

Predictive Inference of Machine Learning for Official Statistics

Li-Chun Zhang¹

Abstract

Machine learning (ML) methods are increasingly important for estimation in official statistics, which targets descriptive summaries of finite populations, where models are viewed as ‘statistical algorithms’ with specified input and algorithmically generated outputs. Such uses of ML notably do not rely on the ideal that a true data model exists, nor that one is able to acquire the true data model should it exist. To make valid predictive inference (i.e. statistical inference of descriptive targets), one often needs to accommodate arbitrary ML models that are patently not ‘true’. In this paper, we highlight assumption-lean predictive inference for ML applications, discuss some recent developments for supervised learning and transfer learning in this perspective, and point out some topics for future development.

Keywords: algorithmic model, design-based prediction, Neyman-Fisher consistency, training test sample-split, Rao-Blackwellisation, prior information

¹ 1Statistisk sentralbyrå, Norway & University of Southampton, UK, e-mail: L.Zhang@soton.ac.uk.