ChordGAN: Symbolic Music Style Transfer with Chroma Feature Extraction

Conan Lu, Washington

ABSTRACT

This project introduces ChordGAN, a generative adversarial network that transfers the style elements of tonal music genres. Early attempts at symbolic music style transfer have faced the challenge of preserving specific content features like chords while simultaneously changing style, which includes aspects like rhythm and ornamentation. ChordGAN seeks to learn the rendering of harmonic structures into notes using a novel process that embeds chroma feature extraction within the training process. In noted music, the chroma representation approximates content features as it only takes into account the pitch-class of musical notes, representing multiple notes collectively as a density of pitches over a short time period.

The base logic is that a piece should retain its style features by modifying the note configurations. Think: pop song as a sonata. Our project introduces ChordGAN, a generative adversarial network that transfers the style elements of tonal music genres. Early attempts at symbolic music style transfer have faced the challenge of preserving specific content features like chords while simultaneously changing style, which includes aspects like rhythm and ornamentation. ChordGAN seeks to learn the rendering of harmonic structures into notes using a novel process that embeds chroma feature extraction within the training process. In noted music, the chroma representation approximates content features as it only takes into account the pitch-class of musical notes, representing multiple notes collectively as a density of pitches over a short time period.

The base logic is that a piece should retain its style features by modifying the note configurations. Think: pop song as a sonata.

MODEL ARCHITECTURE

The Conditional GAN architecture features two networks: the generator and discriminator. The generator’s objective is to create the most realistic music sample, while the discriminator’s task is to distinguish between real and generated samples.

**Base GAN**

The Conditional GAN architecture features two networks: the generator and discriminator. The generator’s objective is to create the most realistic music sample, while the discriminator’s task is to distinguish between real and generated samples.

**Unique Additions**

- Conditional GAN uses additional parameter (chroma) for discriminator.
- Discriminator is trained jointly.
- Discriminator architecture.
- Generator input (z) changed from random noise to chroma.
- Chromatic distance.
- ChordGAN.
- ChordGAN: Symbolic music style transfer with chroma feature extraction.
- Table 1
- Table 2
- With independent data, genre prediction accuracy was 80%. Once classifier was finished training, fifty post-transfer samples were given as input. Results shown in Table 2.

PROPOSAL

ChordGAN accomplishes style transfer in the symbolic domain of music using a novel technique, combining conditional generative adversarial network (GAN) architecture and the chromagram representation of music.

**Implementation**

**Dataset**

Datasets used genres of pop, general classical, general jazz, Bach, Mozart, and Haydn. Pop, jazz, and classical datasets were used for training purposes. To evaluate the success of the transfer, I used two metrics: Tonnetz distance, to measure harmonic similarity, and a separate genre classifier, to measure transfer realism. The success of the transfer is evidenced by the high independent genre classifier accuracy rate and near-zero Tonnetz distance. Using chroma feature extraction offers significant advantages, as the format improves transfer consistency while requiring less data to train. ChordGAN can be utilized as a tool for musicians to study compositional techniques. In addition, experiments using lead sheets in conjunction with ChordGAN present the possibility of automatic music generation.

**Evaluation**

Content Evaluation: Tonnetz

I use Tonnetz to measure the content (defined as harmonic structure) similarity of pieces before and after transfer. Tonnetz is a graph representation of the harmonies within a piece of music, and the distance between two Tonnetz matrices is the Tonnetz distance. A lower Tonnetz distance between two pieces of music signifies more similar harmonic structures.

**Style Evaluation: Genre Classifier**

Style evaluation is measured via an independent genre classifier.

<table>
<thead>
<tr>
<th>Model</th>
<th>Pop Accuracy</th>
<th>Jazz Accuracy</th>
<th>Classical Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>80%</td>
<td>68%</td>
<td>74%</td>
</tr>
<tr>
<td>ChordGAN</td>
<td>80%</td>
<td>74%</td>
<td>64%</td>
</tr>
</tbody>
</table>

**Conclusion**

Based on style and content evaluations, ChordGAN presents a successful novel approach to style transfer in the symbolic domain. It specifically targets the style involved in rendering harmonic structures to notes. Limitations include narrow scope of style—no long-term structure—and uneven training for genres like jazz.

This modality presents significant implications for accessible computational creativity, as it requires relatively small amounts of data to draw inferences while maintaining consistency.

One major application of this work is in automatic music generation, especially in collaboration with humans. By substituting chroma for lead sheets and initial chords created by a user, ChordGAN has the capacity to be a useful compositional tool. Beyond music, the novel split between style and content features can improve models for areas like speech synthesis and image manipulation.

REFERENCES


doi: 10.1007/978-3-319-30913-4_3

