Reified Context Models
Jacob Steinhardt  Percy Liang
{jsteinhardt,pliang}@cs.stanford.edu

Structured Prediction Task
input $x$: 1 0 1 C a 0 1 C
output $y$: volcanic

Expressivity and Coverage
V 0 1 C a
DP: V o l C
beam search:
Key idea: contexts!

Reifying Contexts
input $x$: 1 0 1 C a 0 1 C
output $y$: volcanic
context $c$: r r o r a r l r a l c v r r a r l r a l c
"context sets"

Challenge: how to trade off contexts of different lengths?
→ Reify contexts as part of model!

Reified Context Models
Given:
• context sets $C_1, \ldots, C_L$
• features $\phi_i(c_{i-1}, y_i)$
Define the model
$$p(y_1, c_1, \ldots, c_L) \propto \exp \left( \sum_{i=1}^L \theta^T \phi_i(c_{i-1}, y_i) \right) \cdot p(y, c)$$
Graphical model structure:

Adaptive Context Selection
• Select context sets $C_i$ during forward pass of inference
• Greedily select contexts with largest mass
  $$C_1 \supseteq C_2 \supseteq \ldots$$

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, etc.
output $y$: 13 5 54 13 5
Goal: determine cipher
Fit 3rd-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:

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

Reproducible experiments on Codalab:
codalab.org/worksheets