Exploring Compositional High Order Pattern Potentials for Structured Output Learning

Yujia Li, Daniel Tarlow*, Richard Zemel

University of Toronto
*Now at Microsoft Research Cambridge
June 25, 2013
Structured Output Learning

• Lots of real world applications require structured outputs
  – Image segmentation, pose estimation, sequence labeling, etc.

Figures from Weizmann horse dataset
Structured Output Learning

• Lots of real world applications require structured outputs
  - Image segmentation, pose estimation, sequence labeling, etc.

• Standard model – pairwise MRF/CRF

\[ E(y) = \sum_{i} f_{i}^{u}(y_{i}) + \sum_{i,j} f_{ij}^{p}(y_{i}, y_{j}) \]
  - Sparse connections – easier to learn and do inference
  - Overly simplistic – only modeling up to 2nd order correlation in outputs

Figures from Weizmann horse dataset
Moving to More Expressive Models

- Densely connected CRFs [P. Krahenbuhl et al. NIPS’12]
  - Still 2nd order connections but densely connected

- Robust High Order Potentials [P. Kohli et al. CVPR’08]
  - Smoothness in a region

- Global Connectivity Potentials [S. Nowozin et al. CVPR’09]
  - Require the output to be connected

- Pattern Potentials [C. Rother et al. CVPR’09]
  - Consistency between the output and learned patterns
Pattern Potentials

• Penalize linearly if output deviates from a pattern

\[ d(y) = \sum_i abs(w_i)I[y_i \neq Y_i] \]

\[ f(y) = \min\{d(y) + \theta_0, \theta_{\text{max}}\} \]

• Multiple base pattern potentials can be combined to form more expressive composite pattern potentials

Pattern and weight figures: C. Rother et al. CVPR'09
Restricted Boltzmann Machines (RBMs)

- **RBM probabilistic model**
  
  \[ E(y, h) = - \sum_{ij} w_{ij} y_i h_j - \sum_i b_i y_i - \sum_j c_j h_j \]

  \[ p(y, h) = \frac{1}{Z} \exp (-E(y, h)) \]

  - Sum out \( h \), RBM becomes a high order potential on \( y \)

- **Some success modeling object shape**
  - The Shape Boltzmann Machine [S. M. Ali Eslami et al., CVPR'12]
  - Masked RBMs [N. Heess et al. ICANN'11]
**CHOPP**

- Compositional High Order Pattern Potential (CHOPP)

\[
f_T(y) = -T \log \left( \sum_h \exp \left( \frac{1}{T} \sum_i \left( c_j + \sum_{i} w_{ij} y_i \right) h_j \right) \right)
\]

Interpolate between RBMs and PPs

Combine all patterns

Compatibility with a pattern

---

**Sparsity on \( h \)**

- **1-of-J**
  - Pattern Potential "min" composition
  - Extremely Sparse RBMs

- **Cardinality potential?**
  - Pattern Potential "sum" composition
  - Standard RBMs

- **No sparsity**

---

\( T = 0 \)
- Min out \( h \)
- Multiple modes?

\( T = 1 \)
- Sum out \( h \)
**CHOPP-Augmented CRF**

- **Compositional High Order Pattern Potential (CHOPP)**

  \[ f_T(y) = -T \log \left( \sum_h \exp \left( \frac{1}{T} \sum_j \left( c_j + \sum_i w_{ij}y_i \right) h_j \right) \right) \]

- **CHOPP-augmented CRF Energy function**

  \[-E(y|x) = f^u(y|x) + f^p(y|x) + T \log \left( \sum_h \exp \left( \frac{1}{T} \sum_j \left( c_j + \sum_i w_{ij}y_i \right) h_j \right) \right) \]
“EM” Inference Algorithm

• Making predictions

\[ y^* = \arg\max_y \{-E(y|\mathbf{x})\} \]

**E-step:** fix \( y \) compute \( h \)

**M-step:** fix \( h \) find optimal \( y \)

Hidden variables \( h \)

Posterior inference

Labels \( y \)

The impact of \( h \) factorizes

Just a pairwise CRF
Use Graph Cuts

The image shows a diagram with nodes representing hidden variables and labels. The diagram illustrates the process of making predictions using the EM algorithm. The E-step involves fixing the labels and computing the hidden variables, while the M-step involves fixing the hidden variables and finding the optimal labels. The impact of the hidden variables is shown to factorize, and the diagram also mentions the use of graph cuts for just a pairwise CRF.
An Example for the “EM” Inference Algorithm

Initialize EM with this
An Example for the “EM” Inference Algorithm

Original Image  
Unary Prediction  
Ground Truth  
Unary+ Pairwise  
Initialization EM with this  

Iteration #1

Graph Cuts

Compute h
An Example for the “EM” Inference Algorithm

Original Image

Unary Prediction

Ground Truth

Unary+

Pairwise

Initialize EM with this

$y$ computed by Graph Cuts

\[
p(y | \mathbb{E}_q[h])
\]

Iteration #1

#2

#3

Convergence
Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective
  \[ L = \sum_{y} p(y|x; \theta) \ell(y, y^*) \]
- Estimate gradient using a set of samples from \( p(y|x) \)
Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective

\[ L = \sum_y p(y|x; \theta) \ell(y, y^*) \]

- Estimate gradient using a set of samples from \( p(y|x) \)
Learning by Minimizing Expected Loss

- Contrastive Divergence does not work well
- Expected loss objective
  \[ L = \sum_y p(y|x; \theta) \ell(y, y^*) \]
- Estimate gradient using a set of samples from \( p(y|x) \)
Datasets and Settings

- Weizmann horse dataset
- PASCAL VOC 2011: image inside the bounding box
  - Class “person” and class “bird”
- All images resized to 32x32
- $T=1$, Intersection Over Union (IOU) performance measure
Experiment I

- Train RBM independently (unsupervised)

<table>
<thead>
<tr>
<th>Method</th>
<th>Horse IOU</th>
<th>Bird IOU</th>
<th>Person IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unary Only</td>
<td>0.5119</td>
<td>0.5055</td>
<td>0.4979</td>
</tr>
<tr>
<td>iPW</td>
<td>0.5736</td>
<td>0.5585</td>
<td>0.5094</td>
</tr>
<tr>
<td>iPW+RBM</td>
<td>0.6722</td>
<td>0.5647</td>
<td>0.5126</td>
</tr>
</tbody>
</table>

- Adding an RBM always helps
  - But not equally on different datasets
Experiment I Analysis: Dataset Variability

- **Dataset variability measure**
  
  Clustering → Intra-cluster entropy → Weighted average

- Person & Birds are harder than horses

### Graphs

- **Real Datasets**
  - Delta IOU from Unary Only
  - Variability (K=32)
  - Horses, Birds, Person

- **Synthetic Datasets**
  - Delta IOU from Unary Only
  - Variability (K=32)
Experiment II and III

- Jointly learning RBM parameters by minimizing expected loss

<table>
<thead>
<tr>
<th>Method</th>
<th>Horse IOU</th>
<th>Bird IOU</th>
<th>Person IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPW+RBM</td>
<td>0.6722</td>
<td>0.5647</td>
<td>0.5126</td>
</tr>
<tr>
<td>iPW+jRBM</td>
<td><strong>0.6990</strong></td>
<td><strong>0.5773</strong></td>
<td><strong>0.5253</strong></td>
</tr>
</tbody>
</table>
Experiment II and III

- Jointly learning RBM parameters by minimizing expected loss

<table>
<thead>
<tr>
<th>Method</th>
<th>Horse IOU</th>
<th>Bird IOU</th>
<th>Person IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>iPW+RBM</td>
<td>0.6722</td>
<td>0.5647</td>
<td>0.5126</td>
</tr>
<tr>
<td>iPW+jRBM</td>
<td><strong>0.6990</strong></td>
<td><strong>0.5773</strong></td>
<td><strong>0.5253</strong></td>
</tr>
</tbody>
</table>

- Making the RBM hidden bias conditioned on the image

<table>
<thead>
<tr>
<th>Method</th>
<th>Bird IOU</th>
<th>Person IOU</th>
</tr>
</thead>
<tbody>
<tr>
<td>PW</td>
<td>0.5321</td>
<td>0.5082</td>
</tr>
<tr>
<td>iPW</td>
<td>0.5585</td>
<td>0.5094</td>
</tr>
<tr>
<td>iPW+jRBM</td>
<td>0.5773</td>
<td><strong>0.5253</strong></td>
</tr>
<tr>
<td>iPW+ijRBM</td>
<td><strong>0.5858</strong></td>
<td>0.5252</td>
</tr>
</tbody>
</table>
Examples

Most Improvement

Average Improvement

Least Improvement
Conclusion and Future Work

• Theoretical contribution
  – Relationship between RBMs and Pattern Potentials

• Algorithmic contribution
  – Inference and learning algorithms for CHOPP-augmented CRFs

• Empirical contribution
  – Dataset variability measure

• Looking forward:
  – Convolutional and deeper models
  – Fully explore the variants of CHOPP
  – Challenge: lack of labeled data
Learned Patterns

(a) Horse filters

(b) Bird filters

(c) Person filters.