

A Bayesian Dual Network Clustering Approach for Data Aggregation and Parameter Heterogeneity

While there is significant extant research in the field of Bayesian statistics and quantitative marketing on model selection (e.g. using BIC, marginal likelihood, Bayes factors, etc...) each of these methods both conditions on the observed data (and its selected granularity) and parameter heterogeneity (selected by the researcher). That is, unknowingly, researchers think they are doing model selection but they are really doing data and heterogeneity “conditional model selection”. In this research, we present a novel approach where each observation is represented as a node in a data network and a corresponding parameter network in which we use a Bayesian clustering proceed to sample (data clusters and parameter clusters) from the posterior distribution to infer the appropriate level of data aggregation and parameter heterogeneity. To accomplish this, we develop a novel extension of the Distance-Dependent Chinese Restaurant Process (ddCRP) prior and a merge-and-split MCMC sampler to perform posterior sampling. We apply our method to two canonical problems in marketing (SKU aggregation and household clustering), compare it to extant methods (e.g. latent class and unsupervised learning) and demonstrate that “one approach fits all”.