Predictive Clustering for Image Annotation & Retrieval

Sašo Džeroski
Jozef Stefan Institute, Ljubljana, Slovenia
(joint work with Ivica Dimitrovski, Dragi Kocev and Suzana Loškovska)
• Predictive clustering
  • From predictive modeling and clustering to predictive clustering
  • Predictive clustering for predicting structured outputs
  • Learning predictive clustering trees
  • Ensembles of predictive clustering trees

• Image annotation with PCTs and ensembles
  • Taxonomical classification of diatom images
  • Hierarchical annotation of medical images

• Visual codebook construction with PCTs and ensembles
  • Supervised for image annotation
  • Unsupervised for image retrieval
Predictive models focus on a target variable and predict its value from the values of input variables.

Classical problem: Medical diagnosis

An example: Neurodegenerative diseases

Target variable: Diagnosis; Possible values:
- CN - Cognitively Normal (0)
- SMC - Significant Memory Concern
- EMCI - Early Mild Cognitive Impairment
- LMCI - Late Mild Cognitive Impairment
- AD - Alzheimer‘s Disease (4)

Descriptive vars.: genetic and image markers
Predictive Modelling

• Input: A table of data, a row is an object, single target

<table>
<thead>
<tr>
<th></th>
<th>Descriptive space</th>
<th>Target space</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender</td>
<td>Fusiform</td>
</tr>
<tr>
<td>Example 1</td>
<td>F</td>
<td>16471</td>
</tr>
<tr>
<td>Example 2</td>
<td>M</td>
<td>20680</td>
</tr>
<tr>
<td>Example 3</td>
<td>F</td>
<td>18751</td>
</tr>
<tr>
<td>Example 4</td>
<td>M</td>
<td>22895</td>
</tr>
<tr>
<td>Example 5</td>
<td>F</td>
<td>18446</td>
</tr>
<tr>
<td>Example 6</td>
<td>F</td>
<td>16056</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

• Output: A predictive model for the target, e.g. decision tree
Top-Down Induction of Classification (Regr.) Trees

To construct a tree $T$ from a training set $S$:

• If all the examples belong to the same class $C$ (the values of the target have low variance), construct a leaf labeled with the class value $C$ (the target average)

• Otherwise:
  • Select the best attribute $A$ with values $v_1, \ldots, v_n$, which reduces the most the impurity of the target
  • Partition $S$ into $S_1, \ldots, S_n$ according to $A$
  • Recursively construct subtrees $T_1$ to $T_n$ for $S_1$ to $S_n$
  • Result: a tree with root $A$ and subtrees $T_1, \ldots, T_n$
Clustering/ Unsupervised L.

- Partition a set of objects into clusters of similar objects
- High similarity of objects within individual clusters, low similarity between objects from different clusters
- Minimize intra-cluster variance (ICV)
- Distance/similarity measure in the example space
Basic Clustering Approaches

• K-Means clustering

  • Randomly assign instances to k clusters, then repeat:
    • Calculate centroids of clusters, reassign instances to clusters
    • Until convergence (i.e., cluster assignment doesn’t change)

• Hierarchical agglomerative clustering

  • Start with each instance as a cluster, then repeat
    • Merge the two closest clusters
    • Until all instances are in one single cluster
Predictive Clustering

• Combines prediction and clustering

• We can have hierarchical clustering (trees) and flat/overlapping clusterings (rules)

• With each cluster, predictive clustering provides
  • A description of the cluster
  • A prediction of the selected targets for that cluster

• The output of PC can be viewed both as a clustering and as a predictive model (cf. next example)
Example Task: Cluster Alzheimer’s Patients wrt. Clinical Scores

1. CDRSB – Clinical Dementia Rating Sum of Boxes
2. ADAS13 – AD assessment scale
3. MMSE – Mini Mental State Examination
4. RAVLT (immediate, learning, forgetting, perc. forgetting) – Rey Auditory Verbal Learning Test (4 features)
5. FAQ – Functional Assessment Questionnaire
6. MOCA – Montreal Cognitive Assessment
7. EcogPt (Memory, Language, Visuospatial Abilities, Planning, Organization, Divided Attention, Total score) – Everyday cognition questionnaire – filled in by patient (7 features)
8. EcogSP (Memory, Language, Visuospatial Abilities, Planning, Organization, Divided Attention, Total score) – Everyday cognition questionnaire – filled in by study partner (7 features)
Example Predictive Clustering Tree for Multi-Target Regression

- Attributes used to descr. clusters: Biomarkers (as above)
- Targets attributes for clustering: diagnosis (0-4)+ clinical msmnts/scores (23 of them)

- DX
- CDRSB
- ADAS13
- MMSE
- ...

Cluster 1
N=154
2.00
1.00
14.9
27.9
...

Cluster 2
N=408
1.37
0.70
10.7
28.8
...

Cluster 3
N=108
2.31
1.97
18.2
27.1
...

Cluster 4
N=126
3.23
3.38
26.4
24.8
...
Multi-Label Classification

- Special case of multi-target prediction (incl. MTR & MTC)
- Learning models that simultaneously predict several binary target variables (a set of labels)
- Input: A vector of descriptive variables (as for STC/STR)

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Temperature</th>
<th>K&lt;sub&gt;2&lt;/sub&gt;Cr&lt;sub&gt;2&lt;/sub&gt;O&lt;sub&gt;7&lt;/sub&gt;</th>
<th>NO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Cl&lt;sup&gt;-&lt;/sup&gt;</th>
<th>CO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>Cladophora sp.</th>
<th>Gongrosira incrustans</th>
<th>Oedogonium sp.</th>
<th>Stigeoclonium tenue</th>
<th>Melosira varians</th>
<th>Nitzschia paelea</th>
<th>Aurouinella chalybea</th>
<th>Erythrobella octoculata</th>
<th>Gammarus fossarum</th>
<th>Baetis rhodani</th>
<th>Hydropsyche sp.</th>
<th>Rhyacophila sp.</th>
<th>Simulium sp.</th>
<th>Tubifex sp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID1</td>
<td>0.66</td>
<td>0.00</td>
<td>0.40</td>
<td>1.46</td>
<td>0.84</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ID2</td>
<td>2.03</td>
<td>0.16</td>
<td>0.35</td>
<td>1.74</td>
<td>0.71</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ID3</td>
<td>3.25</td>
<td>0.70</td>
<td>0.46</td>
<td>0.78</td>
<td>0.71</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Multi-Label Classification Example

- A decision tree for multi-label classification

- $NO_2 > 0.12$
  - $K_2Cr_2O_7 > 0.86$
    - yes
      - $Cl > 0.35$
        - yes
          - $Temperature > 2.55$
            - Cladophora sp.
            - Oedogonium sp.
            - Nitzschia pal.
            - Erpobdella oc.
            - Gammarus fo.
            - Hydropsyche sp.
          - no
            - Melosira var.
            - Nitzschia pal.
            - Gammarus fo.
            - Hydropsyche sp.
        - no
          - $Cl > 0.22$
            - yes
              - $CO_2 > 0.0$
                - yes
                  - Rhyacophila sp.
                - no
              - no
          - no
            - Melosira var.
            - Audouinella ch.
            - Gammarus fo.
            - Baetis rh.
            - Rhyacophila sp.
            - Cladophora sp.
            - Melosira var.
            - Gammarus fo.
            - Hydropsyche sp.
Hierarchical Multi-Label Classification (HMC)

- Labels organized in a hierarchy
- Taxonomic classification of diatoms
- From microscopic images
- Taking into account the existing taxonomy of diatoms

<table>
<thead>
<tr>
<th>image</th>
<th>features/descriptors</th>
<th>taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heuristic shape descriptors</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>24 59 66 37 ...</td>
<td>olivaceum</td>
</tr>
<tr>
<td>36</td>
<td>25 53 45 15 ...</td>
<td>minutissimum</td>
</tr>
<tr>
<td>35</td>
<td>25 56 52 19</td>
<td>exigua</td>
</tr>
<tr>
<td>...</td>
<td>... ... ... ...</td>
<td>...</td>
</tr>
</tbody>
</table>

![Taxonomic diagram of diatoms]

13
Top-down induction of PCTs

To construct a tree $T$ from a training set $S$:

• **If the examples in $S$ have low variance,**
  construct a leaf labeled $\text{target}(\text{prototype}(S))$

• **Otherwise:**
  • Select the best attribute $A$ with values $v_1, \ldots, v_n$, which **reduces the most the variance** (measured according to a given distance function $d$)
  • Partition $S$ into $S_1, \ldots, S_n$ according to $A$
  • Recursively construct subtrees $T_1$ to $T_n$ for $S_1$ to $S_n$
  • Result: a tree with root $A$ and subtrees $T_1, \ldots, T_n$
Learning PCTs

• Recursively partition data set into subsets (clusters) with low intra-cluster variance
  • Variance = avg. squared distance to prototype
    \[ ICV(S) = \sum_{y_j \in S} d(y_j, p(S))^2 \]
  • For the variance, the distance is measured
    • In standard clustering, along all dimensions
    • In prediction, along a single target dimension
    • In predictive clustering, along a structured target, e.g., several target dimensions
**Predictive clustering:** A divides data into clusters 1 and 2 coherent along two dimensions.
Distances/variances for SOP tasks

• The algorithm

• Variance for MT regression

\[ \text{Var}(E) = \sum_{i=1}^{T} \text{Var}(Y_i). \]

• Variance for MT classification

\[ \text{Var}(E) = \sum_{i=1}^{T} \text{Entropy}(E, Y_i) \]

• Variance for HMLC

\[ \text{Var}(E) = \frac{1}{|E|} \cdot \sum_{E_i \in E} d(L_i, \bar{L})^2 \]

\[ d(L_1, L_2) = \sqrt{\sum_{i=1}^{|L_i|} w(c_i) \cdot (L_{1,i} - L_{2,i})^2} \]

\textbf{procedure} BestTest($E$)

1: \((t^*, h^*, \mathcal{P}^*) = (\text{none}, 0, \emptyset)\)

2: \textbf{for each} possible test \(t\) \textbf{do}

3: \(\mathcal{P} = \text{partition induced by } t \text{ on } E\)

4: \(h = \text{Var}(E) - \sum_{E_i \in \mathcal{P}} \frac{|E_i|}{|E|} \text{Var}(E_i)\)

5: \textbf{if} \((h > h^*) \land \text{Acceptable}(t, \mathcal{P})\) \textbf{then}

6: \( (t^*, h^*, \mathcal{P}^*) = (t, h, \mathcal{P})\)

7: \textbf{return} \((t^*, h^*, \mathcal{P}^*)\)
Ensembles of PCTs

- Ensembles of PCTs use several methods for constructing base classifiers
  - Bagging & Random forests
  - Random subspaces & Bagged Random subspaces

- PCTs and Ensembles of PCTs implemented in SW package CLUS, jointly developed by JSI, Ljubljana and KULeuven, Belgium

- Written in Java

- Open source, available for download from http://sourceforge.net/projects/clus
Ensembles of PCTs: Bagging

Ensembles of PCTs:

Training set

\[ 1 \quad 2 \quad 3 \quad \ldots \quad n \]

\( n \) bootstrap replicates

\( n \) classifiers

\( n \) predictions

Test set

\[ L_1 \quad L_2 \quad L_3 \quad \ldots \quad L_n \]

vote

Ensembles of PCTs: Bagging
SSL: Incomplete Annotations

- Some examples have labels, some don’t, some incmpl.

<table>
<thead>
<tr>
<th>Example</th>
<th>Descriptive space</th>
<th>Target space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>1</td>
<td>TRUE 0.49</td>
</tr>
<tr>
<td>Example 2</td>
<td>2</td>
<td>FALSE 0.08</td>
</tr>
<tr>
<td>Example 3</td>
<td>1</td>
<td>FALSE 0.08</td>
</tr>
<tr>
<td>Example 4</td>
<td>2</td>
<td>TRUE 0.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>Descriptive space</th>
<th>Target space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>1</td>
<td>TRUE 0.49</td>
</tr>
<tr>
<td>Example 2</td>
<td>2</td>
<td>FALSE 0.08</td>
</tr>
<tr>
<td>Example 3</td>
<td>1</td>
<td>FALSE 0.08</td>
</tr>
<tr>
<td>Example 4</td>
<td>2</td>
<td>TRUE 0.49</td>
</tr>
<tr>
<td>Example 5</td>
<td>3</td>
<td>TRUE 0.49</td>
</tr>
<tr>
<td>Example 6</td>
<td>4</td>
<td>FALSE 0.08</td>
</tr>
</tbody>
</table>
Semi-Supervised Learning w. PCTs

• New definition of variance that includes both targets and attributes, e.g., for MTR

\[
Var(E) = \frac{1}{T + D} \cdot \left( w \cdot \sum_{i=1}^{T} Var(Y_i) + (1 - w) \cdot \sum_{j=1}^{D} Var(X_j) \right)
\]

• \(T = \text{#target attributes, } D = \text{#descriptive attributes}\)
• \(w = \text{weight parameter, trades-off focus on}\)
  • Prediction \((w=1)\)
  • Clustering \((w=0)\)
• \(w\) tuned by internal cross-validation on labeled part
SSL: Calculating Variance for Attributes with Missing Values

Variances of individual target \((Y_i)\) and descriptive \((X_i)\) attributes:

\[
Var(Y_i) = \frac{\frac{N - 1}{K_i - 1} \cdot \sum_{j=1}^{N} (y_{ij})^2 - N \cdot \left( \frac{1}{K_i} \cdot \sum_{j=1}^{N} y_{ij} \right)^2}{N}
\]

\(N\) = number of examples,

\(K_i\) = number of examples with non missing values

In extreme cases \((K = 0)\), est. var. with var. of parent node:

(1) leafs of the decision tree may contain only unlabeled examples

(2) in a leaf, some descr. attributes may have only missing values.
Image Annotation and Retrieval with PCTs
Taxonomic Identification of Diatoms from Microscope Images

• Automated diatom classification
  • image processing (feature extraction from images)
  • image classification (labels and groups the images)

• Labels organized in a hierarchy

• Predict all different levels in the hierarchy of taxonomic ranks: genus, species, variety, and form

• Goal of the complete system: assist a taxonomist in identifying a wide range of different diatoms
Feature Extraction from Images

- Contour extraction, then
- Simple geometric properties
- Length, width, size and the length-width ratio
- Simple shape descriptors: rectangularity, triangularity, compactness, ellipticity, and circularity
- Fourier descriptors (30 coefficients)
- SIFT histograms (key-point detection+)
  - Invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint
  - Cluster key-points, assign KPs to clusters, hist.
Diatom Classification Results

- Predictive performance of the different feature sets and their combinations

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Descriptors</th>
<th># features</th>
<th>55 diatom taxa</th>
<th>48 diatom taxa</th>
<th>37 diatom taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Overall recognition rate [%]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bagging</td>
<td>Geometric and shape descriptors</td>
<td>9</td>
<td>76.3</td>
<td>76.7</td>
<td>77.2</td>
</tr>
<tr>
<td></td>
<td>Fourier descriptors</td>
<td>30</td>
<td>86.7</td>
<td>88.1</td>
<td>88.6</td>
</tr>
<tr>
<td></td>
<td>SIFT histograms</td>
<td>200</td>
<td>88.4</td>
<td>89.2</td>
<td>91.3</td>
</tr>
<tr>
<td></td>
<td>Geometric and shape desc.+Fourier desc.+SIFT hist.</td>
<td>239</td>
<td>96.2</td>
<td>98.1</td>
<td>98.8</td>
</tr>
<tr>
<td>Random Forests</td>
<td>Geometric and shape descriptors</td>
<td>9</td>
<td>76.3</td>
<td>76.7</td>
<td>77.2</td>
</tr>
<tr>
<td></td>
<td>Fourier descriptors</td>
<td>30</td>
<td>86.6</td>
<td>88.1</td>
<td>88.7</td>
</tr>
<tr>
<td></td>
<td>SIFT histograms</td>
<td>200</td>
<td>88.2</td>
<td>87.9</td>
<td>91.1</td>
</tr>
<tr>
<td></td>
<td>Geometric and shape desc.+Fourier desc.+SIFT hist.</td>
<td>239</td>
<td>96.2</td>
<td>98.1</td>
<td>98.7</td>
</tr>
</tbody>
</table>
Medical Image Annotation

• ImageCLEF2009 Challenge
  • 12677 annotated x-ray images; 1733 non-annotated images

• Hierarchical classification according to two labeling sets:
  • ImageCLEF2007: 116 IRMA codes
  • ImageCLEF2008: 196 IRMA codes
IRMA Coding System

• Four axes marked with \{0, ..., 9, a, ..., z\}
  • T (Technical): image modality
  • D (Directional): body orientation
  • A (Anatomical): body region
  • B (Biological): biological system

• IRMA code: TTTT – DDD – AAA – BBB

• The code is strictly hierarchical
  5      uropoietic system
  51     uropoietic system, kidney
  512    uropoietic system, kidney, renal pelvis
Medical Image Annotation

- Set of images with their visual descriptors and annotations
- Annotations with IRMA codes, hierarchical

<table>
<thead>
<tr>
<th>image</th>
<th>features/descriptors</th>
<th>annotations/labels</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Image" /></td>
<td>48 24 59 66 37 ...</td>
<td>cervical spine@musculoskeletal system</td>
</tr>
<tr>
<td><img src="image.png" alt="Image" /></td>
<td>36 25 53 45 15 ...</td>
<td>middle abdomen@renal pelvis</td>
</tr>
<tr>
<td><img src="image.png" alt="Image" /></td>
<td>35 25 56 52 19</td>
<td>lumbar spine@musculoskeletal system</td>
</tr>
<tr>
<td>...</td>
<td>... ... ... ... ... ... ...</td>
<td>...</td>
</tr>
</tbody>
</table>
Feature Extraction

• Local Binary Pattern (LBP) histograms
• Edge Histogram Descriptor (EHD)
• Scale Invariant Feature Transform (SIFT) histograms
• Raw pixel representation (RPR)
  • Scale the image to a common size (32x32 pixels)
  • Represent the image by a feature vector that contains image pixel values
Local Binary Patterns

• Binary code to describe the local texture pattern in a circular region thresholding each neighborhood on the circle by the gray value of its center

<table>
<thead>
<tr>
<th>75</th>
<th>99</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>45</td>
<td>63</td>
</tr>
<tr>
<td>74</td>
<td>36</td>
<td>31</td>
</tr>
</tbody>
</table>

threshold

1 1 0
1 1 1
1 0 0

binary code 11010011

• Circular symmetric neighborhood with different radius R and number of points P

(R=1,P=4)  (R=1,P=8)
LBP Histograms

- Image divided in 4x4 parts
- From each sub-image extract ULBP(1,8)
Edge Histogram Descriptor

- Sharp change of luminous intensity
- Information about the shapes of the objects
- Frequency and the directionality of the brightness changes in the image
SIFT: Bag of Visual Words

- Extract local SIFT features
- Construct visual word dictionary
- Using K-means clustering
- Vocabulary size – number of visual words
- Local feature histogram
Medical Image Annotation

Comparative study of ensembles of PCTs for HMC and collections of SVMs, one per label

Summary of results

• Ensembles (RFs) of PCTs for HMC perform better
  • Lower hierarchical error measure
  • Higher overall recognition rate
  • Best results on these datasets so far

• RFs of PCTs for HMC are also much more efficient/faster
Constructing BOW Codebooks w PCTs

• Visual codebook construction
  • Unsupervised for image retrieval
  • Supervised for image annotation

• Image annotation with hierarchically structured labels (medical X-ray images) and general images

• We used (small) ensembles of PCTs for constructing BOW codebooks

• We learned to annotate using collections of SVMs
Bag-of-Visual-Words (BoVW)

1. Extract features
   • Select key points/patches/regions

![Images of different objects with selected patches highlighted]

• Calculate descriptors/features of the selected patches

<table>
<thead>
<tr>
<th>Patch 1</th>
<th>Patch 2</th>
<th>Patch 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Patch 1 Image" /></td>
<td><img src="image2.png" alt="Patch 2 Image" /></td>
<td><img src="image3.png" alt="Patch 3 Image" /></td>
</tr>
<tr>
<td>12 45 78 ...</td>
<td>34 56 124 ...</td>
<td>1 6 84 ...</td>
</tr>
</tbody>
</table>
2. Learn a visual codebook

- Input: Set of descriptors
- Output: Clusters (Visual Words)
BoVW: Representing Images

1. Extract features
2. Learn a visual codebook
3. Represent the images by histograms (distribution of the patches over the visual words)
Related Work

• Construction of a visual codebook is a bottleneck in the bag-of-visual-words approach
• k-means to cluster local image regions into visual words
  • Serious limitations for large scale object retrieval
• Hierarchical k-means, approximate k-means and extremely randomized tree ensembles
  • Improve the efficiency at the cost of decrease of the discriminative power of the obtained codebook
• **Our method**: Visual codebook construction using predictive clustering trees to alleviate the efficiency issues and increase the predictive power
Codebook: Random forest of PCTs

• Here we use a small number of trees in the forest

• Large scale object retrieval
  • Random forest of PCTs for multi-target regression
  • Descriptive and target space are the same

• Multi-label image annotation
  • Random forest of PCTs for multi-label classification
  • Use the annotations of the images to guide the construction of the visual codebooks
Visual Codebook

• Each tree leaf is a visual word
• Each image is described with a histogram of the number of regions per visual word

• PCTs are computationally efficient in both construction and prediction, but rather unstable: small random forest of PCTs to obtain the overall codebook
• Concatenation of the codebooks of each PCT
Data Description

• Oxford5k dataset: 5062 high-resolution images of Oxford landmarks
• Paris dataset: 6412 high-resolution images
• Pythia: 5555 high-resolution images
• PASCAL VOC 2007: 9963 images, 20 labels, 1.46 labels per image
• ImageCLEF@ICPR: 8000 images, 53 labels, 8.68 labels per image
• ImageCLEF 2010: 8000 images, 93 labels, 12.06 labels per image
• Oxford100K: 100K images from Flickr by searching the 145 most popular tags
• Oxford1M: 1M images from Flickr by searching the 450 most popular tags
• Challenges: substantial variations in scale, viewpoint and lighting conditions of the images and the objects
Unsupervised Codebook Constr.
Unsupervised PCTs

- The descriptive space is simultaneously used as a target space
Large Scale Object Retrieval: Performance & Scalability

Comparison of the retrieval performance (given as mean average precision)

<table>
<thead>
<tr>
<th>Image dataset</th>
<th>Without Spatial re-ranking</th>
<th>With Spatial re-ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AKM</td>
<td>ExtraTrees</td>
</tr>
<tr>
<td>Oxford5K</td>
<td>0.680</td>
<td>0.675</td>
</tr>
<tr>
<td>Paris</td>
<td>0.687</td>
<td>0.661</td>
</tr>
<tr>
<td>Pythia</td>
<td>0.164</td>
<td>0.172</td>
</tr>
</tbody>
</table>

- Spatial re-ranking of a short-list of top ranked results to further boost the retrieval performance
- Better results with larger codebooks and when considering more descriptors
- The retrieval performance of our method is better than the one of both approximate k-means and ensembles of extremely randomized trees
- We are also more efficient than the competition
  - 24.5 times faster than k-means
  - 1.6 times faster than AKM
Supervised Codebook Construction

1. Train images
   - Sampling strategy + SIFT descriptors
   - Set of SIFT descriptors
   - PCTs for MLC
   - Visual codebook
   - Set of histograms
   - visual codebook construction part
   - classification part
   - Learning algorithm

2. Test images
   - Sampling strategy + SIFT descriptors
   - Set of SIFT descriptors
   - Set of histograms
   - SVM classifier
   - Annotations/labels
Supervised Codebook Construction

Art: Mona Lisa

Transport: Bicycle

Musical Instrument: Violin
Multi-Label Image Annotation

<table>
<thead>
<tr>
<th>Image database</th>
<th>Efficiency [s]</th>
<th>Performance [MAP]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$-Means</td>
<td>PCTs for MLC</td>
</tr>
<tr>
<td>PASCAL VOC 2007</td>
<td>12334.820</td>
<td>456.114</td>
</tr>
<tr>
<td>ImageCLEF@ICPR 2010</td>
<td>11977.230</td>
<td>466.829</td>
</tr>
<tr>
<td>ImageCLEF 2010</td>
<td>11209.750</td>
<td>544.740</td>
</tr>
</tbody>
</table>

- The visual codebook constructed with random forests of PCTs for MLC outperforms the one constructed with $k$-means on all three databases: It is more discriminative.
- The improvement is larger for the databases with a larger average number of labels per image.
- Dimitrovski et al., Pattern Recognition Letters 2013
Codebooks Learnt from more KPs have Better Performance

- PCTs for MLC are \(~40\) times more efficient than k-means
- Codebooks using larger number of key-points can be constructed
- Codebooks of 4000 words, diff. no. of KPs, diff. no. of trees in forest
Acknowledgements and Announcement

We acknowledge European Commission support through the grants

• MAESTRA: Learning from Massive, Incompletely annotated, and Structured Data, grant 612944
• HBP SGA1: The Human Brain Project, grant 720270
• LANDMARK: LAND Management: Assessment, Research, Knowledge base, grant 635201

As well as the Slovenian Research Agency through

• P2-0103 Knowledge technologies
• L2-7509 Structured output prediction ...

And announce ...
Thank you ...

• For your attention.
• Questions welcome!