Learning from Candidate Labeling Sets

Luo Jie and Francesco Orabona

- A **semi-supervised** learning framework that descends from Multiple Label Learning and Multiple Instance Learning
- Encoding **prior-information** in a principle way
- An ambiguous loss function and a large-margin discrimination formulation
- **Efficient** algorithm based on stochastic sub-gradient descend
Applications

- Learning from weakly supervised data (Images and Captions)

  President Barack Obama (B) and first lady Michelle Obama (M) wave from the steps of Air Force One as they arrive in Prague, Czech Republic.

<table>
<thead>
<tr>
<th></th>
<th>Face_{left}</th>
<th>Face_{right}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z₁</td>
<td>B</td>
<td>M</td>
</tr>
<tr>
<td>Z₂</td>
<td>M</td>
<td>B</td>
</tr>
<tr>
<td>Z₃</td>
<td>B</td>
<td>null</td>
</tr>
<tr>
<td>Z₄</td>
<td>null</td>
<td>B</td>
</tr>
<tr>
<td>Z₅</td>
<td>M</td>
<td>null</td>
</tr>
<tr>
<td>Z₆</td>
<td>null</td>
<td>M</td>
</tr>
</tbody>
</table>

  null: not shown in the caption

- Learning from multiple annotators

- Clustering with constraints
We introduce an ambiguous loss which upper bounds the true loss in expectation (by a proportionality factor)

\[ \ell_{\text{max}}(\mathcal{X}, \mathcal{Z}; w) = \max_{\tilde{z} \notin \mathcal{Z}} \left( \ell^A_\Delta (\tilde{z}, \mathcal{Z}) + F(\mathcal{X}, \tilde{z}; w) \right) - \max_{z \in \mathcal{Z}} F(\mathcal{X}, z; w) + \right. \\
\]

where \( \ell^A_\Delta (z, \mathcal{Z}) = \min_{z' \in \mathcal{Z}} \ell_\Delta (z, z') \)

Every instance \( x_m \) from a bag \( \mathcal{X} \) is considered jointly at prediction, the prior knowledge between instances is encoded in the labeling vector \( y \)

\[ F(\mathcal{X}, y; w) = \sum_{m=1}^{M} F(x_m, y_m) = \sum_{m=1}^{M} w \cdot \phi(x_m) \otimes \psi(y_m) \]

We propose a large margin optimization problem based on the ambiguous loss and solve the problem using the \textbf{Constrained Concave-Convex Procedure (CCCP)}

Solve the CCCP-relaxed optimization problem using a stochastic sub-gradient descent algorithm based on the \textbf{Pegasos} framework
Recognition accuracy on Yahoo! News Dataset

- Performance comparable to fully-supervised SVM algorithms (the 1st and 2nd columns)
- Outperform other weakly-labeled learning baselines (the 3rd and 4th columns)