Relative Attributes

Team mob-psycho

- Kanav Gupta
- Anurag Mehta
- Parth Partani

Introduction

Introduction

When dealing with recognition tasks, human-nameable visual "attributes" can be beneficial, but:

- Sometimes it is hard to make a binary decision on whether an image satisfies an attribute.

- Comparisons are easier.

Introduction

Relative Attributes

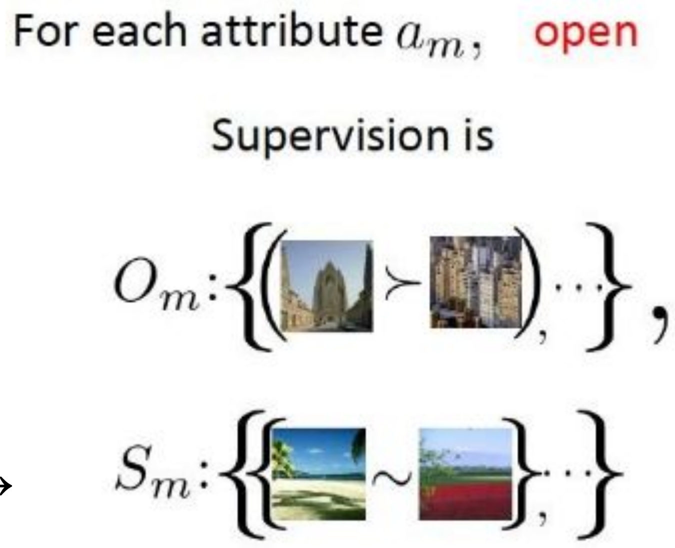
- Great to describe and compare objects in the world.
- Indicate the strength of an attribute in an image with respect to other images.
- Allow relating images and categories to each other.

A is more natural than B
C is less natural than B

Learning Relative Attributes

Learning Relative Attributes

Model Relative Attributes: Learn ranking function for each attribute



Set of ordered pair of images →

Set of unordered pair of images →

Learning Relative Attributes

Learn a scoring function $r_m(\boldsymbol{x_i}) = \boldsymbol{w_m^T x_i}$

Image features

Learned parameters

that best satisfies constraints:

$$\forall (i,j) \in O_m : \boldsymbol{w_m^T x_i} > \boldsymbol{w_m^T x_j}$$

$$\forall (i,j) \in S_m : \boldsymbol{w_m^T x_i} = \boldsymbol{w_m^T x_j}$$

Learning Relative Attributes

Max-margin learning to rank formulation

$$\min \quad \left(\frac{1}{2}||\boldsymbol{w}_{\boldsymbol{m}}^T||_2^2 + C\left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2\right)\right)$$

$$\text{s.t} \quad \boldsymbol{w}_{\boldsymbol{m}}^T(\boldsymbol{x_i} - \boldsymbol{x_j}) \geq 1 - \xi_{ij}, \forall (i,j) \in O_m$$

$$|\boldsymbol{w}_{\boldsymbol{m}}^T(\boldsymbol{x_i} - \boldsymbol{x_j})| \leq \gamma_{ij}, \forall (i,j) \in S_m$$

$$\xi_{ij} \geq 0; \gamma_{ij} \geq 0$$

Based on [Joachims 2002]

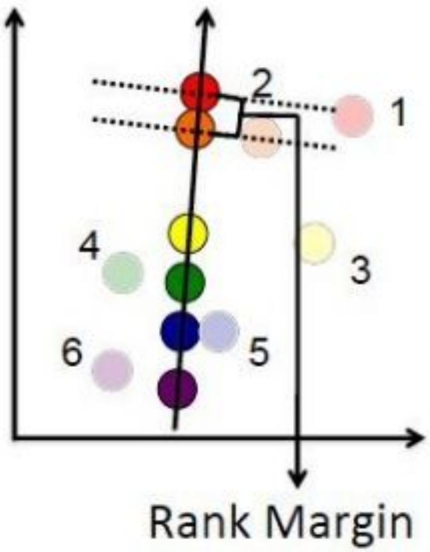
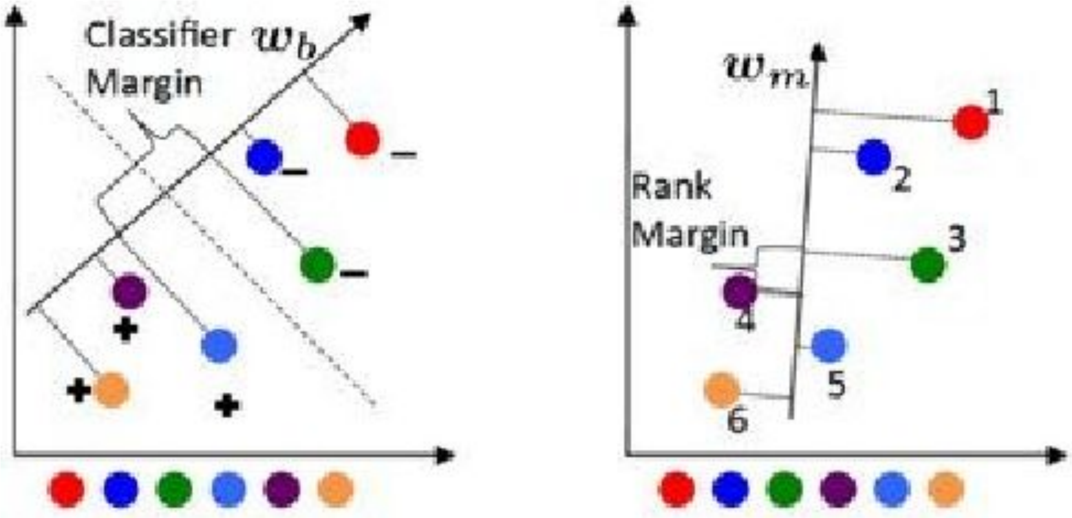


Image → Relative Attribute Score

Learning Relative Attributes

Wide-margin **binary classifier** VS Wide-margin **ranking function**



Applications

Applications

- **Zero Shot Learning:** Train a generative model (like GMM) to predict values for new class based on its relationship with current classes and no training images
- **Describing Images:** Automatically generating relative description of the images (like a particular image is more smiling and young than other image)

1. Zero Shot Learning

Divide Classes into 2 Categories:

1. **Seen (S):**
    - Image set is available
    - Relative description of attributes is present

2. **Unseen (U):**
    - No images available
    - Some attributes described wrt seen categories

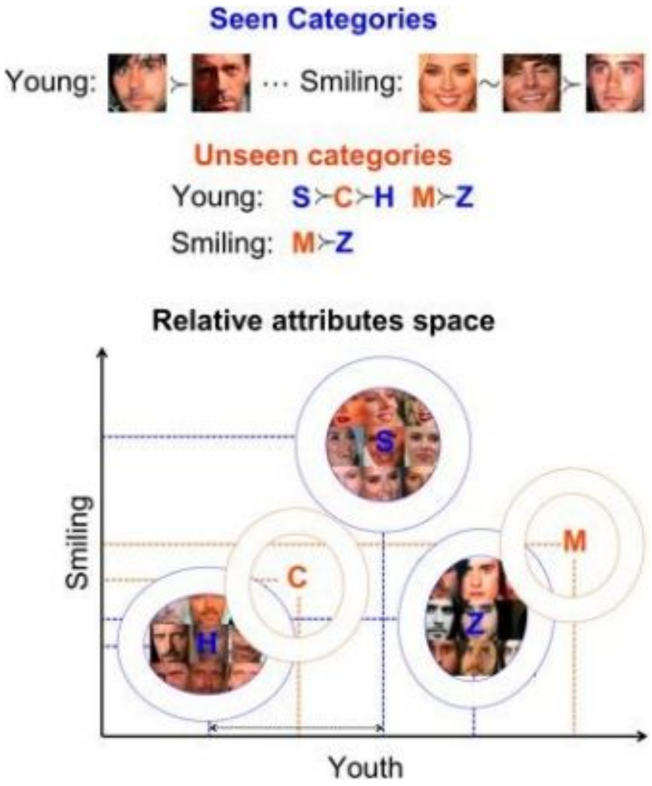
1. Zero Shot Learning

Training

1. Train a set of relative attributes using S categories.
   - Learn all M relative attribute am.
   - Represent each image as m-vector

2. Build GMM for each S category using the responses of the relative attributes.

3. Infer the parameters of the generative models of U categories by using their relative descriptions.



1. Zero Shot Learning

Test (classify new images)

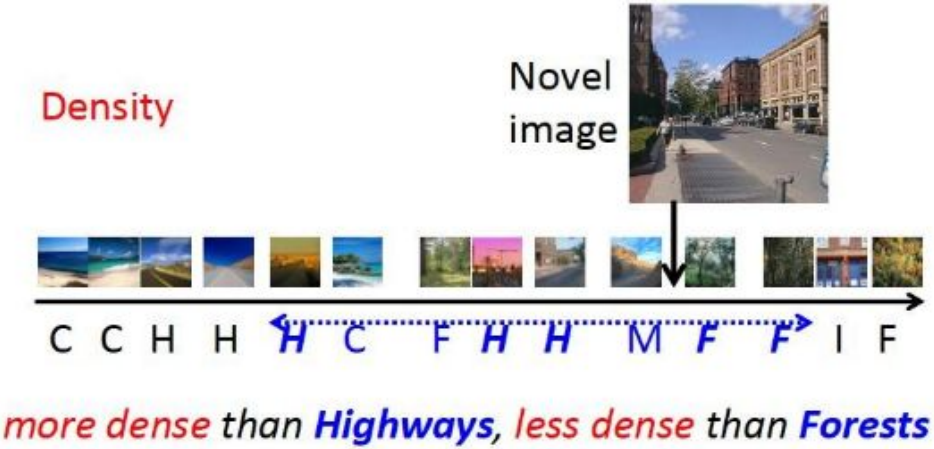
1. Compute xi indicating its relative attribute ranking-scores for the image.
2. Assign it to the seen or unseen category that assigns it the highest likelihood:

$$c^* = \underset{j \in \{1,\ldots,N\}}{\operatorname{argmax}} \; P(\tilde{x}_i \mid \mu_j, \Sigma_j).$$

2. Describing images

Given an image I to be described:

1. We evaluate all learnt ranking functions on I.
2. Identify 2 reference images from which image I will be described.
3. Can also describe an image relative to other categories.



Results Obtained

Results Obtained - Smiling Attribute

```
Attrbute values for image 1 and 2 :
```

Value =[[0.12842465]]

Value =[[0.08821574]]

Results Obtained - Big Lips Attribute

Value =[[-0.18477103]]

Value =[[-1.07309342]]

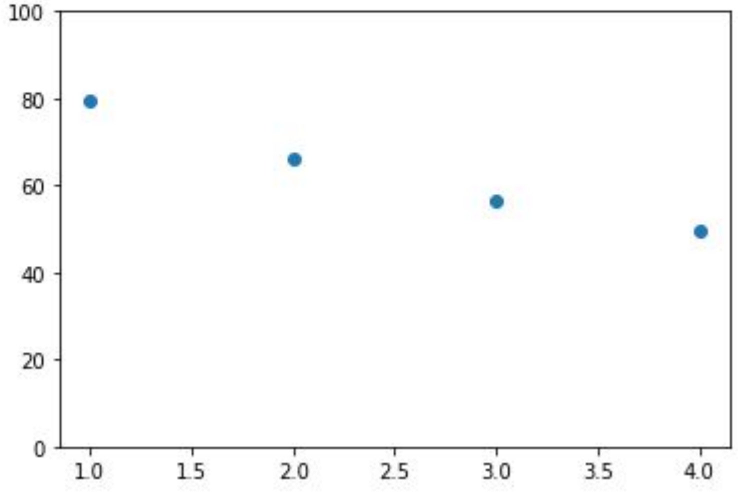
Results Obtained - Chubby Attribute

Value =[[-0.20013611]]

Value =[[-1.43647949]]

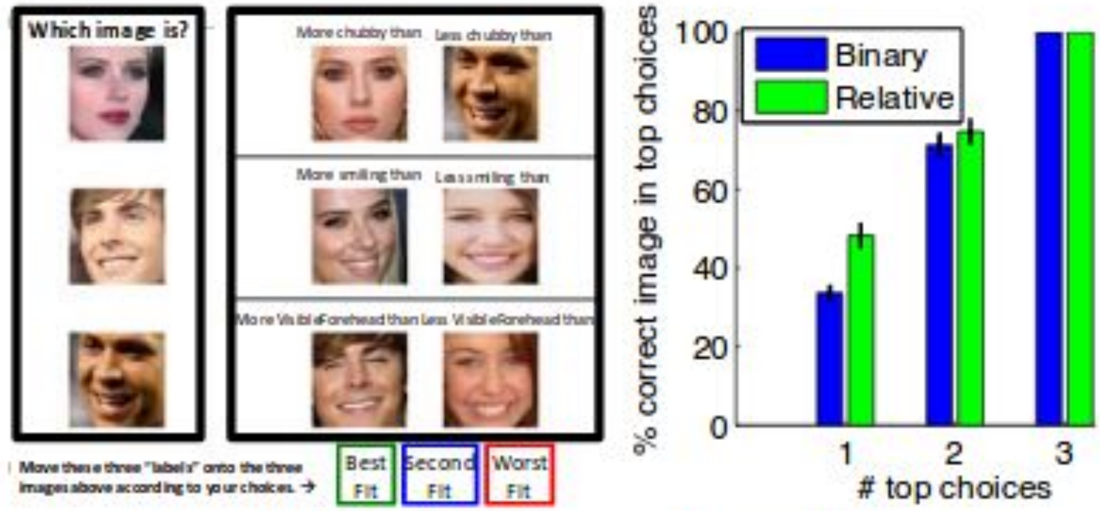
Results Obtained - Zero Shot Learning

Number of Unseen Classes	Accuracy
1	79.583 %
2	66.204 %
3	56.592 %
4	49.481 %



Results Obtained - Describing Images

Analysis conducted by Authors



(a) Human Study Interface (b) Results (both datasets)

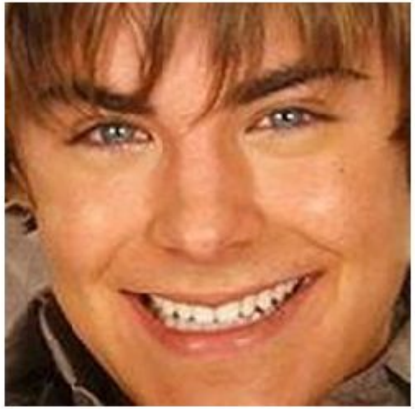
Results Obtained - Describing Images

Attribute = Smiling

Value =-0.298639147492

Value =[[0.12842465]]

Value =0.601809617635



Results Obtained - Describing Images

Attribute = Big Lips

Value =-0.755304992066

Value =[[-0.25540029]]

Value =0.664094646003

`Image is more BigLips than ScarlettJohansson_100.jpg and less BigLips than CliveOwen_89.jpg`

Results Obtained - Describing Images

Attribute = Male

Value =-0.482523724035

Value =[[-0.02635795]]

Value =0.541561750807

`Image is more Male than CliveOwen_119.jpg and less Male than ViggoMortensen_146.jpg`

Thank You