

AUTOMATIC MUSIC RECOMMENDATION BASED ON MUSIC GENOME PROJECT ATTRIBUTES

Jonathan Lessinger¹, Faculty Advisor: Stan Zdonik²

¹*jlessin@cs.brown.edu, jonathan.lessinger@gmail.com*

²*sbz@cs.brown.edu*

Abstract

Music recommendation based on musical content rather than just track metadata is notoriously difficult. Some companies, like Pandora, build recommendation systems using musical attributes from the Music Genome Project, which are hand-labeled by trained musicians. In order to automate this, I look at using lower-level “raw” audio features that can be generated algorithmically from an audio file to estimate these higher-level, more musically-meaningful attributes. The raw features come from Spotify’s EchoNest service - specifically, from a dataset collected several years ago by Columbia called the Million Song Dataset. For training and testing, I use the CAL10k dataset (formerly known as Swat10k), which is a set of tracks with tags similar to Music Genome Project attributes. I also begin to explore the power of these learned attributes to predict how much a user likes a particular track. By incorporating user-track-playcount triples from Last.fm, I can estimate playcounts for new user-track pairs. I do this using the previously learned attributes, and compare this against predictions from the same (or a similar) model trained on the CAL10k and EchoNest data.