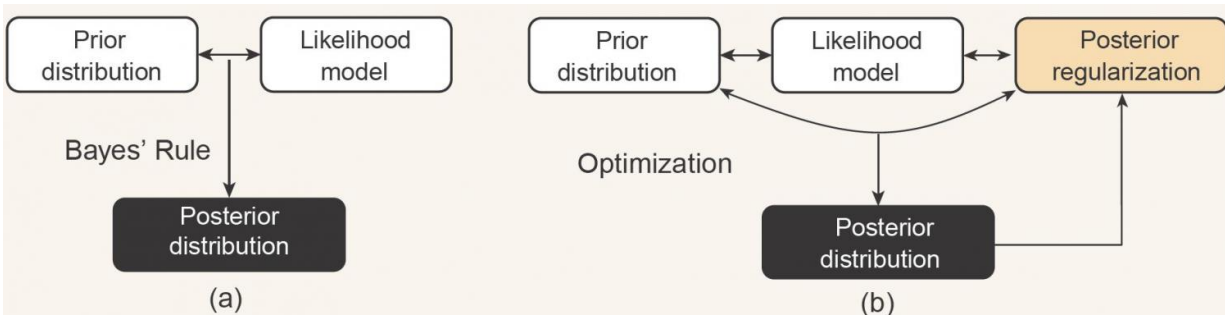


Advances in Bayesian methods for big data

June 2 2017



(a) This is an illustration of the Bayesian inference with Bayes' rule. (b-f) Regularized Bayesian inference under an optimization framework. Credit: ©Science China Press

In the Big Data era, many scientific and engineering domains are producing massive data streams, with petabyte and exabyte scales becoming increasingly common. Besides the explosive growth in volume, Big Data also has high velocity, high variety, and high uncertainty. These complex data streams require ever-increasing processing speeds, economical storage, and timely response for decision making in highly uncertain environments, and have raised various challenges to conventional data analysis.

With the primary goal of building intelligent systems that automatically improve from experiences, machine learning (ML) is becoming an increasingly important field to tackle big data challenges, with an

emerging field of "Big Learning," which covers theories, algorithms and systems on addressing big data problems.

Bayesian methods have been widely used in machine learning and many other areas. However, skepticism often arises when we talking about Bayesian methods for Big Data. Practitioners also note that Bayesian methods are often too slow for even small-scaled problems, owing to many factors such as the non-conjugacy models with intractable integrals. Nevertheless, Bayesian methods have several advantages.

First, Bayesian methods provide a principled theory for combining prior knowledge and uncertain evidence to make sophisticated inference of hidden factors and predictions.

Second, Bayesian methods are conceptually simple and flexible—hierarchical Bayesian modeling offers a flexible tool for characterizing uncertainty, missing values, latent structures, and more. Moreover, regularized Bayesian inference (RegBayes) further augments the flexibility by introducing an extra dimension (i.e., a posterior regularization term) to incorporate domain knowledge or to optimize a learning objective.

Finally, there exist very flexible algorithms (e.g., Markov chain Monte Carlo) to perform posterior inference.

In a new overview published in the Beijing-based *National Science Review*, scientists at Tsinghua University, China present the latest advances in Bayesian methods for Big Data analysis. Co-authors Jun Zhu, Jianfei Chen, Wenbo Hu, and Bo Zhang cover the basic concepts of Bayesian methods, and review the latest progress on flexible Bayesian methods, efficient and scalable algorithms, and distributed system implementations.

These scientists likewise outline the potential development directions of future Bayesian methods.

"Bayesian methods are becoming increasingly relevant in the Big Data era to protect high capacity models against overfitting, and to allow models adaptively updating their capacity. However, the application of Bayesian methods to [big data](#) problems runs into a computational bottleneck that needs to be addressed with new (approximate) inference methods."

The scientists overview the recent advances on nonparametric Bayesian methods, regularized Bayesian inference, scalable algorithms, and system implementation.

The scientists also discuss on the connection with deep learning, "A natural and important question that remains under addressed is how to conjoin the flexibility of deep learning and the learning efficiency of Bayesian methods for robust learning," they write.

Finally, the scientists say, "The current machine learning methods in general still require considerable human expertise in devising appropriate features, priors, models, and algorithms. Much work has to be done in order to make ML more widely used and eventually become a common part of our day to day tools in data sciences."

More information: Jun Zhu et al, Big Learning with Bayesian Methods, *National Science Review* (2017). [DOI: 10.1093/nsr/nwx044](https://doi.org/10.1093/nsr/nwx044)

Provided by Science China Press

Citation: Advances in Bayesian methods for big data (2017, June 2) retrieved 29 January 2023

from <https://phys.org/news/2017-06-advances-bayesian-methods-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.