

Project Proposal: Computer-executed Genre Classification of Music Computer Systems Lab 2009-2010

Alex Stabile

October 26, 2009

1 Purpose and Scope of the Research Project

The purpose of this project is to write code that could identify key characteristics of a given training set of music, determining what makes it similar stylistically. It would then be used to assign samples of music to a specific genre.

2 Background and review of current literature/research in this area

Current research often uses statistical models to determine how a given piece of music should be categorized. This is sometimes done with a focus on instrumentation; however, this approach is not as effective for solo works. Other projects have tried to classify a piece of music based on a short digital sample by analyzing frequencies and timbral textures. None of these methods have yielded perfectly successful results even many people cannot successfully place music (especially from a short sample) into its correct genre.

3 Computer Language and Software

The code will be written in python, and testing results will display data using the PyLab package.

4 Algorithms and Methods to be Researched

Code will obtain note and rhythm information about a piece of music by reading a midi file. To sort the music, I will apply a Nave Bayes classifier. This probability model treats each probability vector as independent, but can still provide useful results. It is often used to categorize bodies of text based on what words are found within. In this case, the probability vectors are calculated based on the occurrence of words in the text. After analyzing a set of training data, the resulting model can attempt to categorize a new body of text. To apply the Bayes classifier to music, probability vectors would be calculated by analyzing the notes and rhythms a piece of music contains.

5 Testing and Analysis

Creating a useful probability model requires a large body of training data. Fortunately, there are large libraries of music in the public domain that are available online in midi format. The classifier can analyze a large amount of this music to create a statistical model, and then use different pieces for testing. The success of the model will be evaluated based on how often it categorizes the music correctly.

6 Expected Results and Value to Others

This project could be used as a valuable tool in organizing large libraries of music. Genres are often generalities that do not necessarily fit music exactly, but a successful statistical model could determine which pieces are most similar and therefore should be grouped together. This could also be applied to if you liked this, you may also like situations that are so important to many music-oriented businesses.