Unsupervised learning of Categories from Sets of Partially Matching Features

Kristen Grauman and Trevor Darrell
CVPR 2006

Presented By
Shilpa Gulati and Muhammad Zubair Malik
4-12-2007

Motivation
- Automatically learn object categories from unlabeled images.
- Allow optional supervision.
- Train classifiers for new images.

Background
The Pyramid Match Kernel

The Need of a good Kernel
- Discriminative classifiers require a good similarity measure
- Given a good similarity measure we can train a whole range of classifiers (perceptrons, pca, hyper-plane or any complicated shape classifiers)
- Of particular interest are hyper-plane, maximum margin classifiers: Support Vector Machines
- Looking at support vectors is far efficient than calculating all pair-wise comparisons (NN)

The Notion of similarity
- Dot Product works well for linearly separable data that is not noisy
- If the data is not linearly separable and if we can find a symmetric semi-positive definite function of it, we can use its dot product
- The Kernel matrix encodes this information
- The eigen functions of Mercer kernel act as features
- Issues
  - Computational complexity increases (high dimension)
  - Risk Over fitting the data

Image representation
- An image is a set local features.

\[ X = \{ \bar{x}_1, \ldots, \bar{x}_m \} \quad Y = \{ \bar{y}_1, \ldots, \bar{y}_n \} \]
- The sets X and Y can have different sizes.
What is a good kernel in this case?

- Kernel matrix is the information bottleneck, and must be selected discreetly
  - A diagonal kernel matrix means we cannot discriminate, all data points are orthogonal to each other and no clusters exist
- In this case we have to consider:
  - Each instance is unordered set of vectors
  - Varying number of vectors per instance

Partial matching for sets of features

- Compare sets by computing a *partial matching* between their features.

Pyramid match

- Place multi-dimensional, multi-resolution grid over point sets
- Consider points matched at finest resolution where they fall into same grid cell
- Approximate similarity between matched points with worst case similarity at given level

*No explicit search for matches!*

Pyramid match overview

- Pyramid match kernel measures similarity of a partial matching between two sets:
  - Place multi-dimensional, multi-resolution grid over point sets
  - Consider points matched at finest resolution where they fall into same grid cell
  - Approximate similarity between matched points with worst case similarity at given level

How to build a multi-dimensional, multi-resolution Histogram?

- We Need to build L such histograms
- Informally O(dmL)
- Given the tree template, we make d decisions for m elements in L such trees

Copied from Grauman Slides
How to apply the Kernel

\[ H(X) \]

\[ H(Y) \]

\[ \mathcal{I}(H(X), H(Y)) = \sum_{j=1}^{n} \min(H(X)_j, H(Y)_j) \]

\[ H(Y) = 6 \]

\[ = 5 \]

\[ w = 1/(2^{L-1}) \]

PMK is a Mercer Kernel

- Clearly histogram intersection is a positive definite function (min of two positive numbers cannot be less than zero)
- Mercer Kernels have good modularity properties.
- Given two kernels K1 and K2, and constants a,b
  - K1+K2 is also a mercer kernel
  - aK1 is also a kernel
  - aK1+bK2 is also a kernel

Application in Software Engineering

Given Example Structures find the Structural Invariants and output heuristics to check if a test structure satisfies them

Use PMK And train classifier

Use PMK

View as Graph

RepOk
Back to unsupervised learning of categories

Similarity between two images

- Similarity between two sets of features:
  - Use pyramid match kernel.
  - A scalar value for each pair of images.
- Partial Matching
  - The points (features) of smaller set are mapped to subset of points of larger set.
  - Not all features are matched.

Similarity Matrix

\[
\begin{array}{cccc}
X_1 & X_2 & \ldots & X_N \\
X_1 & K_{11} & K_{12} & \ldots & \ldots \\
X_2 & K_{21} & K_{22} & \ldots & \ldots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
X_N & K_{N1} & K_{N2} & \ldots & K_{NN}
\end{array}
\]

\(K_{ij}\) is a scalar representing the similarity between sets \(X_i\) and \(X_j\).

Clustering

- Partition the data into two sets
  - Shi and Malik’s efficient approximation of Normalized Cuts algorithm.
- Optimal partitioning
  - Maximizes intra-cluster similarity.
  - Minimizes inter-cluster similarity.
- For \(k > 2\) clusters
  - Recursively partition.
Partial Matching: Again

- Partial Matching: a key feature.
  - Not all features of a set are matched.

- This implies
  - Robustness to clutter and occlusion.
    - These features most likely will not be matched to anything.
  - But images with similar backgrounds may match.
    - Since some of the features match!

Overcoming the limitations

- Within each cluster, for each cluster member:
  - Identify features that form matching with other cluster members.
  - Greater number of matchings imply greater weight for that feature.
Inferring feature masks

Refining intra-cluster matches

Refining intra-cluster matches

Refining intra-cluster matches

Selecting category prototypes
Selecting category prototypes

Inferred feature masks

Unsupervised recovery of category prototypes

Unsupervised recovery of category prototypes

Semi-supervised category labeling

Conclusions

- Main contribution
  - Efficient unsupervised (semi-supervised) category learning.
  - Co-occurrences among features are naturally preserved.
  - Feature masks inferred automatically.
    - Roughly identify location of objects.