Folk Music Classification Using Hidden Markov Models

Wei Chai
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA, U.S.A.

Barry Vercoe
Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA, U.S.A.

Abstract

Automatic music classification is essential for implementing efficient music information retrieval systems; meanwhile, it may shed light on the process of human's music perception. This paper describes our work on the classification of folk music from different countries based on their monophonic melodies using hidden Markov models. Music corpora of Irish, German and Austrian folk music in various symbolic formats were used as the data set. Different representations and HMM structures were tested and compared. The classification performances achieved 75%, 77% and 66% for 2-way classifications and 63% for 3-way classification using 6-state left-right HMM with the interval representation in the experiment. This shows that the melodies of folk music do carry some statistical features to distinguish them. We expect that the result will improve if we use a more discriminable data set and the approach should be applicable to other music classification tasks and acoustic musical signals. Furthermore, the results suggest to us a new way to think about musical style similarity.

Keywords

Music classification, hidden Markov model, music perception

1. Introduction

There are both scientific and practical reasons for building computer systems that can identify music style. We are interested in finding out how a piece of music is put together, what aspects of music characterize a style, and what it is that distinguishes the musical sound of one culture from that of another. However, there are currently no developed scientific theories about how humans can make rapid judgment about the music's style from very short examples [11], yet there are many applications in which music style identification by computer would be useful. For example, we would like to build computer systems that can annotate musical multimedia data, which will benefit music information retrieval; we would also like to build systems for music theory study and teaching. By building such systems, we can learn a great deal about how the human system accomplishes this task.

This paper describes our work on the classification of folk music from different countries based on their monophonic melodies using hidden Markov models. The goals of this research are: (1) to explore whether there exists significant statistical difference among folk music from different countries based on their melodies; (2) to compare the classification performances using different melody representations; (3) to study how HMM's perform for music classification as a time series analysis problem.

The remainder of the paper is organized as follows. Section 2 describes the data set, representations, models and algorithms. Experimental results obtained for the classification of folk music from three different countries are presented in section 3. Section 4 presents our explanations based on the results. Conclusions are drawn and some future work is proposed in section 5.

2. Approach

2.1 Data Set

We chose folk music as our experiment corpus, because (1) most folk music pieces have obvious monophonic melody lines which can be easily modeled by HMM. "Monophonic" here means only one single tone is heard at a time, and there is no accompaniment or multiple lines going simultaneously. (2) melodies of folk music from different countries may have some significant statistical difference, which should be able to be captured by an appropriate statistical model.

Bruno Nettl mentioned in his book [9] that melody is the aspect of music that has been of the greatest interest to folk music study, but is also probably the most difficult part.

In the classification experiment, 187 Irish folk music pieces, 200 German folk music pieces and 104 Austrian folk music pieces were used. We don't have specific reasons for choosing these three countries except for the availability of data. The data were obtained from two corpora:

(1) Helmut Schaffrath's Essen Folksong Collection which contains over 7,000 European folk melodies encoded between 1982 and 1994;
(2) Donncha Ó Maidín's Irish Dance Music Collection.

All of them have monophonic melodies encoded in either of the two symbolic formats: **kern (a subset of Humdrum format) and EsAC (Essen Associative Code).

We developed a tool to extract the pitch and duration information from files in the above formats based on the CPN View implemented by University of Limerick [8]. CPN View (Common Practice Notation View) is a library in C++ for manipulating representations of notated scores. Currently, CPN View supports symbolic formats including ALMA (Alphameric Language for Music Analysis), **kern, EsAC and NIFF (Notation Interchange File Format).

2.2 Representations

The most obvious way to convert each melody into a sequence is using the pitch sequence. However, there are two problems to consider: (1) Should we use the absolute pitch sequence or the interval sequence? (2) Should we incorporate rhythmic information and how to do so if necessary?

In the experiment, we represented melodies in four different ways.

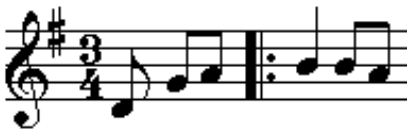
(A) Absolute pitch representation. A melody is converted into a pitch sequence by normalizing the pitches into one octave from C4 (Middle C) to B4, that is, pitches in different octaves but of same chroma are encoded with the same symbol in the sequence and thus there are totally 12 symbols.

(B) Absolute pitch with duration representation. To incorporate rhythm information, we make use of the concept behind the representation for rhythmic sequences employed in [1]. Briefly, we simply repeat each note multiple times to represent the duration, e.g., in our experiment, how many half-beats the note lasts.

(C) Interval representation. A melody is converted into a sequence of intervals, which mean the difference of the current note and the previous note in semitones. There are 27 symbols indicating –13 to 13 semitones (intervals larger than 13 semitones are all indicated by +/-13).

(D) Contour representation. This is similar to Interval representation, but we quantize interval changes into five levels, 0 for no change, +/- for ascending/descending 1 or 2 semitones, ++/-- for ascending/descending 3 or more semitones. Thus, there are totally 5 symbols. This representation is fairly compact and fault-tolerant in melody identification applications [6].

For example,



The sequence in representation A will be {2,7,9,11,11,9}. The sequence in representation B will be {2,7,9,**11**,**11**,11,9}. The sequence in representation C will be {5,2,2,0,-2}. The sequence in representation D will be {++, +, +, 0, -}.

2.3 HMM's

HMM (Hidden Markov Model) is a very powerful tool to statistically model a process that varies in time. It can be seen as a doubly embedded stochastic process with a process that is not observable (hidden process) and can only be observed through another stochastic process (observable process) that produces the time set of observations. A HMM can be fully specified by (1) N, the number of states in the model; (2) M, the number of distinct observation symbols per state; (3) $A = \{a_{ij}\}$, the state transition probability distribution; (4) $B = \{b_j(k)\}$, the observation symbol probability distribution; and (5) $\Pi = \{\pi_i\}$, the initial state distribution. [10]

Because the number of hidden states and the structure may impact the classification performance, we use HMM's with different number of hidden states and different structures to do classification, and then compare their performances. Here are the different structures used and compared in our experiment (Figure 1).

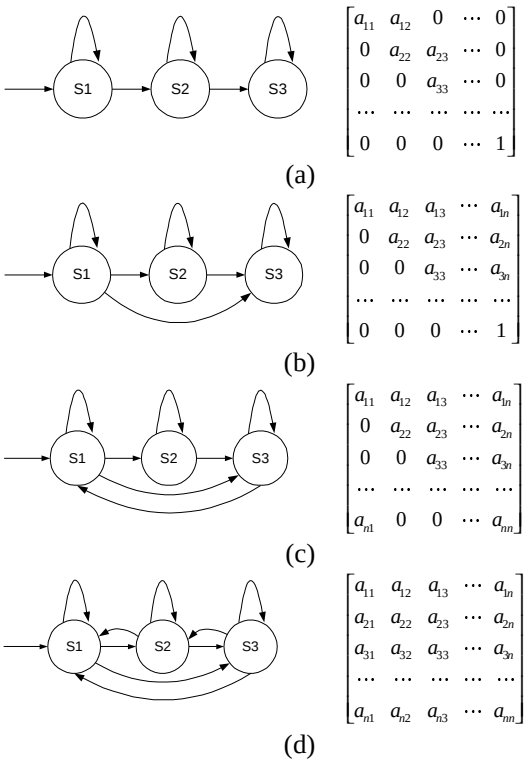


Figure 1: *HMM's used in the experiment (a) A strict left-right model, each state can transfer to itself and the next one state. (b) A left-right model, each state can transfer to itself and any state right to it. (c) Additional to (b), the last state can transfer to the first state. (d) A fully connected model.*

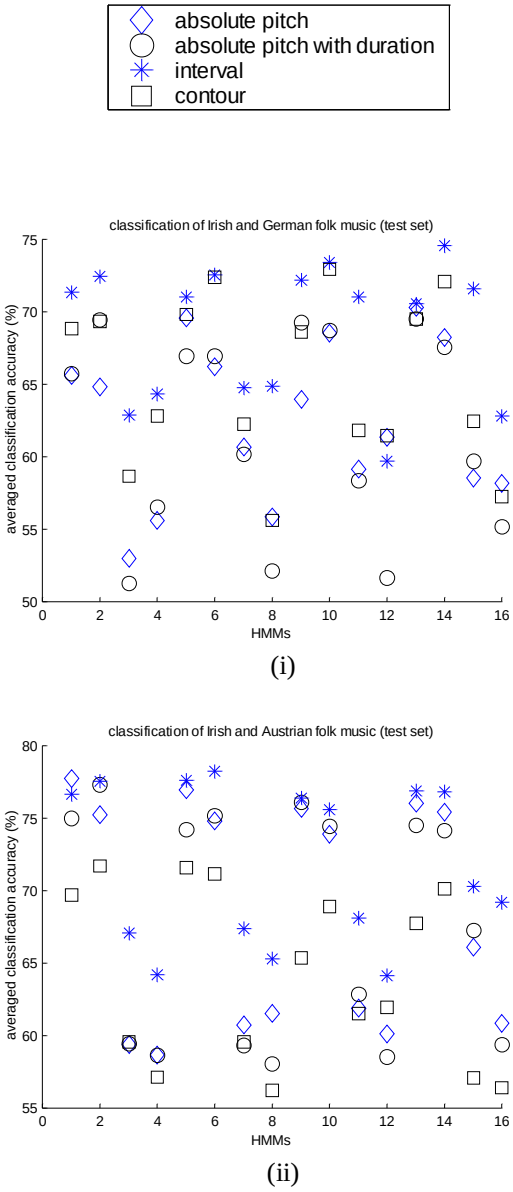
2.4 Classification Algorithm

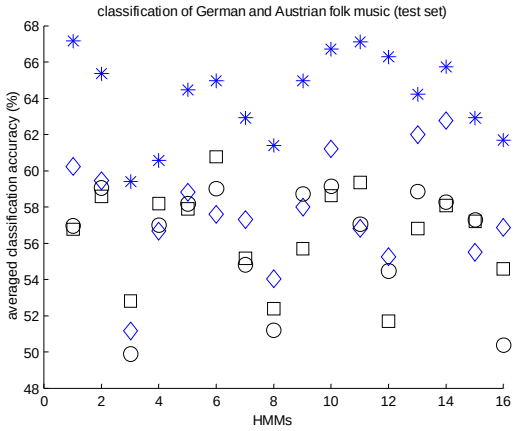
The Baum-Welch reestimation method was implemented to train a hidden Markov model for each country using the training set. To identify the country of an unknown melody, the Viterbi algorithm was used to decode the sequence and compute its log probabilities respectively using HMM's trained for different countries. We then assign the melody to the country with the highest log probability.

3. Results

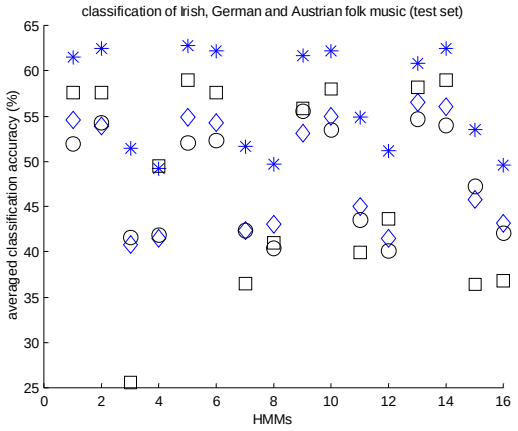
The following results were obtained in this way: All the data were split randomly into training set (70%) and test set (30%). Each result was cross-validated with 17 trials using the 70%/30% splits.

The classification performances are shown in Figure 2 (i) – (iv). The first three figures show the generalization performances of 2-way classifications. The last figure shows the generalization performance of 3-way classification. The X-axis in all the figures indicates different HMM's, whose corresponding structures are shown in Table 1.





(iii)



(iv)

Figure 2: *Classification performances using different representations and HMMs. (i), (ii) and (iii) correspond to 2-way classifications. (iv) corresponds to the 3-way classification.*

Table 1: *16 HMM's with different structures and number of hidden states used in the experiment (see Figure 1 for the description of the structures).*

HMM	1	2	3	4	5	6	7	8
N	2	2	2	2	3	3	3	3
STRUC	a	b	c	d	a	b	c	d
HMM	9	10	11	12	13	14	15	16
N	4	4	4	4	6	6	6	6
STRUC	a	b	c	d	a	b	c	d

The results show that, in general, the state number (2, 3, 4 or 6) didn't impact the classification performance significantly. The strict left-right HMM's (a) and the left-right HMM's (b) outperformed the other two HMM's (c/d). The representation C generally performs better than the representation A, B or D. The performances of 6-state left-right HMM, for example, are shown in Table 2. It achieved classification accuracy of 75%, 77% and 66% for 2-way classifications and 63% for the 3-way classification using representation C.

Table 2: *Classification performances of 6-state left-right HMM using different representations. The first three rows correspond to 2-way classifications. The last row corresponds to the 3-way classification. I: Irish music; G: German music; A: Austrian music.*

Classes	rep. A	rep. B	rep. C	rep. D
I-G	68%	68%	75%	72%
I-A	75%	74%	77%	70%
G-A	63%	58%	66%	58%
I-G-A	56%	54%	63%	59%

4. Discussions

The performances of 2-way classifications are consistent with our intuition that German folk music and Austrian folk music are less discriminable between each other than those with Irish folk music. Therefore, we expect that the result will improve if we use a more discriminable data set.

The results suggest to us a new way to think about musical style similarity. Nettl [9] pointed out that it is very hard to state concretely just how much difference there is between one kind or style of music and another. As he suggested, one way of telling that a musical style is similar to another one, the second of which you already recognize, is that the first of the styles also appeals to you. This has to do with the fact that folk music styles, like languages, exhibit greater or lesser degrees of relationship. Just as it is usually easier to learn a language that is closely related in structure and vocabulary to one's own, it is easier to understand and appreciate a folk music style similar to one that is already familiar. Here we presented a method to measure the musical style similarity quantitatively. The two styles that are less discriminable in classification are deemed more similar. It can be based on the classification accuracy, as was done here, or the

distance of their statistical models directly, for example, the distance of two HMM's [5].

The representation is very important for classification. The choice between the absolute pitch representation and the interval representation is also consistent with humans' perception of melody. Although the absolute pitch method can represent the original work more objectively, the interval method is more compatible with human's perception, since when people memorize, distinguish or sing a melody, they usually do it based only on the interval information.

The experiment shows that the contour representation was significantly worse than the interval representation for folk music classification. This indicates that although contour-based representation is fairly compact for identifying a melody [6], the quantization procedure may cause the features for style discrimination to be reduced.

The fact that the representation with duration did not outperform the representation without duration is not what would be expected. It seems to be inconsistent with humans' perception. We argue that it doesn't mean rhythmic information is useless for classification; instead, we suggest that the rhythmic encoding used (through repeated notes) in fact destroyed some characteristics of the melody, thus reducing the discrimination.

5. Conclusions

It has been shown how hidden Markov models could be used to build classifiers based on melody information of folk music. Folk music from different countries does have significant statistical difference in their respective melodies. The interval representation generally performs better than the absolute pitch representation or the contour based representation.

For further research, our approach should be applicable to other music classification tasks, for example, classifying music by different composers, of different ages or genres, etc. Furthermore, we will explore the possibility of combining our method in symbolic representations with signal processing techniques to build music classification systems on acoustic musical signals. One thing we need to mention at the end is that although melody is an important feature, it is not sufficient for music classification on its own. The classification performance can be greatly improved if we combine other significant features, for example, harmony, instrumentation, performance style, etc.

References

[1] Carpinteiro, Otávio Augusto S. and Itajubá, Escola Federal de Engenharia de Itajubá. *"A self-organizing map model for analysis of musical time series."* In Proc. Vth Brazilian Symposium on Neural Networks, 1998.

[2] Dannenberg, Roger; Thom, Belinda and Watson, David. *"A machine learning approach to musical style recognition."* In Proc. International Computer Music Conference, 1997.

[3] Foote, Jonathan. "*Methods for the automatic analysis of music and audio*". FXPAL Technical Report FXPAL-TR-99-038.

[4] Foote, Jonathan. *"An overview of audio information retrieval"*. In Multimedia Systems, vol. 7 no. 1, pp. 2-11, ACM Press/Springer-Verlag, January 1999.

[5] Juang, B. and Rabiner, L. "*A Probabilistic Distance Measure for Hidden Markov Models*". The Bell System Technical Journal, vol. 64, pp. 391—408, 1985.

[6] Kim, Youngmoo; Chai, Wei; Garcia, Ricardo and Vercoe, Barry. *"Analysis of a contour-based representation for melody"*. In Proc. International Symposium on Music Information Retrieval, Oct. 2000.

[7] Lomax, Alan. *Folk Song Style and Culture.* American Association for the Advancement of Science, 1968.

[8] Maidín, Donncha Ó. "*Common practice notation view users' manual*". Technical Report UL-CSIS-98-02, University of Limerick.

[9] Nettl, Bruno. *Folk and Traditioanl Music of the Western Continents.* 2d ed. Prentice-Hall, 1973.

[10] Rabiner, Lawrence R. *"A tutorial on hidden Markov models and selected applications in speech recognition"*. In

Proc. of the IEEE Volume: 77 2, Feb. 1989, Page(s): 257-286.

[11] Scheirer, Eric D. *Machine-Listening Systems.* Unpublished Ph.D. Thesis, Massachusetts Institute of Technology, 2000.

[12] Selfridge-Field, Eleanor. *"Conceptual and representational issues in melodic comparison"*. In Melodic Similarity, Concepts, Procedures, and Applications, MIT Press, 1998.

[13] Tzanetakis, George and Cook, Perry. *"A framework for audio analysis based on classification and temporal segmentation"*. In Proc. Euromicro, Workshop on Music Technology and Audio processing, Milan, September 1999.

[14] Wold, E.; Blum, T.; Keislar, D. and Wheaten, J. *"Content-based classification, search, and retrieval of audio."* IEEE Multimedia Volume: 33, Fall 1996, Page(s): 27–36.