Machine-Learning Semantics for Webscale NLP

Mark Steedman

27th October 2016
The Problem of Content

• We have (somewhat) robust wide coverage parsers that work on the scale of Bn of words (e.g. Clark and Curran 2004; Lewis and Steedman 2014a). They can read the web (and build logical forms) much faster than we can ourselves.

• So why can’t we ask them questions like “What are recordings by Miles Davis without Fender Rhodes piano”, and get a helpful answer?

• The central problem of QA is that there are too many ways of asking and answering questions, and we have no idea of the semantics that relates them.
Too Many Ways of Answering The Question

- Question: Did Google buy YouTube?
- The Text:
  1. Google purchased YouTube.
  2. Google's purchase of YouTube.
  3. Google acquired every company.
  4. YouTube may be sold to Google.
  5. Google will buy YouTube or Microsoft.
  6. Google didn’t take over YouTube.
The Approach

• Use the semantic parsers to Machine-Read multiple relations over Named Entities in web text.

• Capture relations of Entailment over relations between NEs of the same types (Lewis and Steedman, 2014b; Lewis, 2015).
  – If you read somewhere that a person—say, Obama—was elected to an office—say, President—than you are highly likely to also read somewhere that that person ran for that office.
  – —but not the other way round

• Redefine the parser semantics in terms of entailments and paraphrases, and reparse and index the entire text for IR.
Local Entailment Probabilities

• The typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments using Weeds precision asymmetric similarity (Weeds and Weir, 2003):

  a. $p(\text{conquer} \ xy \Rightarrow \text{invade} \ xy) = 0.9$
  b. $p(\text{invade} \ xy \Rightarrow \text{attack} \ xy) = 0.8$
  c. $p(\text{conquer} \ xy \Rightarrow \text{attack} \ xy) = 0.4$
  d. $p(\text{bomb} \ xy \Rightarrow \text{attack} \ xy) = 0.7$
  e. $p(\text{bomb} \ xy \Rightarrow \text{conquer} \ xy) = 0.2$

  (etc.)
Global Entailments

- The local entailment probabilities are used to construct an entailment graph using integer linear programming with a prior $p = 0.25$ with the global constraint that the graph must be closed under transitivity.

- Thus, (c) will be included despite low observed frequency, while other low frequency spurious local entailments will be excluded.

- Cliques within the entailment graphs are collapsed to a single paraphrase cluster relation identifier, as in the previous approach.
• A simple entailment graph for relations between countries.
Lexicon

- The new semantics obtained from the entailment graph

\[
\text{attack} := (S\setminus NP)/NP : \lambda x \lambda y \lambda e.\text{rel}_1 x y e
\]
\[
\text{bomb} := (S\setminus NP)/NP : \lambda x \lambda y \lambda e.\text{rel}_1 x y e \land \text{rel}_4 x y e
\]
\[
\text{invade} := (S\setminus NP)/NP : \lambda x \lambda y \lambda e.\text{rel}_1 x y e \land \text{rel}_2 x y e
\]
\[
\text{conquer} := (S\setminus NP)/NP : \lambda x \lambda y \lambda e.\text{rel}_1 x y e \land \text{rel}_2 x y e \land \text{rel}_3 x y e
\]
\[
\text{annex} := (S\setminus NP)/NP : \lambda x \lambda y \lambda e.\text{rel}_1 x y e \land \text{rel}_2 x y e \land \text{rel}_3 x y e
\]

- These logical forms support correct inference under negation, such as that
  conquered entails attacked and didn't invade entails didn't conquer
Results

• Examples:

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>From Unseen Sentence:</th>
</tr>
</thead>
<tbody>
<tr>
<td>What did Delta merge with?</td>
<td>Northwest</td>
<td>The 747 freighters came with Delta’s acquisition of Northwest</td>
</tr>
<tr>
<td>What spoke with Hu Jintao?</td>
<td>Obama</td>
<td>Obama conveyed his respect for the Dalai Lama to China’s president Hu Jintao during their first meeting</td>
</tr>
<tr>
<td>What arrived in Colorado?</td>
<td>Zazi</td>
<td>Zazi flew back to Colorado. . .</td>
</tr>
<tr>
<td>What ran for Congress?</td>
<td>Young</td>
<td>. . . Young was elected to Congress in 1972</td>
</tr>
</tbody>
</table>

• Full results in Lewis and Steedman (2013)
# More Examples

<table>
<thead>
<tr>
<th>Premise</th>
<th>Hypothesis</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama want to boost the defense budget</td>
<td>Obama increase the defense budget</td>
<td>False</td>
</tr>
<tr>
<td>The thieves make off with TVs</td>
<td>The thieves manage to steal TVs</td>
<td>True</td>
</tr>
<tr>
<td>My son be terrified of him</td>
<td>My son have a fear of him</td>
<td>True</td>
</tr>
</tbody>
</table>
The Next Step: Generalize to Temporal Semantics

A simple entailment graph for relations over events does not capture relations of causation and temporal sequence.
Learning from Timestamped Data

- One source of information concerning these hidden relations is timestamped news, of the kind available in the University of Washington NewsSpike corpus of 0.5M newswire articles (Zhang and Weld, 2013).

- In such data, we find that statements that so-and-so is visiting, is in and the perfect has arrived in such and such a place, occur in stories with the same datestamp, whereas is arriving, is on her way to, occur in preceding stories, while has left, is on her way back from, returned, etc. occur in later ones.

- This information provides a basis for inference that visiting entails being in, that the latter is the consequent state of arriving, and that arrival and departure coincide with the beginning and end of the progressive state of visiting.
Another Project: Web to Semantic Net

• We would like to interrogate huge databases such as the Google knowledge graphs, a.k.a. Semantic Nets (Reddy et al., 2014)
• There is a mismatch between the semantics delivered by parsers and the language of the knowledge graph.
• So let’s build our own knowledge graph using the clustered entailment semantics of the parser, so that we can query it directly in natural language.

⚠️ This is a potentially a much bigger graph than the Knowledge Graph.

• We will need techniques to limit exponential growth in the costs of loading and interrogating this graph.
• Pilot experiments by Harrington and Clark (2009); Lao et al. (2012) suggest this can be done by spreading activation (Collins and Loftus, 1975).
References


Zhang, Congle and Weld, Daniel, 2013. “Harvesting Parallel News Streams to