船井情報科学振興財団 留学報告書

川口賢司

http://www.mit.edu/~kawaguch

My recent research focuses mostly on theoretical aspects of deep learning. In order to illustrate a reason why deep learning might be important in a casual manner, I will share the story of how I ended up working on deep learning.

I never thought that I would work on deep learning in 2014, when I started some research on classical machine learning and reinforcement learning. In 2014, I was happy with theory in classical machine learning. However, the trouble was, every time I tested theory in experiments, it seemed that what I, a human, choose (the basis, representation or corresponding kernel) was the dominant factor to determine the performance of *machine* learning. As I was excited when I studied beautiful classical machine learning theory including kernel theory based on functional analysis, I was disappointed by the observation that humans seemed to be doing most of the work instead of the machine. In 2015, I was still working on RKHS embedding theory, but at the same time, I started to look at *representation learning* as a flexible alternative to force the machine to do more work. In representation learning, what humans need to choose a priori (the basis, representation or corresponding kernel) in classical machine learning is also learned by the machine based on a dataset and some prior information given by the human. Therefore, to me, representation learning seemed to be a natural answer to the above issue. I then learned that "deep" structure in deep learning seems to be one of natural prior information to give to the machine (e.g., it captures hierarchical nature), in addition to its practical success. To this day, this story captures one of the motivations of why I am interested in deep learning.

Research update

There have been several research activities since the last research update in this series of reports. A complete list of references is presented at the end of this report.

AAAI 2018 – In February, I presented our accepted paper at The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), which took place in New Orleans, Louisiana:

Kenji Kawaguchi*, Bo Xie*, and Le Song. (2018). Deep Semi-Random Features for Nonlinear Function Approximation.

Cambridge University Press – I was personally contacted and invited to write a book chapter for *Mathematics of Deep Learning*, which will be published by Cambridge University Press:

Kenji Kawaguchi, Leslie Pack Kaelbling and Yoshua Bengio. Generalization in Deep Learning.

The code for the paper can be found here: code.

Microsoft Research, Redmond – In June, I was invited to visit Microsoft Research in Redmond, Washington. I spent three weeks as a Traveler/Lecture to discuss our work related to Deep Learning without Poor Local Minima and Generalization in Deep Learning. While there I had the opportunity to collaborate with leading researchers in the field of theoretical machine learning.

Theoretical Foundation of Deep Learning 2018 – In October I was sent to the workshop on Theoretical Foundation of Deep Learning conference at Georgia Tech. I attended the three-day workshop on request from my lab at MIT and it was very interesting. I was able to listen to many different researchers in the field and came up with some new ideas. While there I had the idea for the following paper:

Kenji Kawaguchi and Yoshua Bengio. (2018). [Depth with Nonlinearity Creates No Bad Local Minima in ResNets.]

In this paper, we have solved an open question stated in a paper in the conference on Neural Information Processing Systems (NIPS) 2018 (Shamir, 2018). In particular, because of our paper's result, it is now theoretically supported to prefer the use of *nonconvex* deep learning models over classical *convex* machine learning models, in terms of local minima in optimization. Figure 1 shows an illustration of local minima, a global minimum, and a saddle point. Intuitively, what we want to find would be a global minimum. However, local algorithms, such as (stochastic) gradient descent, can become stuck near a bad local minimum because of non-convexity and existence of potentially bad local minima. In this paper, we have shown that a practically successful type of deep neural networks (ResNets) induces a nonconvex optimization problem but *no* bad local minima, relative to convex optimization problem induced by corresponding classical machine learning models.

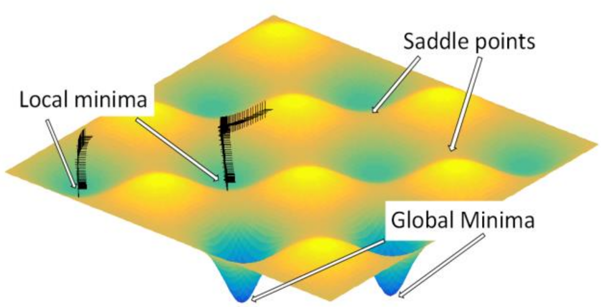

Figure 1: Illustration of local minima, a global minimum, and a saddle point

Beyond ResNets – I have also written another new paper, which goes beyond ResNets:

Kenji Kawaguchi, Jiaoyang Huang and Leslie Pack Kaelbling. (2018). Effect of Depth and Width on Local Minima in Deep Learning.

In this paper, the effects of depth and width on the quality of local minima are analyzed *without strong over-parameterization and simplification assumptions*. Among other results, this paper shows that without strong over-parametrization and simplification assumptions, the quality of local minima can improve towards the global minimum value as depth and width increase.

Another Collaboration with Professor Poggio's lab – I have been collaborating with Professor Poggio's lab for some time. This time, I have provided some theoretical results for the following paper:

Tomaso Poggio, Kenji Kawaguchi, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Xavier Boix, Jack Hidary and Hrushikesh Mhaskar. Theory of Deep Learning III: explaining the non-overfitting puzzle.

Other invitations – Among others, I have been invited to be a reviewer for Journal of Machine Learning Research (JMLR), Neural Information Processing Systems (NIPS), AAAI Conference on Artificial Intelligence (AAAI), Neural Computation (MIT press), and IEEE Transactions on Neural Networks and Learning Systems.

Ongoing research – I have recently obtained new theoretical results on a method of eliminating bad local minima, for which I am currently writing a new paper. I am also collaborating with Professor Yoshua Bengio on a topic of deep learning. In addition, I have been discussing other topics on deep learning with Professor Tomaso Poggio at MIT, Alex lamb and Vikas Verma at MILA, and Jiaoyang Huang at Harvard University. I am looking forward to reporting new results once they are publicly available.

MIT Winterfest 2018

MIT holds a Winterfest party every year. It's a social event to start the holiday season. Many departments also have holiday parties, but Winterfest is open to all MIT. This year it was held in the Charles M. Vest Student Street and the Koch Institute Lobby.

There were many different holiday foods and drinks available for everyone at MIT to enjoy. There were also many decorations, and holiday music was playing. Students and faculty seemed to be enjoying this event. MIT has many really fun events for students to enjoy, but the holiday events always seem to be the best. Everyone is busy this time of year with their own holiday plans, end of semester projects and finals, so it is very nice that MIT provides so many opportunities for students to relax and have fun.

I am definitely planning to attend some of the other department parties before the end of the holiday season.

船井情報科学振興財団には心より感謝しております。本当にありがとうございま す。

References

Kenji Kawaguchi*, Bo Xie*, and Le Song. (2018). Deep Semi-Random Features for Nonlinear Function Approximation. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI).

Kenji Kawaguchi, Leslie Pack Kaelbling and Yoshua Bengio. Generalization in Deep Learning. To be included in Mathematics of Deep Learning, Cambridge University Press. Prepint avaliable at: arXiv preprint arXiv:1710.05468, 2017.

Kenji Kawaguchi and Yoshua Bengio. Depth with Nonlinearity Creates No Bad Local Minima in ResNets. arXiv preprint arXiv:1810.09038, 2018.

Kenji Kawaguchi, Jiaoyang Huang and Leslie Pack Kaelbling. Effect of Depth and Width on Local Minima in Deep Learning. arXiv preprint arXiv:1811.08150, 2018.

Tomaso Poggio, Kenji Kawaguchi, Qianli Liao, Brando Miranda, Lorenzo Rosasco, Xavier Boix, Jack Hidary and Hrushikesh Mhaskar. Theory of Deep Learning III: explaining the non-overfitting puzzle. Massachusetts Institute of Technology CBMM Memo No. 73, 2018.

Ohad Shamir. Are resnets provably better than linear predictors? (2018). In Advances in Neural Information Processing Systems.