Traditional and Folk Melody Classifier on Culture Style using Markov Models and Neural Networks

Chamila Karunatilake
University of Aizu
Aizu-Wakamatsu City
Fukushima, 965-8580
Japan
d8121106@u-aizu.ac.jp

Satoshi Nishimura
University of Aizu
Aizu-Wakamatsu City
Fukushima, 965-8580
Japan
nisim@u-aizu.ac.jp

Minetada Osano
University of Aizu
Aizu-Wakamatsu City
Fukushima, 965-8580
Japan
osano@u-aizu.ac.jp

ABSTRACT

Music plays a vital role in any culture despite whether it is primary or modern and it is a good indicator reflecting the nature of the culture where it has been produced. Music traditions were developed even in the pre-historic periods when people did not have a proper method of documentation and scores of music.

Melodies of different culture styles exhibit immense differences. Analyzing those differences is essential in various fields particularly ethnomusicology which studies non-western music based on cultural context. This paper presents an attempt of culture based melody classification using pitch.

With respect to that, a prototype has been developed in Java, which utilizes a Markov chains and Neural Networks. Experiments were conducted with several datasets which were chosen from the traditional music styles such as Indian and Japanese.

Categories and Subject Descriptors

H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing - *Modeling*, *Systems*

General Terms

Algorithms, Design, Experimentation

Keywords: Culture-based music classification, Kohonen Self Organizing Maps, Markov models

1. INTRODUCTION

It is believed that the music was originated in Africa, the cradle of humankind. The primary shape of music could be naturally occurred sounds and rhythms. Many observations have been reported on many species imitate sounds they hear. Therefore, the most primitive human-like species could have had their own kind of simple music which was used as a method of communicating and exhibiting their strong emotions [1]. Obliviously the first musical instruments could be the voice and clapping [2]. The earliest evidence of a tool which is believed to be a music instrument is dated to at least 35000 years ago. With the dispersal of early modern human from Africa, reaching all habitable continents around the world, they took the music with them. As a result, every tribe, even the most isolated ones, has their own style of music [3].

After formulating different cultures, different music styles evolved separately and sometimes without influencing one another. Because of this, the creation, performance, and even the definition of music vary according to culture and social context.

However, this uniqueness cannot be captured directly by human analysis since the complexity of the correlations. Computer based approaches are essential in this scenario and they can produce methods to examine problems beyond human analysis; particularly the artificial neural networks were designed to achieve this. In this research we propose a computer based approach to address this problem.

Various efforts have been made in the area of music classification and they have produced effective results [4] [5] [6] [7].

2. CULTURAL CONTEXT OF MELODY

As cultures and societies developed, the melodies they were produced were influenced by various environmental and social factors. For example, in many cultures, folk music was originated in parallel to the day-to-day work of ordinary people. The melodies inadvertently reflected the nature of the work that was being done. Religion also greatly influenced the structure of the melody, as pitch sequences were selected to evoke different emotions. Significant variations can be seen in melodies that evolved in different religious contexts, such as Gregorian chants, Japanese Buddhist chants and Hindu Mantras [8][9]. Because of these reasons, a particular style of music related to a certain culture produces melodies with similarities.

When it comes to the different cultures, melodies exhibit significant differences. This consequence can be seen when we listen to music. One argument oppose to this idea, is that the reason for this, is musical instruments used in the performance; however even we do it in different instruments such as piano or violin or human voice, the same outcome can be observed, therefore it should be an aspect related to the melody. One of the key characteristics of a melody is the association of the notes or pitch values which plays a vital role of making the originality.

Today, music is a medium which has globalized. Through advances in communication technologies, music is becoming more complex where the cultural context is concerned. Composers are influenced by many musical styles, so the cultural context of music is becoming unclear. As an example, new age

*HCCE'*12, March 8–13, 2012, Aizu-Wakamatsu, Fukushima, Japan.

musicians, such as Enigma, use components of Indian chants, Chinese music and Gregorian chants. Cultural background of this kind of music is a complex and sometimes useless.

Also when it comes to the western music, it has highly diverted from the cultural context it evolved. As an example, two western melodies which are created in Germany and Korea do not show much difference since both shares the same style of music. Hence, analyzing the cultural perspective of the melodies of western music is a very complex task. Here, the 'western' does not imply the geographical west where many folk and traditional music styles can be found. Those styles are also considered as non-western music excluded from the well defined Western music style.

Because of the above reasons, both western and contemporary music is not considered in this research.

3. APPROACH

A melody can be considered as a sequence of notes (Figure 1). The fundamental note association can be observed with adjacent notes and it can be modeled with Markov chains. Similarly, the note association of each melody is converted in to a comparable form by using Markov models. Then, those models are classified using Kohonen Self Organizing Maps (SOM), a neural network which has the ability of unsupervised learning.

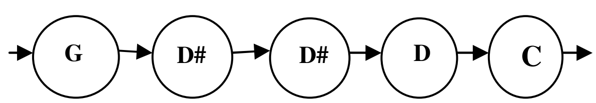

Figure 1. Melody as a Sequence of Notes

System takes a set of melodies which are chosen from two styles (cultures) of music as the input. These melodies are input to the system in a random manner and their culture identities are unknown to the system. Output of the system is the two classes which are created by the system by analyzing the pitch value combination patterns of the input melodies (Figure 2). Finally, the classification correctness of the output is calculated.

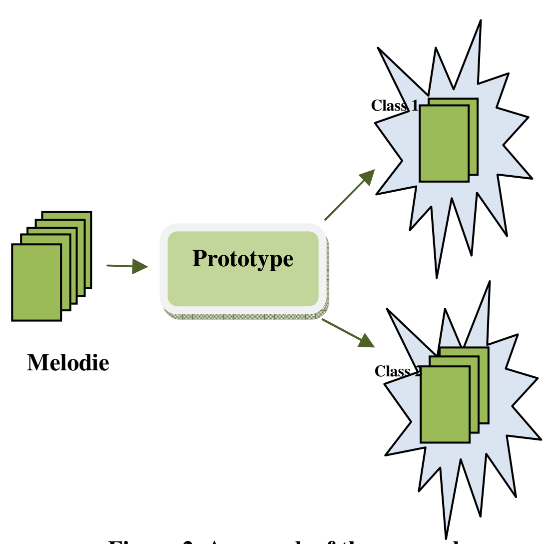

Figure 2. Approach of the research

4. MARKOV MODEL OF MELODY

Markov chains are stochastic processes based on the Markov property. A stochastic process consists of a set of states and a probability distribution which represents the transition of these states. The Markov property says that the future states of the stochastic process depend only on the present state, and are independent of past states.

Suppose we have a sequence of random variables $X_1, X_2, X_3, \ldots, X_n$

The above idea can be given with the following equation.

$$\Pr(X_{n+1}=x|X_1=x_1, X_2=x_{2\ldots.} X_n=x_n) = \Pr(X_{n+1}=x|X_n=x_n)$$

In our approach, the melody is considered as a sequence of notes (Figure 1). When melody is considered as a sequence of notes, the fundamental note association can be observed with adjacent notes and it can be modeled with Markov chains. The set of all possible notes which are taking part of this sequence is called the state space. For each note in the state space, the probabilities of being the next note are calculated by traversing the note sequence of the melody. This process produces a matrix which is known as the probability matrix for the particular melody. For all the melodies in the input dataset, the probability matrices are created. This outcome is a comparable model of the adjacent note associations of the melodies. This comparison cannot be done in a straightforward manner as the complexity of the matrices therefore it is handed over to neural networks which are proved to be thrived in these cases. Here, Kohonen self organizing maps were used.

5. KOHONEN SELF ORGANIZING MAPS

Kohonen Self Organizing Map (SOM) is an artificial neural network which utilizes unsupervised learning. They are capable of representing multidimensional data in much lower dimensional spaces which is called a map. These neural networks are used to discover some underlying structure of a given dataset. The specialty of these networks is that they use a neighborhood function to preserve the topological properties of the input space so they are also called topology-preserving map. The most common application of SOMs is for classifications or clustering complex patterns.

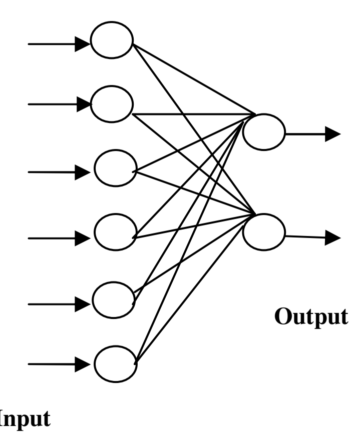

Figure 3. Structure of a SOM

A structure of a SOM is given in the Figure 3. This has two layers, input and output. The number of nodes or neurons in the output layer should be much lower than the number of input nodes. Here in this example, it is two, hence this network classify input patterns into two classes or clusters. The number of input nodes is determined by the number of data (properties) in the input patterns. In our approach the input is the probability matrix which is the outcome of the above Markov chain analysis process; hence the number of input nodes depends on the size of the probability matrix.

Each node of the input layer is connected to the each of the output node and every connection has a weight which is a decimal value. At first, these weights are assigned with initial values and then probability matrices of input melodies are submitted to the SOM one at a time. This is called the input vector.

The basic theory of SOM is assigning each of these input vectors to one of the output nodes which are considered as class labels of the clustering. These assignments are based on the degree of 'closeness' of the input vector to the weight vector of output node which is calculated through the Euclidian distance.

Euclidean distance = $\|x - W_i\|$

Here, x is the input vector and W_i is the weight vector of the output node *i*.

Each time, a new input vector is taken through the input layer; difference of the values of the input vector and the related weights of the output node weight vector is calculated. Then this distance checking is done for the next output node as well. The node which has the minimum distance is selected and named as the "winner" node and the input pattern is granted to it.

if $\|x - W_m\| = \min_{i=1,2,\ldots p} \| x - W_i\|$

The *m*th neuron is the winner

Then, the weight vector of the winner is updated as follows

$W_i = W_i + \alpha(x - W_i)$

α is the learning rate.

Usually, $\alpha \in [0.1,0.7]$

To the weights of the other nodes (non-winners), do nothing.

Hence, this learning rule is called "winner-take-all" learning. Generally, Kohonen SOMs use this rule.

The extraordinary characteristic of SOM is its unsupervised learning capability which allows network to work totally by itself to find the solution.

6. USING MIDI MUSIC FILES

One of the key requirements of this prototype is to use a common method of representing pitch values of the melodies. To achieve this, MIDI note values were used as it is an industry standard and facilities provided by programming languages are freely available.

MIDI (Musical Instrument Digital Interface) is a standard protocol adopted by the electronic music industry for enabling devices, such as keyboard controllers, computers and other electronic equipment, to communicate. MIDI does not actually transmit the sound signal itself but values for the note's pitch, length, and volume. It can also include additional characteristics, such as attack and delay time.

All MIDI compatible devices or software follow the same MIDI 1.0 specification and hence are capable of interpreting any given MIDI message the same way and communicate with each other. For example, if a note is played on a MIDI controller, it will sound at the right pitch on any MIDI instrument.

MIDI systems communicate using MIDI messages; furthermore they use MIDI channel messages to carry performance information. There are three steps are involved in the atomic event of this process. As an example, in a MIDI system, a one note is played using an instrument (here it is key board), then following sequence is taken place.

1. The key (middle C) is pressed with a specific velocity, a "Note-on" message is sent by the instrument.
2. The pressure applied on the key is changed while holding it down, which is known as after-touch, a special technique in musical performance. In this case, one or more after-touch messages are sent.
3. The key is released in a particular velocity and a "Note-off" message is sent.

All of these are channel messages and in Note-On and Note-Off, the MIDI specification defines a number (from 0–127) for every possible note pitch (C, C#, D etc.), and this number is included in the message. Other than this, many performance parameters can be transmitted with channel messages, however the concern of this research is only note pitch numbers which are also called MIDI note numbers. Table 1 shows some MIDI numbers for several notes.

Table 1. MIDI note numbers

Symbol	MIDI note number
C4	60
C#4	61
D4	62
D#4	63
E4	64
F4	65
F#4	66
G4	67

7. IMPLEMENTATION

The prototype which has been developed classification has three components.

1. Controller
2. MelodyMMBuilder
3. KohonenSOM

Controller contains the method (main ()) and it takes the leading role of the system by calling methods of other classes required to run the application.

MelodyMMBulder is the unit which transforms melodies in to a comparable form which is a Markov model for this prototype.

KohonenSOM compares melodies and does the classification using Artificial Neural Networks methodologies.

The execution sequence of the prototype starts by sending melodies to a file reader, where each melody is extracted as a sequence of notes. The sequences are then transferred to the MelodyMMBuilder to make the Markov transition probability matrices. They are then transformed to vectors, and an array of these vectors is submitted to the KohonenSOM for the classification (Figure 4).

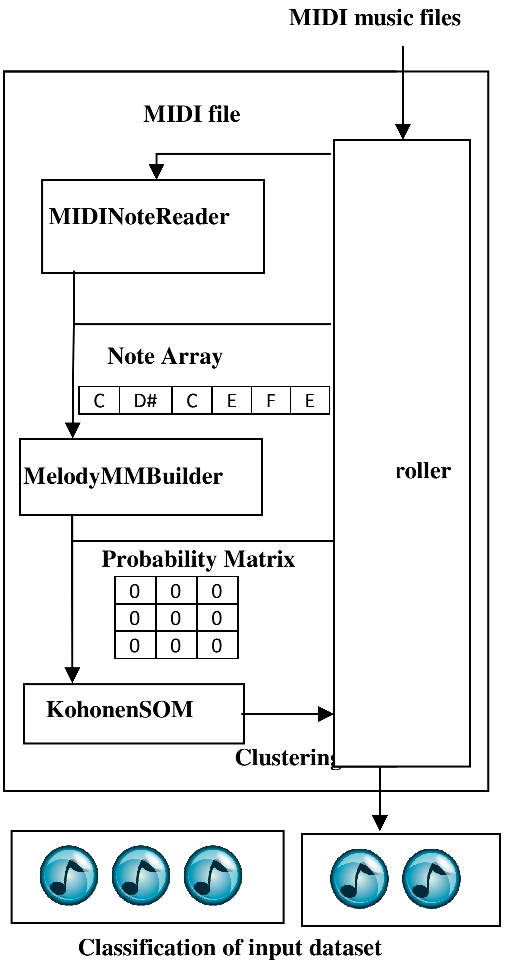

Figure 4. Components of the extended prototype

For this prototype the melodies were directly submitted to the MelodyMMBuilder in text files that contain the melodies as sequences of notes. The states are notes or pitch values, and the notes between G3 and E5 (22 notes) were used as the state space. In Markov Model point of view, a melody is a sequence of states that belongs to this state space (Figure 4).

In each melody, the preceding note is analyzed for each note, and a vector of preceding notes is created. Then, the probability of each proceeding note in this vector is calculated and a probability vector is constructed completing a transition probability matrix (Figure 5). This is the output of the MelodyMMBuilder and the input to the Neural Network.

The transition probability matrices are given to the Neural Network for the classification process.

		Current Note				
		…	**60**	**61**	**62**	**…**
	…	…	…	…	…	…
Next Note	**60**	…	0.1	0.6	0.3	…
	61	…	0.2	0	0.5	…
	62	…	0.7	0.3	0	…
	…	…	…	…	…	…

Figure 5. Transition probability matrix

The classification is done using Kohonen Self Organizing Maps (SOM). However, it expects vectors as input. Therefore, the transition probability matrices should be transformed into vector space before submitting.

In this prototype, a two-layer Neural Network was used. Here, the number of the output layer neurons of the network is decided by the number of clusters expected to be produced. This parameter should be set externally when the first time submitting the data to the system. Number of Input neurons depends on the size of the probability matrix.

The winner-take-all learning rule was used in this implementation. First step of this rule is to assign random values for the connection weight vector. Then, for the input vector, the Euclidian distance is calculated with respect to the weight vector. Neuron which has the minimum distance is selected as the winner.

Then the input is assigned to the winner neuron which can be considered as the class or the cluster for that input. Then weight vector between the winner neuron and the input neurons are assigned with new updated weights. Likewise, for each input, the cluster is determined.

The algorithm can be summarized as follows:

1. Assign random values to weight vectors
2. Get an input vector
3. Traverse each output node
    a. Use Euclidean distance formula to find similarity between the input vector and the node's weight vector
    b. Track the node that produces the smallest distance (this node is the best matching unit, BMU) and the node is considered as the winner
4. Update the nodes in the neighborhood of BMU by pulling them closer to the input vector
5. Go to the next input vector until all the input vectors are finished.

8. EXPERIMENT

Experiments were conducted using two styles, traditional Indian and Japanese music. Mono-phonic music was used including vocal and instrumental from both styles. Instruments such as Japanese Koto, Indian Harmonica and Sitar had been used. However, after converting into the MIDI format, those instrumental signatures had disappeared. So the classifications are totally depending on the melodies.

Melodies were selected from a pool of 100 melodies 50 Indian and 50 Japanese. The experiment was conducted 3 times independently selecting 60 random melodies at a time which include 30 melodies from each style. The overall correctness of the classification was 77.5%.

9. CONCLUSION

One of the most important aspects of our approach was using only pitch therefore this is a pure pitch based analysis. We have not considered at least time information of the melody. Some systems have used many other properties of music such as tempo and timbre. With our results shows that melody of different styles has considerable uniqueness.

Our prototype achieves the correctness of 77.5 % for the two different styles. Datasets are smaller since the difficulty of finding the traditional and folk melodies. Further enhancements of this prototype using music signals would resolve these problems and are expected to provide better results. Those enhancements have been carrying out at the movement.

10. REFERENCES

[1] Choi, Charles Q. 2009. *Monkey Drumming Suggests the Origin of Music*. DOI= http://www.livescience.com/animals/091016-monkey-drumming.html.

[2] Schellenberg, Murray. 2009. Singing in a Tone Language: Shona, *39th Annual Conference on African Linguistics*, ed. Akinloye Ojo and Lioba Moshi, 137-144.

[3] Herzog, George. 1934. Speech-Melody and Primitive Music. The Musical Quarterly, Vol. 20, No. 4.

[4] Liu, Yuxiang. , Xiang, Qiaoliang. , Wang, Ye. and Cai, Lianhong. 2009. Cultural Style based Music Classification of Audio signals. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (2009), 57-60.

[5] Pandey,G. , Mishra,C., and Ipe,P. 2003. TANSEN: A System for Automatic Raga Identification. *Indian International Conference on Artificial Intelligence*.

[6] Chordia P. and Rae A. 2007. Automatic Raag Classification Using Pitch-class and Pitch-class Dyad Distributions.*ISMIR*.

[7] Chai W., and Vercoe B. 2001. Folk Music Classification Using Hidden Markov Models. *IC-AI*.

[8] RT. Rev. MSGR. Spence, Charles E. 1952. *Chants of the church – Selected Gregorian Chants – edited and compiled by the Monks of Solesmes*. Gregorian Institute of America.

[9] Braun, Yehezkel. 1992. Aspects of melody: an examination of the structure of Jewish and Gregorian chants, *Companion to Contemporary Music Thought*, Vol. 2.