Music Demo

Melody-to-Chord using paired model and multi-task learning language modeling

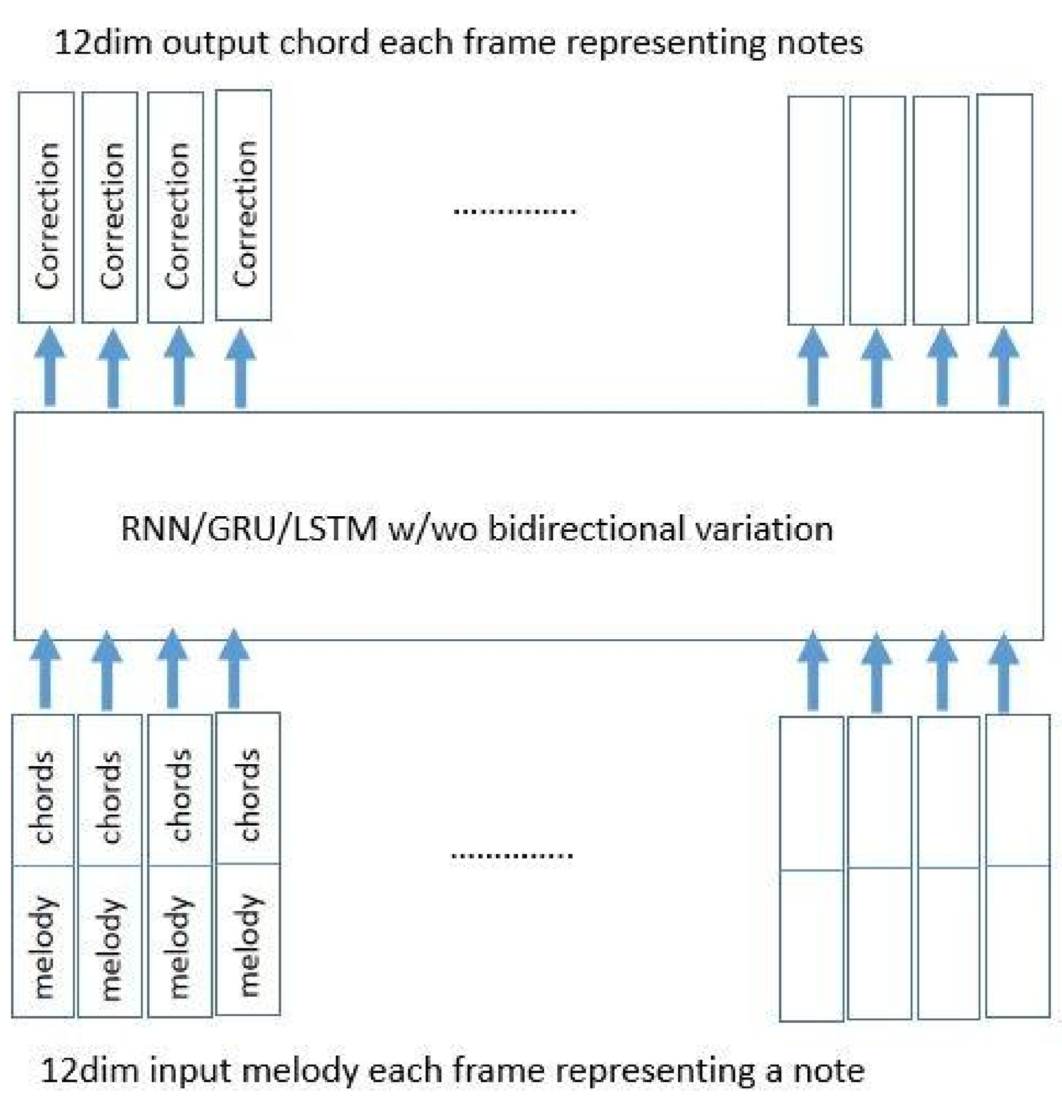
Mu-Heng Yang, Wei-ting Hsu, Nicholas Huang

Problem Definition

We designed a system to generate chords matching a given melody using several models, drawing analogy between **natural language** and **music**. We mainly focused on two models: a **proposed paired** model and a **multi-task language** model. Both models generated chords matching melody in harmony. We evaluate the result using L1 error. We found the one-hot model generates better result in evaluation scheme but proposed paired model generates more pleasing music.

Pair Model

- Given a melody/chord pair, the task is to determine whether the melody and chords are compatible or a good match.
- predict the "correction" from the input chord to ground truth chord in vector form

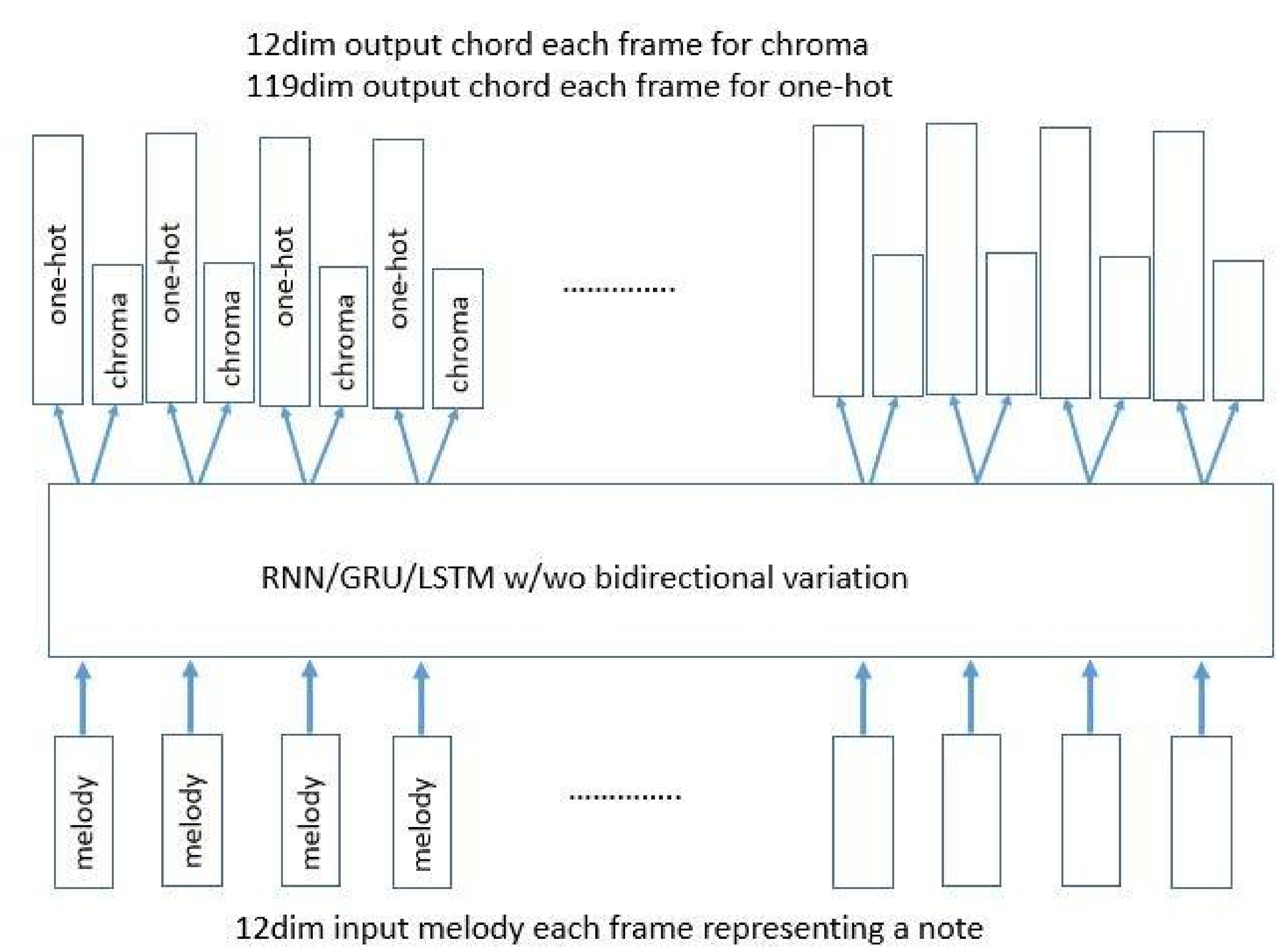


Dataset

- Collected datasets of MIDI files, each with 8 measures of melody and chords.
- partitioned MIDI into a melody and chord track, and parse into 12 dim vector for each timestep, where each timestep represent a sixteenth note (semiquaver)
- Randomly sampled 10% clips as testing data set, and used the remainders as the training data set.
- Also converts chord to one-hot representation

Language Model

- Treats melody as input and chords as label at each timestep and attempts to predict the most suitable chords to an unseen melody. Musical chords, the ones our model will predict, can be described in both one-hot representation and chroma format.
- Use multitask learning strategy to learn the relationship between melody and chord in both the chroma and one-hot representations.



Results

The paired model has lower L1 error per time frame, however, it can often generate more pleasing music, The multi-task model is better at driving down the L1 error per time

Figure 1 shows error update for multi-task learning in language modeling. We expect to see the U-curve loss plot. However, because one-hot dimension >> 12, it's difficult to balance the contributions, and one-hot representation dominates the shared weights. We also experimented with late-fusion ensemble learning on both output. Nonetheless, it was not enough to offset the imbalance no improvement

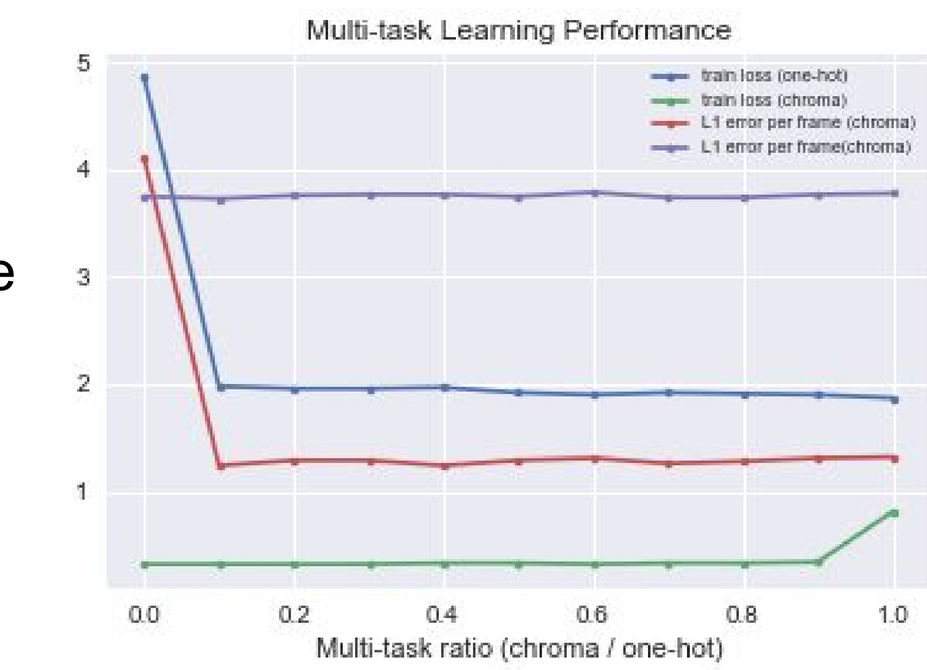


Figure 1: Final error of different multitask ratio. 0 means only training chroma task, 1.0 means only training one-hot

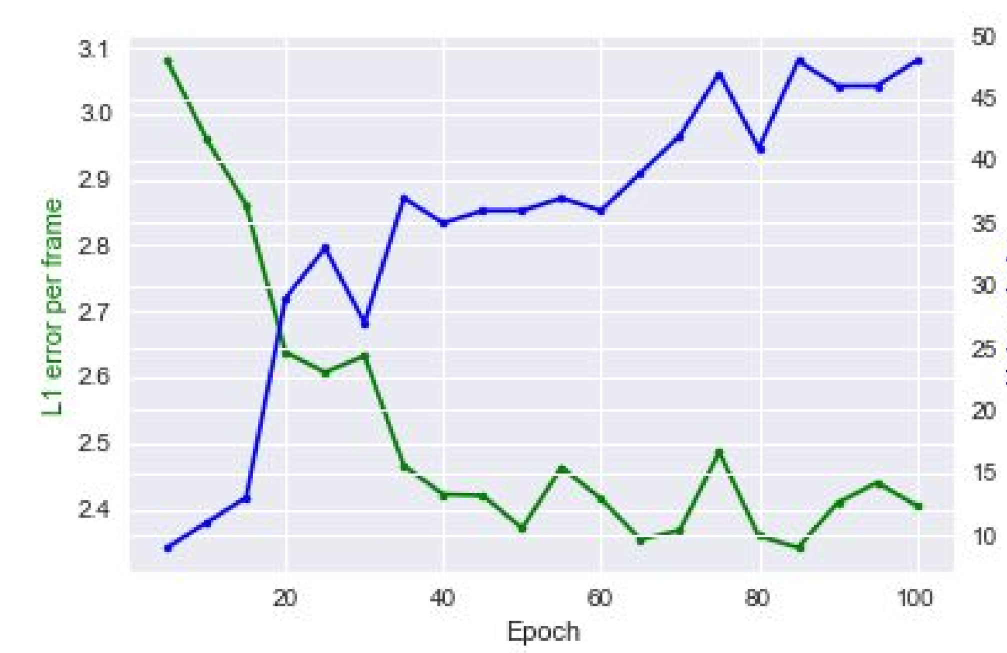


Figure: Error of Multi-task Learning (ratio = 0.5)

Challenges

- Most music are protected by copy right, leading to only few dataset available for research
- expensive human-labor to precisely label the chords and melody
- Music has time-notion, lanuage model has a dificulty deciding the best time for chord transition.
- Music is subjective, no absolute metric to define good match