Tutorial 5

Using Correlation Analysis and Big Data to Identify and Predict Musical Behaviors

Jeff C. Smith (Smule)

Abstract

New and significant repositories of musical data afford unique opportunities to apply data analysis techniques to ascertain insights of musical engagement. These repositories include performance, listening, curation, and behavioral data. Often the data in these repositories also includes demographic and/or location information, allowing studies of musical behavior, for example, to be correlated with culture or geography. Historically, the analysis of musical behaviors was limited. Often, subjects (e.g. performers or listeners) were recruited for such studies. This technique suffered from issues around methodology (e.g. the sample set of subjects would often exhibit bias) or an insufficient number of subjects and/or data to make reliable statements of significance. That is to say the conclusions from these studies were largely anecdotal. In contrast to these historical studies, the availability of new repositories of musical data allow for studies in musical engagement to develop conclusions that pass standards of significance, thereby yielding actual insights into musical behaviors. This tutorial will demonstrate several techniques and examples where correlation and statistical analysis is applied to large repositories of musical data to document various facets of musical engagement. Web site: https://ccrma.stanford.edu/damp/ Stanford University has created a new corpus of amateur music performance data, the Stanford Digital Archive of Mobile Performances, or DAMP, to facilitate the study of musical engagement through application of correlation and statistical analysis.