

G_002

音声情報によるニューラルネットワークを用いた夜行性野鳥の識別

Nocturnal bird species identification by their calls' information using neural networks

東谷 幸治†
Koji Higashitani三田 長久‡
Nagahisa Mita牧野 洋平‡
Yohei Makino

1. 研究背景

近年、急激な技術革新に伴い、環境問題が深刻になっている。それにつれて、環境評価の研究が盛んに行われている。本研究では、環境の変化に敏感な野鳥について、鳴き声を使って種類識別を行っている。鳴き声を使っているのは、山の中などでは、野鳥が見えない時が多く、画像で識別するよりも活躍の場が多いと考えられるためである。実際、環境省が実施した鳥類繁殖分布調査（平成16年）などでは、夜行性鳥類が著しく減少している可能性が示唆されているが、夜は調査員に負担がかかり、また、目で判断することが困難なため、保護対策を講じることが難しい。そこで本研究では、対象を夜行性野鳥とし、鳴き声の識別、及び、長時間録音データから鳴き声の部分の抜き出しを行う。

2. ニューラルネットワーク

ニューラルネットワークとは、ある入力値が入力層に入力されたとき、所望の出力が出力されるように、入力層から中間層、中間層から出力層への重みを学習させ、それを利用して、識別を行うものである。今回は、3層のニューラルネットワークを利用し、誤差逆伝播法で学習を行った。入力ユニットは32、出力ユニットは野鳥の種類である12種類で行っている。ニューラルネットワークの構成図を図1に示す。

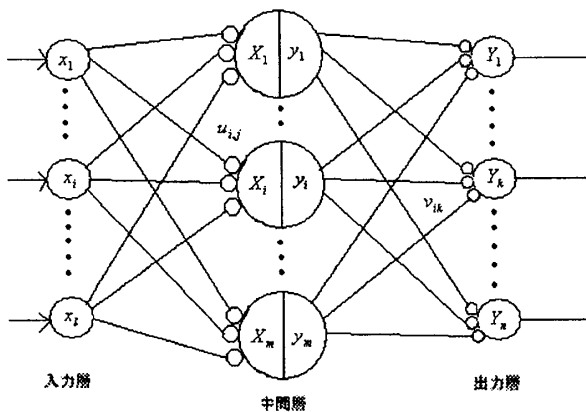


図1 3層ニューラルネットワークの構成図

ニューラルネットワークにおいて重要なのは、学習係数（重みの一回の修正量）や中間ユニット数を経験的に決めなくてはならないところである。数を増やすとかわかって識別率が低下したり、学習に膨大な時間がかかったりしてしまう。しかし、今回対象となる野鳥は夜行性の鳥であり、地域などにより数が限定される。そのため、学習時間に膨大な時間がかかっても、一度学習が終われば、短時間での識別が可能であると考え、ニューラルネットワークによって識別を行うことにした。

3. 周波数帯域パワーの計算方法

今回、主に使用しているパラメータは周波数帯域パワーである。ここで、その計算方法を説明する。

周波数帯域に変換された音声データを $X(p)$ と表し、周波数 $f_n \sim f_{n+1}$ の領域に含まれるパワーを P_n とすると次式で周波数帯域パワーを求める。

$$P_n = \sum_{p=f_n}^{f_{n+1}} |X(p)|^2 \quad (1)$$

ここで重要になるのが周波数の分割の方法である。夜行性の野鳥は低い周波数領域で鳴く野鳥が多いため、低い周波数の分割数を大きく、高い周波数の分割数を少なくした。決定した分割を示す。

0.25~1[kHz]	0.25[kHz]間隔
1~3[kHz]	0.4[kHz]間隔
3~5[kHz]	0.5[kHz]間隔
5~8[kHz]は、	0.75[kHz]間隔

また、野鳥の声の高さの個体差に対応するため、周波数帯域は方形窓にはせず、図2による窓関数をかけて(1)を計算した。

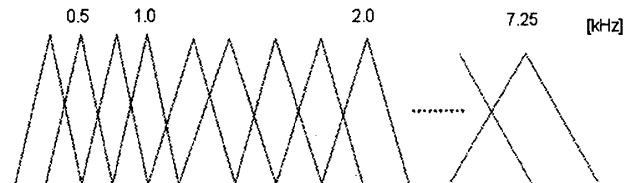


図2 周波数帯域パワーを求める時に使用した窓関数の図

† 熊本大学大学院 自然科学研究科 電気システム専攻

‡ 熊本大学大学院 自然科学研究科 情報電気電子工学専攻

4. 周波数帯域パワーの計算方法

ニューラルネットワークの入力データを野鳥の鳴き声か作る方法を説明する。流れを以下に示す。

- ① 野鳥の音声データ 1秒を準備する。
- ② ハミングウィンドウ (4096 ポイントを半分ずつずらす) でフーリエ変換
- ③ ②のウィンドウごとに1つのデータ、計 20 個のデータを得ることができる。それを前章で説明した方法で周波数帯域パワーを求める。
- ④ 20 個のウィンドウからそれぞれ③の周波数帯域パワー16 個が求まるので、この平均と標準偏差をそれぞれ計算する
- ⑤ それぞれ正規化を行い、計 32 個 (平均 16, 標準偏差 16) を入力データとする。なお、平均は、最大値で正規化を行い、標準偏差は平均を求める前の周波数帯域パワー (13×20) の中の最大値を求め、それで全てを正規化したあと、標準偏差を求める。

この操作をすべての野鳥の鳴き声で行うことにより、ニューラルネットワークに入力するデータを作成する。ここで、標準偏差を用いた理由は、平均の場合、鳴き声の持続時間の影響がデータに投影されないと考えたからである。そのため、標準偏差を用いて、各周波数帯域パワーの平均との差を測ることにより、持続時間の変化にも対応する入力データを作成できると考えた。周波数帯域の最初のピークを 0.25[kHz]からにしているのは、雑音は低い周波数に多くみられる傾向があるため、その対策として低周波を抑圧するためである。

5. シミュレーション

シミュレーション条件について述べる。今回は、12 種類の野鳥の識別を行った。対象は、アオバズク、アカハラ、オオコノハズク、コノハズク、トラツグミ、トラフズク、フクロウ、ホトトギス、マミジロ、ミゾゴイ、ヤマシギ、ヨタカの 12 種類の夜行性鳥類である。音源は、サンプリング周波数 44.1[kHz]の CD 音源を使用している。学習方法は Leave-one-out 法を用いた。この学習方法の詳細は、各野鳥 3 種類の個体から 5 つを学習に使用、残り 1 つのデータをテストデータとして、識別を繰り返した。ニューラルネットワークのパラメータを以下に示す

学習データ	60 個
中間ユニット	60
学習係数	0.05
学習回数	20000

今回中間ユニットを 60 としたのは、いくつか試したところ最もよい数字となったためである。しかし、中間ユニットによる識別率の変動はほとんどなかった。

6. 結果

前章のシミュレーション条件による結果を以下に示す。

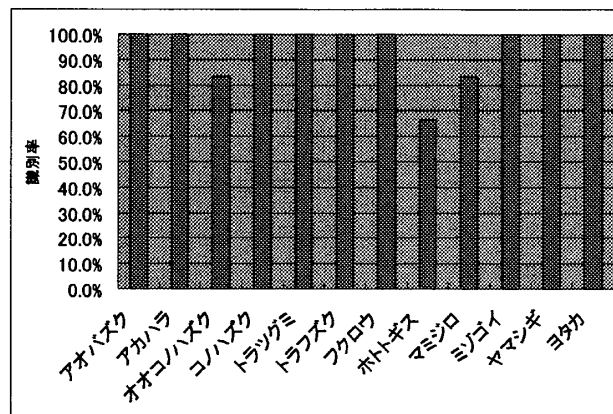


図3 識別対象種類毎の識別率

学習時間は 5 分 40 秒前後だった。(Pentium4 HT 3.0GHz) なお、ニューラルネットワークのユニットの値が最大になったユニットを識別結果とするが、閾値を 0.5 とし、全てのユニットがそれ以下になった場合は結果なしとした。

全体の識別率は 94.29%で、高い識別率を得ることができた。誤識別の検討だが、誤識別している野鳥は鳴き方のバリエーションがあり、学習のサンプル数が十分でないと推測される。そのため、別の周波数帯域にピークが現れることもあるので、例えばテストデータにおいて、周波数を少し上下に振らすといった対策が考えられる。

6. まとめ

今回提案した周波数帯域パワーを用いたニューラルネットワークによる識別は、夜行性鳥類において高い識別率を得ることができ、有用性が確認できた。今回は地域性を考慮していなかったが、同じ品種でも地域により鳴き方が異なったりするため、今後はサンプルを増やし、地域ごとに識別器を構成する必要がある。また、実用化を目指し、長時間録音データから音声部分を抜き出す必要があるため、今後は雑音環境下における抜き出しの検討を行う。

なお、本研究は一部を環境技術開発等推進費の補助を受けて実施したものである。