Automatic Music Generation and Machine Learning based Evaluation

Semin Kang, Soo-Yol Ok, and Young-Min Kang†

TONGMYONG University
535, Yongdang-Dong, Nam-Gu, Busan, Korea, 608-711
{semins,ymkang,sooyol}@tu.ac.kr

Abstract. This paper describes the automatic music composition system and automatic music evaluation system . The system composes short pieces of music by choosing some factors in music, such as pitch interval. timbre, tempo, rhythm. The most important features of the composition system include using the concept of mode, and density. mode control the pitch interval and densyty control the rhythm of music. We use Neural Network algorithm for automatic evaluation system of music. especially we used Back Propagation algorithm for objectification of user's taste.

Key words: Machine Learning, Automatic Music Composition, Automatic Music Evaluation, Backpropagation

1 Introduction

The development of computer technologies and smart devices brought a turning point in music creation. Many individual and independent developers are in needs of their own music for their own apps, and increasing people try to express their own personalities with their own styles of ring tones. Therefore, the individual users are now trying to create their own music. However, the composition of music requires deep knowledge about music which cannot be easily obtained. The ordinary people have unfortunately little time to learn music composition techniques. Therefore, easy and automated composition of music for individual purpose will become more and more important.

There have been various efforts to automatically compose music. The general approaches are categorized into two classes: 1) rule-based automatic composition with musical grammar and 2) procedural composition with mathematical models such as fractal, cellular automata or L-system. However, the previous approaches required users to have detailed knowledge about music. Otherwise, the previous methods could not generate music with various styles. Therefore, the high level control over the automated composition of various music has not been successfully achieved.

It is obvious that the music is not the composition of random sounds. The fact that we usually predict the next tune after the current tune demonstrates

†: Corresponding Author

the music obeys certain rules such as harmonics and counterpoint. However, the variety and subjectivity of music cannot be easily expressed with simple combination of known rules. Moreover, the evaluation of music depends on complex and subjective understanding so that the quantative and objective measurement cannot be easily modeled.

The automated composition of music can be achieved anyhow with music theory, mathematical models, or randomness. However, the more important issue in the composition of desired music lies rather in controllability of the music generation. The previous research efforts usually focused on the music generation itself and often neglected the evaluation of the procedurally generated music. Therefore, the high level control over the mood or style of music has not been successfully implemented.

In this paper, we propose an automated music composition system that enables users to create their own music for their own purpose by controlling the music with high level control parameters. The system is based on the supervised machine learning approach to evaluate music to take the subjective tastes of users into account.

2 Previous work

The earliest form of automatic music composition is to compose the fragments of melodies in accordance with randomly generated numbers. Since Mozart threw dice to select melody fragments for some minuets [2], various random number based approaches have been proposed and studies. The most common approaches to the automated music composition is to utilize mathematical models or some simple rules known in music literature.

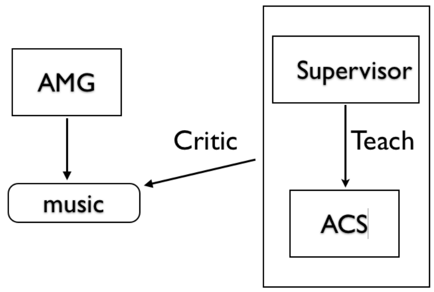
Puente [5] used musical grammar to procedurally generate music. Boenn[1] also used musical grammars such as harmonics to implement 'Anton', an automated composition system. The system actually produced plausible music. However, the variety of the obtained music was not satisfactory enough.

Many researchers proposed various methods based on fractal, L-system, or other mathematical models. Pressing[4] employed mathematical equations to generate music, and Worth [7] utilized L-systerm. However, the mathematical models could not successfully model the actual music composition rules that is customary and has been accumulated during the long time in human history. Therefore, the previous methods were sometimes useful only when people wants to create really new and fresh styles because the generated music tends to be unfamiliar and strange. In contrast, the methods usually failed to create ordinary music peaple are comfortable with.

There exist no common rules for music so that discords sometime sound fresh and nice. In the consequence, the automatic composition of music eventually fails when the musical grammar is overstressed. The well-defined grammars can be a predefined and limited set of musical pieces.

The limitation of the previous approaches is caused by the ignorance or neglectance of the importance of evaluation.

3 Automatic Music Generator



AMG: Automatic Music Generator
ACS: Automatic Classification System

Fig. 1. system structure :

A music piece is composed of various components such as pitch intervals, beats, and timbre. However, we usually regard different music pieces with the same melody as an identical piece even when the it is played with different instruments in different rhythms. Such experience demonstrates that the most important factor in distingushing music pieces is the melody.

The melody can be defined as time series of notes. Therefore, the music generation can be essentially reduced to the note determination at each time step. The most compelling example of such concept is 'music box'. The music box spins a cylinder with several tracks assigned to differently tuned teeth of a steel comb. The cylinder has pins placed on the surface, and the pins pluck the tuned teeth to produce sounds. The different tunes of the teeth are the states, and the time when the music box turns into each state is determined by the rotating speed. The music generated with our method is based on the model inspired by the music box.

Although the melody is the key factor in recognizing the identity of music pieces, the mood of an identical music can vary in accordance with other factors. It is often experience that an identical piece generates completely different moods when played with different instruments. In contrast, we can produce the variety of emotions with an identical instruments. Therefore, to control the mood and style of music, we took into account the 4 different factors: pitch interval, rhythm, tempo, and timbre. The variety of music was obtained by combining different instances of the four factors, and the evaluation was also performed on the domain defined by the factors. structure of the system is shown in Fig. 1.

3.1 The automated generation of music

The music composition in this paper is essentially reduced to determining the tunes of 16 eighth notes respectively. We have 16 time steps where the state must be determined. Each state can be one of 13 possible tunes or a mute sound. The note determined to be mute plays a role as a rest note. An example of decided states sequence is shown in Fig2. After the states of all the notes are determined, the timbre and tempo of the music is decided and the generated music is played with the tune color of the timbre in the speed assigned to the tempo.

C	⊗	G	F	C	⊗	C	⊗	D	F	E	⊗	G	G	C	⊗

Fig. 2. An example sequence of decided states

The random selection from available timbres or the random determination of tempo between the pre-defined minium and the maxium do not produce weird music. However, the ramdomly selected pitch intervals sequences hardly sounds like real music. In most cases, the sequence creates awkward and unstable melody. In order to avoid such problem, we employed 'mode' in the automated music composition.

The mode restricts the availability of the tunes. Therefore, our composition system selects one of the limited set of tunes in accordance with the current mode and previous tunes. By employing the mode, the music composition based on randomness can produce more reasonable set of music pieces, and the weird music pieces are avoided in the generation process.

The rhythm of music in the proposed system is determined by the sequence of mute notes and audible notes. Once the rhythm of a music piece is set, the probability of audible notes in the melody is also computed. The different rhythms of music pieces can be essentially regarded the density of the audible notes in the piece. Therefore, we employed 'density' parameter that determines the rhythm and actually controls the probability of the appearance of audible notes.

Mode The seven modes are known, and the mode is similar to the scale of music in that it restricts the available tunes. For instance, the mode 'C Ionian' is similar to the scale 'C Major'. In the C Ionian mode, the scale is generate as C-D-E-F-G-A-B-C where the pitch intervals between the 3rd and the 4th notes and between the 6th and the 7th notes are semitones. The different modes have different locations where the semitones intervals appear in the scale. For example, semitones appear between 2nd and 3rd and between 5th and 6th notes in Dorian mode. The interval sequence corresponding to the modes are listed in Tab 1.

If the music composition is restricted to use the scale defined in accordance with the specific mode, the discords drastically decrease and plausible and stable music pieces are usually generated. Moreover, each mode produces different mood of music. In the literature of music, the different moods by the modes are interpreted and proposed as shown in Tab. 2. The system proposed in this paper employed the mode not to produce strange and uncomfortable music pieces which are usually obtained when random selection of sounds.

modes	interval sequence
Ionian	T-T-e-T-T-T-e
Dorian	T-e-T-T-T-e-T
Phrygian	e-T-T-T-e-T-T
Lydian	T-T-T-e-T-T-e
Mixolydian	T-T-e-T-T-e-T
Aeolian	T-e-T-T-e-T-T
Locrian	e-T-T-e-T-T-T

Table 1. Seven modes and corresponding interval sequences: T represents whole tones and e represents semitones

modes	D'arezzo	Fulda
Ionian	serious	any feeling
Dorian	sad	sad
Phrygian	mystic	vehement
Lydian	harmonious	tender
Mixolydian	happy	happy
Aeolian	devout	pious
Locrian	angelical	of youth

Table 2. correlation of modes and feeling

Density The 'density' parameter of our method controls the probability of the generation of audible notes. Strictly speaking, the density is not exactly the same as the rhythm. However, the model of our method is based on music box where the rhythmic effect is generated by the alternation of audible notes and mute notes. Therefore we can suppose that the density is indirectly related to the rhythm. The difference between the high density music and the low density music is shown in Fig. 3.

C	⊗	G	F	C	⊗	C	A	D	F	E	D	G	F	C	⊗

(a) example of high density music piece

C	⊗	⊗	⊗	⊗	C	⊗	⊗	⊗	⊗	E	⊗	⊗	⊗	⊗	⊗

(b) example of low density music piece

Fig. 3. Difference between High Density & Low Density

3.2 Automated genration of drum beats

In general, famous music pieces are remixed by various musicians and the remixed versions have different feels. In order to create new mood of the music, drum beats are usually the most likely to be modified. For example, hip-hop musicians frequently use sampling, and the samples are often from classic or rock music pieces. However, the mood of the hip-hop music is usually quite different from that of the original music, and it seems the drum beats are the main reason of the difference.

Basic elements for the drum beats are hi-hat, snare and bass. The combination of those elements creates various rhythms such as bossanova, shufle, dub, and drum'n bass. The automatic generation of the drum beats in this paper is essentially reduced to determining the states of 16 sixteenth notes respectively. We have 16 time steps where the state must be determined. Each state can be 'Touched' or a mute sound.

In order to determine one drum beat, 16 sixteenth notes were generated for each element (i.e, hi-hat, snare and bass). The beat determination is performed for each adjacent two notes. Therefore, the 16 notes are grouped as 8 states variables. Each state can have one of four possible values such as 00, 01, 10 and 11. We grouped the adjacent beats to control the probability each possible values so that the generated beats are more reasonable and plausible. For example, the hi-hat beats should have more frequent beats than snare or bass. However, snare or bass should have lower probability for the states such as 11, 01 and 10. Fig. 3 shows how the shuffle rhythm can be generated with our approach.

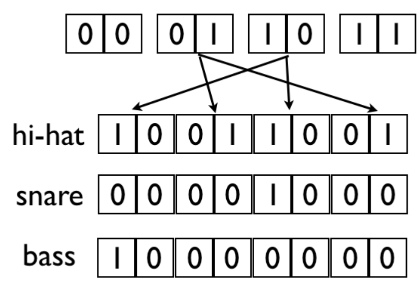


Fig. 4. composition of shufle rhythm

3.3 Automated classification of generated music

It is hard to quantify the evaluation of music because the understanding or emotional acceptance of music pieces is extremely subjective process. Therefore, it might be almost impossible to design a well-defined functions or measurement for the evaluation of music. Therefore, we employed the machine learning based evaluation system based on backpropagation neural network.

The backpropagation is an example of supervised learning system. According to the delta learning rule, the network adjusts the weights to obtain optimized connections to recognize the data set of which recoginition result is given. After the learning phase, the system can be utilized to recognize any data set based on the previous learning. Therefore, the method is proper to be quantify the subjective evaluation of music based on the user tastes, and can be utilized for the evaluation of various music pieces.

The machine learning of the backpropagation method is based on the input vector and weights of the links that connects the input to the output vector. Therefore, it is very important to determine what parameters involved in music generation to be selected as the part of input vector[6, 3].

We selected mode, timbre, tempo and pitch interval to be used as input vector. The parameters have 7, 16, 2, and 128 possible cases respectively, and expressed with 3, 4, 1, 8, and 4 bits in the gene sequence. The melody and drum beats are also coded in the gene sequence with 60 and 48 bits, so the total size of the gene sequence is 128 bits long.

The input layer with 128 notes are linked to hidden layer with 30 nodes, and they are also linked to the final output layer with 3 nodes. The learning final output was translated back to high level meaning of the input music.

The actual interface of the system is shown in Fig. 5. Users can easily generate various music and teach system to recognize the music in terms of high level description such as happiness. The supervised learning is utilzed further to recognize or classfy the music pieces which have never been evaluated before.

4 Conclusion and future work

In this paper, we proposed an automated generation and evaluation of music is proposed. The composition of music is based on the randomness. However, the automated composition with pure randomness seldom produces reasonable or plausible music pieces so that a method that restrict the possible output space to be limited within the plausible set of music pieces. In order to achieve such goal, we employed the concept of 'mode' in the generation process.

The automated composition of music requires two different conflicting capabilities: 1) the generation of restricted set of plausible music pieces and 2) the generation of variety of music pieces. While the mode is employed to restrict the generation, we also employed density, tempo, and timbre to increase the variety of the generated music pieces. The controlling parameters are converted into binary code to be given to a learning system based on backpropagation. The machine learning enabled users to automatically evaluate the music pieces generated by our system.

The evaluation system made it possible to classify the huge amount of automatically generated music pieces. The automated evaluation will be succefully utilized for the customized or purpose-based music generation.

The binary sequence expression of the music can be also successfully utilized to integrate our system to any evolutionary approach to automatically generate desired music or discover the unknown music fit to user defined objectives.

Unfortunately, the proposed system is not closely interoperable with the possible services that generate and control the user created music with high level

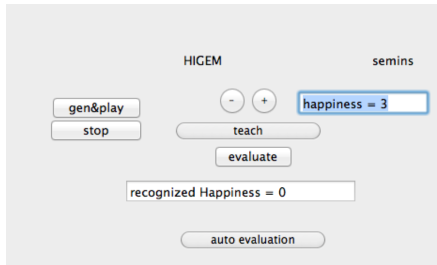


Fig. 5. system interface

parameters. We are intensively trying to integrate our system into the high level music composition system where the quntified evaluation and evolution-based music composition is implemented.

Acknowledgments. This work was supported in part by the Ministry of Knowledge Economy (MKE), Korea, under the Information Technology Research Center (ITRC) support program supervised by the National IT Industry Promotion Agency (NIPA) (NIPA-2011-(C-1090-1021-0006)), and this work (Grants No.C0033371) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2012.

References

1. Boenn G. : Anton: Answer Set Programming in the Service of Music. Proceedings of the Twelth International Workshop on Non-Monotonic Reasoning. Sydney: University of New South Wales, pp. 85-93.. (2008)
2. Alpern A. : Tecnique for Algorithmic Composition of Music. Hampshire College. (1995)
3. Gose E, .et al :Pattern recognition and image analysis. Prentice Hall Press PTR. (1996)
4. Pressing A. : Nonelinear maps as generator of muscal design . computer music Journal. (1988)
5. Puente A.et al :Automatic composition of music by means of grammatical evolution . Proceedings of the 2002 conference on APL, pp 148-155. (2002)
6. Tzanakou E. : Supervised and unsupervised pattern recognition. CRC Press. (2000)
7. Worth P, .et al :Growing Music: Musical Interpretations of L-Systems Lecture Notes in Computer Science. (2005)