Leverage Loosely–tagged Images and Inter–Object Correlations for Tag Recommendation

Yi Shen, Jianping Fan
UNC–Charlotte, USA

Presentation Outlines
- Research Motivations
- Instance Label Identification
- Object Correlation Network
- Multi–Task Multi–Label Structured SVM Learning
- Experimental Results

Research Motivations: Harnessing Large–Scale Loosely Tagged Images
- Loosely–tagged images:
  - Multiple object tags are loosely given at the image level rather than at the object level with specific object locations and areas, ….

Why we need Loosely–tagged Images
- They are large and thus they can represent various visual properties of the object classes more sufficiently.
- They can be obtained easily by leveraging the collaborative efforts of large numbers of Internet users
- Both their tags and their visual properties are diverse, which gives a real–world point of departure for object detection and scene recognition

Goals of Our Research
- Leverage both the loosely tagged images and the inter–object correlations for object classifier training
  - Transform each loosely–tagged image into bag of instances (image regions) and express its semantic ambiguity (multiple tags) explicitly in the instance space
  - Identify the instance labels automatically
  - Structure learning for exploiting the inter–object correlations to achieve more effective learning of a large number of inter–related object classifiers

Proposed Solutions
- Instance Label Identification for Multiple Instance Learning
- Object Correlation Network for Determining the Inter–related Learning Tasks
- Multi–Task Multi–Label Structured Learning
**Key Components of our System**

- Each loosely-tagged is first partitioned into a set of regions and multiple segmentations are integrated for obtaining more meaningful regions.
- Each image region is treated as one instance
- Multi-modal visual features are extracted from each image instance to characterize its various visual properties more sufficiently

**Instance Label Identification**

- **Image Instance Clustering**
  - Instance Clustering via Affinity Propagation
  - Identify Relevant instances and irrelevant instances
    - "relevant" means strongly related with the given tag and their semantics can be interpreted precisely by the given tag
- **Relevant Cluster Identification**

**Relevant Cluster Identification**

- **Inter–Cluster Visual Correlation**

**Bags of Instances**

- Each loosely-tagged is first partitioned into a set of regions and multiple segmentations are integrated for obtaining more meaningful regions.
- Each image region is treated as one instance
- Multi-modal visual features are extracted from each image instance to characterize its various visual properties more sufficiently

**Relevant Cluster Identification**

- **Inter–cluster visual similarity**

$$\hat{d}(G_i, G_j) = \frac{1}{|G_i| \cdot |G_j|} \sum_{(u, v) \in G_i} [\hat{d}(u, v) + \hat{d}(v, u)]$$

- The best–matched cluster pair between the positive bags and the negative bags

$$\hat{d}(G_i, G_j) = \max \{\hat{d}(G_i, G_j) | G_i \in \Omega, G_j \in \Omega\}$$

**Assumption:** For a given tag, its relevant clusters in the positive bags should be far away from the instance clusters in the negative bags.
**Positive Clusters vs. Negative Clusters**

- Positive Clusters
- Negative Clusters

**Object Correlation Network**

- Object Correlation Network
  - Nodes: Object classes
  - Edges: Inter-object correlations
- Inter-object Correlations
  - Inter-object Visual Similarity Context
  - Object Co-occurrence Correlation

**Inter-object Correlations**

- Inter-object Visual Similarity Context
  
  \[ \gamma(C_i, C_j) = \frac{1}{2|C_i||C_j|} \sum_{a \in C_i} \sum_{v \in C_j} \frac{d(a, v) + d(a, v)}{d(a, v) + d(a, v)} \]

- Object Co-occurrence Correlation
  
  \[ \rho(C_i, C_j) = -P(C_i) \log \frac{P(C_i | C_j)}{P(C_i) + P(C_j)} \]

- Inter-object Correlations
  
  \[ \phi(C_i, C_j) = \eta \cdot \gamma(C_i, C_j) + (1 - \eta) \cdot \rho(C_i, C_j) \]
Inter-object Correlations

- **Inter-Object Correlation**
  
<table>
<thead>
<tr>
<th>object pair</th>
<th>φ</th>
<th>object pair</th>
<th>φ</th>
<th>object pair</th>
<th>φ</th>
<th>object pair</th>
<th>φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>grass-building</td>
<td>0.66</td>
<td>knee-row</td>
<td>0.18</td>
<td>cat-dog</td>
<td>0.36</td>
<td>cat-bicycle</td>
<td>0.53</td>
</tr>
<tr>
<td>bird-antelope</td>
<td>0.56</td>
<td>eagle-antelope</td>
<td>0.25</td>
<td>blue-grass</td>
<td>0.01</td>
<td>sky-row</td>
<td>0.51</td>
</tr>
<tr>
<td>sign-building</td>
<td>0.64</td>
<td>selfphone</td>
<td>0.19</td>
<td>body-dog</td>
<td>0.03</td>
<td>body-cat</td>
<td>0.79</td>
</tr>
<tr>
<td>road-building</td>
<td>0.72</td>
<td>road-cat</td>
<td>0.10</td>
<td>blue-antelope</td>
<td>0.13</td>
<td>mountain-car</td>
<td>0.02</td>
</tr>
<tr>
<td>cat-grass</td>
<td>0.99</td>
<td>trunk-cat</td>
<td>0.12</td>
<td>body-dog</td>
<td>0.08</td>
<td>sky-book</td>
<td>0.01</td>
</tr>
<tr>
<td>boat-sky</td>
<td>0.89</td>
<td>antelope-mountain</td>
<td>0.03</td>
<td>mountain-water</td>
<td>0.71</td>
<td>body-sky</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Object Correlation Network

- It can characterize the inter-object correlations explicitly and provide a good environment to identify the inter-related learning tasks **directly in the feature space**.
- It can provide a good environment to integrate the training instances from multiple inter-related object classes for training their classifiers jointly and bring more powerful inference schemes to enhance their discrimination and adaptation power.

Structured Learning

- For a given object class, its first-order nearest neighbors on the object correlation network are strongly correlated and their training instances may share similar visual properties.

Multi-Task Multi-Label Structured SVM

- **Joint Objective Function**
  
  \[ J = \frac{1}{2} \| W \|^2 + \sum \sum \alpha_i \| V_i \|^2 + \sum \sum \sum \lambda_i (\sum \sum \sum \theta_i \Phi_i (x)) \cdot \sum \sum \sum \theta_i \Phi_i (x) \]

- Subject to
  
  \[ \forall C_i, C_j \in \{ 1 \ldots n \} \cdot \| W_i + W_j \cdot \Phi_i (x) \| > 0 \]
  
  \[ \forall C_i, C_j \in \{ 1 \ldots n \} \cdot \| W_i + W_j \cdot \Phi_i (x) \| > 0 \]

- **Multi-Task Multi-Label Structured SVM**

  \[ f_{C_j} (x) = \sum_{C_i \in \mathcal{T}} g_i (W_i + V_i) \Phi_i (x) \]

- \( W \): common regularization term shared among the classifiers for multiple inter-related object classes centered by \( C \).
- \( V \): individual regularization term for the classifier between the given object class \( C_j \) and its neighbor \( C_t \).

Joint Objective Function

- **Lagrange of joint objective function**
  
  \[ L = J + \sum \sum \sum \alpha_i (W_i + V_i) \Phi_i (x) + \sum \sum \Gamma_i \]

- **Partial Difference of Lagrange**
  
  \[ \frac{\partial L}{\partial W} = \sum \sum \sum \sum \alpha_i (W_i + V_i) \Phi_i (x) + \sum \sum \sum \sum \Gamma_i \]

  \[ \frac{\partial L}{\partial V} = \lambda \sum \sum \sum \sum \alpha_i (W_i + V_i) \Phi_i (x) + \sum \sum \sum \sum \Gamma_i \]

  \[ \frac{\partial L}{\partial \theta} = \lambda \sum \sum \sum \sum \alpha_i (W_i + V_i) \Phi_i (x) + \sum \sum \sum \sum \Gamma_i \]

  \[ \frac{\partial L}{\partial \alpha_i} = \lambda \sum \sum \sum \sum \alpha_i (W_i + V_i) \Phi_i (x) + \sum \sum \sum \sum \Gamma_i \]
Dual Problem of Objective Function

- Dual form of multi-task structured SVM

\[
\begin{align*}
\min_{\mathbf{W}, \mathbf{b}, \mathbf{g}} & \quad \frac{1}{2} \mathbf{W}^T \mathbf{W} + \mathbf{b}^T \mathbf{b} + \mathbf{g}^T \mathbf{g} \\
\text{subject to} & \quad \sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{A}_{ij} \mathbf{x}_{ij} = \mathbf{y}_i, \quad \forall i = 1, \ldots, N \\
& \quad -\sum_{i=1}^{N} \sum_{j=1}^{M} \mathbf{A}_{ij} \mathbf{x}_{ij} \leq \mathbf{b}, \quad \forall j = 1, \ldots, M
\end{align*}
\]

Experiment Setting

- Dataset
  - MSRC dataset
    - 523 images, 3184 image regions, 23 object classes
  - Corel 30k dataset
    - 30000 images, 98 object classes

- Reference Method
  - Structured SVM
  - MILES
  - Multi-Label MIL

Performance Measurement

- Symbols
  - TP: True Positive
  - TN: True Negative
  - FP: False Positive
  - FN: False Negative

- ROC curve
  - True Positive Rate
    - \( \text{TPR} = \frac{TP}{TP + FN} \)
  - False Positive Rate
    - \( \text{FPR} = \frac{FP}{TN + FP} \)

- AUC Score
  - Area under ROC curve

Multi-Label Image Annotation Results
Experiment Result

- Multi-Label Image Annotation Results

![Image](image1.jpg)

ROC Curve on MSRC dataset

![ROC Curve](roc_curve.png)

AUC Score on Corel 30k dataset

- Numerical Experimental Results

![AUC Score](auc_score.png)

AUC Score on Corel 30k dataset

- Numerical Experimental Results

![Average AUC Score](average_auc_score.png)
Conclusions

- A multi-task multi-label structured SVM algorithm is developed to leverage both the inter-object correlations and the loosely-tagged images for achieving more effective training of a large number of inter-related object classifiers.
- Our experimental results on a large number of object classes have provided very positive results.