

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

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

October 4, 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。

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

自動作曲分野の概要

3/17

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

RNN

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.

自動作曲分野の概要 3

5/17

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 → 改善版: **LSTM**
- 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

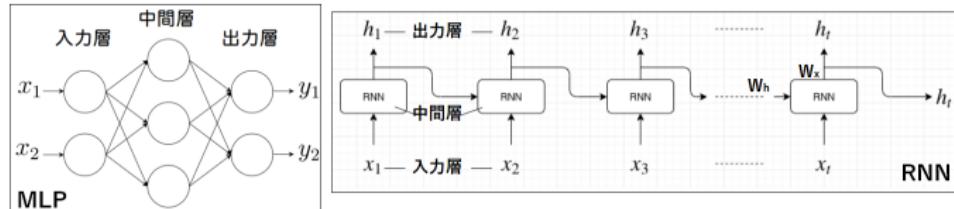


図 1: MLP と RNN

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

自動作曲の応用先

6/17

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

※研究は応用手前である機械学習手法を扱っていく(予定)

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

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

研究のメインになりそうなところ

8/17

ジャンルの偏りについて

- MIDI 形式で最新最大「MetaMIDI Dataset¹」：収録数 436,631
→ほとんどが「pop」「classic」に集中
- データ数の少ないジャンルを補うような特徴量を組み込む
→曲調の偏りを防ぐ

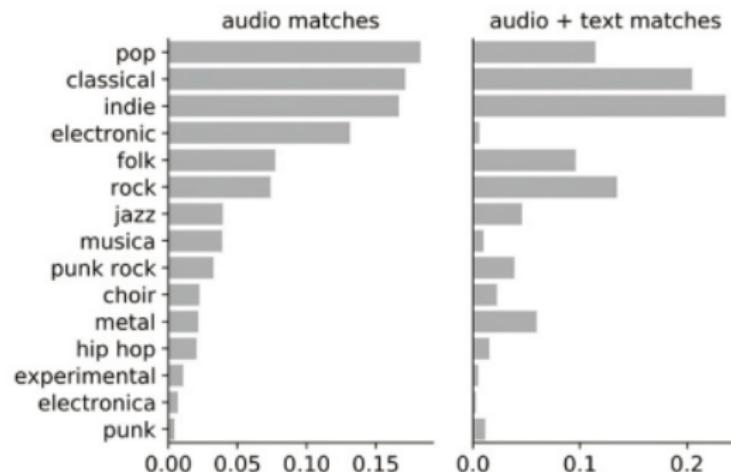


図 2: MetaMIDI のジャンルカバー率

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

ジャンルごとの特徴抽出と組み込み

9/17

Spotify API

- Spotify 上の楽曲データを Python のデータフレームに保存可能
- 様々な特徴量を取得できる

	Name	Artist	Release Date	Length	minutes	Sec	Key	Mode	key_sig	key_sig_num	BPM	danceability	Acousticness	energy	loudness	speechiness	valence
0	Lemon	Kenshi Yonezu	2018-03-14	4分16秒	4.3	256	B	メジャー	#×5	11	87	53.2	36	66.1	-5	2	44.6
1	115万キロのフィルム	Official HIGE DANDISM	2018-04-11	5分24秒	5.4	324	E♭	メジャー	♭×3	3	97	51.6	0	79.1	-2	3	48.0
2	SUN	Gen Hoshino	2015-12-02	4分1秒	4.0	241	D♭	メジャー	♭×5	1	108	78.2	2	68.4	-6	3	85.8
3	点描の唄	Mrs. GREEN APPLE	2018-07-31	5分7秒	5.1	307	C	メジャー	#/♭×0	0	152	50.2	44	63.2	-5	2	30.8
4	湘南のメモリー	Kyoko Koizumi	2017-05-17	4分5秒	4.1	245	C	マイナー	♭×3	3	156	30.1	29	72.2	-6	3	39.4

図 3: Spotify API によるデータ取得

やってみたいこと

- Spotify API でその曲のジャンルに見られる共通項を特徴として抽出
→その特徴を公開されているデータセットに組み込んで再学習
→表現の幅が広がるか検討

表現の幅についての評価

10/17

用いる手法

- ・LSTM を用いて自動生成
- ・従来の LSTM 研究に比べて表現の幅が上がったどうかの比較？

表現の幅とは？

何を持って表現の幅が上がっていると評価するのかは要検討

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

クラス分類の精度を利用

- (1) 各楽曲ジャンルのクラス分類問題を適用
- (2) データセットの少ないジャンル・多いジャンル全てのクラスにおいて、新規に作った特徴量を組み込む
- (3) 先行研究に比べて分類精度が上がったかどうかを見る

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

評価方法案 2

12/17

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

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

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

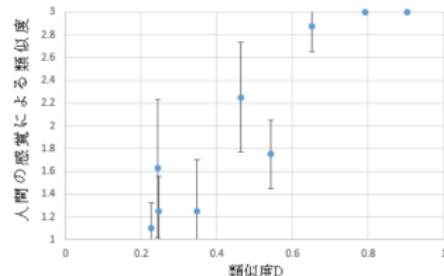


図2. 類似度Dと人間の感覚の相関

方向性

だいたい確定？

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)
- 表現の幅についての評価方法を検討

はじめに

音楽情報処理分野

研究課題

まとめ

RNN

素材置き場 先行研究の一例

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 4: 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

15/17

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 5: ABC 記譜法のテキストデータ

RNN の構造

Recurrent Neural Network(RNN)：1 個前の中間層の学習結果を次の中間層に伝達→時系列データに対応

$$\text{【時刻 } t \text{ の出力】 } h_t = \tanh(h_{t-1}W_h + x_tW_x + b) \quad (1)$$

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (2)$$

W_h : 中間→中間層への重み, W_x : 中間→出力層への重み, b : バイアス

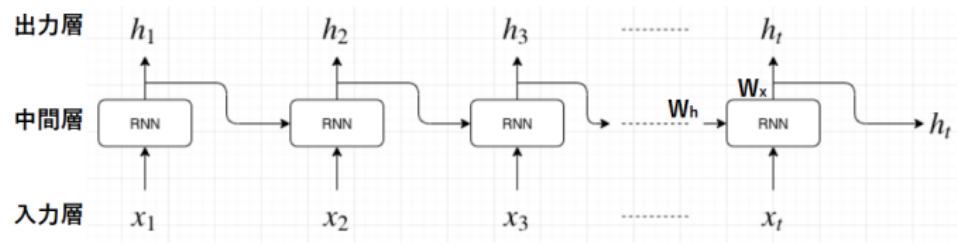


図 6: RNN の再帰学習構造 ([1] より一部改変)

1 ギークなエンジニアを目指す男，“単語と図で理解する自然言語処理（word2vec, RNN, LSTM）後編”，

<https://www.takapy.work/entry/2019/01/09/080338#RNN>, 閲覧日 2022 年 7 月 3 日

評価方法

17/17

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

RNN