Using LSTM-RNN to Generate Novel Salsa Tracks

Stephanie Brito, Joel Ramirez, Luis Govea

sbrito@stanford.edu joel101@stanford.edu lgovea@stanford.edu

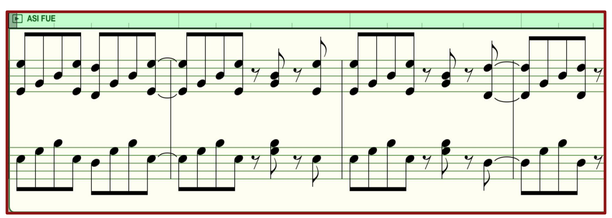
PROBLEM

We are focusing on the space of **computer-generated music**. Most music generation tasks focus on classical music or single instrument music (such as piano). Our team wanted to take on the challenge of generating novel music by **training on salsa tracks.** Salsa, originating form Afro-Cuban culture, is a popular and complex genre. Salsa tracks can have up to 10 instruments with different rhythms, but most salsa tracks follow some **hemiola structure**. To approach this task, we will attempt to isolate different instruments from salsa tracks to generate novel music.

DATASET

- **Data Collection:**

We acquired 13 Salsa music songs that we believed were representative of the genre. In order to facilitate training, the songs were formatted as midi files.



Score for “Asi Fue” - Juan Gabriel

- **Data Processing:**

All songs were broken down into a list of notes and chords. Pitch information was obtained from each object and stored in an array that we utilized to train our model.

EXPERIMENTAL RESULTS

- **Developing Novel Track**

The model was trained for a total of 500 epochs utilizing a batch size of 32. During development of new tracks, input tracks were broken down into 4-note sequences and fed into the neural network.

- **Measuring Accuracy**

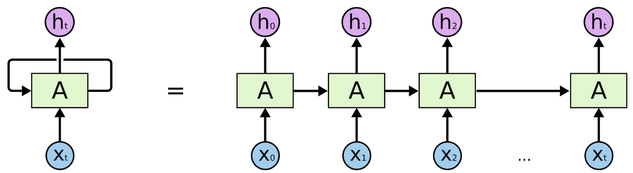
As a way to measure the success of our model we compared the presence of each 4-note sequence from the input to its corresponding output song. We encountered a high degree of similarity between input and output, suggesting our model was able to accurately extract patterns from the input songs.

Song	Accuracy
Mujer de Arena	23.8%
La Salsa	18.8%
Rumbera	16.1%
Usted Abuso	16.1%
Lloraras	10.8%
Mi Tierra	10.7%
Bamboleo	5.5%

METHODS

- **LSTM-RNN**

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) that enables the model to recreate information based on patterns that are fed into the network through the use of feedback connections that enables the network to remember information for a long period of time in order to create a new prediction.



- **Architecture**

We decided to implement a 4-layered fully-connected network for this project. We utilized a dense network to allow retention of previous output (notes) to generate future outputs in the model. For the last layer, we utilized a softmax function in order to acquire individual probabilities for all notes to allow us to choose the most likely note to follow the input sequence.

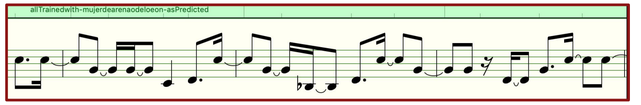
- **Loss Function**

Due to the fact that each of our outputs belongs to an individual class and we have more than 2 classes, we decided to utilize a categorical-cross entropy loss function.

$$CE = -\sum_{c=1}^{M} y_{i,c} \log(p_{i,c})$$

DISCUSSION

1. Through trial-and-error we discovered that it is highly complex to train the model on multiple instruments simultaneously to generate complex salsa music.
2. Model is widely impacted by the start and end portions of the training songs which have a high degree of variation in salsa music.
3. Looking at our results, we believe that most of the predictions failed due to the propensity of the model to replicate our data as opposed to generating novel results. More data is needed to develop new, rhythmically pleasant salsa music.



Score for LSTM generated track

FUTURE WORK

1. Acquire more salsa midi files to avoid overfitting of model.
2. Develop a way to train model with different instruments simultaneously and output more harmonic structure rather than melodic.
   a. This will enable outcome to more closely resemble salsa music
3. Continue tuning hyperparameters to ensure minimum loss and production of coherent salsa music.

REFERENCES

1. Payne, Christin. “MuseNet.” OpenAI, 25 Apr. 2019, openai.com/blog/musenet
2. “WaveNet: A Generative Model for Raw Audio.” Deepmind, https://deepmind.com/blog/article/wavenet-generative-model-raw-audio