

Learning from Candidate Labeling Sets



Luo Jie and Francesco Orabona

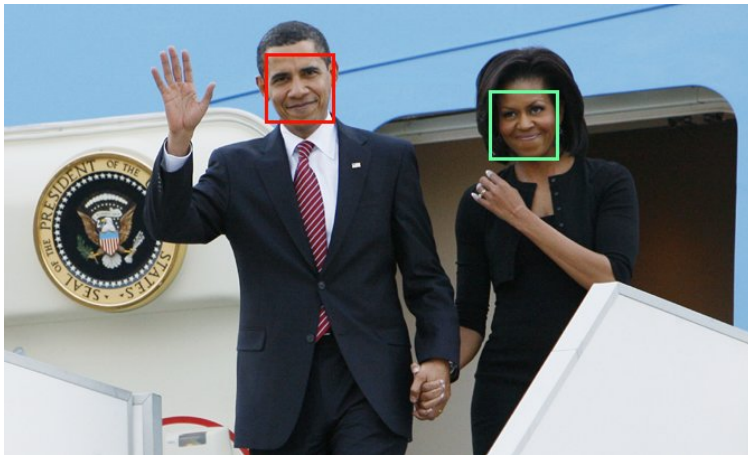
POSTER M85

- 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.

	Face _{left}	Face _{right}
Z ₁ :	B	M
Z ₂ :	M	B
Z ₃ :	B	null
Z ₄ :	null	B
Z ₅ :	M	null
Z ₆ :	null	M

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)

$$\ell_{\max}(\mathcal{X}, \mathcal{Z}; \mathbf{w}) = \left| \max_{\bar{z} \notin \mathcal{Z}} (\ell_{\Delta}^A(\bar{z}, \mathcal{Z}) + \mathbf{F}(\mathcal{X}, \bar{z}; \mathbf{w})) - \max_{z \in \mathcal{Z}} \mathbf{F}(\mathcal{X}, z; \mathbf{w}) \right|_+$$

$$\text{where } \ell_{\Delta}^A(z, \mathcal{Z}) = \min_{z' \in \mathcal{Z}} \ell_{\Delta}(z, z')$$



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 \mathbf{y}

$$\mathbf{F}(\mathcal{X}, \mathbf{y}; \mathbf{w}) = \sum_{m=1}^M F(\mathbf{x}_m, y_m) = \sum_{m=1}^M \mathbf{w} \cdot \phi(\mathbf{x}_m) \otimes \psi(y_m)$$

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

Solve the CCCP-relaxed optimization problem using a stochastic sub-gradient descent algorithm based on the **Pegasos** framework

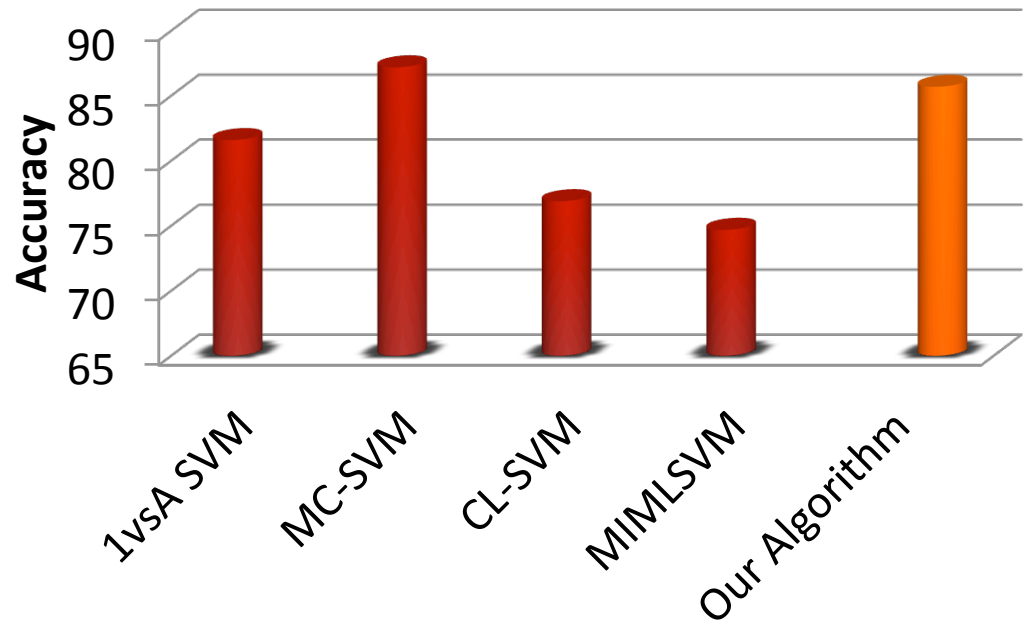
Poster Location: M85

Code
Available

Yahoo! News dataset
[Berg, et.al. NIPS04]



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)