Dropout as an Implicit Gating Mechanism for Continual Learning

Iman Mirzadeh[1] Mehrdad Farajtabar[2], Hassan Ghasemzadeh[1]

[1]Washington State University
{seyediman.mirzadeh, hassan.ghasemzadeh}@wsu.edu

[2]DeepMind
farajtabar@google.com

June 2020

Agenda

Continual Learning & Catastrophic Forgetting

- Neural networks suffer from the *catastrophic forgetting* problem when they face a sequence of tasks.

Continual Learning & Catastrophic Forgetting

- Neural networks suffer from the *catastrophic forgetting* problem when they face a sequence of tasks.
- They forget older tasks more and more as they learn newer ones.

Continual Learning & Catastrophic Forgetting

- Neural networks suffer from the *catastrophic forgetting* problem when they face a sequence of tasks.
- They forget older tasks more and more as they learn newer ones.

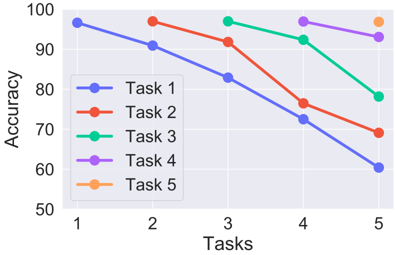
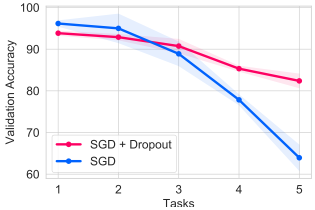


Figure: Catastrophic Forgetting in Continual Learning

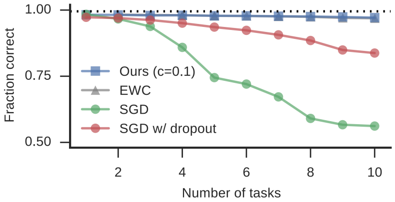
Motivation

Observation

Networks trained with dropout tend to forget at a slower rate.



(a) Our Paper

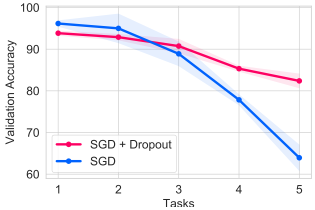


(b) Zenke et al., 2017

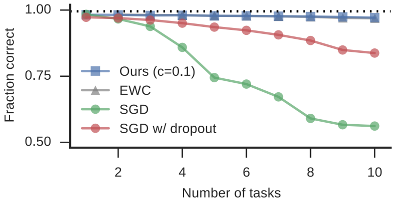
Motivation

Observation

Networks trained with dropout tend to forget at a slower rate.



(a) Our Paper



(b) Zenke et al., 2017

Question

But why?

Dropout and Gating (1)

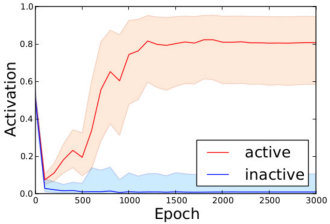
Our Claim

Training with dropout and learning rate decay implicitly creates a gating behavior in network such that for different tasks, different paths of the network are active.

Dropout and Gating (1)

Our Claim

Training with dropout and learning rate decay implicitly creates a gating behavior in network such that for different tasks, different paths of the network are active.



(a) The three phases of learning. For a particular input, a typical active neuron (red) starts out with low variance, experiences a large increase in variance during learning, and eventually settles to some steady constant value. In contrast, a typical inactive neuron (blue) quickly learns to stay silent.

Figure: From Baldi and Sadowski, 2013

Dropout and Gating (2)

1. Dropout regularization helps to create gates in the network by pushing the neurons to be either highly active or highly inactive during the learning experience.

Dropout and Gating (2)

1. Dropout regularization helps to create gates in the network by pushing the neurons to be either highly active or highly inactive during the learning experience.
2. When facing new tasks, the regularization mechanism will change the semi-active neurons more compared to active or inactive neurons, which helps to preserve the task-specific pathways when learning subsequent tasks.

Dropout and Gating (2)

1. Dropout regularization helps to create gates in the network by pushing the neurons to be either highly active or highly inactive during the learning experience.
2. When facing new tasks, the regularization mechanism will change the semi-active neurons more compared to active or inactive neurons, which helps to preserve the task-specific pathways when learning subsequent tasks.
3. The learning rate decay, also helps preserving gates throughout the continual learning experience.

Results (1): Dropout and Gating

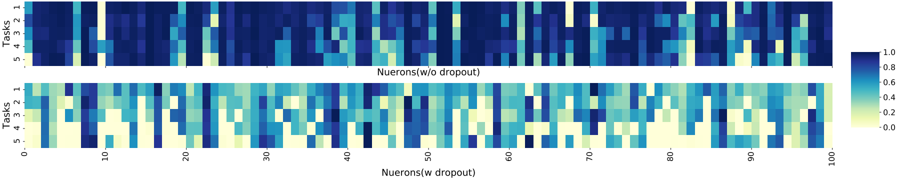


Figure: The effect of dropout on the activation(firing) pattern of neurons

Results (1): Dropout and Gating

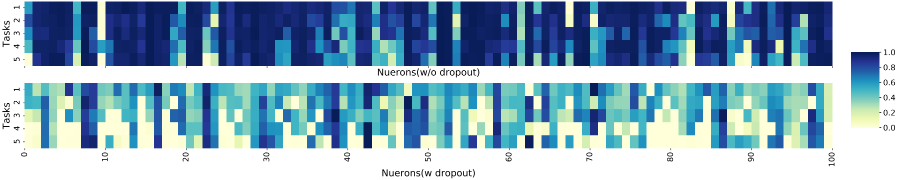


Figure: The effect of dropout on the activation(firing) pattern of neurons

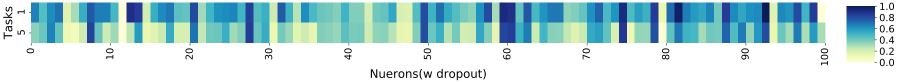


Figure: Consistency between activation patterns of neurons for task 1, after learning task 1 and task 5

Results (2): Dropout Increases the Network Stability

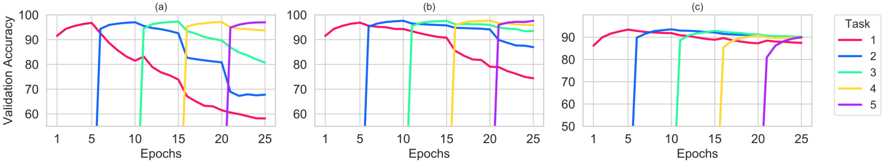


Figure: Increasing the stability and reducing the plasticity from left to right by increasing the the dropout rate and learning rate decay.

Comparison with Other Methods (20 Tasks)

Method	Memoryless	Permuted MNIST		Rotated MNIST		Split CIFAR100	
		Accuracy	Forgetting	Accuracy	Forgetting	Accuracy	Forgetting
Naive SGD	✓	44.4 ($\pm$2.46)	0.53 ($\pm$0.03)	46.3 ($\pm$1.37)	0.52 ($\pm$0.01)	40.4 ($\pm$2.83)	0.31 ($\pm$0.02)
EWC	✓	70.7 ($\pm$1.74)	0.23 ($\pm$0.01)	48.5 ($\pm$1.24)	0.48 ($\pm$0.01)	42.7 ($\pm$1.89)	0.28 ($\pm$0.03)
A-GEM	✗	65.7 ($\pm$0.51)	0.29 ($\pm$0.01)	55.3 ($\pm$1.47)	0.42 ($\pm$0.01)	50.7 ($\pm$2.32)	0.19 ($\pm$0.04)
ER-Reservoir	✗	72.4 ($\pm$0.42)	0.16 ($\pm$0.01)	69.2 ($\pm$1.10)	0.21 ($\pm$0.01)	46.9 ($\pm$0.76)	0.21 ($\pm$0.03)
Stable SGD	✓	**80.1 ($\pm$0.51)**	**0.09 ($\pm$0.01)**	**70.8 ($\pm$0.78)**	**0.10 ($\pm$0.02)**	**59.9 ($\pm$1.81)**	**0.08 ($\pm$0.01)**
MTL	N/A	86.5 ($\pm$0.21)	0.0	87.3($\pm$0.47)	0.0	64.8($\pm$0.72)	0.0

Table: Comparison of the average accuracy and forgetting of several methods on three datasets.

Conclusion

- We studied an open question in continual learning: "*Why does dropout networks forget at a slower rate?*"

Conclusion

- We studied an open question in continual learning: "*Why does dropout networks forget at a slower rate?*"
- We showed that one explanation could be the gating mechanism due to the dropout regularization

Conclusion

- We studied an open question in continual learning: "*Why does dropout networks forget at a slower rate?*"
- We showed that one explanation could be the gating mechanism due to the dropout regularization
- To verify our hypothesis, we studied the activation pattern of dropout network compared to non-dropout networks.

Conclusion

- We studied an open question in continual learning: "*Why does dropout networks forget at a slower rate?*"
- We showed that one explanation could be the gating mechanism due to the dropout regularization
- To verify our hypothesis, we studied the activation pattern of dropout network compared to non-dropout networks.

Thanks

ArXiv Link:
https://arxiv.org/abs/2004.11545.pdf

Github Code:
https://github.com/imirzadeh/stable-continual-learning

Feel free to contact me regarding any further questions.

References

Baldi, P., & Sadowski, P. J. (2013). Understanding dropout (C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger, Eds.). In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26*. Curran Associates, Inc. http://papers.nips.cc/paper/4878-understanding-dropout.pdf

Zenke, F., Poole, B., & Ganguli, S. (2017). Continual learning through synaptic intelligence. *Proceedings of machine learning research*, *70*, 3987–3995.