

## ニューラルネットによる音源の特性についての検討

Abdullah Almatrafi\* 平松 綾子\*\* 能勢 和夫\*\*

\*大阪産業大学大学院工学研究科情報システム工学専攻 \*\*大阪産業大学デザイン工学部情報システム学科

### 1. はじめに

平成 24 年度の交通事故発生件数は、66 万 5138 件[1]と前年に比べて減少傾向にあるが、依然として高水準である。交通事故の 42.3%が市街地の交差点、またはその付近で発生している。そのため、一部の交差点では交通事故の発生を自動的に記録する交通事故自動記録装置が設置されている。この装置は事故の発生を音で判断し、事故の発生前後の映像を記録する。しかしながら、単に音の大きさだけで事故の発生を判断すると誤認識が起こる。例えば、装置のセンサー感度を上げると事故音以外の騒音に反応してしまい、感度を下げると事故が発生しているにもかかわらず、事故が記録されないという不具合が生じる可能性がある。そのため、環境音の識別が必要である。

本研究では、ニューラルネットワーク (以下 NN と表記) を利用した環境音の識別を目的とする。NN の入力となる音源の特徴量としては音声認識技術を参考として周波数分布をベースとした方法を検討する。

### 2. 音響データ

音響データはサンプリング周波数 44.1kHz で記録し、音響データに含まれている無音部分や雑音部分はあらかじめ取り除く。交差点で発生する音響を 13 種類のカテゴリ (音源) に分類し、衝突音等の 3 種類のカテゴリを「危険な事例」、サイレン等の 10 種類を「危険でない事例」に大別した。危険な事例の音響が発生したとき、事故が発生したと判別する。危険な音響はカテゴリ数が少ないため、データに重みをつけるために、危険なカテゴリは危険でないカテゴリの 2 倍の事例を用意した。

### 3. 先行研究

日常生活でよく聞こえる環境音 (水、雷、飛行機、など) の分類識別に関する先行研究として、文献[2][3]がある。文献[2]では MFCC と NN を組み合わせて 20 種類の自然音の分類を行い、70%程度の精度を得ている。文献[3]では、MFCC と DTW を組み合わせて 8 種類の音の識別を行い、70%程度の精度を得ている。

Study on Environmental Sound Classification Using Neural Network

\*Abdullah Almatrafi \*\*Ayako Hiramatsu \*\*Kazuo Nose

\*Graduate School, Osaka Sangyo University

\*\*Faculty of Design Technology, Osaka Sangyo University

### 4. 識別システム

#### 4.1 システムの構成

音響データに対し前処理を行い、そこから得られた特徴量を用いて NN に学習させ音源識別を行う。(図 1)

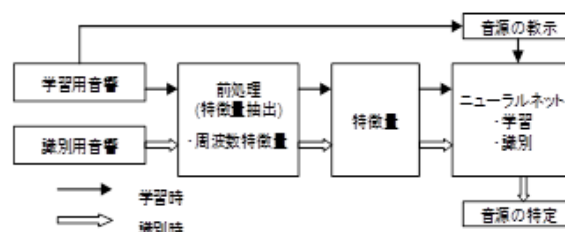


図1 音源識別システム

本研究で使用した NN は入力層・中間層・出力層の 3 層からなる階層型 NN である。学習には「誤差逆伝搬法」を用いた。

出力は 0 から 1 の範囲の数字が出力され、1 に近いほどそのカテゴリである可能性が高いことを示している。

#### 4.2 メル周波数

音声認識で用いているメル周波数を周波数分布に対する新しい特徴量算出の基礎として用いた[4]。

$$mel(x) = \frac{1000}{\ln\left(\frac{1000}{Y} + 1\right)} \ln\left(\frac{x}{Y} + 1\right) \quad (1)$$

$x$  は周波数成分を表している。メル周波数では、低周波数成分が引き伸ばされ、高周波数成分が圧縮される。一般的に  $Y=1000$  とされることが多いが、 $Y$  を変化させ、適切な値を検討する。

#### 4.3 特徴量ベクトル

NN への入力となる特徴量ベクトルの次元を  $N$  とし、次のようにして特徴量ベクトルを求める。  
① 周波数範囲を  $N$  等分する。② 各等分点 ( $n=0, \dots, N$ ) の周波数に対応した高調波の次数  $k(n)$  を求める。③ 次数  $k(n)$  と  $k(n+1)$  の間で FFT で得られた周波数を平均化する。④ 平均化された  $N$  個の値を特徴量ベクトルとする。ただし、②の際にメル周波数を用いてより低周波に重みを持た

せた分割を行う。

さらに、特徴量ベクトルはNNの入力とする際に、各成分の大きさの自然対数、常用対数、シグモイド関数を用いた補正を検討する。

#### 4.4 カテゴリの識別

各音響データの区間抽出については区間音圧が最大となる区間から続く4区間の計5区間抽出し、区間で抽出した特徴量を用いてNNの学習および識別を行う。

1つの音響データにつき5区間で識別を行うため、得られた5個の識別結果をもとにその音響のカテゴリを判定する。今回の実験ではそれぞれの識別結果を足しこみ、もっとも高い数値が得られた出力ユニットに属する音源をNNの特定した音源としてみなすようにした。

### 5. 実験

#### 5.1 実験方法

今回の実験では、前処理で抽出した特徴量をNNの入力に用い、さらには学習回数を変化させることによって、識別率が最適となる回数などを検討する。検証方法として、4fold交差検証を用いる。式(1)のYの値を700, 1000, 2000のそれぞれの場合で実験を行った。なお、NNのパラメータとして、中間層を32、学習係数を0.2とした。

#### 5.2 実験結果

$N=16$ の場合の識別精度を表1に示す。いずれも交差検証を100回行った結果の平均識別精度である。表中のFFTはメル周波数を用いずに等分した場合の結果である。全体的に学習回数が少ない場合に、メル周波数を用いることで精度が向上している。補正方法ではシグモイド関数を用いることで効果は見られなかった。特に、 $Y=1000$ 、常用対数を用いた場合最も良い結果が得られた。

#### 5.3 改良の検討

上記の対数を用いた補正では、特徴量に負値を持つ場合がある。特に、0に近い値を補正した場合、絶対値が非常に大きな負値となり、強調されてしまう。そこで、補正前の値に1を足すことで、正值のみの特徴量へと補正した。実験結果を表2に示す。

表2より少し改善が見られる場合もあるが、全体として効果が見られなかった。

### 6. おわりに

環境音の識別を目的として、NNの入力となる音源の特徴量を検討した。メル周波数を利用することで、少しではあるが識別精度が向上した。今後さらに特徴量の検討を進める予定である。

#### 参考文献

- [1] 警察庁交通局：平成24年度中の交通事故死者数について
- [2] P. Khunarsal : "Very short time environmental sound classification based on spectrogram pattern matching", Information Sciences, Vol.243, pp.57-74 (2013)
- [3] Michael Cowling, Renate Sitte: "Comparison of techniques for environmental sound recognition", Pattern Recognition Letters, 24, pp.2895-2907 (2003)
- [4] P.H. リンゼイ, D.A. ノーマン: 情報処理心理学入門 I, サイエンス社 (1983)

表1 識別精度の比較

| 学習回数 | 補正方法 | 識別精度(%) |          |             |         |
|------|------|---------|----------|-------------|---------|
|      |      | FFT     | Mel-2000 | Mel-1000    | Mel-700 |
| 500  | A    | 44.2    | 65.8     | 67.6        | 61.6    |
|      | B    | 56.3    | 59.1     | 61.0        | 56.9    |
|      | C    | 42.8    | 45.3     | 51.5        | 53.3    |
| 1500 | A    | 61.0    | 68.1     | <b>69.4</b> | 64.1    |
|      | B    | 61.8    | 64.8     | <b>69.0</b> | 63.4    |
|      | C    | 46.1    | 48.0     | 53.1        | 54.7    |
| 2500 | A    | 62.6    | 66.7     | 68.8        | 63.9    |
|      | B    | 60.5    | 65.8     | 68.7        | 64.2    |
|      | C    | 47.1    | 47.8     | 53.2        | 53.3    |
| 5000 | A    | 62.2    | 65.9     | 67.3        | 62.8    |
|      | B    | 62.0    | 66.1     | 69.0        | 64.3    |
|      | C    | 48.8    | 48.7     | 52.8        | 52.3    |

A: 常用対数 B: 自然対数 C: シグモイド

表2 正值のみの特徴量を用いた結果

| 学習回数 | 補正方法 | 識別精度(%) |          |             |         |
|------|------|---------|----------|-------------|---------|
|      |      | FFT     | Mel-2000 | Mel-1000    | Mel-700 |
| 500  | A    | 43.1    | 60.5     | 61.4        | 56.2    |
|      | B    | 51.1    | 62.6     | 64.6        | 58.3    |
| 1500 | A    | 60.2    | 67.5     | <b>68.4</b> | 62.2    |
|      | B    | 60.2    | 65.9     | <b>67.9</b> | 61.8    |
| 2500 | A    | 62.9    | 66.2     | 68.6        | 62.2    |
|      | B    | 59.9    | 65.2     | 68.2        | 61.7    |
| 5000 | A    | 62.4    | 64.8     | 67.3        | 61.5    |
|      | B    | 60.6    | 64.5     | 67.5        | 60.8    |

A: 常用対数 B: 自然対数