

細分連鎖ニューラルネットワークによるトーナメント形認識システムによる楽器同定

Instruments Identification of Tournament Type Recognition System with Subdivided Chain Neural Networks.

F-19

渡辺 泰誠[†] 秋山 稔[†]Taisei WATANABE[†], and Minoru AKIYAMA[†]

1. まえがき

近年の様々な音楽需要により、演奏された音楽から譜面を起こす作業等が早急に行われる必要性が増加している。しかし、大抵の場合において、人間が採譜を行っているのが現状である。人間が採譜を行うには、絶対または相対音感、経験、知識等が必要であり、これらは一朝一夕に身に付くものではない。そこで最近では、自動採譜に関する研究が盛んに行われている^[1]。自動採譜では演奏されているすべての楽器を一つの譜面上に書き記す事は比較的容易であるが、楽器別に採譜を行う事は困難である。ゆえに、様々な方向から検討がなされている^[2]。本論文では細分連鎖ニューラルネットワークを用いたトーナメント形認識システム(TTRS)を提案し、実験によりその有効性を確認する。

2. システム概要

本認識システムは3種以上の複数楽器の楽器認識問題を、認識の最小単位である2楽器毎に細分化し、細分化したネットワークを結合し、認識を行う。用いるニューラルネットワークは、無フィードバックの単純な形態を使用し、これらのニューラルネットワークをトーナメント形に組み上げ、ニューラルネットワークの出力結果を次のニューラルネットワークへ反映させることにより、楽器認識を行うシステムである。本システムでは、音階毎に学習させた2楽器単位のニューラルネットワークを用意し、トーナメントの下位からの認識状況に応じ、適宜ニューラルネットワークのウエイトデータを入れ替えることにより楽器認識を行う。

本手法では、組み合わせ数も総当たりではなく、トーナメント制のため比較回数が少ない。また、複雑なニューラルネットワークではないため、学習が容易に行えることが期待される。一つのネットワークに全ての認識対象楽器を学習させる手法では、楽器追加の度に学習結果を破棄する事になるが、本手法では新たな楽器と、既存の楽器の組み合わせを学習したウエイトデータを用意する事により楽器追加を容易に行うことが可能である。また、音程毎にニューラルネットワークを用意するため、音程によるスペクトルの差違を吸収することが可能となり、認識率の向上も期待される。

3. 構成要素

本認識システムは、周波数同定部、倍音構成処理部、音源同定部の3部より構成される。

3.1 周波数同定部

周波数同定部には、複素スペクトル内挿法を用いた。複素スペクトル内挿法は、FFT後の音素のある前後で位相が反転することと、ベクトルの絶対値の逆数が一次直線上に並ぶことを利用して同定する手法である。

3.2 倍音構成処理部

倍音構成処理部では、楽器音の倍音構成が調波構造をとることを前提に、FFTにより得られる各周波数毎のスペクトル絶対値をグループ化した楽器の倍音列(以降倍音グループと呼ぶ)を作成する。一般的に6~8次の倍音により楽器の音色は決定されるといわれている。そこで、本検討では8次倍音までを採用する。周波数同定部により判定された周波数成分の存在する点の情報をもとに、低周波数から、高周波数へと基本波を検索する。検出した点を基本波と仮定し、8倍音まで検索を行う。検索した結果、8倍音全ての成分が存在した場合、基本周波数の点に演奏された音程が存在すると見なす。検索した倍音をグループ化し、音源同定部への入力として扱う。

3.3 楽器同定部

楽器同定部では、倍音構成処理部で作成した倍音グループをニューラルネットワークへの入力とし楽器認識を行う。音源同定部には本稿で提案する細分連鎖ニューラルネットワークによるトーナメント形認識システム(TTRS(Tournament Type Recognition System with subdivided chain neural networks))を用いる。

3.3.1 細分連鎖ニューラルネットワークを用いたトーナメント形認識システム(TTRS)

提案するTTRSは、認識の最小単位である2楽器を学習させたウエイトデータを扱うニューラルネットワークをトーナメント形に配置したものである(図1)。一般的にニューラルネットワークの層数決定等は経験と勘に頼る部分が多いといわれているため、一つのネットワークに全てを学習させる方式では、層数決定、中間素子数決定等に大幅な手間と時間が割かれる事となる。そこで、本手法では二種の楽器の判定しか行わないという小規模なニューラルネットワークをトーナメント形に組み合わせる方式を提案する。本方式では、一つのニューラルネットワークに全てを学習させる方式に比べ容易に学習目標に達する可能性が高いと

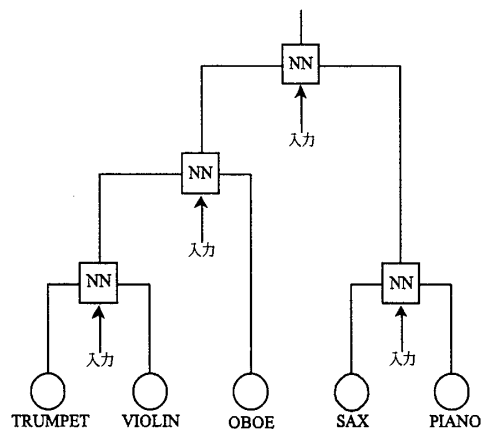


図1 トーナメント形認識システム

[†] 芝浦工業大学 システム工学部 電子情報システム学科

考えられる。ニューラルネットワークの学習にはBP法を用いた。用いたニューラルネットワークはフィードバックの無い基本的で簡素なものである。また、ネットワークを複数有するため、楽器を追加する際に、追加楽器と既存の楽器との組み合わせを学習させる事により追加が可能であるため、全ての楽器を再度、学習させる必要が無く、楽器の追加が容易なことが挙げられる。

3.3.2 TTRSの基本動作

TTRSは、通常のトーナメント戦と同じく楽器を出場チーム、ニューラルネットワークを試合という形に置き換えたものである。上位に進むにつれ、楽器が絞られ、最終的に勝ち残った楽器が演奏された楽器であると認識する方式である。学習された二種の楽器のどちらに近いかを個々のニューラルネットワークが出力し、その出力楽器を組み合わせたウエイトデータを持つニューラルネットワークにより次の認識を行う。この作業をくり返すことにより最終的に残った楽器が認識対象となる楽器であると判断する。トーナメント中、解の楽器を学習時に含まないニューラルネットワークが存在する事となるが、この場合、出力は不定であり、学習を行った二種の楽器のいずれかが出力されることとなり、不定の出力対解の楽器という組み合わせでは、解の楽器が出力される。

4. ベンチマーク 実験

実験は、後述の諸条件により学習させたウエイトデータを用いて実験用のデータを入力し行った。フーリエ観測区間毎に観測区間を移動しながら TTRS に入力を行い、出力結果を一拍毎に統計処理することにより、楽器を推定する。

4.1 諸条件

学習目標値を誤差0.001と設定し、最大学習回数を10000回とした。用いるニューラルネットワークは図2に示した3層(入力層1, 中間層1, 出力層1)、入力素子数8, 出力素子数1, 中間素子数6である。楽器認識結果は興奮性、抑制性の中心点である0.5を境にどちらの領域に属するかにより決定する。学習に用いた楽器は、ピアノ、オーボエ、サクソ、バイオリン、トランペットの5楽器である。また、トーナメントにおける楽器の並び順については、表1に示したとおりである。並びは学習データの距離によって決定した。実験には電子楽器(Roland 社製 SoundCanvasSC-55)を用い、シーケンサを用いてMIDIによって演奏したものをサンプリングし用いる。演奏時のペロシティ値は80固定とした。作成するデータは49音程を音階順に演奏したデータを用いた。サンプリング周波数44.1kHz, 量子化ビット数16である。FFT時の観測区間はlow interval limitを考慮し2048points

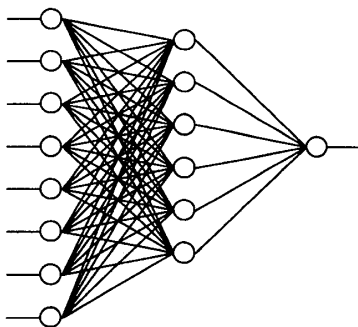


図2 ニューラルネットワーク構成

とした。

4.2 入力データ

学習、想起ともに入力データは0.1~0.9の範囲内に収まるよう一次変換を行う。FFT後の*i*を倍音位置、倍音グループ毎スペクトル絶対値データ列を x_i , 変換後のデータ列を y_i , 倍音グループ毎スペクトル絶対値 x の最大値を x_{max} , 最小値を x_{min} とした場合、式(1)のように表される。この操作はニューラルネットワークで使用している興奮性、抑制性の値と同じ最大値、最小値を入力として与えるために行うものである。

$$y_i = \left(\frac{(x_i - x_{min}) \times 0.8}{x_{max} - x_{min}} \right) + 0.1 \quad (1)$$

4.3 学習

学習データは入力データ、教師信号の組をそれぞれの音程、楽器の組み合わせ数用意し行う。一つのウエイトデータには一音程、一組の楽器(2楽器)を学習させる。よって、本実験で用意するウエイトデータは、 ${}_5C_2 \times 49$ (4オクターブ)=490のウエイトデータを用意する。

4.4 実験結果と考察

実験の一部を表2に示す。一番左端の数字が出力された楽器を示す番号である。表2の例では表1の楽器の対応表からSAXと認識されている。単音による実験では、一音符を最小単位とした場合では、認識率は100%という結果であった。一観測区間毎に正答であるかを判定した場合においても認識率97%という高い認識率を示した。3%は音素の立ち上がり、減衰時での誤同定が大多数であった。

表1 音源対応表

1	trumpet
2	violin
3	oboe
4	sax
5	piano

表2 データ例(一部抜粋)

楽器	周波数	スペクトル絶対値 (8倍音)
4	0.000001383.673	0.3454295417,0.5974201159,0.4691661078,1.0000000000,0.1322963146,0.0720429135,0.0875560977,0.2861574257,
4	0.000001381.107	0.3995817240,0.6066117912,0.3879223073,1.0000000000,0.1605398557,0.0759748164,0.0486386150,0.2639187999,
4	0.000001383.328	0.3472921761,0.5969991834,0.4471056887,1.0000000000,0.1291247692,0.0825383362,0.0877310718,0.2335427823,
4	0.000001382.901	0.3633444622,0.6172396562,0.4412268772,1.0000000000,0.142329951,0.0833375179,0.081144304,0.2149782209,
4	0.000001383.104	0.3547334069,0.6285036752,0.4348169004,1.0000000000,0.1497426682,0.0741239475,0.0493515518,0.2122022922,
4	0.000001383.594	0.3332844497,0.6219010853,0.4366228840,1.0000000000,0.1468369428,0.0750928291,0.0546678437,0.2451265584,

5. むすび

ニューラルネットワークを単体で使用するのではなく、認識楽器数を最小単位に分解し、単純なニューラルネットワークを複数、トーナメント形に配置した、細分連鎖ニューラルネットワークを用いたトーナメント形認識システム(TTRS)による楽器認識について提案し、実験によりその有効性を確認した。本方法を用いることで、単楽器を高精度に認識可能であることが実験により証明された。また、今後の課題としては、同時認識音数の拡張や、減衰音の処理等が挙げられる。

文献

[1]長島 洋一, 橋本 周司, 平賀 謙, 平田 圭二編:音楽とコンピュータの世界,bit8月号別冊,1998.Aug
 [2] Jae-Hoon Jeong,Dou-Suk Kim, and Soo-Young Lee:Musical timber recognition with neural networks..pp.869-872