Large-Scale Object Recognition using Label Relation Graphs

Jia Deng¹,² Nan Ding², Yangqing Jia², Andrea Frome², Kevin Murphy², Samy Bengio², Yuan Li², Hartmut Neven², Hartwig Adam²

University of Michigan¹, Google²
Object Classification

- Assign semantic labels to objects

Corgi

Dog ✓
Corgi ✓
Puppy ✓
Cat ✗
Object Classification

• Assign semantic labels to objects
Object Classification

- Assign semantic labels to objects
Object Classification

- **Independent binary classifiers: Logistic Regression**
  
  - Corgi: 0.8
  - Puppy: 0.6
  - Cat: 0.4

  No assumptions about relations.

- **Multiclass classifier: Softmax**
  
  - Corgi: 0.4
  - Puppy: 0.3
  - Cat: 0.1

  Assumes mutual exclusive labels.
Object labels have rich relations

Softmax: all labels are mutually exclusive 😞
Logistic Regression: all labels overlap 😞
Goal: A new classification model

Respects real world label relations

![Diagram of a neural network classifying animals with probabilities]

- Dog: 0.9
- Corgi: 0.8
- Puppy: 0.9
- Cat: 0.1
Assumption in this work: Knowledge graph is given and fixed.
Agenda

• Encoding prior knowledge (HEX graph)
• Classification model
• Efficient Exact Inference
• Experiments
• Conclusion and Future Work
Agenda

- Encoding prior knowledge (HEX graph)
- Classification model
- Efficient Exact Inference
- Experiments
- Conclusion and Future Work
Hierarchy and Exclusion (HEX) Graph

- Hierarchical edges (directed)
- Exclusion edges (undirected)
Examples of HEX graphs

- Mutually exclusive
- All overlapping
- Combination
State Space: Legal label configurations

Each edge defines a constraint.

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
State Space: Legal label configurations

Each edge defines a constraint.

Hierarchy: (dog, corgi) can’t be (0,1)
State Space: Legal label configurations

Each edge defines a constraint.

Hierarchy: (dog, corgi) can’t be (0,1)

Exclusion: (dog, cat) can’t be (1,1)
Agenda

• Encoding prior knowledge (HEX graph)
• Classification model
• Efficient Exact Inference
• Experiments
• Conclusion and Future Work
HEX Classification Model

• Pairwise Conditional Random Field (CRF)

\[ \Pr(y \mid x) = \frac{1}{Z(x)} \prod_{i} \phi_{i}(x_{i}, y_{i}) \prod_{i,j} \psi_{i,j}(y_{i}, y_{j}) \]

where

- \( x \in \mathbb{R}^{n} \) is the input scores.
- \( y \in \{0, 1\}^{n} \) is the binary label vector.
- \( Z(x) \) is the partition function.
- \( \phi_{i}(x_{i}, y_{i}) \) are feature functions for each node.
- \( \psi_{i,j}(y_{i}, y_{j}) \) are pairwise interaction terms.

The diagram illustrates the structure of the CRF model with nodes representing input scores and edges showing interactions between labels.
HEX Classification Model

• Pairwise Conditional Random Field (CRF)

\[ x \in \mathbb{R}^n \]

Input scores

\[ y \in \{0, 1\}^n \]

Binary Label vector

\[
\Pr(y| x) = \frac{1}{Z(x)} \sum_{i} \phi_i(x_i, y_i) \sum_{i,j} \psi_{i,j}(y_i, y_j)
\]

Unary: same as logistic regression

\[
\phi_i(x_i, y_i) = \begin{cases} 
\text{sigmoid}(x_i) & \text{if } y_i = 1 \\
1 - \text{sigmoid}(x_i) & \text{if } y_i = 0
\end{cases}
\]
HEX Classification Model

• Pairwise Conditional Random Field (CRF)

\[ x \in \mathbb{R}^n \]
Input scores

\[ y \in \{0, 1\}^n \]
Binary Label vector

\[
\Pr(y|x) = \frac{1}{Z(x)} \prod_{i} \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)
\]

Unary: same as logistic regression

\[
\phi_i(x_i, y_i) = \begin{cases} 
\text{sigmoid}(x_i) & \text{if } y_i = 1 \\
1 - \text{sigmoid}(x_i) & \text{if } y_i = 0 
\end{cases}
\]

Pairwise: set illegal configuration to zero

\[
\psi_{i,j}(y_i, y_j) = \begin{cases} 
0 & \text{If violates constraints} \\
1 & \text{Otherwise}
\end{cases}
\]
HEX Classification Model

- Pairwise Conditional Random Field (CRF)

\[ x \in \mathbb{R}^n \]

Input scores

\[ y \in \{0, 1\}^n \]

Binary Label vector

\[
\Pr(y \mid x) = \frac{1}{Z(x)} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)
\]

\[
Z(x) = \prod_{y \in \{0, 1\}^n} \prod_i \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)
\]

Partition function: Sum over all (legal) configurations
HEX Classification Model

• Pairwise Conditional Random Field (CRF)

\[ x \in \mathbb{R}^n \]
Input scores

\[ y \in \{0, 1\}^n \]
Binary Label vector

\[
\Pr(y \mid x) = \frac{1}{Z(x)} \prod_{i} \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(y_i, y_j)
\]

Probability of a single label: marginalize all other labels.

\[
\Pr(y_i = 1 \mid x) = \frac{1}{Z(x)} \prod_{y_i = 1} \phi_i(x_i, y_i) \prod_{i,j} \psi_{i,j}(\overline{y}_i, \overline{y}_j)
\]
### Special Case of HEX Model

#### Softmax

- **Car**
- **Bird**
- **Dog**
- **Cat**

- Mutually exclusive

#### Logistic Regressions

- **Red**
- **Shiny**
- **Round**
- **Thick**

- All overlapping

\[
\Pr(y_i = 1 \mid x) = \frac{\exp(x_i)}{1 + \sum_j \exp(x_j)}
\]

\[
\Pr(y_i = 1 \mid x) = \frac{1}{1 + \exp(-x_i)}
\]
Learning

Maximize marginal probability of observed labels

Label: Dog

Loss = -\log \Pr(Dog = 1)
Agenda

• Encoding prior knowledge (HEX graph)
• Classification model
• Efficient Exact Inference
• Experiments
• Conclusion
Naïve Exact Inference is Intractable

• Inference:
  – Computing partition function
  – Perform marginalization

• HEX-CRF can be densely connected (large treewidth)
Observation 1: Exclusions are good

- Lots of exclusions $\rightarrow$ Small state space $\rightarrow$ Efficient inference
- Realistic graphs have lots of exclusions.
- Rigorous analysis in paper.

Number of legal states is $O(n)$, not $O(2^n)$. 
Observation 2: Equivalent graphs
Observation 2: Equivalent graphs

Sparse equivalent
• Small Treewidth 😊
• Dynamic programming

Dense equivalent
• Prune states 😊
• Can brute force
HEX Graph Inference

1. Sparsify (offline)

2. Build Junction Tree (offline)

3. Densify (offline)

4. Prune Clique States (offline)

5. Message Passing on legal states (online)
Agenda

• Encoding prior knowledge (HEX graph)
• Classification model
• Efficient Exact Inference
• Experiments
• Conclusion and Future Work
Exp 1: Learning with weak labels

- Many basic category labels
- Few fine-grained labels

Weak labels: No information on subcategories.
Exp 1: Learning with weak labels

Hypothesis: HEX models can improve fine-grained recognition using basic level labels.

Loss = \(-\log \Pr(Dog = 1)\)

Pr(Dog = 1)
Exp 1: Learning with weak labels

- ILSVRC 2012: “relabel” or “weaken” a portion of fine-grained leaf labels to basic level labels.
- Evaluate on fine-grained recognition
Exp 1: Learning with weak labels

• ILSVRC 2012: “relabel” or “weaken” a portion of fine-grained leaf labels to basic level labels.
• Evaluate on fine-grained recognition.
• Consistently outperforms baselines.

<table>
<thead>
<tr>
<th>relabeling</th>
<th>softmax-leaf</th>
<th>softmax-all</th>
<th>logistic</th>
<th>ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>50.5(74.7)</td>
<td>56.4(79.6)</td>
<td>21.0(45.2)</td>
<td>58.2(80.8)</td>
</tr>
<tr>
<td>90%</td>
<td>26.2(47.3)</td>
<td>52.9(77.2)</td>
<td>9.3(27.2)</td>
<td>55.3(79.4)</td>
</tr>
<tr>
<td>95%</td>
<td>16.0(32.2)</td>
<td>50.8(76.0)</td>
<td>5.6(17.2)</td>
<td>52.4(77.2)</td>
</tr>
<tr>
<td>99%</td>
<td>2.5 (7.2)</td>
<td><strong>41.5</strong> (68.1)</td>
<td>1.0(3.8)</td>
<td><strong>41.5</strong> (68.5)</td>
</tr>
</tbody>
</table>

Top 1 accuracy (top 5 accuracy)
Exp 2: Zero-Shot Recognition using Object-Attribute Knowledge

- Animals with Attribute (AwA) dataset (Lampert et al. 2009)
- Training:
  - Observe only a subset of animal labels.
  - Given all animal-attribute relations
  - Indirectly learns attributes.
- Test: predict new classes with no images in training.

<table>
<thead>
<tr>
<th></th>
<th>DAP (Lampert et al.)</th>
<th>IAP (Lampert et al.)</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40.5%</td>
<td>27.8%</td>
<td>38.5%</td>
</tr>
</tbody>
</table>
Related Work

• Multilabel Annotation & Hierarchy
  [Lampert et al. NIPS’11] [Hwang et al. CVPR’11]
  [Chen et al. ICCV’11] [Kang et al. CVPR’06]
  [Bi & Kwok, NIPS’12] [Marszalek & Schmid CVPR’07]
  [Bucak et al. CVPR’11] [Zweig & Weinshall CVPR’07]

  Ours: Unifies hierarchy and exclusion.

• Transfer learning & Attributes
  [Rohrbach et al. CVPR’10] [Farhadi et al. CVPR’10]
  [Lampert et al. CVPR’09] [Lim et al. NIPS’11]
  [Kuettel et al. ECCV’12] [Yu et al. CVPR’13]
  [Akata et al. CVPR’13] [Fergus et al. ECCV’10]

  Ours: A classification model that allows transferring.

• Extracting Common Sense Knowledge
  [Chen et al. ICCV’13] [Zhu et al. ECCV’14]
  [Zitnick & Parikh CVPR’13] [Fouhey & Zitnick CVPR’14]

  Ours: Assumes knowledge is given.
Conclusions

• A unified framework for single object classification
  – Generalizes standard classification models
  – Leverages a knowledge graph
  – Efficient exact inference

• Future work
  – Non-absolute relations
  – Spatial relations between object instances