

動画の感情表現に合わせたBGM自動生成の試み

原田楓・小高知宏・黒岩丈介 (福井大学大学院工学研究科)
白井治彦 (福井大学工学部)・諏訪いずみ (仁愛女子短期大学)

1 はじめに

動画の内容に適したBGMの選択は、動画作成において重要な要素である。現在、多種多様なジャンルの動画が動画サイトやSNSで数多く投稿されていることから、今後BGMの最適な選択はさらに重要になってくると考えられる。

そこで本研究では、動画から得られる感情表現に合わせたBGMの自動生成を行う。感情は、動画と音楽の親和性を生み出す要素の一つである。動画と音楽、どちらの分野においても感情に関する研究が行われてきた[1]。長短期記憶(LSTM)を用いて感情からBGMを生成し、生成されたBGMがどの程度動画の感情に適しているかを評価する。

2 生成手法と入力データの設定

LSTMとは、リカレントニューラルネットワーク(RNN)の一種である。LSTMの特徴として、中間層がメモリと3種類のゲートを持つブロック(LSTM Block)で構築されている。3種類のゲートには、入力やメモリ情報を選択的に伝える働きがある。この中間層によってLSTMは、通常のRNNでは出来ない長期間の情報保持が出来る構造になっている。そのため、時系列データの長期学習が可能である。LSTMを用いた音楽生成は先行研究でも数多く行われているため[2]、本研究でもLSTMを生成手法として用いる。

本研究では、動画から得られる感情を入力データとして、BGMの生成を行うことを目的とした。今回は、手動で動画から得られる感情を指定した。動画素材提供サイトpixabayにて、指定する感情をキーワードとして検索を行った。感情の種類としては楽しい、穏やか、悲しい、恐ろしいの4種類である。検索結果より、感情と関連性のあるとされた動画を収集した。

3 コード進行の生成実験

BGMの生成方法として、まず感情とコード進行のデータセットをLSTMに学習させた。BGM配布サイトDOVA-SYNDROMEにて、2章で指定した4種類の感情をキーワードに検索を行い、検索結果から得られたBGMのコード進行を音楽解析ソフトwavetoneを用いて解析した。4種類の感情からそれぞれ得られたBGMのコード進行をデータセットとした。

次に、図2のブロック図の流れでBGMの生成を行った。コード進行を学習させたLSTMに感情の入力を行う。今回は、感情に合わせて初期コードを設定しており、LSTMは

初期コードに続く新しいコード進行を生成する。生成されたコード進行はテキストデータとして出力される。得られたテキストデータを演奏・変換ソフトTexChord 2に入力し、音データに変換して出力する。

その後、出力された音データが入力した感情とどれくらい一致しているか評価する。評価方法としては、入力した感情と関連性のある動画のBGMに生成した音データを使用する。音データがBGMとして合っているか判定する。実験の詳細な結果は当日発表する。

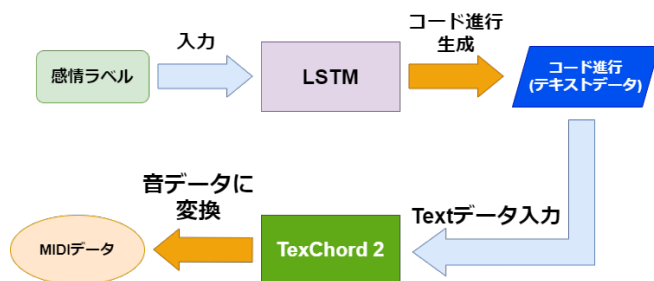


図1: 生成システムのブロック図

4 考察とまとめ

今回、LSTMを用いて感情からコード進行を生成し、生成されたコード進行の評価を行った。コード進行自体は生成することが出来たが、評価は現時点で、十分な結果が得られていない。

今後の展望として、生成されるコード進行のテンポやリズムを入力する感情に合わせて変更することで、より感情に適した生成が可能になると考える。また、コード進行に合わせたメロディーラインも生成できると、よりBGMとしての実用性が上がると思われる。

参考文献

- [1] Yi-Hsuan Yang, Yu-Ching Lin, Ya-Fan Su, Homer H.Chen, "A Regression Approach to Music Emotion Recognition", IEEE Transactions on Audio, Speech, and Language Processing Vol.16, Issue 2, pp.448 - 457, 2008
- [2] Douglas Eck, Juergen Schmidhuber, "A First Look at Music Composition using LSTM Recurrent Neural Networks", Technical Report No.IDSIA-07-02, Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, 2002