Using Neural Networks for Musical Genre Classification

Cory McKay
Faculty of Music, McGill University
555 Sherbrooke Street West
Montreal, Quebec, Canada H3A 1E3
cory.mckay@mail.mcgill.ca

ABSTRACT

This paper presents a system that automatically classifies MIDI files into hierarchally organized parent genres and sub-genres. A novel configuration of neural networks makes it possible to independently classify files by parent genre and sub-genre before combining the results with a blackboard system. Success rates of 85% for parent genre and 65% for sub-genre were achieved using twenty primarily original features.

KEYWORDS

Genre, style, classification, neural nets, MIDI

INTRODUCTION

Browsing and searching by genre can be very effective tools for users of the rapidly growing networked music archives. The current lack of a generally accepted automatic genre classification system necessitates manual classification, which is both time-consuming and inconsistent.

Most existing studies have focused on accomplishing the difficult task of feature extraction from audio data (Grimaldi et al., 2003; Koshina 2002; Tzanetakis & Cook 2002; Whitman & Smaragdis 2002), with less emphasis placed on pattern recognition techniques. This study is intended to build on previous work by concentrating on the latter aspect of genre recognition. The nebulous and changing nature of genre definitions makes the task well suited to machine learning systems such as neural networks. Blackboard systems provide a useful means of coordinating groups of such systems.

It was decided to classify MIDI files rather than audio files, as beginning with a high-level representation made it possible to focus on classification techniques rather than signal processing. MIDI files are commonly available and it is relatively easy to translate other representations such as Humdrum or GUIDO into MIDI. This makes it possible to apply the system to unrecorded scores if desired. Aside from the work of Shan and Kuo (2003), there has been a paucity of genre research related to MIDI. This paper uses a different approach to genre identification than that of Shan and Kuo, who focused on chord and melody-based features rather than the types of features discussed below.

CLASSIFICATION SYSTEM

Eighty MIDI files were used as training data and a further twenty files were used as test data. The classification of files was based primarily on information from www.allmusic.com. The files were divided hierarchally into three parent genres and nine sub-genres. This allowed a comparison of how well the system distinguished between fairly dissimilar music from the parent genres compared to the more similar sub-genres. The parent genres were Classical (with sub-genres of Baroque, Romantic and Modern Classical), Jazz (with sub-genres of Swing, Funky Jazz and Cool Jazz) and Pop (with sub-genres of Rap, Country and Punk Rock).

A total of twenty features were extracted from each MIDI file (see Table 1). Many of these features were extracted by constructing tempo and pitch histograms of the type developed by Tzanetakis and Cook (2002). The tempo histogram consisted of beats-per-minute bins that were constructed using autocorrelation to derive the frequencies of lags between MIDI note-ons.

The features were classified using an array of feed-forward neural networks that consisted of four networks for identifying parent genre and four networks for identifying sub-genre. Each network had a single hidden layer. This division into two groups made it possible to classify parent genre independently from sub-genre. The input units of each network took in different groups of features, thus making it possible to study the relative success of the different features in classifying the test data. 5000 epochs were used to train each of the networks.

Feature	Explanation
Orchestration	Which of the 128 MIDI instruments are played
Number of instruments	Total number of instruments played
Percussion prevalence	Fraction of note-ons belonging to an unpitched instrument
Dominant pitch prevalence	Fraction of note-ons corresponding to the most common pitch
Dominant pitch class prevalence	Fraction of note-ons corresponding to the most common pitch class
Dominant interval	Number of semi-tones between the two most common pitch classes
Adjacent fifths	Number of consecutive pitch classes separated by perfect 5ths that represent at least 9% of the notes
Pitch class variety (common)	Number of pitch classes that represent at least 9% of the notes
Pitch class variety (rare)	Number of pitch classes played at least once
Register variety	Number of pitches played at least once
Range	Difference between highest and lowest pitches
Pitchbend fraction	Number of pitch bends divided by total number of note-ons
Dominant tempo	Frequency of the highest tempo bin
Second dominant tempo	Frequency of the second highest tempo bin
Combined dominant tempos	Combined frequency of the two highest tempo bins
Dominant tempo strength ratio	Ratio of the frequencies of the two highest tempo bins
Dominant tempo ratio	Ratio of the tempos of the two highest tempo bins
Number strong tempos	Number of tempo bins with normalized frequency > 0.1
Number moderate tempos	Number of tempo bins with normalized frequency > 0.01
Number relatively high tempos	Number of tempo bins with frequencies at least 25% as high as the highest frequency

Table 1: Features extracted from MIDI files and fed into neural networks.

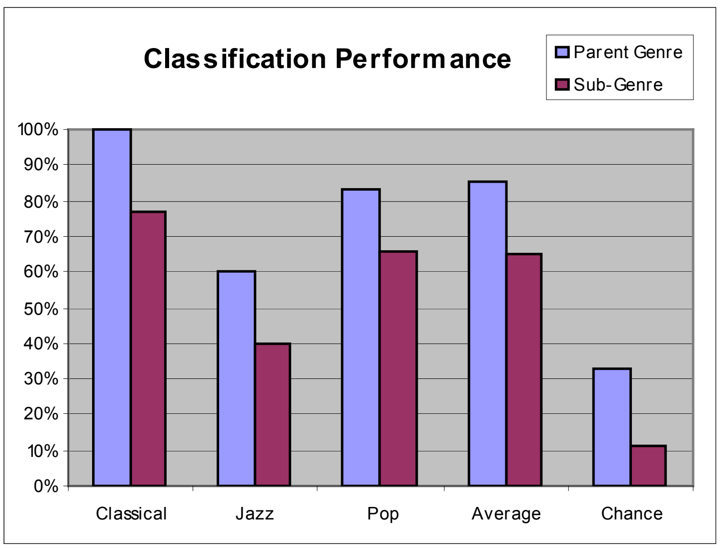

Figure 1: Classification success rates on test set. The sub-genre bars give the average rates of the sub-genres belonging to the corresponding parent genre.

A blackboard system was used to combine and coordinate the selections of each of the networks. This system attempted to ensure that each selected parent genre and sub-genre matched. The blackboard system also suggested alternative classifications when it was not confident with its results.

RESULTS

Each of the networks was moderately successful in identifying genres when considered individually. The blackboard system was effective in improving results, leading to an increase of 15% in the average classification rate as compared to classifications where the selections of each of the networks were simply averaged. The blackboard system also lowered the number of alternative classifications suggested by 28% and improved their accuracy from 33% to 42%when the primary result was wrong.

As can be seen from Figure 1, the test set was classified at a success rate significantly higher than chance in all cases. Classification of classical music was the most successful, mainly because the system often confused Funky Jazz with Country or Rap and vice versa. The average success rate was 85% for parent genre and 65% for sub-genre.

CONCLUSIONS

The classification system performed with success rates comparable to those in the literature. This is encouraging, given the small number of training samples and broad genre categories used. Success under these conditions shows that the system is good at making generalizations when given limited examples. The system configuration was also successful in examining the effectiveness of parent genre classifications relative to sub-genre classifications.

The next step in this research will be to greatly increase both the number of training samples and the number of genre categories. Deeper genre hierarchies will also be considered. Future work will involve experimenting with new features and optimizing their performance with the blackboard system.

REFERENCES

Grimalidi, M., A. Kokaram, and P. Cunningham. 2003. *Classifying music by genre using a discrete wavelet transform and a round-robin ensemble*. Work report. Trinity College, University of Dublin, Ireland.

Koshina, K. 2002. *Music genre recognition*. Diploma thesis. Technical College of Hagenberg, Austria.

Shan, M. K., and F. F. Kuo. 2003. Music style mining and classification by melody. *IEICE Transactions on Information and Systems* E86-D (3). 655-659.

Tzanetakis, G., and P. Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing* 10 (5). 293-302.

Whitman, B., and P. Smaragdis. 2002. Combining Musical and Cultural Features for Intelligent Style Detection. *Proceedings of the International Symposium on Music Information Retrieval*. 47-52.