COS 529 Assignment 2: Zero-Shot Recognition

Due 11:59pm 03/11/2018 ET

Collaboration policy This assignment must be completed individually. No collaboration is allowed.

1 Zero-Shot Recognition with Attributes

In this assignment, you will be building a model to perform *zero-shot recognition.* In zero-shot recognition, the training and test classes are disjoint. In other words, your model will be tested on classes which it hasn't seen before. For example, your training set may contain images of $Y_{train} = \{\text{seahorse}, \text{frog}, \text{land snail}\}$ while $Y_{test} = \{\text{sea pig}, \text{armadillo}, \text{mantis shrimp}\}$. Here, $Y_{train} \cap Y_{test} = \emptyset$

Now, without any additional information, assigning labels to unseen classes in not possible. One way to overcome this problem is indirect attribute prediction. As shown in the figure below, each class can be assigned a set of attributes (i.e. color, shape, natural habit).

Instead of training a classifier to predict the type of animal directly, we can instead predict attributes. Assuming we know the attributes of the test classes, we can use our predicted attributes to find the best match. In this setting, we can use attributes as an intermediate between images and labels. Since attributes are shared between images in the training and testing set, our attribute prediction model can still be used at test time—using attributes as an intermediate.

2 Dataset

For this assignment, you will be working with the Animals with Attributes dataset `https://cvml.ist.ac.at/AwA2/`. The dataset contains 37322 images of 50 different animal classes, each with 85 labeled attributes. The dataset is divided into a training set and a test set with disjoint animal classes. You may assume that you have access to the test class attributes at test time.

In the attached folder, we have provided the following documents

- trainclasses.txt: A list of the training classes
- testclasses.txt: A list of the test classes
- classes.txt: A list of all the classes in order of their assignment
- predicate-matrix-binary.txt: the matrix associating each class to its list of attributes
- test_images.txt: the list of images which you will use to test your model
- sample_submission.txt: an example submission file
- eval_awa.py: A python script to test the performance of your model.

```
python eval_awa.py --gt test_images.txt --pred example_submission.txt
```

If you wish to use the full resolution version of the dataset, you can download the images from this link: `https://cvml.ist.ac.at/AwA2/`. If you would instead like to use downsampled images (128x128 resolution), they can be downloaded using this the following link:

```
wget https://www.dropbox.com/s/33v1s9ri85o21x7/JPEGImages_128x128.zip?dl=0
```

3 Baseline

For comparision, we have created a simple baseline solution consisting of the following steps:

1. *features:* We used visual bag-of-words features using the KAZE feature detector and descriptor. We used a dictionary of 2000 words obtained by kmeans clustering of the detected features in the training set.

2. *classifier:* Using the extracted features, we trained a seperate linear SVM for each of the 85 attributes.

3. *zero-shot recognition:* At test time, we used our classifier to predict test attributes. We then found the nearest neighbor test class using the Hamming distance (i.e. count of differences) between attribute vectors. But there are much better and more interesting ways to do this: https://cvml.ist.ac.at/AwA2/.

Using our baseline system, we achieve an accuracy of 22.5%. **In order to receive full credit on this assignment, you must achieve an accuracy of at least 20%.**

4 Task

Your task is to train your own zero-shot recognition model on the animals-with-attributes dataset. This is an open-ended assignment, so feel free to be creative with your solutions. You can think of this assignment as a mini-project. You may choose to use any classifier to predict attributes, including neural-networks (more information about GPU access has been posted to Piazza).

Code Restrictions: In your project, you *may* use external libraries and pre-trained models which are not specific to the Animals with Attributes dataset. Examples of acceptable usage include: feature extraction using OpenCV, SciKit Learn for training classifiers, or using networks pre-trained on ImageNet. You *may not* use any code specifically written for the Animals with Attributes dataset. For example, if you find that a paper which tested on the AwA dataset has open-sourced their code, you may not use their code in your implementation.

You will turn in a 4-page report describing your method. You may include whatever information you feel is important, but your report must contain the following:

1. **Description of your approach.** (a) What method did you use to classify the attributes? (b) What method did you use to do zero-shot recognition? Provide enough details so a fellow student might be able to reproduce your results. (c) Include references to papers, codebases, or other resources you consulted when coming up with your approach. A good place to get started is the AwA project page https://cvml.ist.ac.at/AwA2/ which includes a table of other works which test on the dataset.

2. **Accuracy.** What was the accuracy of the zero-shot recognition system?

3. **Error analysis.** (a) Show the confusion matrix of the zero-shot recognition system. Which classes are more or less confused with each other? (b) What was the accuracy of the individual attributes? Which ones were

more or less difficult to recognize? (c) What are the biggest sources of errors in your zero-shot recognition system? Briefly explain why you think your model performs worse on some classes as opposed to others.

4. **Reflection.** (a) Name at least one thing you tried that didn't initially work in your system. What did you do to get it to work?

5. **Next steps.** (a) What would be the next steps you would try to further improve the accuracy? (b) How much of an improvement do you get, or think you could get? You can also experiment directly with next steps; and providing results and careful interpretation will be sufficient.

   Important: be precise and carefully motivate both of your answers. Not sufficient: "(a) The attribute classifier is not accurate enough and should be improved. (b) I think I can get 5% improvement in accuracy." Sufficient: "(a) I hypothesize the number of clusters in the bag of words model is the limiting factor in accuracy. The training error of the attribute classifier model is high (25%), on par with the test error (30%), suggesting that the model is not powerful enough and is underfitting the data. Increasing the number of clusters would increase the discriminative ability of the model. (b) I ran a simple experiment to evaluate this. Since increasing the number of clusters would be computationally expensive, I instead *decrease* the number in half, and observed a drop in attribute accuracy of 3.8% on average, and an overall drop in zero-shot recognition accuracy of 3.2%. This leads me to conclude that in fact there is a strong correlation between the number of clusters and the accuracy of the model. From these results, increasing the number of clusters could be expected to yield 2-4% improvement in accuracy of the attribute classifiers and a 1-3.2% improvement in zero-shot recognition accuracy.