Relative Attributes

Experiments

Sanmit Narvekar
Department of Computer Science
The University of Texas at Austin

October 19, 2012
1. How does the type of “pairs” supervision given affect how well an attribute is learned?
2. Do we need a continuous relative ranking, or would discrete work better?
3. How do we know whether the attributes are learning the features they correspond to?
Analyzing Type of Supervision

- **Category-level training pairs**
  - Easy to obtain more pairs, which may not all be “correct”
- **Instance-level training pairs**
  - Harder to obtain, but more “correct”

Categories are different people or scene types
Analyzing Type of Supervision

- Compare **which attributes perform better for which type of supervision**
- Masculinity and smiling
- Naturalness and openness
- Evaluated on 10 random pairs of images

### Accuracies

<table>
<thead>
<tr>
<th></th>
<th>Categorical</th>
<th>Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masculinity</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>Smiling</td>
<td>0.70</td>
<td>0.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Categorical</th>
<th>Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalness</td>
<td>0.90</td>
<td>0.80</td>
</tr>
<tr>
<td>Openness</td>
<td>0.80</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Faces Dataset

Scenes Dataset
Miley usually smiles more than Alex, so the categorically trained classifier got confused.

Attributes that vary within classes are trained better on instances.
Occlusion interferes with the inference.
But, we know Miley usually smiles more than Alex. Does this count?
Masculinity and Smiling

Masculinity is technically a categorical attribute. However, even categorical attributes can vary intra-class in unexpected ways.
And some things inevitably come down to taste
Need for Relative Attributes

• Do we really need continuous relative attributes?

OR

• Do some attributes form distinct groups?
  – male vs. female
  – natural vs. artificial
  – Could be more than 2 groups...
  – Then use a discrete ranking system?

Analyze the histogram of rankings across attributes and their mean shift cluster centers
Most rankings have a Gaussian-like distribution, suggesting attributes are more amenable to representation by relative rankings rather than binary or discrete rankings.
In distributions where a lot of the mass is in the middle, binary attribute labels (representing the extrema) could be inappropriate.
Attribute Localization

- How do you know whether the attributes learned correspond to their semantic meanings?
  - Especially when no labels, bounding boxes, etc. given

Object recognition  
Learning airplane or sky?

Attribute-based recognition  
Learning high heels or no laces?

Seems more problematic in attribute-based recognition, since each attribute has semantic meaning, and is a part of a whole that can be hard to identify.
Attribute Localization

- **Task**: Determine whether the ranker is learning the attribute "high heels" in a dataset of shoes

- **Approach**:

  Compare results of rankers trained on these different types
Attribute Localization

• Evaluate on 10 random pairs of images
• Images are automatically flipped if facing the wrong way
• Compare how well each method ranks high heels given
  – Image descriptor of the whole image
  – Image descriptor of only the heel area
  – Image descriptor of everything except the heel area

<table>
<thead>
<tr>
<th>Accuracies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole Image</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>

Suggests some contextual information was used for classification
Find the Highest Heel

<table>
<thead>
<tr>
<th></th>
<th>Whole</th>
<th>Relevant</th>
<th>Irrelevant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6742</td>
<td>-0.0342</td>
<td>-0.1146</td>
</tr>
<tr>
<td></td>
<td>-0.6160</td>
<td>-0.1440</td>
<td>-0.0074</td>
</tr>
</tbody>
</table>

The “whole” and “relevant” descriptors both saw the missing heel in the right-side shoe. The straps might have mislead the “irrelevant area” classifier?
The ranker fed the whole image descriptor could probably reason about heel height from the sole, since the heel itself was occluded. Attribute captured, not captured, or assisted?
Summary

We looked at:

• Types of supervision, and its effects on attributes intrinsic to a class (masculinity) and where they can vary (smiling)
  – Category-level supervision
  – Instance-level supervision

• Need for continuous relative attributes, or whether attributes form “discrete” groups
  – How that affects different classes

• Attribute localization
  – Are we learning what we think we are?
References


• Links to existing code and data used:
  – Rank SVM: http://ttic.uchicago.edu/~dparikh/relative.html#code
  – Categorical and Instance Pair labels, extracted feature representations: http://www.cs.utexas.edu/~grauman/research/datasets.html

• Links to primary datasets used:
  – PubFig: http://www.cs.columbia.edu/CAVE/databases/pubfig/
Questions?