

AIを用いた楽曲制作に関する検討

1532117 秋場 翼

1532151 松元 孝樹

指導教員：中村 直人 教授

平成31年度

目次

第1章	序論	1
1.1	研究の背景と目的	1
1.2	本論文の構成	2
第2章	理論	3
2.1	AIを用いた楽曲作成	3
2.1.1	MIDI	3
2.1.2	Magenta	3
2.2	開発環境の構築	4
第3章	実験内容	5
3.1	モデルによる違い	5
3.1.1	MelodyRNN	5
3.1.2	PolyphonyRNN	6
3.2	学習回数による違い	6
3.3	ノード数による違い	6
第4章	楽曲制作	7
4.1	NoteSequenceについて	7
4.2	Melody_rnnを使用して学習モデルを使用	8
4.2.1	NoteSequenceの作成	8
4.2.2	学習の開始	8
4.2.3	音楽データの作成	9
4.3	事前に学習済のモデルを使用	10
4.4	Polyphony_rnnを使用して学習モデルを使用	10
第5章	結論	11
5.1	今後の課題	12
	謝辞	13
	参考文献	14

目 次

1.1	magenta による MIDI 音楽データ生成までのプロセス	1
2.1	magenta による MIDI 音楽データ生成までのプロセス	3
3.1	NoteSequence の作成	5
4.1	NoteSequence の作成	8
4.2	NoteSequence を学習用と評価表に分割	8
4.3	BasicRNN を使用した学習の開始	9
4.4	学習モデルを使用し 10 曲を作成	9
4.5	事前学習済モデルの使用	10

表 目 次

2.1 開発環境	4
--------------------	---

第1章 序論

1.1 研究の背景と目的

近年, AI 分野は急速な発展を続けている. スマートスピーカなどの対話型の AI が Google や Amazon, IBM によって商品化され, 現在ではスマートフォンにも Siri という AI が搭載されるなどその存在は非常に身近になっており, その種類も非常に多岐にわたる.



図 1.1: magenta による MIDI 音楽データ生成までのプロセス

また囲碁や将棋, チェスなどの競技においても, プロに AI が勝利するなどその精度は以前から高いが, その AI は一つの競技でしか使用できない特化型人工知能 (AGI) でありった. しかし, 英 DeepMind が発表した AlphaZero という様々なボードゲームに対応できる汎用性を持った AI が発表され, 汎用人工知能 (GAI) の成長も著しい.

自然言語処理を用いた芸術の分野では, 2012 年にスタートした人工知能を使って小説を生成するプロジェクトが「星新一賞」の第一審査を通過した. また, 絵画や音楽に関しても AI が作成した肖像画が米競売大手クリスティーズのオークションで 43 万 2500 ド

ル（約 4900 万円）で落札され, AI を用いて新しい作品を作るものが出回っている。

このように AI の発展は様々な分野においてその成果を上げており, 今後は業務の効率化や補助だけにとどまらず, 自動車の自動運転や医療の現場でも人間の手よりも高精度なものとして活躍することが期待されている。

本研究では AI による楽曲生成についての実証実験を行う。Google brain によって公開されている Tensorflow のライブラリである Magenta は AI Duet やそのライブラリを用いて学習データやノード数による楽曲の生成結果の違いを比較, 検証し, AI による楽曲制作が有用なものか調査する。

1.2 本論文の構成

本論文の構成は以下の通りである。

第 1 章では本論文の背景と目的について述べている。

第 2 章では本論文で利用する理論について述べている。

第 3 章では実験内容について述べている。

第 4 章では楽曲制作について述べている。

第 5 章では AI を用いた楽曲制作についての本研究の結論について述べている。

第2章 理論

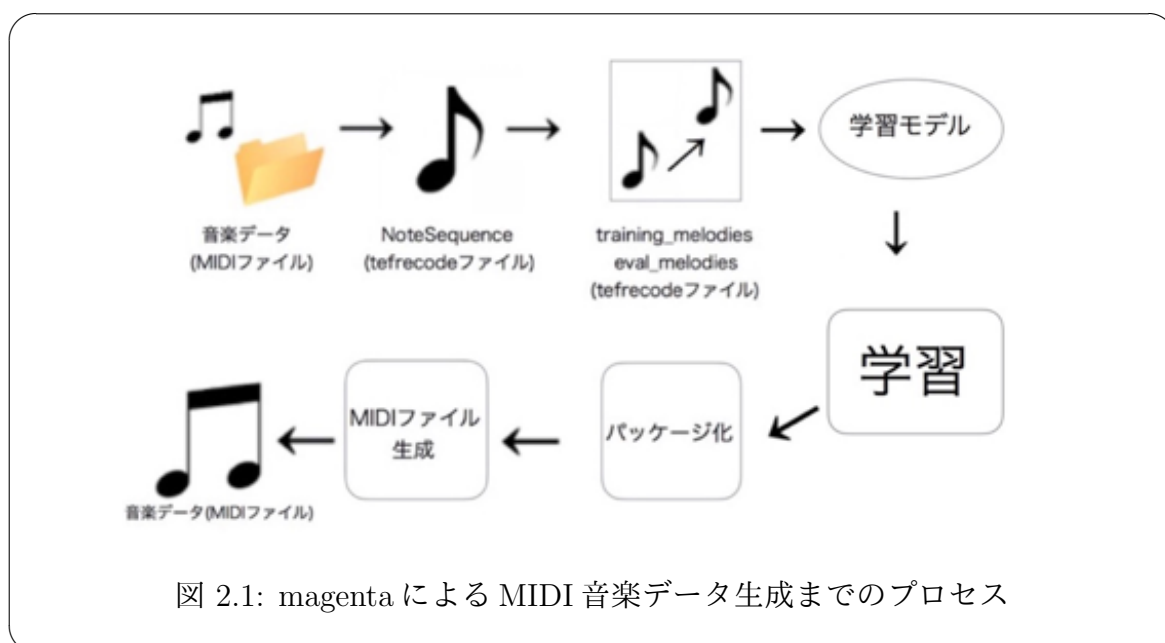
2.1 AIを用いた楽曲作成

2.1.1 MIDI

AIによる曲制作では主に MIDI ファイルの音楽データを使用する. MIDI ファイルには実際の音ではなく音楽の演奏情報（音の高さや長さなど）である. 本研究で用いる AI はこの MIDI ファイルの情報を元に学習をする. また入出力の際もこの規格を用いる.

2.1.2 Magenta

本研究では Magenta[1] を使用する. これは音楽などを TensorFlow を使って機械学習するライブラリであり, Google Brain が GitHub 上に公開している. Magenta ではまず学習させたい音楽の MIDI データを NoteSequence (magenta が扱うファイル形式) とよばれるデータフォーマットに変更する. それを学習用データセットと評価用データセットに変換したあと学習を行う. このとき, 一度に学習させるデータの数, 学習を行う回数, ノード数を設定する. これをパッケージ化し, MIDI ファイルとして新たに楽曲を生成するという流れである. これを図 2.1 に示す.



2.2 開発環境の構築

開発環境の構築には二つ方法があり, 一つは Docker というコンテナ型の仮想化環境を構築できるオープンソースソフトウェアを用いる方法と, ローカル環境に Python のパッケージ管理システムである pip を用いて構築する方法の 2 つがある.

Docker を用いることで DockerHub という仮想化環境をクラウド上で共有できるサービスを使用できるため, 短時間でパッケージのインストールをおこなわずに環境を構築できる. しかし, Docker に GPU を割り当てることは Docker に精通していないと難しいため, 本研究ではローカル環境に開発環境を構築した. 本システムの開発環境を表 2.1 に示す.

表 2.1: 開発環境

OS	OS X Yosemite
CPU	Intel Core i5
メモリ	8GB
使用ライブラリ	TensorFlow , magenta

本システムは Macbook Pro を使用し OS は OS X Yosemite を使用した. 使用するライブラリとして, ニューラルネットワークを構築できる Tensorflow を利用した.

第3章 実験内容

3.1 モデルによる違い

本研究では Magenta で用意されている 2 つの学習モデルを用いた。以下に示すモデルを用いて楽曲制作をそれぞれ行い，比較，検証する。

3.1.1 MelodyRNN

MelodeRNN は楽曲のメロディを制作するモデルである。MelodyRNN である 3 つのモデルを以下に示す。

(1) basic_rnn

前の状態を保持し，これを記憶または忘却する。時系列を学習することにより，次の音の予測を可能にしている。Lookback_rnn と Attention_rnn はこれを基に機能を追加したものである。

```
convert_dir_to_note_sequences \  
  --input_dir=$INPUT_DIRECTORY \  
  --output_file=$SEQUENCES_TFRECORD \  
  --recursive
```

図 3.1: NoteSequence の作成

(2) lookback_rnn

Basic_rnn を基に，1 小節前と 2 小節前の音，拍数，前の小節の繰り返しかどうかの情報を与え，音楽の流れを掴もうとするもの。

(3) attention_rnn

basic_rnn を基に，過去の情報を予測結果に加えてこれによる繰り返しを捉えるもの。

3.1.2 PolyphonyRNN

複数の同時音のモデリングが可能になっており、複数音の響きを1つのかたまりとして捉えて学習しているモデルである。このモデルを使用することで、伴奏も含めた楽曲の生成が可能である。

上記のモデルを用いて制作を行い、それぞれの違いと有用性について検証する。

3.2 学習回数による違い

学習回数を変更して楽曲制作を行い、それぞれの違いと有用性について検証する。

3.3 ノード数による違い

ノード数を変更して楽曲制作を行い、それぞれの違いと有用性について検証する。

第4章 楽曲制作

4.1 NoteSequence について

NoteSequence とは MIDI データから作成されるプロトコルバッファである。プロトコルバッファとは Google が 2008 年にオープンソース化したバイナリベースのデータフォーマットである。既存の技術としては XML や JSON などのテキストベースのデータフォーマットがあるが、プロトコルバッファはバイナリフォーマットであるので、アプリケーション間でデータ構造の送受信をする際に少ないデータ量ですむという特徴がある。スキーマ言語はなぜ重要なのか、そういう時代になったからだ。我々が単一 RDB に接続する単一の Web アプリだけを書いている時代はとうの昔に終わった。データはあちこちのいろんなストレージ技術で保存されているかもしれないし、バックエンドも単一サービスではなくて分割されているかもしれないし、クライアントは web 版, iOS 版, Android 版があってひょっとしたらそれぞれ別の言語で実装されており、外部開発者向けに API も公開しなければならない。だから、そこら中でデータをシリアルライズするしデシリアルライズするし、通信の両端で解釈に矛盾が無いようにすり合わせる必要がある。でも、すりあわせの目的でいちいち人間と自然言語で会話するのは苦痛なので、私たちは機械処理可能なコードで語りたい。だから、どういうデータがやってくるのかきちんと宣言的 DSL で定義しておきたい。そこでスキーマ言語だ。JSON Schema が広がりつつあるのもたぶんそういう理由だし、そもそも Protobuf 自体も Google が社内で同じ問題にぶち当たって発明されたんだったような気がする。自分の管轄領域内に閉じたデータ構造であれば、スキーマとか型宣言とか細かい縛りなしで軽量に進めるのもありだ。しかし、他人の領域との界面はしっかり定義しておいた方がよい。スキーマ定義という税金を支払わないとその分のツケはどこかで回ってきて、1 時間掛かる E2E テストがこけるとか、週次変更レビュー合同会議の発足とか、そういうやつで支払うことになる。スキーマ定義さえあれば、web クライアント開発用の、バックエンド開発用の、アプリ開発用の言語に向けて対応するデータ構造定義を自動生成して、矛盾なくすべてをメンテナンスできる。通信の両端でスキーマさえ合致していれば、シリアルライゼーション形式を合わせるのはむしろ間違えづらい。だから、JSON でも Protobuf のデフォルトのそれでも、MessagePack でも XML でも好きにすればよい。ちなみに、Protobuf で定義したデータ構造は Protobuf 標準形式の他に JSON にもきちんとシリアルライズできる。

4.2 Melody_rnn を使用して学習モデルを使用

4.2.1 NoteSequence の作成

NoteSequence の作成は図 4.1 に示すコマンドで作成できる。
-input_dir で学習させる MIDI データのディレクトリの絶対パスを指定し、-output_file で NoteSequence の出力先のディレクトリを指定する。

```
convert_dir_to_note_sequences \  
  --input_dir=$INPUT_DIRECTORY \  
  --output_file=$SEQUENCES_TFRECORD \  
  --recursive
```

図 4.1: NoteSequence の作成

次に作成した NoteSequence のデータセットを学習用と評価用に分割するために、図 4.2 に示すのコマンドを実行する。

-config で使用する RNN を指定する。-input_dir で NoteSequence の絶対パスを指定し、-output_file で分割した NoteSequence の出力先のディレクトリを指定する。-eval_ratio で NoteSequence のデータを何パーセント学習用に用いるかを指定する。コマンドの場合は 10% が学習用のデータになる。

```
melody_rnn_create_dataset \  
  --config=<one of 'basic_rnn', 'lookback_rnn', or 'attention_rnn'> \  
  --input=/tmp/notesequences.tfrecord \  
  --output_dir=/tmp/melody_rnn/sequence_examples \  
  --eval_ratio=0.10
```

図 4.2: NoteSequence を学習用と評価表に分割

4.2.2 学習の開始

作成した NoteSequence から学習モデルを作成するために図 4.3 のコマンドを実行する。

`-config` で学習に使用する学習モデルを指定,`-rundir` で学習のために用意した Notesequence を指定し,`-sequence_examplefile` で学習モデルの出力先のディレクトリを指定する. `-hparams` でメモリの使用量を指定し,`-rnn_layer_size` で中間層のノード数を指定し,`-num_trainingsteps` で学習回数を設定する.

```
melody_rnn_train \  
--config=attention_rnn \  
--run_dir=/tmp/melody_rnn/logdir/run1 \  
--sequence_example_file=/tmp/melody_rnn/sequence_examples/training_melodies.tfrecord \  
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \  
--num_training_steps=20000
```

図 4.3: BasicRNN を使用した学習の開始

4.2.3 音楽データの作成

図 4.4 に示すコマンドで学習モデルに入力する.

`-config` で学習に使用する学習モデルを指定,`-rundir` で学習済みのモデルを指定し,`-output_dir` で音楽データの出力先のディレクトリを指定する. `-num_outputs` で生成する音楽データの個数を指定し,`-num_steps` で`-hparams` でメモリの使用量を指定し,`-rnn_layer_size` で中間層のノード数を指定し,`-primer_melody` で学習モデルに入力する最初の音程を MIDI の形式で指定する.

```
melody_rnn_generate \  
--config=attention_rnn \  
--run_dir=/tmp/melody_rnn/logdir/run1 \  
--output_dir=/tmp/melody_rnn/generated \  
--num_outputs=10 \  
--num_steps=128 \  
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \  
--primer_melody="[60]"
```

図 4.4: 学習モデルを使用し 10 曲を作成

4.2.4 事前に学習済のモデルを使用

また,Magenta プロジェクトにすでに学習済のモデルが存在するのでそれを使用して音楽データを作成することもできる. 生成には mag バンドファイルが必要になるので magenda の Github に公開されているので,それをダウンロードしてくる. その後図 4.5 に示すコマンドを実行する事で生成することができる.

-config で学習に使用する学習モデルを指定,-rundir で学習済みのモデルを指定し,-output_dir で音楽データの出力先のディレクトリを指定する. -num_outputs で生成する音楽データの個数を指定し,-num_steps で-hparams でメモリの使用量を指定し,-rnn_layer_size で中間層のノード数を指定し,-primer_melody で学習モデルに入力する最初の音程を MIDI の形式で指定する.

```
melody_rnn_generate \  
--config=${CONFIG} \  
--bundle_file=${BUNDLE_PATH} \  
--output_dir=/tmp/melody_rnn/generated \  
--num_outputs=10 \  
--num_steps=128 \  
--primer_melody="[60]"
```

図 4.5: 事前学習済モデルの使用

4.3 Polyphony_rnn を使用して学習モデルを使用

h

第5章 結論

5.1 今後の課題

aaa

謝辭

参考文献

- [1] <https://www.zaikai.co.jp/article/20160323/299468.html>
- [2] <https://headlines.yahoo.co.jp/hl?a=20181026-00010002-afpbbnews-int>
- [3] <https://news.mynavi.jp/article/20080718-protocolbuffer/>
- [4]