Fuzzy Music Composition

Yan Zhu and Taylor Rosenfeld

Music lies at the intersection of imagination and shrewd methodology. Musicians must master a multitude of scales, harmonic structures, chord progressions, and countless other musical tools to create meaningful works of art. Balancing careful decision-making and innovative thinking is central to achieving noteworthy musicianship.

Jazz is a compelling musical genre due to its emphasis on both methodology and creativity. One defining characteristic of jazz is the deliberate exploitation of chordal harmony to inform note selection when composing and improvising. In other words, note choice is conditioned on the underlying chord progression. This interaction between chords and melody provides a compelling structure for many musical compositions and lends itself quite well to a computational intelligence application.

We present a neural network/fuzzy hybrid system (NNFHS) for composing music based on a given chord progression. The system produces melodies to accompany chords in a defined progression. The neural network used is an improvisation recurrent neural network (ImprovRNN) adapted from the open-source Google Magenta project. The fuzzy system learns the ImprovRNN mapping and introduces a rule structure. Introducing the fuzzy system allows us to "look under the hood" and gain insight into the ImprovRNN's decision-making process.

SECTION I. Research

There are several powerful neural network based approaches for generating music. Recurrent Neural Networks (RNN) are ideal network structures due to their ability to predict sequences based on previous values [1] [2]. Short memory windows plague RNNs in various applications and limit their use. Long short-term memory (LSTM) networks enhance the RNN memory window and deal with the notorious vanishing gradients issue. Variation of LSTM shows more interesting results on piano, jazz, and pop music [3][4].

Variational Auto-Encoders (VAEs) are another model used for image and music generation [5] [6]. VAEs learn the latent representation of training data based on prior latent variables. Language modeling can enhance VAEs by capturing more complex temporal and melodic structures (variational recurrent auto-encoders, VRAEs) [7]. This is a compelling approach due to the common interpretation of music as a language. Despite this, recent research suggests VRAEs show no significant improvement over LSTMs and introduce more parameters to tune during the training phase.

We settled on an LSTM for its ease of implementation and promising performance. The Google Magenta project provides several open-source LSTM models for music generation. The ImprovRNN specifically conditions melodies on chord progressions and is ideal for our application.

SECTION III. Recurrent Neural Network Used

Our implementation of the ImprovRNN utilizes 3 layers with 256 LSTM at each layer and 0.5 dropout keep probability. Lead sheets were the main source of data and parsed to create input/output pairs of chords and melodies. The model input is a one-hot vector encoding the events at previous note as well as a one-hot vector encoding the current chord. The target output is the next note event. This rich encoding structure is particularly attractive for jazz due to its emphasis on both the previous note, target note, and chord. Moreover, this model sets softmax cross entropy as loss function and uses an Adam optimizer. This model was pre trained on thousands of lead sheets.

SECTION IV. Fuzzy System Implementation

The fuzzy system component of the design used the neural network training data to learn the underlying mapping. Lead sheets are parsed by measure into note events where each note is represented by its pitch, rhythm, and the present chord. Melodies are represented as binned note sequences in order to reduce the problem dimensionality. The fuzzy system then learns the mapping from chords to notes. The current optimal configuration utilizes six Gaussian rules and trained on over two thousand lead sheets.

SECTION V. Evaluation

Soundtracks produced by pre trained neural network models sound close to real music. With fine tuning on the pre trained model, we would be able to train on small number of data on specific genre or artist to generate more interesting music pieces. Yet, we are still working on this part due to the complexity of Magenta's project structure. Besides the complexity of the built project structure, this model is unexplainable from input note sequence to output melody. The last note on neural network will be its long training time while we do hope using a small dataset to fine tune model would alleviate this problem without compromising quality of music generate.

Works Cited

[1] Chu, H., Urtasun, R., and Fidler, S. (2016). “Song From PI: A Musically Plausible Network for Pop Music Generation”. *arXiv preprint,* arXiv:1611.03477.

[2]. Colombo, F., Muscinelli, S. P., Seeholzer, A., Brea, J., and Gerstner, W. (2016). “Algorithmic Composition of Melodies with Deep Recurrent Neural Networks”. *arXiv preprint* arXiv:1606.07251.

[3] Johnson, Daniel D., and Robert M. Keller. "Learning to Create Jazz Melodies Using a Product of Experts." *Iccc 2017* : n. pag. Print.

[4] Malik, Iman, and Carl Henrik Ek. "Neural Translation of Musical Style." N.p., 11 Aug. 2017. Web. 20 Nov. 2017.

[5] K. Gregor, I. Danihelka, A. Graves, and D. Wierstra, “DRAW: A Recurrent Neural Network For Image Generation,” *Icml-2015*, pp. 1– 16, 2014.

[6] R. Sabathé, E. Coutinho and B. Schuller, "Deep recurrent music writer: Memory-enhanced variational autoencoder-based musical score composition and an objective measure," *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, 2017, pp. 3467-3474.

[7] Tikhonov, Alexey and Ivan P. Yamshchikov. “Music generation with variational recurrent autoencoder supported by history.” *CoRR* abs/1705.05458 (2017): n. Pag.

[8] "Leonard Bernstein at 100." *Bernstein and Jazz* | Composer | About | Leonard Bernstein. N.p., n.d. Web. 20 Nov. 2017.