

はじめに

データの概要

RNN

提案手法

結果

結果

結果

今後の展望

【論文紹介】Music Composition using Recurrent Neural Networks

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 5, 2022

はじめに ~RNN を用いた自動作曲~

2/12

研究背景

- 音楽初心者への作曲支援→機械学習 (特に RNN)
- 【従来】音声スペクトル、MIDI ファイル使用
→重要な成分 (特徴) 抽出→RNN 学習 (音主体)
- 学習モデル：音の組み合わせのみを学習
→学習した特徴を満たすどんな楽譜も生成できてしまう
⇒作曲者の意図を上手くモデル化できていない

目的

- 楽譜と直接対応するテキストデータを RNN に学習
→人間の作曲プロセスにより合ったモデリング (学習) を行う

用いるテキストデータ

3/12

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 1: 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

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

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

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

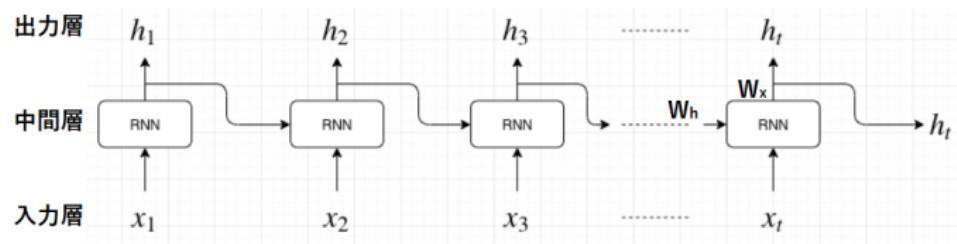


図 2: RNN の再帰学習構造 ([1] より一部改変)

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

RNNの構造

損失関数の勾配を最小化することで、パラメータ(重み、バイアス)を更新

【全損失関数】 $L = \sum_{t=1}^T L^t \quad (L^t = (h_t - \hat{h}_t)^2)$ (3)

【勾配】 $\frac{\partial L^t}{\partial W_h}$ (4)

L^t : 時刻 t における損失関数(出力誤差), h_t : 予測出力, \hat{h}_t : 教師データ

はじめに

データの概要

RNN

提案手法

結果

結果

今後の展望

自然言語処理(単語の予測)

6/12

はじめに
データの概要
RNN
提案手法
結果
結果
今後の展望

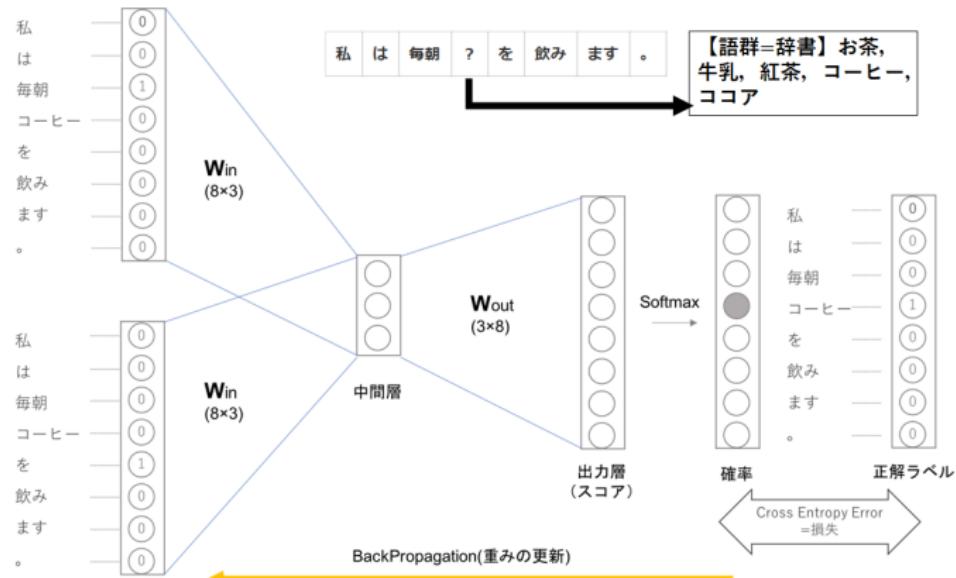


図 3: 自然言語処理における単語予測

自然言語処理(単語の数値化)

7/12

はじめに
データの概要
RNN
提案手法
結果
結果
今後の展望

One-Hotベクトル

私	1	0	0	0	0	0	0
は	0	1	0	0	0	0	0
毎朝	0	0	1	0	0	0	0
を	0	0	0	1	0	0	0
飲み	0	0	0	0	1	0	0
ます	0	0	0	0	0	1	0
。	0	0	0	0	0	0	1

単語の重みづけ(中間層)

男	0.01	0.46	0.34	...
東京	0.23	0.78	0.08	...
Ruby	0.88	0.04	0.54	...

図 4: 自然言語処理における単語予測

提案手法

8/12

seq2seq モデル

- 【従来】前の単語学習⇒後ろの単語予測 + テキスト外の単語も含めて全て学習
→全体の文脈を意識していない
→学習時間が長い
- 【提案】前の1文予測(前単語⇒後ろ単語)
⇒後ろの1文予測 + テキスト内の単語のみ
→文脈を意識 + 学習時間短縮

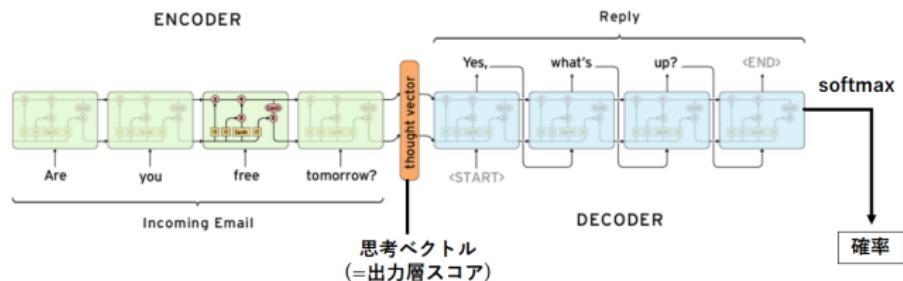


図 5: 提案手法のモデル

データの内容

- ABC 楽譜を取り扱うサイトから 34,400 種のデータを取得
→ジャンル：ジグ，ワルツ，中国系の民謡
- フォーマット・整数エンコーディング後，
学習：テスト：検証=8：1：1 で，データ分割

実験内容

- 従来手法(単語-単語予測)との予測精度比較
- 実際に人が作曲した曲の前半を学習→後半を予測
- 聴者による評価

結果 1

10/12

従来手法との比較

- 従来:20%→提案 65.5%

実際の曲の進行予測



図 6: 進行予測結果

はじめに

データの概要

RNN

提案手法

結果

結果

今後の展望

結果 2

11/12

被験者評価

- 48 人に対して検証
- 従来モデル、提案モデル、人作曲の各 3 曲、合計 9 曲
- 提案手法の 2 曲で 50% 以上が人口作曲と間違われた → リアリティのある音楽

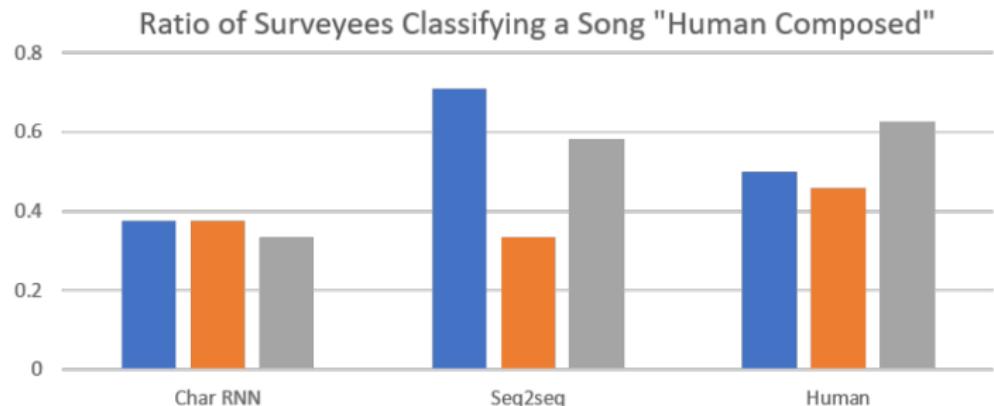


図 7: 提案手法のモデル

まとめと今後の展望

12/12

まとめと今後の課題

- ABC 記譜法を用いた自動作曲の有効性を示した
- デュエット曲などのもっと幅広いジャンルへの適用

自分の研究に対して

- 機械学習よりの自然言語処理を勉強できた（テキストベースの自動作曲）
- まだもう少し、時系列予測の前提知識を勉強する必要性

はじめに

データの概要

RNN

提案手法

結果

結果

今後の展望