Int. Computer Symposium, Dec. 15-17, 2004, Taipei, Taiwan.

Using Back Propagation Model to Design a MIDI Music Classification System

Yo-Ping Huang and Guan-Long Guo
Dept. of Computer Science and Engineering
Tatung University
Taipei, Taiwan 10451
Email: yphuang@ttu.edu.tw

Chang-Tien Lu
Dept. of Computer Science
Virginia Tech
Falls Church, VA 22043
Email: ctlu@vt.edu

Abstract-*The main purpose of this paper is to investigate how to develop an effective classification system that can first categorize the characteristics in MIDI music files and then search similar music in the Internet. In this system, back propagation network is applied to train and categorize the characteristics in MIDI music. Many search engines now can provide efficient ways to search music. However, those search engines only search the files by the names of music, and cannot categorize and compare the music according to the characteristics of music. In this paper, we select representative songs of eight specific music categories to construct a module that can identify the types of music by means of back propagation network. We introduce the theoretical basis of music classification and present the experiment results to validate the effectiveness of the proposed model.*

Keywords: MIDI music, back propagation network, musical and percussion instruments.

1. Introduction

MIDI is an abbreviation of Musical Instrument Digital Interface [13-14]. MIDI is a kind of communication specification proposed in January 1983. Because of this specification, we can exchange music files that different electronic instruments produce with each other and prompt the development of electronic instruments rapidly and conveniently. The unit that an electronic instrument uses to make sounds is a channel. Before a channel can make a sound, it has to specify the kind of instrument, when to make a sound, the volume of the sound, the musical scale, and how long this sound will last.

The only function of MIDI is to send the commands to an electronic instrument and the instrument will create the sound according to these commands. The better the electronic instrument is, the more accurate it can regenerate the original music and show the style of music completely. Although MIDI defines the functions of the electronic instruments, different electronic instruments may have different codes for different instruments. For example, code 1 probably means a piano for electronic instrument *A*, but could be a trumpet for *B*. In this case, even the music created with *A* can be replayed with *B*, the style of music will be totally different and could not faithfully represent the original flavor of melody. To solve this kind of problem, we construct the general MIDI instrument map. All electronic instruments will follow this table to create sounds and replay the music just as they were. The only difference is the tone and characteristics of each electronic sound.

GM has defined 128 kinds of musical instruments, 47 kinds of percussion instruments, and classified them into 16 categories [13-14]. It seemed sufficient in the beginning, but recently a lot companies added tones of music into electronic instruments as much as they can with the hope that their products can be more competitive in the market and distinguishable from others. Because MIDI only records music data, the size of its file is much smaller than the one that records wave data. This kind of music is popular on the network for the requirement of limited bandwidth.

There are a lot of characteristics in melody. In order for users to compare and search similar music, we have to categorize the characteristics in a more systematic way. In general, these characteristics in melody include:

(1) Tempo: We can easily recognize whether the music belongs to slow or fast tempo. This is an obvious characteristic.

(2) The kind of musical instrument used: It is also an important characteristic. If a song uses bright piano or perceptual violin, this is also an important feature for a song's style.

(3) The number and proportion of each musical instrument used in a melody: Most people have different feeling about solo or large-scale symphony orchestra. Therefore, the number and proportion of each musical instrument used in a melody should also be considered in music classification.

We investigate the above characteristics to determine the style of a song and construct a model that can compare and search similar music files conveniently than from most of the search engines. The next question is how to construct an effective model to fulfill such a goal. Artificial neural network is a kind of data processing system that can imitate the neural network of human beings. This system can continually evolve its way of thinking, that is, think like human beings and learn from its experience.

Back-propagation model is one of the most well-known artificial neural networks [1-3,7-12]. Its basic principle is to use the concept of gradient steepest descent method to minimize the error function. By introducing the hidden layer concept into the network, the back propagation model has a good capability in mapping the relationship between inputs and desired outputs. We thus train the back propagation model to establish a classification system that can categorize the MIDI music files.

2. The Classification of MIDI Music Files

MIDI files have two kinds of formats, i.e., format 0 and format 1. Format 0 is an early format that has only one area for music data. Most of the music files on the network use this format. On the contrary, format 1 has many areas for music data and can be used for complex electronic instruments. We use format 0 in this paper to analyze the music. Because most mobile phones use MIDI as their rings format, lots of MIDI files are created specially for mobile phone users. Since there is only one track and contains incomplete music data in these MIDI files, in order to retrieve the real characteristics of music files, we did not use the kind of MIDI files that mobile phones used.

There are 128 kinds of MIDI musical instrument sounds, which can be divided into 16 categories according to their characteristics. Table 1 lists these 16 kinds of MIDI classes. There are 47 kinds of percussion instrument sounds of MIDI and can be divided into 5 categories, based on their special types. Table 2 lists the general MIDI percussion map.

Table 1. General MIDI instrument map (Channel 1-16, except 10).

Prg#	Instrument
1st kind：keyboards	
001	Acoustic Grand Piano
002	Bright Acoustic Piano
003	Electric Grand Piano
004	Honky-Tonk Piano
005	Electric Piano 1
006	Electric Piano 2
007	Harpsichord
008	Clavinet
2nd kind: Chromatic Percussion	
3rd kind: Organ	
4th kind: Guitar	
5th kind: Bass	
6th kind: Strings	
7th kind: Ensemble	
8th kind: Brass	
9th kind: Reed	
10th kind: Pipe	
11th kind: Synth Lead	
12th kind: Synth Pad	
13th kind: Synth Effects	
14th kind: Ethnic	
15th kind: Percussive	
16th kind: Sound Effects	

Table 2. General MIDI percussion map (Channel 10).

Key#	Percussive Instrument
1st kind: Bass Drum	
35	Acoustic Bass Drum
36	Bass Drum 1
64	Low Conga
66	Low Timbale
68	Low Agogo
2nd kind: Tom	
41	Low Floor Tom
43	High Floor Tom
45	Low Tom
47	Low-Mid Tom
48	Hi-Mid Tom
50	High Tom
54	Tambourine
62	Mute Hi Conga
63	Open Hi Conga
65	Hi Timbale
67	Hi Agogo
3rd kind: Snare	
37	Side Stick
38	Acoustic Snare
40	Electric Snare
60	Hi Bongo
61	Low Bongo
4th kind: Hat	
42	Closed Hi-Hat
44	Pedal Hi-Hat
46	Open Hi-Hat
49	Crash Cymbal 1
51	Ride Cymbal 1
52	Chinese Cymbal
55	Splash Cymbal
57	Crash Cymbal 2
59	Ride Cymbal 2
5th kind: Others	
39	Hand Clap

53	Ride Bell
56	Cowbell
58	Vibraslap
69	Cabasa
70	Maracas
71	Short Whistle
72	Long Whistle
73	Short Guiro
74	Long Guiro
75	Claves
76	Hi Wood Block
77	Low Wood Block
78	Mute Cuica
79	Open Cuica
80	Mute Triangle
81	Open Triangle

3. The Construction of System Module

MIDI files record a lot of music information and each file contains different kinds of formats. We have to understand those music information before we can analyze them. By simple statistical analysis, the system will calculate the values of these characteristics. Then we will determine the tempo of music, the number of tracks and tones in this music, and know what kind of music instrument used in the music. Through the analysis, we can obtain practical data that can be used to analyze the music.

To establish a classification model for the music files, we first need to find out what are the representative songs. These songs are regarded as the training samples for the model. The basic steps to classify and retrieve the music files are stated as follows:

(1) Analyze the MIDI structure: We can analyze the original MIDI file structure to extract the characteristic value.

(2) Select the representative training samples: The representative MIDI music songs are regarded as the training samples for the back propagation model.

(3) Find a better set of system parameters: We can repeat the training processes to find a better set of system parameters, such as the number of hidden nodes and learning rates, for the back propagation model.

(4) Retrieve other MIDI files: The well-trained model is then used as the basis to search and analyze other MIDI files in the Internet.

There are a lot of characteristics in MIDI music, for example, the beat, quantity of sound rail, classified timbre, etc. These characteristics cannot be directly fed to the training network. Instead, those characteristic values need to be normalized into [-1, 1] before being used for the back propagation model.

In our back propagation model, there are 23 input nodes, 15 hidden nodes, and 8 output nodes. Those 23 characteristic values of the music are briefly summarized in Table 3. Based on different trials from our simulation results, we use 15 nodes in the hidden layer. We use 8 nodes in the output layer to represent the 8 music categories as listed in Table 4. For example, the music 1_LULLAB.MID has the corresponding output "Blue" (output 1). The overall structure for the proposed back propagation model is shown in Fig. 1. Table 5 lists a training sample of input-output pattern for the network.

Table 3. The meaning for the 23 characteristic values for the input nodes in back propagation model.

<table>
<tr><th colspan="2">The serial number for input</th><th>Meaning</th></tr>
<tr><td colspan="2">1</td><td>Tempo</td></tr>
<tr><td colspan="2">2</td><td>Quantity of sound rail</td></tr>
<tr><td rowspan="16">Channel 1-16, except 10</td><td>3</td><td>Proportion of Keyboards classification</td></tr>
<tr><td>4</td><td>Proportion of Chromatic Percussion classification</td></tr>
<tr><td>5</td><td>Proportion of Organ classification</td></tr>
<tr><td>6</td><td>Proportion of Guitar classification</td></tr>
<tr><td>7</td><td>Proportion of Bass classification</td></tr>
<tr><td>8</td><td>Proportion of Strings classification</td></tr>
<tr><td>9</td><td>Proportion of Ensemble classification</td></tr>
<tr><td>10</td><td>Proportion of Brass classification</td></tr>
<tr><td>11</td><td>Proportion of Reed classification</td></tr>
<tr><td>12</td><td>Proportion of Pipe classification</td></tr>
<tr><td>13</td><td>Proportion of Synth Lead classification</td></tr>
<tr><td>14</td><td>Proportion of Synth Pad classification</td></tr>
<tr><td>15</td><td>Proportion of Synth Effects classification</td></tr>
<tr><td>16</td><td>Proportion of Ethnic classification</td></tr>
<tr><td>17</td><td>Proportion of Percussive classification</td></tr>
<tr><td>18</td><td>Proportion of Sound Effects classification</td></tr>
<tr><td rowspan="5">Channel 10</td><td>19</td><td>Proportion of Bass Drum classification</td></tr>
<tr><td>20</td><td>Proportion of Tom classification</td></tr>
<tr><td>21</td><td>Proportion of Snare classification</td></tr>
<tr><td>22</td><td>Proportion of Hat classification</td></tr>
<tr><td>23</td><td>Proportion of Others classification</td></tr>
</table>

Table 4. The 8 music categories.

The serial number for output	Meaning
1	Blue

2	Classical
3	Dance
4	Country
5	Funk
6	Jazz
7	Pop
8	Rock

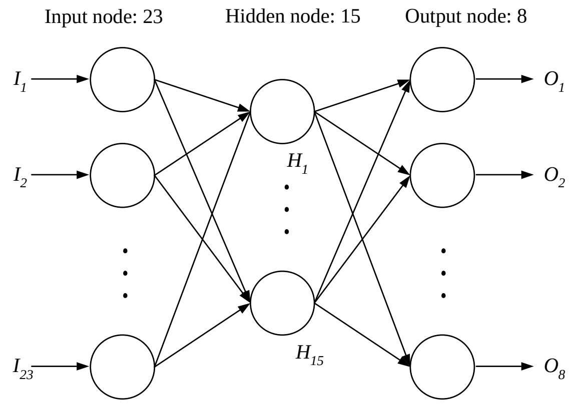

Fig. 1. The overall structure of the back propagation model.

Table 5. The sample of Blue, the name of the shelf: 1 _ LULLAB.MID.

Serial number	Value
Input 1	-. 883721
Input 2	-. 571429
Input 3	-. 133333
Input 4	-1.000000
Input 5	-1.000000
Input 6	-. 571429
Input 7	-. 666667
Input 8	-1.000000
Input 9	-1.000000
Input 10	-1.000000
Input 11	-. 333333
Input 12	-1.000000
Input 13	-1.000000
Input 14	-1.000000
Input 15	-1.000000
Input 16	-1.000000
Input 17	-1.000000
Input 18	-1.000000
Input 19	-1.000000
Input 20	1
Input 21	0
Input 22	-1.000000
Input 23	-1.000000
Output 1	1
Output 2	0
Output 3	0
Output 4	0
Output 5	0
Output 6	0
Output 7	0
Output 8	0

4. Experimental Results and Discussions

Since there is no definite method to decide an appropriate learning rate for the back propagation model, we performed several simulations to analyze how the learning rates affect the converging speed. Table 6 compares the final training errors from different learning rates under the same initial condition and 10,000 training iterations.

Based on the final training errors from the experiments, we found the larger the learning rate, the smaller the training errors. As a result, it seems that it is better to select a larger learning rate for the back propagation model. However, a smaller training error for the training samples may not correspond to a better test result. Based on our experience, an acceptable set of network parameters should be good for both the training and test patterns. In our model, we can have the best classification result when the learning rate is equal to 2.5. Fig. 2 plots the correct training result for this case. When the learning rate is set to 3.0, we may obtain an incorrect result as shown in Fig. 3.

The well-trained back propagation model is then used to search the similar music files. For example, in Fig. 4, when users click the 01.MID filename in the left hand side, the system can search the similar music files and list in descending order of similarities. By carefully analyzing the result, despite the tempo and middle musical instrument have a little difference, most characteristics of the 01.MID and 7_WALKLI.MID files are the same. Fig. 5 compares such a result. For comparison, if we use 01.MID to search for navigation and tempo, then the close one is BROWNJUG.MID as given in Fig. 6. In Fig. 7, we can see that both music files have good match except the slight differences for some musical instruments. Although BROWNJUG.MID has been classified as classical music, it also partially belongs to the rock category as shown in Fig. 8.

Table 6. The training errors from different learning rates in back propagation model.

Learning rate	Training errors
0.1	0.009247748342085
0.5	0.001570179186564
1	0.000717116696915
1.5	0.000497303467918

2	0.000370886425268
2.5	0.000307177201952
3	0.000103514806934
3.5	0.0000288618708 28

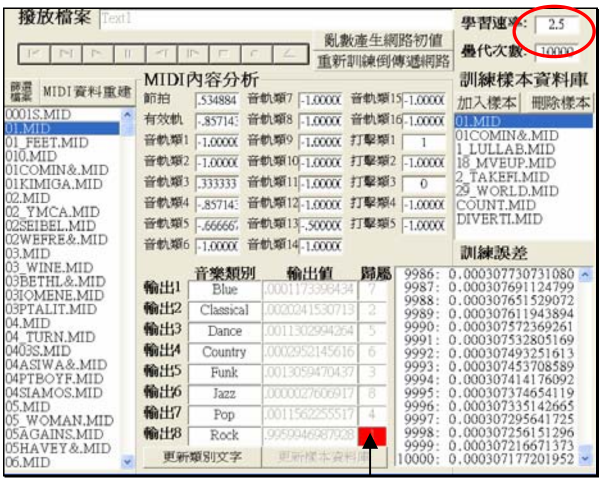

Fig. 2. A correct training result when the learning rate is 2.5.

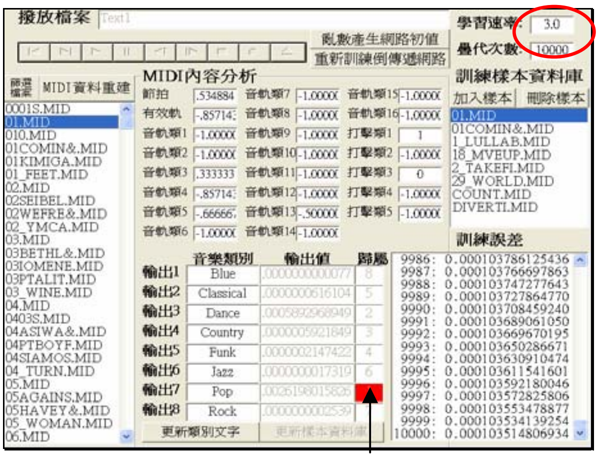

Fig. 3. An incorrect training result when the learning rate is 3.0.

Fig. 4. Using 01.MID file to find the similar 7_WALKLI.MID file.

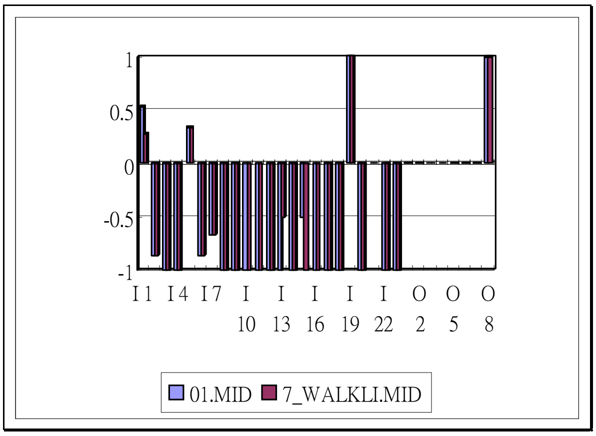

Fig. 5. Despite the tempo and middle musical instrument have a little difference most characteristics of 01.MID and 7_WALKLI.MID are the same.

Fig. 6. If we use 01.MID to search for navigation and tempo, then the close one is BROWNJUG.MID

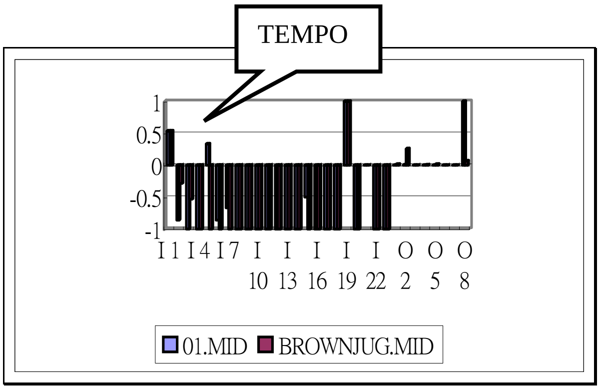

Fig. 7. The good match between 01.MID and 7_WALKLI.MID except the slight differences for some musical instruments.

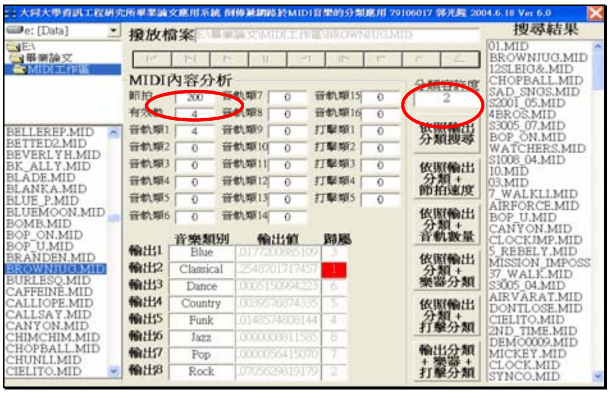

Fig. 8. Although BROWNJUG.MID has been classified as classical music, it also partially belongs to the rock category.

5. Conclusion

With the advent of advanced network technology, multimedia files can easily circulate in the Internet. Our goal is to design a system that can effectively classify the music files and apply this technology to searching multimedia files. It can advance the classification accuracy faced by most search engines that rely on the input keywords to query the files. After analyzing the characteristics of music, we identify the key factors for music classification which include 16 musical instruments and 5 percussion instruments, and propose a classification model to categorize the MIDI files based on the well-trained back propagation model. Users can select a favorite music to search similar music files without using the music filename. Experimental results verified that the proposed system can fulfill the goal of providing a satisfactory MIDI classification model.

The future work can focus on training the proposed model with large music classes so that the classification model can cope with the music classification in the changing world. In addition, fuzzy classification techniques can be used in the model to improve the performance of partial query problem, such as 0.4 degree belonging to POP and 0.6 degree to ROCK. We can also provide additional functionalities in the user interface for users to input their preferences of instrument types or tempos to simplify the query.

Acknowledgment

This work is supported by National Science Council, Taiwan, R.O.C. under Grants NSC92-2213-E-036-017, NSC92-2516-S-036-001, and by Tatung University under Grant B9208-I02-025.

References

[1] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computing*, Addison Wesley Publishing Company, Redwood City, CA, USA, 1991.

[2] An Introduction to Back-Propagation Neural Networks, Http://www.seattlerobotics.org/encoder/nov98/neural.html.

[3] H. White, "Economic prediction using neural networks: the case of IBM daily stock returns," *IEEE Int. Conf. on Neural Networks*, San Diego, CA, vol. 2, pp.451-458, July 1988.

[4] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, CA, 2001.

[5] B.Y. Ricardo and R.N. Berthier, *Modern Information Retrieval*, Addison-Wesley/ACM Press, New York, USA, 2002.

[6] Y.H. Ke, *An Efficient Inference Model for Personalized Data Mining System*, Department of Computer Science and Engineering, Tatung University, Taipei, Taiwan, 2002.

[7] W.J. Hsieh, *The Analysis and Application of Grey Model and Back-propagation Network to the Premium Rate Service*, Department of Computer Science and Engineering, Tatung University, Taipei, Taiwan, 2003.

[8] P. Werbos, *Beyond Regression: New Tool for Prediction and Analysis in the Behavioral Sciences*, Ph.D. Thesis, Harvard University, 1974.

[9] G.A. Carpenter and S. Grossberg, “ART2: self-organization of stable category recognition codes for analog input patterns,” *Applied Optics*, vol. 26, pp.4919-4930, 1987.

[10] T.P. Hong, C.S. Kuo, and S.C. Chi, “A fuzzy data mining algorithm for quantitative values,” *3rd Int. Conf. on Knowledge-Based Intelligent Information Engineering Systems*, pp.480-483, 1999.

[11] A. Kaufmann and M.M. Gupta, *Fuzzy Mathematical Models in Engineering and Management Science*, Amsterdam: North-Holland, 1988.

[12] R.R. Yager and D.P. Filev, *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons Inc, USA, 1994.

[13] Musical Instrument Digital Interface (MIDI), Http://www.indiana.edu/~emusic/MIDI.html#References.

[14] Standard MIDI Files 1.0, Http://ourworld.compuserve.com/homepages/mark_clay/MIDIfile.htm.