

はじめに

音楽情報処理分野

研究課題

まとめ

研究の進捗

武藤 克弥 (Katsuya Mutoh)
u255018@st.pu-toyama.ac.jp

富山県立大学 大学院 電子・情報工学専攻 情報基盤工学部門

July 27, 2022

進捗

- まだ分野研究(音楽情報処理業界)している段階

背景

- 音楽の創作に機械学習が使用される例が、年々増加している
- ルールベースから深層学習ベースへの発展に伴い、音楽の自動生成技術も高まっている
- 音楽は文脈の時系列的変化という点で文章と構造が似ており、自然言語処理と親和性が高い
- 機械学習で生成される音楽の表現の幅には未だ限界がある

本研究の目的

- 【予定】学習時に新たな特徴量を追加し、既存研究に比べて表現力が向上しているか評価する。

自動作曲分野の概要

3/12

音楽情報処理分野の概要

- メインの自動作曲は音楽情報処理分野の一つに位置づけられる
- 他にも音響解析, 感情分析, 視聴時の生理的反応分析など多岐にわたる
- 歌詞, 音符配列によるテキストマイニングもあれば機械学習ベースの研究もある

→今は自動作曲に焦点を当てる

はじめに

音楽情報処理分野

研究課題

まとめ

対象となる学習データ (自動作曲)

- 音楽信号： 波形データをスペクトルに変換した特徴量を使用
→ CD のサンプリングレート (44.1kHz) をそのまま使うとなると非現実的
 - MIDI: コンピュータ上で扱うための電子楽譜, 一番使われる
→ 現在, 約 44 万曲分のデータセット (MMD¹) があるが, ジャンルの偏りが大きい
 - ABC 記法 (テキスト): ABC 記法の楽譜を自然言語処理
→ 基本的に単旋律しか扱われないという問題²
 - コード進行: コードの遷移を文章のようにとらえ自然言語処理
- 今のところ MIDI か ABC ベースのデータを使う

1 Ens, Jeff, and Philippe Pasquier. 2021. "MetaMIDI Dataset", Zendo. <https://zenodo.org/record/5142664#.Yt-joHbP02z>, accessed 26 July 2022.

2 Nao Tokui. 2017. "Deep Learning を用いた音楽生成手法のまとめ [サーベイ]", Medium.

<https://naotokui.medium.com/deep-learning を用いた音楽生成手法のまとめ-サーベイ-1298d29f8101>, accessed 25 July 2022.

自動作曲での機械学習手法

- MLP(多層パーセプトロン)：よく見る深層 NN
RNN と異なり、一音ずつではなく一気に全部作成
- RNN(再帰型 NN)：コンテキストに沿って、一音ずつ時系列的に作成
- GAN(敵対的生成ネットワーク)：生成器と識別機の競争
- Transformer：長い文章の文脈を考慮できるようにした自然言語処理手法
→ RNN よりもさらに長い時系列的依存関係に対応できる

※どの手法を選ぶかは要検討

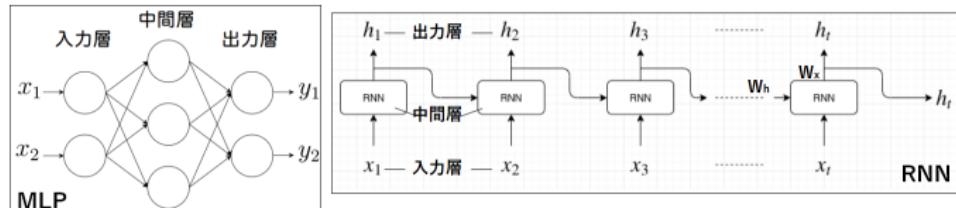


図 1: MLP と RNN

自動作曲の応用先

6/12

作業系 BGM

- 「集中できる曲」「ドライブ風景に合った曲」「天候に応じた曲」といったテーマの音楽生成

コンテンツ融合

- 「Youtube 動画内での BGM」「写真のスライドショーに流す音楽」のための自動作曲

作曲支援

- 専門知識のない人向けの簡単な作曲
- プロの作曲家に向けたアシスト など

※本研究では、応用手前である機械学習手法の部分を扱う（予定）

はじめに

音楽情報処理分野

研究課題

まとめ

学習時間

- 数分の音楽生成に日単位かかる可能性
→リアルタイム利用の難しさ

表現の幅

- データセットがクラシック系に偏っている
- クラシックよりの曲調になりがち
→MIDIにBPM, 音階といった特徴を加える
→データの水増し

本研究では、表現の幅(特徴量考慮)の部分に絞る

はじめに

音楽情報処理分野

研究課題

まとめ

自動作曲の評価

学術的評価が難しい
→妥当な評価方法の検討↓

- 教師データを上手く学習しているか
- ただの模倣になっていないか

評価手法 (例 1)

- 元の音楽との比較

はじめに

音楽情報処理分野

研究課題

まとめ

評価手法 (例 2)³

作成曲とオリジナル曲の類似度 + 人間評価 「似ているかどうか」

- 曲 u と v の類似度 D を求める
(→各シークエンスにおける、音形 + 音高の類似度平均)
- 10 人のアンケート調査によって「似ていないし別の曲」「似ているが別の曲」「ほとんど同じ曲」の 3 段階評価 → 20 曲分で平均
- 類似度 D と人間評価の相関で評価

$$D = \max \left\{ \frac{1}{2N} \left(\sum_{t=0}^N \delta_{y_u(t), y_v(t+i)} + \sum_{t=0}^N \delta_{z_u(t), z_v(t+i)} \right) \mid i \right\}$$

δ : クロネッカーデルタ関数,
 $y_u(t)$: 時刻 t における曲 u の音形,
 $z_u(t)$: 時刻 t での曲 u の音高

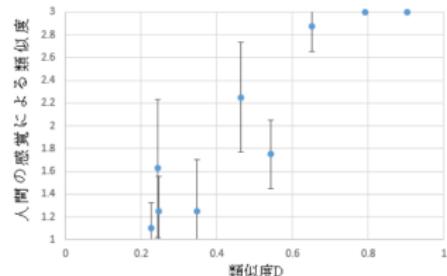


図2. 類似度Dと人間の感覚の相関

方向性

ほぼ決まったと考えている。

今後の予定

- 先行研究の調査 (どんな特徴量を組み込んでいるか)
- 提案手法の捻出 (表現の幅を広げる特徴量の検討)
- 提案手法を考えた際の評価方法の検討 (例. 教師データを上手く学習し, かつ, 完全な模倣でない)

はじめに

音楽情報処理分野

研究課題

まとめ

先行研究の一例

11/12

LSTM(Long Short-Term Memory) を用いた自動作曲¹

- LSTM：時系列、自然言語処理に重きを置いた深層ニューラルネットワーク
- 創作活動といえるレベルの自動作曲（被験者評価）

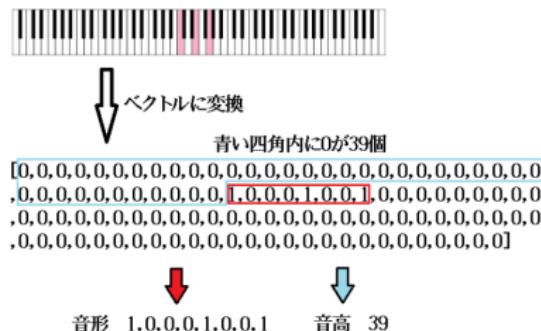


図 2: MIDI データのベクトル変換¹

¹ 姫野 雅大, "LSTM による自動作曲システムの構築", 東京大学大学院 情報理工学系研究科 修士論文, 2016.

用いるテキストデータ

12/12

ABC 記譜法

- 決まった楽譜情報に対応するテキストデータ
→これを自然言語処理して、より作曲者の意向にそった音楽を生成できる

ACupOfTea



Header行

X : 管理番号
 T : タイトル
 M : ～拍子
 L : 音符の長さの基準
 Q(X/Y):X分間のY音符の個数
 →BPM
 K(Key) : ～調

最終行(Body)=実際の音符配置

X:1
 T:ACupOfTea
 R:reel
 M:4/4
 L:1/8
 K:Amix
 Q:1/4=100
 |:eA(3AAAg2fg|eA(3AAABG
 Gf|eA(3AAAg2fg|1afged2gf:
 |2afged2cd||:eaagefgf|eaa
 gedBd|eaagefge|afgedgfg:|

図 3: ABC 記譜法のテキストデータ