Music Genre Classification

Thomas Lillis

Student  
9962 Bear Creek Apartments  
Boulder, CO 80303   
+1 303 590 5985

thli2089@colorado.edu

Joseph Marylander

Student  
1025 Adams Circle, E215  
Boulder, CO 80303  
+1 303 475 8838

joma1054@colorado.edu

Connor McGuinness

Student  
805 29th St. #557  
Boulder, CO 80303  
+1 203 521 9512

comc3361@colorado.edu

**ABSTRACT**

Genres within music are culturally agreed upon categorical descriptions used to filter songs and artists. As music production has boomed with technology, the list of genres is well beyond reach for a single person to know all of them. In this work, a classification process that eventually will add tags to previously unclassified data containing artist and song information. In detail, the process will analyze certain musical elements like rhythm, time signature, ‘danceability’, or key in order to create tags such as ‘pop’ and ‘rap’ or even ‘happy’ and ‘creepy’. In order to develop such tags, the classification process will utilize methods currently being discussed in class.

Bayesian classification will be one of the first test classification methods used due to it’s easy to use structure and good fit with our project goal. For example, setting pre-defined probabilities for each genre tag and then applying a value for ‘rhythm’ which, depending on it’s amount, will have a stronger or lesser affect on the probability of the genre tag. Determining probabilities for each genre tag is where our classification process will become our own unique algorithm. Therefore, during the development of our classification process, we will work to minimize any biases towards a certain value affecting a probability. For example, just because a song has a high value for ‘beats per minute’ does not automatically give the song a tag such as ‘happy’ or ‘exciting’ since another value may reduce the probability the song would receive such a tag.

As well as testing a handful of classification methods, classification accuracy needs to be analyzed as well. Using a clustering algorithm, we will able to more easily check the accuracy of the tags of each song produced during the classification process.

Based on the classification accuracy and process, a user interface will be developed on either a web or mobile platform. This interface will allow an unfamiliar user to interact with the dataset and filter songs based on the pre-defined tags. Users will be able to select from a list of tags and once submitted, be presented with a list of songs that have one or more of those tags.

**1. MOTIVATION**

Though how we interpret music genres as humans is a subjective classification, there are still some metrics which specifically fit genres in which most can agree. We will attempt to find and use the best attributes extracted from raw audio files to classify the audio files into commonly accepted, broad, generic, genres. This type of work could be useful for many different types of applications. It may be used for a Pandora type website for finding songs with a similar genre classification that the user might enjoy or could be used by radio stations to find more music they could play for their station. For this project we plan on using several different types of multi-class classification to see which method is most effective with our dataset. If no method is as informative as would like we will extend the project to extracting more attributes from the audio files to get more accurate classification.

# 2. LITERATURE SURVEY

In the paper [1] “Exploring different approaches for music genre classification” Antonio Jose Homsli Goulart, et al. describe how they extracted various features from a dataset of songs and how they attempted to classify them. They focused only on three genres of music: Blues, Classical, and Lounge.  In contrast to their research, we plan on classifying our data into as many genres as we can while being accurate. However, their classifications techniques are useful in determining how we should proceed with our own classification.

Similar to our project goal, this team proposed a musical genre classification system using attributes such as rhythmic structure and instrumentation [2]. Their work aims to solve the problem of the large amount of music being created without an automated genre classification system. As well, the team created two graphical interfaces to allow unfamiliar users to interact with their classification system; something our team wants to do. Understanding George Tzanetakis’ and his team’s approach to genre classification will help guide us in creating a similar project but with a different genre classification process.

Thierry Bertin-Mahieux [3] takes a simplified approach to musical genre classification, which is something our team wants to do as well. Instead of allowing tags to take multiple forms like ‘American pop’, ‘indie pop’, or ‘European pop’, they chose to include all of these tags into the tag ‘pop’. Thierry, B. M genre classification process only allows for the following 10 genres: classic pop and rock, folk, dance and electronica, jazz and blues, soul and reggae, punk, metal, classical, pop, and hip-hop. As well as only including these genres, Thierry, B. M excludes other popular genre tags within their dataset. This allows them to focus on artists or songs that are only within ‘pop’ instead of both ‘pop’ and ‘electronic’ for example. Our team’s aim is to follow a similar approach in simplifying the genres we choose, but also

# 3. PROPOSED WORK

The project will have multiple subtasks grouped into main tasks. Initially, main tasks will focus on obtaining and organizing the [Million Song Dataset](http://labrosa.ee.columbia.edu/millionsong/). Determining interesting attributes within the dataset and then filtering these attributes will be an important first step. The first phase of the project will be broken down into a collections of subtasks that aim to ensure the data is accurate and organized, and that there are enough values within each attribute to find meaningful trends.

Next, the second phase of the project will be focused on parsing the dataset and developing and trying different types of classification methods. Subtasks will consist of developing a parsing script, determining several classification methods, and implementing/testing the classification methods. This phase’s main goal is to create a well thought out classification method and then implement it. The third and final phase of the project will be analyzing the types of meta tags outputted from the classification process.  By analyzing these meta tags in we should, for example, be able to correctly predict other songs in our database that are similar thus creating a music recommendation service.

There are a number of ways we are considering classifying our data into specific genres.  Since we are also being provided labels for our data thereby allowing us to create a training set we will be determining which one of the following techniques provides the best accuracy: DAG SVM, neural network, k-means, k-nn.  However, we are also interested in running our data against a clustering algorithm in order to see what type of clusters are generated and how they compare to the actual labels we have.

# 4. MILESTONES

* Obtain the Million Song dataset.
* Determine interesting attributes within the dataset.
* Filter any noise or unneeded attributes within dataset that do not contribute to classification requirements.
* Create a list of ‘tags’ that relate to the chosen attributes.
* Test a couple of classification methods such as Bayesian or using a clustering algorithm on a small sample size of the data.
* Learn ‘scipy/sklearn’ to create a data parsing script.
* Implement the chosen classification methods in Python.
* Test implementation against a small sample size.
* Ensure implementation can handle the full dataset.
* Test accuracy of results by analyzing tag and song combinations.
* Store results in a SQL database.
* Develop a user interface to search the database.
* Make the user interface easy to use for an unfamiliar user.

# 5. EVALUATION

In order to determine the accuracy of our classification algorithm, the team will require a set of labels that will objectively classify the genre of the song. Evaluation will be done by comparing the team’s classification labels to the labels in the dataset. We will use Yajie Hu et al.’s genre dataset[4]. This dataset is a mirror of the original “Million Song Dataset” with all the features removed except for the song name and their genre label. By using their collection of genre labels in the team’s training and evaluation set, it is guaranteed that the labels assigned to each song will match with this dataset. The ideal way to compare the accuracy for each of the proposed classification tools is to use a confusion matrix. It should be very easy to take the team’s results from each of these algorithms along with the correct labels found in this dataset and run them through sklearn’s confusion\_matrix function to gather the results.

The team also plans on classifying this dataset using an unsupervised clustering algorithm in order to see what genre clusters emerge and how they compare to the genres assigned by humans. It will be challenging to evaluate the quality of the clusters because there are not many good labels that are assigned to theses genres. The team will have to parse through theses clusters and attempt to rename the labels to have similar names to that of the dataset­ [4]. This is not an easy process because the team will have to determine what the main common features are of the set and compare that to the predefined genres in order to best rename the clusters to the appropriate label.

Furthermore, it is very possible with the team’s clustering algorithm, that some genres will be lost and as a result two or more genres are merged into a single cluster due to similarities of the song features. In this event there is no accurate way to evaluate these clusters because they do not represent any data. However, these clusters may provide insight into how genres are similar and how they could have influenced each other over the course of time. It is very likely that a clustering algorithm will choose to cluster on elements that do not directly relate to a genre, such as ‘danceability’ or the energy of a song. In this event evaluating accuracy is also very challenging but it has further applications in areas like song suggestions based on mood or “feel” of a song.

# 6. SUMMARY OF PEER REVIEW

During the peer review session there were several good recommendations given to our group. Since the project is fairly simple in concept, most of the ideas expanded past the scope of our initial ideas. There were mostly ideas on how to extend our project or how to extract more audio features. One idea was to dig further into the audio analysis. When starting out with the project, we plan on using attributes mainly for our classification given by our dataset that are already extracted from the raw audio files. One of the other groups suggested some ideas if we were to extract some more features from the audio ourselves. One of their main ideas was doing a Fast Fourier transformation to do frequency analysis on the signals. This is the way many applications like Shazam work and it is an interesting way to deconstruct the audio signal. This could be an interesting way to extend our project if we are unable to do accurate classification using the attributes we already have.

Other ideas that were mentioned to us had to do with the application of what we can do with our final project. Because genres can be fluid categories, they had an idea of using the information we gathered to make music mixes that are dynamic over time. It could be for something like a work out mix where the music starts not intense and then builds up in intensity throughout the workout. There were also ideas about categorizing songs not just into genres but others like ‘danceability’ and ‘energy’ which we can pull from the attributes within the dataset. The other groups also recommended just clustering in general to see what kind of interesting clusters could be formed with the data we have. They recommended the ability to be able to take two songs and look at the similarities of the two. Last of all, the other teams thought it may be interesting to have some actual application derived from our research. They suggested it would be helpful to have a website similar to Pandora but with more extended features and different types of recommendation algorithms. If we have a large amount of time left after completing the project, this could be a stretch functionality for us to implement.

# 7. REFERENCES

1. Goulart Homsi Jose, A., Guido Capobianco, R., and Dias Maciel, C. 2012. Exploring different approaches for music genre classification. *Egyptian Informatics Journal.* 13, 2 (July 2012), 59-63. DOI= <http://dx.doi.org/10.1016/j.eij.2012.03.001>.
2. Tzanetakis, G., Essl, G., and Cook, P. 2001. *Automatic Musical Genre Classification of Audio*. Technical Report. Princeton University.
3. Thierry, B. M., Ellis, D., Whitman, B., and Lamere, P. 2011. The Million Song Dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference* (ISMIR 2011).
4. Yajie Hu and Mitsunori Ogihara, “Genre Classification for Million Song Dataset Using Confidence-Based Classifiers Combination”, *in Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, Portland, Oregon, USA, 2012, pages 1083 – 1084