

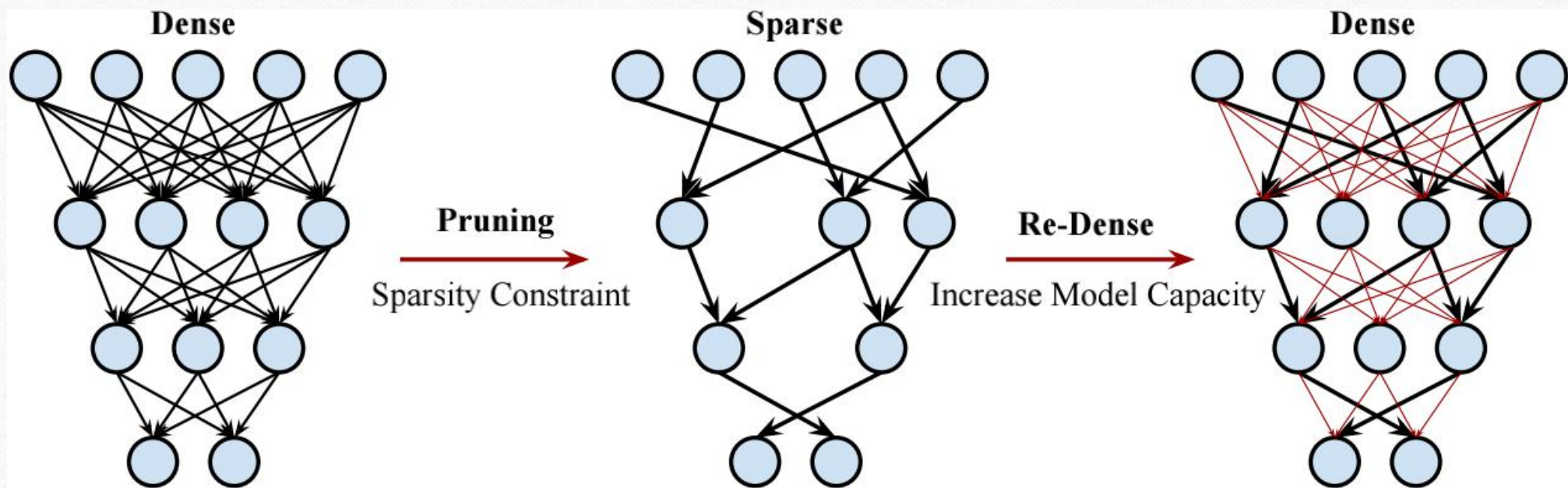
本周工作

学生：屠晓涵

DSD: Dense-Sparse-Dense Training for Deep Neural Networks

- 论文阅读和实验
 - DSD: Dense-Sparse-Dense Training for Deep Neural Networks
 - ICLR顶级会议

Dense-Sparse-Dense 训练



相比于之前的剪枝，多了个恢复连接，效果大大提升

算法过程

Algorithm 1: Workflow of DSD training

Initialization: $W^{(0)}$ with $W^{(0)} \sim N(0, \Sigma)$

Output: $W^{(t)}$.

— *Initial Dense Phase* —

while not converged do

$\tilde{W}^{(t)} = W^{(t-1)} - \eta^{(t)} \nabla f(W^{(t-1)}; x^{(t-1)});$
 $t = t + 1;$

end

— *Sparse Phase* —

// initialize the mask by sorting and keeping the Top-k weights.

$S = \text{sort}(|W^{(t-1)}|); \lambda = S_{k_i}; \text{Mask} = \mathbf{1}(|W^{(t-1)}| > \lambda);$

while not converged do

$\tilde{W}^{(t)} = W^{(t-1)} - \eta^{(t)} \nabla f(W^{(t-1)}; x^{(t-1)});$
 $\tilde{W}^{(t)} = \tilde{W}^{(t)} \cdot \text{Mask};$
 $t = t + 1;$

end

— *Final Dense Phase* —

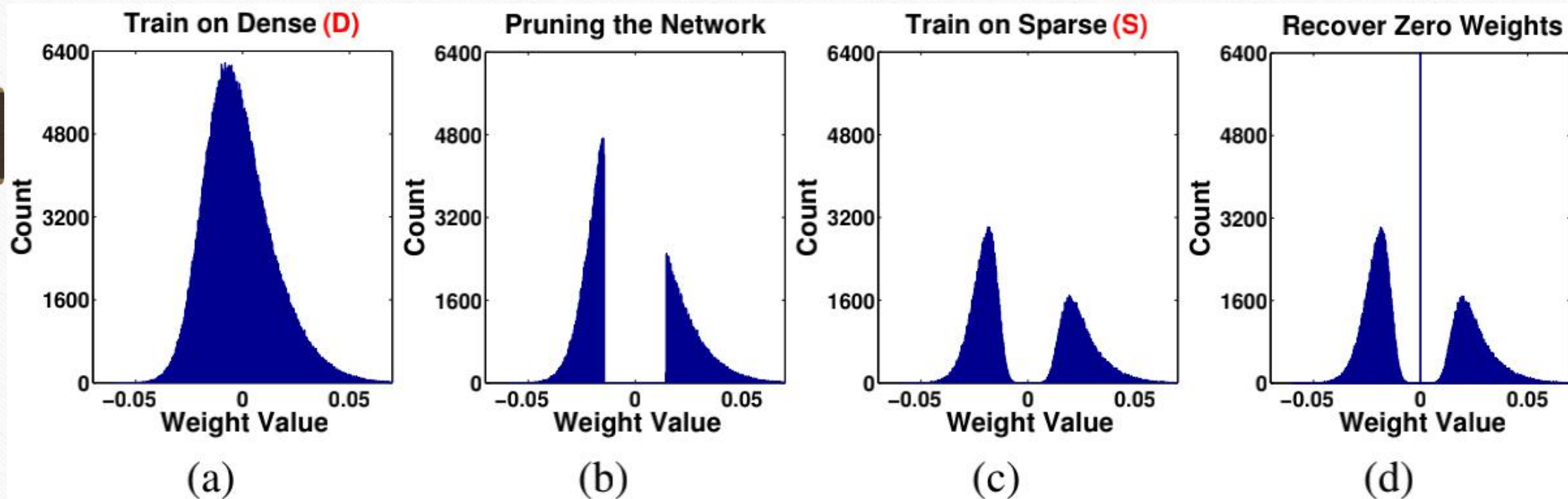
while not converged do

$\tilde{W}^{(t)} = W^{(t-1)} - \eta^{(t)} \nabla f(W^{(t-1)}; x^{(t-1)});$
 $t = t + 1;$

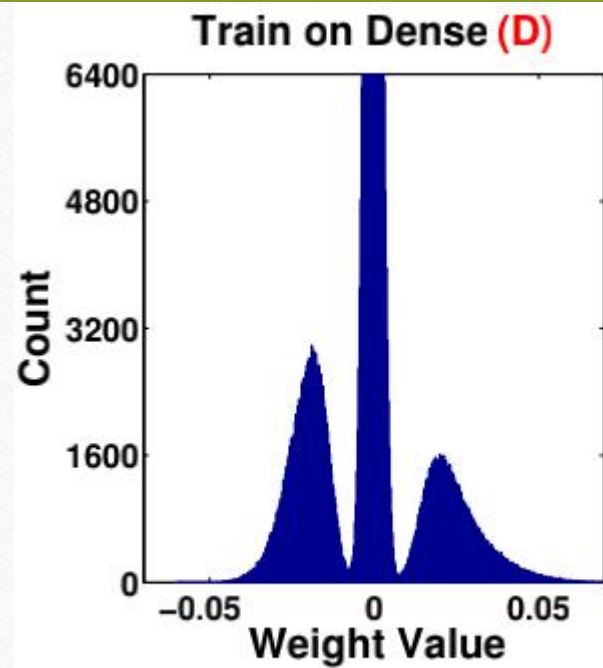
end

goto *Sparse Phase* for iterative DSD;

权重分布



权重分布



(e)

效果1

Neural Network	Domain	Dataset	Type	Baseline	DSD	Abs. Imp.	Rel. Imp.
GoogLeNet	Vision	ImageNet	CNN	31.1% ¹	30.0%	1.1%	3.6%
VGG-16	Vision	ImageNet	CNN	31.5% ¹	27.2%	4.3%	13.7%
ResNet-18	Vision	ImageNet	CNN	30.4% ¹	29.3%	1.1%	3.7%
ResNet-50	Vision	ImageNet	CNN	24.0% ¹	23.2%	0.9%	3.5%
NeuralTalk	Caption	Flickr-8K	LSTM	16.8 ²	18.5	1.7	10.1%
DeepSpeech	Speech	WSJ'93	RNN	33.6% ³	31.6%	2.0%	5.8%
DeepSpeech-2	Speech	WSJ'93	RNN	14.5% ³	13.4%	1.1%	7.4%

效果2

NeuralTalk	BLEU-1	BLEU-2	BLEU-3	BLEU-4	Sparsity
Baseline	57.2	38.6	25.4	16.8	0%
Sparse	58.4	39.7	26.3	17.5	80%
DSD	59.2	40.7	27.4	18.5	0%
Improvement (abs)	2.0	2.1	2.0	1.7	-
Improvement (rel)	3.5%	5.4%	7.9%	10.1%	-

谢谢观看指导！