DDR: Dance Dance Robotics

By, Austin Lawrence (Ablarry91@gmail.com)

EECS-349: Machine Learning, Northwestern University

Abstract

In the interest of achieving noticeable advancement in the field of artificial intelligence, classifying emotion remains a mystifying challenge with potential to connect humans with robots on a more meaningful level. Similar to emotion is the method in which music is conveyed: particularly, the genre. This project focuses towards capturing the genre of unclassified musical files with the ultimate goal of providing an appropriate style in which an accompanying robot can 'dance'. Preprocessed audio data is gathered from the Million Song Dataset and prepared into an SQLite database for subsequent Weka machine learning classifications. With a sub-dataset of 6866 instances over 8 potential genre classifications, a 43.7% accuracy is reported using a Bayes Network approach. It is recommended to seek new musical features through audio processing and explore methods in which similar genres can be more distinctly differentiated.

Background

Musical classification is not a new topic of study, though it remains a problem yet to be satisfiably solved. Several instances of music classification have emerged in research, addressing multiple subtopics concerning the overall motive of emotional classification. For example, Carnegie Mellon's Hideki Kozima developed a small toy-like robot named Keepon (Figure 1), complete with embedded cameras, microphones, speakers, and three actuators that enable the machine to playfully interact with humans. It's found particular success in engaging mentally disabled children and serves as a teaching utility for educators. From a more computational perspective, Roger Dannenberg, also of Carnegie Mellon, has developed an autonomous dancing robot platform stimulated by Hidden Markov Model thresholding, Support Vector Regression classification, and actuation of a NAO humanoid robot. Interestingly, emotion is classified based on an adopted Thayer 2-dimensional Activation-Valence model, and is capable of achieving very high accuracies as a result.

Figure 1: Carnegie Mellon's 'Keepon' musically interactive robot.

Method

Data is gathered using a subset of the Million Song Dataset (MSD). The MSD is based off of a research initiative to accelerate music analysis through the distribution of common musical features. It works closely with the Echo Nest project, a Spotify-acquired organization, which develops musical analysis software. These features are derived from metadata (such as artist name, release date) and analytical (calculated tempo, loudness). The dataset is true to its name in that data for one million songs are available (>250 gigabytes), though a 1% sample is also available for convenience. To date, all analysis has been conducted with the 1%, 10,000 song dataset.

There are over 50 unique attributes to an individual song that could be utilized within this dataset. A number of features consist of large lists and were thrown out. An alternative utilization of these features could be the length of lists, but no analysis has been made in that regard. The features being utilized presently are:

- Duration: the length of a song
- Artist familiarity: an estimation of how familiar the artist is to the world
- Artist hotness: an estimation of how popular an artist is
- Song hotness: an estimation of how popular a song is
- End of fade in: an estimation of when the song finishes fading in
- Start of fade out: an estimation of when the song begins to fade out
- Loudness: Overall loudness of a song in dB
- Mode: whether a song is in major or minor key
- Key: an estimation of the song's key
- Tempo: an estimation of how fast the song is
- Time signature: an estimation of a song's time signature

The information from the MSD is compiled into a SQLite database using Python scripts to query the song, extract features using commands provided by the MSD's API, and adding to the database. From there, the database is viewed through an SQL visualizer and filtered for incomplete and irrelevant data.

To reduce the number of unique genres from which to classify, an associated weight for each tag is provided and was utilized to reduce the quantity of unique genres down to 62. To accomplish this, each genre of each song was evaluated against a frequency table and given precedent to the genre whose product of genre frequency and weight are highest. From this point on, manual classification of the dataset was performed, utilizing Last.FM's genre search tool to seek similarities that may best fit one of 8 possible genre classifications:

- hip hop
- blues
- pop
- electronica
- rock
- jazz
- country
- metal

Should the song being queried differ drastically from the genres listed above, it is thrown out. Musical categories, such as samba, kongo, and alternative were ruled out in the interest of designing a focused dataset for analysis. 6866 instances of the original 10,000 song dataset persisted and were reclassified in a more discrete context. A script was written to convert the SQL database to a Weka-friendly .arff file format, where analysis of machine learning techniques may take place.

Weka was utilized to visualize the data in a graphical context, as well as quickly implement popular machine learning techniques. Analytical features, such as tempo, duration, and loudness are continuous in nature, and were subsequently binned into 50 unique values so as to mitigate the complexity of various classifiers. Multiple classification methods were utilized, including, but not limited to:

- J48 Decision Trees
- Naïve Bayes
- Bayes Networks
- Clustering
- SVM

Results

Table 1 illustrates accuracies for various learning methods and compares results against a ZeroR accuracy of 20.3% (rock).

Tables 2,3, and 4 report the classification results of the best method in this study, a Bayes Network classifier.

Figure 2 displays a relatively typical distribution of data for this testing set. Of particular interest is the reasonable separability of genres electronica and hip hop using features artist_hotness and key.

Table 1: Classification results, 10x cross-validation

Learning Method	Accuracy
Bayes Network	43.7%
J48 Decision Tree	39.7%
Naïve Bayes	33.0%
Clustering	17.4%
SVM	23.1%
ZeroR	20.3%

Table 2: Confusion Matrix of genre classification using Bayes Net classifier.

hip hop	blues	pop	electronica	rock	jazz	country	metal	
383	41	58	69	130	68	21	33	**hip hop**
47	416	72	72	140	87	60	20	**blues**
58	121	378	102	171	41	45	41	**pop**
75	79	92	416	171	65	24	49	**electronica**
89	128	134	121	769	116	58	77	**rock**
50	99	34	76	100	225	41	27	**jazz**
12	95	50	27	81	29	202	9	**country**
45	37	58	56	131	20	14	211	**metal**

Table 3: Bayes Net pattern recognition scores for each class.

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.477	0.062	0.505	0.477	0.49	0.835	**hip hop**
0.455	0.101	0.409	0.455	0.431	0.817	**blues**
0.395	0.084	0.432	0.395	0.412	0.782	**pop**
0.428	0.089	0.443	0.428	0.436	0.795	**electronica**
0.515	0.172	0.454	0.515	0.483	0.76	**rock**

0.345	0.069	0.346	0.345	0.345	0.809	**jazz**
0.4	0.041	0.434	0.4	0.416	0.855	**country**
0.369	0.041	0.452	0.369	0.406	0.826	**metal**
0.437	0.095	0.437	0.437	0.436	0.802	**average**

Table 4: Classification statistics for the Bayes Net classifier.

Correctly Classified Instances	3000	43.7%
Incorrectly Classified Instances	3866	56.3%
Kappa statistic	0.343	
Mean absolute error	0.158	
Root mean squared error	0.303	
Relative absolute error	73.4%	
Root relative squared error	92.3%	
Total number of instances	6866	

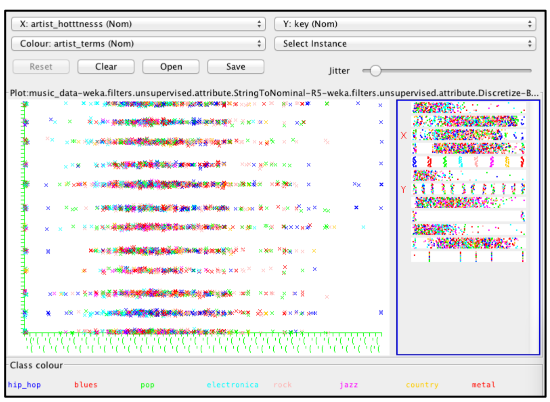

Figure 2: Graphical display of artist hotness against song key.

Discussion

One outstanding issue in this database is the overclassification of genres for a single song. For example, Daft Punk's "Get Lucky" single is provided with 32 genre classification, ranging from "French Disco" to "Alternative" to "Funk Fusion". Among the original 10,000 song instances, a total of 3,502 unique genres and over a quarter million tags were contained within this dataset. The amendment of manual classification may have simplified the learning procedure, though differentiating various genres into a simpler range of one of eight potential values proved to be difficult. For example, given the genre "blue jazz", is it more appropriate to categorize as the musical genre jazz or blues? The results of Table 2 may suggest a flaw in oversimplifying songs that are more appropriately hybrid combinations: for example, the relatively high number of country songs that classified as blues may more appropriately be diagnosed as a new category, bluegrass.

Though the constructed Bayes Network determined the best results for this particular test, it is probable that the provided dataset does not do enough to properly distinguish musical genres for the learning method to be fully effective. Judging from the typical results of Figure 2, distinguishing one class from another in homogenous data like this is very difficult. This is justifiably why clustering saw such poor results. Decision trees were relatively suitable for the application, though constructed unsuitably large trees of greater than 5000 nodes to make these assumptions. Reducing the magnitude of the tree confidence factor does greatly reduce the overall size of the tree, but at the cost of lowering accuracy past ZeroR. Bayes Networks likely performed best due to their full consideration for the distribution of the data in the form of a full joint probability matrix. Utilizing this method appears to best identify the trends that may persist in various genres of music.

Conclusion

Over the course of this study, a 10,000 song dataset of the Million Song Dataset was extracted for analytical features. The classifier, genre, was reduced to a discrete range of 8 potential values through a weighted frequency approach followed by a manual binning procedure with assistance of Last.FM genre correlations. The data was compiled into an SQLite database for ease of viewing and subsequently scripted into a Weka-friendly .arff file format for learning analysis. At a ZeroR accuracy of 20.3% (rock), the best classifier observed was a Bayes Network, at 43.7% accuracy.

Though the results of this study still warrant room for improvement, it does serve as proof that an informed guess can be made as to the genre of a particular song given specific feature criteria. Further study into the topic will be made in hopes of identifying new combinations of musical features that have not been examined in this paper, as well as transitioning work towards the interest of assessing emotions of songs, rather than their specific genre identifier. Doing so will certainly cater towards the holistic interest of developing an interactive dancing robot for human interaction.