



香港中文大學統計學系

Department of Statistics

THE CHINESE UNIVERSITY OF HONG KONG

DISTINGUISHED LECTURE

Repro Samples Method and Principled Random Forests



Professor **Min-ge XIE**

Distinguished Professor
Department of Statistics,
Rutgers University

Director, Rutgers Office of Statistical Consulting

Editor, The American Statistician (TAS)

Co-Editor-in-Chief, The New England Journal of Statistics in Data
Science (NEJSDS)

Fellow, IMS, ASA

Elected Member, ISI

Date: 7 June 2024 (Friday)

Time: 10:30 am — 11:30 am

Venue: LT6, Lady Shaw Building,
The Chinese University of Hong Kong

Abstract

Repro Samples method introduces a fundamentally new inferential framework that can be used to effectively address frequently encountered, yet highly non-trivial and complex inference problems involving discrete or non-numerical unknown parameters and/or non-numerical data. In this talk, we present a set of key developments in the repro samples method and use them to develop a novel machine learning ensemble tree model, termed *principled random forests*. Specifically, repro samples are artificial samples that are *reproduced* by mimicking the genesis of observed data. Using the repro samples and inversion techniques stemmed from fiducial inference, we can establish a confidence set for the underlying ('true') tree model that generated, or approximately generated, the observed data. We then obtain a tree ensemble model using the confidence set, from which we derive our inference. Our development is principled and interpretable since, firstly, it is fully theoretically supported and provides frequentist performance guarantees on both inference and predictions; and secondly, the approach only assembles a small set of trees in the confidence set and thereby the model used is interpretable. The development is further extended to handle tree-structured conditional average treatment effect in a causal inference setting. Numerical results have demonstrated superior performance of our proposed approach than existing single and ensemble tree methods.

The repro samples method provides a new toolset for developing interpretable AI and for helping address the blackbox issues in complex machine learning models. The development of the principle random forest is our first attempt on this direction.

★ ★ ★ ★ ★ All are welcome ★ ★ ★ ★ ★

For enquiries please contact Miss Esther TAM (Tel: 3943 7931)
<https://www.sta.cuhk.edu.hk/news-and-events/>

