メロディを対象とした生成Deep Learningモデルの比較

駒澤大学　平井辰典

本日の発表内容

- 新しい提案はありません
- 音楽生成に関連した生成ディープラーニング技術に注目し，それらを比較して，問題提起
- 特筆すべき結論はありません（研究の経過報告的な発表）

※ 当初実施予定だった評価実験を実施しないことに

- 今日は論文には書かないような内容をメインに発表します（一部主観的な内容もあります）

研究背景

- 生成ディープラーニング技術の急速な進化


GAN，2014


DCGAN，2015


COGAN，2016


ProGAN，2017


VQ-VAE，2019

⇒　年々，生成品質が向上している！

研究背景

- 生成ディープラーニング技術の音楽生成への応用
  - Jukebox [Dhariwal+, OpenAI, 2020]：VQ-VAE [Demo]
  - MuseNet [Payne+, OpenAI, 2019]：Sparse Transformer [Demo]
  - MusicTransformer [Huang+, ICLR2018]：Transformer [Demo], [Demo]
  - MuseGAN [Dong+, AAAI2018]：GAN [Demo]
  - MusicVAE [Roberts+, ICML2018]：VAE + RNN [Demo]
  - DeepBach [Hadjeres+, ICML2017]：LSTM [Demo]

問題：
この「？」が取れない

⇒　年々，良い音楽(?)が生成されつつある(?)

音楽生成研究を行う上での問題点

- SOTAが明確でない
  ⇒ これまでDeep Learning技術により更新されてきた各種識別タスクのような明確な評価基準が音楽生成には存在しない

そのため

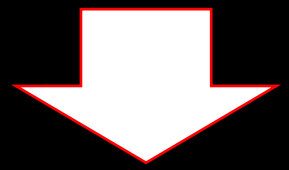
MuseNetを改善するアプローチがいいのか,
MusicVAEを改善するアプローチがいいのか,
新規研究のスタート地点がわからない

現在の音楽生成Deep Learning研究の流れ

- 他分野で成功を収めた手法を音楽生成へと応用するアプローチ

  ⇒ VAE，RNN，GAN，Transformer，VQ-VAE等

※ もちろんその応用が大変な研究課題であったりもする



- モデルの学習という観点では進歩していることは明白
- 「良い音楽」の生成という面で進歩しているのかは不明

※ おそらく打率は上がってきている

- 既存手法に新たに一層追加すれば新規手法になるのか？ すら不明

そこで本研究では,

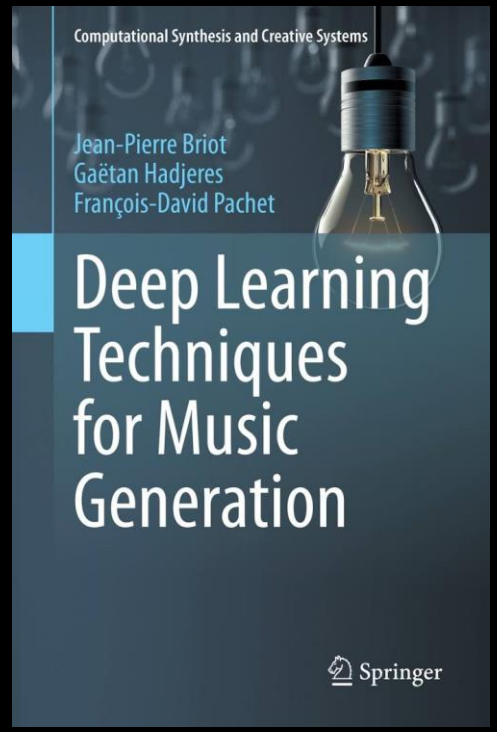
基本に立ち返り,

各種生成Deep Learningモデルを比べてみました

先行研究

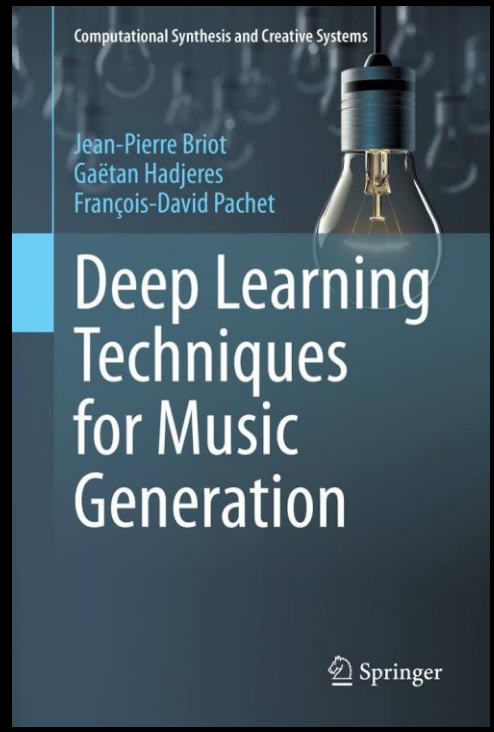
- Deep Learning Techniques for Music Generation (2020, Springer) / Jean-Pierre Briot, Gaëtan Hadjeres, François-David Pachet

- Deep Learningベースの生成モデルのみに注目した全284ページの詳細なサーベイ
  ⇒ 近年の重要な技術の概要は粗方載っている



先行研究

- Deep Learning Techniques for Music Generation (2020, Springer) / Jean-Pierre Briot, Gaëtan Hadjeres, François-David Pachet



- 各種手法を以下の５つの観点からまとめている
  - Objective：生成対象
  - Representation：データ表現
  - Architecture：ネットワークの構造
  - Challenge：どういった問題に取り組むか
  - Strategy：生成の方法

⇒ 手法の比較や評価法についてはほぼ触れていない

精度等の

本研究の目的

- 音楽のメロディに注目して生成Deep Learningモデルを比較

比較の方針：Objective

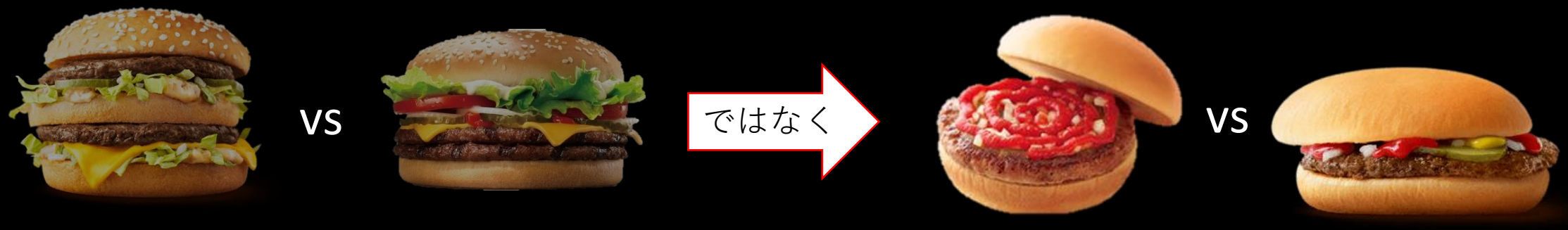
- 対象（Objective）をメロディ（単旋律）に限定

- なぜメロディか？
  ⇒ 生成対象（Objective）が違えば比較は不可能だから
  （例：リハーモナイズとメロディ生成は比較できない）

比較対象

- 生成Deep Learningモデルの基本構成要素に限定
  i.e. DNN (FFN), VAE, LSTM, Attention, GAN

- なぜ基本構成要素に限定するか？
  ⇒ MuseNetとMusicTransformerを比較しても，どの要素が効いて結果が変わったのかを判断することが困難なため



比較項目

- メロディを生成する上での特徴
  ⇒ 入力形式の特徴，データ表現の制約等

- 学習の難しさ（困難さ）
  ⇒ 実装後のハイパーパラメータチューニング等の試行錯誤がどの程度必要となりそうか？

- おまけ程度の項目
  - 処理時間
  - 生成結果（メロディそのもの）

比較条件

- 手法：ランダム生成，DNN，VAE，Seq2seq（LSTM，Attention），GAN
- データセットを固定：
  Webから収集されたMIDIデータから約1万曲分のメロディを抽出し学習
- 各手法の最小構成で，なるべくスタンダードな実装
  （チュートリアルや参考書籍に掲載されてものを参考に実装）
- ハイパーパラメータのチューニング等なし：
  ロスに注目して学習が正常に進んでいるかのみを確認
- 生成の際に入力メロディを受け付けられるモデルの場合は，
  童謡「キラキラ星」のメロディを入力

- その他の詳細については予稿をご参照ください

比較結果：モデルの特徴

- クイックな比較結果のまとめ
  - VAE，GANは，（そのままでは）入力メロディに基づいて新たなメロディを生成するということはできない
  - GANは学習がなかなかうまくいかない
    学習が進んでも，生成結果が固定されてしまったりする
  - DNN，Seq2seq（LSTM），Seq2seq（LSTM+Attention）では，
    最初の数エポックでロスが下がった後にはほとんど学習が進まなくなる

- 時間の都合上，詳細については予稿をご覧ください！

比較結果（おまけ）：学習に要した時間

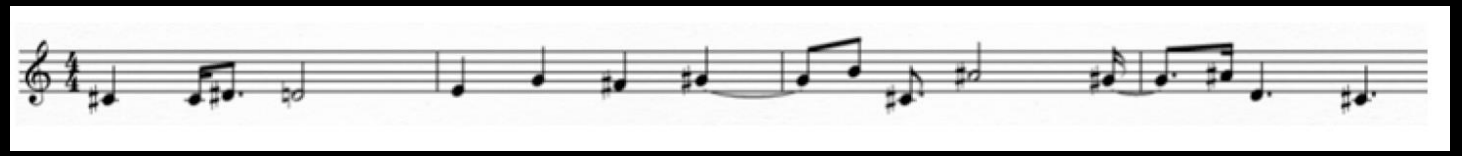
- VAE < GAN < DNN < Seq2seq (LSTM) < Seq2seq (LSTM+Attention)

DNNを1.0としたときの1エポックあたりの学習に要した時間の比率

モデル	DNN (FFN)	VAE	Seq2seq (LSTM)	Seq2seq (+Attention)	GAN
処理時間	1.0	0.0095	2.1	5.8	0.072

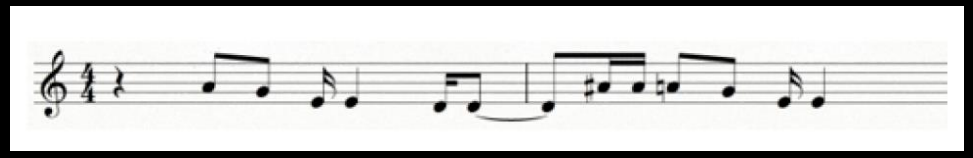
比較結果（おまけ）：生成されたメロディ

ランダム生成

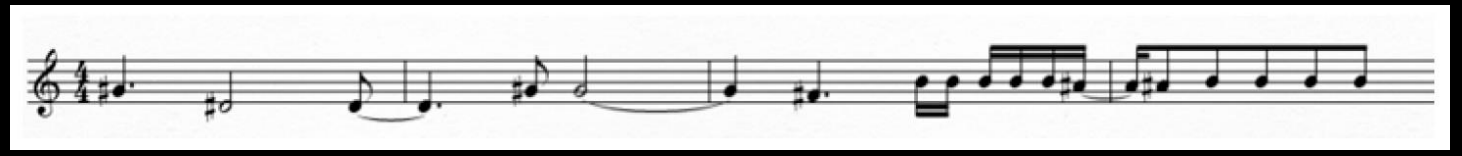


DNN

VAE

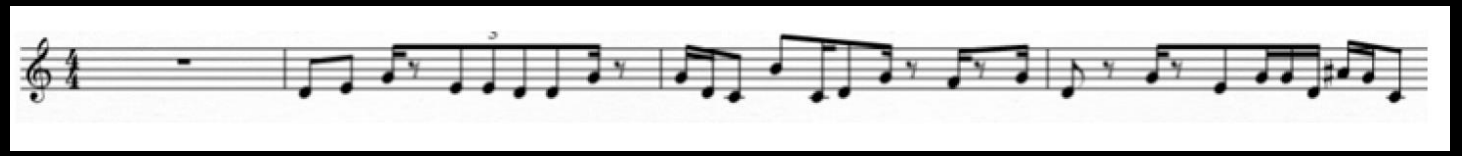


Seq2seq（LSTM）



Seq2seq（+Attention）

GAN



主観評価実験による評価について

- 当初は主観評価実験により生成結果を比較する計画……だった
  ⇒ 様々な事情を勘案して方針転換

主観評価実験の問題点について

- 再現性がない（……信頼できる評価手順を踏んでいれば大丈夫）
- 被験者に聞かせるサンプルによって結果が変わる
- 被験者の音楽経験や嗜好によって結果が変わる
  - DeepBach [Hadjeres+, 2017]において，システムによる生成結果がバッハ本人による曲かどうかを区別するような主観評価実験を1,272人に対して実施
    ⇒ バッハ本人による曲と判定された割合が多かった
    が，音楽経験が豊富な被験者の多くは区別がついていた

⇒ 限られた被験者に対する限られたサンプルに基づく主観評価実験結果を基に結論を出すことは難しい
⇒ （今回は）評価実験の実施は断念

モデルの公平な評価のために必要なこと

- メロディ生成タスクの定義，評価基準が必要
- 既存データセットに対する予測精度，Perplexityなどは，
  音楽の良さとは切り離されるべき指標
  （タスク次第で大きく変わる上に音楽の良さとの関係は不明）

⇒ 今後，SOTAを明確にできるような評価の枠組みの制定が必要

※ MIREXでは過去に結果を集めてリスニングテストで比較していたことも

- 音楽の良し悪しの評価は音楽情報処理研究のGrand Challenge

その他に考えられる今後の研究の方向性

- 良い音楽の定義は人によって異なるため，万人にとって良い音楽を生成するという方針自体に無理がある
- 個人の嗜好に基づいて欲しい音楽を生成できるようなパーソナライズが容易なモデルが望まれる
  - MidiMe：Dinculescu+, 2019
- 生成モデルを用いた音楽生成支援インタフェース
  - Melody Composition with Human-in-the-Loop Optimization：Zhou+, 2020
  - Magenta Studio：Roberts+, 2018
  - Flow Machine：Pachet+, 2012～
- いかに利用して，いかに創作/鑑賞を支援するかに注目した技術

音楽生成研究において解決すべき課題

- クリエータが利用可能な音楽生成モデル構築環境の実現

  ⇒ハンバーガーの具のカスタマイズのように誰もがモデル構築

  - Magenta.js：Roberts+. 2018
- 各モデル/モジュールのExplainability向上（脱ブラックボックス化）
- 高速な学習の実現

  ⇒パラメータを調整したらすぐにフィードバックが得られるシステム（e.g., 既存の音楽制作ツールのようなスピード感）

  - オンライン学習
  - fine-tuning
  - and more?

まとめ

- DNN (FFN)，VAE，LSTM，Attention，GAN等の
  生成Deep Learningモデルにおける重要な構成要素を実装し，
  実際にメロディを生成することで各種モデルを比較
- 音楽生成のためのDeep Learning研究の課題と
  今後考えられる応用研究の方向性を提案