

タイミング情報のサブバンド分解を用いた RNN による音楽表情付け

菊地浩太郎[†] 西村憲[†]

会津大学大学院コンピュータ理工学研究科[†]

1. はじめに

音楽の自動表情付けは音楽情報処理における主要なテーマの一つであり、近年ニューラルネットワーク(neural network, NN)を使用した研究が多く行われるようになった。本稿では、再帰型ニューラルネットワーク(recurrent neural network, RNN)を用いて、楽譜に書かれたピアノ楽曲への表情付けを行う。

NNを利用したものとして、強弱の表情付け[1]や、テンポの揺らぎの表情付け[2][3]が存在している。いずれも双方向 RNN を基盤としたモデルが使用され、自動表情付けに対して有効であると示している。しかし、これらのテンポの揺らぎの表情付けでは、和音構成音の発音時間のずれなど細部の表現は考慮されていない。

本稿では、RNN を用いて表情付けを行う際、タイミング情報に対しサブバンド分解を施すことを提案する。サブバンド分解により、楽譜上の発音時刻と演奏上の時刻との関係が、巨視的なテンポの揺らぎと各音における微視的な時間偏差に分解される。これによりテンポの揺らぎだけでなく、和音やリズムパターンにおける各構成音の時間偏差を表現でき、特にポリフォニック楽曲に対して、豊かなアゴーギク表現の生成が期待できる。

2. 処理の概要

本手法では、RNN の一種である双方向 LSTM(long short-term memory)を用いて、16 分音符毎のテンポと、各音における時間偏差を出力させる。

図 1 に本手法の処理の概略を示す。学習では、まず演奏情報と楽譜情報で各音の対応付けを行い、そこからテンポカーブを抽出する(①)。次に、抽出したテンポカーブに対してサブバンド分解を施し、平滑化をかけたテンポカーブと各音毎の時間偏差の情報を取得する(②)。ここでの時間偏差とは、楽譜上の各音における発音・消音の時刻を平滑化テンポカーブによって演奏時間に変換したものと、実際の演奏時刻との差分を意味する。平滑化には、スペクトルの平滑

Generating music expression with RNNs using the sub-band decomposition of timing information

Kotaro KIKUCHI, Satoshi NISHIMURA[†]

Graduate School of Computer Science and Engineering, University of Aizu[†]

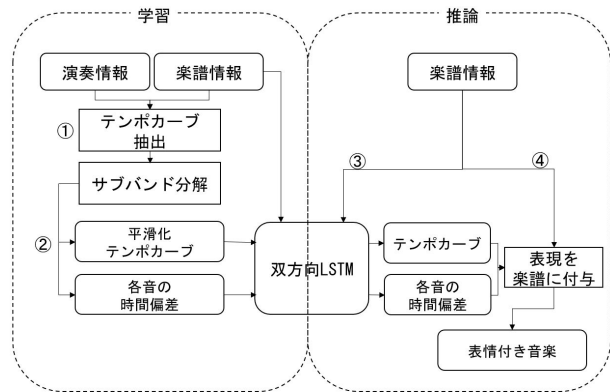


図 1: 処理の概要

化でよく使われる Savitzky-Golay 法[4]を用いた。

楽譜情報の NN への入力には Malik ら[1]の方法に基づいている。各 MIDI ノート番号について、発音開始、発音持続中、発音終了、無音の 4 つのいずれであるかを示す符号を 16 分音符毎に入力する。この楽譜情報に加えて、平滑化テンポカーブ、各音の時間偏差を双方向 LSTM へ入力し学習を行う。

推論においては、新たな楽曲の楽譜を学習済みモデルに入力する(③)。学習済みモデルからテンポカーブと各音の時間偏差が出力され、これらを楽譜に対して付与することで表情付き音楽が生成される(④)。

3. 実験

実験では古典派の楽曲を使用し、Wolfgang Amadeus Mozart と Ludwig van Beethoven のピアノソナタから 58 楽章分を学習用に選定した。演奏・楽譜情報としては、音や発音持続時間などが記述されている標準 MIDI ファイルを用いた。演奏情報の MIDI ファイルは e-competition での録音[5]から、楽譜情報の MIDI ファイルは KernScores [6]より取得した。データは±6 半音ずつ移調してデータの水増しを行った。一度に 4 小節ずつ入力を行い、2 小節ずつ重ねながら入力をずらしながら学習を行った。双方向 LSTM の構成は参考文献[1]を参考にした。

学習の際、時間偏差が 4 分音符以上の長さのものは演奏情報と楽譜情報の対応付けミスと判断し、時間偏差がないものとして扱った。また、テンポの揺らぎは、4 小節ずつ、その先頭のテン

自然であるか	窓枠・次元数	ピアノ未経験者		ピアノ経験者		音楽大学ピアノ専攻	
		N>R	N>M	N>R	N>M	N>R	N>M
Mozart: 第1番 ハ長調 K. 279 (189d) 第1楽章	11, 3	75%	25%	100%	50%	100%	0%
Mozart: 第18(17)番 ニ長調 K. 576 第2楽章	11, 3	75%	0%	25%	50%	67%	0%
Beethoven: 第1番 ヘ短調 作品2-1 第1楽章	11, 3	100%	75%	50%	25%	33%	33%
Beethoven: 第3番 ハ長調作品2-3 第2楽章	5, 3	50%	50%	50%	50%	67%	33%
Beethoven: 第11番 変ロ長調 作品22 第1楽章	11, 3	100%	25%	100%	25%	67%	0%
人間らしいか							
Mozart: 第1番 ハ長調 K. 279 (189d) 第1楽章	11, 3	50%	25%	50%	75%	33%	100%
Mozart: 第18(17)番 ニ長調 K. 576 第2楽章	11, 3	75%	50%	75%	50%	67%	67%
Beethoven: 第1番 ヘ短調 作品2-1 第1楽章	11, 3	75%	75%	25%	50%	33%	67%
Beethoven: 第3番 ハ長調作品2-3 第2楽章	5, 3	50%	50%	25%	75%	33%	67%
Beethoven: 第11番 変ロ長調 作品22 第1楽章	11, 3	100%	50%	75%	100%	100%	100%

図 3: NN, 乱数, 表情なしの音源の比較アンケートの結果. “自然であるか” についての回答 (上). “人間らしいか” についての回答 (下). 表情を NN で生成したものを N, 乱数で生成したものを R, 表情がないものを M とし, 例えば N>R は乱数で生成したものより NN で生成したものがより自然である (あるいは人間らしい) ことを意味する. 窓枠・次元数は使用したサブバンド分解のパラメータである.

ポ(1 分間あたりの 4 分音符数)を基準とし, ± 30 以上の変動がある場合は ± 30 を上限, 下限として扱った.

4. 評価結果

本手法の表情付けを評価するため, 比較アンケートを実施した. 評価には, Mozart および Beethoven のピアノソナタのうち, 学習データセットには含まれていない楽章から 20-30 秒のフレーズを 5 つ使用した. 対象者は音楽大学ピアノ専攻の学生 3 人, および, 会津大学の 1 年以上のピアノ経験者 4 人とピアノ未経験者 4 人の計 11 人とした.

アンケートでは, NN で表情を生成した音源, 正規分布乱数によって表情を生成した音源, 表情を与えていない音源の 3 種で比較してもらった. 比較時, 参考になるよう人間による演奏も聴くことを可能にした上で, “自然であるか”, および “人間らしいか” について順位付けを行ってもらった.

NN から生成されたもののうち, 表情があまりにも強く反映され不自然な場合や生成値が小さく変化を感じられない場合は, 生成値に定数倍をかけ調整した. 本手法の表情付き音楽の例を図 2 に示す. 乱数の分散値は, NN で生成した表現から外れ値を除いて得られる分散値と一致するようにした. 乱数により生成されたテンポは, 大きく変動しないように平滑化を行った.

アンケートの結果として図 3 の評価が得られた. 曲によって傾向が異なるが, 概ね NN による表情付けは, 乱数より自然ではあるが表情なしよりは不自然に感じられるという結果となった. しかし, 人間らしいかについては表情なし音楽よりも優位であることが分かった.

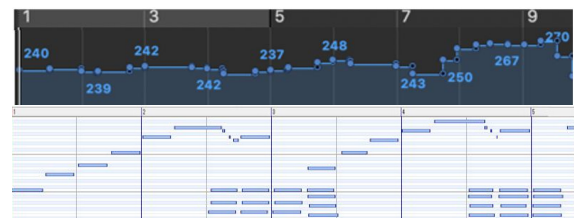


図 2: 本手法による表情付けの例. Beethoven ピアノソナタ第 1 番第 1 楽章, 第 1 小節から第 9 小節までのテンポの揺らぎ (上). 同楽曲, 第 1 小節から第 4 小節までの時間偏差を付与した表情付き音楽 (下).

5. まとめ

本稿では, 双方向 LSTM に基づいた RNN を用い表情付けを行う際, タイミング情報に対しサブバンド分解を施すことを提案した. 自然でないと感じられた曲でも人間らしさは感じ取れるといった結果となった. 今後の課題として, サブバンド分解の効果についてさらに調査を行い, 多重解像度分解による表情付けへと発展させていきたい.

参考文献

- [1] I. Malik and C. H. Ek, “Neural Translation of Musical Style,” CoRR, arXiv:1708.03535, 2017.
- [2] M. K. Jędrzejewska, A. Zjawinski and B. Stasiak, “Generating musical expression of MIDI music with LSTM neural network,” Proceedings of the 11th International Conference on Human System Interaction, pp.132–138, 2018.
- [3] M. Grachten and C. E. Cancino-Chacón, “Temporal dependencies in the expressive timing of classical piano performances,” In M. Lesaffre, M. Leman, and P. J. Maes (Eds.), The Routledge companion of embodied music interaction, pp. 360–369, 2017.
- [4] A. Savitzky and M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” Analytical Chemistry, Vol. 36, no. 8, pp. 1627–1639, 1964.
- [5] Yamaha Corporation, “Disklavier Education Network,” <https://www.yamahaden.com/midi-files>.
- [6] CCARH, “Kern Scores,” <https://kern.humdrum.org/>.