Feature-Cost Sensitive Learning
with Submodular Trees of Classifiers

Matt J. Kusner, Wenlin Chen, Quan Zhou, Zhixiang (Eddie) Xu, Kilian Q. Weinberger, Yixin Chen
Department of Computer Science & Engineering, Washington University in St. Louis, USA
Department of Automation, Tsinghua University, Beijing, China

Budgeted Learning
[Chen et al., 2012; Xu et al., 2013]

\[
\begin{align*}
D &= \{ (x^{(i)}, y^{(i)}) \}_{i=1}^{n} \subset \mathcal{X} \times \mathcal{Y}, \\
\min \ell(\beta, \{ (x^{(i)}, y^{(i)}) \}_{i=1}^{n}) \quad \text{s.t.} \quad c(\beta) \leq B
\end{align*}
\]

Cost-Sensitive Tree of Classifiers
[Xu et al., 2014]

Expected Tree Loss
\[
E[\ell(T)] = \frac{1}{n} \sum_{v \in V} \sum_{i=1}^{n} p_i(y^{(i)} - \beta^T x^{(i)})^2
\]

Expected Tree Cost
\[
E[c(T)] = \sum_{v \in L} \left[ \sum_{a \in \pi} c_a \sum_{v \in \pi} |\beta_a^T|_0 \right]
\]

Objective
\[
\min_{\beta^T \cdots , \beta^T} \quad E[\ell(T)] + \lambda E[c(T)]
\]

Results
- state-of-the-art performance
- expensive global optimization
- requires continuous relaxation
- difficult to implement

Approximately Submodular Tree of Classifiers
Greedy Tree Building

Greedy Feature Selection (node k)
\[
\ell_k(A) = \min_{\beta^T} \frac{1}{n} \sum_{i=1}^{n} p_i(y^{(i)} - \beta^T x^{(i)})^2
\]

New Objective
\[
\max_{A} \ell_k(A) \quad \text{s.t.} \quad c(A) \leq B
\]

Fast Selection via QR-Decomposition
\[
\frac{\ell_k(A) - \ell_k(A \cup a)}{c_a} = \frac{(q^T y)^2}{c_a}
\]

where
\[
q = \frac{X_a - QQ X}{\|X_a - QQ X\|_2}, \quad X_a = QR
\]

Results

![Yahoo! Forest CIFAR MiniBooNE](images)

Figure 3. Plot of ASTC, CSTC, and a cost-sensitive baseline on a real-world feature-cost sensitive dataset (Yahoo!) and three non-cost sensitive datasets (Forest, CIFAR, MiniBooNE). For Yahoo! circles mark the CSTC points that are used for training time comparison, otherwise, all points are compared.

Table 1. Training speed-up of ASTC over CSTC for different cost budgets on all datasets.

<table>
<thead>
<tr>
<th>Cost Budget</th>
<th>10</th>
<th>50</th>
<th>86</th>
<th>169</th>
<th>468</th>
<th>800</th>
<th>1495</th>
<th>3</th>
<th>5</th>
<th>8</th>
<th>13</th>
<th>21</th>
<th>50</th>
<th>9</th>
<th>24</th>
<th>180</th>
<th>239</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASTC</td>
<td>119x</td>
<td>52x</td>
<td>41x</td>
<td>21x</td>
<td>15x</td>
<td>9.2x</td>
<td>6.6x</td>
<td>8.4x</td>
<td>7.0x</td>
<td>6.3x</td>
<td>4.9x</td>
<td>3.1x</td>
<td>1.4x</td>
<td>3.6x</td>
<td>2.3x</td>
<td>0.68x</td>
<td>0.23x</td>
<td>0.14x</td>
</tr>
<tr>
<td>ASTC, SOFT</td>
<td>121x</td>
<td>48x</td>
<td>46x</td>
<td>18x</td>
<td>15x</td>
<td>8.2x</td>
<td>6.4x</td>
<td>8.0x</td>
<td>6.4x</td>
<td>5.7x</td>
<td>4.5x</td>
<td>2.8x</td>
<td>1.5x</td>
<td>5.3x</td>
<td>2.3x</td>
<td>0.62x</td>
<td>0.27x</td>
<td>0.13x</td>
</tr>
</tbody>
</table>

References