

深層学習を用いた楽曲からのフレーズ抽出システムの設計

2018SE032 鬼頭 幸佑
指導教員：野呂昌満

1 はじめに

リズムゲームには二種類ある。「太鼓の達人」などで見られる、「判定音」「タップ音」と呼ばれる楽曲とは関係のない効果音が出るものと、「beatmania」シリーズなどで見られる、「キー音」と呼ばれる楽曲中の特徴的な音を切り出し用いるものである。本研究ではキー音を用いたリズムゲームを対象とする。

リズムゲームの譜面もとい楽曲は人の手で作られている。プレイヤーから良譜面だと判断されるには作り手の経験やセンスが問われる。

AIによる楽曲の自動生成はすでにシステムとして設計されているが、まだ精度の向上が見込められる。

本研究では、楽曲から音響特徴量と時系列データを抽出し、曲の一部をフレーズとして切り出し、新たな楽曲として配置することを目的とする。

本研究内のフレーズとは、その楽曲内の特徴的な小節を指し、上記にあるキー音に近いものである。

本研究の課題は以下の通りである。

1. 楽曲の音響特徴量の抽出
2. 譜面ファイルから時系列データ取得
3. 抽出したデータから楽曲のどこをフレーズとして切り出すかをリアルタイムで推定・配置

本研究により、リズムゲームのための特徴的なフレーズを切り出し適切に配置するシステムが提案され、そのシステムが適切かどうかを評価することができる。

2 先行研究および既存技術

福永ら[1]は、Weinerら[2]によって提唱された与えられた文字列の全ての接尾辞を表した木構造のデータ構造であり、主に文字列検索などに利用されている、suffix tree を用いて音響特徴量と時系列データから学習を行うことでキー音とBGMの分類を自動化した。接尾辞とは、ある文字列 X があるとする。すると X の接尾辞は先頭から 0 文字以上削ったものを指す。

また、楽曲には様々な個性が存在することが辻野らによって示唆されている。[3] これにより福永らはクラスタリングを行うことで個性をモデリングし、楽曲自動生成の機械学習モデルを作成した。[4]

香川ら[5]はFBKCE法によりMIDIファイルから重要音を抽出した。

Donahueら[6]は「Dance Dance Revolution」について、楽曲のスペクトルデータとニューラルネットワークにより自動的に楽曲を生成する手法を提案・実装した。

以下に既存技術を示す。

2.1 FBKCE法

FBKCEとはFrequency-Based Key Component Extractionの略で、楽曲中の音符が持つ絶対時間・チャンネル番号・音高・

音の長さの四つの情報から一部を利用し、任意の音符列の音数と出現回数を数え二次元グラフ化する。そして、しきい値を音符に与え、その値以上の音数や出現回数の音符を重要音として抽出する。

2.2 Dance Dance Convolution

CNNとLSTMを組み合わせたC-LSTMという手法を用いて提案された。

楽曲の生成には二つのサブタスクがある。一つは「時刻決定タスク」である。これは音符を意味するノートと呼ばれるものをどのタイミングで配置するかを推定。ここではスペクトログラムに対してCNNを用いるonset detection技術を応用している。[7] もう一つは「向き決定タスク」である。これは時刻決定タスクで得た情報を基に、どこにどの向きのノートを配置するか決定する。Donahueらは楽曲内の64ステップを入力し、C-LSTMモデルを使って65ステップ目を推定した。

3 提案方法

フレーズの推定には抽出した音響特徴量と譜面ファイルから得た時系列データを基に学習データを作成し、pythonを用いてリアルタイムに音の切り出しを行う。

それと同時に、DonahueらのC-LSTMモデルを応用し、時系列データの学習可能なモデルを構築する。

本研究の提案方法の全体構造を図1に示す。

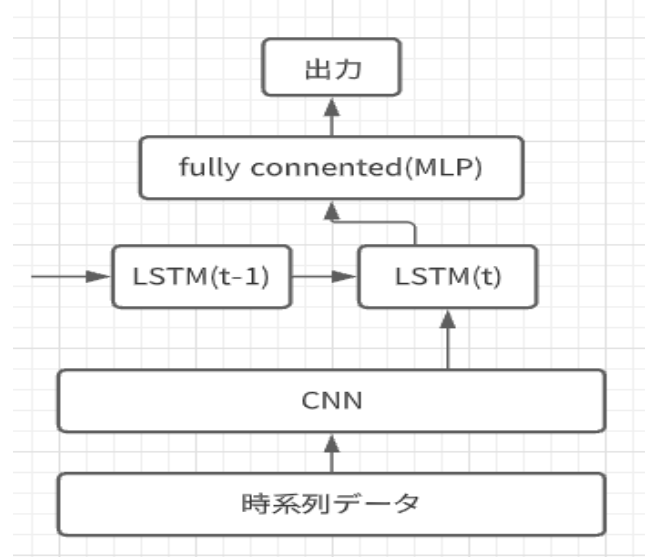


図1 提案方法の全体構造

4 設計と実現

4.1 設計

4.1.1 音響特徴量の抽出

楽曲中の音響特徴量抽出には、音声・音楽情報処理向けの特徴量抽出ソフトであるopenSMILE[8]を使用する。

4.1.2 譜面ファイルから時系列データ取得

時系列データ取得には、楽曲フォルダ内にある譜面ファイルのテキスト解析を行う。

譜面ファイルはメモ帳で開くことができ、その一部を図 2 に示す。図 2 で示した部分はヘッダで、ここにはその曲のタイトルや作者、BPM などの情報が記されている。

メイン部分はヘッダの次に記されており、そこで楽曲のノーツを配置している。

これらの情報からあるタイミングでのノーツの位置、チャンネル記述、BPM、秒単位を読み取る。

```
*----- HEADER FIELD
#PLAYER 1
#GENRE ProgressiveTrance / Electro
#TITLE Black Rainbow
#ARTIST rilym
#BPM 130
#PLAYLEVEL 4
#RANK 3
#TOTAL 250
#STAGEFILE title.jpg

#WAV02 bomb_000.wav
#WAV03 basschu_000.wav
#WAV04 basschu_001.wav
#WAV05 basschu_002.wav
#WAV0F opencymbal_000.wav
#WAV0G closecymbal_000.wav
#WAV0H closecymbal_000.wav
```

図 2 譜面ファイルの一部

4.2 実装

openSMILE を実装した際のコードを図 3 に示し、ある曲に対して抽出した特徴量の一部を表 1 に示す。

```
import opensmile

smile = opensmile.Smile(
    feature_set=opensmile.FeatureSet.eGeMAPSv01b,
    feature_level=opensmile.FeatureLevel.Functionals,
)
output_features = smile.process_file("[rilym BGA.8 to 7] Black Rainbow.wav")
print(output_features)
```

図 3 openSMILE コード

表1 使用する特徴量

特徴量の名称	説明	結果
LogEnergy	音量(対数)	4.225634
F0	基本周波数	20.399567
SpectralFlux	スペクトルの変化量	2.547059

時系列データの取得には本研究では主成分分析を用いた。譜面ファイルから得た時系列データの例を図 4 に示す。

[Black Rainbow.wav: 1.0.0, 01, 130.0, 1.36]

図 4 時系列データの例

5 考察

先行研究では対象のノーツから 10 個分過去のノーツについて機械学習をしているが、一曲には 1000 個以上ノーツが存在するので深層学習が有効と考えられている。

抽出した音響特徴量とテキスト解析で得た時系列データで学習データを作成し、Donahue らの C-LSTM を使用すれば、LSTM を用いているので、曲全体の流れから、ある程度の曲のパターンとその時々ノーツの流れを記憶することができ、さらにキー音の推定精度を上げることができると考えた。

今回「キー音」を用いたリズムゲームについて研究を行ったが、提案手法を「タップ音」を用いたリズムゲームに使用した場

合、キー音を用いたリズムゲームの楽曲を生成するには楽曲からフレーズを抽出し、キー音となるオブジェクトを推定して適切な位置に配置するという三つの過程が必要である。

それに対してタップ音を用いたリズムゲームではフレーズを抽出することには変わりはないがキー音と BGM を分類する必要がない。よって提案手法の推定する部分が空回りすることになり、あまり効果がないと考える。

6 おわりに

本研究では深層学習を利用し、リズムゲームにおけるフレーズの抽出を目的とした。

今後の課題として本研究の提案手法を用いてニューラルネットワークを構築し実験を進め、ユーザ評価を行い、その結果を基にさらなる精度の向上を目指していく。

ユーザ評価では、ある曲において提案手法によって自動生成された譜面と、BMS[9]内に実装されている楽曲の2つを用意し、その楽曲 IT を無作為に選んだ被験者 10 名にプレイしてもらい、「譜面の難しさ」・「譜面の楽しさ」・「曲に合っているか」の3つについて5段階で評価しようと考えている。

参考文献

- [1] 福永大輝, 越智景子, 大淵康成 : リズムアクションゲームにおけるキー音の自動推定, 芸術科学会論文誌, Vol.18, No. 1, pp. 10-18 (2019).
- [2] P. Weiner: Linear pattern matching algorithms, SWAT '73 Proceedings of the 14th Annual Symposium on Switching and Automata Theory, pp. 1-11 (1973).
- [3] 辻野雄大, 山内良典, 山下洋一, 井本桂右 : ダンスゲーム譜面の特性分析とクラスタリングに基づく特徴的な譜面の自動生成, エンターテインメントコンピューティングシンポジウム 2019 論文集, pp. 96-103 (2019).
- [4] 福永大輝, 越智景子, 大淵康成 : キー音を用いたリズムアクションゲームにおける譜面の自動生成, 情報処理学会研究報告, Vol 55, No. 10, 1-4 (2020).
- [5] 香川俊宗, 手塚宏史, 稲葉真理 : 音楽の重要な構成要素の抽出の提案 -音楽ゲーム用譜面自動生成のために-, エンターテインメントコンピューティングシンポジウム, EC2015, pp. 326-333 (2015).
- [6] Chris Donahue, Zachary C. Lipton, Julian McAuley, Dance Dance Convolution, ICLR 2017 Workshop (2017).
- [7] {村上真} onset (音の開始点), (2020-6-21), <https://makotomurakami.com/blog/2020/06/21/5790/>, (閲覧日 2021-9-23).
- [8] Florian Eyben, Felix Weninger, Florian Groß, Björn Schuller, Recent Developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor, Proceedings of ACM Multimedia 2013, Barcelona, Spain, ISBN 978-1-4503-2404-5, pp. 835-838, 2013.
- [9] {シア(id:Xia_Cookie)} 万物快樂理論, “超絶神無料音ゲー”BMS”やりませんか?”, (2018-9-14), <https://midnightsky.hatenablog.com/entry/2018/09/14/232901>, (閲覧日 2021-9-10).