

音響信号における多重解像度 SOM に関する研究

学籍番号 23413525 氏名 烏田 光一郎

指導教員名 黒柳 奨

1 はじめに

近年、ニューラルネットワークを用いた音声、音響信号に関する研究が盛んに行われており、本研究室でもニューラルネットワークによるパターン認識手法を用いた研究を行っている。一般的に音響信号によるパターン認識手法においては、周波数スペクトル情報を特徴ベクトルとして与えることが多いが、構造的特徴はスペクトル化することで失われてしまう問題がある。ところで画像処理において多重解像度表現という手法がある。多重解像度表現とは、入力画像に Harr-Wavelet 変換を施すことで、原画像を段階的な解像度画像に分解し、解像度毎に別々の SOM へ与えることにより、低解像度画像を与えた SOM では画像の大域的な特徴、高解像度では部分的な特徴を学習することができる。そこで、スペクトログラムにおいて画像処理における多重解像度表現を用いて、段階的な解像度で表現することで、スペクトルパターンの「構造」を把握できるようになり、先ほど述べた救急車のサイレン音やカッコウの鳴き声など、周波数パターンが変化する音種の学習・識別がこの手法で実現できると考え、音響信号における多重解像度表現を用いた SOM の実装及び検証を本研究では行う。

2 画像処理における多重解像度表現

東京農工大の秋元らの「自己組織化マップによる多重解像度マップのクラスタ関係の解析」[1]より画像処理における多重解像度表現について説明する。画像処理における多重解像度表現は、原画像に Haar-Wavelet 変換を施し、生成画像を更に Haar-Wavelet 変換することで多重解像度表現を実現する。画像における Haar-Wavelet 変換は座標方向の輝度における二組の正規直交関数列によって変換される。一つは低周波数成分を出す Scaling 関数と、もう一つは高周波数成分を出す Wavelet 関数である。1 回の変換を行うと画像の縦横方向と Scaling 関数、Wavelet 関数の二関数の組み合わせで濃淡画像 1 枚(LL), テ

クスチャ画像 3 枚(HL, LH, HH)に画像が出力され、計 4 枚の画像が出力される。1 回の変換レベルを圧縮レベル 1 とし、繰り返し濃淡画像を Haar-Wavelet 変換することによって、階層的な各解像度の分解画像の和として原画像を表現できる。圧縮レベルが高くなるほど画像が低解像度となり、低いほど高解像度となる。すなわち、高圧縮レベル画像を学習した SOM は、低解像度の画像を学習することで大域的特徴の学習、低圧縮レベル画像を学習した SOM は、高解像度の画像を学習することで部分的特徴の学習が行われる。

3 音響信号における多重解像度 SOM(提案手法)

音響信号における多重解像度 SOM の学習の概要を図 1 に示す。

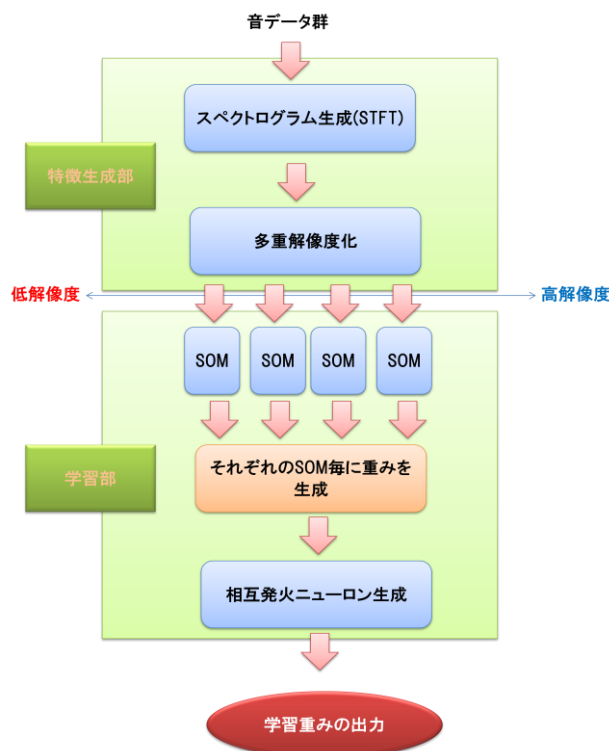


図 1 : 音響信号における多重解像度 SOM

表 1 : 実験結果

SOMニューロン数	平滑化あり	平滑化なし
4×4	95%	100%
5×5	95%	100%
6×6	95%	95%
7×7	80%	85%
8×8	95%	72.5%

SOM ニューロンの数により識別率が大きく変化している。これは、相互発火ニューロンが各解像度の SOM のニューロン番号を学習しているためであると考えられる。近い入力を与えても小さなノイズなどの要因で別の近傍ニューロンが発火し、相互発火ニューロンが異なるパターンになってしまうからであると考えられる。解決策としては、ある程度近い値を持つ近傍ニューロン同士を自動的に結合するという方法などが挙げられる。

5 まとめ

一般的な音響信号のパターン認識は、周波数スペクトルを特徴ベクトルとして用いるものが多く、音響信号の構造的特徴が失われる問題があった。そこで、画像処理における多重解像度表現を音響信号に適用し、学習識別実験を行った。SOM ニューロンの数や SOM の重みを平滑化することで識別率の向上ができた。しかし、似た特徴の音データでも微小な差により相互発火ニューロンが異なる問題があり、細かなパラメータ調整が必要という問題がある。今後の課題として、自動でパラメータを最適な値に設定するような機構の導入、また、音種を増やした実験を行う必要がある。

6 参考文献

- [1] 秋元 信二, 関根 優年 : “自己組織化マップによる多重解像度マップのクラスタ関係の解析”, 日本神経回路学会 2003 年 9 月

まず、学習する音データ群を特徴生成部でスペクトログラム化する。そして生成されたスペクトログラム(長さ X)に対し、 $1/2X$, $1/4X$, $1/8X$, …の長さの矩形窓を用意し、窓をスペクトログラムの時間軸方向に対し、1 点ずつシフトしながら窓内の値を平均化し、移動平均をとる。窓長毎に走査を行うことで長い窓では低解像度、短い窓では高解像度のスペクトルが出力される。これらが、画像処理における Haar-Wavelet 変換の役割を果たすことができると考えられる。学習は SOM を用い、解像度毎に別々の SOM で学習を行う。学習で得られた解像度毎の重みの発火情報を元に相互発火ニューロンを生成する。相互発火ニューロンは、時系列的に対応する入力を各解像度毎の SOM に与え、発火した SOM ニューロンに応じて対応した相互発火ニューロンが発火する。

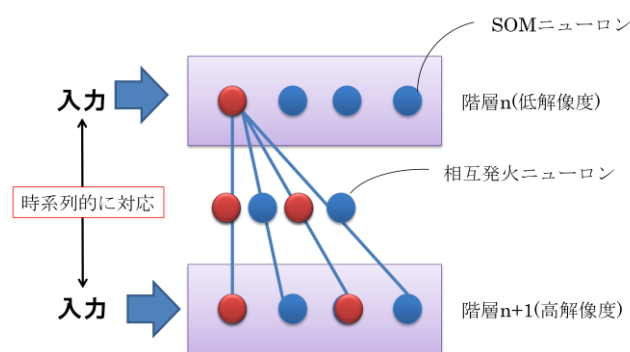


図 2 : 相互発火ニューロン

識別については、音データを学習時と同じように特徴生成部にて特徴生成を行う。その後、SOM の重み群を用いて発火パターンベクトルを生成する。そして、発火パターンベクトルと相互発火ニューロンを k-nearest neighbor を用いて識別を行う。

4 提案手法による実験

提案手法による学習識別実験を行った。学習識別に使用した音は救急車のサイレン、カッコウの鳴き声、純音 1000Hz, 500Hz の 4 種である。SOM は 2 次元 SOM を用い、ニューロン数 $4 \times 4 \sim 8 \times 8$ 、及びニューロンの平滑化あるなしに関して全ての組み合わせで学習識別実験を行った。実験結果を表 1 に示す。