Uncertainty-guided Continual Learning with Bayesian Neural Networks

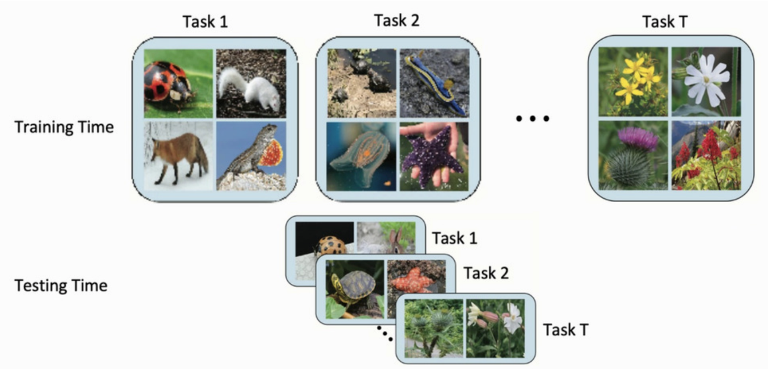
Sayna Ebrahimi, Mohamed Elhoseiny, Trevor Darrell, Marcus Rohrbach

ICLR 2020

Presenter: Jiaru Zhang

Introduction

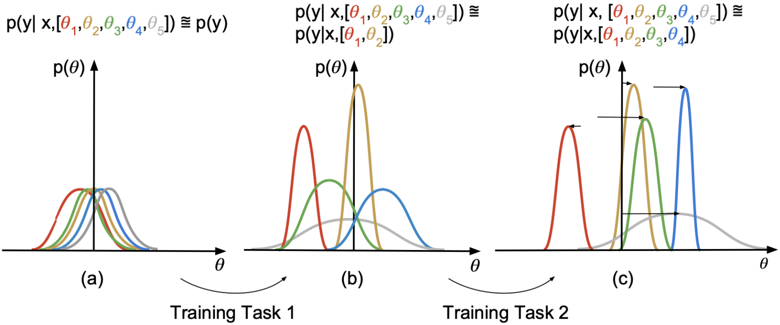
- Continual learning: **Sequentially learning tasks without forgetting**



- **What not to forget?** Important parameters.
- **How not to forget?** Minimize the change in important parameters.

Importance vs. Uncertainty

How do we define importance?



The more uncertain a parameter is, the more learnable it can be.

Uncertainty-guided Continual Learning with Bayesian NNs (UCB)

Each parameter is modeled by mean μ and variance ρ.

Learning rate regularization:

UCB

$$\alpha_\mu \leftarrow \alpha_\mu / \Omega_\mu, \tag{1}$$

$$\alpha_\rho \leftarrow \alpha_\rho / \Omega_\rho, \tag{2}$$

where Ω represents the importance.

$$\Omega_\mu = 1/\sigma, \tag{3}$$

$$\Omega_\rho = 1, \tag{4}$$

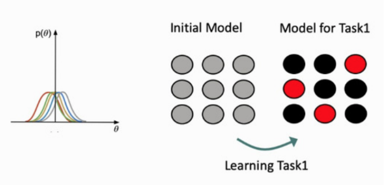
are the best settings empirically found. Here σ is the standard deviation.

UCB using weight pruning (UCB-P)

- Use the **signal-to-noise ratio (SNR)** as the importance for each parameter:

$$\Omega = \text{SNR} = |\mu|/\sigma \tag{5}$$

- After training on a task, we
  - Freeze the important parameters.
  - Prune the unimportant parameters.



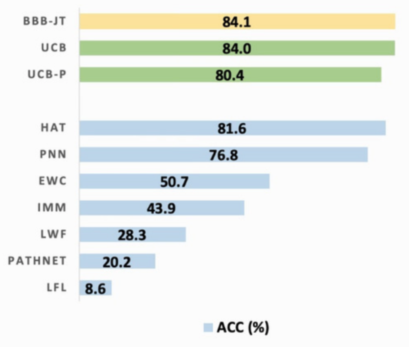
- Pros: Recovering pre-pruning performance.
- Cons: Saving masks per task; Require task information at test time.

Results: Sequence of 8 datasets

Datasets: FaceScrub, MNIST, CIFAR100, NotMNIST, SVHN, CIFAR10, TrafficSigns, and FashionMNIST .

Average Accuracy:

$$\text{ACC} = \frac{1}{n}\sum_{i=1}^{n} R_{i,n} \qquad (6)$$



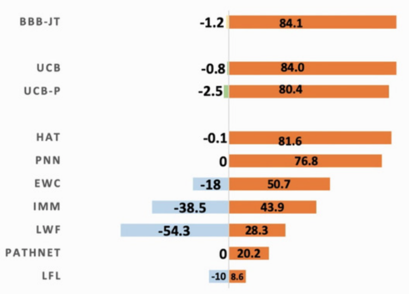
Results: Sequence of 8 datasets

Datasets: FaceScrub, MNIST, CIFAR100, NotMNIST, SVHN, CIFAR10, TrafficSigns, and FashionMNIST .

Backward Transfer (BWT):

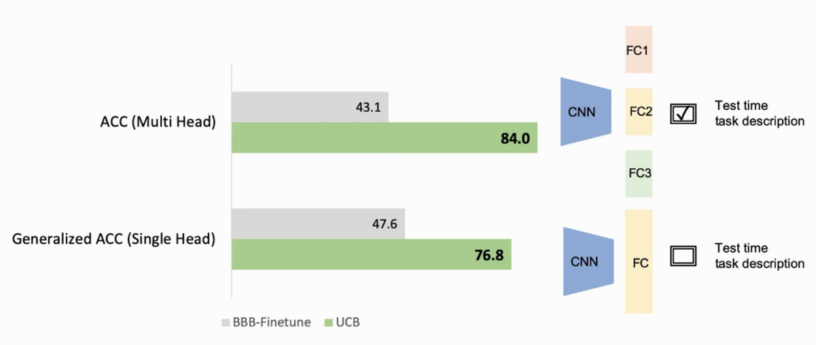
$$\mathrm{BWT} = \frac{1}{n}\sum_{i=1}^{n} R_{i,n} - R_{i,i}. \qquad (7)$$

It indicates how much learning new tasks has influenced the performance on previous tasks.



Results: Task-free

UCB can be used even if **the task information is not given** at test time.



Conclusion

- UCB regularizes the learning rate with the uncertainty measured by Bayesian NNs.
- The more uncertain the parameter is, the higher the learning rate should be.
- UCB can be task free.
- State-of-the-art results on image classification benchmarks.