Supplementary material for
Multi-class Classification from Multiple Unlabeled Datasets with Partial Risk Regularization

Yuting Tang
Tang@ms.k.u-tokyo.ac.jp
Nan Lu
lu@ms.k.u-tokyo.ac.jp
Tianyi Zhang
zhang@ms.k.u-tokyo.ac.jp
The University of Tokyo/RIKEN, Japan.

Masashi Sugiyama
sugi@k.u-tokyo.ac.jp
RIKEN/The University of Tokyo, Japan.

Editors: Emtiyaz Khan and Mehmet Gönen

1. Experiments on large-scale datasets

We further tested our method on Kuzushiji-49 dataset\(^1\), which has 49 classes (28x28 grayscale, 270,912 images) of Hiragana characters, and SVHN dataset\(^2\), which has 10 classes (32x32 RGB, 531,131 images) of printed digits cropped from house number plate photos. A 5-layer fully-connected network is used for the Kuzushiji-49 dataset, and ResNet-20 is used for the SVHN dataset. We ran each experiment five times, and we trained the model for 100 epochs on all datasets. Other experimental setups followed Section 5.1 in the paper. The experiments on the symmetric class-prior matrix are reported in Table 1.

Table 1: Experimental results on large-scale datasets. Means (standard deviations) of the classification error (Err) and the drop (\(\Delta E\)) over five trials in percentage. The best and comparable methods based on the paired t-test at the significance level 5% are highlighted in boldface.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>(a,b)</th>
<th>Biased</th>
<th>Prop</th>
<th>Prop-CR</th>
<th>Unbiased</th>
<th>U-correct</th>
<th>U-flood</th>
<th>U-PRR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a,b)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
<td>Err(\Delta E)</td>
</tr>
<tr>
<td>Kuzushiji-49</td>
<td>0.51, 0.01</td>
<td>54.59 (0.52)</td>
<td>30.11 (0.46)</td>
<td>45.76 (0.89)</td>
<td>1.05 (0.84)</td>
<td>38.90 (0.80)</td>
<td>0.12 (0.17)</td>
<td>53.27 (0.45)</td>
</tr>
<tr>
<td>SVHN</td>
<td>0.5, 0.05</td>
<td>12.79 (2.40)</td>
<td>3.47 (2.12)</td>
<td>27.75 (2.37)</td>
<td>0.28 (0.44)</td>
<td>22.58 (1.62)</td>
<td>0.01 (0.01)</td>
<td>88.22 (6.45)</td>
</tr>
</tbody>
</table>

The experimental results show the proposed U-PRR method outperforms baselines and successfully mitigates the overfitting of the Unbiased method.

1. http://codh.rois.ac.jp/kmnist/