

小規模データセットにおける音の分類モデルの高精度化に関する研究

Research on Improving the Accuracy of Sound Classification Models with Small Datasets

平岡 忠志*
HIRAOKA Tadashi

抄 録

機械学習における教師あり分類問題について、小規模データセットにおいても高精度な音の分類モデルを作成するために、疑似データを用いた学習を検証した。旋削加工時に発生するびびりの有無を判断する2クラス分類問題において、物理モデルを利用した疑似データ生成手法により作成したびびり有1,000サンプル、びびり無1,000サンプルを用いてファインチューニングしたモデルで現実のびびり有66サンプル、びびり無294サンプルを分類したところ、精度100%を達成した。これにより、小規模データセットにおける音の分類モデルの高精度化に、物理モデルにより生成された疑似データを活用することの有効性が示された。

1 はじめに

近年、人工知能の進歩により従来の技術では不可能であったタスクが遂行可能となった。高精度なモデルを構築するためには、ファインチューニング(大規模データセットで学習されたモデルを個別の小規模データセットで再学習すること)が有効であるが、小規模データセットの収集には多大な労力が必要であり、十分なデータを収集できない場合も多い。音の分類モデルの開発においてもこの問題は例外ではない。運用時に発生する音のパターンを網羅した小規模データセットを作成することは、現実の環境では難しい場合がある。

この課題を解決する方法として、音の疑似データの活用が考えられる。疑似データ生成には、正弦波の振幅と位相を調整してその和をとる調和生成法、録音した音を短い時間の断片に分割し、それらを組み合わせるテクスチャ生成法、音の周波数成分を分析し、その成分を元に生成するスペクトル分析法、振動する物体をモデル化して生成する物理ベース生成法がある¹⁾。特に物理ベース生成法では、振動する物体の寸法、剛性等のパラメータを基に、運用時に発生しうる音のパターンを網羅した疑似データを生成できる。本研究ではこの物理ベース生成法を利用した。

製造現場におけるニーズとしては、旋削加工におけるびびり有無、工具折損の有無、摩耗による工具

寿命の判断、転造加工におけるダイスの欠けの有無、ワークサイズの種類、特殊ベアリングの異常箇所特定等がある。これらの音の原因は全て剛体の振動に起因しており、この振動を基にした疑似データ生成手法とモデル作成手法を開発することで幅広い応用が可能になると考えた。そこで、本研究では、剛体の振動に起因した音の1つである旋削加工におけるびびり音について、その分類モデルの開発を課題とし、物理ベース生成法による疑似データを用いて音分類モデルをファインチューニングし、その有効性を検証した。

2 方法

2・1 びびりの発生メカニズム

旋削加工中に発生するびびり音を物理ベース生成法により生成するためには、その発生メカニズムを理解する必要がある。びびりの発生メカニズムの解明及びその対策は古くから研究されており、それらをまとめた資料としては鈴木による解説がある²⁾。びびりはワークまたは工具のいずれかが非常に大きな振幅で振動する現象であり、仮に工具がワークに比べて柔軟な場合は工具にびびりが発生する。

まず、ワークと工具が接触したとき、ワークから工具へ作用する力の影響で工具が振動する。この時、工具に働く外力は正弦波ではなく、インパルス波のように様々な振動数成分を含む力であるため、工具はその固有振動数で振動すると考えられる。旋削加工における工具は片持ち梁の曲げのモデルが適用で

* 機械技術担当

き、工具の振動は主に1次固有振動数の影響を受ける。

次に、工具が振動した状態でワークが旋削されると、その振動が加工面に転写される。1回転後、振動に影響された加工面が再び旋削されると、切込み深さが変動し、それに比例して加工力も変動するため、工具の振動が増幅される場合がある。この現象を再生効果といい、回転毎に工具の振動が増幅されると、最終的には非常に大きな振動となり、作業者は音や加工面の状態からびびりの発生を確認できる。

びびりが発生したとき、工具の1次固有振動数およびその倍数の振動数の振動が非常に大きくなる。1次固有振動数以外にその倍数の振動数の振動が出現する理由として、工具に働く外力が正弦波でないことが考えられる。例えば、その外力が三角波であれば工具は三角波で振動する。この時、三角波をフーリエ変換すると基本となる振動数とその倍数の振動数成分が大きく現れる。現実録音したびびり音は正弦波ではなく、三角波に近い形状を示すことが多かった。

工具の1次固有振動数 f は次の式(1)で計算できる。ここで、 L は工具突出長、 E はヤング率、 I は断面2次モーメント、 ρ は密度、 A は断面積である。

$$f = \frac{3.516}{2\pi L^2} \sqrt{\frac{EI}{\rho A}} \quad (1)$$

工具は完全に剛体に固定されているわけではないため、現実の振動数は理論値よりも低くなる。本研究では、一宮の文献³⁾を参考にして、理論値の約0.8倍の振動数を用いることとした。1次固有振動数の倍数の振動数については単純にその振動数を2倍、3倍、4倍にして算出した。

2・2 疑似データ生成

工具のような硬い棒の振動を物理的にモデル化して音を生成する手法としてバンドウェーブガイドがある⁴⁾。これはモーダル合成とウェーブガイド合成のハイブリッド手法である。モーダル合成は振動する物体の振動モードをモデル化したものであり、物体の複数のモードに対して、それぞれ振動数、減衰率、振幅を設定し、音を生成する手法である。ウェーブガイド合成は、弦の振動を表現した波動方程式を、左右方向に伝播する2つの波に分解して表現す

ることで効率的に解き、音を生成する手法である。プログラミング言語 ChucK⁵⁾には、バンドウェーブガイドが ModalBar クラスとして実装されており、これを利用して工具の音を生成することができる。本研究では、工具の振動を片持ち梁の曲げでモデル化し、その1次固有振動数(モード0)を ModalBar クラスの Freq に、その倍数の振動数(モード0,1,2,3)の相対値を ModalBar クラスの modeRatio に設定した(表1)。1次固有振動数については、工具突出長 L は0.040~0.063mの範囲で乱数を、ヤング率 E は205GPaを、断面2次モーメント I と断面積 A は0.014mの正方形から算出した値を、密度 ρ は7850kg/m³を設定し、式(1)を用いて算出した。また、振幅の減衰率を示す modeRadius、振幅を示す modeGain を各モードとびびり有無に応じて、現実の録音データを基にその範囲を想定した値を設定した。以上の設定で1次固有振動数の逆数の時間だけ ModalBar クラスを再生することを繰り返すことで工具の振動から発生する音を生成した。

旋削加工における工具の振動による音以外の機械音等の生成には白色ノイズにローパスフィルタを適用したピンクノイズを利用した。具体的には、プログラミング言語 ChucK の Noise クラスと Low Pass Filter (LPF) クラスを利用した。LPF クラスのカットオフ周波数を示す Freq は900~1100の範囲で乱数を設定した。

以上の技術を用いて、工具突出長と倍音のゲインを変動させて、びびり有1,000サンプル、びびり無1,000サンプルを生成し、これらを学習用データセットとした。

表1 ModalBar の設定値

項目	設定値
modeRatio 0,1,2,3	1,2,3,4 (固有振動数の倍数)
Freq	算出した1次固有振動数
modeRadius 0 (びびり有)	0.998~0.999の乱数
modeRadius 0 (びびり無)	0.998~0.999の乱数
modeRadius 1,2,3 (びびり有)	0.900~0.999の乱数
modeRadius 1,2,3 (びびり無)	0.800~0.999の乱数
modeGain 0,1,2,3 (びびり有)	全て100
modeGain 0,1,2,3 (びびり無)	75.0,0.0,0

2・3 現実の加工音

当センターにおいて録音した旋削加工の条件を表2に示した。約10秒のびびり有の録音データを7個、

約10秒のびびり無の録音データを27個得た。また、県内企業の協力を得て、約5秒のびびり有の録音データを2個得た。これらのデータを1秒間のwavファイルに分割し、びびり有66サンプル、びびり無294サンプルを作成し、これらを評価用データセットとした。

表2 旋削加工の条件

項目	品番, 加工条件
汎用旋盤	昌運工作所, HB-500×1000
回転数	1150rpm
送り	3mm/s
切込み深さ	0.1mm
マイク	ASUS, Zenfone8 付属
工具	三菱ドリル, S16MSTFCR11
高さ&幅	14mm
突出し長	20~60mm

2・4 モデルの学習方法と評価方法

音の分類モデルに Audio Spectrogram Transformer (AST) を利用した⁶⁾。ASTは10ms毎の25msハミング窓と128次元対数メルフィルタバンクで計算された音のスペクトログラムを入力とするモデルである。このため、ASTの入力はt秒間の音であれば128×100tの画像となる。また、ASTは自然言語処理で事実上の標準であるTransformerを取り入れた画像分類モデルData efficient image Transformers (DeiT)を音の分類用に改良したモデルである。このため、DeiTのパラメータはASTに転用できる。ASTの出力はびびり有と無の2クラスとした。画像分類用の大規模データセットであるImageNet-1kで学習されたDeiTのパラメータを利用して、音分類用の大規模データセットであるAudioSet⁷⁾でASTをファインチューニングした。AudioSetで学習されたASTを疑似的に生成された学習用データセットでさらにファインチューニングした。この時の学習条件は、学習率1e-5、バッチサイズ24、エポック数10とした。最後に、現実の録音データである評価用データセットで学習されたモデルを評価した。

3 結果

疑似データによる学習用データセットで音の分類モデルASTをファインチューニングし、現実データによる評価用データセットでそのモデルを評価したところ、精度100%を達成した。

現実データのびびり有のスペクトログラムを図1に、びびり無のスペクトログラムを図2に示した。横軸は時間で左から右に0~1秒、縦軸は周波数で下から上に0~20kHzの線形目盛で示した。また、ある時点のある周波数での強度を画像の明るさで示した。図1から3.3kHzとその倍数に強い音があり、びびりの特徴が確認できた。図1と図2に共通して観測できるモーター回転等の機械音は低周波で強く、高周波で弱い音として現れていた。びびり有と無の違いは基本周波数とその倍数の周波数に強い横線が現れるかどうかで、強い横線が現れた場合はびびり有、現れない場合はびびり無として判断できる。びびり無の場合でも弱い横線は現れる場合がある。

疑似データのびびり有のスペクトログラムを図3に、びびり無のスペクトログラムを図4に示した。両者に共通して観測される機械音はChucKのNoiseクラスとLPFクラスで生成したピンクノイズでうまく表現できたことがわかる。また、びびり有の特徴である基本周波数とその倍数の周波数の強い横線はChucKのModalBarクラスでうまく表現されていることがわかる。

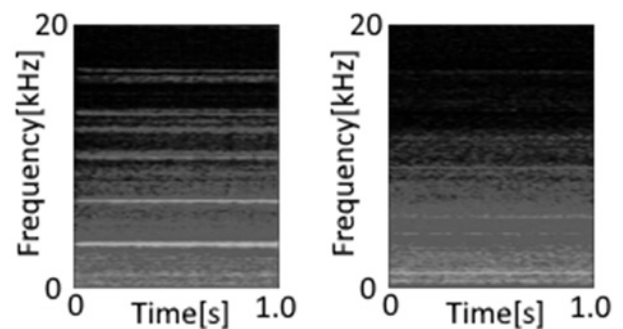


図1. 現実/びびり有 図2. 現実/びびり無

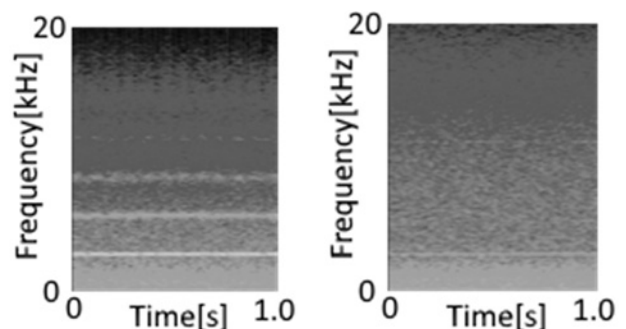


図3. 疑似/びびり有 図4. 疑似/びびり無

4 まとめ

旋削加工におけるびびり音分類モデルの開発を課

題とし、以下の成果を得た。

(1) びびりの発生メカニズムを解明し、プログラミング言語 ChucK の ModalBar クラスを利用して疑似データを生成する手法を開発した。本手法は、振動する剛体の各モードの固有振動数、減衰率、振幅の範囲を把握していれば、工具以外の剛体にも同様に適用可能である。

(2) 生成された疑似データから構成された学習用データセット（びびり有 1,000 サンプル、びびり無 1,000 サンプル）を用いて、大規模データセット AudioSet で事前学習された音分類用モデル AST をファインチューニングした結果、現実の加工音から構成された評価用データセット（びびり有 66 サンプル、びびり無 294 サンプル）において、分類精度 100%を達成した。

これにより、小規模データセットにおける音の分類モデルの高精度化に向け、物理モデルにより生成された音の疑似データを活用することの有効性が示された。

参考文献

1) Shiguang Liu, Dinesh Manocha. “Sound Synthesis, Propagation, and Rendering: A Survey”. arXiv:2011.05538. 2020.

2) 鈴木教和. “切削加工におけるびびり振動(前編)”. 精密工学会誌. 2010, Vol. 76, No. 3, p. 280-284.

3) 一宮亮一, 知久一佳, 相地誠. “切削加工における騒音の解析(第 1 報)”. 精密機械. 1983, Vol.49, No.3, p.377-382.

4) Perry Cook. “Real Sound Synthesis for Interactive Applications”. Routledge. 2002, 263 p.

5) Spencer Salazar, Ajay Kapur, Ge Wang, Perry Cook “Programming for Musicians and Digital Artists: Creating music with ChucK”. Manning. 2014, 309 p.

6) Yuan Gong, Yu-An Chung, James Glass. “AST: Audio Spectrogram Transformer”. Proc. Interspeech. 2021, p.571-575.

7) Jort Gemmeke, Daniel Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, Channing Moore, Manoj Plakal, Marvin Ritter. “Audio Set: An Ontology and Human-labeled Dataset for Audio Events”. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017.