# MUSIC STYLE MINING AND CLASSIFICATION BY MELODY

*Man-Kwan Shan, Fang-Fei Kuo, Mao-Fu Chen*

Department of Computer Science
National Cheng Chi University
Taipei, Taiwan, ROC

**ABSTRACT**

Music style is one of the features that people used to classify music. Discovery of music style is helpful for the design of content-based music retrieval system. In this paper we investigated the mining and classification of music style by melody from a collection of MIDI music. We extracted the chord from the melody and investigate the representation of extracted features and corresponding mining techniques for music classification. Experimental results show that the classification accuracy achieved about 70% to 84% for 2-way classification.

## 1. INTRODUCTION

With the development of multimedia technology, digital music is now in widespread use. Content-based music retrieval (CBMR) has attracted much interest in recent years. CBMR allows users query by music content rather than metadata. However, even with the capability of query by humming [5,7], the effectiveness of CBMR system suffers from the ability of query content expression for people without music training.

Music style is one of the features that people used to classify music. In human perception of music, the music style is the semantic description of feeling. For example, Chopin's music is romantic and the New Age music feels peaceful. What music features make the romantic feeling? Can we find the syntactic description of the New Age music? In this paper, we investigate the data mining approach to find out the syntactic description of music style. Our work makes the following applications:

(1) It can be used in the personalized CBMR system based on users' preference [9]. CBMR system can learn users preferred music style by monitoring the users' retrieval activities and discovering the syntactic patterns from the accessed music. The retrieval effectiveness could be improved by considering the users preference.
(2) With the discovered syntactic description of music style, it is possible to perform the automatic classification and annotation of music style.
(3) It allows users to query music by style in spite of the composer. For example, the user can search for the music with both Bach and Mozart style composed by other composers.
(4) It helps the automatic computer composer to compose specific style music according to the discovered patterns.

There are several essentials that influence the styles and feeling of music, such as melody, instrument, rhythm and tempo. Above all, melody makes music memorable and enables people to distinguish one work from another [10].

In this paper we investigated the discovery of melody style from a collection of music recorded as MIDI. In particular, we extracted the chord from the melody based on the harmony. We investigated different representations to describe melody feature and present corresponding patterns to discover the music style. Then, classification of music is performed based on the discovered patterns.

## 2. RELATED WORK

Little work has been done on the discovery of music style. The work developed in MIT Media Lab. [3] employed hidden Markov model to model and classify the melodies of Irish, German and Austrian folk song. Melodies are represented as a sequence of absolute pitches, absolute pitches with duration, intervals and contours. The classification performances achieved 65% to 77% accuracy with the interval representation. Another research in CMU uses the naïve classifier, linear and neural network respectively to recognize music style for interactive performance systems [4]. Thirteen statistical features derived from MIDI are identified for learning of music style.

## 3. MELODY MINING AND CLASSIFICATION

### 3.1. Rationale

There are three issues need to be considered for melody mining and classification. One is the melody feature extraction. Another is the representation of the extracted features. The other is the mining and classification technique.

Some possible features for melody include the sequence of pitches, the interval contour, and the distribution of pitch.

0-7803-7304-9/02/$17.00 ©2002 IEEE

However, these features are not adequate for music style mining. For example, Figure 1 shows two measures of melody from two music segments. These two segments are similar in terms of either melody sequence or interval contour. However, they sound quite different.

We propose the approach to utilize the chord based on the harmony to extract the features for melody style mining. The harmony is depends on the melody. It is the chords with which a melody was accompanied. Therefore, we can find the chords according to the melody sequence.


Figure 1. Examples of segments and the assigned chords.

A chord is a combination of three or more notes which sound simultaneously. Triad and 7th chord are two common used chord forms. A triad is composed of three different notes, and there are four triad chord types – major, minor, diminished and augmented triad. The 7th chords consist of four notes and there are five types of 7th chords commonly used in harmony – dominant, major, minor, half diminished and fully diminished 7th chords.

The chords on the scale notes can be related to the positions in the scale. If we use the representation of the chords described above, such as C, $E_m$, the collocation chords of a song would be different when this song modulates. To solve this problem, in musicology, Roman numerals – Ⅰ, Ⅱ, Ⅲ, Ⅳ,…Ⅶ are used to represent note name of the chords respective to the scale. For example, triads in C major scale are I, Ⅱm, Ⅲm, Ⅳ…Ⅶdim.

### 3.2. Melody Extraction

MIDI contains more than one channel, and there are a number of notes which sound at the same time. Four melody extraction methods have been proposed in [11]:

(1) All-mono: Merge all channels and remove all notes which sound simultaneously except for the highest note.
(2) Entropy-channel: Keep the highest note of each channel, and select the channel with largest entropy.
(3) Entropy-part: Use heuristic to segment each channel into parts and select the highest entropy parts.
(4) Top-channel: Keep channel with highest average pitch and remove all other notes sound simultaneously except for the highest note.

Experimental result has shown that all-mono algorithm is more accurate among these four methods [11]. However, all-mono algorithm does not consider some useful information contained in the MIDI files, such as the instrument or volume. We improve the all-mono to get more precise melody sequence; the modified method contains three steps:

A. Remove channels of instruments, which are unlikely for melody performing. For example, drum and cymbal only perform fixed pitch and we can remove these channels.
B. For each measure, select the channel of the largest volume. The reason is that volume of melody is usually the largest, and it does not change channel constantly, so melodies of first-half and second-half measure are in the same channel.
C. For the selected channel of each measure, keep the highest note while several notes sound simultaneously.

In Figure 2, the first staff shows an example of a music segment [11]. The second staff shows the extracted melody using all-mono. Our method extracts the accurate melody, which is shown in the third staff of Figure 2.


Figure 2. Examples for melody extraction.

### 3.3. Chord Assignment

To match up the chords and melody, we develop the chord assignment method based on the music theory and guitar chords progression rules. We select 60 common chords introduced in section 3.1 as the candidates, and count score of each candidate according to following steps:

**Stage 1**: Each score of the 60 chords is initialized to zero. Then the chord sampling unit is determined. First, find the prevailing note, which has the longest total performance length. If the prevailing note is half note, quarter note, and eighth note, the sampling unit is four measures, two measures and one measure respectively. And then get the scale K of the song and add ten points to the chord Ⅰ of the sampling unit which sounds first.

**Stage 2**: For each sampling unit, do the following steps.

A. For each kind of pitch, add one point to those chords that contain it.
B. If total duration of the longest pitch in this unit is longer than half duration of this unit, add two points to the chords that contain this pitch. Otherwise, add one point to these chords.
C. If there are no sharps or flats in the unit, discard the chords except the chord Ⅰ which contains sharps and flats, and vice versa.

**Stage 3**: For each sampling unit, if the chord with the highest score is unique, we assign it to the sampling unit; otherwise take the relationship between adjacent units into consideration and do the following steps. Let $P$ denote chord of the previous unit.

A. Add two points to the chords whose root is descending fifth, descending third or ascending second of that of *P*.
B. If *P* belongs to set of Dominant 7 chords, add points according to the following rules:
  a. If *P* is chord Ⅰ7, add two points to the Ⅳ chord.
  b. If *P* is chord Ⅱ7, add two points to the Ⅴ chord.
  c. If *P* is chord Ⅲ7, add two points to the Ⅵ chord.
  d. If *P* is chord Ⅳ7, add two points to the Ⅴ chord.
  e. If *P* is chord Ⅴ7, add two points to the Ⅰ and Ⅵ chords, also add one point to the V chord.
  f. If *P* is chord Ⅵ7, add two points to the II chord.
  g. If *P* is chord Ⅶ7, add two points to the Ⅲ chord.
C. If the chord with the highest score is unique, this chord is the answer; otherwise do the following steps.
  a. Find the lowest pitch in the sampling unit, add two points to the chords whose root is the lowest pitch.
  b. Assign the chords with the highest score to this unit.

Note that it is possible to assign a set of chord, named as the *chord-set*, to a given unit. Figure 1 shows the result of chord assignment using our selection method.

### 3.4. Melody Feature Representation

After determining the chord-set respective to each sampling unit, melody can be treated in the following different ways.
(1) Set: melody is represented as a set of chord-sets.
(2) Bi-gram: melody is represented as a set of bi-grams of chord-sets. A bi-gram is an adjacent pair of chord-sets extracted from a sequence of chord-sets.
(3) Sequence: melody is represented as a sequence of chord-set. In this way, a melody with *n* units is actually an *n*-gram.

Moreover, in terms of music form, melody may include the introduction, verse, chorus, interlude, bridge and ending. The chorus usually repeats several times and is most memorable. We also extract the chorus which is the repeating pattern in a melody sequence. The extracted chorus is therefore associated with the chord-sets and represented in the above ways the same as the whole melody. We implemented the repeating pattern finding algorithm proposed in [6]. This algorithm constructs the correlative matrix from melody sequence and uses this matrix to find out the repeating patterns.

### 3.5. Melody Mining

**Frequent Itemset** If the feature of a melody is represented as a set of chord-sets (or a set of bi-grams of chord-set), to obtain the interesting hidden relationships between chords and music styles, we utilize the concept of frequent itemset in the association rule mining [2]. In our work, each item is a chord set or a bi-gram of chord sets respective to each sampling unit. An itemset is frequent if percentage of the number of training songs that contain this itemset to the total number of songs is no less than a threshold called as *minimum support*. Assume there is a frequent itemset { Ⅰ, Ⅵ, Ⅴ} of the lyric-style music, this represents that a great part of lyric-style music consist of chords I, Ⅵ and Ⅴ chords simultaneously. We implemented the Apriori algorithm [1] to find the frequent itemsets.

**Frequent Substring** If the feature of melody is represented as a sequence of chord-sets, to find the ordered patterns, we propose a new pattern, *frequent substring*, by modifying the concept of *sequential patterns* [2] in sequence data mining techniques. A substring is frequent if percentage of the number of training songs that is the superstring of this substring to the total number of training songs is no less than the minimum support. The frequent substring differs from the sequential patterns in that frequent substring is consecutive. We modified the join step of the Apriori-based sequential mining algorithm to find frequent substring.

### 3.6. Melody Classification

Our music classification method adopts the approach of associative classification [8]. In associative classification, a rule is of the form $l \Rightarrow y$, where $l \in \bigcup_k L_k$, *l* is a frequent itemset and *y* is a class. A rule has confidence *c* if *c*% of training samples that satisfy *l* belong to class *y*. A training sample conforms to *l* if its corresponding feature *m* contains *l*. In our work, if the discovered pattern is frequent substring, then *l* is a frequent substring and *m* conforms *l* if *m is* the superstring of *l*.

The classification contains two steps: First, all the frequent itemsets or substrings are found. Then, a classifier is constructed several passes over the training data using heuristic method. Format of the classifier is $< r_1, r_2, ..., r_n, default\_class >$, where rules $r_i$ are ranked by the confidence and support. The first rule that satisfies the test data is used to classify it. If there are no rules satisfying the test data, the test data is classified according to the default_class.

## 4. EXPERIMENTAL RESULTS

### 4.1. Experiment Set Up

We collected three categories of MIDI files from the Internet. One is the music of New Age singer Enya collected from various web sites of Enya. Another is the album of Beatles accessed from http://www.geocities.com /SunsetStrip/Studio/7779/. The other is the Chinese folk song downloaded from http://ingeb.org. After pruning some files incompatible with our MIDI parser, each category keeps 35 to 50 MIDI files. Average length of the Enya, the Beatles, and the Chinese folk songs is 74.1, 79 and 24.5 sampling units respectively.

### 4.2. Performance Analysis

The performance of our experiment is measured by the accuracy of the 2-way classification. The accuracy is defined as the percentage that we classify the test songs to correct category. To measure the accuracy of our classification method, 5-fold cross-validation is utilized. In other words, each category of songs is randomly partitioned into five equal-sized mutually exclusive subsets (folds). Training and testing is performed five times. In each time one of the folds is selected as the test set while the other four folds are collected to derive the classifier. The accuracy is therefore measured as the average accuracy over the five tests.

Factors that affect the performance include minimum support, unit of feature extraction (whole melody or chorus), and feature representation (set, set of bi-gram, or sequence). We made the experiments by considering the effect of these factors. Table 1 shows the result where A, B, C denotes the category of the Enya, the Beatles and the Chinese folk songs respectively. Note that in Table 1, no results are shown for the bigram feature with minimum support larger than 10%. In these cases, no bigram rules are generated and the accuracy is 0%.

Table 1. Performance of the experiment.

| | Minimum Support | Whole Melody | | | Minimum Support | Chorus | | |
|---|---|---|---|---|---|---|---|---|
| | | A–B | A–C | B–C | | A–B | A–C | B–C |
| Set | 20% | 63.0% | 82.1% | 80.0% | 20% | 68.6% | 51.9% | 61.4% |
| | 30% | 66.0% | 77.7% | 76.4% | 30% | 68.3% | 53.5% | 59.4% |
| | 40% | 67.2% | 72.5% | 77.3% | 40% | 63.5% | 53.5% | 55.3% |
| | 50% | 63.8% | 69.3% | 63.6% | 50% | 57.7% | 53.5% | 51.5% |
| Bi-gram | 10% | 64.2% | 81.8% | 75.5% | 10% | 51.4% | **59.6%** | 52.4% |
| | 20% | 66.2% | 73.4% | 69.1% | 20% | - | - | - |
| | 30% | 62.1% | 46.6% | 71.8% | 30% | - | - | - |
| | 40% | 38.9% | 8.5% | 35.5% | 40% | - | - | - |
| Sequence | 10% | 66.2% | 80.8% | **81.8%** | 10% | **69.2%** | 52.8% | **75.7%** |
| | 20% | 67.1% | **84.2%** | 79.1% | 20% | 68.6% | 51.9% | 62.3% |
| | 30% | **70.2%** | 78.7% | 75.5% | 30% | 68.3% | 53.5% | 59.4% |
| | 40% | 66.1% | 76.8% | 75.5% | 40% | 63.5% | 53.5% | 55.3% |
| | 50% | 63.8% | 69.3% | 65.5% | 50% | 57.7% | 53.5% | 51.5% |

The performances are consistent with our intuition that the sequence representation is more discriminating than the set and the bigram representation. The discovered patterns of the sequence representation are more discriminating that consists of one item, two consecutive items (bigram), and three consecutive items. The bigram representation did not perform better than the set representation while the discovered patterns of the former are likely to be included in that of the latter.

Moreover, it is observed that the songs of the Enya and the Beatles (A-B) are less discriminating between each other. This is also consistent with our intuition.

An interesting observation is that the accuracy of chorus for most pairs of categories is not significantly worse than that of whole melody. This suggests that the features of chorus can be used to represent the music style for the efficiency issue.

## 5. CONCLUSIONS

In this paper we investigated the melody style mining and classification technique. We extracted the chords from melody and use the chords to represent the feature of melody. Experiment showed that the sequence representation with frequent substring mining perform well than the set representation. Above all, the chorus, which is the repeating pattern in a melody, may be utilized for the compact representation of the melody style. Future work includes the application of the discovered patterns for personalized music retrieval system and automatic computer music composer.


## ACKNOWLEDGEMENTS

We would like to thank Mr. Ming Jing Ho for the implementation of the MIDI parser and the repeating pattern finding algorithm.


## REFERENCES


[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules." *Proc. of International Conference on Very Large Data Bases VLDB'94*, 1994.

[2] R. Agrawal and R. Srikant, "Mining Sequential Patterns," *Proc. of International Conference on Data Engineering ICDE'95*, 1995.

[3] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models," *Proc. of the International Conference on Artificial Intelligence IC-AI'01,* 2001

[4] R. B. Dannenberg, B. Thom, and D. Watson, "A Machine Learning Approach to Musical Style Recognition," *Proc. of International Computer Music Conference ICMC'97*, 1997.

[5] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by Humming Musical Information Retrieval in an Audio Database," *Proc. of ACM International Multimedia Conference MM'95*, 1995.

[6] J. L. Hsu, C. C. Liu, and A. L. P. Chen, "Discovering Nontrivial Repeating Patterns in Music Data," *IEEE Transactions on Multimedia*, Vol. 3, No. 3, 2001.

[7] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima, "A Practical Query-By-Humming System for a Large Music Database," *Proc. of ACM International Multimedia Conference MM'00*, 2000.

[8] B. Liu, W. Hsu, and Y. Ma, "Integrating Classification and Association Rule Mining," *Proc. of ACM International Conference on Knowledge Discovery and Data Mining KDD'98*, 1998.

[9] P. Y. Rolland, "Adaptive User Modeling in a Content-Based Music Retrieval System," *Proc. of 2nd International Symposium on Music Information Retrieval ISMIR'01*, 2001.

[10] E. Selfridge-Field, "Conceptual and Representational Issues in Melodic Comparison," in *Melodic Similarity, Concepts, Procedures, and Applications*, MIT press, 1998.

[11] A. Uitdenbogerd and J. Zobel, "Manipulation of Music for Melody Matching," *Proc. of ACM International Multimedia Conference MM'98*, 1998.