CMU: Arc-Factored, Discriminative Semantic Dependency Parsing

Sam Thomson, Brendan O'Connor, Jeffrey Flanigan, David Bamman, Jesse Dodge, Swabha Swayamdipta, Nathan Schneider, Chris Dyer and Noah A. Smith
Language Technologies Institute, Carnegie Mellon University

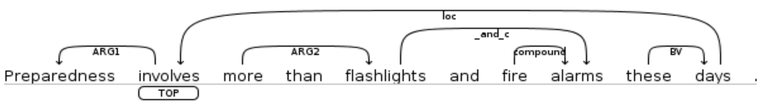
SemEval Shared Task 8: Broad-Coverage Semantic Dependency Parsing

Goal:

- Given a natural language sentence, produce a labeled, directed graph that represents its semantic relationships.

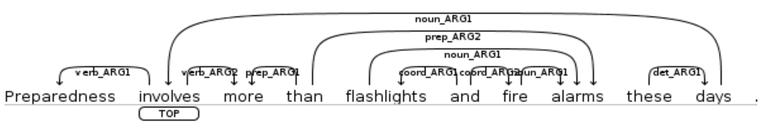
Formalisms

DM



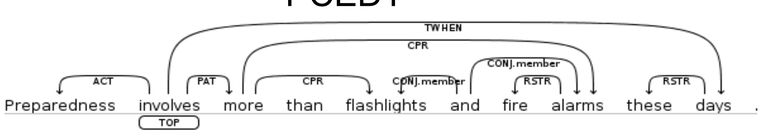
- Minimal Recursion Semantics
- 90% Connected
- 99% deterministic

PAS



- Enju predicate-argument-structures
- 100% deterministic

PCEDT



- Tectogrammatical layer of Prague Czech-English Dependency Treebank 2.0
- 85% deterministic

General approach:

- Learn an arc-factored statistical model that scores each labeled, directed edge.
- Choose subgraph with maximum sum of edge weights, subject to linguistically-motivated graph constraints.
- Similar to MSTParser, but
  semantic graphs are not trees

Semantic Well-Formedness Constraints

- Only one of each core argument allowed per predicate
- At most one arc between each pair of words.
- Connected?

Edge Models

- Multiclass Logistic Regression

$$P(\ell \mid \boldsymbol{\phi}, x, i, j) = \frac{\exp\{\boldsymbol{\phi} \cdot \boldsymbol{f}(x, i, j, \ell)\}}{\sum_{\ell' \in L} \exp\{\boldsymbol{\phi} \cdot \boldsymbol{f}(x, i, j, \ell')\}}$$

  - best model when edges considered independently

- Structured SVM

$$-\boldsymbol{\psi}^\top \boldsymbol{f}(x_i, y_i) + \max_y \boldsymbol{\psi}^\top \boldsymbol{f}(x_i, y) + cost(y, y_i)$$

  - easy to incorporate graph constraints

Edge Features

- Tokens
- Lemmas
- POS tags
- Linear Order (left or right)
- Linear Distance
- Dependency Path
- Up/Down Dependency Path
- Up/Down/Left/Right Dependency Path
- Is Parent
- Dependency Path Length
- POS Context
- Subcategorization Sequence
- Subcategorization Sequence with POS
- POS Path
- Distance Thresholds

Results

	LP	LR	LF	LM
DM	0.8446	0.8348	0.8397	0.0875
PAS	0.9078	0.8851	0.8963	0.2604
PCEDT	0.7681	0.7072	0.7364	0.0712
Average	0.8402	0.8090	0.8241	0.1397

Discussion

Optimal choice of the constraints depends on the formalism, and evaluation metric

Missing (important) higher-order features