Towards Robust Learning with Different Label Noise Distributions

Diego Ortego, Eric Arazo, Paul Albert, Noel E. O’Connor and Kevin McGuinness

Code on github https://git.io/JJ0PV
Exploiting large data collections

- DNN training with more data → Better results
- Labeling large data collections is expensive
- Possible solution: collect web images and infer the labels from the metadata
Label noise

Studied in a controlled manner by flipping labels to incorrect classes (synthetic in-distribution (ID) noise)

Methods designed to perform label correction

Label noise

Studied in a controlled manner by flipping labels to incorrect classes (synthetic in-distribution (ID) noise)

Methods designed to perform label correction

But, in real-world noise...
(WebVision [1] examples)

Out-of-distribution (OOD) noise highly present

Important to consider different noise types

Impact of label noise distribution

Small loss trick (low loss=clean) often used is not straightforward to apply for all noise distributions.

- Uniform ID
- Uniform OOD
- Non-uniform ID
- Non-uniform OOD
Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):
1) Label noise detection.
2) Semi-supervised learning (SSL)
Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):
1) Label noise detection.  2) Semi-supervised learning (SSL)

- **Pre-train** with high learning rate to learn clean data pattern without memorizing label noise
Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):
1) Label noise detection.       2) Semi-supervised learning (SSL)

Relabel all samples with network predictions [2]. This stage is key to reveal a discriminative measure for noise detection.

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):
1) Label noise detection. 2) Semi-supervised learning (SSL)

Noise detection based on fitting a Beta mixture model [3] to the discriminative measure and thresholding the posterior distribution

Proposed method

Label noise **Distribution Robust Pseudo-Labeling** (DRPL):
1) Label noise detection.
2) Semi-supervised learning (SSL)

SSL using [4] where clean=labeled and noisy=unlabeled
Repeat noise detection to refine the clean and noisy sets

Proposed method

Label noise **Distribution Robust Pseudo-Labeling (DRPL):**
1) Label noise detection. 2) Semi-supervised learning (SSL)

Final SSL [4] training with the refined clean/noisy sets

Discriminative measure during relabeling

Cross-entropy (CE) between current predictions and old labels

Disagreement between the new and old noise patterns

Clean (Low CE)
Keep the same labels in both patterns

Noisy (High CE)
Exhibit different labels in both patterns
Evaluation (ImageNet32/64, CIFAR-10/100, mini-WebVision)

Uniform (U) and non-uniform (NU) for ID and OOD noise

<table>
<thead>
<tr>
<th></th>
<th>ImageNet32</th>
<th></th>
<th>ImageNet64</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NU-ID</td>
<td>U-ID</td>
<td>NU-OOD</td>
</tr>
<tr>
<td>30% 50% 40% 80%</td>
<td>30% 50% 40%</td>
<td>30% 50% 40% 80%</td>
<td>30% 50% 40%</td>
</tr>
<tr>
<td>FW [24]</td>
<td>54.22</td>
<td>43.38</td>
<td>52.06</td>
</tr>
<tr>
<td>R [14]</td>
<td>67.24</td>
<td>63.62</td>
<td>62.98</td>
</tr>
<tr>
<td>M [40]</td>
<td>67.14</td>
<td>51.96</td>
<td>61.98</td>
</tr>
<tr>
<td>DB [16]</td>
<td>62.88</td>
<td>52.20</td>
<td>67.62</td>
</tr>
<tr>
<td>DRPL</td>
<td>73.46</td>
<td>68.18</td>
<td>73.48</td>
</tr>
</tbody>
</table>

mini-WebVision

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73.88</td>
<td>74.68</td>
<td>76.52</td>
<td>80.76</td>
<td>74.28</td>
<td>79.68</td>
<td>73.96</td>
<td>79.96</td>
<td>78.16</td>
<td>82.08</td>
</tr>
</tbody>
</table>

✔ State-of-the-art results
✔ Consistent across noise levels, distributions, and image resolutions
Real-world (RW) vs synthetic (SY)

**Take home message:**
Most SY noises are different from RW noise
RW noise contains OOD samples
OOD samples should be treated differently (mixup success)
Representation learning

ImageNet64

- M: Mixup, O: Ours (DRPL)
- Linear model evaluation for features at different depths
- Degradation concentrates at the end

Take home message: Even if label noise is memorized, discriminative low-level and mid-level features emerge
Where DNNs look for memorizing?

Activation maps

Noisy label: Keyboard
True label: Theater curtain

Take home message: DNNs skip relevant areas for the true class, while focusing on areas that help explaining the noisy label
Conclusions

• DRPL: robust image classification models in the presence of label noise

• Robustness comes from an effective label noise detection for different noise distributions

• We analyze different label noise distributions from multiple perspectives (similarities with RW noise, representation learning, attention maps...) leading to important conclusions that help in better understanding label noise effect
Thank you!

Q/A?