Sparsity in Dependency Grammar Induction

Jennifer Gillenwater[1] Kuzman Ganchev[1] João Graça[2]
Ben Taskar[1] Fernando Pereira[3]

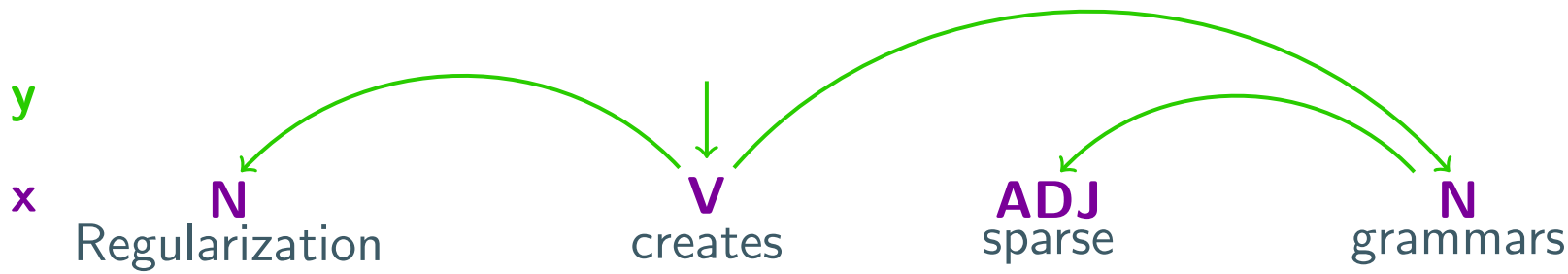
[1]Computer & Information Science
University of Pennsylvania

[2]L[2]F INESC-ID, Lisboa, Portugal

[3]Google, Inc.

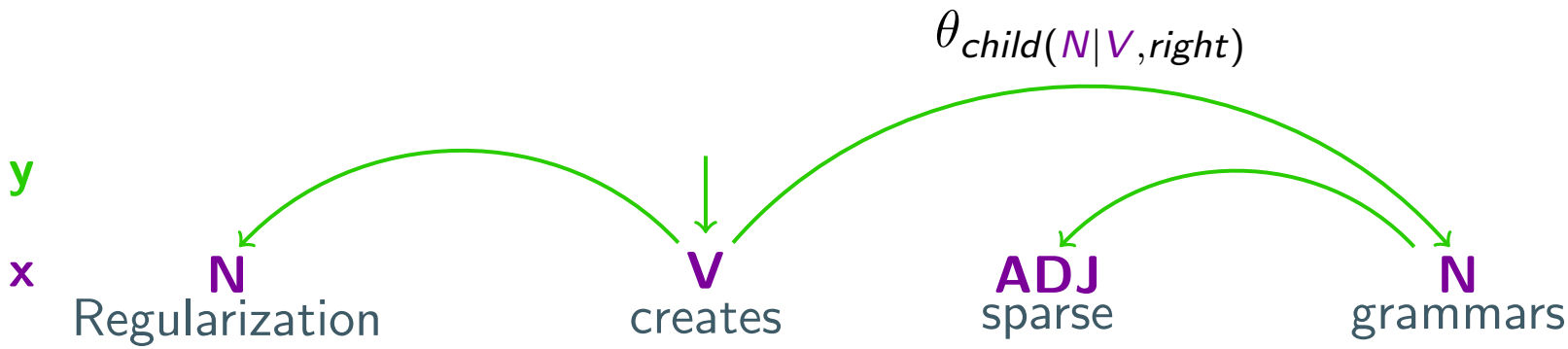
Dependency model with valence (Klein and Manning, ACL 2004)

$$p_\theta(\mathbf{x}, \mathbf{y})$$



Dependency model with valence (Klein and Manning, ACL 2004)

$$p_\theta(\mathbf{x}, \mathbf{y})$$



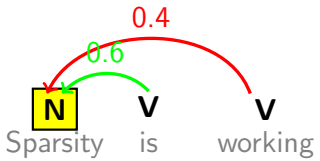
Traditional objective optimization

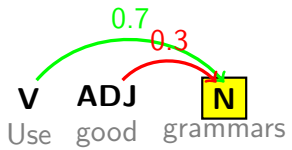
- **Traditional optimization**: expectation maximization (EM)

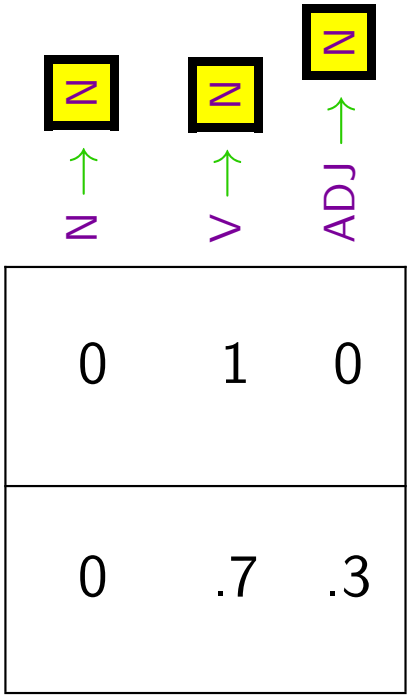
Traditional objective optimization

- **Traditional optimization**: expectation maximization (EM)
- **Problem**: EM may learn a very ambiguous grammar
  - V $\rightarrow$ N should have non-zero probability, but ...
  - V $\rightarrow$ DET, V $\rightarrow$ JJ, V $\rightarrow$ PRP$, etc. should be 0

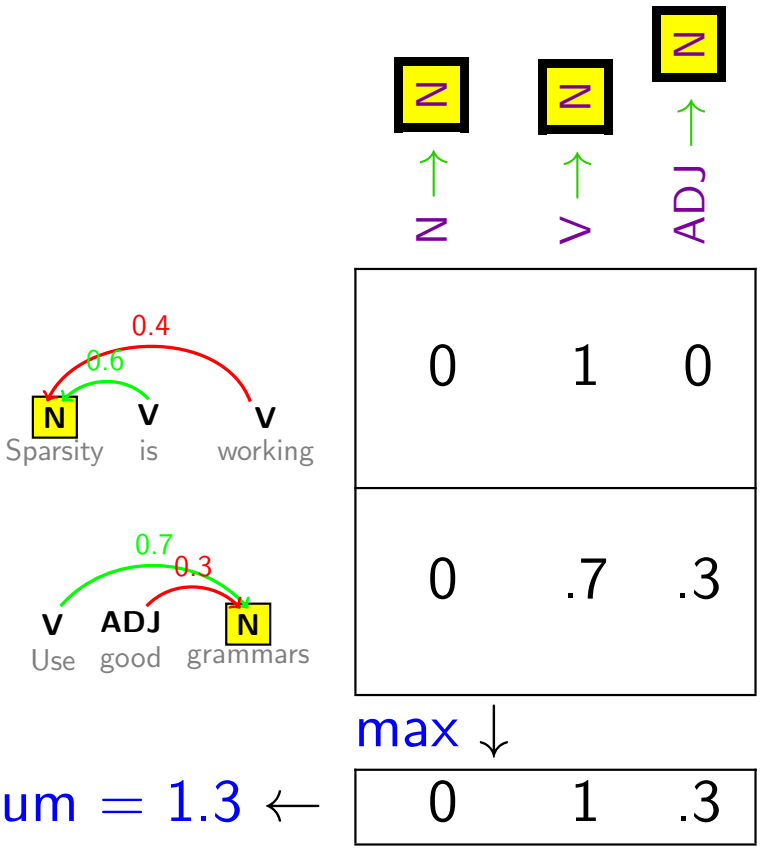
Measuring ambiguity on distributions over trees







Measuring ambiguity on distributions over trees



Minimizing ambiguity through posterior regularization

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

Minimizing ambiguity through posterior regularization

E-Step $$q^t(\mathbf{y} \mid \mathbf{x}) = \arg\min_{q(\mathbf{y}|\mathbf{x})} KL(q \parallel p_{\theta^t}) + \sigma L_{1/\infty}(q(\mathbf{y} \mid \mathbf{x}))$$

Experimental results

- English from Penn Treebank: state-of-the-art accuracy

	Accuracy		
Learning Method	≤ 10	≤ 20	all
EM	45.8	40.2	35.9
Sparsifying Dirichlet Prior	46.4	40.9	36.5
PR ($\sigma = 140$)	**62.1**	**53.8**	**49.1**

Experimental results

- English from Penn Treebank: state-of-the-art accuracy

	Accuracy		
Learning Method	≤ 10	≤ 20	all
EM	45.8	40.2	35.9
Sparsifying Dirichlet Prior	46.4	40.9	36.5
PR ($\sigma = 140$)	**62.1**	**53.8**	**49.1**

- 11 other languages from CoNLL-X:
  - Dirichlet prior baseline: 1.5% average gain over EM
  - Posterior regularization: 6.5% average gain over EM

Experimental results

- English from Penn Treebank: state-of-the-art accuracy

	Accuracy		
Learning Method	≤ 10	≤ 20	all
EM	45.8	40.2	35.9
Sparsifying Dirichlet Prior	46.4	40.9	36.5
PR ($\sigma = 140$)	**62.1**	**53.8**	**49.1**

- 11 other languages from CoNLL-X:
  - Dirichlet prior baseline: 1.5% average gain over EM
  - Posterior regularization: 6.5% average gain over EM
- **Come see the poster for more details**