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 somwhere 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 assymetric similarity (Weeds and Weir, 2003):

  a. $p(\mathit{conquer}\,x\,y \Rightarrow \mathit{invade}\,x\,y) = 0.9$

  b. $p(\mathit{invade}\,x\,y \Rightarrow \mathit{attack}\,x\,y) = 0.8$

  c. $p(\mathit{conquer}\,x\,y \Rightarrow \mathit{attack}\,x\,y) = 0.4$

  d. $p(\mathit{bomb}\,x\,y \Rightarrow \mathit{attack}\,x\,y) = 0.7$

  e. $p(\mathit{bomb}\,x\,y \Rightarrow \mathit{conquer}\,x\,y) = 0.2$

  (etc.)

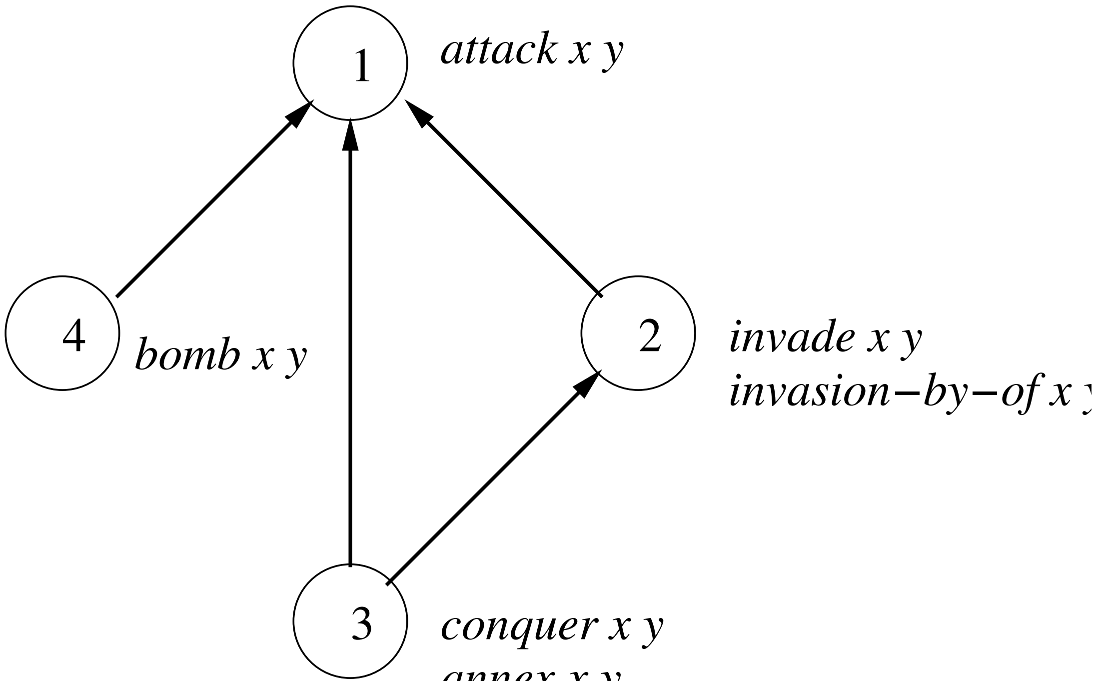
Global Entailments

- 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 paraphase cluster relation identifier, as in the previous approach.

Entailment graph



- A simple entailment graph for relations between countries.

Lexicon

- The new semantics obtained from the entailment graph

  attack $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\,x\,y\,e$
  bomb $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\,x\,y\,e \wedge rel_4\,x\,y\,e$
  invade $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\,x\,y\,e \wedge rel_2\,x\,y\,e$
  conquer $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\,x\,y\,e \wedge rel_2\,x\,y\,e \wedge rel_3\,x\,y\,e$
  annex $:= (S\backslash NP)/NP : \lambda x\lambda y\lambda e.rel_1\,x\,y\,e \wedge rel_2\,x\,y\,e \wedge rel_3\,x\,y\,e$

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

Results

- Examples:

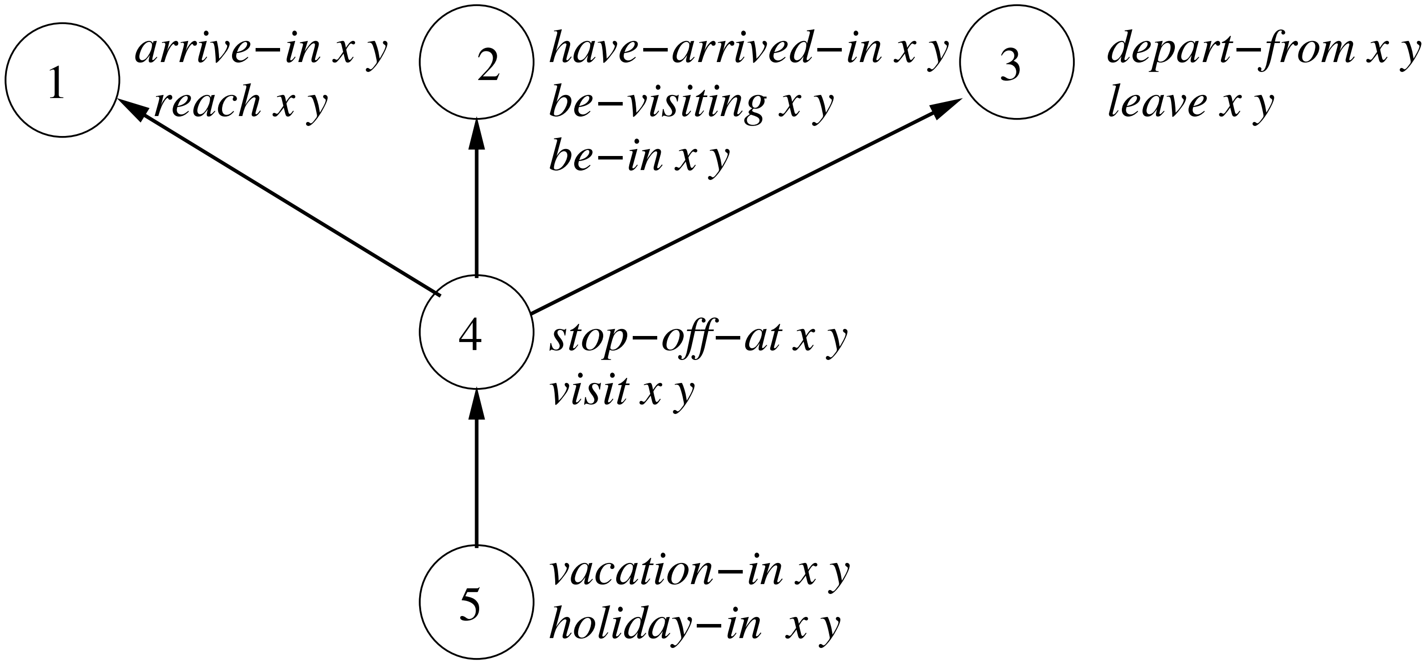
Question	Answer	From Unseen Sentence:
What did Delta merge with?	Northwest	The 747 freighters came with Delta's acquisition of Northwest
What spoke with Hu Jintao?	Obama	Obama conveyed his respect for the Dalai Lama to China's president Hu Jintao during their first meeting
What arrived in Colorado?	Zazi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

- Full results in Lewis and Steedman (2013)

More Examples

Premise	Hypothesis	Answer
Obama want to boost the defense budget	Obama increase the defense budget	False
The thieves make off with TVs	The thieves manage to steal TVs	True
My son be terrified of him	My son have a fear of him	True

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 lets 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

Clark, Stephen and Curran, James R., 2004. "Parsing the WSJ using CCG and Log-Linear Models." In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. Barcelona: ACL, 104–111.

Collins, Allan and Loftus, Elizabeth, 1975. "A Spreading Activation Theory of Semantic Processing." *Psychological Review* 82:407–428.

Harrington, Brian and Clark, Stephen, 2009. "ASKNet: Creating and Evaluating Large Scale Integrated Semantic Networks." *International Journal of Semantic Computing* 2:343–364.

Lao, Ni, Subramanya, Amarnag, Pereira, Fernando, and Cohen, William, 2012. "Reading the Web with Learned Syntactic-Semantic Inference Rules." In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. ACL, 1017–1026.

Lewis, Mike, 2015. *Combined Distributional and Logical Semantics*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike and Steedman, Mark, 2013. "Combined Distributional and Logical Semantics." *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Steedman, Mark, 2014a. "A^* CCG Parsing with a Supertag-factored Model." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL, 990–1000.

Lewis, Mike and Steedman, Mark, 2014b. "Combining Formal and Distributional Models of Temporal and Intensional Semantics." In *Proceedings of the ACL Workshop on Semantic Parsing*. Baltimore, MD: ACL, 28–32. Google Exceptional Submission Award.

Reddy, Siva, Lapata, Mirella, and Steedman, Mark, 2014. "Large-scale Semantic Parsing without Question-Answer Pairs." *Transactions of the Association for Computational Linguistics* 2:377–392.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Weeds, Julie and Weir, David, 2003. "A General Framework for Distributional Similarity." In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. ACL, 81–88.

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

Generate Paraphrases of Event Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1776–1786.