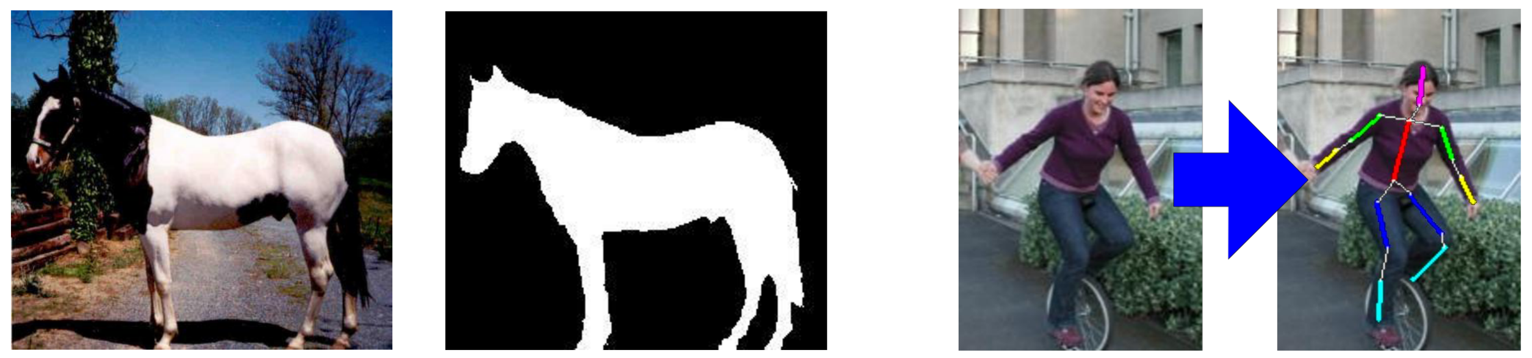
Yujia Li
Daniel Tarlow
Richard Zemel
University of Toronto, Canada

Exploring Compositional High Order Pattern Potentials for Structured Output Learning

Structured Output Learning

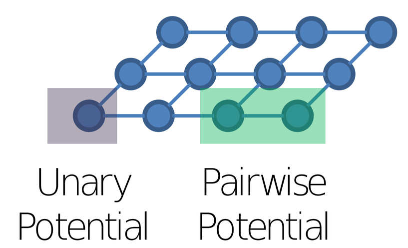
- Real applications require structured prediction



Figures: Weizmann horses dataset and http://www.vision.ee.ethz.ch/~hpedemo/fullhpedemo.png

- Standard Model: Pairwise MRF/CRF

$$E(\mathbf{y}) = \sum_i f_i^u(y_i) + \sum_{ij} f_{ij}^p(y_i, y_j)$$



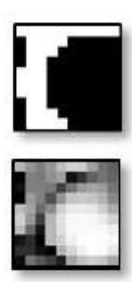
  - Sparse connection - easier learning and inference
  - Overly simplistic - only modeling pairwise correlations

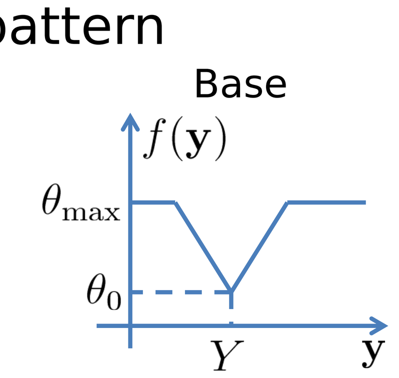
Pattern Potentials

- Penalize linearly if output deviates from a pattern

$$d(\mathbf{y}) = \sum_i abs(w_i)\mathbf{I}[y_i \neq Y_i]$$

$$f(\mathbf{y}) = \min\{d(\mathbf{y}) + \theta_0, \theta_{\max}\}$$



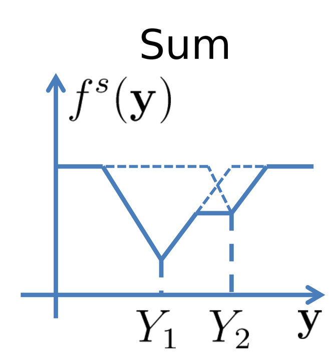


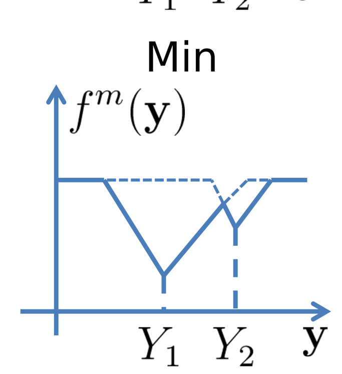
- Combine base models
  - Sum

$$f^s(\mathbf{y}) = \sum_j \min\{d_j(\mathbf{y}) + \theta_j, \theta_{\max}\}$$

  - Min

$$f^m(\mathbf{y}) = \min\{d_1(\mathbf{y}) + \theta_1, ..., d_J(\mathbf{y}) + \theta_J, \theta_{\max}\}$$





RBMs are like Pattern Potentials

$$E(\mathbf{y}, \mathbf{h}) = -\sum_{ij} w_{ij} y_i h_j - \sum_i b_i y_i - \sum_j c_j h_j$$

$$F(\mathbf{y}) = -\sum_i b_i y_i - \log\left(\sum_{\mathbf{h}} \exp\left(\sum_j \left(c_j + \sum_i w_{ij} y_i\right) h_j\right)\right)$$

Standard RBM

$$-\sum_j \log(1 + \exp(c_j + w_{ij} y_i))$$

$$\min\{x, 0\} \approx -\log(1 + \exp(-x))$$

1-of-J constraint

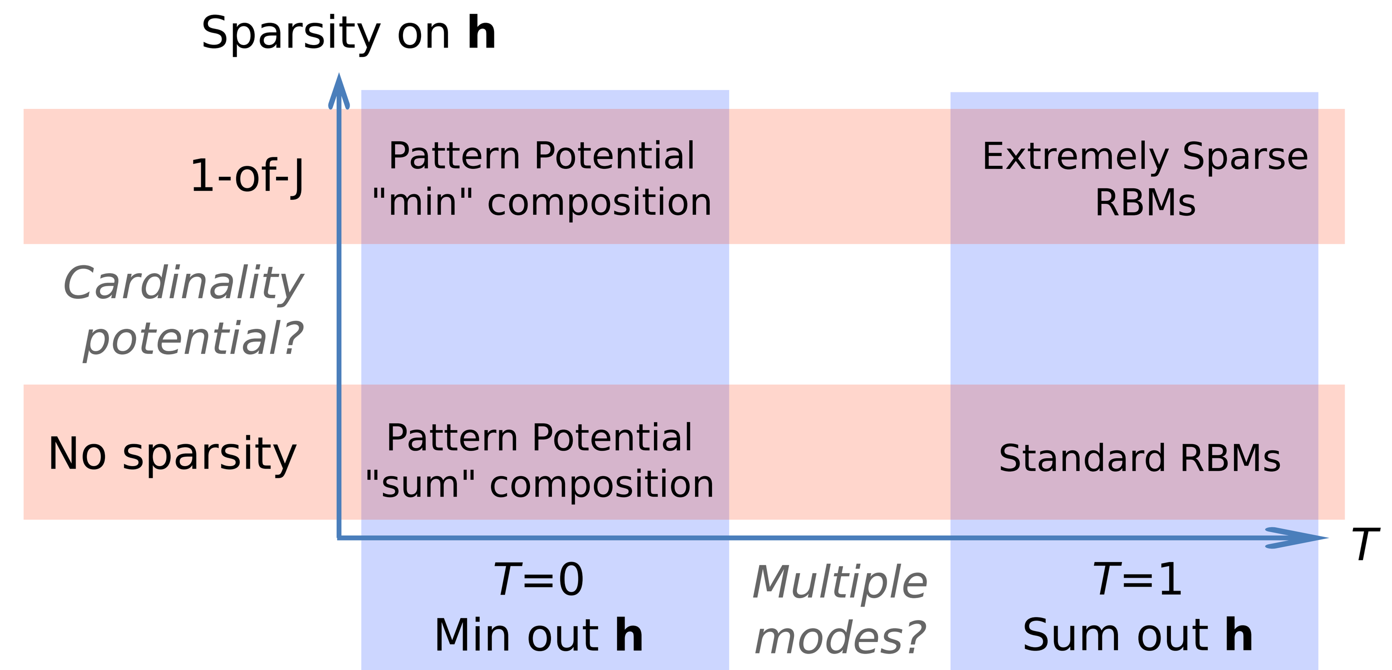
$$-\log\left(1 + \sum_j \exp(c'_j + w'_{ij} y_i)\right)$$

$$\min\{x_1, ..., x_J, 0\} \approx -\log\left(1 + \sum_j \exp(-x_j)\right)$$

CHOPP

$$f(\mathbf{y}; T) = -T \log\left(\sum_{\mathbf{h}} \exp\left(\frac{1}{T}\sum_j \left(c_j + \sum_i w_{ij} y_i\right) h_j\right)\right)$$

- CHOPP interpolates RBMs and Pattern Potentials, as well as different composition strategies



- CHOPP-augmented CRFs

$$-E(\mathbf{y}; T) = f^u(\mathbf{y}) + f^p(\mathbf{y}) + \sum_i b_i y_i + T \log\left(\sum_{\mathbf{h}} \exp\left(\frac{1}{T}\sum_j \left(c_j + \sum_i w_{ij} y_i\right) h_j\right)\right)$$

MAP Inference with the "EM" Algorithm

- Variational bound

$$-E(\mathbf{y}; T) \geq f^u(\mathbf{y}) + f^p(\mathbf{y}) + \sum_i b_i y_i + \sum_{\mathbf{h}} q(\mathbf{h}) \sum_j \left(c_j + \sum_i w_{ij} y_i\right) h_j + H(q)$$

- E-step: compute optimal q(**h**) with **y** fixed

$$q(\mathbf{h}) = \frac{\exp\left(\frac{1}{T}\sum_j (c_j + \sum_i w_{ij} y_i) h_j\right)}{\sum_{\mathbf{h}} \exp\left(\frac{1}{T}\sum_j (c_j + \sum_i w_{ij} y_i) h_j\right)}$$

- M-step: change **y** with q fixed

$$\sum_i \left(b_i + \sum_j w_{ij} \mathbb{E}_q[h_j]\right) y_i + f^u(\mathbf{y}) + f^p(\mathbf{y})$$

New unary potential

This is a pairwise CRF!

- The EM algorithm always increases the bound

Learning CHOPP Parameters

- Minimize expected loss

$$L = \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \ell(\mathbf{y}, \mathbf{y}^*)$$

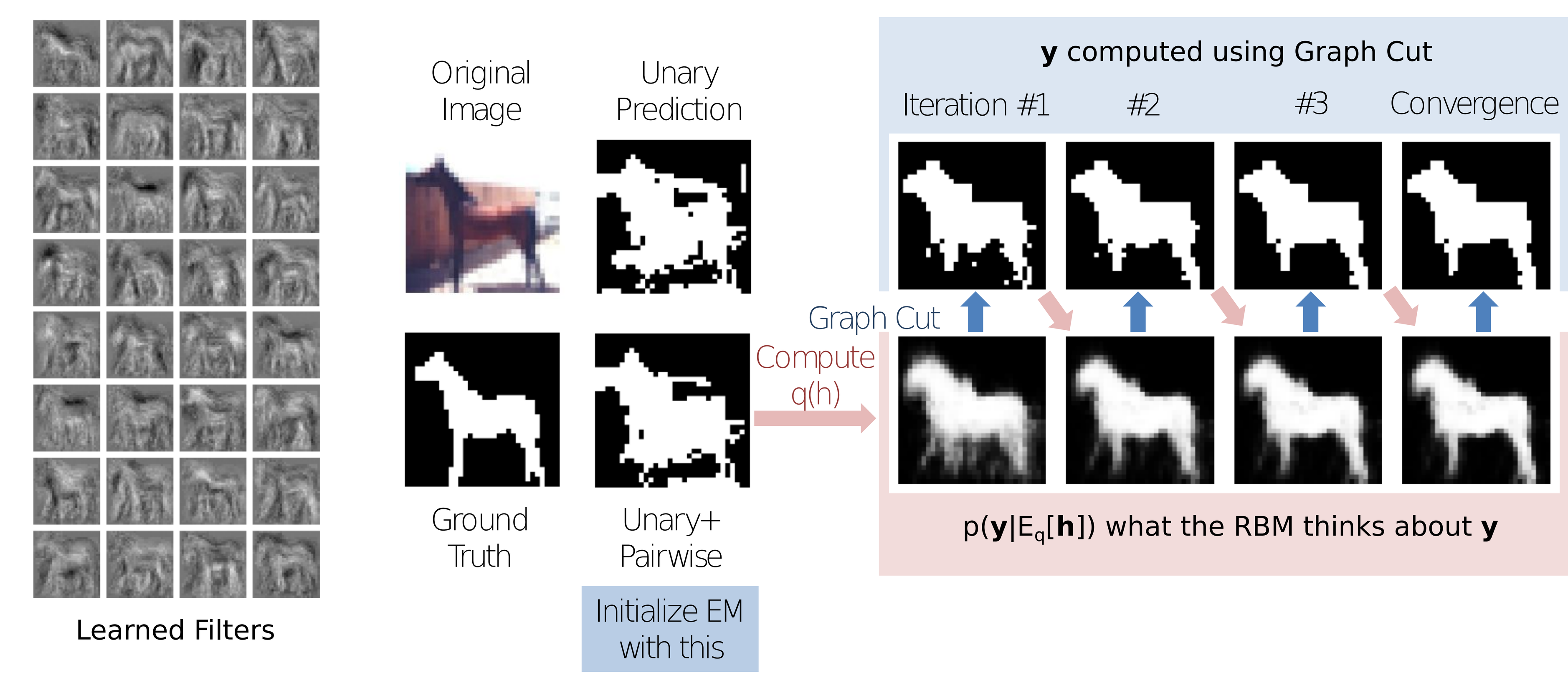
- Follow the negative gradient estimated by a set of samples

$$\frac{\partial L}{\partial \theta} \approx \frac{1}{N-1} \sum_{n=1}^{N} \left(\ell(\mathbf{y}^n, \mathbf{y}^*) - \frac{1}{N}\sum_{n'=1}^{N} \ell(\mathbf{y}^{n'}, \mathbf{y}^*)\right)\left(-\frac{\partial E(\mathbf{y}^n)}{\partial \theta}\right)$$

  - Increase energy for samples with high loss
  - Decrease energy for samples with low loss

Experiments

- An example using RBM trained with CD



- More experiments

Method	Horse IOU	Bird IOU	Person IOU
Unary Only	0.5119	0.5055	0.4979
iPW	0.5736	0.5585	0.5094
iPW+RBM	0.6722	0.5647	0.5126
iPW+jRBM	**0.6990**	**0.5773**	**0.5253**