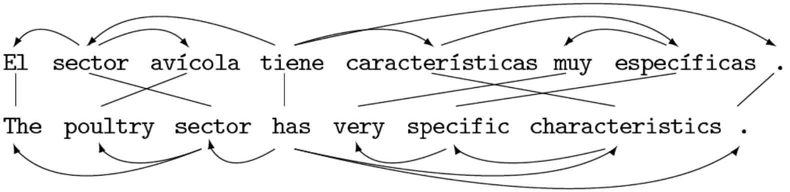
Dependency Grammar Induction via Bitext Projection Constraints

Kuzman Ganchev   Jennifer Gillenwater   Ben Taskar

Computer & Information Science
University of Pennsylvania

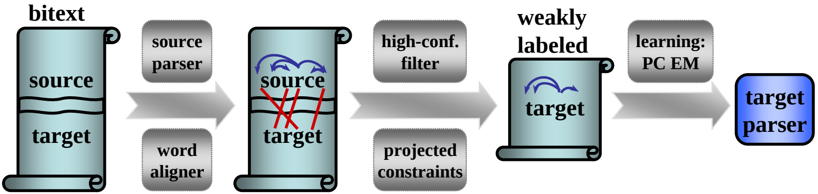
May 11, 2009

Overview



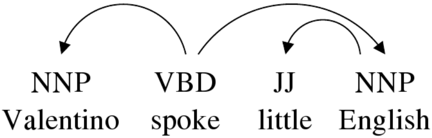
- Goal: Automate creation of linguistic resources
- Method: Use parallel corpora to bootstrap parser learning

Overview



Require most projected dependencies be exhibited in learned parses

Generative Model



$$
\begin{gathered}
p_\theta(x, y) = \theta_{root(VBD)} \\
\cdot \theta_{continue(VBD, right, false)} \cdot \theta_{child(VBD, right, NNP)} \\
\cdot \theta_{stop(VBD, right, true)} \cdot \theta_{stop(NNP, right, false)} \\
\cdot \theta_{continue(VBD, left, false)} \cdot \theta_{continue(NNP, left, false)} \\
\cdot \theta_{child(VBD, left, NNP)} \cdot \theta_{child(NNP, left, JJ)} \\
\cdot \theta_{stop(NNP, right, false)} \cdot \theta_{stop(VBD, left, true)} \\
\cdot \theta_{stop(JJ, right, false)} \cdot \theta_{stop(JJ, left, false)} \\
\cdot \theta_{stop(NNP, left, false)} \cdot \theta_{stop(NNP, left, true)}
\end{gathered}
$$

N. Smith. Ph.D. thesis, 2006.

Discriminative Model

$$p_\theta(\mathbf{y} \mid \mathbf{x}) \propto \prod_{y \in \mathbf{y}} e^{\theta \cdot \phi(y, \mathbf{x})}$$

Example feature vector ϕ

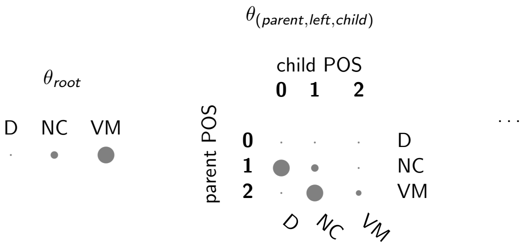
$$\begin{pmatrix} \vdots \\ \mathbf{1}(\text{child-POS} = \text{D, child-word} = \text{el, parent-POS} = \text{NC}) \\ \mathbf{1}(\text{child-POS} = \text{D, between-POS} = \text{AQ, parent-POS} = \text{NC}) \\ \mathbf{1}(\text{pre-child-POS} = \text{VM, child-POS} = \text{D, parent-POS} = \text{NC}) \\ \mathbf{1}(\text{child-POS} = \text{D, parent-POS} = \text{NC, post-parent-POS} = \text{VM}) \\ \vdots \end{pmatrix}$$

R. McDonald et. al. *Online Large-Margin Training of Dependency Parsers*, 2005.

Expectation Maximization

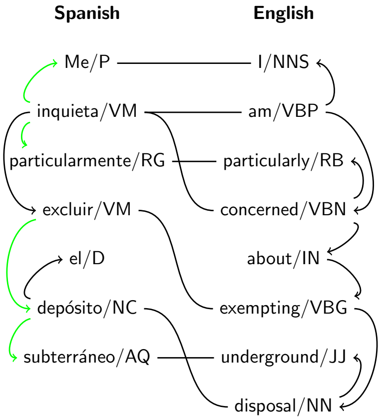
E-Step $\arg\min_{q(y)} \mathrm{KL}(q(y) \parallel p_{\theta^t}(y \mid x)) = p_{\theta^t}(y \mid x)$

$q(y) =$



M-Step $\arg\max_{\theta} E_X[q(y) \log p_\theta(x, y)]$

Constrained EM



- $E_q[f(x,y)] \geq c$
- $f(x,y) = \#$ of projected dependencies realized in parse y
- $c =$ lower limit on feature expectation

Constrained EM

E-Step

$$\underset{q(y) \in \mathcal{Q}(x)}{\arg\min} \; \mathrm{KL}(q(y) \parallel p_{\theta^t}(y \mid x))$$

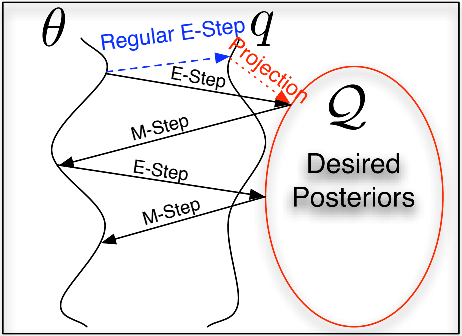
M-Step

$$\arg\max_{\theta} E_X[q(y) \log p_\theta(x, y)]$$

- $x =$ words and POS tags, $y =$ dependency parse
- $\theta =$ model parameters
- $f =$ a feature, $c =$ lower limit on feature expectation
- $\mathcal{Q} = \{q : E_q[f(x, y)] \geq c\}$

J. Graca, K. Ganchev, B. Taskar. *EM and Posterior Constraints*, 2008.

Constrained EM



Objective: $$\arg\max_{\theta} \left(L(\theta) - E_X[\mathrm{KL}(\mathcal{Q}(x) \parallel p_\theta(y \mid x))] \right)$$

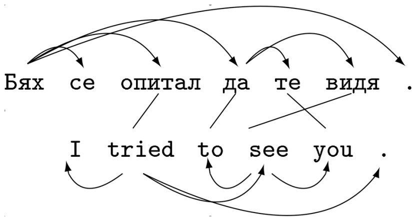
J. Graca, K. Ganchev, B. Taskar. *EM and Posterior Constraints*, 2008.

Alignment Pre-Processing

- Corpora — Bulgarian subtitles, Spanish Europarl
- Remove alignments if POS don't belong to same category

Me/P	I/NNS
inquieta/VM	am/VBP
particularmente/RG	particularly/RB
excluir/VM	concerned/VBN
el/D	about/IN
depósito/NC	exempting/VBG
subterráneo/AQ	underground/JJ
	disposal/NN

Corrective Rules for Bulgarian



- "da" should dominate words until next verb, and adopt their children
- Auxiliary verb should be parent of main verb
- Similar rules for 5 more words like "da"

Corrective Rules for Spanish

Ésa	es	la	primera	cuestión
p	vs	d	ao	nc

He	recibido	algunas	preguntas	sobre	los	billetes	de	banco
va	vm	d	nc	sp	d	nc	sp	nc

- Main verb should be parent of auxiliary verb
- First element in adjective-noun or noun-adjective pair should be parent of other, and adopt other's children

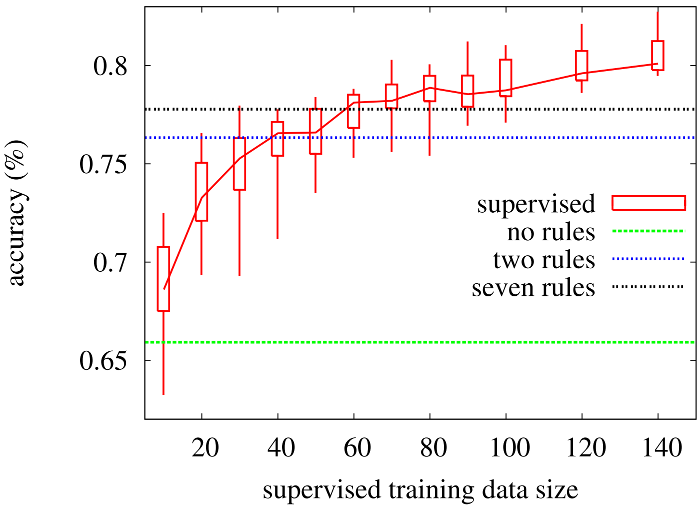
No Rules Results

Constraint: In expectation, at least 70% of projected dependencies must appear in Bulgarian/Spanish parses.

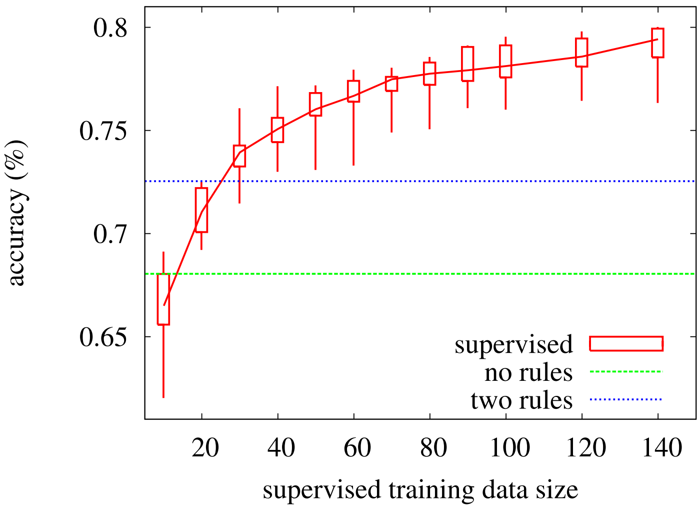
Language	Link-left	Gener.	Discrim.
Bulgarian	33.8 %	61.9%	65.9%
Spanish	27.9 %	55.6%	68.1%

- Train sets: 10k parallel sentences of length ≤ 20
- Test sets: CoNLL train, sentences of length ≤ 10

Bulgarian Discriminative Results



Spanish Discriminative Results



Top Bulgarian Errors

child POS			parent POS	
	acc(%)	errors		errors
N	75.1	1839	N/V	1078
P	70.2	1223	V/V	607
V	84.4	1004	R/V	533
R	79.0	678	V/N	482

- V verb, N noun, P pronoun, R preposition, T particle
- Accuracies are by child or parent truth/guess POS tag

Related Work on Spanish

Hwa et. al.

- Special projection for each of one-to-many, many-to-one, and many-to-many alignments
- Filtered sentences where
  - $<$ 30% of words aligned
  - one-to-many alignment was too unbalanced
- Used extensive set of language-specific rules (only 37% accuracy before rules)

Best performance $\approx$ 72% for both methods (though corpora differ)

Hwa et. al. *Bootstrapping Parsers via Syntactic Projection across Parallel Texts*, 2004

Conclusion

- Equivalent of supervised methods with limited training data
- Using constrained EM allows for
  - Fewer language-specific rules
  - Learning from partial projected parses
- Further improvement by adding more complex constraints?
  - Grandparent or other long-range chains
  - Surface length for a particular POS tag