Efficient Genre Classification using Qualitative Representations

Morteza Dehghani
Northwestern University
EECS Dept 2145 Sheridan Rd
Evanston, IL 60208-0834
morteza@cs.northwestern.edu

Andrew M. Lovett
Northwestern University
EECS Dept 2145 Sheridan Rd
Evanston, IL 60208-0834
andrew@cs.northwestern.edu

Abstract

We have constructed a system that can compute a qualitative representation of music from high-level features extracted from MusicXML files. We use two cognitively motivated computational models called SME and SEQL to build generalizations of musical genres from these representations. We then categorize novel music pieces according to the generalizations. We demonstrate the feasibility of the system with training sets much smaller than those used in previous systems.

Keywords: Genre Classification, Symbolic Representation of Music.

1. Introduction

The problem of automatic genre classification has received a lot of attention in recent years, as it plays an important role in the field of music information retrieval [5]. The main approach to this problem has been focused on using low-level features extracted from audio files as a starting point for classification. At present, there is no precise way to extract many high-level features from polyphonic audio [5,6].

With only a small amount of computation, music features can be easily extracted from symbolic formats such as MusicXML. There have been several attempts to perform genre classification using high-level features extracted from symbolic formats [1,5,7]. McKay and Fujinaga [5] used a corpus of 950 MIDI files and achieved a success rate of 98% for 3-way classification. Chai and Vercoe [1] trained hidden Markov models on a corpus of 491 pieces and achieved the performance of 65%-77% for 2-way classification. Shan and Kuo [7] achieved an accuracy of 70%-84% for 2-way classification. One disadvantage of these systems is that they require a large database of files for training. This requirement is a result of the machine learning algorithms used by their systems.

In this paper, we suggest an alternative method for representing, comparing, and classifying pieces of music. Our system builds qualitative representations of musical pieces based on high-level features in symbolic MusicXML files. It compares two pieces through a process called structural alignment, which is based on a psychological model of analogy and similarity in humans [3]. It learns to classify pieces by building generalizations for each genre containing the structure found in most or all the examples of that genre.

2. System Description

The music conversion pipeline for our system starts with MIDI files. After converting MIDI's to MusicXML's, the following information is extracted from the files: the instrument name, part name, time signatures, and key signatures for each track; and the chord names.

2.1 Qualitative Representation (QR)

We use the data extracted from the MusicXML files to produce a qualitative representation of each musical piece. The representation contains an entity for each track in the musical piece. The entity receives attributes representing the instrument, time signature, and key signature for that track. The system also encodes the chord intervals in the piece. These intervals represent the distance from key name to root pitch of the chord.

2.2 Comparison and Generalization

We compare representations using the Structure-Mapping Engine (SME) [2]. SME is a computational model of similarity and analogy based on Gentner's [3] structure-mapping theory of analogy in humans. SME takes as input two cases. It finds all possible correspondences between entities, attributes, and relations in the two cases. It combines consistent correspondences to produce mappings between the cases.

Our system learns categories of objects using SEQL, a model of generalization built on SME [4]. The idea behind SEQL is that humans form a representation of a category by abstracting out the common structure in all the exemplars of that category. In its default mode, SEQL works in the following way: when it encounters a new case, it uses SME to compare that case to the known generalizations. If the new case aligns with a sufficient amount of the structure in one of the generalizations, the case is added to that generalization. Any part of the generalization's structure that does not align with the new case is removed, so that the generalization continues to represent only the structure found in all of its exemplars.

A recent update to SEQL associates a probability with each fact in the generalization. When a new case is added to a generalization, those parts of the generalization that do not align with the case are not automatically removed, but instead have their probability decreased.

3. Experimental Section

Our dataset includes 85 MIDI files. These files are from two root genres: pop and classical. There are two leaf genres under the pop root genre: Rock and Bluegrass; and three leaf genres under the classical root genre: Baroque, Classical, and Romantic. There are 17 pieces for all leaf genres.

3.1 Evaluation Method

To evaluate the system, we randomly divided the files into a test and a training set 20 different times. The training sets contained 9 pieces from each genre, and the test sets contained the remaining 8. In each run, our system used SEQL to produce a single generalization for all the pieces in each leaf genre's training set. Pieces in the test set were classified by using SME to compare each piece's representation to all of the genre generalizations and returning the genre that matched most closely. Matches were scored according to what percentage of expressions in the generalization were matched by SME to expressions in the test piece's representation.

Two results were measured: successful classification into the correct leaf genre, and classification into any genre falling under the correct root genre. The results were averaged over all 20 trials.

3.2 Results

Our best results were achieved when only instrument names and chord intervals were encoded. With this representation scheme, our system was able to classify a musical piece into the correct root genre 88% of the time and into the correct leaf genre 58% of the time. Performance using only instrument names was identical for root genre classification but slightly lower for leaf genre classification, suggesting that chord intervals provided some additional information that helped to distinguish between leaf genres within a root genre

Finally, because our focus was on learning from a minimal amount of data, we evaluated the performance when the size of the training set was decreased. Table 1 shows the results when the training set size varied from 2 pieces per genre to 9 pieces per genre. Interestingly, performance was well above chance even with only 2 pieces for each genre. Performance increased as the training set size increased. However, it appeared to level off at about 8 pieces per genre. This may indicate that, at least for our dataset, there is no significant advantage to training on more than half of the 17 pieces in each genre.

4. Conclusions and Future Work

Our system was able to efficiently learn generalizations based on training sets up to 1/10 smaller than training sets

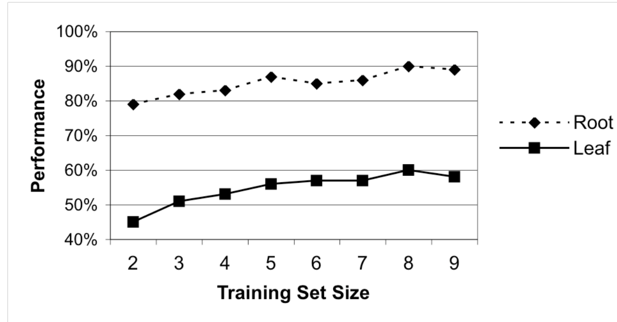

Table 1. Performance with different training sets

used in systems which utilize other machine learning techniques [1,5], and still achieve an accuracy of 88% and 58%, respectively, for 2-way and 5-way classification. Even when we further limited the training set size to as few as 2 pieces per genre, we were still able to achieve accuracy rates of 79% and 45%, levels comparable to other systems. We believe our system can be useful for classifying genres in situations where there is a constraint on the number of music pieces in the training set.

5. References

[1] Chai, W. and B. Vercoe. 2001. Folk music Classification using Hidden Markov Models. In *Proceedings of the International Conference on Artificial Intelligence*.

[2] Falkenhainer, B., Forbus, K. and Gentner, D. 1989. The Structure-Mapping Engine: Algorithms and Examples. *Artificial Intelligence* 41: 1-63

[3] Gentner, D. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 7: 155-170.

[4] Kuehne, S., Forbus, K., Gentner, D. and Quinn, B. 2000. SEQL: Category Learning as Progressive Abstraction Using Structure Mapping. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 770-775. Philadelphia, PA.

[5] McKay, C., & Fujinaga, I. 2004. Automatic Genre Classification Using Large High-Level Musical Feature Sets. In *Proceedings of the 5th International Conference on Music Information Retrieval*. Barcelona, Spain.

[6] Pampalk, E., Flexer, A., & Widmer, G. 2005. Improvements of Audio-Based Music Similarity and Genre Classification. In *Proceedings of the 6th International Conference on Music Information Retrieval.* London, UK.

[7] Shan, M. K., and F. F. Kuo. 2003. Music style Mining and Classification by Melody. *IEICE Transactions on Information and Systems E86-D (3)*: 655–69.