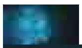
Dec 16 | Dr. Hari Koduvely

Bayesian Research in NeurIPS2018



Bayesian Research at NeurIPS2018

Dr. Hari Koduvely

Neural Information Processing Systems conference (NeurIPS or NIPS previously) is considered as the most reputed academic conference in Artificial Intelligence. This year's conference, NeurIPS2018 was held in the beautiful city of Montreal from 3rd to 8th December. Thanks to Facebook, the conference was live streamed and videos of all lectures is available on NIPS Foundation Facebook page. There were some very interesting keynote lectures. The most fascinating talk I found was Bioelectric Computation Outside the Nervous System, Primitive

Cognition and Synthetic Morphology by Professor Micheal Levin. Other noticeable keynote lectures were Reproducible, Reusable and Robust Reinforcement Learning by Professor Joelle Pineau, Making Algorithms Trustworthy by David Spiegelhalter and Designing Computer Systems for Software 2.0 by Kunle Olukotun. There was also a very interesting workshop on Machine Learning for Molecules and Materials for discussing applications in Physical Sciences.

I was following NeurIPS2018 mainly for understanding latest research going on in **Bayesian Inference** and its use in Machine Learning and Artificial Intelligence. As expected there were quite a few papers, more than 70, spanning the 6 day conference. Day 0 tutorial on Scalable Bayesian Inference by Professor David Dunson was a very useful one. This tutorial contains an overview of state-of-the-art approaches for analyzing massive data sets using Bayesian statistical methods. Professor Dunson discussed simple approaches for scaling up commonly used Markov chain Monte Carlo (MCMC) algorithms. Some examples discussed were embarrassingly parallel (EP) MCMC, approximate MCMC, stochastic approximation, hybrid optimization & sampling, and modularization. These methods have applications in areas such as computational advertising, genomics and neurosciences.

There were also two workshops related to **Bayesian Inference**. One was on Bayesian Deep Learning and the other was on Non-Parametric Bayesian Inference both are very active areas of research currently.

The oral presentations and posters covered several aspects of **Bayesian Inference**, both theoretical advancements and applications in machine learning. The topics covered includes Bayesian Deep Learning, Bayesian Reinforcement Learning, Bayesian Optimization, Variational Inference, Variational Auto-Encoder, Markov Chain Monte Carlo (MCMC) Methods, Representation Learning or Meta Learning, Cognitive Science, Differential Privacy, Approximate Bayesian Methods and Bayesian Networks. This wide set of topics shows the importance of Bayesian methods in Machine Learning and Artificial Intelligence.

I will here summarize some of the talks/posters I found interesting, full list of talks/posters can be found on NeurIPS2018 Proceedings page.

Neural Architecture Search with Bayesian Optimization and Optimal Transport

In this work, the authors develop NASBOT, a Gaussian process based Bayesian Optimization framework for neural architecture search. The authors develop a distance metric in the space of neural network architectures, which can be computed efficiently via an optimal transport program.

[Constructing Deep Neural Networks by Bayesian Network Structure Learning]
In this paper the authors introduce a principled approach for unsupervised structure learning of deep neural networks. They propose a new interpretation for depth and inter-layer connectivity, where conditional independencies in the input distribution are encoded hierarchically in the network structure, which enables the depth of the network to be determined inherently. The proposed method casts the problem of neural network structure learning as a problem of Bayesian network structure learning.

[Explaining Deep Learning Models — A Bayesian Non-parametric Approach]
In this work, the authors propose a novel technical approach that augments a Bayesian non-parametric regression mixture model with multiple elastic nets. Using the enhanced mixture model, one can extract generalizable insights for a target model through a global approximation.

[Bayesian Adversarial Learning]
Deep neural networks are vulnerable to adversarial attacks and the standard defensive approaches is through formulating it as a robust optimization problem. Here one minimizes a point estimate of the worst-case loss derived from an *adversary data-generating distribution*. In this work, a novel robust training framework called *Bayesian Robust Learning*, is proposed, in which a distribution is put on the adversarial data-generating distribution to account for the uncertainty of the adversarial data-generating process.

[Bayesian Distributed Stochastic Gradient Descent]
This work discusses a high-throughput algorithm for training deep neural networks on parallel clusters. This algorithm uses amortized inference in a deep generative model, to perform joint posterior predictive inference of mini-batch gradient computation times, in a compute cluster specific manner. Particularly, this algorithm mitigates

the straggler effect in synchronous, gradient-based optimization by choosing an optimal cutoff beyond which mini-batch gradient messages from slow workers are ignored.

Bayesian Model-Agnostic Meta-Learning

In this paper, the authors propose a novel Bayesian model-agnostic meta-learning method for learning from a small data sets. The proposed method combines efficient gradient-based meta-learning with nonparametric variational inference in a principled probabilistic framework.

Beauty-in-averageness and its contextual modulations: A Bayesian statistical account

Understanding how humans perceive the likability of high-dimensional objects'' such as faces is an important problem in both cognitive science and AI/ML. It is well known from psychology literature that human assessment of facial attractiveness is context-dependent. In this paper, the authors hypothesize that human preference for an object is increased when it incurs lower encoding cost, in particular when its perceived *statistical typicality* is high, in consonance with Barlow's seminal efficient coding hypothesis.

Bayesian Inference of Temporal Task Specifications from Demonstrations

When observing task demonstrations, human apprentices are able to identify whether a given task is executed correctly long before they gain expertise in actually performing that task. The authors present *Bayesian Specification Inference*, a probabilistic model for inferring task specification as a temporal logic formula. Authors incorporate methods from probabilistic programming to define their priors, along with a domain-independent likelihood function to enable sampling-based inference.

Predictive Approximate Bayesian Computation via Saddle Point

Approximate Bayesian computation (ABC) is an important methodology for Bayesian inference when the likelihood function is intractable. In this paper, the authors introduce an optimization-based ABC framework that addresses deficiencies in existing methods. Leveraging a generative model for posterior and joint distribution matching, the authors show that ABC can be framed as saddle point problems, whose objectives can be accessed directly with samples.

Reinforcement Learning with Multiple Experts: A Bayesian Model Combination Approach
In this paper, the authors apply Bayesian Model Combination with multiple experts in a way that learns to trust a good combination of experts as training progresses.

Variational Bayesian Monte Carlo
Many probabilistic models of interest in scientific computing and machine learning are intractable and require access to the gradient or a large number of likelihood evaluations. The authors introduce here a novel sample-efficient inference framework, *Variational Bayesian Monte Carlo* (VBMC). VBMC combines variational inference with Gaussian-process based, active-sampling Bayesian quadrature, using the latter to efficiently approximate the intractable integral in the variational objective.

Inference in Deep Gaussian Processes using Stochastic Gradient Hamiltonian Monte Carlo
Deep Gaussian Processes (DGPs) are hierarchical generalizations of Gaussian Processes that combine well calibrated uncertainty estimates with the high flexibility of multilayer models. One of the biggest challenges with these models is that exact inference is intractable. In this work, the authors provide evidence for the non-Gaussian nature of the posterior and they apply the Stochastic Gradient Hamiltonian Monte Carlo method to generate samples from posterior distribution.

Algorithmic Assurance: An Active Approach to Algorithmic Testing using Bayesian Optimization
In this work, the authors introduce algorithmic assurance, the problem of testing whether machine learning algorithms are conforming to their intended design goal. The authors mathematically formulate this task as an optimization problem of an expensive, black-box function. They use an active learning approach based on Bayesian optimization to solve this optimization problem.

Discretely Relaxing Continuous Variables for tractable Variational Inference
The authors explore a new research direction in Bayesian variational inference with discrete latent variable priors where they exploit Kronecker matrix algebra for efficient and exact computations of the evidence lower bound (ELBO). This results in posterior samples consist

of sparse and low-precision quantized integers which permit fast inference on hardware limited devices.

Wasserstein Variational Inference
This paper introduces *Wasserstein Variational Inference*, a new form of approximate Bayesian inference based on optimal transport theory. Wasserstein variational inference uses a new family of divergences that includes both f-divergences and the Wasserstein distance as special cases. This technique results in a very stable likelihood-free training method that can be used with implicit distributions and probabilistic programs.

Learning Latent Subspaces in Variational Autoencoders
It is often difficult to interpret the latent space representations learned using *Variational Autoencoders* (VAE). The authors propose a VAE-based generative model, which is capable of extracting features correlated to binary labels in the data and structuring it in a latent subspace, which is easy to interpret.

Constrained Graph Variational Autoencoders for Molecule Design
With an emphasis on applications in chemistry, the authors explore the task of learning to generate graphs that conform to a distribution observed in training data. They propose a Variational Autoencoder model in which both encoder and decoder are graph-structured. They show that by using appropriate shaping of the latent space, this model allows to design molecules that are (locally) optimal in desired properties.

I might have omitted here several other important work on **Bayesian Inference** presented at NeurIPS2018. This is partly because of my personal choices and also I might not have noticed them among the several hundreds of papers and posters presented at the conference. I would encourage the readers to search through the NeurIPS2018 Proceedings to find other interesting papers.

In summary the extensive list of work on Bayesian Inference presented at NeurIPS2018 shows the relevance of the subject in the modern era of Machine Learning and Artificial Intelligence. There has been some important advances last year in the state-of-art Bayesian Inference in terms of better algorithms for posterior density estimation, and application in problems ranging from interpretation of Deep Learning models to design of new molecules.
