Content-Based Image Retrieval Using Multiple-Instance Learning

Qi Zhang
We Yu
Sally A. Goldman
Jason E. Fritts

Image Retrieval Systems

- Image Indexing using keywords – earliest IR systems.
- Lipson et. al., 1997:
  - hand-crafted templates can be used to classify natural scenes.
- Search by content
  - *search-by-similarity*
  - *target search*

Content Based Image Retrieval

- Retrieve images based of automatically derived features such as color, texture and shape.
- Existing CBIR Systems:
  Require user to specify salient regions in the query image.
  - QBIC system [Flickner et. al., 1995]

Multiple Instance Learning

- Traditional Supervised Learning
  - What if the teacher can only label collection of instances, and not individual instances?
- Multiple-Instance Learning (MIL) [Dietterich et. Al., 1997]
  - Learn a concept given a collection of instances labeled positive or negative.

A Natural Scene Classification Example

Given a picture containing a waterfall, what is it about the image that causes it to be labeled as a waterfall?

At least one of the objects in the image is a waterfall.

Given a number of images (labeled waterfall or non-waterfall), the authors attempt to find the intersections within the waterfall images, that do not appear in the non-waterfall images.

Problem Description

- Each bag corresponds to a single image and may contain many instances.
- Each instance in a bag corresponds to a description of some sub-region of the image.
- A bag is labeled positive if it contains *at least one* positive instance and negative otherwise.
- From a collection of labeled bags, the learner tries to induce a concept that will label unseen bags correctly.

Steps for CBIR using MIL

1. Image Processing (extracting image features and generating bags).

2. Multiple Instance Learning using Diverse Density algorithm.

Image Processing Steps

1. Color representation: using one of the following systems
   - RGB
   - YCrCb

2. Bag Generator:
   - Segmentation
   - Subsampling

Segmentation

- Divide the image into regions called *blobs*.
- Each *blob* is represented by 6 values.
  - < R, G, B, HL(Y), LH(Y), HH(Y) > or
    < Y, Cr, Cb, HL(Y), LH(Y), HH(Y) >
- K-means segmentation algorithm is used to segment the image.
- So we have one bag for each image, one 6-D point for each segment in the bag.

Subsampling

- Smoothes the image followed by subsampling.

- Applies a fixed segmentation method to generate the points within the bag for each image.

- Single blob with neighbors (sbn) to extract features.

Multiple-Instance Learning

- Training Data:
  $D = \{<B_1, l_1>, \ldots, <B_m, l_m>\}$
  m bags where bag B_i has label l_i.

  If bag $B_i = \{ B_{i1}, \ldots, B_{ij}, \ldots B_{in}\}$, then B_{ij} is the jth instance in B_i.
  Positive Bags: B_i^+

Diverse Density Algorithm [Maron and Lozano-Perez, 1998]

- Main idea:
  Find a point in feature space that have a *high Diverse Density* –
  - High density of positive instances
  - Low density of negative instances
  - Higher diverse density = higher probability of being the target concept.

Diverse Density

Assuming that the target concept is a single point t and x is some point in feature space,

$\Pr(x = t \mid B_1^+, \ldots, B_n^+, B_1^-, \ldots B_m^-)$ …………(1)

represents the probability that x is the target concept given the training examples.

We can find t if we maximize the above probability over all points x.

Probabilistic Measure of Diverse Density

- Using Bayes' Rule, maximizing (1) is equivalent to maximizing $\Pr(B_1^+, \ldots, B_n^+, B_1^-, \ldots B_m^- \mid x = t)$
- Further assuming that the bags are conditionally independent given t, the best hypothesis is $\text{argmax}_x \prod_i \Pr(B_i^+ \mid x = t) \prod_i \Pr(B_i^- \mid x = t)$

General Defn. Of DD

- Again using Bayes' rule, the above term is equivalent to $\text{argmax}_x \prod_i \Pr(x = t \mid B_i^+) \prod_i \Pr(x = t \mid B_i^-)$
- x will have high Diverse Density if every positive bag has an instance close to x and no negative bags are close to x.

Noisy-or model

The causal probability of instance j in bag B_i

$\Pr(x = t \mid B_{ij}) = \exp(-\| B_{ij} - x \|^2)$

A positive bag's contribution:

$\Pr(x = t \mid B_i^+) = 1 - \prod_j (1 - \Pr(x = t \mid B_{ij}^+))$

A negative bag's contribution:

$\Pr(x = t \mid B_i^-) = \prod_j (1 - \Pr(x = t \mid B_{ij}^-))$

Feature Relevance

- "closeness" depends on the features
- Problem: Some features might be irrelevant, and some others might be more important than the others.
  $\| B_{ij} - x \|^2 = \text{sigma}_k\, w_k\, (B_{ijk} - x_k)^2$
- Solution: "weight" the features depending on their relevance. Find the best weighting of the features by finding the weights that maximize Diverse Density.

Finding the maximum DD

- Use gradient ascent with multiple starting points.

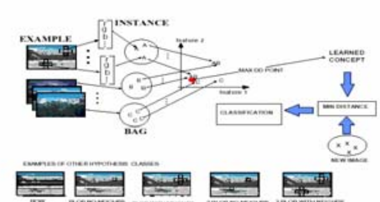


Figure 1: System Diagram

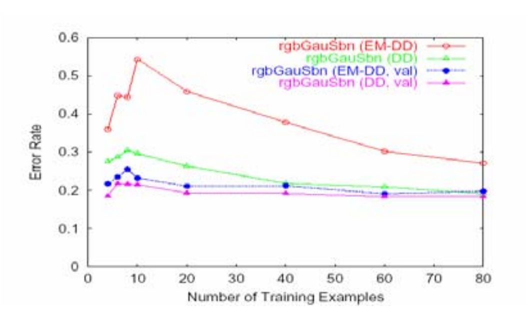
EM-DD

- Expectation Maximization algorithm (Dempster, Laird and Rubin (1977))
  - Start with *h* set to some appropriate instance from a positive bag.
  - E-Step: *h* is used to pick one instance from each bag that is most likely to be responsible for its label.
  - M-Step: two-step gradient ascent search to find a new *h* that maximizes *DD(h)*.

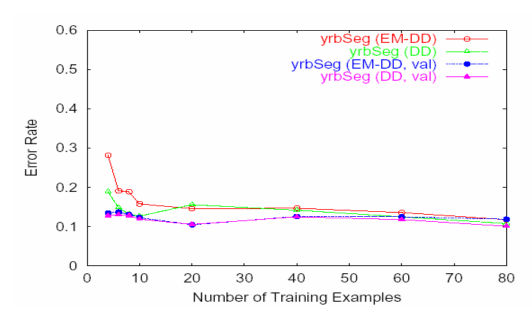
Results

- Maron and Ratan's algorithm used as a benchmark.
- Standard learning curves, *recall* curves, *precision* versus *recall* curves.

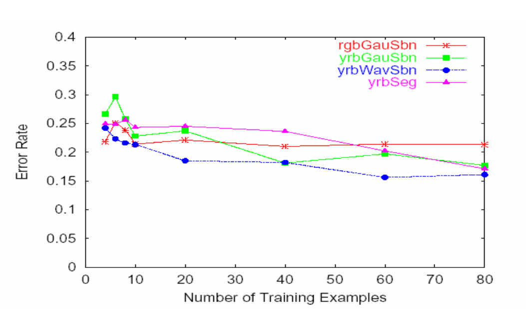
Comparing DD and EM-DD



Comparing Different Image Processing Techniques



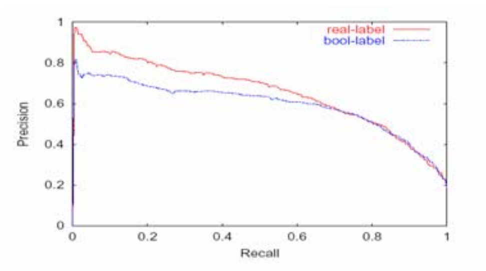
EM-DD using different feature selection methods



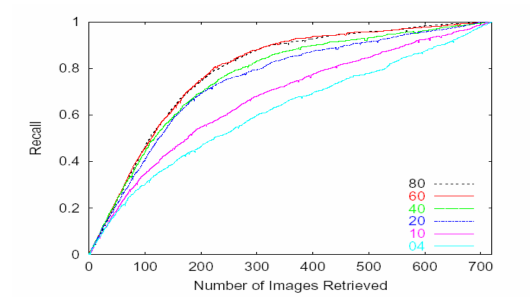
Recall and Precision Plots

- Recall: $\frac{\text{number of +ve images retrieved}}{\text{total number of +ve images in the set}}$
- Precision: $\frac{\text{number of +ve images retrieved}}{\text{number of images retrieved}}$

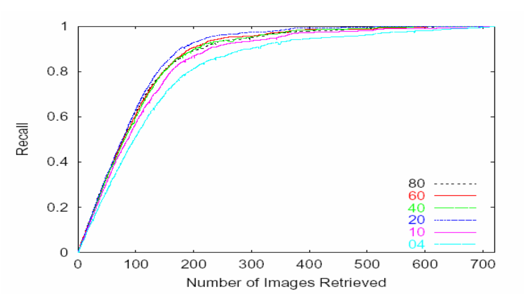
Real-valued versus Boolean Labels



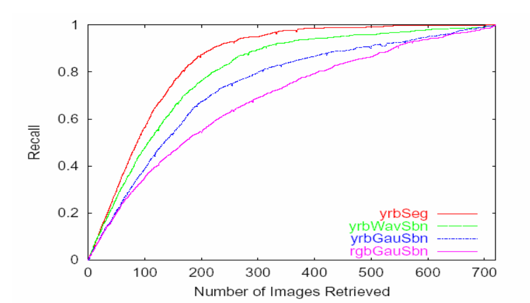
Recall Plot for RGB with Gaussian filter with sbn fixed segmentation



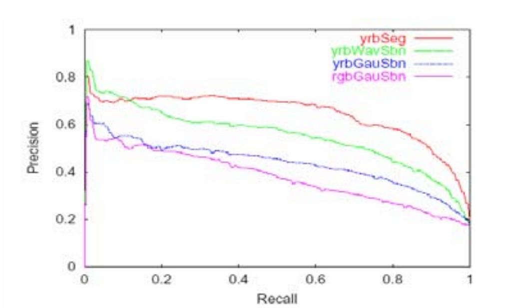
Recall Plot for YCrCb with Segmentation Feature Selector



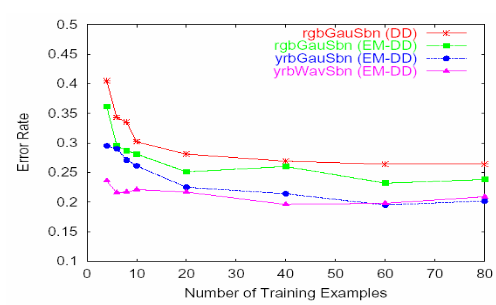
Recall Plot with 6 training examples



Precision Plot with 6 training examples



Less variety in images included in test and training data



Conclusion

- EM-DD performed better than DD for the drug discovery case.
- Future Work:
  - Explore other methods for feature selection (two-blob with neighbor-bag generator)
  - Hierarchical segmentation technique.
  - Alternate ways to select best hypothesis.

References

- Content-Based Image Retrieval Using Multiple Instance Learning by Zhang, Yu, Goldman, Fritts
- Maron & Ratan (1998). Multiple Instance Learning for Natural Scene Classification
- Maron & Lozano-Perez (1998), A Framework for Multiple Instance Learning

Assumptions of Noisy-or model

- for x to be the target concept, it is caused by one of the instances in the bag.

- probability of j not being the target is independent of any other instance not being the target.

Content-Based Image Retrieval Using Multiple-Instance Learning

paper by
Qi Zhang, Wei Yu,
Sally Goldman, Jason Fritts

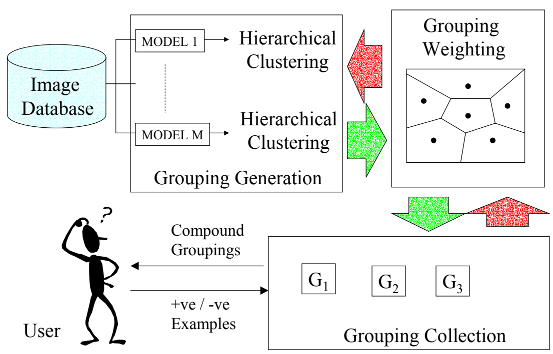
presented by
Navdeep Kaur

comments by
Siddharth Patwardhan

An Interactive Learning Model

- **"Four eyes" – an interactive learning system.**
- **The system learns *groupings* of images or image "patches" based on positive and negative examples.**
- **It interactively provides the user with examples and improves the *groupings* based on user feedback.**

Schematic



Multiple Instance Learning

- User doesn't need to select a particular region of an image.
- Learning algorithm learns the region of interest and its characteristics, based on feedback.

References

[1] Zhang Q. et al. Content-Based Image Retrieval Using Multiple-Instance Learning. *International Conference on Machine Learning (ICML 2002),* Sydney, Australia, July 2002.

[2] Minka T. and Picard R. Interactive Learning Using a "Society of Models". *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR-1996).* June 18 - 20, 1996 San Francisco, Ca.

Content-based Image Retrieval Using Multiple-instance Learning

Qi Zhang, We Yu, Sally A.Goldman
Jason E. Fritts

Presented by: Navdeep Kaur

Bag Generator

- Key part of the system
- Mechanism – takes image and generates a set of instances
- Instance – possible description of the image

Observation- Better bag generator leads to simpler learning algorithm

Multiple-instance Learning

- Bag - each training image
- Instances – description of various sub images
- Several ways to describe an instance

Hypothesis Classes

- *Row*: row's mean color and color difference in rows above and below it
- *Single blob with neighbors*: mean color of 2 x 2 blob & color difference with its 4 neighboring blobs
- *Single blob with no neighbors*: color of each pixel in a 2 x 2 blob
- *Disjunctive blob with neighbors*: same as *single blob with neighbors* – concept learned is disjunction of two single blob concepts

Hypothesis Classes

- *Disjunctive blob with no neighbors*: same as *single blob with no neighbors* – concept learned is disjunction of two single blob concepts
- *Two blob with neighbors* – mean color of two descriptions of two *single blob with neighbors* and relative spatial relationships
- *Two blob with no neighbors* – mean color of two descriptions of *two single blob with no neighbors* and their spatial relationship

References

- Multiple-Instance Learning for Natural Scene Classification – Oded Maron, Aparna Lakshmi Ratan