MaxDropout: Deep Neural Network Regularization Based on Maximum Output Values

Claudio F. G. Santos, Mateus Roder, João Paulo Papa

Universidade Federal de São Carlos, Universidade Estadual Paulista “Júlio de Mesquita Filho”

I – Introduction

Different techniques have emerged in the deep learning scenario, such as Convolutional Neural Networks, Deep Belief Networks, and Long Short-Term Memory Networks, to cite a few. In lockstep, regularization methods, which aim to prevent overfitting by penalizing the weight connections, or turning off some units, have been widely studied either. In this paper, we present a novel approach called MaxDropout, a regularizer for deep neural network models that works in a supervised fashion by removing (shutting off) the prominent neurons (i.e., most active) in each hidden layer. The model forces fewer activated units to learn more representative information, thus providing sparsity. Regarding the experiments, we show that it is possible to improve existing neural networks and provide better results in neural networks when Dropout is replaced by MaxDropout.

II – Proposed Approach

Figure 1: Simulation using grayscale (a)-(c) and colored images (d)-(f): (a) original grayscale image and its outcomes after (b) Dropout and (c) MaxDropout transformations, respectively, and (d) original colored image and its outcomes after (e) Dropout and (f) MaxDropout transformations, respectively. In all cases, the dropout rate is 50%.

III – Algorithm

Pseudocode for MaxDropout training algorithm.

while training do
  for each layer do
    rate ← U(0, r)
    normTensor ← L2Normalized(Tensor)
    max ← Max(normTensor)
    Kept_idx ← idx_of(maxTensor)
    returnTensor ← Tensor + Kept_idx
  end for
end while

<table>
<thead>
<tr>
<th>Approach</th>
<th>CIFAR-100</th>
<th>CIFAR-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet18</td>
<td>24.50 ± 0.19</td>
<td>5.17 ± 0.18</td>
</tr>
<tr>
<td>ResNet18+RandomErasing</td>
<td>24.03 ± 0.19</td>
<td>4.31 ± 0.07</td>
</tr>
<tr>
<td>ResNet18+Cutout [1]</td>
<td>21.96 ± 0.24</td>
<td>3.99 ± 0.13</td>
</tr>
<tr>
<td>ResNet18+MaxDropout [2]</td>
<td>21.94 ± 0.23</td>
<td>4.63 ± 0.11</td>
</tr>
</tbody>
</table>

Table 1: Results of MaxDropout over the WRN.

<table>
<thead>
<tr>
<th>Regularizer</th>
<th>CIFAR-100</th>
<th>CIFAR-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutout [1]</td>
<td>21.96 ± 0.24</td>
<td>3.99 ± 0.13</td>
</tr>
<tr>
<td>MaxDropout [2]</td>
<td>21.94 ± 0.23</td>
<td>4.63 ± 0.11</td>
</tr>
<tr>
<td>MaxDropout + Cutout [2]</td>
<td>21.82 ± 0.13</td>
<td>3.76 ± 0.08</td>
</tr>
</tbody>
</table>

Table 2: Results of the MaxDropout combined with Cutout.

<table>
<thead>
<tr>
<th>Model</th>
<th>CIFAR-100</th>
<th>CIFAR-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRN [4]</td>
<td>19.25</td>
<td>4.00</td>
</tr>
<tr>
<td>WRN + Dropout [4, 3]</td>
<td>18.85</td>
<td>3.89</td>
</tr>
<tr>
<td>WRN + MaxDropout [2]</td>
<td>18.81</td>
<td>3.84</td>
</tr>
</tbody>
</table>

Table 3: Results of Dropout and MaxDropout over the WRN.

<table>
<thead>
<tr>
<th>MaxDropout Rate (r)</th>
<th>CIFAR-100</th>
<th>CIFAR-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>22.05 ± 0.17</td>
<td>4.76 ± 0.09</td>
</tr>
<tr>
<td>0.10</td>
<td>22.06 ± 0.32</td>
<td>4.71 ± 0.09</td>
</tr>
<tr>
<td>0.15</td>
<td>22.16 ± 0.20</td>
<td>4.63 ± 0.11</td>
</tr>
<tr>
<td>0.20</td>
<td>21.99 ± 0.21</td>
<td>4.70 ± 0.08</td>
</tr>
<tr>
<td>0.25</td>
<td>21.94 ± 0.23</td>
<td>4.70 ± 0.06</td>
</tr>
<tr>
<td>0.30</td>
<td>22.08 ± 0.24</td>
<td>4.67 ± 0.12</td>
</tr>
<tr>
<td>0.35</td>
<td>22.10 ± 0.29</td>
<td>4.71 ± 0.16</td>
</tr>
<tr>
<td>0.40</td>
<td>22.17 ± 0.34</td>
<td>4.79 ± 0.20</td>
</tr>
<tr>
<td>0.45</td>
<td>22.31 ± 0.29</td>
<td>4.71 ± 0.11</td>
</tr>
<tr>
<td>0.50</td>
<td>22.33 ± 0.23</td>
<td>4.75 ± 0.10</td>
</tr>
</tbody>
</table>

Table 4: Ablation results concerning MaxDropout over ResNet18.

IV – Results

In this paper, we introduced MaxDropout, an improved version of the original Dropout method. Experiments show that it can be incorporated into existing models, working along with other regularizers, such as Cutout, and can replace the standard Dropout with some accuracy improvement.

V – Conclusion

With relevant results, we intend to conduct a more in-depth investigation to figure out the best drop rates depending on the model and the training data. Moreover, the next step is to re-implement MaxDropout and make it available in other frameworks, like TensorFlow and MXNet, and test in other tasks, such as object detection and image segmentation. Nonetheless, we showed that MaxDropout works very well for image classification tasks. For future works, we intended to perform evaluations in other different tasks such as natural language processing and automatic speech recognition.

Acknowledgments

The authors are grateful to São Paulo Research Foundation - FAPESP (#2013/07375-0, #2014/12236-1, #2017/25908-6), and #2019/07825-1, Brazilian National Council for Scientific and Technological Development - CNPq (#307066/2017-7, and #427968/2018-6), and Petrobras (#2017/00285-6).

References