ASER: Building a Commonsense Knowledge Graph by Higher-order Selectional Preference

Yangqiu Song
Department of CSE, HKUST, Hong Kong
2020.07
Contributors and Acknowledgements

• PhD/MPhil Students
  • Hongming Zhang, Xin Liu, Haojie Pan, Changlong Yu

• Undergraduate Students
  • Hantian Ding, Jiefu Ou, Xinran Zhao

• Other Collaborators
  • Cane Wing-Ki Leung, Daniel Khashabi, Dan Roth, Wilfred Ng, Lifeng Shang
Outline

• Motivation: NLP and commonsense knowledge

• Consideration: selectional preference

• New proposal: large-scale and higher-order selectional preference

• Application on the Winograd Schema Challenge

• Extensions
Natural language conversation requires a lot of commonsense knowledge. Interacting with humans involves a lot of commonsense knowledge:
- Space
- Time
- Location
- State
- Causality
- Color
- Shape
- Physical interaction
- Theory of mind
- Human interactions

Judy Kegl, The boundary between word knowledge and world knowledge, TINLAP3, 1987
Ernie Davis, Building AIs with Common Sense, Princeton Chapter of the ACM, May 16, 2019
Commonsense Knowledge is the Key

• How to define commonsense knowledge? (Liu & Singh, 2004)
  • “While to the average person the term ‘commonsense’ is regarded as synonymous with ‘good judgement’, ”
  • “in the AI community it is used in a technical sense to refer to the millions of basic facts and understandings possessed by most people.”
  • “Such knowledge is typically omitted from social communications”, e.g.,
    • If you forget someone’s birthday, they may be unhappy with you.
How to collect commonsense knowledge?

• **ConceptNet5** (Speer and Havasi, 2012)
  - Core is from *Open Mind Common Sense (OMCS)* (Liu & Singh, 2004)
  
  ![ConceptNet5 Diagram](image)

• Essentially a crowdsourcing based approach + text mining
The Scale

• “A founder of AI, Marvin Minsky, once estimated that ‘...commonsense is knowing maybe 30 or 60 million things about the world and having them represented so that when something happens, you can make analogies with others.’” (Liu & Singh, 2004)

• ConceptNet
  • 2004: 1.6 million relations among 300,000 nodes
  • 2017: 21 million edges over 8 million nodes
    • 1.5 million nodes are English

Slides credit: Haixun Wang
What contribute to ConceptNet5.5 (21 million edges and over 8 million nodes)?

- Facts acquired from Open Mind Common Sense (OMCS) (Singh 2002) and sister projects in other languages (Anacleto et al. 2006)
- Information extracted from parsing Wiktionary, in multiple languages, with a custom parser (“Wikiparsec”)
- “Games with a purpose” designed to collect common knowledge (von Ahn, Kedia, and Blum 2006) (Nakahara and Yamada 2011) (Kuo et al. 2009)
- Open Multilingual WordNet (Bond and Foster 2013), a linked-data representation of WordNet (Miller et al. 1998) and its parallel projects in multiple languages
- JMDict (Breen 2004), a Japanese-multilingual dictionary
- OpenCyc, a hierarchy of hypernyms provided by Cyc (Lenat and Guha 1989), a system that represents commonsense knowledge in predicate logic
- A subset of DBPedia (Auer et al. 2007), a network of facts extracted from Wikipedia infoboxes

Most of them are entity-centric knowledge, there are only 116,097 edges among 74,989 nodes about events.

Speer, Chin, and Havasi, ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. AAAI 2017.
Most Existing KBs are Entity-centric

• Many large-scale knowledge graphs about **entities** and their **attributes** (property-of) and **relations** (thousands of different predicates) have been developed
  • **Millions** of entities and concepts
  • **Billions** of relationships

Google Knowledge Graph (2012)
570 million entities and 18 billion facts
However,

• Semantic meaning in our language can be described as ‘a finite set of mental primitives and a finite set of principles of mental combination (Jackendoff, 1990)’.

• The primitive units of semantic meanings include
  • Thing (or Object, Entity, Concept, Instance, etc.),
  • Property,
  • Place,
  • Path,
  • Amount,
  • Activity,
  • State,
  • Event,
  • etc.

Outline

• Motivation: NLP and commonsense knowledge

• Consideration: selectional preference

• New proposal: large-scale and higher-order selectional preference

• Application on the Winograd Schema Challenge

• Extensions
“Linguistic description – grammar = semantics”

The lower bound of a semantic theory (Katz and Fodor, 1963)

- Disambiguation needs both “the speaker's knowledge of his language and his knowledge” (Katz and Fodor, 1963)

- Compare semantic meanings by fixing grammar
  - Syntactically unambiguous

Principle #1

Selectional Preference (SP)

• The need of language inference based on ‘partial information’ (Wilks, 1975)
  • The soldiers fired at the women, and we saw several of them fall.
  • The needed partial information: hurt things tending to fall down
    • “not invariably true”
    • “tend to be of a very high degree of generality indeed”

(hurt, X) connection (X, fall)

• Selectional preference (Resnik, 1993)
  • A relaxation of selectional restrictions (Katz and Fodor, 1963) and as syntactic features
    (Chomsky, 1965)
  • Applied to isA hierarchy in WordNet and verb-object relations

A Test of Commonsense Reasoning

• Proposed by Hector Levesque at U of Toronto
• An example taking from Winograd Schema Challenge
  • (A) The fish ate the worm. It was hungry.
  • (B) The fish ate the worm. It was tasty.

• On the surface, they simply require the resolution of anaphora
  • But Levesque argues that for Winograd Schemas, the task requires the use of knowledge and commonsense reasoning

http://commonsensereasoning.org/winograd.html
https://en.wikipedia.org/wiki/Winograd_Schema_Challenge
Why is it a challenge?

• Must also be carefully written not to betray their answers by selectional restrictions or statistical information about the words in the sentence

• Designed to be an improvement on the Turing test

• (A) The fish ate the worm. It was hungry.
  • (B) The fish ate the worm. It was tasty.
SP-10K: A Large-scale Evaluation Set of Selectional Preference

• 72 out of 273 questions satisfying nsubj_amod and dobj_amod relations
  • Jim yelled at Kevin because he was so upset.
  • We compare the scores
    • (yell, upset object) following nsubj_amod
    • (upset object, yell) following dobj_amod

• Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Correct</th>
<th>Wrong</th>
<th>NA</th>
<th>Accuracy (predicted)</th>
<th>Accuracy (overall)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stanford</td>
<td>33</td>
<td>35</td>
<td>4</td>
<td>48.5%</td>
<td>48.6%</td>
</tr>
<tr>
<td>End2end (Lee et al., 2018)</td>
<td>36</td>
<td>36</td>
<td>0</td>
<td>50.0%</td>
<td>50.0%</td>
</tr>
<tr>
<td>PP* (Resnik, 1997)</td>
<td>36</td>
<td>19</td>
<td>17</td>
<td>65.5%</td>
<td>61.8%</td>
</tr>
<tr>
<td>SP-10K</td>
<td>13</td>
<td>0</td>
<td>56</td>
<td>100%</td>
<td>59.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dobj_amod</th>
<th>Plausibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>(lift, heavy object)</td>
<td>9.17</td>
</tr>
<tr>
<td>(design, new object)</td>
<td>8.00</td>
</tr>
<tr>
<td>(attack, small object)</td>
<td>5.23</td>
</tr>
<tr>
<td>(inform, weird object)</td>
<td>3.64</td>
</tr>
<tr>
<td>(earn, rubber object)</td>
<td>0.63</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>nsubj_amod</th>
<th>Plausibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>(evil subject, attack)</td>
<td>9.00</td>
</tr>
<tr>
<td>(recent subject, demonstrate)</td>
<td>6.00</td>
</tr>
<tr>
<td>(random subject, bear)</td>
<td>4.00</td>
</tr>
<tr>
<td>(happy subject, steal)</td>
<td>2.25</td>
</tr>
<tr>
<td>(sunny subject, make)</td>
<td>0.56</td>
</tr>
</tbody>
</table>

*PP: posterior probability for SP acquisition using Wikipedia data

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL, 2019.
Outline

• Motivation: NLP and commonsense knowledge

• Consideration: selectional preference

• New proposal: large-scale and higher-order selectional preference

• Application on the Winograd Schema Challenge

• Extensions
Higher-order Selectional Preference

• The need of language inference based on ‘partial information’ (Wilks, 1975)
  • The soldiers fired at the women, and we saw several of them fall.
  • The needed partial information: hurt things tending to fall down
  • Many ways to represent it, e.g.,

  \((\text{hurt}, \, X)\) connection \((X, \, \text{fall})\)

• How to scale up the knowledge acquisition and inference?
ATOMIC

- Crowdsourcing 9 Types of IF-THEN relations

- All personal entity information has been removed to reduce ambiguity

- Arbitrary texts

KnowlyWood

• Perform information extraction from free text
  • Mostly movie scripts and novel books

• Four relations: previous, next, parent, similarity

• Only verb+object

A New Knowledge Graph: ASER
Activities, States, Events, and their Relations

• Use verb-centric patterns from dependency parsing
  • Principle #1: to compare semantics by fixing syntax (Katz and Fodor, 1963)
• Maintain a set of key tags and a set of auxiliary tags
  • Principle #2: to obtain frequent ‘partial information’ (Wilks, 1975)

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

• Using **patterns** to collect partial information

• Six relations are also kept but treated as **auxiliary edges**
  • advmod,
  • amod,
  • nummod,
  • aux,
  • compound,
  • neg

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Code</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>n1-nsubj-v1</td>
<td>s-v</td>
<td>'The dog barks'</td>
</tr>
<tr>
<td>n1-nsubj-v1-dobj-n2</td>
<td>s-v-o</td>
<td>'I love you'</td>
</tr>
<tr>
<td>n1-nsubj-v1-xcomp-a</td>
<td>s-v-a</td>
<td>'He felt ill'</td>
</tr>
<tr>
<td>n1-nsubj-(v1-iobj-n2)-dobj-n3</td>
<td>s-v-o-o</td>
<td>'You give me the book'</td>
</tr>
<tr>
<td>n1-nsubj-a1-cop-be</td>
<td>s-be-a</td>
<td>'The dog is cute'</td>
</tr>
<tr>
<td>n1-nsubj-v1-xcomp-a1-cop-be</td>
<td>s-v-be-a</td>
<td>'I want to be slim'</td>
</tr>
<tr>
<td>n1-nsubj-v1-xcomp-n2-cop-be</td>
<td>s-v-be-o</td>
<td>'I want to be a hero'</td>
</tr>
<tr>
<td>n1-nsubj-v1-xcomp-v2-dobj-n2</td>
<td>s-v-v-o</td>
<td>'I want to eat the apple'</td>
</tr>
<tr>
<td>n1-nsubj-v1-xcomp-v2</td>
<td>s-v-v</td>
<td>'I want to go'</td>
</tr>
<tr>
<td>(n1-nsubj-a1-cop-be)-nmod-n2-case-p1</td>
<td>s-be-a-p-o</td>
<td>'It' cheap for the quality'</td>
</tr>
<tr>
<td>n1-nsubj-v1-nmod-n2-case-p1</td>
<td>s-v-p-o</td>
<td>'He walks into the room'</td>
</tr>
<tr>
<td>(n1-nsubj-v1-dobj-n2)-nmod-n3-case-p1</td>
<td>s-v-o-p-o</td>
<td>'He plays football with me'</td>
</tr>
<tr>
<td>n1-nsubjpass-v1</td>
<td>spass-v</td>
<td>'The bill is paid'</td>
</tr>
<tr>
<td>n1-nsubjpass-v1-nmod-n2-case-p1</td>
<td>spass-v-p-o</td>
<td>'The bill is paid by me'</td>
</tr>
</tbody>
</table>
Distribution

- Frequency characterizes selectional preference, e.g.,
- ‘The dog is chasing the cat, it barks loudly’
  - ‘dog barks’ appears 12,247
  - ‘cat barks’ never appears
Eventuality Relations: Pattern Matching + Bootstrapping

- Seeds from Penn Discourse Treebank (PDTB) (Prasad et al., 2007)
- 14 relations taking from CoNLL shared task
  - More frequent relations
- Less ambiguous connectives
  - ‘so that’ 31 times only in ‘Result’ relations
- Some are ambiguous
  - ‘while’: Conjunction 39 times, Contrast 111 times, Expectation 79 times, and Concession 85 times

<table>
<thead>
<tr>
<th>Relation Type</th>
<th>Seed Patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precedence</td>
<td>E1 before E2; E1, then E2; E1 till E2; E1 until E2</td>
</tr>
<tr>
<td>Succession</td>
<td>E1 after E2; E1 once E2</td>
</tr>
<tr>
<td>Synchronous</td>
<td>E1, meanwhile E2; E1 meantime E2; E1, at the same time E2</td>
</tr>
<tr>
<td>Reason</td>
<td>E1, because E2</td>
</tr>
<tr>
<td>Result</td>
<td>E1, so E2; E1, thus E2; E1, therefore E2; E1, so that E2</td>
</tr>
<tr>
<td>Condition</td>
<td>E1, if E2; E1, as long as E2</td>
</tr>
<tr>
<td>Contrast</td>
<td>E1, but E2; E1, however E2; E1, by contrast E2; E1, in contrast E2; E1, on the other hand, E2; E1, on the contrary, E2</td>
</tr>
<tr>
<td>Concession</td>
<td>E1, although E2</td>
</tr>
<tr>
<td>Conjunction</td>
<td>E1 and E2; E1, also E2</td>
</tr>
<tr>
<td>Instantiation</td>
<td>E1, for example E2; E1, for instance E2</td>
</tr>
<tr>
<td>Restatement</td>
<td>E1, in other words E2</td>
</tr>
<tr>
<td>Alternative</td>
<td>E1 or E2; E1, unless E2; E1, as an alternative E2; E1, otherwise E2</td>
</tr>
<tr>
<td>ChosenAlternative</td>
<td>E1, E2 instead</td>
</tr>
<tr>
<td>Exception</td>
<td>E1, except E2</td>
</tr>
</tbody>
</table>

Nianwen Xue, Hwee Tou Ng, Sameer Pradhan, Rashmi Prasad, Christopher Bryant, Attapol T. Rutherford. The CoNLL-2015 Shared Task on Shallow Discourse Parsing.
Eventuality Relations: Pattern matching + Bootstrapping

• Bootstrapping: incrementally self-supervised learning
• For each instance $x = (E1; E2; \text{sentence})$
  • Use three bidirectional LSTMs
• Reduce the confident rate by iterations to reduce error propagation
Scales of Verb Related Knowledge Graphs

- FrameNet (Baker et al., 1998)
- ACE (Aguilar et al., 2014)
- PropBank (Pustejovsky et al., 2005)
- TimeBank (Pustejovsky et al., 2003)
- ConceptNet (Liu & Singh, 2004)
- Event2Mind (Smith et al., 2018)
- Propora (Dalvi et al., 2018)
- ATOMIC (Sap et al., 2018)
- Knowlywood (Tandon et al., 2015)
- ASER (core)
- ASER (full)

1000X larger
100X larger

#Eventualities vs. #Relations
Multi-hop Reasoning based on Selectional Preference

• One-hop
  • frequency(`sing’-nsubj-`singer’-) > frequency(`sing’-nsubj-`house’)
  • frequency(`eat’-dobj-`food’) > frequency(`eat’-dobj-`rock’)

• Two-hop
  • frequency(`eat’-nsubj-X-amod-`hungry’) > frequency(`eat’-dobj-Y-amod-`hungry’)

• Multi-hop
  • frequency(`X eat dinner’->Causes-‘X be full’) > frequency(`X eat dinner’->Causes-‘X be hungry’)

Outline

• Motivation: NLP and commonsense knowledge

• Consideration: selectional preference

• New proposal: large-scale and higher-order selectional preference

• Application on the Winograd Schema Challenge

• Extensions
Inference for Winograd Schema Challenge

Question

97. The fish ate the worm. It was hungry.

98. The fish ate the worm. It was tasty.

Extracted Eventualities

The fish: (‘X ate Y’, ‘X was hungry’)
the worm: (‘X ate Y’, ‘Y was hungry’)

The fish: (‘X ate Y’, ‘X was tasty’)
the worm: (‘X ate Y’, ‘Y was tasty’)

ASER Knowledge

ASER(‘X ate Y’, ‘X was hungry’) = 18
ASER(‘X ate Y’, ‘Y was hungry’) = 1

ASER(‘X ate Y’, ‘X was tasty’) = 0
ASER(‘X ate Y’, ‘Y was tasty’) = 7

Prediction

The fish
the worm
Results on Cases Consistent with Our Patterns

- We selected a subset of 165 questions
  - The sentence does not have a subordinate clause
  - The targeting pronoun is covered by a pattern we used

<table>
<thead>
<tr>
<th>Methods</th>
<th>Correct</th>
<th>Wrong</th>
<th>NA</th>
<th>Predicted Accuracy</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Guess</td>
<td>83</td>
<td>82</td>
<td>0</td>
<td>50.30%</td>
<td>50.30%</td>
</tr>
<tr>
<td>Deterministic (Raghunathan et al., 2010)</td>
<td>75</td>
<td>71</td>
<td>19</td>
<td>51.40%</td>
<td>51.20%</td>
</tr>
<tr>
<td>Statistical (Clark &amp; Manning, 2015)</td>
<td>75</td>
<td>78</td>
<td>12</td>
<td>49.00%</td>
<td>49.10%</td>
</tr>
<tr>
<td>Deep-RL (Clark &amp; Manning, 2016)</td>
<td>80</td>
<td>76</td>
<td>9</td>
<td>51.30%</td>
<td>51.20%</td>
</tr>
<tr>
<td>End2end (Lee et al., 2018)</td>
<td>79</td>
<td>84</td>
<td>2</td>
<td>48.50%</td>
<td>48.50%</td>
</tr>
<tr>
<td>Knowledge Hunting (Emami et al., 2018)</td>
<td>94</td>
<td>71</td>
<td>0</td>
<td>56.90%</td>
<td>56.90%</td>
</tr>
<tr>
<td>LM (single) (Trinh &amp; Le, 2018)</td>
<td>90</td>
<td>75</td>
<td>0</td>
<td>54.50%</td>
<td>54.50%</td>
</tr>
<tr>
<td>SP (human) (Zhang et al., 2019)</td>
<td>15</td>
<td>0</td>
<td>150</td>
<td>100%</td>
<td>54.50%</td>
</tr>
<tr>
<td>SP (PP) (Zhang et al., 2019)</td>
<td>50</td>
<td>26</td>
<td>89</td>
<td>65.80%</td>
<td>57.30%</td>
</tr>
<tr>
<td>ASER</td>
<td>63</td>
<td>27</td>
<td>75</td>
<td>70.00%</td>
<td>60.90%</td>
</tr>
</tbody>
</table>

Hongming Zhang, Hantian Ding, and Yangqiu Song. SP-10K: A Large-Scale Evaluation Set for Selectional Preference Acquisition. ACL, 2019.
Overall Results based on Fine-tuning

<table>
<thead>
<tr>
<th>Methods</th>
<th>Supervision</th>
<th>Overall Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Guess</td>
<td>NA</td>
<td>50.2%</td>
</tr>
<tr>
<td>Knowledge Hunting (Emami et al., 2018)</td>
<td>NA</td>
<td>57.3%</td>
</tr>
<tr>
<td>LM (single) (Trinh &amp; Le, 2018)</td>
<td>NA</td>
<td>54.5%</td>
</tr>
<tr>
<td>LM (Ensemble) (Trinh &amp; Le, 2018)</td>
<td>NA</td>
<td>61.5%</td>
</tr>
<tr>
<td>SP (human) (Zhang et al., 2019)</td>
<td>NA</td>
<td>52.7%</td>
</tr>
<tr>
<td>SP (PP) (Zhang et al., 2019)</td>
<td>NA</td>
<td>54.4%</td>
</tr>
<tr>
<td>GPT-2 (Radford et al., 2019)</td>
<td>NA</td>
<td>70.7%</td>
</tr>
<tr>
<td>BERT (Kocijan et al., 2019)</td>
<td>NA</td>
<td>61.9%</td>
</tr>
<tr>
<td>BERT+WSCR (Kocijan et al., 2019)</td>
<td>WSCR</td>
<td>71.4%</td>
</tr>
<tr>
<td>ASER (inference)</td>
<td>NA</td>
<td>56.6%</td>
</tr>
<tr>
<td>BERT+ASER</td>
<td>WSCR</td>
<td>64.5%</td>
</tr>
<tr>
<td>BERT+WSCR+ASER</td>
<td>WSCR+ASER</td>
<td>72.5%</td>
</tr>
</tbody>
</table>

WSCR: Rahman and Ng’s dataset (2012)
ASER: Automatically constructed patterns as training examples
Outline

• Motivation: NLP and commonsense knowledge
• Consideration: selectional preference
• New proposal: large-scale and higher-order selectional preference
• Application on the Winograd Schema Challenge
• Extensions
  • ASER 2.0
  • ASER-EEG
  • TransOMCS
Partial Information Aggregation

• “hurt things tending to fall down”

(hurt, X) connection (X, fall)

• “stocks price may increase when company X acquire a start-up”

(company, acquire, start-up) result-in (stock, increase)
Conceptualization: The Goal

Google acquire Deepmind

Apple acquire Drive.ai

Microsoft acquire Github

company acquires startup company
## Normalization

<table>
<thead>
<tr>
<th>Input</th>
<th>Output</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>He, she, I, Bob, ...</td>
<td><em><strong>PERSON</strong></em></td>
<td>1.0</td>
</tr>
<tr>
<td>1996, 2020, 1949, ...</td>
<td><em><strong>YEAR</strong></em></td>
<td>1.0</td>
</tr>
<tr>
<td>23, 20, 333, ....</td>
<td><em><strong>DIGIT</strong></em></td>
<td>1.0</td>
</tr>
<tr>
<td><a href="http://www.google.com">www.google.com</a>, ...</td>
<td><em><strong>URL</strong></em></td>
<td>1.0</td>
</tr>
</tbody>
</table>
Conceptualization with ProBase

Concepts are the glue that holds our mental world together.

Gregory L. Murphy
NYU

Probase is a large, universal, probabilistic knowledge base with an extremely large concept space.

Data are available at https://concept.research.microsoft.com/
Conceptualization with ProBase

Typicality

\[ P(\text{concept} \mid \text{instance}) = \frac{\#(\text{concept, instance})}{\#(\text{instance})} \]

• Robin

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>character</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>songbird</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>common bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>small bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Penguin

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>species</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flightless bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>seabird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>diving bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data are available at https://concept.research.microsoft.com/
Yangqiu Song, Haixun Wang, Zhongyuan Wang, Hongsong Li, Weizhu Chen: Short Text Conceptualization Using a Probabilistic Knowledgebase. IJCAI 2011: 2330-2336
A Running Example

Obama

(politician, 0.0855)
(democrat, 0.0560)
(liberal, 0.0560)

(Obama, have, dog)

(obama have animal, 0.2811)
(obama have pet, 0.1377)
(politician have dog, 0.0855)
(democrat have dog, 0.05604)
...
(politician have animal, 0.0240)
(democrat have animal, 0.01575)
...

dog

(animal, 0.2811)
(pet, 0.1377)
(domestic animal, 0.0525)

\[
\prod_{i=1}^{N} P(C_{i,k} | E_i)
\]

\[
P\text{(politician} \mid \text{Obama}) \times P\text{(animal} \mid \text{dog})
\]

\[
= 0.0855 \times 0.2811 = 0.0240
\]
A Running Example

Obama

(politician, 0.0855)
(democrat, 0.0560)
(liberal, 0.0560)

(Obama, have, dog)

(obama have animal, 0.2811)
(obama have pet, 0.1377)
(politician have dog, 0.0855)
(democrat have dog, 0.05604)
...
(politician have animal, 0.0240)
(democrat have animal, 0.01575)
...

Number of ASER-concepts:

\[ C_N^1 \times K + C_N^2 \times K^2 + \cdots + C_N^K K^N \]

K is Top K probase-concept for each entity, N is #entity in an eventuality

dog

(animal, 0.2811)
(pet, 0.1377)
(domestic animal, 0.0525)


\[
P(\text{ResultIn} | \text{person, have, animal}, \text{positive-emotion, come}) = 0.281 \times 3 \times 0.087 + 0.333 \times 2 \times 0.125 = 0.157
\]
ASER 2.0

• Rule based extraction (14 Eventuality Patterns, Improved Version)

<table>
<thead>
<tr>
<th>Data</th>
<th>#Eventualities</th>
<th>#Unique Eventualities</th>
<th>#Relations</th>
<th>#Unique Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>349,296,240</td>
<td>34,212,258</td>
<td>65,997,575</td>
<td>15,339,027</td>
</tr>
<tr>
<td>Full</td>
<td>587,290,657</td>
<td>272,206,675</td>
<td>265,681,802</td>
<td>205,758,398</td>
</tr>
</tbody>
</table>

• Discourse Parser (18 Eventuality Patterns + Wang and Lan 2015)

<table>
<thead>
<tr>
<th>Data</th>
<th>#Eventualities</th>
<th>#Unique Eventualities</th>
<th>#Relations</th>
<th>#Unique Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core</td>
<td>477,383,662</td>
<td>42,964,177</td>
<td>120,995,415</td>
<td>25,880,127</td>
</tr>
<tr>
<td>Full</td>
<td>799,191,666</td>
<td>364,772,181</td>
<td>463,640,100</td>
<td>368,635,332</td>
</tr>
</tbody>
</table>

• Conceptualization Core:
  • Concepts: 65,837,819 (1.5 times larger)
  • Concept Relations: 289,735,387 (11 times larger)

Outline

• Motivation: NLP and commonsense knowledge
• Consideration: selectional preference
• New proposal: large-scale and higher-order selectional preference
• Application on the Winograd Schema Challenge

• Extensions
  • ASER 2.0
  • ASER-EEG
  • TransOMCS
Incorporating More Relations

Two Issues:
1. Concept Transitivity
2. Verb’s Entailment Relations
Entailment Graph Construction

![Entailment Graph Construction Diagram]

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Reference</th>
<th>#Graphs</th>
<th>#Nodes</th>
<th>#Edges</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typed Predicate</td>
<td>Berant et al., ACL, 2011</td>
<td>2,303</td>
<td>10,672</td>
<td>263,756</td>
<td>Place/disease</td>
</tr>
<tr>
<td></td>
<td>Hosseini et al. TACL, 2018</td>
<td>363</td>
<td>101K</td>
<td>66M</td>
<td>News</td>
</tr>
<tr>
<td>Open IE Proposition</td>
<td>Levy et al., CoNLL, 2014</td>
<td>30</td>
<td>5,714</td>
<td>1.5M</td>
<td>Healthcare</td>
</tr>
<tr>
<td>Eventuality</td>
<td>Ours</td>
<td>473</td>
<td>10M</td>
<td>103M</td>
<td>Commonsense</td>
</tr>
</tbody>
</table>
Three-step Construction

1. Eventuality pre-processing
2. Local Inference
3. Global Inference

Changlong Yu, Hongming Zhang, Yangqiu Song, Wilfred Ng, Lifeng Shang. Enriching Large-Scale Eventuality Knowledge Graph with Entailment Relations. AKBC. 2020.
Results

- We can generate 10 times of edges

<table>
<thead>
<tr>
<th>Eventuality</th>
<th># Eventuality</th>
<th># ER (global)</th>
<th># ER (local)</th>
<th>Acc (local)</th>
<th>Acc (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s-v ⊨ s-v</td>
<td>3.3M</td>
<td>32.7M</td>
<td>10.7M</td>
<td>89.1%</td>
<td>85.7%</td>
</tr>
<tr>
<td>s-v-o ⊨ s-v-o</td>
<td>5.3M</td>
<td>45.2M</td>
<td>14.8M</td>
<td>90.1%</td>
<td>89.3%</td>
</tr>
<tr>
<td>s-v-p-o ⊨ s-v-p-o</td>
<td>1.9M</td>
<td>12.6M</td>
<td>5.3M</td>
<td>88.3%</td>
<td>87.4%</td>
</tr>
<tr>
<td>s-v-o-p-o ⊨ s-v-o</td>
<td>0.5M</td>
<td>0.8M</td>
<td>0.8M</td>
<td>91.4%</td>
<td>90.0%</td>
</tr>
<tr>
<td>s-v-p-o ⊨ s-v-o</td>
<td>1.1M</td>
<td>2.7M</td>
<td>0.9M</td>
<td>88.5%</td>
<td>87.2%</td>
</tr>
<tr>
<td>s-v-o ⊨ s-v-p-o</td>
<td>0.9M</td>
<td>5.4M</td>
<td>2.2M</td>
<td>87.8%</td>
<td>86.7%</td>
</tr>
<tr>
<td>s-v-o-p-o ⊨ s-v-o-p-o</td>
<td>2.4M</td>
<td>3.2M</td>
<td>2.1M</td>
<td>89.4%</td>
<td>88.4%</td>
</tr>
<tr>
<td>s-v-a ⊨ s-be-a</td>
<td>0.2M</td>
<td>0.1M</td>
<td>0.1M</td>
<td>97.9%</td>
<td>97.9%</td>
</tr>
<tr>
<td>s-be-a-p-o ⊨ s-be-a</td>
<td>0.8M</td>
<td>0.4M</td>
<td>0.4M</td>
<td>96.0%</td>
<td>95.8%</td>
</tr>
<tr>
<td>s-be-a-p-o ⊨ s-be-a-p-o</td>
<td>0.1M</td>
<td>0.1M</td>
<td>0.1M</td>
<td>95.1%</td>
<td>94.7%</td>
</tr>
<tr>
<td>Overall</td>
<td>10.0M</td>
<td>103.2M</td>
<td>37.4M</td>
<td>91.4%</td>
<td>90.3%</td>
</tr>
</tbody>
</table>
Outline

• Motivation: NLP and commonsense knowledge
• Consideration: selectional preference
• New proposal: large-scale and higher-order selectional preference
• Application on the Winograd Schema Challenge
• Extensions
  • ASER 2.0
  • ASER-EEG
  • TransOMCS
ASER is Essentially a Knowledge Graph based on Linguistics

How is it transferrable from linguistic knowledge to existing definition of commonsense knowledge?
Revisit the Correlations of SP and OMCS

(sing, song) (dobj, 9.25)
(song, UsedFor, sing)

(phone, ring) (nsubj, 8.75)
(phone, CapableOf, ring)

(cold, water) (amod, 8.86)
(water, HasProperty, cold)

(create, new) (dobj_amod, 8.25)
(create idea, UsedFor, invent new things)

(hungry, eat) (nsubj_amod, 10.00)
(eat, MotivatedByGoal, are hungry)
Revisit the Correlations of ASER and OMCS

[Heatmap diagram showing correlations between various categories such as "HasPrerequisite", "Causes", "MotivatedByGoal", and "HasSubevent", with a color scale indicating percentage values.]
Can we Discover more OMCS Knowledge from ASER?

Step 1: Pattern mining by heuristic scoring

Step 2: Learning to rank from 1,000 annotated tuples in each relation

Hongming Zhang, Daniel Khashabi, Yangqiu Song, and Dan Roth. TransOMCS: From Linguistic Graphs to Commonsense Knowledge. IJCAI. 2020.
<table>
<thead>
<tr>
<th>Model</th>
<th># Vocab</th>
<th># Tuple</th>
<th>Novel (Tuple)</th>
<th>Novel (Concept)</th>
<th>ACC (Novel)</th>
<th>ACC (Overall)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMET 1.2K ConceptNet Test Set (Greedy)</td>
<td>715</td>
<td>1,200</td>
<td>33.96%</td>
<td>5.27%</td>
<td>58%</td>
<td>90%</td>
</tr>
<tr>
<td>COMET 1.2K ConceptNet Test Set (10 Beams)</td>
<td>2,232</td>
<td>12,000</td>
<td>64.95%</td>
<td>27.15%</td>
<td>35%</td>
<td>44%</td>
</tr>
<tr>
<td>COMET 24K ASER Sampled Graphs (Greedy)</td>
<td>3,912</td>
<td>24,000</td>
<td>99.98%</td>
<td>55.56%</td>
<td>34%</td>
<td>47%</td>
</tr>
<tr>
<td>COMET 24K ASER Sampled Graphs (10 Beams)</td>
<td>8,108</td>
<td>240,000</td>
<td>99.98%</td>
<td>78.59%</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>LAMA 1.2K ConceptNet Test Set (Top 1)</td>
<td>328</td>
<td>1,200</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>49%</td>
</tr>
<tr>
<td>LAMA 1.2K ConceptNet Test Set (Top 10)</td>
<td>1,649</td>
<td>12,000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20%</td>
</tr>
<tr>
<td>LAMA 1.2K ASER Sampled Graphs (Top 1)</td>
<td>1,443</td>
<td>24,000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>29%</td>
</tr>
<tr>
<td>LAMA 1.2K ASER Sampled Graphs (Top 10)</td>
<td>5,464</td>
<td>240,000</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>10%</td>
</tr>
<tr>
<td>TransOMCS overlapped with 1.2K ConceptNet</td>
<td>33,238</td>
<td>533,449</td>
<td>99.53%</td>
<td>89.20%</td>
<td>72%</td>
<td>74%</td>
</tr>
<tr>
<td>TransOMCS (Top 1%)</td>
<td>37,517</td>
<td>184,816</td>
<td>95.71%</td>
<td>75.65%</td>
<td>86%</td>
<td>87%</td>
</tr>
<tr>
<td>TransOMCS (Top 10%)</td>
<td>56,411</td>
<td>1,848,160</td>
<td>99.55%</td>
<td>92.17%</td>
<td>69%</td>
<td>74%</td>
</tr>
<tr>
<td>TransOMCS (Top 30%)</td>
<td>68,438</td>
<td>5,544,482</td>
<td>99.83%</td>
<td>95.22%</td>
<td>67%</td>
<td>69%</td>
</tr>
<tr>
<td>TransOMCS (Top 50%)</td>
<td>83,823</td>
<td>9,240,803</td>
<td>99.89%</td>
<td>96.32%</td>
<td>60%</td>
<td>62%</td>
</tr>
<tr>
<td>TransOMCS (no ranking)</td>
<td>100,659</td>
<td>18,481,607</td>
<td>99.94%</td>
<td>98.30%</td>
<td>54%</td>
<td>56%</td>
</tr>
<tr>
<td>OMCS in ConceptNet 5.0</td>
<td>36,954</td>
<td>207,427</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>92%</td>
</tr>
</tbody>
</table>
Distribution of Relations and Accuracy

[Bar chart showing the distribution of different relations with logarithmic y-axis and linear x-axis.]
# Case Studies

<table>
<thead>
<tr>
<th>“human” <code>CapableOf</code></th>
<th></th>
<th>“love” <code>Causes</code></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COMET</strong></td>
<td><strong>LAMA</strong></td>
<td><strong>TransOMCS</strong></td>
</tr>
<tr>
<td>1. kill other person</td>
<td>1. be</td>
<td>1. happiness</td>
</tr>
<tr>
<td>2. kill other human</td>
<td>2. fly</td>
<td>2. be happy</td>
</tr>
<tr>
<td>3. kill other sentient be</td>
<td>die</td>
<td>3. get marry</td>
</tr>
<tr>
<td>4. feel emotion</td>
<td>4. talk</td>
<td>4. death</td>
</tr>
<tr>
<td>5. kill other human be</td>
<td>kill</td>
<td>5. you get marry</td>
</tr>
<tr>
<td>6. make wine</td>
<td>6. speak</td>
<td>6. you feel good</td>
</tr>
<tr>
<td>7. hate</td>
<td>7. breathe</td>
<td>7. pain</td>
</tr>
<tr>
<td>8. love</td>
<td>8. eat</td>
<td>8. love</td>
</tr>
<tr>
<td>10. die</td>
<td>10. see</td>
<td>10. war</td>
</tr>
</tbody>
</table>

(a) Original Setting

<table>
<thead>
<tr>
<th><strong>COMET</strong></th>
<th><strong>LAMA</strong></th>
<th><strong>TransOMCS</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. chaos</td>
<td>1. be friendly</td>
<td>1. be friendly</td>
</tr>
<tr>
<td>2. pain</td>
<td>2. be happy</td>
<td>2. be happy</td>
</tr>
<tr>
<td>3. problems</td>
<td>3. pain</td>
<td>3. pain</td>
</tr>
<tr>
<td>4. love</td>
<td>4. marriage</td>
<td>4. marriage</td>
</tr>
<tr>
<td>5. trouble</td>
<td>5. be quaint</td>
<td>5. be quaint</td>
</tr>
<tr>
<td>6. death</td>
<td>6. be unhappy</td>
<td>6. be unhappy</td>
</tr>
<tr>
<td>7. fear</td>
<td>7. be allergic</td>
<td>7. be allergic</td>
</tr>
<tr>
<td>8. happiness</td>
<td>8. be desperate</td>
<td>8. be desperate</td>
</tr>
<tr>
<td>9. war</td>
<td>9. be apart</td>
<td>9. be apart</td>
</tr>
<tr>
<td>10. conflict</td>
<td>10. be silly</td>
<td>10. be silly</td>
</tr>
</tbody>
</table>

(b) Extended Setting
Conclusions

• We extended the concept of selectional preference for commonsense knowledge acquisition

• Many potential extensions
  • More links with knowledge base completion and population
  • Many downstream tasks

• Project Homepage
  • https://hkust-knowcomp.github.io/ASER/

Thank you 😊
Eventuality Extraction Results

• Extract examples from 11-billion tokens from Yelp, NYT, Wiki, Reddit, Subtitles, E-books

• Evaluate about 200 examples in each pattern using Amazon Turk
Relation Extraction Results

- Left: number of relations and overall accuracy
- Right: accuracy of each relations for the last iteration
- Each point is annotated with 200 examples by Amazon Turk