Bayesian Efficient Multiple Kernel Learning

Mehmet Gonen  
mehmet.gonen@aalto.fi  
http://users.ics.aalto.fi/gonen

Helsinki Institute for Information Technology HIIT, Department of Information and Computer Science, Aalto University

**Motivation**

- To obtain a better similarity measure and to integrate information from different sources
- To develop a computationally feasible Bayesian MKL algorithm without sampling

**Existing Bayesian Methods**

- Nonconjugacy between Dirichlet and normal distributions requires a sampling method
- Nonlinear dependency between random variables when calculating prediction scores

**Proposed Method**

- Combination is formulated in a novel way
- Intermediate outputs are introduced as auxiliary variables
  
  - Kernel weights are assumed to be normally distributed without any constraints
  - Sample- and kernel-level sparsities can be adjusted using gamma priors on precisions

**Graphical Model**

**Probabilistic Model**

\[
q(f) = \prod_{i=1}^{N} \mathcal{N}(f_i; e_i \cdot \tilde{g}_i + \tilde{\omega}_i, 1, f_i y_i > \nu)
\]

**Large-Margin Learning**

- \( \nu > 0 \) corresponds to placing a margin between two classes

**Extensions**

- Multiclass learning is done by sharing kernel weights in one-versus-all classification
- Semi-supervised learning using truncated normals is left for future research

**Illustration on a Toy Data Set**

- 8 benchmark data sets from UCI repository
- Inference takes less than a minute with large numbers of kernels, from 91 to 793

**Benchmark Data Sets**

<table>
<thead>
<tr>
<th></th>
<th>sparse</th>
<th>non-sparse</th>
</tr>
</thead>
<tbody>
<tr>
<td>pima</td>
<td>Test Acc.</td>
<td>68.1 ± 1.2</td>
</tr>
<tr>
<td>BEMKL (one-versus-all)</td>
<td>71.5 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>BEMKL (multiclass)</td>
<td>71.2 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>Oxford Flowers102 data set</td>
<td>0.969 ± 0.069</td>
<td></td>
</tr>
</tbody>
</table>

**MKL Data Sets**

- 4 comparison data sets for MKL methods
- Protein fold recognition data set

**Conclusions**

- A Bayesian MKL framework with a novel kernel combination formulation is introduced
- Fully conjugate probabilistic model leads to a very efficient variational approximation
- Matlab implementation is available at http://users.ics.aalto.fi/gonen/bemkl

**References**
