YEMING WEN

Vector Institute, 661 University Ave Suite 710 ⋄ Toronto, Canada, M5S 3G4
Email: ywen@cs.toronto.edu

RESEARCH INTEREST

My research interests lie in the development of efficient learning algorithms for deep neural networks, with a focus on large batch training, ensemble methods and uncertainty modelling. Recently, I am particularly interested in lifelong learning, which enables deep neural network with human adaptability.

EDUCATION

University of Toronto, M.Sc. in Computer Science Sept. 2018 - Present
Machine Learning Group and Vector Institute
Awards: Vector Institute Scholarships in AI, University of Toronto Tuition Fellowship
Advisor: Jimmy Ba and Roger Grosse

University of Toronto, B.Sc. in Mathematics and Computer Science Sept. 2013 - June 2017
Cumulative GPA: 3.98/4.0, Average: 92.7/100
Awards: University of Toronto Excellence Awards (UTEA), The Dorothy Walters Scholarship (Four scholarships are awarded annually)
Competition: Quantathon Data Mining Competition Finalist (Top 5 in over 50 teams)

PREPRINTS AND PUBLICATIONS

Yeming Wen, Dustin Tran, Jimmy Ba. *BatchEnsemble: An Alternative Approach to Efficient Ensemble and Lifelong Learning.* Bayesian Deep Learning Workshop at NeurIPS-2019. (Submitted to ICLR-2020, https://openreview.net/pdf?id=Sklf1yrYDr)

Yeming Wen*, Kevin Luk*, Maxime Gazeau*, Guodong Zhang, Harris Chan, Jimmy Ba. *Interplay Between Optimization and Generalization of Stochastic Gradient Descent with Covariance Noise.* arXiv preprint arXiv:1902.08234, 2019. (Submitted to AISTATS-2020)

Tingwu Wang, Xuchan Bao, Ignasi Clavera, Jerrick Hoang, **Yeming Wen**, Eric Langlois, Shunshi Zhang, Guodong Zhang, Pieter Abbeel, Jimmy Ba. *Benchmarking Model-Based Reinforcement Learning.* arXiv preprint arXiv:1907.02057, 2019. (Submitted to ICLR-2020)

Yeming Wen, Paul Vicol, Jimmy Ba, Dustin Tran, Roger Grosse. *Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches.* International Conference on Learning Representations (**ICLR**), 2018.

RESEARCH EXPERIENCE

Research Internship at Google August 2019 - Dec 2019
Advisor: Dustin Tran *Toronto, Canada*

· Rank-1 Net: An Alternative Approach to Efficient Ensembles and Lifelong Learning

- Extended BatchEnsemble (Rank-1 net) to more complicated lifelong learning set-up, including a new benchmark dataset SPLIT-ImageNet.
- Demonstrated that Rank-1 Net is capable of learning a large number of lifelong learning tasks (up to 100) without forgetting, which no previous methods can achieve.
- Experiments in uncertainty modelling showed that Rank-1 Net is orthogonal to existing ensemble methods. Combining Rank-1 net with existing ensemble methods such as MC-dropout leads to better uncertainty predictions.

M.Sc. Research Project March 2019 - Dec 2019
Advisor: Prof. Jimmy Ba *Toronto, Canada*

· BatchEnsemble: Ensembles of Neural Networks in a Mini-Batch Friendly Way

- Proposed an efficient ensemble method which is mini-batch friendly. It incurs negligible computational and memory costs.
- Demonstrated its effectiveness in image classification and machine translation. BatchEnsemble also captures model uncertainty in contextual bandits task and achieves compelling calibrated predictions on CIFAR-10 corrupted dataset.
- Demonstrated BatchEnsemble can be used in large-batch training and continual learning.

Research Intern at Borealis AI Sept 2018 - Feb 2019
Advisor: Prof. Jimmy Ba *Toronto, Canada*

· Large-Batch Stochastic Optimization with Curvature Noise

- Explored different intrinsic noise structures in SGD optimization.
- Analytically showed that the convergence rate of noisy SGD optimization not only depends on the marginal variance of the noise but also the Frobenius norm of the noise matrix.
- Empirically verified the above conclusion and showed that adding diagonal Fisher noise to large batch gradient leads to better generalization without increasing the number of training iterations.

University of Toronto Excellence Awards May 2017 - Sept 2017
Research Assistant, Advisor: Prof. Roger Grosse *Toronto, Canada*

· Flipout: Efficient Pseudo-Independent Weight Perturbations on Mini-Batches

- Analytically showed that Flipout is unbiased and gives lower gradient variance than naive stochastic neural networks.
- Implemented the Flipout upon multiplicative perturbation algorithm with various neural network architectures, such as MLP, LeNet, VGG. Empirically evaluated that Flipout achieves an ideal variance reduction effect.
- Extended the algorithm to Bayesian neural networks (trained with Bayes by Backprop) and evolution strategies in both supervised learning and reinforcement learning. Evaluated by MNIST data set and Mujoco environment.

OTHERS

Reviewer	ICLR2020, ICML2019, NeurIPS2019
Programming Languages	Python, Matlab, R
Frameworks & Tools	Tensorflow, MXNet, PyTorch
Teaching	TAed Calculus, Theory of Computation, Probability and Statistics