World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:04, 2022

A Large Dataset Imputation Approach Applied to Country Conflict Prediction Data

Authors : Benjamin Leiby, Darryl Ahner

Abstract : This study demonstrates an alternative stochastic imputation approach for large datasets when preferred commercial packages struggle to iterate due to numerical problems. A large country conflict dataset motivates the search to impute missing values well over a common threshold of 20% missingness. The methodology capitalizes on correlation while using model residuals to provide the uncertainty in estimating unknown values. Examination of the methodology provides insight toward choosing linear or nonlinear modeling terms. Static tolerances common in most packages are replaced with tailorable tolerances that exploit residuals to fit each data element. The methodology evaluation includes observing computation time, model fit, and the comparison of known values to replaced values created through imputation. Overall, the country conflict dataset illustrates promise with modeling first-order interactions while presenting a need for further refinement that mimics predictive mean matching.

Keywords : correlation, country conflict, imputation, stochastic regression
Conference Title : ICASDA 2022 : International Conference on Applied Statistics and Data Analytics
Conference Location : Boston, United States
Conference Dates : April 21-22, 2022