Bach to the Future

Gabby Asuncion [gasuncio], Sophia Chen [schen96],
Kelvin Yang [kyang35], Stanley Yip [syip2]

November 2019

1 Introduction

What problem are you trying to solve and why?

1. If you are implementing an existing paper, describe the paper's objectives and why you chose this paper.
   We will be implementing an existing paper titled "Deep Learning in Music." The objectives of this paper are to (1) represent music notes as vectors - that is, translate language learning algorithms like word2vec to music and (2) build network architectures that can learn/reflect concepts like harmony and melody. The authors describe earlier papers that have approached this problem to limited success (e.g., fails to distinguish melody vs harmony, fails to learn a long term musical structure, etc.), and seek to expand on past work in this area.

2. What kind of problem is this? Classification? Regression? Structured prediction? Reinforcement Learning? Unsupervised Learning? etc.
   The problem we have set out to solve is structured prediction in that the model will train on existing melodies from our data and generate music based on the melodic patterns that it sees in the training set.

2 Related Work

1. Are you aware of any, or is there any prior work that you drew on to do your project?
   We are aware of several projects that have implemented a similar model in the attempt to reproduce musical composition. Some examples that we have seen are Magenta, a research project exploring machine learning in the context of creative expression, and MuseNet, a neural network that generates four minutes compositions comprised of ten different instruments and combining several different styles.

3 Data

1. What data are you using (if any)?
   http://www-etud.iro.umontreal.ca/ boulanni/icml2012
   http://piano-midi.de/
   http://classicalmidiresource.com/
   http://musedata.org/

2. If you're using a standard dataset (e.g. MNIST), you can just mention that briefly. Otherwise, say something more about where your data come from (especially if there's anything interesting about how you will gather it).

   We will be using the corpus files linked in the paper we are referencing, which are standard representations of music that can be processed and used for training our model. There are two formats – raw files and piano roll representation, both of which are in MIDI format. The piano roll representation can be thought of as "positionally invariant" pieces, all being transposed to the same key.

3. How big is it? Will you need to do significant preprocessing?
   Our data can be as large as we'd like. To follow the paper, we'll end up picking 417 of Bach's pieces for a total of about 1.6 million tokens. We will need to pickle the MIDI files, but the first link above describes the process.

4 Methodology

1. What is the architecture of your model?
   This model uses a 2-layered LSTM RNN architecture that treats notes as characters/words to predict the next note in a sequence. One thing we'd like to try is to run the output through convolution layers and calculate loss with the convolved version of an input and an output. This way, hopefully the model can learn both the general structure and the specific rules.

2. How are you training the model?
   The method of training will be similar to previous projects we've completed for this class. We will create an embedding matrix that maps learned "tokens" (i.e., notes) into vector representations and feed these embeddings into the LSTM.

3. If you are implementing an existing paper, detail what you think will be the hardest part about implementing the model here.
   I think the hardest part about implementing the model will be the preprocessing the data. In this project, we won't really have any guidance on how to preprocess the data, but other parts of the project are described in great detail in the paper.

5 Metrics

1. What constitutes "success?"
   We will be successful if we can feed a model the same sequence of some number of measures as an existing Bach piece. We can compare the structure of the convolved output to the structures of other Bach pieces, and if the structure is within a certain similarity, we are successful in mimicking the structure of Bach pieces. We will also try to attempt the qualitative measurement that the researchers in the paper used to measure success.

2. For assignments, we have looked at the accuracy of the model. Does the notion of accuracy apply for your project?
   The notion of accuracy applies when we're talking about comparing the structure of the pieces our model outputs and the structure of Bach's pieces.

3. If you are implementing an existing project, detail what the authors of that paper were hoping to find and how they quantified the results of their model.

   One of the major measures of success that the authors of the paper used was based on evaluating the musical aesthetic of the pieces produced. They measured this based on a blind experiment where 26 volunteers listened to 3 samples of generated music, and rated them on a 1 to 10 scale. The following metric was used to quantify the scale:

   - 1 rating - "completely random noise"
   - 5 rating - "musically plausible"
   - 10 rating - "composed by a novice composer"

   The first sample was a clip of an actual Bach composition, the second sample was a clip from the output of the RNN model trained on raw MIDI files, while the third sample was a clip from the output of the RNN model trained on the positionally invariant piano-roll MIDI files.

6 Ethics

Choose 2 of the following bullet points to discuss; not all questions will be relevant to all projects so try to pick questions where there's interesting engagement with your project. (Remember that there's not necessarily a ethical/unethical binary; rather, we want to encourage you to think critically about your problem setup.)

1. Why is Deep Learning a good approach to this problem? What broader societal issues are relevant to your chosen problem space?

   We are not entirely sure if Deep Learning in this application is necessarily a good approach to the problem. This project challenges the idea of what constitutes art and whether or not a neural network can replace a human composer. Of course, our intention is not necessarily to replace the artist but to further explore the use of Deep Learning in generating sound.

2. What is your dataset? Are there any concerns about how it was collected, or labeled? Is it representative? What kind of underlying historical or societal biases might it contain?

   Our dataset is a set of Bach's pieces. We don't see any concerns about how it was collected or labeled, as these pieces have been well preserved. Our model will be biased towards Bach and Bach's composing style, but this is the goal of our project.

3. Who are the major "stakeholders" in this problem, and what are the consequences of mistakes made by your algorithm?
   Some of the major stakeholders in this problem include composers and music theorists. If the algorithm makes mistakes, this may not matter, but if the algorithm does really well, it poses a threat to people who compose music, as well as theorists who study music, particularly if our algorithm breaks traditional music theory rules of the time.

4. How are you planning to quantify or measure error or success? What implications does your quantification have?

5. Add your own: if there is an issue about your algorithm you would like to discuss or explain further, feel free to do so.

   In the future, as these models and architectures become more complex, they may become better and better at mimicking creativity. Is creativity something that a machine can learn? Is there a difference between human creativity and mechanical creativity? These are questions that we'd like to keep in mind and think about during this project.

7 Division of Labor

Briefly outline who will be responsible for which part(s) of the project.

Because we've done projects similar to this in the past, we think it'd be most beneficial if we all worked together to write the code. Then, we can individually modify hyperparameters to improve our model.