Learning Discriminative Latent Attributes for Zero-Shot Classification

Huajie Jiang[1,2,3], Ruiping Wang[1], Shiguang Shan[1], Yi Yang[4], Xilin Chen[1]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2]Shanghai Institute of Microsystem and Information Technology, CAS, Shanghai, 200050, China
[3]ShanghaiTech University, Shanghai, 200031, China
[4]Huawei Technologies Co., Ltd., Beijing, 100085, China
huajie.jiang@vipl.ict.ac.cn, {wangruiping, sgshan, xlchen}@ict.ac.cn, yangyi16@huawei.com

This document gives detailed analysis for parameter settings and the sparse property of attributes combination coefficients.

1. Parameter Settings

In this part, we analyze some parameters used in our framework. The parameters α and β in the objective function Eq.(4) control the relative importance of semantic and discriminative constraints, which can be tuned by the cross-validation approach. γ and λ are regularization parameters used in Eq.(13) and Eq.(14). They can also be tuned. For simplicity, we set them to be 1. Dictionary size decides the dimension of latent attributes. The values of these five parameters used in our paper are shown in Table 1. Due to the fact that latent attribute dictionary items are randomly initialized, the recognition rates may have small difference by different initializations. Detailed implementations can be found in our source code.

Table 1. Parameter settings on the four datasets. 'DS' denotes the dictionary size.

Dataset	α	β	γ	λ	DS
aP&Y	10	0.1	1	1	500
AwA	20	10	1	1	300
CUB-200	20	50	1	1	500
SUN-A	20	30	1	1	800

2. Analysis of Sparsity

Since our latent attributes can be viewed as different combinations of attributes, it is desirable that the combination coefficients should be sparse. That is to say, a latent attribute should have strong activations on a small number of attributes. Although we do not explicitly impose sparsity restrictions on the combination coefficients matrix W in the objective function Eq.(4), the combination coefficients are found to be mostly sparse in our experimental study. To demonstrate this, here we take **AwA** for example, where the dictionary size is set to be 300 and the attributes number is 85. In order to visualize the combination coefficients W, which is a 300*85 matrix, we get the absolute value for each element of W and observe the activations. Figure 1 displays the activation values of W, of which each row represents the activations of each latent attribute (300 latent attributes in total) on all the attributes (85 attributes in total). It can be figured out that most values are near zero. This indicates that the learned combination coefficients are implicitly ensured to be sparse.

To get an intuitive impression on the activation values, we make statistics for each row (*i.e.* each latent attribute) of W and put the values into 10 bins. Figure 2 shows the statistical histogram results, where each color represents a latent attribute. We can see that most of the combination coefficients are less than 0.1, which indicates that only small numbers of attributes have strong activations on the latent attributes. While the latent attributes are automatically learned in our framework, the discriminative constraint in our objective function enables to learn the most discriminative combinations, which inherently ensures the sparsity of the attribute combinations to some extent.

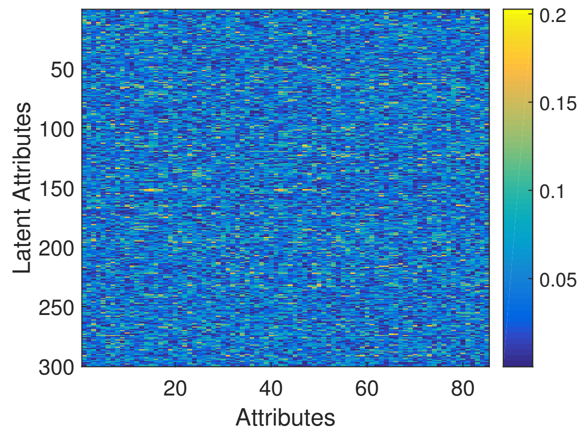

Figure 1. Visualization of the combination coefficients W on **AwA**.

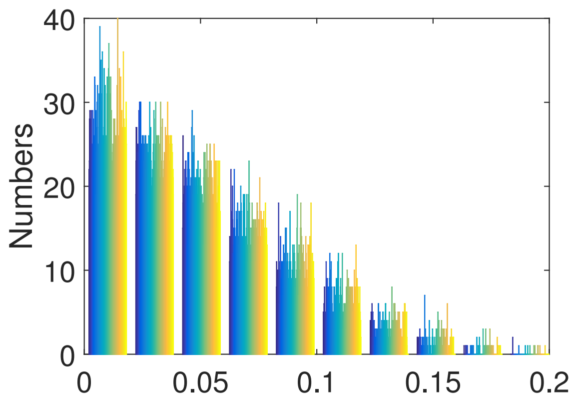

Figure 2. Attribute combination coefficients statistic on **AwA**. Different colors represent different latent attributes.