

構造部材破壊音判定における NN と SVM の比較

東京理科大学 学生員 ○和泉田 健夫

東京理科大学 正会員 佐伯 昌之

1. 序論

本研究の最終目的は、地震時において構造部材が破壊したかどうかを自動判定するシステムを開発することである。

近年、機械学習の再燃により音声認識技術の発展が目覚ましく¹⁾、日常生活で利用される場面が多くなっている。また、防犯目的を主として、音によって起きている状況を認識する技術が開発されており、カメラ映像だけでは見えにくい場所での音による状況確認の仕組みも開発されている²⁾。

また、現在ではスマートフォンが多岐にわたって利用されており、我々の生活の利便性を向上させている。地震災害に関しては、地震観測データの収集を目的として、スマートフォンを用いた揺れの観測技術の開発も行われている³⁾。

本研究では、スマートフォンに搭載されているような安価なマイクで構造部材の破壊音を検出することを目的とし、そのための基礎的な検討を行う。

2. 使用したデータと機械学習手法

静穏な実験室内において、木材を曲げ破壊したり、相欠き継手の結合部を破壊したり、割裂させることで「バキッ」という破壊音を発生させた。その音圧を、小野測器社製の積分平均型精密騒音計 LA-4440 および音響ポータブルレコーダ DR-7100 で記録した。騒音計であるため可聴域を対象としており、サンプリング周波数は 51.2 kHz である。

機械学習する際に、音圧波形をそのまま学習データとして使用方法と、音響特徴量に変換したデータを使用する方法があるが、本研究では、音響特徴量としてメル周波数ケプストラム係数 (Mel Frequency Cepstral Coefficients, 以下 MFCC) を用いることとした。メルは、低周波数帯の音は細かく、高周波数帯の音は粗く分解するといった人間の聴覚特性をよく表す尺度である。ケプストラムは、ス

ペクトルをフーリエ変換したものであり、周波数帯により振幅がどのように変化しているかを表す量である。

MFCC の計算過程を図 1 に示す^{4) 5)}。まず、音圧データから 20 ms 程度を切り出し、それを一つのフレームとする。1 フレームの音圧データにハミングの窓関数をかけ、パワースペクトルを求める。求めたパワースペクトルに、メル周波数軸上で等間隔となるような 20 個の三角フィルタ (Mel filter bank) を適用し、各フィルタの通過エネルギーを算出する。各フィルタ通過エネルギーの対数をとったものを、離散コサイン変換 (DCT) することで MFCC を得る。以上の操作により、20 ms の音圧データは、20 成分のベクトルに変換される。

MFCC は、録音した音の音圧が高い区間とそうでない区間に分類した。さらに、木材の破壊音から得られる MFCC で音圧の最大振幅を含むフレームを BROKEN とし、木材の破壊音以外の音から得られる MFCC を OTHERS と分類した。

BROKEN と分類された MFCC は 78 個あり、学習させる際に、正解値を 1 とした。また、お皿が割れる音や会話の音など 78 個の MFCC を OTHERS

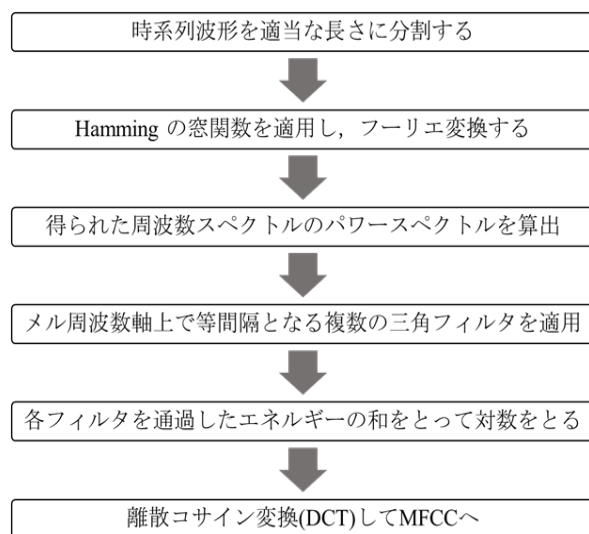


図 1 MFCC の計算フロー図

キーワード：損傷判定指標、MFCC、機械学習

から選び、学習時に正解値を 0 とした。

3. 破壊音検知手法の比較

機械学習に用いたモデルは、3 層のニューラルネットワーク (Neural Network, 以下 NN) とサポートベクターマシーン (Support Vector Machine, 以下 SVM) である。

実際に学習を行う際には、Keras や scikit-learn のモジュールを用いてコーディングしたプログラムを用いた。また、フレームによって MFCC のノルムが大きく異なるため、学習前にベクトルの正規化を行った。

(1) NN による 2 値分類

学習時のバッチサイズは 20, 25, 30, 35, 学習繰り返し回数 (エポック数) は 1000, 2000, 4000 で設定し、12 通りの学習を 156 データ用いて行った。NN の中間層のノード数は、20 の場合と 40 の場合とで設定した。学習の結果を表 1 に示す。表中に示すのは、12 通りの学習をした際に、BROKEN と OTHERS をそれぞれ正しく学習したデータの割合の平均値と最大値と最小値である。中間層のノード数が 20 のときでも 40 のときでも、BROKEN のデータの約 80% を常に正しく学習できている。中間層のノード数が 20 の場合と 40 の場合で学習精度は 0.3 ほどしか変わっていないため、3 層の NN においては、中間層のノ

表 1 NN における学習結果

(a) 中間層のノード数 : 20			
	BROKEN	OTHERS	TOTAL
AVE	81.6	100.0	90.8
MAX	82.1	100.0	91.0
MIN	80.8	100.0	90.4

(b) 中間層のノード数 : 40			
	BROKEN	OTHERS	TOTAL
AVE	81.9	93.9	87.9
MAX	84.6	100.0	91.0
MIN	80.8	26.9	55.8

表 2 SVM における学習結果

BROKEN	OTHERS	TOTAL
47.4	100.0	73.7

ド数の違いが木材の破壊音の認識精度に与える影響は非常に小さいと考えられる。

(2) SVM による 2 値分類

SVM を用いて 156 データを学習したときの、BROKEN と OTHERS の認識精度を表 2 に示す。SVM ではバッチサイズやエポック数を設定しない。SVM の学習では、BROKEN に対する学習の精度が 50% を下回っており、木材の破壊音の半分以上を正しく学習できていない。

4. まとめ

構造部材として木材に着目し、木材の破壊音検知のためのモデルについて検討を行った。

NN と SVM では、同じ学習データセットに対しての認識精度が NN の方が高かった。しかし、すべてのデータを正しく学習できていないことや、別のデータを用いた検証ができていないため、学習したモデルの信頼性が乏しい。

今後は、学習時にどの MFCC が正しく学習できていないかを確認することが必要である。さらに、NN に加え、別の音響検知手法の検証や、学習データの増加が必要であると考えられる。

参考文献

1) 河原達也：音声認識技術，電子情報通信学会誌，Vol.98, No.8, p710-717, 2015

2) 谷真宏，小松達也，ナリセッティチャイタニヤ，近藤玲史：NEC の音状況認識技術とシンガポールにおける実証実験への取り組み，電子電機情報通信学会技術研究報告，117 巻 106 号，p29-32, 2017 年 6 月 22 日・23 日

3) 国立研究開発法人防災科学技術研究所：首都圏を中心としたレジリエンス総合力向上プロジェクト サブプロジェクト(b) 「官民連携による超高密度地震動観測データの収集・整備」，2018 年 5 月

4) 久保陽太郎：音声認識のための深層学習，人工知能 29 巻 1 号，p62-71, 2014 年 1 月

5) Beth Logan：Mel Frequency Cepstral Coefficients for Music Modeling, 2000