Melody Generation Using RecurrentNeural Network

Sarthak Somvanshi[1], Vinit Vaidya[2], Chinmay Vartak[3], Dnyaneshwar Kapse[4]

[1,2,3 –] Student, [4 -] Guide

[1,2,3,4]Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai,Maharashtra, India.

Abstract

In recent years, neural networks have been used to generate symbolic melodies. However, the long-term structure in the melody has posed great difficulty for designing a good model. In this paper, we present ahierarchical recurrent neural network for melody generation, which consists of three Long-Short-Term- Memory (LSTM) subnetworks working in a coarse-to- fine manner along time. Specifically, the three subnetworks generate bar profiles, beat profiles and notes in turn, and the output of the high-level subnetworks are fed into the low-level subnetworks are fed into the low-level subnetworks, serving as guidance for generating the finer-time-scale melody components in low-level subnetworks. Two human behaviour experiments demonstrate the advantage of this structure over the single-layer LSTM which attempts to learn all hidden structures in melodies.

Keywords*-Machine Learning, Melody, MelodyGeneration Using Recurrent Neural Network, Literature Survey.*

Date of Submission: 28-04-2021 Date of acceptance: 12-05-2021

I. INTRODUCTION

Automatic music generation using neural networks has attracted much attention. There are two classes ofmusic generation approaches, symbolic music generation and audio music generation. In this study, we focus on symbolic melody generation, which requires learning from sheet music. Many music genres such as pop music consist of melody and harmony. Since usually beautiful harmonies can be ensured by using legitimate chord progressions whichhave been summarized by musicians, we only focus on melody generation, similar to some recent studies. This greatly simplifies the melody generation problem. Melody is a linear succession of musical notes along time. It has both short time scale such as notes and long time scale such as phrases and movements, which makes the melody generation a challenging task. Existing methods generate pitches and rhythm simultaneously or sequentially using Recurrent Neural Networks (RNNs), but they usuallywork on the note scale without explicitly modeling the larger time-scale components such as rhythmic patterns. It is difficult for them to learn long-term dependency or structure in melody. Theoretically, an RNN can learn the temporal structure of any length in the input sequence, but in reality, as the sequence gets longer it is very hard to learn long-term structure. Different RNNs have different learning capability, e.g., LSTM (Hochreiter and Schmidhuber 1997) performs much better than the simple Elman network. But any model has a limit for the length of learnable structure, and this limit depends on the complexity ofthe sequence to be learned. To enhance the learning capability of an RNN, one approach is to invent a newstructure.

In this work we take another approach: increase thegranularity of the input. Since each symbol in the sequence corresponds to a longer segment than the original representation, the same model would learn longer temporal structure. To implement this idea, we propose a Hierarchical Recurrent Neural Network (HRNN) for learning melody. It consists of three LSTM-based sequence generators — Bar Layer, Beat Layer and Note Layer. The Bar Layer and Beat Layer are trained to generate bar profiles and beat profiles, which are designed to represent the high-level temporal features of melody. The Note Layer is trained to generate melody conditioned on the bar profile sequence and beat profile sequence output by the Bar Layer and Beat Layer. By learning on different time scales, the HRNN can grasp the general regular patterns of human composed melodies in different granularities, and generate melodies with realistic long-term structures.

This method follows the general idea of granular computing in which different resolutions of knowledge or information are extracted and represented for problem solving. With the shorter profile sequences to guide the generation of note sequence, the difficulty of generating note sequencewith well organized structure is alleviated.

II. LITERATURE SURVEY

Florian Colombo, Alexander Seeholzer, and Wulfram Gerstner [1] “DeepArtificialComposer: A Creative Neural Network Model for Automated Melody Generation” , in this paper we assume that music is a form of language that can be expressed with musical grammar, and rules in such grammar are numerous and universal. Since the grammar is large, we hypothesize that different composers utilize different subsets of these rules. Bela Bartok,a well-known 20th century composer, can be used to illustrate this hypothesis. Bartok’s work has beenanalyzed in detail; for example, Antokoletz proposed a system that could serve as a means to formulate and organize Bartok’s music. One characteristic of Bartok’s work is the use of certaincell structures (patterns of tonal music), which are found more frequently in his work than those of other composers . Based on this analysis, we may design a system to simulate Bartok’s music by usinga common set of musical grammar rules and aspecific set of rules about the cell structures characteristic to Bartok. We believe that the resultsfrom this system will sound, to an extent, like Bartok’s music.

Florian Colombo, Alexander Seeholzer, and WulframGerstner, [2] “Learning to Generate Music with BachProp” Here, we present BachProp, an algorithmic composer that can generate music scoresin many styles given sufficient training data. To adapt BachProp to a broad range of musical styles, we propose a novel representation of music and train a deep network to predict the note transition probabilities of a given music corpus. In this paper, new music scores generated by BachProp arecompared with the original corpora as well as with different network architectures and other related models. We show that BachProp captures important features of the original datasets better than other models and invites the reader to a qualitative comparison on a large collection of generated songs.

Jenif D’Souza W S et.al, [3] “Melody Generator: A Device for Algorithmic Music Construction” The method followed consists of four stages: 1) selection of music-theoretical insights, 2) translation of these insights into a set of principles, 3) conversion of the principles into a computational model having theform of an algorithm for music generation, 4) testing the “music” generated by the algorithm to evaluate the adequacy of the model. As an example, the method is implemented in Melody Generator, an algorithm for generating tonal melodies. The programhas a structure suited for generating, displaying, playing and storing melodies, functions which are allaccessible via a dedicated interface.

PING-HUAN KUO, TZUU-HSENG S. LI [4]

“Development of an Automatic Emotional Music Accompaniment System by Fuzzy Logic and Adaptive Partition Evolutionary Genetic Algorithm” For different emotions, a fuzzy logic controller is designed to adjust the tempo of the music, and an adaptive partition evolutionary genetic algorithm is developed to create corresponding melodies. The chord progressions are generated via music theory, and the instrumentation is disposed by the conception of the probability. What is noteworthy is that all the processes can be output by Virtual Studio Technology in real time so that users can listen directly to the composing results from any emotions. From the experimental results, the proposed adaptive partition evolutionary genetic algorithm performs better than other optimal algorithms in such topics.

Bana Handaga et.al, [5] “Performance of Three Slim Variants of The Long Short-Term Memory (LSTM) Layer”. The Long Short-Term Memory (LSTM) layer is an important advancement in the field of neural networks and machine learning, allowing for effective training and impressive inference performance. LSTM-based neural networks have been successfully employed in various applications such as speech processing and language translation. The LSTM layer can be simplified by removing certain components, potentially speeding up training and runtime with limited change in performance. In particular, the recently introduced variants, called SLIM LSTMs, have shown success in initial experiments to support this view. Here, we perform computationalanalysis of the validation accuracy of a convolutional plus recurrent neural network architecture using comparatively the standard LSTM and three SLIM LSTM layers. We have found that some realizations of the SLIM LSTM layers can potentially perform as well as the standard LSTM layer for our considered architecture..

Emilios Cambouropoulos, [6] “From MIDI to Traditional Musical Notation” A system that attempts to extract the musical surface (i.e. a symbolic representation of notes in terms of quantised onsets and durations, and correctly spelled pitches) from a polyphonic MIDI performance is herein described. This system was developed as a means for obtaining the scores (in asymbolic machine-readable format) of a largenumber of performed piano works in the context ofthe project: ‘Artificial Intelligence Models of Musical Performance’. Working on a score-extraction project would not only provide a useful tool for obtaining symbolic scores (optical recognition techniques are probably a more obviouscandidate for this task) but would additionally give rise to invaluable insights into the relation of a musical performance to its corresponding musical score. In general, however, score-extraction techniques are indispensable for a plethora of applications that process performed MIDI input (e.g. music notation packages, interactive musical performance systems etc.)..

Benjamin Genchel [7] “Explicitly Conditioned Melody Generation: A Case Study with Interdependent RNNs” Deep generative models forsymbolic music are typically designed to modeltemporal dependencies in music so as to predict thenext musical event given previous events. In manycases, such models are expected to learn abstractconcepts such as harmony, meter, and rhythm fromraw musical data without any additional information. In this study, we investigate the effects of explicitlyconditioning deep generative models with musicallyrelevant information. Specifically, we study theeffects of four different conditioning inputs on the performance of a recurrent monophonic melody generation model. Several combinations of theseconditioning inputs are used to train different modelvariants which are then evaluated using threeobjective evaluation paradigms across two genres ofmusic. The results indicate musically relevant conditioning significantly improves learning andperformance, and reveal how this information affectslearning of musical features related to pitch andrhythm. An informal subjective evaluation suggests a corresponding improvement in the aesthetic quality ofgenerations.

Yuan L et al, [8] “Melody Generation System based on a Theory of Melody Sequence” on the Implication- Realization Model (IRM) of music theory. The IRM is a music theory, which was proposed by Eugene Narmour. The IRM abstracts music. It then expressesmusic according to symbol sequences based on information constituting the music pitch, rhythm, andrests. Previous melody generation systems are mostlybased on tone transition models, which do not have a function of abstracting melodies observed in trainingdata. In those previous systems, generated melodies do not reflect tone sequences that do not exist in training data. However, it is obviously required that amelody generation system is able to abstract melodies in training data and to output certain melodies, which is rarely observed in the training data. Our melody generation approach properly abstracts melodies in training data based on the IRM. The IRM expresses contexts of melodies using symbol sequences. Our melody generation system consists of two models; that of symbol sequence transition and that of generating tones from symbols. With the former model, the symbol transition probability model is trained with the results of the IRM analysis. s. The system then generates an optimal symbol sequence according to the probability model. Then, from a set of tones, each symbol sequence generates a melody.

Hsiao-Tzu Hung †, Chung-Yang Wang† Yi-Hsuan Yang†‡, Hsin-Min∗ Wang∗, [9] “Improving

Automatic Jazz Melody Generation by Transfer Learning Techniques”. Jazz is one of representative types of music, but the lack of Jazz data in the MIDIformat hinders the construction of a generative model for Jazz. Transfer learning is an approach aiming to solve the problem of data insufficiency, so as to transfer the common feature from onedomain to another. In view of its success in other machine learning problems, we investigate whether,and how much, it can help improve automatic musicgeneration for under-resourced musical genres. Specifically, we use a recurrent variational autoencoder as the generative model, and use a genre-unspecified dataset as the source dataset and a Jazz-only dataset as the target dataset. Two transfer learning methods are evaluated using six levels of source-to-target data ratios. The first method is to train the model on the source dataset, and then fine-tune the resulting model parameters on the target dataset. The second method is to train the model on both the source and target datasets at the same time, but add genre labels to the latent vectors and use a genre classifier to improve Jazz generation. The evaluation results show that the second method seems to perform better overall, butit cannot take full advantage of the genre- unspecified dataset.

III. RESULTS

The results of our literature survey have been divided in five points that ensure designing an efficient automated attendance system with highestaccuracy. These factors have been chosen after thorough analysis of the problems that occur in processing of images in face recognition.

A. Machine Learning Platform

We trained and evaluated the Deep Artificial Composer (DAC, see Methods) network with a corpus comprising 2180 Irish folk melodies and 600Klezmer tunes. The songs in the corpus containedon average 250 notes per melody for a total of695’000 note transitions. The predictive performance, or accuracy, of the DAC corresponds to the percentage of transitions where the model gave the highest probability to the true upcomingevent. On average, the accuracy of the trained DACnetworks on melodies from the validation/trainingset is 50%/80%. The average predictive performance of the duration model is 80%/85%

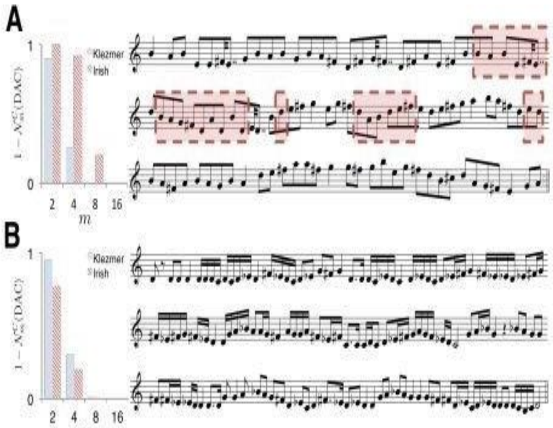

Fig. 1. Examples of DAC-generated melodies

[1] Above Fig. 1 example melodies imagined by the DAC present a very well defined rhythmic and melodic structure that can be related to either Irish or Klezmer style. Notably, the modes of the two melodies fit either style. For both melodies, the artificial composer learned to end on the fundamental and display a well defined structure, which can only be achieved if DAC networks were able to internally represent the mode through their hidden states

[2] In Fig 2 We compare the novelty profiles for all models with respect to the original Chorales corpus with which each model was trained. We compare the different profiles with the auto-novelty of the reference corpus. The auto-novelty is the novelty profile for each song in the reference corpus with respect to the same corpus without the song for which the novelty score is computed. It reflects, how similar is the music within the original corpus and is consequently the distribution to match for an ideal generative model of music. Here, the only model that is clearly outside the target distribution is the MLP model. While the IndepBP and MidiBP models match the target distributions, their novelty distributions for bigger pattern sizes islower than the original corpus auto-novelty. This isan indicator that these models are generating musicexamples that are too similar to the original data.

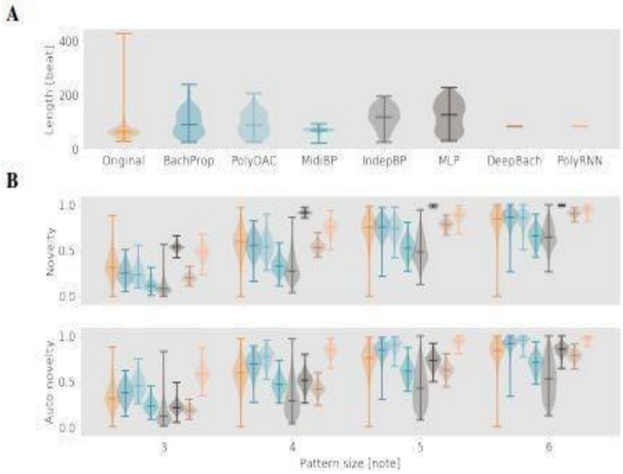

Fig. 2 Song lengths and novelty profiles.

A LSTM Variant Performance Based on the comparison of average validation accuracy across learning rates and activation functions, LSTM3 appears to have the best average accuracy out of allof the reduced LSTM variants, and additionally does not appear to vary significantly from the base LSTM layer’s performance. While some tests indicate that LSTM3 was the best variant overall, training variance was high enough that the results merely suggest that LSTM3 isn’t strictly better thanthe base LSTM

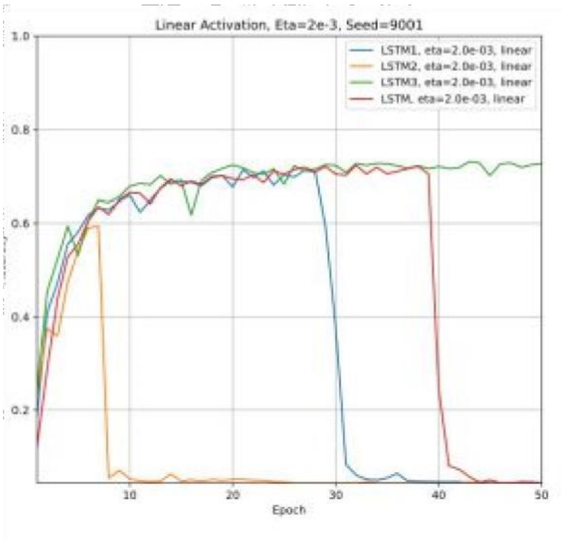

Fig. 3 LSTM Training, Linear Activation, eta=2e-3

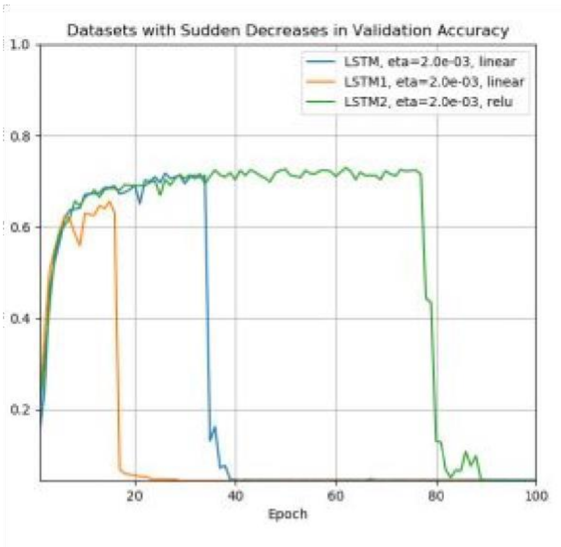

Fig. 4: LSTM Training, LinearActivation, eta=2e-3

terms of validation performance and loss. Still, if this quality holds up in other architectures, it could provide a basis for using LSTM3 by default in performance-critical roles. The LSTM3 layer does not apriori impose structural form on the gating signal. The adaptive process for the parameters willinvolve the input signal profile, hidden units. In contrast, the standard LSTM makes the imposition of a definite structure that may not be convenient inall experiments or datasets. Of course, the choice ofthe “optimal" hyper-parameters in each LSTM variant has the potential of achieving strong performance in each variant.

B. Dataset

We collected 3,859 lead sheets with the time signature of 4/4 in MusicXML format from http://www.wikifonia.org. We have made these lead sheets publicly available3 90% of the lead sheets were used as training set and the other 10% were used as validation set. The speed of most music pieces in the dataset is 120 beats per minute.To guarantee the correct segmentation of melodies,all melodies started with weak beats were removedso that we can take bar as a basic unit.[11]

Hsiao-Tzu Hung 9 we have collected a clean Jazz- only dataset et al as[ ]the, target dataset, and a genre-unspecified dataset as the source dataset.. The target dataset, referred to as the CY+R datasethereafter, consists of two small Jazz music collections. The first collection consists of 575 four-bar melody phrases composed by one of the authors, who is a well-trained musician. All phrases are soft Jazz music. The second collection comes from the Jazz Realbook,3 which contains 240 unique songs.

Tian Cheng et al, [12]. We conduct the experiment on the RWC Music Database (Popular Music) [12]. There is a subjective similarity study undertaken on80 songs of the RWC Music Database. In this study27 participants are asked to vote the similarity (on melody, rhythm, vocals and instruments, respectively) for 200 pairs of clips after listening tothem. Each clip lasts for 30 seconds (starting from the first chorus starting time). For these pairs of clips, the similarity votes range from 0 to 27. 2 Thelarger the vote is, the more similar the clips are. Themelodic similarity matrix is shown in Figure 1, indicating the similarity scores of 200 pairs of clips.The matrix is symmetric because if a is similar to b,it means that b is similar to a as well. There are 400non-zero values in the matrix (twice of 200 becauseof the symmetry).

IV. CONCLUSION

In this project, LSTM models are capable of learning harmonic and melodic rhythmic probabilities from polyphonic MIDI files of Bach. The model design was explained, with an eye to key functional principles of flexibility and generalizability. The underlying logic and method of training and generation of algorithmic music were presented. Further, the outputs of the model were analyzed in a quantitative and qualitative fashion. Some suggestions were then put forward for future work.

REFERENCES

[1]. Florian Colombo, Alexander Seeholzer, and Wulfram Gerstner, "DeepArtificialComposer: ACreative Neural Network Model for Automated Melody Generation," 2016 Laboratory of Computational Neuroscience, Brain Mind Institute, School of Life Science, School of Informatics and Communication,EPFL.

[2]. Florian Colombo, Alexander Seeholzer, and Wulfram Gerstner,“Learning to Generate Music with BachProp” 2015 School of Computer Science and School of Life Sciences École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland Dirk-Jan Povel, “Melody Generator: A Device for Algorithmic Music Construction” J. SoftwareEngineering & Applications, 2010, 3, 683-695 Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands.

[3]. PING-HUAN KUO, TZUU-HSENG S. LI, , YA-FANG HO,AND CHIH-JUI LIN] “Development of an Automatic Emotional MusicAccompaniment System by Fuzzy Logic andAdaptive Partition Evolutionary GeneticAlgorithm” 2015, aiRobots Laboratory, Department of Electrical Engineering, NationalCheng Kung University, Tainan 701, Taiwan

[4]. Daniel Kent and Fathi Salem “Performance of Three Slim Variants of The Long Short-Term Memory (LSTM) Layer”. 2019 ireless and VideoCommunications (WAVES) Lab Circuits, Systems, and Neural Networks (CSANN) Lab Department of Electrical and Computer Engineering Michigan State University East Lansing, Michigan, United States of AmericaEmilios Cambouropoulos, “From MIDI to Traditional Musical Notation” Austrian Research Institute for Artificial Intelligence Schottengasse 3, A-1010 Vienna, Austria

[5]. Benjamin Genchel, Ashis Pati, Alexander Lerch“Explicitly Conditioned Melody Generation: A Case Study with Interdependent RNNs” 2017 Center for Music Technology Georgia Institute of Technology Atlanta, GA 30332 USA

[6]. YuanLl,,Sakurako Yazawa, Masatoshi Hamanaka “Melody Generation System based on a Theory of Melody Sequence” 2017 GraduateSchool of Systems and Information Engineering, University of Tsukuba, Tsukuba,Japan.

[7]. Hsiao-Tzu Hung, Chung-Yang Wang Yi-Hsuan Yang, Hsin-Min Wang, “Improving Automatic Jazz Melody Generation by Transfer Learning Techniques Institute of InformationScience, Academia Sinica, Taipei, TaiwanTaiwan AI Labs, Taipei, Taiwan ‡ Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan

[8]. Madhuram M., B. P. Kumar, L. Sridhar, N. Prem, V. Prasad, “Face Detection and RecognitionUsing OpenCV”, 2018

[9]. Jian Wu, Changran Hu, Yulong Wang, XiaolinHu, and Jun Zhu “A Hierarchical Recurrent Neural Network for Symbolic MelodyGeneration”, ‘http://www.wikifonia.org’

[10]. Tian Cheng, Satoru Fukayama, Masataka Goto ,” COMPARING RNN PARAMETERS FOR MELODIC SIMILARITY”, “http://staff.aist.go.jp/m.goto/RWC-MDB/AIS T-Annotation/SSimRWC/.”

Sarthak Somvanshi, et. al. "Melody Generation Using Recurrent Neural Network." *International Journal of Computational Engineering Research (IJCER*), vol. 11, no.4, 2021, pp 01-06.