5D-06

深層学習による歌詞を考慮したコード進行の自動生成

世良　拓也†　大井　雄介†　浜野　佑介†　谷口　浩平†
日本電気株式会社†

1. はじめに

楽曲において、コード進行は楽曲の雰囲気やオリジナリティに影響する重要な要素である。コード進行を含む楽曲制作には、音楽理論や経験が必要であり初心者にとってそのハードルは高い。また、歌詞の韻律に基づいた自動作曲［1］に関する研究は存在するが、歌詞を入力としたコード進行の生成に関する研究[2]は途上の段階である。

そこで、歌詞を考慮したコード進行の自動生成手法を提案する。具体的には、Seq2Seq を用いて歌詞を入力し、コード進行を出力するモデルを構築する。また、生成されたコード進行の定量的・定性的な評価手段についても検討する。

2. 提案手法

2.1 データセット

Web 上からスクレイピングして歌詞とコードのペアを取得し、各曲の key を C に統一した。また、″C,G,F,Am,Em,Dm,G7,Dm7,Em7,FM7,Am7,CM7,Bm7-5,Bdim″のみが出現するコードと歌詞のペアのみに絞り込みを行い、207,891 件のコードと歌詞のセットをデータとして用いた。データセットの例を以下に示す。

歌詞:冬の朝は寒い

コード:CFG

2.2 モデル

本研究の特徴は、attention 機構を付与した Sequence-to-sequence モデル(Seq2seq)[3]を用いて、歌詞を考慮したコード進行の生成を試みた点にある。

Seq2seq は系列データ(文章など)を入力として別の系列データを出力するモデルであり、主に機械翻訳や文章要約に用いられる。Seq2seq はエンコーダとデコーダから構成されている。エンコーダは入力された系列データから情報を抽出した「コンテキスト」と呼ばれるベクトルを生成し、デコーダへと送る。デコーダはコンテキストから出力要素を生成する。

本研究では、エンコーダ、デコーダに GRU[4]を採用する。エンコーダへ対象となる歌詞を入力し、歌詞のコンテキストを生成、デコーダへと送る。デコーダでコンテキストから歌詞に応じてコード進行の生成を行う。

3. 評価

3.1 評価の目的

2 章の手法により生成されたコード進行について、以下 2 つの観点での評価を試みた。

■定量的評価

収集したデータを学習データと評価データに分割し、評価データとその予測結果におけるコードの再現率、編集距離の評価を行う。

正解となる評価データに対し、生成結果のコードのうち位置、コードが共に一致したものの割合を再現率とする。正解のコードを予測できていたかの指標として、再現率を用いる。

$$\text{再現率} = \frac{\text{予測と正解が一致していたコード数}}{\text{正解のコード長}}$$

また、正解のコードに対して予測のコードがどれくらい近いかの指標として編集距離[5]を用いる。

本来、文字列の操作となるが本研究では Dm7 のような複数文字からなるコードも、1 コードを 1 文字として扱う。

編集距離では、予測コードから正解コードに削除・置換・挿入の操作を何回繰り返して戻せるかで評価する。しかし、予測したコードの数が一致していると最大コード長の操作で正解コードに戻せてしまうため、本研究では置換を除いた削除・挿入の操作何回で正解のコードに戻せるかを評価の指標として用いる。編集距離が 0 となる場合、予測が完全に一致したことを示す。

■定性的評価

入力した文章と生成されたコード進行を提示し、その文章の印象に合っているかアンケート評価を行う。

3.2 評価データ

学習データと評価データの分布を表 1 に示す。

表 1 データの分布

	学習データ	評価データ
データ数	166,356	41,535
平均コード長	3.18	3.09

Automatic generation of lyric-aware chord progression using deep learning
Sera Takuya†, Oi Yusuke†, Hamano Yusuke†, Taniguch Kohei†
†{takuya-sera, y-oi-cd, yusuke-hamano, kohei-taniguchi}@nec.com
†NEC Corporation

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

4. 結果

4.1 定量的評価

epoch ごとの再現率平均、編集距離の推移を図 1 に示す。横軸は epoch、左側の縦軸は再現率平均と編集距離が 0 となるデータの割合、右側の縦軸は編集距離平均を示している。epoch18 の時点で、再現率平均は 0.49、編集距離 0 割合は 0.31、編集距離平均は 2.92 となった。

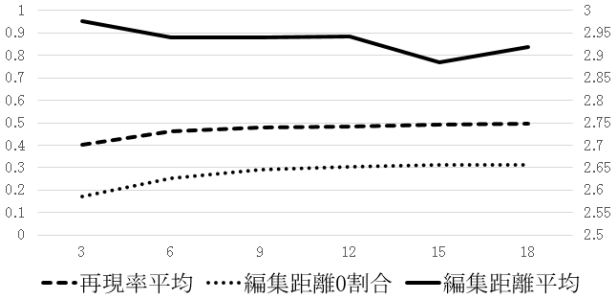

図 1 再現率、編集距離の推移

正解のコードに対して、評価データから無作為に抽出したコードを予測結果とするランダム評価を行ったところ、再現率平均は 0.18、編集距離平均は 4.27 となった。

4.2 定性的評価

ある文章を入力して生成されたコード進行を再生し、その文章の印象に合っているかアンケートを行った。参加者は 20~50 代会社員 289 名である。結果を図 2 に示す。

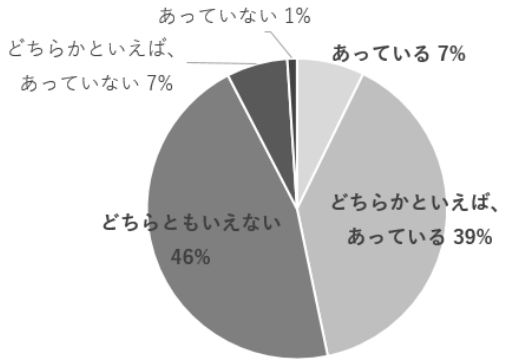

図 2 アンケート結果

5. 考察

5.1 評価結果

定量的評価では、再現率平均、編集距離平均共にランダム評価の精度を提案手法が上回る結果となった。学習が進むにつれて再現率平均、編集距離が 0 となる予測結果の割合が増加していることが確認できる。平均編集距離は 2.95 前後を推移しており、明確な精度向上は確認できない。これは、編集距離が 0 となる割合が増加するものの、コード長が長い正解に対する予測を誤ることで、編集距離が大きいデータが増加してしまうことなどが原因であると考える。

定性的評価では、肯定的な回答が合計 46%、否定的な回答が 8%となった。よって、今回の生成結果は人が持つ文章の印象と大きく異なってはいないと考える。

5.2 課題

今回の手法ではコード進行を絞って学習を行っているためバリエーションが少なく、生成結果がコントロールできないという機能的な課題が存在する。また、精度向上に向けて今後は Transformer をはじめとした他のアルゴリズム、データ収集・学習の手法の検討を行いたい。

今回の精度評価には再現率平均、編集距離を用いたが、一般的に生成タスクで使用される他の手法には BLEU、METEOR などがある。今後はこれらの手法も使用して評価を行いたい。また、定性的評価においても、印象に関するより詳細なアンケート項目について検討したい。

6. まとめ

本研究では、Seq2Seq を用いて歌詞を考慮したコード進行の生成を試み、2 つの観点で評価を行った。結果として再現率平均は 0.49、編集距離平均は 2.92 となり、ランダム評価より高い精度となった。

今後はコードのバリエーション増加、結果のコントロール、評価方法の更なる検討を行いたい。

参考文献

[1] 深山覚, 中妻啓, 米林裕一郎, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹."Orpheus:歌詞の韻律に基づく自動作曲システム",情報処理学会研究書報告, MUS-076,pp.179-184, 2008.

[2] a2kiti. "歌詞からコード進行の生成 Attention による可視化".Qiita. https://qiita.com/a2kiti/items/d75b093d8328f3e382b6,(参照 2022-01-05).

[3] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. arXiv preprint arXiv:1409.0473

[4] Cho,Kyunghyun,et al. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.

[5] GusfieldDan.(1997).Algorithms on strings, trees, and sequences. Cambridge University Press.

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.