Motivations

• Samples from different sources can be interpreted as corresponding to different views of the same target.
• Learning in the multi-view setting is challenging:
  – Information Fusion
  – Parsimony vs. Robustness
  – Unlabeled samples
• In this work, we proposed a stochastic consensus-based multi-view learning framework [7] to handle these issues.

Comparison to previous work

• Comparison of conventional fusion methods and the proposed consensus-based multi-view learning method (in bold)

<table>
<thead>
<tr>
<th>fusion stage</th>
<th>parsimony</th>
<th>semi-sup.</th>
<th>noise tol</th>
<th>Bayes</th>
<th>views</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCA feature</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>Bi-DAE feature</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>SVM-2K feature</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>Bayesian-Fusion decision</td>
<td>√</td>
<td>x</td>
<td>√</td>
<td>√</td>
<td>≥2</td>
</tr>
<tr>
<td>Boosting decision</td>
<td>√</td>
<td>x</td>
<td>√</td>
<td>x</td>
<td>≥2</td>
</tr>
<tr>
<td>Co-training consensus</td>
<td>√</td>
<td>x</td>
<td>√</td>
<td>x</td>
<td>2</td>
</tr>
<tr>
<td>Bayesian Co-trm consensus</td>
<td>x</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>≥2</td>
</tr>
<tr>
<td>CMV-MED consensus</td>
<td>√</td>
<td>x</td>
<td>√</td>
<td>x</td>
<td>≥2</td>
</tr>
</tbody>
</table>

Our Contributions

1. We learn view-specific posterior distributions as features.
2. The proposed method maximizes the stochastic agreement among different models on unlabeled samples.
3. The proposed information-theoretical consensus measure is robust to noisy samples and outliers.

Notations and Model

• Multiview observation model:

\[ X_1 \times \ldots \times X_{\nu} \times Y; \]

Labeled i.i.d. samples \((x_i, y_i), i \in U: \) labeled set.

Unlabeled i.i.d. samples \(x_i, i \in U: \) unlabeled set.

view-specific model

\[ \log p(y_i | x_i, \theta) \propto \log p(y_i | x_i, \theta) \in M_{\nu}, \theta \equiv \theta(x_i) \in \Theta_i, 1 \leq i \leq V \]

view-specific loss functional

\[ L_{\nu} = \sum_{i=1}^{V} \log p(y_i | x_i, \theta) \in M_{\nu}, \theta \equiv \theta(x_i) \in \Theta_i, 1 \leq i \leq V \]

• Consensus part:

<table>
<thead>
<tr>
<th>Consensus-view model</th>
<th>Parameters of CV model</th>
<th>Pairwise consensus measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>( q(y_{i,a}</td>
<td>\theta_{c}) \in M = \theta_{i,c}, a \in U_c )</td>
<td>( \theta_0 = \theta_0(x_{i,c}), \ldots, \theta_{n_c} )</td>
</tr>
</tbody>
</table>

Variance EM Algorithm

For \( \ell = 1, \ldots \) until convergence

1. E-step: Given \( \tilde{y}_{i,j} = E_{\nu_j}(\cdot | y, \theta) \), \( 1 \leq j \leq V \), the consensus model is the log-average \( \log q(y | \nu_j) = \sum_{i} \log p_i(y | \theta_i) \rightarrow \log Z(y | \nu_j), i \in U \).

2. M-step: Given \( q(y | \theta), \xi \), solve for \( q(\theta, \xi) \) independently. In particular, each view corresponds to a maximum entropy learning problem with maximal margin error constraint \( \xi \). Under proper prior on soft-threshold \( \xi \), it can be solved efficiently using SVM-like dual optimization. See e.g. [4].

Experiments

We compare the proposed CMV-MED model with the SVM-2K [3], the MV-MED [6] as well as the conventional MED for each view on two real multi-view data sets: ARL-footstep[2] and WebKB[4]

Acknowledgements

This research was partially supported by US Army Research Office (ARO) grants W911NF-11-1-0391 and WA11NF-11-1-103A1.

References