論文の内容の要旨

論文題目　Mathematical Analysis of Large-Scale Boltzmann Machines
(大規模ボルツマンマシンの数理的解析)

氏　　名　木脇　太一

While great progress has been achieved with deep discriminative models, a lot of important problems remain unsolved for deep generative models (DGMs). A serious difficulty in DGMs is that they lack an appropriate measure for their representational power. This deficiency makes it difficult to quantify the trade-off between tractability and representational power in various models.

In this thesis, we study the representational power of a DGM by mathematically analyzing Boltzmann machines (BMs). We develop several measures for the representational power of BMs, and theoretically and empirically demonstrate high generative performance of deeply layered BMs.

The thesis consists of three parts. In the first part (in Chapters 1 and 2), we provide a broad review on the whole research area. In Chapter 1, we review recent progress in deep learning research on deep discriminative models as well as DGMs. In Chapter 2, we review BMs in detail. We describe variants of BMs, training and evaluation algorithms for BMs.

In the second part (from Chapter 3 to Chapter 5), we develop methods to evaluate the quality of BM representations. In Chapter 3, we roughly quantify amount of information transmitted from data to representation of Gaussian RBMs (GRBMs), a variant of RBMs. We relate this approximate measure to the quality of a GRBM representation with respect to classification of natural images. In Chapter 4, we develop a method to design an efficient initial distribution for annealing GRBMs using annealed importance sampling (AIS). In Chapter 5, we develop a method to approximate the optimal annealing schedule for a general BM that minimizes the estimation error of AIS. We analytically derive a functional that dominates the AIS estimation error. We propose a method to approximate the optimal schedule by numerically minimizing this functional. We experimentally demonstrate that the proposed algorithm mostly outperforms conventional scheduling schemes with large quantization numbers.

In the final part (in Chapter 6), we turn to the other main theme of this thesis: the representational power of BMs. We quantify the representational power of a BM with a piecewise linear approximation of the negative log likelihood of the BM, or the free energy function of the BM. This approximation bounds the true free energy function from above. We quantify the representational power of a BM with the complexity of this approximating function, which can be measured with the number of linear regions of it. This measure roughly indicates the number of effective mixing components that the BM has. We show bounds of this measure for numerous BM architectures:

RBMs, DBMs, and soft-deep BMs, a newly introduced BM architecture which we dub sDBMs. These analyses show that DBMs do not have large representational power as commonly expected, and sDBMs can have exponentially greater representational power than both RBMs and DBMs. We finally experimentally demonstrate the state-of-the-art generative performance of sDBMs on two benchmark datasets.