
Music Genre Classification

Michael Hagglade

Yang Hong

Kenny Kao

1 Introduction

Music classification is an interesting problem with many applications, from Drinkify (a program that generates cocktails to match the music) to Pandora to dynamically generating images that complement the music. However, music genre classification has been a challenging task in the field of music information retrieval (MIR). Music genres are hard to systematically and consistently describe due to their inherent subjective nature.

In this paper, we investigate various machine learning algorithms, including k-nearest neighbor (k-NN), k-means, multi-class SVM, and neural networks to classify the following four genres: classical, jazz, metal, and pop. We relied purely on Mel Frequency Cepstral Coefficients (MFCC) to characterize our data as recommended by previous work in this field [?, ?]. We then applied the machine learning algorithms using the MFCCs as our features.

2 Our Approach

2.1 Data Retrieval Process

2.2 Mel Frequency Cepstral Coefficients (MFCC)

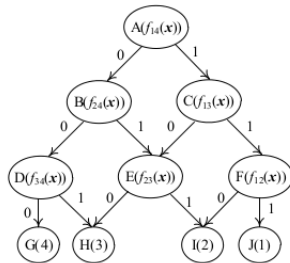
3 Techniques

3.1 Kullback-Lieber (KL) Divergence

3.2 k-Nearest Neighbors (k-NN)

3.3 k-Means

3.4 Multi-Class Support Vector Machine (DAG SVM)



A directed acyclic graph of 2-class SVMs

SVM classifiers provide a reliable and fast way to differentiate between data with only two classes. In order to generalize SVMs to data falling into multiple classes (i.e. genres) we use a directed acyclic graph (DAG) of two-class SVMs trained on each pair of classes in our data set (eg $f_{14}(x)$ denotes the regular SVM trained on class 1 vs class 4) [1]. We then evaluate a sequence of two-class SVMs and use a process of elimination to determine the output of our multi-class classifier.

3.5 Neural Networks

We tried neural networks because it has proved generally successful in many machine learning problems. We first pre-process the input data by combining the mean vector and the top half of the covariance matrix (since it is symmetric) into one feature vector. As a result, we get $15 + (1 + 15) * \frac{15}{2}$

features for each song. We then process the output data by assigning each genre to an element in the set of the standard orthonormal basis in R^4 for our four genres, as shown in the table below:

Genre	Classical	Jazz	Metal	Pop
Vector	(1, 0, 0, 0)	(0, 1, 0, 0)	(0, 0, 1, 1)	(0, 0, 0, 1)

We then split the data randomly by a ratio of 70-15-15: 70% of the data was used for training our neural network, 15% of the data was used for verification to ensure we don't over-fit, and 15% of the data for testing. After multiple test runs, we found that a feedforward model with 10 layers for our neural network model gives the best classification results.

4 Results

Table 1: DAG SVM Results

		Actual			
		Classical	Jazz	Metal	Pop
Predicted	Classical	29	4	1	1
	Jazz	1	20	1	0
	Metal	0	4	26	0
	Pop	0	2	2	29
Accuracy		97%	67%	87%	97%

Table 2: Neural Network Results

		Actual			
		Classical	Jazz	Metal	Pop
Predicted	Classical	14	0	0	0
	Jazz	1	12	4	0
	Metal	0	0	13	0
	Pop	1	0	0	19
Accuracy		88%	100%	76%	100%

Table 3: k-Means Results

		Actual			
		Classical	Jazz	Metal	Pop
Predicted	Classical	14	16	0	0
	Jazz	2	27	1	0
	Metal	0	0	27	3
	Pop	0	1	1	28
Accuracy		88%	61%	93%	90%

Table 4: k-NN Results

		Actual			
		Classical	Jazz	Metal	Pop
Predicted	Classical	26	9	0	2
	Jazz	4	20	4	1
	Metal	0	1	24	0
	Pop	0	0	2	27
Accuracy		87%	67%	80%	90%

Classification accuracy varied between the different machine learning techniques and genres. The SVM had a success rate of only 66% when identifying jazz, most frequently misidentifying it as classical or metal. The Neural Network did worst when identifying metal with a 76% success rate. Interestingly, the Neural Network only ever misidentified metal as jazz. k-Means did well identifying all genres but Jazz, which was confused with Classical 36% of the time. k-NN had difficulty differentiating between Metal and Jazz in both directions. Of its 33% failures identifying Jazz, it misidentifies as Metal 90% of the time. Similarly, k-NN incorrectly predicted that Metal songs would be Jazz in 66% of all its failed Metal identifications.

Overall we found that k-NN and k-means yielded similar accuracies of about 80%. A DAG SVM gave about 87% accuracy and neural networks gave 96% accuracy.

5 Conclusion

5.1 Discussion

5.2 Future Work

TODO alphabetically rearrange sources, and actually add all of them.

References

- [1] *Chen, P., Liu, S.* "An Improved DAG-SVM for Multi-class Classification" <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=0566976>.
- [2] *Marsyas.* "Data Sets" http://marsyas.info/download/data/_sets.