



MaxDropout: Deep Neural Network Regularization Based on Maximum Output Values

¹Claudio F. G. Santos, ²Mateus Roder, ²João Paulo Papa

¹Universidade Federal de São Carlos, ²Universidade 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. The proposed method was evaluated in image classification, achieving comparable results to existing regularizers, such as Cutout and RandomErasing, also improving the accuracy of neural networks that uses Dropout by replacing the existing layer by MaxDropout.

III – Algorithm

Pseudocode for MaxDropout training algorithm. while training do for each layer do rate $\leftarrow U(0, r)$ *normTensor* \leftarrow *L*2*Normalize*(*Tensor*) $max \leftarrow Max(normTensor)$ $Kept_Idx \leftarrow Idx_Of(normTensor, (1 - rate) * max)$ $returnTensor \leftarrow Tensor * Kept_Idx$ end for

II – Proposed Approach



end while

IV – Results

Approach	CIFAR-100	CIFAR-10
ResNet18	24.50 ± 0.19	5.17 ± 0.18
$ResNet18{+}RandomErasing$	24.03 ± 0.19	4.31 ± 0.07
ResNet18+Cutout [1]	21.96 ± 0.24	$\textbf{3.99} \pm \textbf{0.13}$
ResNet18+MaxDropout [2]	$\textbf{21.94} \pm \textbf{0.23}$	4.63 ± 0.11

Table 1: Results of MaxDropout and other regularizers

Regularizer	CIFAR-100	CIFAR-10
Cutout [1]	21.96 ± 0.24	3.99 ± 0.13
MaxDropout [2]	21.94 ± 0.23	4.63 ± 0.11
MaxDropout + Cutout [2]	$\textbf{21.82} \pm \textbf{0.13}$	$\textbf{3.76} \pm \textbf{0.08}$

Table 2: Results of the MaxDropout combined with Cutout.

Model	CIFAR-100	CIFAR-10
WRN [4]	19.25	4.00
WRN + Dropout [4, 3]	18.85	3.89
WRN + MaxDropout [2]	18.81	3.84

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

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%.

References

laxDropout Rate (r)	CIFAR-100	CIFAR-10
05	22.05 ± 0.17	4.76 ± 0.09
10	22.06 ± 0.32	4.71 ± 0.09
15	22.16 ± 0.20	$\textbf{4.63} \pm \textbf{0.11}$
20	21.99 ± 0.21	4.70 ± 0.08
25	$\textbf{21.94} \pm \textbf{0.23}$	4.70 ± 0.06
30	22.08 ± 0.24	4.67 ± 0.12
35	22.10 ± 0.29	4.71 ± 0.16
40	22.17 ± 0.34	4.79 ± 0.20
45	22.31 ± 0.29	4.71 ± 0.11
50	22.33 ± 0.23	4.75 ± 0.10

Table 4: Ablation results concerning MaxDropout over ResNet18.

V – Conclusion

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.

With relevant results, we intend to conduct a more in-depth investigation to figure

T. DeVries and G. W. Taylor.

Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017.

C. F. G. Santos et al.

Maxdropout github repository. https://github.com/cfsantos/MaxDropout-torch, 2020.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov.

Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1):1929–1958, 2014.

S. Zagoruyko and N. Komodakis.

Wide residual networks.

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).