

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
1.1.1	動画共有サイトの発展	1
1.1.2	音楽と著作権	1
1.1.3	AIによる作品と著作権	2
1.1.4	AI 楽曲作成サービス	2
1.1.5	研究の目的	3
1.2	本論文の構成	4
第2章	理論	5
2.1	AI の理論	5
2.1.1	AI の始まり	5
2.1.2	第1次 AI ブーム	6
2.1.3	第2次 AI ブーム	6
2.1.4	第3次 AI ブームの始まり	7
2.1.5	第3次 AI ブームの現在	8
2.2	AI を用いた楽曲作成	9
2.2.1	MIDI	9
2.2.2	Magenta	11
2.2.3	リカレントニューラルネットワーク (RNN)	12
2.3	機械学習に適した開発環境について	13
2.3.1	CUDA	14
第3章	実験内容	15
3.1	モデルによる違い	15
3.1.1	MelodyRNN	15
3.1.2	PolyphonyRNN	18
3.2	学習回数による違い	19
3.3	ノード数による違い	19
第4章	楽曲制作	20
4.1	Melody_rnn を使用して学習モデルを作成	20
4.1.1	NoteSequence の作成	20
4.1.2	basic_rnn を用いて学習を開始	21
4.1.3	lookback_rnn を用いて学習を開始	22
4.1.4	attention_rnn を用いて学習を開始	22
4.1.5	音楽データの作成	23

4.1.6	事前に学習済のモデルを使用	23
4.2	Polyphony_rnn を使用して学習モデルを作成	24
4.2.1	NoteSequence の作成	24
4.2.2	学習の開始	24
4.2.3	音楽データの作成	25
第 5 章	結論	26
5.1	AI による楽曲制作の結果	26
5.1.1	モデルによる生成結果の違い	26
5.1.2	学習回数による生成結果の違い	27
5.1.3	ノード数による生成結果の違い	28
5.2	調査結果	29
5.2.1	学習回数ごとの各モデルの調査結果	30
5.2.2	各モデルごとの学習回数の調査結果	31
5.2.3	ノード数ごとの調査結果	32
5.3	今後の課題	33
	謝辞	34
	参考文献	35

目 次

1.1	Amper Musick	2
2.1	チューリングテストを合格した EugeneGoostman (引用 : http://www.itmedia.co.jp/news/articles/1807/26/news014.html)	5
2.2	クイズ番組に出演する Watson (http://www.swift-web.org/cp-bin/blogn/index.php?e=1023)	7
2.3	多種多様なスマートスピーカー	8
2.4	インプットデータの仕組み	10
2.5	MIDI と音階	10
2.6	GitHub 上に公開されている Magenta	11
2.7	magenta による MIDI 音楽データ生成までのプロセス	12
3.1	basic_rnn のモデル (1)	15
3.2	basic_rnn のモデル (2)	16
3.3	三つのモデルについて	16
3.4	lookback_rnn のモデル	17
3.5	attention_rnn のモデル	18
5.1	学習モデルごとの生成結果	26
5.2	basic_rnn による学習回数ごとの生成結果	27
5.3	lookback_rnn による学習回数ごとの生成結果	27
5.4	attention_rnn による学習回数ごとの生成結果	27
5.5	ノード数ごとの生成結果	28
5.6	調査の様子	29
5.7	学習回数ごとの各モデルの調査結果	30
5.8	各モデルごとの学習回数の調査結果	31
5.9	ノード数ごとの調査結果	32

表 目 次

2.1	開発環境	13
4.1	JSON とプロトコルバッファの比較	20

第1章 序論

1.1 研究の背景と目的

1.1.1 動画共有サイトの発展

スマートフォンの発展とともに2016年の日本のインターネット利用者数は1億84万人、人口普及率は83.5%となった(総務省2017)。インターネット上の映像を容易に再生可能なスマートフォンやタブレットなどのデバイスが普及したことで、DVD等と同等の高品質な映像を視聴、配信可能なYoutube[1]やニコニコ動画[2]などの動画共有サービスの発展も著しく、コンテンツデリバリーサービス(Content Delivery Service:CDS)や各種クラウドサービスの発展により、提供側のサーバなどの構築に必要な費用が非常に低くなる中で、インターネット上に多様な映像配信サービスが提供されている。誰でも利用でき、非常に幅広いコンテンツが投稿、視聴されているが、近年、動画内に使用されている楽曲などの著作権を侵害する動画が投稿されるケースが多く社会問題になっている。

1.1.2 音楽と著作権

著作権法では、著作物を「思想又は感情を」「創作物に」「表現したもの」で、「文芸、学術、美術又は音楽の範囲に属するもの」と定義している。音楽の著作物には、曲のほかに歌詞も含まれ、録音や記譜されている必要はなく、即興演奏のような形で表現されたものも著作物である。(著作権法2条1項1号) また著作権は、演奏権、複製権(コピー)、公衆送信権(インターネットでの配信)など、利用方法ごとに「〇〇権」と権利が定められており、それぞれの権利に対して利用の都度、作者の許諾が必要となっている。(著作権法21条~28条)つまり、動画共有サイト内で動画を配信する際にBGM等で著作物等を使用している場合にはその都度、権利者から許諾を得てから利用しなければならない。

1.1.3 AIによる作品と著作権

著作権法では前項であげたように定義している。しかし現時点の AI では「思想又は感情を」持っているとは認められず、自らの「思想」「感情」に基づいて作品を作っているわけではない。よって AI の作品に著作権は認められず、その音楽は「誰のものでもない」ということになる。これには諸説あるが具体的に法整備が行われておらず、現在ではこれが一般的な考え方である。また、AI の学習の際に用いた解析用データに関しては、電子計算機による情報解析目的のための著作物利用が認められている。(著作権法 47 条 7 項) よって、AI による楽曲制作で有用な結果が得られれば、動画共有サイト等で利用する場合に自由に利用でき、このような社会的問題を解決するために非常に有効であると言える。

1.1.4 AI 楽曲作成サービス

現在、AI を用いた楽曲作成サービスは Jukedek[3] や Amper Music[4] など様々なものが出回っている。例えば Amper Music では「作成したい音楽ジャンル」と「曲の長さ」を指定すれば、AI がオリジナル曲を作曲してくれるという非常に高度なサービスである。作成した曲は、テンポを変えたり、楽器を追加したりと後からも編集できるようにもなっている。このようなサービスでは個人的であれば著作権フリーで利用できるが、商用利用に関してはサービスごとに利用規約で定められており、利用できるものとそうでないものがある。

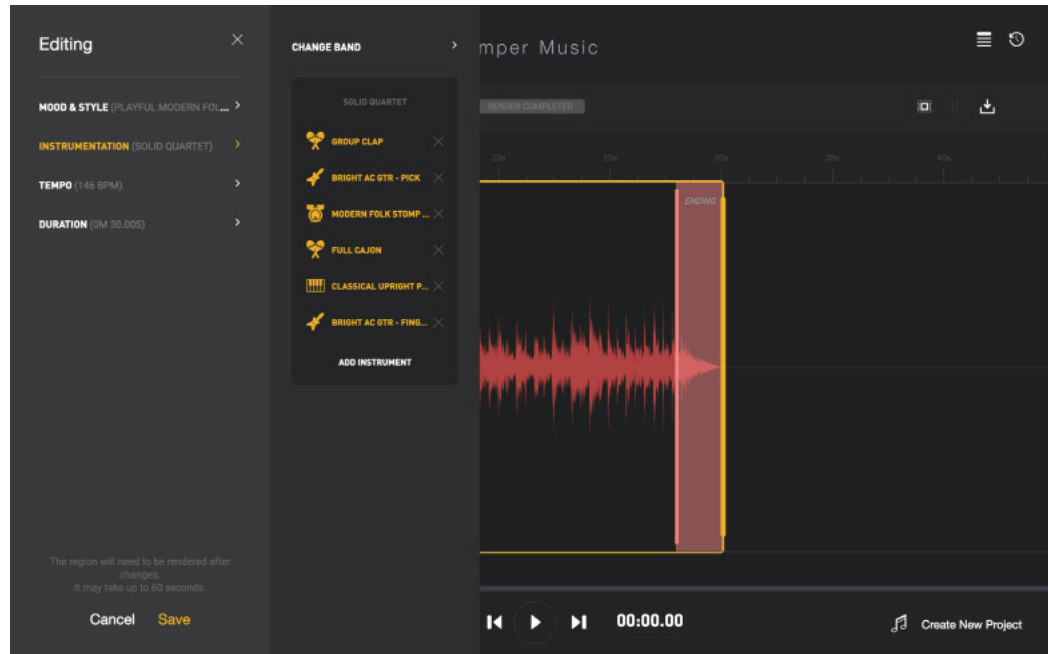


図 1.1: Amper Musick

1.1.5 研究の目的

こうした背景をもとに，本研究では AI による楽曲生成について Google brain によって公開されている Tensorflow のライブラリである Magenta を用いて実証実験を行う．これによって AI による楽曲制作は有用なものなのか，どのような条件で楽曲を生成すれば良い結果を得られるかを調査することが目的である．

実際に人間の手で楽曲を制作する際は音楽理論をもとにコードやスケールに注意して制作をする．また，どの音を使えば気持ちが良いかなど人間の感情的な部分も制作を左右する．しかし AI に感情はなく，気持ちが良い，悪いなどの判断をすることは不可能である．そこで Magenta を用いて AI の楽曲制作ではどこまでの楽曲を制作できるのか，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，これ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の理論

2.1.1 AIの始まり

AIの概念の始まりは,1950年にアラン・チューリングが提唱したチューリングテストである.[5] チューリングテストとは, 男性, 女性, 質問者の3人でおこなわれ, 目的は質問者がどちらが男性でどちらが女性かを当てるゲームである. 会話は文字だけでおこない, 男性は質問者を間違わせるように振舞い, 逆に女性は質問者を助けるように振舞う. 次に, このゲームを男性の代わりに機械がおこない, 質問者が二人の人間を相手にした時と同じくらいの間違いを人間と機械のペアにも起こすことができれば, 機械は知性を持っているとするものである. また, チューリングテストを受ける機械はどんな技術を使っても良く, 機械を作った人がその機械の動作の仕様をきちんと説明できないようなものであってもかまわないということを認めている. この概念が提唱されてから初めてチューリングテストに合格したのは,2014年にレディング大学で開催された「Turing Test 2014」で発表された, ウクライナ在住の13歳の少年が開発した「Eugene Goostman」というプログラムだった.

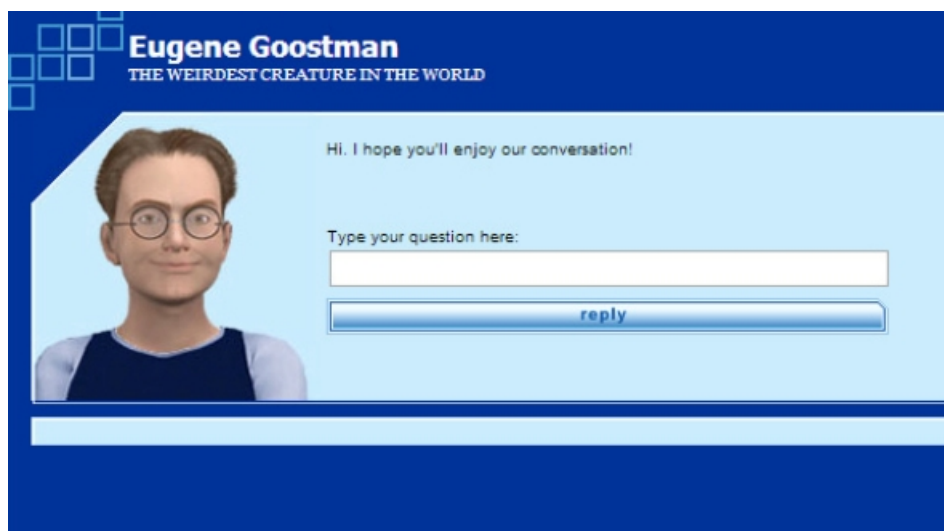


図 2.1: チューリングテストを合格した EugeneGoostman
(引用 : <http://www.itmedia.co.jp/news/articles/1807/26/news014.html>)

2.1.2 第1次 AI ブーム

その後、次の AI ブームは 1950 年代後半から 1960 年代に起きた。この時代になり初めて AI (Artificial Intelligence, 人工知能) という言葉がダートマス会議で用いられた。この会議は 1956 年に米ニューハンプシャー州のダートマス大学で開催され、コンピュータ研究者たちの研究成果を発表し合う研究発表会である。[11] この会議の発起人であるジョン・マッカーシー氏が AI という言葉を The Dartmouth Summer Research Project on Artificial Intelligence[6] の中で用いた。また、この会議で初めての人工知能プログラムと言われる “Logic Theorist” と呼ばれる数学原理をコンピュータで証明するデモンストレーションが行われた。当時のコンピュータはせいぜい四則演算が限界だったので、これは画期的な成果といえた。

他にも、マサチューセッツ工科大学のジョセフ・ワイゼンバウム氏が 1966 年に作成した単純な自然言語処理プログラム「ELIZA」が発表された。ただ、ELIZA はそのプログラムはパターン照合を適用しているので、パターンにない会話や曖昧な事柄に対応できない。

また、人工知能の理解は文字列だけにしか及ばず、画像の特徴を自己で判断し抽出することができないシンボルグラウンディング問題が指摘される。例えば鳩とツノドリの画像を見せ分類するとする時、人間はツノドリのクチバシや足の色などの特徴から分類する。人工知能はこの時、ツノメドリの特徴として何に着目したらいいのか分からないということが起きるため分類に失敗するという問題がおきた。

2.1.3 第2次 AI ブーム

1980 年代に入り、家庭にコンピュータが普及したことにより第二次ブームが発生した。「知識」(コンピュータが推論するために必要な様々な情報を、コンピュータが認識できる形で記述したもの) を与えることで人工知能が実用可能な水準に達し、多数のエキスパートシステムとよばれる、専門分野の知識を取り込んだ上で推論することで、その分野の専門家のように振る舞うプログラムが生み出された。日本では、政府による「第五世代コンピュータ」と名付けられた大型プロジェクトが推進された。しかし、当時はコンピュータが必要な情報を自ら収集して蓄積することはできなかったため、必要となる全ての情報について、人がコンピュータにとって理解可能なように内容を記述する必要がある、世にある膨大な情報全てを、コンピュータが理解できるように記述して用意することは困難なため、実際に活用可能な知識量は特定の領域の情報などに限定する必要があった。こうした限界から、1995 年頃からブームは衰えた。

2.1.4 第3次 AI ブームの始まり

第二次 AI ブームでのエキスパートシステムが壁にぶつかった問題として、日常世界には例外処理や矛盾したルールが非常に多く、知識を教え込む作業が非常に困難というのがあった。これらを解決する手段として「機械学習」や「ディープラーニング」にてコンピュータが自らが学んでいくという手法が第二次 AI ブームの時代から研究されていたが、実用化するためにはコンピュータの性能が追い付いていなかった。

しかし、2000 年代に入り、コンピュータの小型化・性能向上に加えインターネットの普及、クラウドでの膨大なデータ管理が容易となったことで実現可能なレベルとなり、2006 年にはニューラルネットワークの代表的な研究者であるジェフリー・ヒントンの研究チームが、制限ボルトマンマシンによるオートエンコーダの深層化に成功し、再び注目を集めた。この際に発表した論文から、これまでの多層ニューラルネットよりもさらに深いネットワーク構造を意味するディープネットワークの用語が定着し、第三次 AI ブームが沸き起こった。他にも、2011 年に IBM のワトソンが難解な質問と独特の解答方法で知られる人気クイズ番組「ジョパディ！」に出演した。[13] 出演したワトソンが読み込んだ本や映画の脚本、百科事典などは合計 100 万冊にものぼり、公平を期すため、インターネットには接続しておらず、読み込んだデータのみでの勝負となり、歴代チャンピオン 2 人に勝利した。



図 2.2: クイズ番組に出演する Watson

(<http://www.swift-web.org/cp-bin/blogn/index.php?e=1023>)

2012年にはGoogleが発表した機械学習の論文では、事前に猫をネットワークに教えたわけでもなく、猫のラベル付けした画像を与えたわけではない人工知能を発表した.[7] このときまで、機械学習のほとんどは、ラベル付きデータの量に依存していた。この論文により、機械がラベルのない生データでも処理することができ、そしておそらく、人間が予備知識を持たないデータですら処理できることが示された。

2.1.5 第3次 AI ブームの現在

2014年にはAmazon.comからスマートスピーカーと呼ばれる対話型の音声操作に対応したAIアシスタント機能を持つスピーカーであるAmazon Echoが発売され、その後GoogleやAmazon,IBMによって様々なスマートスピーカーが商品化された。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。



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より高いが、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、以降、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.[9] また、芸術の分野

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）で落札されたり、AI を用いて新しい楽曲を作るものが出回っていたりと、成長が著しい.[10]

2.2 AI を用いた楽曲作成

作曲の流れはその構成によって階層化されており、比較的自動化が容易とされている。音楽を構成する 3 要素はスケール（調）とコード（和音）とメロディ（旋律）とされており、スケールが決まればその構成音に合わせてコードも決まる。コード進行は一定のルールがあり、これまでの曲の中で良い進行とされるパターンは数多く蓄積されている。コードは曲のムードを大きく左右し、人間のその時の感情、感じ方に非常に影響を与えていると言われている。メロディも同様にコードの構成音を元にすれば大きく外れることはないが、単調になってしまう傾向があるので、多少のランダムさが必要とされている要素である。コードに対して大きく外さない範囲で変動させれば単調になるのを防ぐこともできる。

2.2.1 MIDI

音楽データには大きく分けて二つある。一つ目はオーディオデータといい波形の情報を記録する形式であり、二つ目が MIDI である。

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。

なお、インプットデータは one-hot Vector で図 2.4 のようになっている。また、楽曲制作の際に音程を指定する場合は図 2.5 の数値を指定する。

❖ インプットデータ・・・One-hot Vector

0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

何消ドレミフソラシドレミフソラシドレ
も音 ア ア
し
ない

例) 低いシの音を鳴らす

図 2.4: インプットデータの仕組み

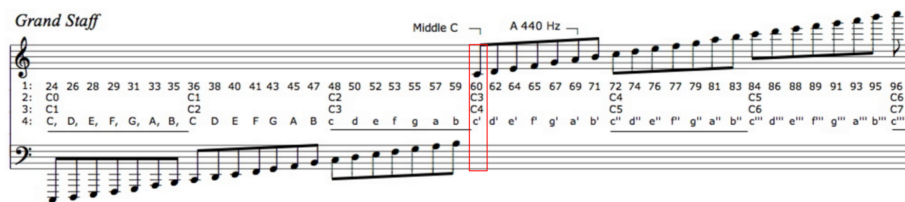


図 2.5: MIDI と音階

2.2.2 Magenta

本研究で使用する Magenta 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 上に公開されている OSS である. これを図 2.6 に示す.

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 2.7 に示す.

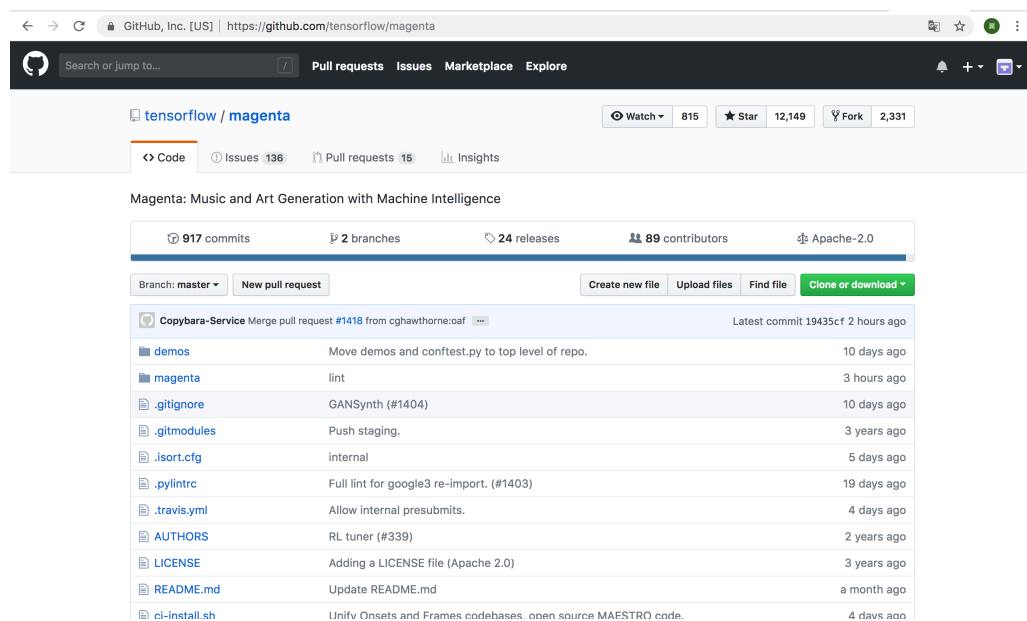


図 2.6: GitHub 上に公開されている Magenta

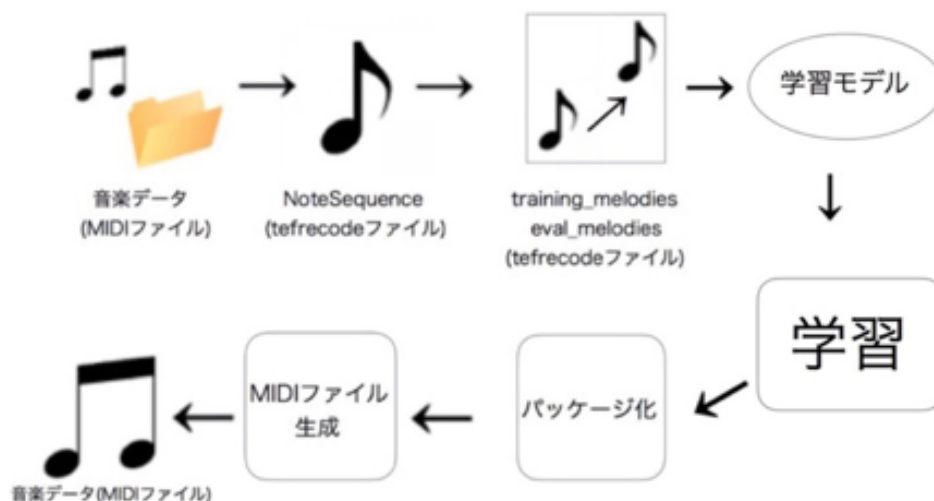


図 2.7: magenta による MIDI 音楽データ生成までのプロセス

2.2.3 リカレントニューラルネットワーク (RNN)

本研究で用いる Magenta のモデルは RNN(Recurrent Neural Network) を取り入れたものである。RNN は、ある層の出力がもう一度その層へ入力される回帰結合を持つニューラルネットワークである。このような結合を持つことで、ニューラルネットは過去の情報を保持することができるようになる。

2.3 機械学習に適した開発環境について

Tensorflow のランタイムとして以下のシステムがサポートされている.

- (1) Ubuntu 16.04 以降
- (2) macOS 10.12.6(Sierra) 以降 (GPU サポートなし)
- (3) Windows7 以降
- (4) Raspbian 9.0 以降

また,GPU を用いて学習を行う時には tensorflow-gpu というパッケージが必要となり,導入には以下のドライバやライブラリが必要である.

- (1) CUDA Toolkit (tensorflow は CUDA9.0 をサポート)
- (2) CUPTI (CUDA Toolkit に付随)
- (3) NVIDEAGPU ドライバ (CUDA9.0 には 384.x 以上が必要)
- (4) cuDNN SDK

Windows10 の環境ではリリースされている CUDA のバージョンは 10 のみであり,Tensorflow のサポートを外れてしまう. そのため,CUDA v9 がインストール可能な Ubuntu を用いることとし, システムの開発環境を表 2.1 に示す.

表 2.1: 開発環境

OS	Ubuntu 16.04 LTS
CPU	Intel Core i3 8100
メモリ	8GB
GPU	GeForce GTX 1060
CUDA	CUDA(9),cuDNN(7.4.2)
ライブラリ	TensorFlow(1.12.0),magenta(0.5.0)

2.3.1 CUDA

CUDA は,NVIDIA が開発している GPU 上でプログラミングをするためのソフトウェアプラットフォームである. 含まれるものとしては,CUDA を実行形式に変えるコンパイラや,それをサポートする SDK, ライブラリ, デバッグツール群である. CUDA を導入することによって, プログラムを複数のプロセッサで動かすだけでなく,無駄なく並列化することができる.

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。

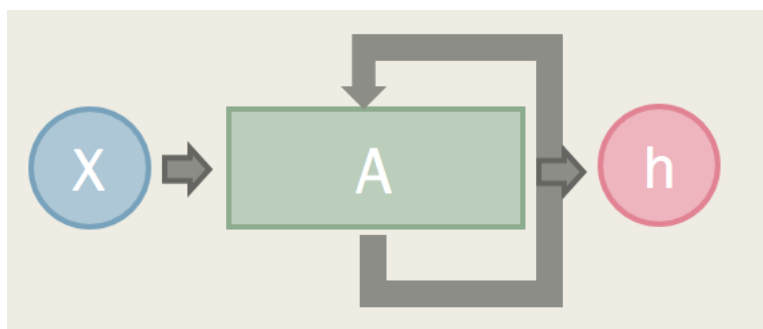


図 3.1: basic_rnn のモデル (1)

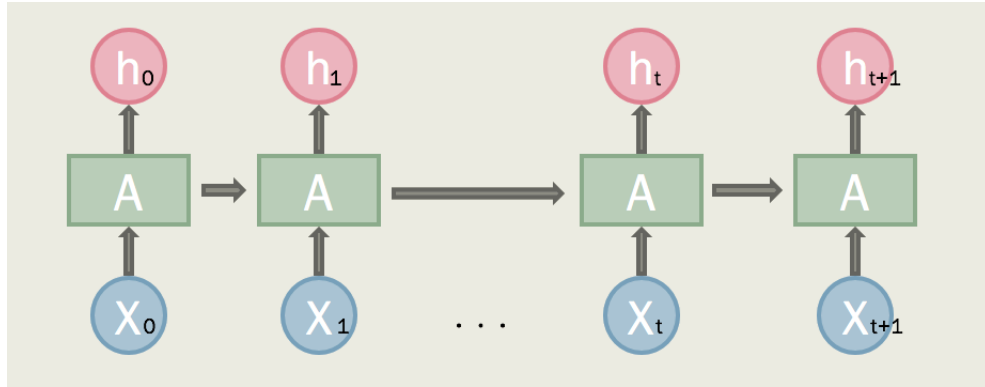


図 3.2: basic_rnn のモデル (2)

basic_rnn

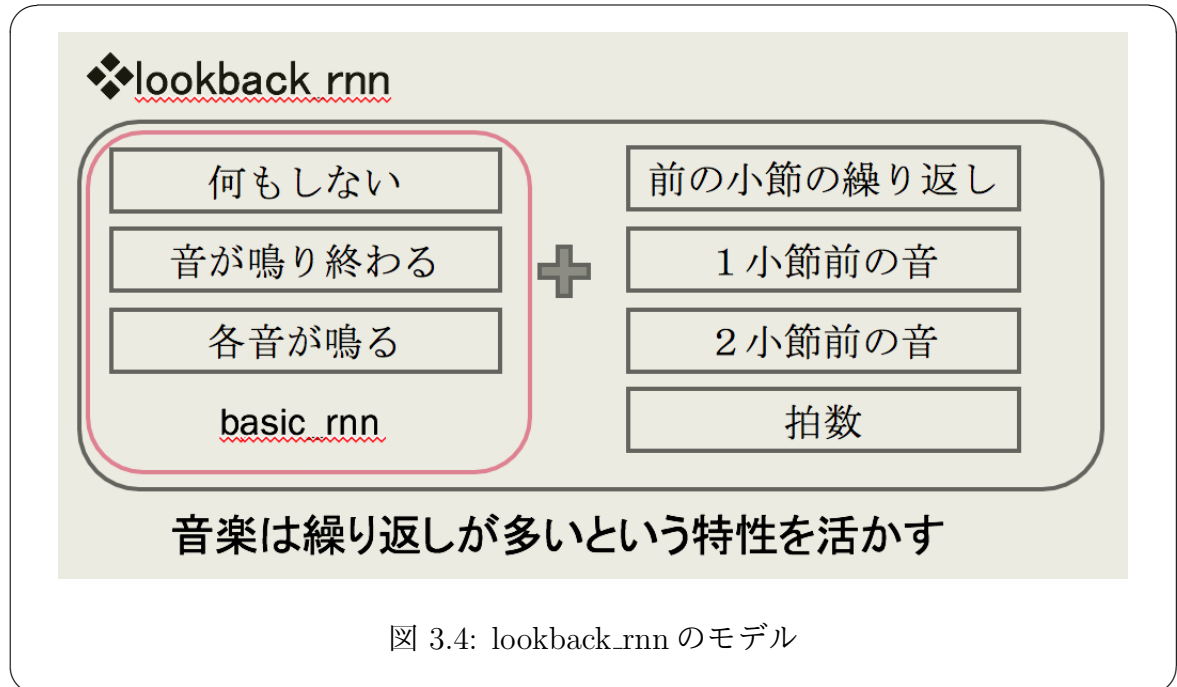


- Recurrent Neural Network というもの
- basic_rnnを拡張したものがlookback_rnnとattention_rnn
- どちらが優位ということはない
- 入出力データはOne-hot vector

図 3.3: 三つのモデルについて

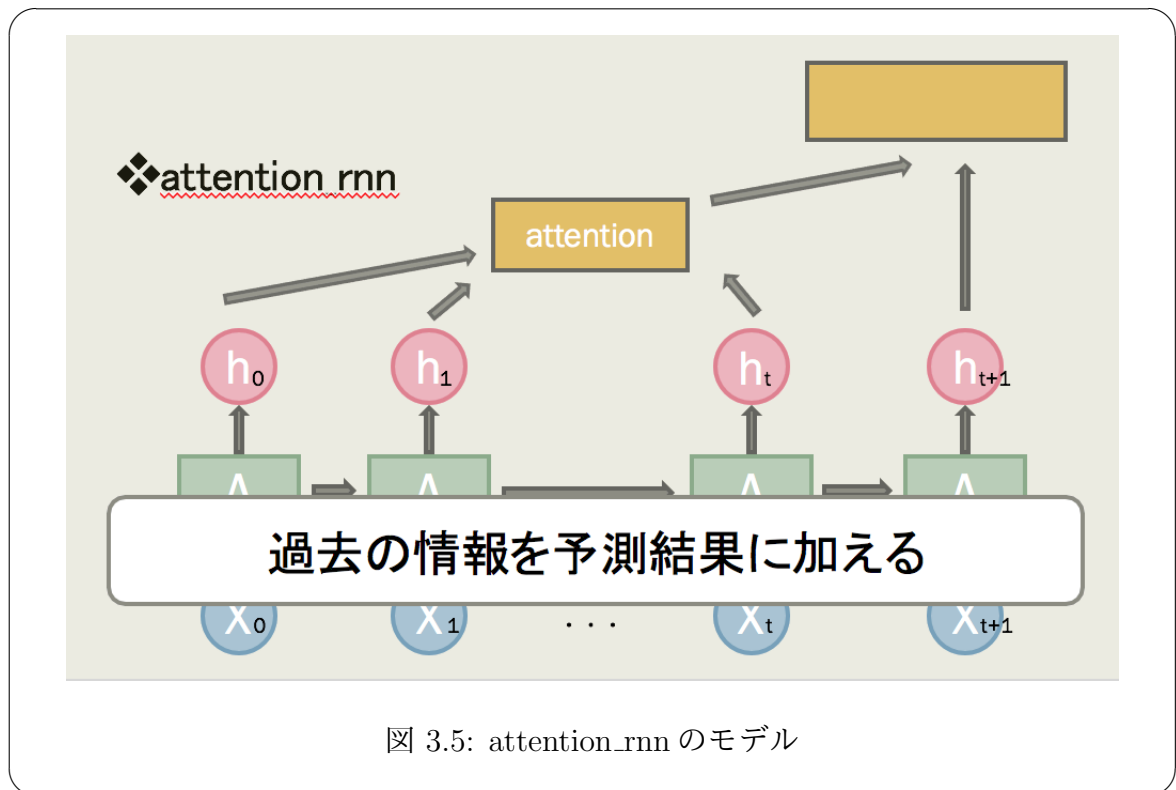
(2) lookback_rnn

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である．これは音楽は繰り返しが多いという特性からなるものであり，これを活かして楽曲生成を行うものである．



(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である．図 3.5 のような流れになっている．



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．

以上のモデルを用いて楽曲制作を行い，それぞれの違いと有用性について検証する．

3.2 学習回数による違い

学習回数を変更して楽曲制作を行う。本研究では500回, 5000回, 20000回でMelodyRNNのbasic_rnn,lookback_rnn,attention_rnnでそれぞれ楽曲を制作した。これらの結果, 生成された楽曲の音程やリズムへの影響やそれぞれの違いを検証し, 有用性について検討する。

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う。本研究では32,64,128でbasic_rnnそれぞれ楽曲を制作した。これらの結果, 生成された楽曲の音程やリズムへの影響やそれぞれの違いを検証し, 有用性について検討する。

第4章 楽曲制作

4.1 Melody_rnn を使用して学習モデルを作成

4.1.1 NoteSequence の作成

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。

表 4.1: JSON とプロトコルバッファの比較

	プロトコルバッファ	JSON
フォーマット	バイナリベース	テキストベース
データ量	少ない	多い
	GPU	GeForce GTX 1060
	CUDA	CUDA(9),cuDNN(7.4.2)
	ライブラリ	TensorFlow(1.12.0),magenta(0.5.0)

NoteSequence の作成は以下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
```

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以下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
```

4.1.2 basic_rnn を用いて学習を開始

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

`-config` で学習に使用する `basic_rnn` を指定、`-rundir` で学習のために用意した NoteSequence を指定し、`-sequence_example_file` 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_training_steps` で学習回数を設定する。

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

4.1.3 lookback_rnn を用いて学習を開始

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

`-config` で学習に使用する `lookback_rnn` 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=lookback_rnn \  
--run_dir=/tmp/melody_rnn/logdir/run1 \  
--sequence_example_file=/tmp/melody_rnn/training_melodies.tfrecord \  
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \  
--num_training_steps=20000
```

4.1.4 attention_rnn を用いて学習を開始

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

`-config` で学習に使用する `attention_rnn` 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/training_melodies.tfrecord \  
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \  
--num_training_steps=20000
```

4.1.5 音楽データの作成

以下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 `attention_rnn` 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]"
```

4.1.6 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. その後以下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]"
```

4.2 Polyphony_rnn を使用して学習モデルを作成

PolyphonyRNN を使用してする手順としては、大まかな流れは同じであるが`-config` で使用する RNN を指定する必要があるという違いがある

4.2.1 NoteSequence の作成

NoteSequence の作成は以下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
```

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以下に示すのコマンドを実行する。

`-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。

```
polyphony_rnn_create_dataset \
--input=/tmp/notesequences.tfrecord \
--output_dir=/tmp/polyphony_rnn/sequence_examples \
--eval_ratio=0.10
```

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以下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。

```
polyphony_rnn_train \
--run_dir=/tmp/polyphony_rnn/logdir/run1 \
--sequence_example_file=/tmp/polyphony_rnn/training_tracks.tfrecord \
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \
--num_training_steps=20000
```

4.2.3 音楽データの作成

以下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.(図 2.5 参照)

```
polyphony_rnn_generate \  
--run_dir=/tmp/polyphony_rnn/logdir/run1 \  
--hparams="batch_size=64,rnn_layer_sizes=[64,64]" \  
--output_dir=/tmp/polyphony_rnn/generated \  
--num_outputs=10 \  
--num_steps=128 \  
--primer_pitches="[67,64,60]" \  
--condition_on_primer=true \  
--inject_primer_during_generation=false
```

第5章 結論

5.1 AIによる楽曲制作の結果

5.1.1 モデルによる生成結果の違い

学習モデルごとの生成結果は単音の単純な羅列が続く `basic_rnn` に比べ `lookback_rnn` や `attention_rnn` の方がリズム的な繰り返しが多いという特徴が見られた。また，`lookback_rnn` に比べ `attention_rnn` の方がの方が単純な繰り返しではなく，展開的な繰り返しをした。図 5.1 はその生成結果である。(GarageBand 上のスクリーンショット)

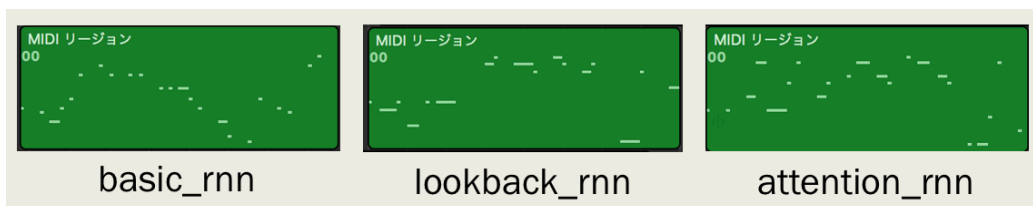


図 5.1: 学習モデルごとの生成結果

5.1.2 学習回数による生成結果の違い

学習回数ごとの生成結果では、学習回数が増えるごとに学習させた楽曲に近づいた。また学習回数が少ないとコードのスケール外の音も入ってしまっていたが、学習回数を上げるとその現象の発生があまり起きなかった。図 5.2, 図 5.3, 図 5.4 はその生成結果である。(GarageBand 上のスクリーンショット)



図 5.2: basic_rnn による学習回数ごとの生成結果



図 5.3: lookback_rnn による学習回数ごとの生成結果



図 5.4: attention_rnn による学習回数ごとの生成結果

5.1.3 ノード数による生成結果の違い

ノード数による生成結果ではノード数32と64に比べ128の方が長い音が増えた。また音階にまとまりが出てあまり高音や低音に移動せず一定の音域の楽曲になった。図5.5はその生成結果である。(GarageBand 上のスクリーンショット)



5.2 調査結果

楽曲制作に関して有用な条件をアンケート形式で調査した。調査方法は学習回数が500回、5000回、20000回のそれぞれで生成した `basic_rnn`、`lookback_rnn`、`attention_rnn` の生成結果、`basic_rnn`、`lookback_rnn`、`attention_rnn` それぞれで生成したの学習回数500回、5000回、20000回での生成結果、`basic_rnn` で生成したノード数32、64、128での生成結果をそれぞれ視聴してもらい、それぞれの項目から一番良いと思うものを選択してもらう形式をとった。

調査の様子を図 5.6 に示す。

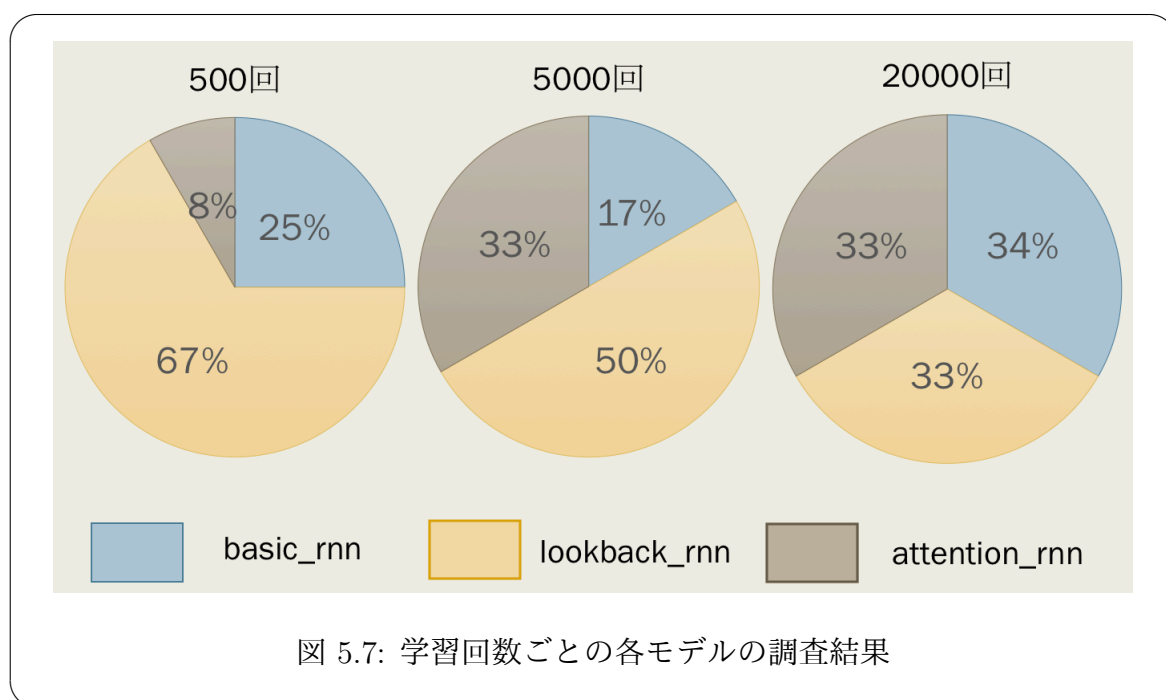


図 5.6: 調査の様子

5.2.1 学習回数ごとの各モデルの調査結果

学習回数ごとの各モデルの調査では学習回数が500回、5000回で `lookback_rnn` が一番良いと答えた人の割合が半数を超えた。これは `lookback_rnn` のモデルが繰り返しに対応していることで人間が聴きやすい楽曲になったことが考えられる。一方で学習回数20000回では全てのモデルの良いと答えた人の割合が同数だった。これは学習回数を多くすることでスケール外の音が入るなどの現象がなくなり、生成される楽曲が似てくるためであると考えられる。

以下の図 5.7 に学習回数ごと各モデルの調査結果を示す。



5.2.2 各モデルごとの学習回数の調査結果

basic_rnn では学習回数が多い方が良いと答える人の割合が増えた．一方で attention_rnn では学習回数が 500 回のものが一番良いと答えた人の割合が増えた．

以下の図 5.8 に各モデルごとの学習回数の調査結果を示す．

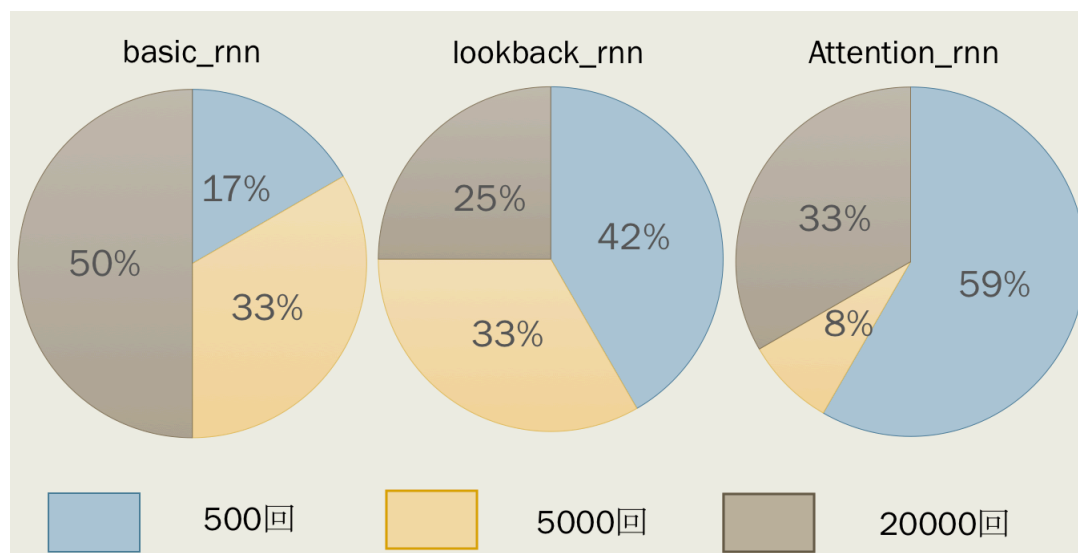
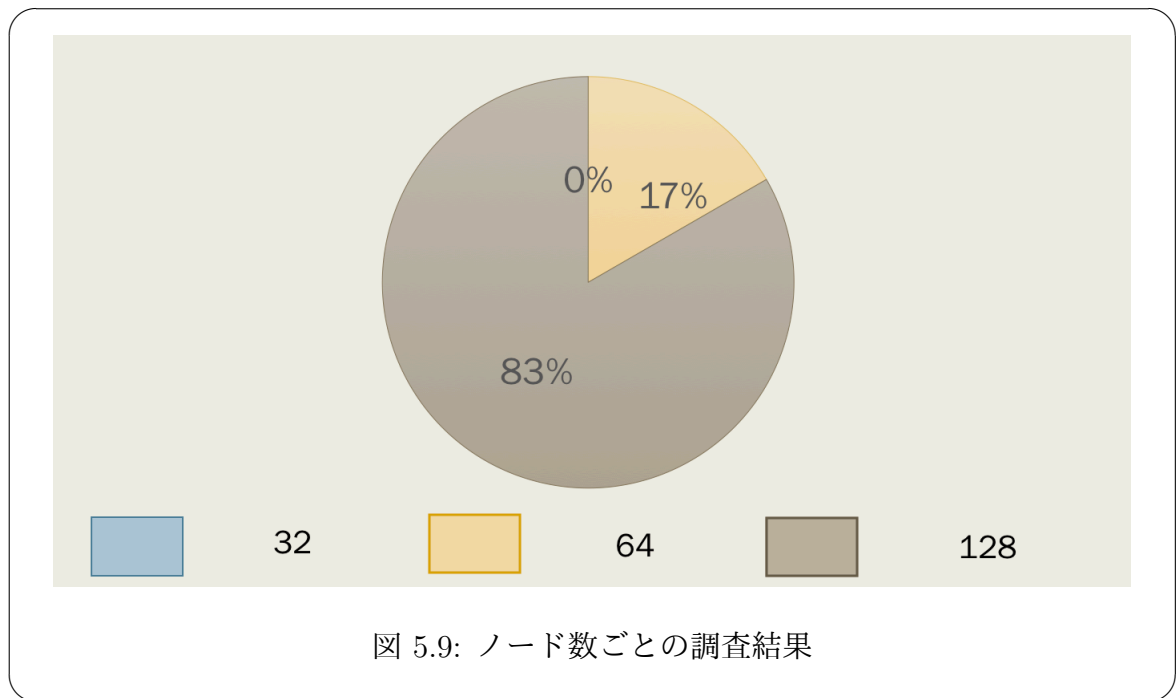


図 5.8: 各モデルごとの学習回数の調査結果

5.2.3 ノード数ごとの調査結果

ノード数ごとの調査結果ではノード数が128で生成したものが良いと答えた人の割合が一番多かった。これはノード数が多い方が長い音が増えたことや音階にまとまりが出たことで視聴者が聴きやすいと感じやすかったからだと考えられる。

以下の図 5.9 にノード数の調査結果を示す。



5.3 今後の課題

今回、様々な条件で楽曲を制作した。特に学習回数を大きくすると楽曲の精度は高まった。しかし音階やスケールは理解できているようだが、メロディに関してはあまり良いものとは言えなかった。要因としては学習に用いる楽曲が少なかったことや学習回数がまだ足りなかったことがあげられる。

また、本実験の調査の結果から AI における楽曲制作において、必ずしも AI の学習回数を上げれば視聴者に有用な楽曲が生成されるわけではないことがわかった。これは人によって良いと感じる音やメロディ、リズムには差があり、AI による正確性があれば良いと一概には言えないためだと思われる。これは学習させる楽曲が一番影響を与えており、これを変えることで実験結果は変わると思われる。

謝辞

本研究に関しまして，熱心かつ丁寧にご指導いただきました，千葉工業大学情報科学研究科情報科学専攻中村直人教授に心から御礼申し上げます。また，中村研究室の皆様，学部生の皆様に心から御礼申し上げます。皆様のおかげで，これまでの研究生活を充実かつ楽しく送ることができました。そして，これまで支えてくれた両親をはじめとする親族各位に改めて御礼申し上げます。ありがとうございました。

参考文献

- [1] YouTube” YouTube”,
<https://www.youtube.com>
- [2] ドワンゴ” ニコニコ動画”,
<https://www.nicovideo.jp>
- [3] Jukedek” Jukedek”,
<https://www.jukedek.com/>
- [4] Amper Music” Amper music”,
<https://www.ampermusic.com/>
- [5] A. M. Turing (1950)” COMPUTING MACHINERY AND INTELLIGENCE”,
<https://www.csee.umbc.edu/courses/471/papers/turing.pdf>
- [6] J. McCarthy,Dartmouth College M. L. Minsky,Harvard University N.
Rochester,I.B.M. Corporation C.E. Shannon,Bell Telephone Laboratories(1955),
”A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE”
<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- [7] Quoc Le Marc’Aurelio Ranzato Rajat Monga Matthieu Devin Kai Chen Greg
Corrado Jeff Dean Andrew Ng(2012) ”Building High-level Features Using Large
Scale Unsupervised Learning”
<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/38115.pdf>
- [8] Google Brain チーム”Magenta”,
<https://github.com/tensorflow/magenta>
- [9] 財経新聞”人工知能を使って執筆した小説が星新一賞一次選考を通過”,
<https://www.zaikai.co.jp/article/20160323/299468.html>
- [10] YAHOO ニュース”AI 絵画、大手オークションで初の落札 予想額の 40 倍超”,
<https://headlines.yahoo.co.jp/hl?a=20181026-00010002-afpbnews-int>

- [11] (社) 人工知能学会 ” 人工知能の話題 ” ,
<https://www.ai-gakkai.or.jp/whatsai/AItopics5.html>
- [12] マイナビニュース”XML はもう不要!? Google 製シリアルライズツール「Protocol Buffer」 ” ,
<https://news.mynavi.jp/article/20080718-protocolbuffer/>
- [13] 日本経済新聞 (2011)”人間にクイズで勝ったコンピューター「ワトソン」の素顔”,
https://www.nikkei.com/article/DGXNASDD2305K_T20C11A3000000/
- [14] Google”GPU support” <https://www.tensorflow.org/install/gpu>
- [15] Google”Install TensorFlow” <https://www.tensorflow.org/install/>