Practical on Zero-shot Learning

Yongqin Xian

Max-Planck Institute for Informatics

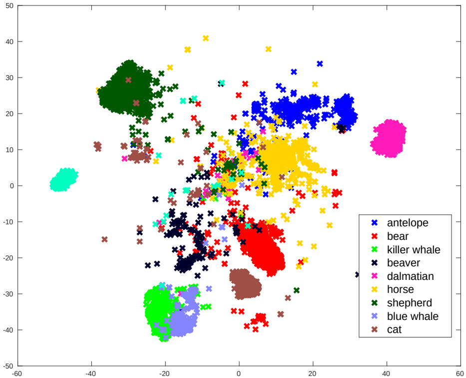
12.09.2016, GCPR

Road Map

- Visualization of Embeddings
- Zero-shot Learning
- Structural Joint Embedding
- Demos
- Latent Embedding
- Demos

Visualization of Embeddings

Figure: t-SNE visualization of GoogleNet image feature on AWA dataset



ZSL Problem

Training phase:

1. Observed data set

$$\mathcal{T} = \{(x, y) | x \in \mathcal{X}, y \in \mathcal{Y}\}$$

   x: image embedding, y: class embedding

2. Using $\mathcal{T}$, learn function

$$f : \mathcal{X} \rightarrow \mathcal{Y}$$

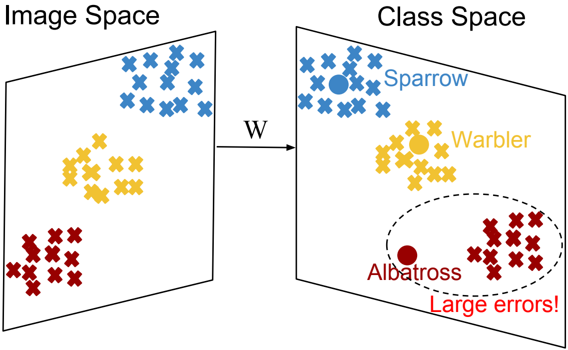
Testing phase:

1. Predict x^* of unseen classes using $f(x^*)$

Structural Joint Embedding

SJE [Akata et.al.'15]: single bilinear compatibility

- Not discriminative enough



$$F(x, y) = x^\top W y$$

Structural Joint Embedding

Prediction function

$$f(x; W) = \arg\max_{y \in \mathcal{Y}} x^T W y,$$

Objective function

$$\min_{W} \frac{1}{N} \sum_{n=1}^{N} \max_{y \in \mathcal{Y} \setminus \{y_n\}} \{0, \Delta(y_n, y) + x_n^T W y - x_n^T W y_n\},$$

Structural Joint Embedding

Algorithm 1 SGD optimization of SJE

Given $\mathcal{T} = \{(x, y) | x \in \mathbb{R}^{d_x}, y \in \mathbb{R}^{d_y}\}$
Initialize W randomly
for $t = 1$ to T **do**
  for $n = 1$ to $|\mathcal{T}|$ **do**
    Draw $(x_n, y_n) \in \mathcal{T}$
    $y^* \leftarrow \underset{y \in \mathcal{Y} \setminus \{y_n\}}{\arg\max}\, x_n^\top W y$
    if $x_n^\top W y^* + 1 > x_n^\top W y_n$ **then**
      $W \leftarrow W - \eta x_n (y^* - y_n)^\top$
    end if
  end for
end for

Demo

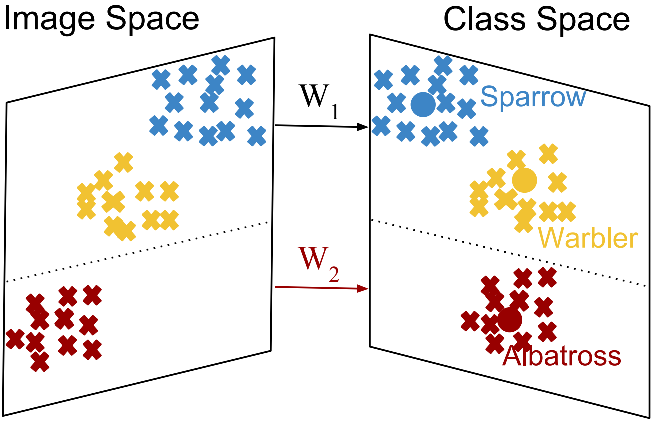
Image Retrieval

$$\arg\max_{i} x_i^\top W y$$

Chimpanzee

Seal

Latent Embeddings Method (LatEm)



$$F(x,y) = \max_{1\le i\le K} x^\top W_i y$$

- Learn a collection of matrices
- Selection of which one to use is latent

Latent Embeddings Method (LatEm)

Loss function

$$L(x_n, y_n) = \sum_{y \in \mathcal{Y}} \left[\Delta(y_n, y) + F(x_n, y) - F(x_n, y_n)\right]_+ ,$$

Objective function

$$\min_{\{W_i\}_{i=1^K}} \frac{1}{N} \sum_{n=1}^{N} \max_{y \in \mathcal{Y} \setminus \{y_n\}} \{0, \Delta(y_n, y) + x_n^T W y - x_n^T W y_n\},$$

Demo