Comparative Study of Musical Performance by Machine Learning

Yashoda N. Sakhare
ME-II Computer Engg
Vidya Pratishthan's College of Engineering
Baramati, India.
veena33@gmail.com

Mr. D. B. Hanchate
Department of Computer Engg.
Vidya Pratishthan's College of Engineering
Baramati, India.
dinesh_b_hanchate@yahoo.co.in

Abstract: This paper deals with the very special domains from computer science viz. Machine learning, genetic algorithms, rule based systems, music and various intelligent systems. Most of the musicians use machine learning approach to improve accuracy of the musical note. Intelligent systems use databases to store monophonic audio recordings performed by the musician of jazz standards. However, these approach use to obtain a model which explain and generate performances of expressive music. Rule based approach gives note level information containing time, dynamics and melody alteration. In this paper, we investigate how all these machine learning techniques work. We also compare their features and performance with evolutionary approach which will help user to get Rule based incremental model. Finally, output will be in a summarized format which gives reference solution. Comparative analysis shows that methods used by *Incremental Rule based Approach* provide full functionality and effectiveness as compared with previous machine learning techniques.

Keywords – *Computer Music, Machine Learning, Intelligent Systems, Computation Model.*

*****

1. INTRODUCTION

Expressive music is important for musicians to convey emotion and to connects with the listener. Traditionally, expressive performance involves manual operations in which a person prepares the hypothesis which captures the music and tested on real audio performances to find its accuracy [1]. Moreover, it is difficult for the person to manage all these tasks efficiently and accurately. Thus, many skilled musicians approaches evolutionary computation systems. Evolutionary computation is one type of machine learning technique to improve the accuracy of the performed music. ML methods successfully been implemented in a range of real-world audio and music applications [8].ML is divided into two phases: training phase and testing phase. Training phase learns data's internal structure from given samples however testing phase takes new samples and perform testing's [6]. A number of different learning strategies exist to train a program. There are three main types of learning: supervised; unsupervised; and semi-supervised [9].In a supervised learning approach, training data consists of pairs of input with corresponding desired output (the goal is known). In an unsupervised learning approach, the training data consists only inputs (the goal is unknown and must be learned from data). Finally, semi-supervised approaches consider examples of pairs of inputs and desired outputs as well as examples comprising only inputs (the goal is partially known and is refined by considering more unlabeled data)[8].In this paper we used supervised machine learning techniques which are configured to achieve specific tasks[8]:

Regression models:-It is a function of input data to create output.

Classification: It categorizes datasets.

Segmentation:-It partitions incoming data into separate regions.

Following challenges observed in other machine learning techniques [1]:

1. Note retrieval measures: Expressive readings of similar type of notes to inexpressive input notes recalled from the memory to generate a jazz solo performance. Most machine learning systems uses the information of case memory parameters, such as tempo, note duration, and so on to retrieve the melody [2]. This system does not allow understanding the way of predictions.

2. Solo music: Most machine learning techniques considered solo music performance. This classical solo music focusing on global tempo and energy transformations. But in jazz, note level timing and energy transformations are also consider.

3. Musical parameters: Tempo of any performed music pieces should remain constant and melody alterations should permit for better performance ,but in classical solo music the parameter does not remains constant and also not permitted to melody alterations using machine learning technique.

To address the above challenges, a new technique an evolutionary technique is proposed which automatically generate musical compositions. This technique uses Narmour's theory using performed music analysis which better understand the sense and knowledge of melodies.Given an audio signal from the user, system analyzes and extracted melodic features. This transformation is carried out by a set of low-level audio processing algorithms which are applied to the audio signal in order to generate the desired transformed melody.

For improving the accuracy of performed music, this technique will use a sequential covering genetic algorithm. This algorithm combines sequential covering [7] and genetic algorithms [1]. The sequential-covering component of the algorithm incrementally constructs a set of rules by learning new rules one at a time, removing the positive examples covered by the latest rule before attempting to learn the next rule [1]. The genetic component of the algorithm learns each of the new rules by applying a genetic algorithm [1]. An, evolutionary approach use linear regression method for each rule to generate linear equatation predicting a numerical value .Select one example covered by the rule having highest fitness value for new melodic context. It obtain a model capable of both explain and automatically generate expressive music performance. In summary, three key features of evolutionary approach can be listed as follows:

1) It will use note-level musical analysis to better understand the knowledge of different melodies.

2) It will utilize a Genetic sequential covering algorithm based on sequential covering algorithm and genetic algorithm to evolve a population of rules with normal mutation and crossover operation.

3)It will generate computational models for different views of expressive musical performance to improve the accuracy of the performed pieces.

Later this proposed evolutionary technique is compared with previously proposed machine learning techniques Comparative analysis shows that evolutionary technique performs better than machine learning techniques like Decision trees, Support vector machine(SVM),Artificial neural network, Logistic regression(LOGREG),KNN, Naïve bayes, Random forests, Bagged trees etc.

2. OVERVIEW OF MACHINE LEARNING TECHNIQUES:-

Machine Learning is a body of statistical methods that achieve tasks by learning from examples [8].These systems classify musical notes properly from a continuous stream of data . We apply these techniques of machine learning to the task of music composition with a large set of musical data from which standard algorithms can learn and predict musical notes. Such existing systems can be described as follows:

2.1 **Artificial neural network** [6]: In this system if the instances are not linear and separable ,learning will never reach a point where all instances are classified properly. Artificial Neural Networks have been created to try to solve this problem. Modler [11] used ANN for classifying hand postures based on data streamed from a glove. Classification output (i.e. the hand posture recognized) is used to control a sound synthesis. ANN allows learning of non-linear relationship between given inputs and their corresponding output.

Drawbacks:

This technique is not able to take into account temporal aspects of input where sample at a certain instant does not dependent on previous samples [5]. It has poor interpretability. Sometimes, due to low number of neurons the size of the hidden layer cannot determine properly. Thus, it leads to poor approximation and generalization capabilities[6].

2.2 Recurrent Neural Network (RNN)[10]:-This is the extension of ANN that takes into account temporal dependency by introducing short term memory .Early work by Lee et al. [10] made use of RNN for recognizing motion from a audio recording .The RNN performed regression and classification for outputs. The classification process used to control a MIDI instrument.

Drawbacks:

In these neural networks, complex relationship requires a large number of examples to train the network and results in a complex topology between inputs, hidden layers and outputs [10].

2.3 Support vector machines (SVM)[5]:-Support Vector Machines (SVM) are a set of related supervised learning methods used for classification and regression. SVM were originally formulated for two-class classification problems, and have quickly been accepted as a powerful tool for developing pattern classification and function approximation systems. SVM has been adapted to classify musical notes from a continuous stream of data. It is used to perform much better when dealing with multi-dimensions and continuous features. Its parameters indicate an algorithm is easy to use. This also predicts non linear models by using non linear attributes mapping. Thus, they provide more flexible predictions.

Drawbacks:

SVMs also have poor interpretability. For SVM a large sample size is required in order to achieve its maximum prediction accuracy.SVM training time is slow as compare to other algorithms. It requires higher computation cost to perform higher computation at higher space[13].

2.4 Bayesian networks [5]:-

The task of learning a Bayesian network is divided into two subparts: learning of
the DAG structure and determination of its parameters[5,6]. If the structure is unknown, approach is to introduce a scoring function that evaluates the "fitness" of networks with respect to the training data. Naive Bayes requires little storage space during both the training and classification stages.This model is sufficient to discriminate between classes. The major advantage of the naive Bayes classifier is its short computational time for training the musical notes.

Drawbacks:-

Naive Bayes networks are usually less accurate than other more sophisticated learning algorithms. They are not suitable for datasets with many features [12]. Construction of very large network is not feasible in terms of time and space . In most cases, the numerical features need to be discretized [5].

3. INCREMENTAL RULE BASED APPROACH:-This section deals with new proposed technique which provides better performance as compared to previous techniques. Steps of the proposed technique are shown below

Preprocessing of audio signal.
Note-level musical analysis to better understand the sense and knowledge of the melodies.
Generation of a computational model for different views of expressive musical performance.

TABLE1.STEPS OF PROPOSED INCREMENTAL RULE BASED APPROACH.

These steps are divided into modules which are further divided into sub modules

3.1 Preprocessing of audio signal:-In this section we will preprocess the audio signal by estimating energy and frequency, note boundaries, note descriptors and note analysis. For analyzing note we will use abstract structure as Narmour structure [2].

3.1.1 Energy and frequency estimation of user's audio signal:- In this section performed melody will extract for audio signals. For each note of the recording we will obtain set of descriptors .By using the values of amplitude spectrum at each analysis frame on spectral domain ,energy descriptor will computed[3].Fundamental frequency will estimate from two way mismatch procedure[14].

3.1.2 Detection of note boundaries:-It is a two step procedure. In first step, energy onsets will detect following a band-wise algorithm that uses psycho-acoustical knowledge [3].In second step, fundamental frequency will detect. By merging both results we will find note boundaries.

3.1.3 Note (event related) descriptor Calculation:-with the help of low level descriptors and note boundaries note descriptors will be calculated. Each note segment constitutes pitch note and fundamental frequency. Pitch histograms calculate note pitch and fundamental frequency. For averaging the frequency values convert into cents, by using the equation.

c = 1200 * {ratio of [logarithmic ratio of frequency to reference frequency] to *log*2}

Where *reference frequency*= 8.176.

After identifying the note pitch, MIDI pitch will be calculated by quantization of frequency mean over the frames in note limits. MIDI (Musical Instruments Digital Interface) is a standard for controlling and communicating with electronic musical instruments.

3.2 Note level musical analysis[3]:- After calculating note descriptors, for providing an abstract structure for recordings Narmour's theory is used to understand the knowledge of melodies.

The Analysis and Cognition of basic Melodic Structures (1989) established a comprehensive foundation for the detail theory. Narmour's theory is a note level phenomenon Narmour define the bases for differentiating between simple melodic[3] patterns in terms of their implications for continuation. He gives the principles based on registral and intervallic implication of messages.A small melodic interval generates registral and intervallic implications of similarity, while a large interval generates intervallic and registral implications of differentiation or reversal. Narmour gives five original models of melody [3] :

(a) Process or iteration (P and D)
(b) Reversal (R)
(c) Registral returns (RR)
(d) Dyad (two-element groupings, the unrealized implications of a and b) and
(e) Monad (one-element groupings, closed or unclosed with no generation of implication).

From the process and reversal, again five prototypical derivatives (models) established each constituting a partial denial of the realizations implied by the initial interval [3] these are:

(a) **Registral process** (VP - small interval to large interval, same registral direction)
(b) **Intervallic process** (IP- a small interval to a similar small interval in a different registral direction)
(c) **Registral reversal** (VR -large interval to even larger interval, different registral direction). The symbol "V" here stands for register.
(d) **Intervallic reversal** (IR – large interval to small interval, same registral direction) and
(e) **Intervallic duplication** (ID - Small interval to same small interval, different registral direction)

On the basis of these principles melodic patterns are identified to violate or satisfy the implication.

ISSN: 2321-8169
347 – 352

Following figure shows the prototypical Narmour structures[1]

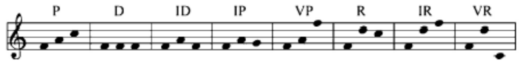

Figure 2 : Prototypical Narmour structures[1].

In musical pieces analysis of each melody in the training data is syntactically analyzed.
Following figure shows the analysis for melody[1].

Figure 3: Analysis for melody[1].

In western music, notes are represented by staff notations .some basic staff notations are listed below.

Staff G clef
Natural note Beams
Flat note Bar line(measure)
Sharp note

3.3 Generation of computational model:-

After preprocessing of audio signal, next step towards generating computation model have capability to explain and automatically produce expressive music performances. For that purpose required training data set, task determination and used algorithm is considered.

3.3.1 Training data set:- The training data set used in this technique is monophonic recordings of four Jazz standards (*Body and Soul*, *Once I loved*, *Like Someone in Love*, and *Up Jumped Spring*) performed by professional musician[1]. The required attributes for each musical piece are tempo, duration, metrical strength, Narmour structure etc. Each note is explained with these attributes and performed features .Number of attributes represents with note and its melodic context. The attributes which represents note are note duration, note interval and metrical strength of the note, while the attributes representing melodic context are tempo, neighboring notes information as well as Narmour group in which note appears[1].

3.3.2. Determination of the task :- In this section we approach to expressive performance alteration i.e. transformation. In note level expressive
Performance different transformations and their classes[2].

Transformation	Class
1. Note duration	1)Lengthen(duration 20% > nominal) 2)shorten 3)same duration (*same-dur*)
2. Onset deviation	1)advance (onset 5% >nominal) 2)Delay 3)same –onset
3. Energy	1)soft 2)Loud (louder than predecessor) 3)Medium
4. Note alteration	1)ornamentation 2)None

TABLE 2: NOTE LEVEL EXPRESSIVE TRANSFORMATIONS AND THEIR CLASSIFICATIONS.

3.3.3 Algorithm:-For getting generative performance model we decide to use sequential covering genetic algorithm . This algorithm is having some advantages over other machine learning techniques such as,
1) We get different models by executing this algorithm several times.
2) It gives the direction to the evolution of the model in a natural way.
3) The established model search and examine while it is “developing”.
4) Genetic Algorithm provides a power full technique for searching large problem spaces [4].
Genetic algorithm evolves a population of rules with usual selection, crossover and mutation operations. Initially it builds up the fitness values. Phrases (chromosomes)are selected with tournament selection process that considers phrase fitness and measure fitness .Then genetic operators are applied and half of each population is replaced by new offspring [4]. The algorithm incrementally constructs a set of rules by learning new rules one at a time, it removes the positive examples covered by the latest rule before attempting to learn the next rule [1].

Once we obtain the set of rules covering all the training examples, we apply linear regression to each rule in order to obtain a linear equation providing a numerical value. For note alteration it is not necessary to compute a numerical value. Instead we simply store the set of covered examples by the rule [1].For reconstruction of melody select one of the covered examples by the rule and adapt the selected example to the new melodic context [1].Hypothesis representation considers the rule set which consist fixed set of attributes. The parameters Ps,cr and mr gives the values of Population size, crossover rate, mutation rate respectively. Different steps of algorithm to generate rules are given as follows[1]:

Begin

SeqCovgenAlg (Class, Fitness, ps, cr, mr,Examples)
1) EBC = examples which belong to Class
2) NBC = examples which do not belong to Class
3) Lr = {}
4) While EBC do
    A) P = generate p hypotheses at random
    B) For each hypothesis h in P,
        Compute fitness (h)
    C) While max (fitness (h)) <400generations do
        Create a new generation Pnew
        Ps = Pnw
    D) For each h in P,
        Compute fitness (h)
5) Nr = hypothesis in Ps having highest fitness value
6) Rs = members of EBC covered by Nr
7) Compute PredictedValue (Rs)
8) NumNR = NewRule with Class replaced by
    Regression (Rpos)
9) Lr = Lr + NumNR
10) EBC = EBC - Rs
    Return Lr

End

Sequential covering genetic algorithm [1].

4. CONCLUSION:

In this paper, we have made a review of different Machine learning techniques. We have explained evolutionary computation technique in which system obtain a model capable of both explain and automatically generate expressive music Performances . Also, we have compared it with existing techniques like Decision trees, Artificial neural network, Logistic regression (LOGREG) ,KNN, Naïve bayes, Random forests,etc. Evolutionary computation technique is more useful to assist the audio signals by providing result in the tabular format by considering accuracy and no of rules in each classes of various transformations . Comparative study shows that evolutionary computation technique performs better as compared to existing machine learning techniques.

Table III shows comparison among different machine learning techniques. Evolutionary computation technique performs better as compared to existing machine learning technique

Machine learning system	**Concept**	**Advantages**	**Limitations**	**Performance**
Artificial neural network	This classifies hand postures based on data streamed from a glove. Classification output is used to control a sound synthesis	ANN allows learning of the non-linear relationship between given inputs and their corresponding output	Not able to take into account temporal aspects of input, poor interpretability	Intermediate
Recurrent Neural Network	This specifies temporal dependency by introducing short term memory .Also it recognize motion from a audio recording.	The RNN performed regression and classification with outputs from the classification process used to control a MIDI instrument.	Complex relationship requires a large number of examples to train the network	Intermediate
Support vector machine (SVM)	This technique classifies musical notes from a continuous stream of data. SVM is formulated for two-class classification problems. A powerful tool for developing pattern classification and function approximation system.	SVM has been adapted to classify musical notes from a continuous stream of data.	Poor interpretability, a large sample size is required in order to achieve its maximum prediction accuracy, training time is slow.	Intermediate
Bayesian networks	This learns the structure of the melody and then determines its parameters.	Naive Bayes requires little storage space during both the training	Less accurate, not suitable for datasets with many features of melody,	Intermediate

	If the structure is unknown, it score the function that evaluates the "fitness" of networks with respect to the training data.	and classification.	not suitable for very large network, Before the induction, the numerical features need to be discretized in most cases	
Incremental Rule based Approach	This technique obtain a models capable of both explain and automatically generate expressive music performance.	Obtaining models captures different possible interpretations of a musical piece.	-	High

TABLE III: COMPARATIVE STUDY OF MACHINE LEARNING TECHNIQUES.

5. REFERENCES

[1] Rafael Ramirez, Esteban Maestre, and Xavier Serra, "A Rule-Based Evolutionary Approach to Music Performance Modeling", IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 16, NO. 1, pp.96-107, FEBRUARY 2012

[2] Maarten Grachten, Josep Llu´ıs Arcos and Ramon L´opez de M´antaras" Melody Retrieval using the Implication/Realization Model", IIIA, Artificial Intelligence Research Institute CSIC, Spanish Council for Scientific Research Campus UAB, 08193 Bellaterra, Catalonia, Spain {maarten,arcos,mantaras}@iiia.csic.es

[3] E. Narmour "Review of The analysis and cognition of melodic complexity: the implication/realization Model"; Univ. of Chicago Press, Chicago, 1992. *Music Perception*, 1995, 12 , pp 486-509.

[4] John A Biles" GenJam:A Genetic Algorithm for Generating JazzSolos" Associate professor, Information Technology Department, Rochester Institite of Technology, jab@cs.rit.edu.

[5] Machine learning: a review of classification and combining techniques S. B. Kotsiantis · I. D. Zaharakis · P. E. Pintelas. Artif Intell Rev DOI 10.1007/s10462-007-9052-3

[6] Zhang G (2000) "Neural networks for classification: a survey", IEEE Trans Syst Man Cy C 30(4):451–462

[7] R. S. Michalski, "On the quasi-minimal solution of the general covering problem," in *Proc. 1st Int. Symp. Inform. Process.*, 1969, pp. 125–128.

[8] Baptiste Caramiaux and Atau Tanaka "Machine Learning of Musical Gestures", Department of Computing Goldsmiths, University of London , Department of Computing Goldsmiths, University of London bc@goldsmithsdigital.com

[9] T. Hastie, R. Tibshirani, and J. Friedman "The elements of statistical learning: data mining, inference, and prediction",Springer Series in Statistics, second edition, 2001.

[10] M. A. Lee, A. Freed, and D. Wessel. "Neural networks for simultaneous classi_cation and parameter estimation in musical instrument control",in Adaptive and Learning Systems, pages 244-255, 1992.

[11] P. Modler. "Neural networks for mapping gestures to sound synthesis", Trends in Gestural Control of Music, pages 301-314, 2000.

[12] Cheng J, Greiner R, Kelly J, Bell D, Liu W (2002) "Learning Bayesian networks from data: an information theory based approach", Artif Intell 137:43–90.

[13] Rafael Ramirez Amaury Hazan Emilia Gomez Esteban Maestre, "A Machine Learning Approach to Expressive Performance in Jazz Standards" Music Technology Group, Pompeu Fabra University Ocata, 1. 08003, Barcelona, Spain.

[14] R. C. Maher and J. W. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," *J. Acoust. Soc.Am.*, vol. 95, no. 4, pp. 2254–2263, 1994.