Turning the pipeline into a loop:
Iterated unsupervised dependency parsing and PoS induction

THE UNIVERSITY of EDINBURGH informatics
ilcc Institute for Language, Cognition and Computation

EPSRC Engineering and Physical Sciences Research Council

Christos Christodoulopoulos
School of Informatics
University of Edinburgh
christos.c@ed.ac.uk

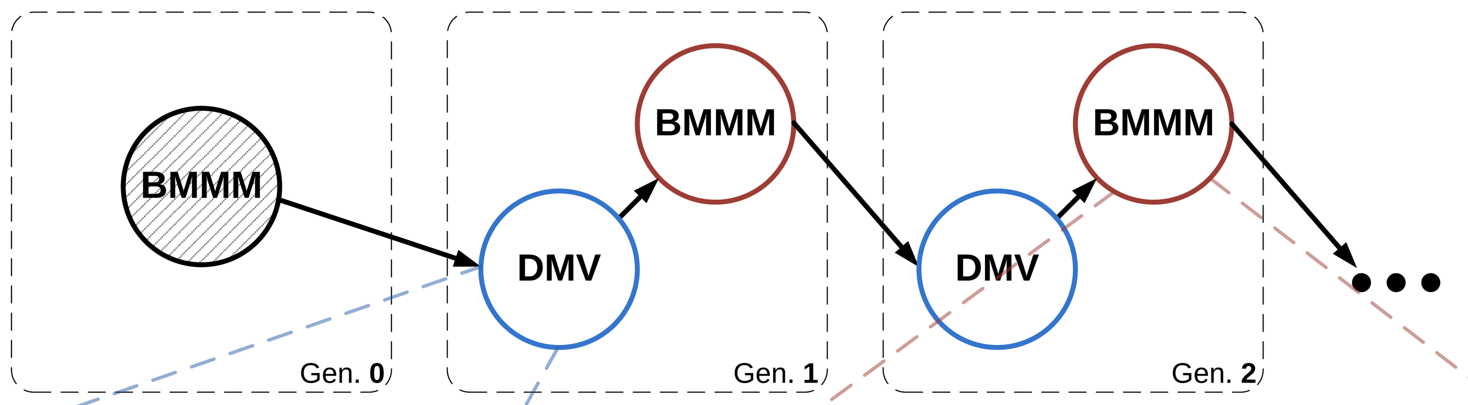
Sharon Goldwater
School of Informatics
University of Edinburgh
sgwater@inf.ed.ac.uk

Mark Steedman
School of Informatics
University of Edinburgh
steedman@inf.ed.ac.uk

Motivation

- Traditional view: PoS → Dependencies
  - Directly, instead of words (Klein and Manning, 2004)
  - Back-off mechanism (Headden et al., 2009)
- A better approach: PoS ↔ Dependencies
  - Evidence from supervised systems (Cohen et al., 2011 *inter alia*)
  - Taggers (in most cases) rely on local contextual features
  - Parsers use richer hierarchical features than taggers
- Our developmental experiments on WSJ show that dependencies can be useful for unsupervised PoS induction

Iterated Learning

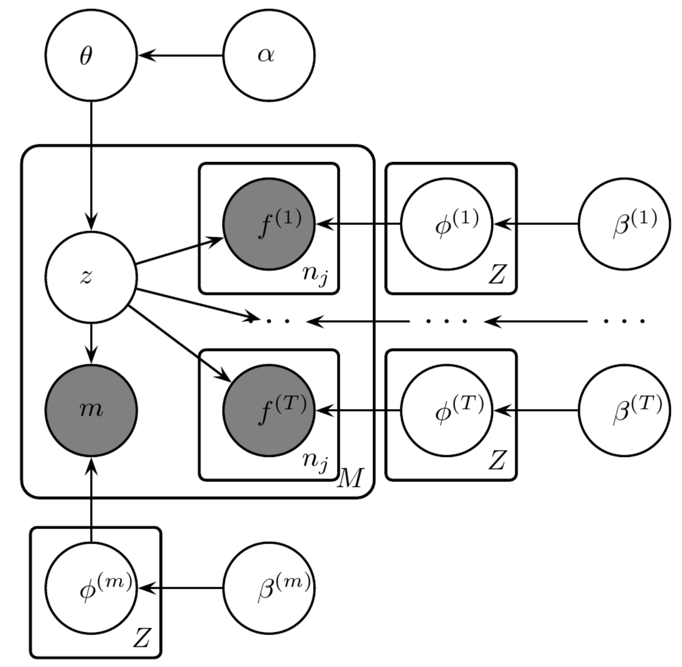


DMV (Klein and Manning, 2004)

- Dependencies generated based on 3 probability decisions (3 probability distributions)
  - P_{ORDER} Direction of children attachment (left or right)
  - P_{STOP} Stop attaching more children (true or false)
  - P_{ATTACH} Attach a specific child node
- $P_{sentence} = \Sigma\, P(D_{ROOT})$ for all derivations headed by **ROOT**
- Basis for most unsup. dependency parsing systems
- Simple, easy to train with Inside-Outside
- Can be used with unsupervised PoS tags as input

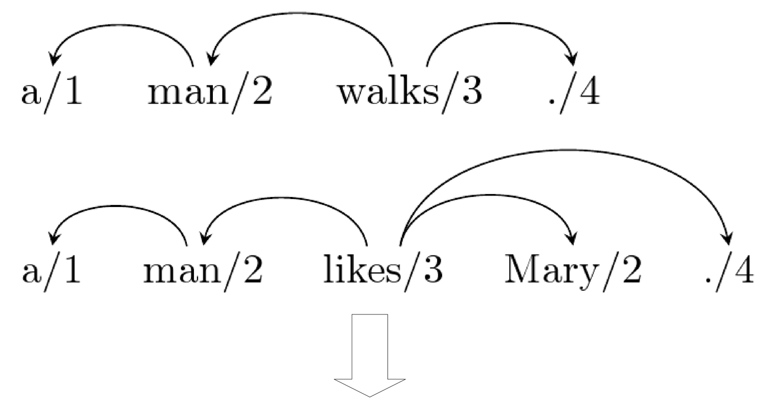
BMMM (Christodoulopoulos et al., 2011)

- Observed variables are:
  - Token-level features (e.g. left-right context, $f^{1 \ldots T}$)
  - Type-level features (morphology, m)
- Mixture of multinomials
  - θ for the word classes
  - $\phi^{(i)}$ for each feature
  - Dirichlet priors (α and β respectively)
- Collapsed Gibbs sampling for inference



Using dependencies as features (gen. 1 and above)

- Assuming DMV parsed corpus (with PoS tags from gen. 0)

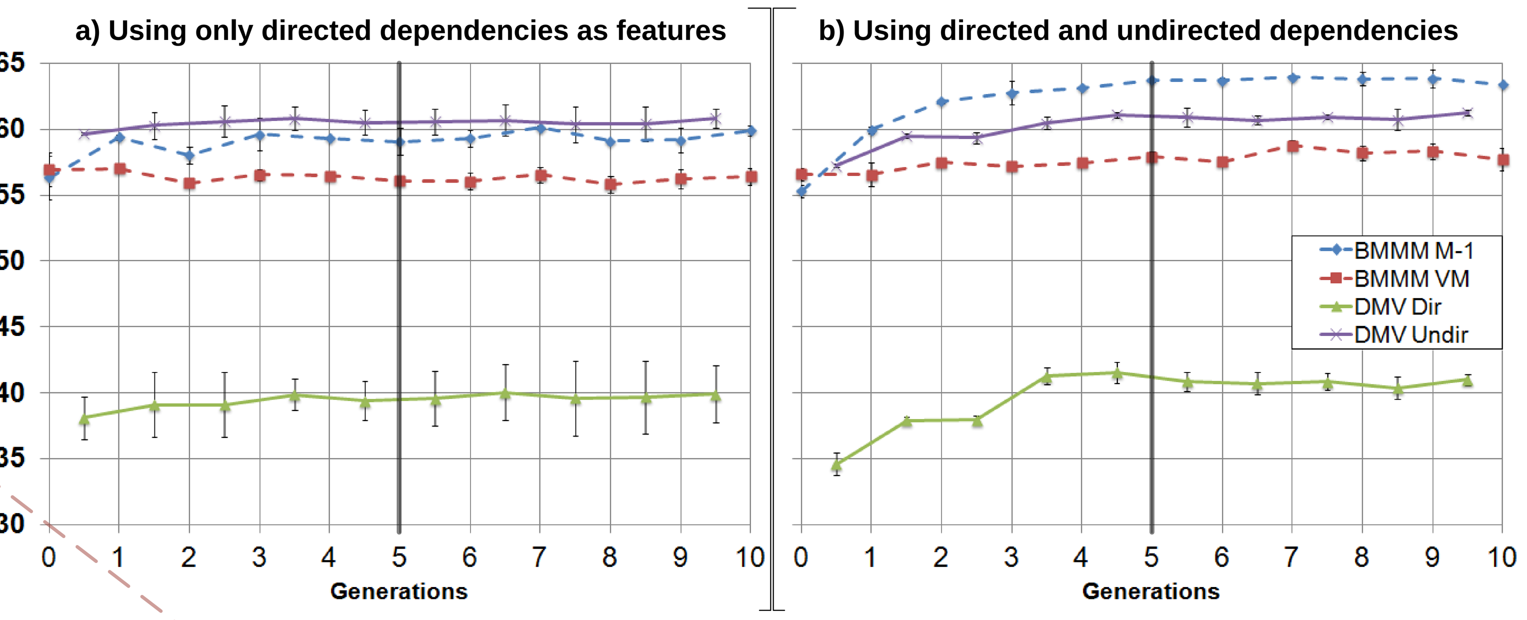


	1	2	3	4
.			2	
a		2		
likes		2		1
man	2		2	
Mary			1	
walks		1		1

- Features are # times $word_1$ heads $word_2$
- Group by PoS tag (to reduce sparsity)
- Use undir. dependencies (adding the reverse counts) – see right panel

Experimental Design

- Developmental results on WSJ10



(a) Using only directed dependencies:
- Some improvement in many-to-1 after 1 generation
- Other metrics no improvement

(b) Using directed and undirected dependencies:

After the first 5 generations:
- 8.5% increase in M-1
- 1.3% increase in VM

for the PoS inducer
- 7% increase in directed dep. accuracy
- 3.8% increase in undirected dep. Accuracy

for the parser

Experimental setup

- 5 generations of the combined system
- For the DMV
  - 20 training iterations
  - "Harmonic initializer" with parameters: $P_{ORDER} = 0.5$, $P_{STOP=T} = 0.25$, $P_{STOP=F} = 0.75$, $P_{ATTACH}(\alpha|h,\cdot) = 1/(1+dist(\alpha,h))$
- For the BMMM:
  - 500 sampling iterations
  - dir. + undir. dependencies
  - ±1 context words (100 most frequent)
  - suffixes extracted from Morfessor (Creutz and Lagus, 2005)
  - extended morphology features (Haghighi and Klein, 2006)

References

Christos Christodoulopoulos, Sharon Goldwater, and Mark Steedman. 2011. A Bayesian mixture model for pos induction using multiple features. In *Proceedings of EMNLP*, pages 638–647.

Shay B. Cohen, Dipanjan Das, and Noah A. Smith. 2011. Unsupervised structure prediction with non-parallel multilingual guidance. In *Proceedings of EMNLP*, pages 50–61.

Mathias Creutz and Krista Lagus. 2005. Inducing the morphological lexicon of a natural language from unannotated text. In *Proceedings of AKRR*, pages 106–113.

Aria Haghighi and Dan Klein. 2006. Prototype-driven learning for sequence models. In *Proceedings of NAACL*, pages 320–327.

William P. Headden, Mark Johnson, and David McClosky. 2009. Improving unsupervised dependency parsing with richer contexts and smoothing. In *Proceedings of NAACL*, pages 101–109.

Dan Klein and Christopher D. Manning. 2004. Corpus Based induction of syntactic structure: models of dependency and constituency. In *Proceedings of ACL*, pages 478–485.