TacoLM**: Ben Zhou, Qiang Ning, Daniel Khashabi, Dan Roth. Temporal Common Sense Acquisition with Minimal Supervision. ACL 2020