Learn to Grow: A Continual Structure Learning Framework for Overcoming Catastrophic Forgetting

Xilai Li\textsuperscript{1*}, Yingbo Zhou\textsuperscript{2*}, Tianfu Wu\textsuperscript{1}, Richard Socher\textsuperscript{2}, and Caiming Xiong\textsuperscript{2}

North Carolina State University\textsuperscript{1}, Salesforce Research\textsuperscript{2}
Continual Learning

- Continual Learning is letting model learn multiple tasks sequentially
- Suffers from Catastrophic Forgetting
Current Approaches

- Fixed structure: Will finally limited by the capacity
- Manually growing is sub-optimal

**Fixed Structure** (Regularization based method, e.g. EWC, Kirkpatrick et. al., 2016)

**Uniformly Growth** for new task (Progressive Nets, Rusu et. al., 2016)
Learn to Grow: A Continual Structure Learning Framework for Overcoming Catastrophic Forgetting Poster #89

Proposed Method

**Fixed Structure** (Regularization based method, e.g. EWC, Kirkpatrick et. al., 2016)

**Uniformly Growth** for new task (Progressive Nets, Rusu et. al., 2016)

Learn-to-Grow

Structure Search from options: “reuse”, “adaptation”, “new”
Transfer knowledge from previously learned tasks (Super Net $\Theta_{t-1}$)

Structure Optimization

$$\theta_t = f(D_{t}^{\text{train}}, \Theta_{t-1})$$

Parameter Optimization

Update “reused” knowledge

Store new knowledge

$\Theta_t = \Theta_{t-1} \cup \theta_t$
Structures found by Learn-to-Grow Are Sensible

- Experiments on Permutated MNIST
  - The structure optimization results in “new” on the first layer and “reuse” for the rest.
  - Ablations experiments validates the search results.

- Experiments on Visual Domain Decathlon (VDD)
  - Qualitative analysis on the searched structure on Task 2 (Task 1: ImageNet)
    - Learned structure is sensible
      - Similar tasks tends to share more structure and parameters
      - Distant tasks share less
Are Forgetting Alleviated in Learn-to-Grow?

- Experiments on Permuted MNIST

- The “tune” higher than “fix” at certain task indicates “positive forward transfer”

- The “tune” curve “goes up” means “positive backward transfer”
Comparison with Other Approaches

Permuted-MNIST

Split-CIFAR100
Come to see our poster!

Today 6:30 – 9:00 PM @ Pacific Ballroom #89