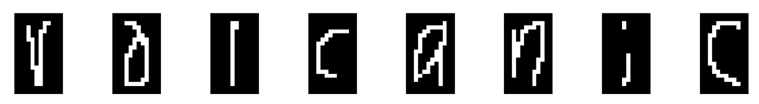
Reified Context Models

Jacob Steinhardt **Percy Liang**

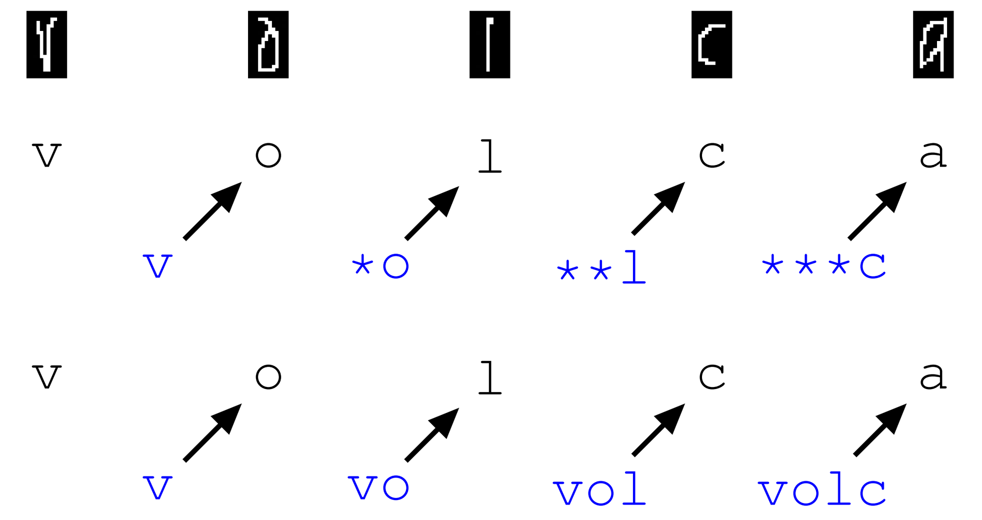
{jsteinhardt,pliang}@cs.stanford.edu

Structured Prediction Task

input x: 

output y: v o l c a n i c

Expressivity and Coverage



Key idea: contexts!

$$\star\text{o} \overset{\text{def}}{=} \left\{ \begin{matrix} \text{ao} \\ \text{bo} \\ \text{co} \\ \vdots \end{matrix} \right\}$$

r ro rol rolc
v ra ral ralc

- expressivity (long contexts)
  - capture complex dependencies

r *o **l ***c
v *a **i ***r

- coverage (short contexts)
  - better uncertainty estimates (precision)
  - stabler partially supervised learning updates

Reifying Contexts

input x: 

output y:	v	o	l	c	a n i	c
context c:	v	*o	*ol	*olc	······	
	r	ro	rol	*olc		
	v	ra	ral	***c	←**"context sets"**	
	y	*o	*ol	***r		
	*	**	***	****		
	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_4$		

Challenge: how to trade off contexts of different lengths?
$\Longrightarrow$ *Reify* contexts as part of model!

Reified Context Models

Given:

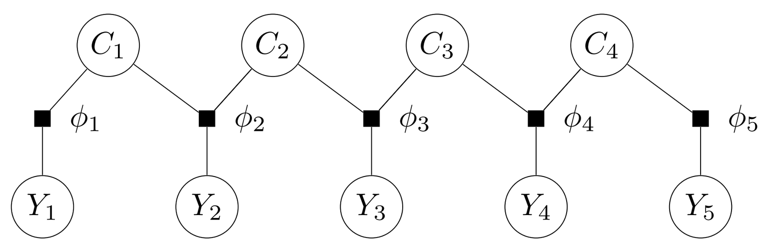
- context sets $\mathcal{C}_1, \dots, \mathcal{C}_L$
- features $\phi_i(c_{i-1}, y_i)$

Define the model

$$p_\theta(y_{1:L}, c_{1:L-1}) \propto \exp\left(\sum_{i=1}^{L} \theta^\top \phi_i(c_{i-1}, y_i)\right) \cdot \underbrace{\kappa(y, c)}_{\text{consistency}}$$

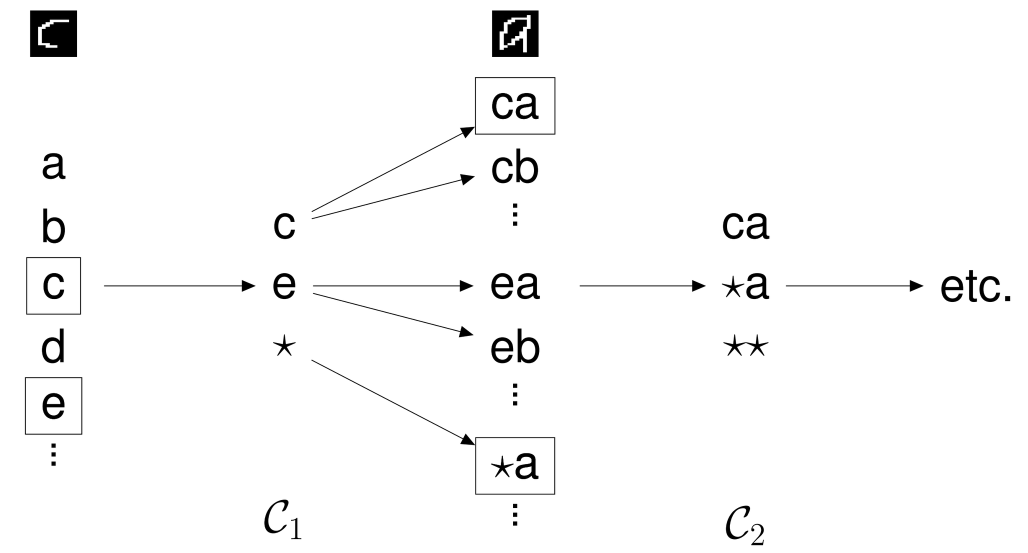
Graphical model structure:

inference via forward-backward!



Adaptive Context Selection

- Select context sets $\mathcal{C}_i$ during forward pass of inference
- Greedily select contexts with largest mass



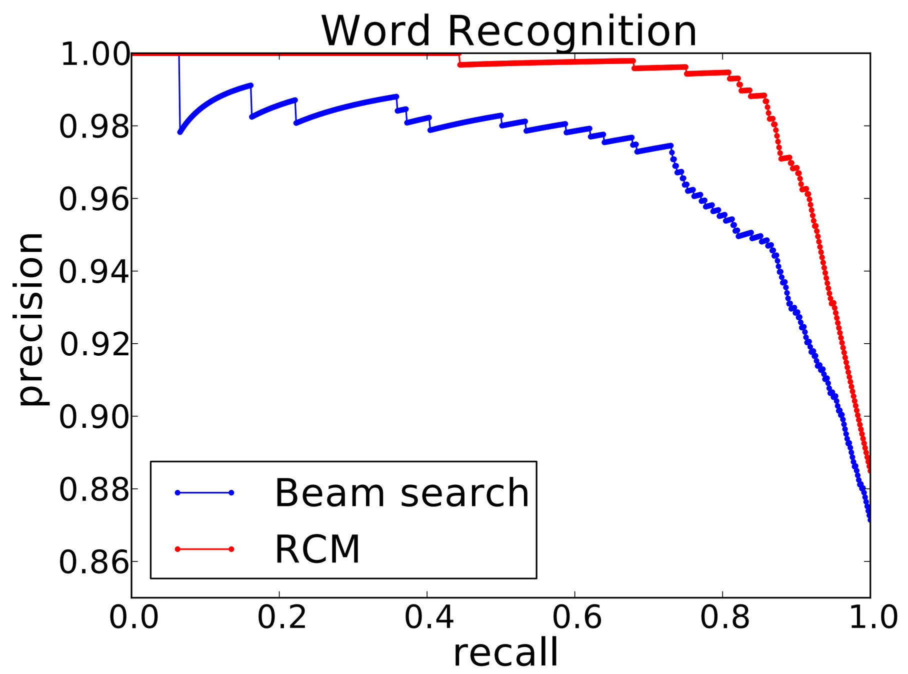
Biases towards short contexts unless there is high confidence.

Precision

Model assigns probability to each prediction, so can predict on most confident subset.

Measure precision (# of correct words) vs. recall (# of words predicted).

- comparison: beam search



Partially Supervised Learning

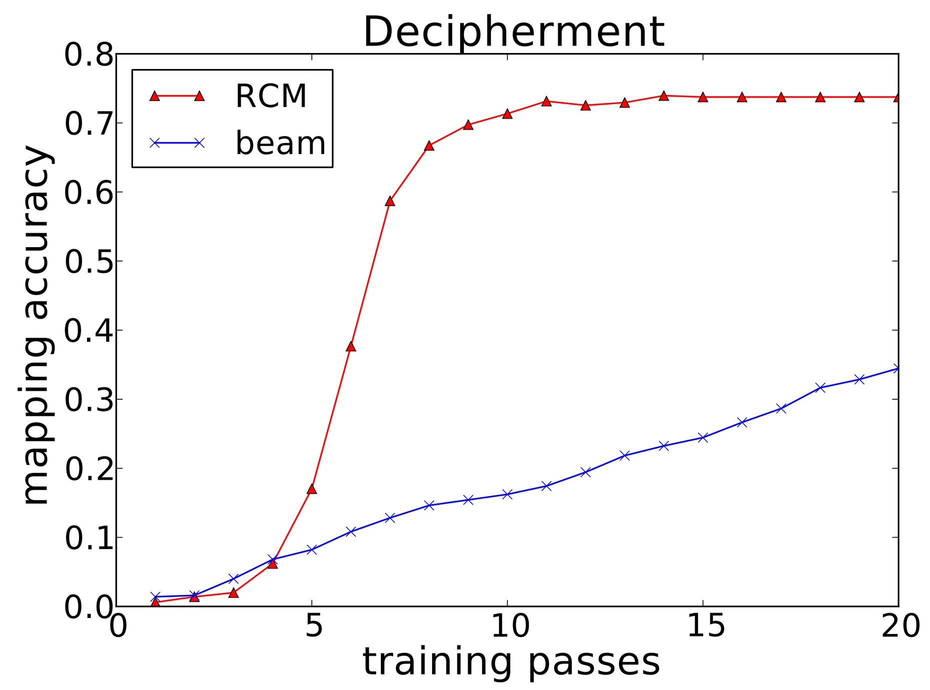
Decipherment task:

cipher	am ↦ 5, I ↦ 13, what ↦ 54, ...				
latent z	I	am	what	I	am
output y	13	5	54	13	5

Goal: determine cipher

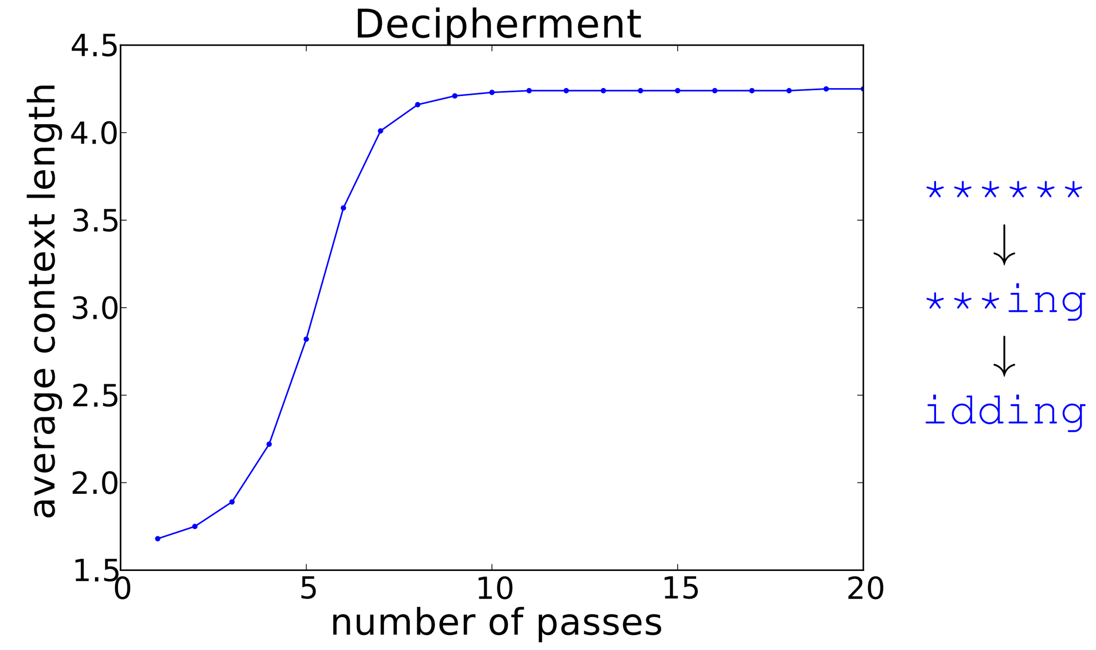
Fit 2nd-order HMM with EM, using RCMs for approximate E-step.

- use learned emissions to determine cipher.
- again compare to beam search



Contexts During Training

Context lengths increase smoothly during training:



Start of training: little information, short contexts.
End of training: lots of information, long contexts.

Discussion

RCMs provide both expressivity and coverage, which enable:

- More accurate uncertainty estimates (precision)
- Better partially supervised learning updates

Reproducible experiments on Codalab: codalab.org/worksheets

The first author was supported by the Hertz foundation and by the NSF.

International Conference on Machine Learning (ICML), Lille 2015