Mining the Web for Object Recognition

Charles Rosenberg
November 16, 1999

Goal

Use machine learning techniques to construct a large number of special purpose object recognizers.

The problem is that this typically this requires hand labeling a large number of examples which is very time consuming and expensive.

Reduce the labeling effort by mining the large number of images on the web that have already been loosely labeled with associated text.

What is different about this work?

These are real world messy images.

Entire images are not clearly labeled as belonging to a particular class.

The result is that each image contains many objects in addition to the object of interest. The assumption is that there is some image region, a subset of the image pixels, which are the "cause" of the specific class label. The algorithm must attempt to discover the statistical properties of that subset which differentiates it from other regions in the image.

This is an example of *multiple instance learning*. In the multiple instance learning framework, the learning algorithm is given "bags" of training examples. A bag is labeled as positive if it contains at least one positive example. A bag is labeled as negative if it contains no positive examples. It is the task of the learning algorithm to learn to distinguish positive from negative examples using this data.

Data Collection Procedure

News web sites were spidered nightly. Images and associated text were collected.

Spider runs were conducted on three web sites:
cnn.com, abcnews.com, dailynews.yahoo.com

The total number of pages visited in a particular site per night was limited because many sites contain large archives. To avoid loops, the same URL was never visited twice in one spidering session.

Description of Data Collected

Approximately 100,000 images were collected which amounted to a total of 3GB of data over a period of 60 days.

Each image and its associated caption was collected as well as the html source for the page which contained it.

Associated text was labeled to be one of the following:

- full marked caption
- half marked caption
- raw text containing the word photo
- raw text containing parenthesis

Data was collected into two databases. One which is a cache of web objects (including both images and html) indexed by their MD5 hash. The second contains pointers to images and the text associated with each image. These are indexed by their MD5 has of the image and caption text.

Care was taken to eliminate duplicates in both databases. This was achieved by pre-filtering html before caching. This is necessary because much of the html today is dynamic, causing the same page to be different each time it is visited.

Proposed Training Process

1) Extract one group of images which contain a specific keyword in their caption and extract a second group of images which do not.
2) For each block in each image calculate the average color in HSV space, discard intensity (V).
3) Quantize the (H,S) pair for each block to one of 200 values.
4) Generate two histograms - one for colors in images containing the keyword and one for images which do not.
5) Use these histograms to build a class conditional probability distribution over colors in and out of the desired class.
6) Segment the image on a block basis based on color using the log odds ratio of the class conditional probability distribution of color.
7) Train an appearance based object recognition algorithm using the segmented examples.

Proposed Recognition Process

1) Use color segmentation to locate sections of the image likely to contain the target object.

2) Scan the trained object recognizer over portions of the image likely to contain the desired object. Use segmented region position and size to achieve some scale and shift invariance.

Future Work

- Use EM to get a better estimate of class conditional color probabilities.
- Use other features - like edges - to aid segmentation.
- Implement region extraction based on segmentation.
- Integrate object recognition algorithm.

References

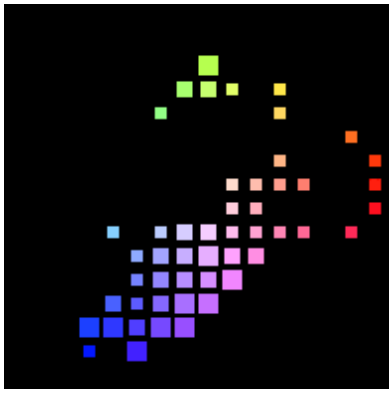
A Framework for Multiple-Instance Learning, by Oded Maron and Tomás Lozano-Pérez, in Neural Information Processing Systems 10, 1998.

Multiple-Instance Learning for Natural Scene Classification, by Oded Maron and Aparna Lakshmi Ratan, in ICML-98.

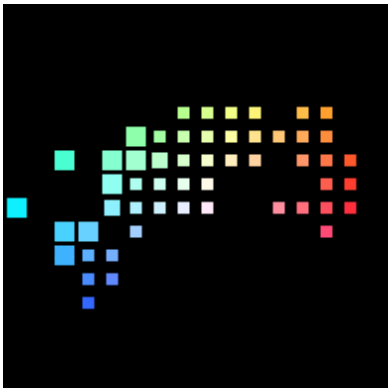
Solving the multiple-instance problem with axis-parallel rectangles, by T. Dietterich, R. Lathrop, and T. Lozano-Pérez, in AI, 89(1-2), pp. 31-71, 1997.

A Note on Learning from Multiple-Instance Examples, by Avrim Blum and Adam Kalai, Machine Learning, 30:23--29, 1998.

Preliminary Segmentation Results



Likelihood ratio for “Clinton” class

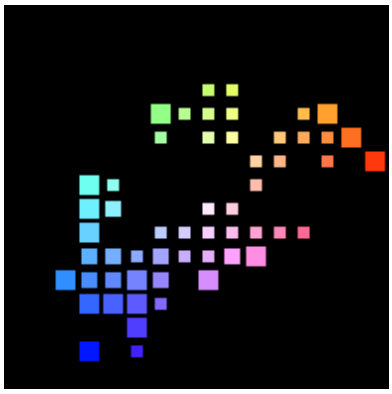


Likeihood ratio for “Not Clinton” class

Colors in the left diagram are more likely to exist in images whose captions contained the word “clinton”. Colors in the right diagram are more likely to exist in images whose captions did not contain the word “clinton”. Colors equally likely to occur in both classes are not shown.

Example color based segementations of images with captions containing the keyword “clinton”. Blocks not considered to be members of the “clinton” class were colored green.

Likelihood ratio for “Flag” class



Likeihood ratio for “Not Flag” class

Colors in the left diagram are more likely to exist in images whose captions contained the word “flag”. Colors in the right diagram are more likely to exist in images whose captions did not contain the word “flag”. Colors equally likely to occur in both classes are not shown.

Example color based segementations of images with captions containing the keyword “flag”. Blocks not considered to be members of the “flag” class were colored green.