How Using Machine Learning Classification as a Variable in Regression Leads to Attenuation Bias and What to Do About It

Abstract

Social scientists have increasingly been applying machine learning algorithms to "big data" to measure theoretical concepts they cannot easily measure before, and then been using these machine-predicted variables in a regression. This seminar first demonstrates that directly inserting binary predictions (i.e., classification) without regard for prediction error will generally lead to attenuation biases of either slope coefficients or marginal effect estimates. Dr Zhang Han then proposes several estimators to obtain consistent estimates of coefficients. The estimators require the existence of validation data, of which researchers have both machine prediction and true values. This validation data is either automatically available during training algorithms or can be easily obtained. Monte Carlo simulations demonstrate the effectiveness of the proposed estimators. Finally, he summarizes the usage pattern of machine learning predictions in 18 recent publications in top social science journals, apply the proposed estimators to two of them, and offer some practical recommendations.