

演奏情報に関する楽曲の特徴抽出システムの作成

関本 陽子 野池 賢二 野瀬 隆 乾 伸雄 小谷 善行 西村 恕彦
(東京農工大学大学院 工学研究科 電子情報工学専攻)

本研究の目的は、旋律度と演奏情報を計算することによって楽曲の特徴を見つけ出すことである。人間が演奏するときに考えることを解析し、それらを評価要素として表した。それらの評価要素に重みがついた値で特徴量を表すために、線形モデル・非線形モデル(ニューラルネットワーク)を仮定した。

オーケストラ編成の曲において、評価要素の重み、また旋律度とヴェロシティを計算した結果、線形モデルの重みの決定係数は0.6以上であったが、ニューラルネットワークは正確な重みを獲得しなかった。評価要素の改良・多くのデータでの実験・ニューラルネットワークの設定のさらなる調査が必要である。

A Characteristic Extractor of Pieces Related to Performance Expressions

Yoko SEKIMOTO, Kenzi NOIKE, Takashi NOSE,
Nobuo INUI, Yoshiyuki KOTANI, Hirohiko NISIMURA
(Tokyo University of Agric. and Tech., Dept. of Computer Science)

The aim of this paper is to find out the characteristics of pieces by computing the degree of melodies and the performance information. we analyzed the thoughts of the humans for the performance, and expressed them as evaluation elements. we assumed that each element had its weight. There are two methods to compute the weights. The first one is to use a least square method, and the second one is to use a neural network.

As a result, the coefficients of the weight calculated by the least square method came to more than 0.6, whereas the neural network could not obtained accurate weights. Further research is necessary to improve the neural networks.

1 はじめに

表情に関する特徴量を計算する研究は盛んに行われているが、楽曲の文法化や楽典的な解析という方針を取る研究では、楽曲に例外なく対応することが困難である。

本研究では、人間が楽譜から演奏情報を推論する過程を明らかにすることを目的とし、表情に関する特徴量を柔軟に計算するモデルを提案した。そのモデルは、このような特徴量が、楽譜から直接得られるいくつかの情報(評価要素)に重みがついた

値で表せるというものである。モデルには、線形モデルと非線形モデル(ニューラルネットワーク)の二つを仮定している。重みを計算するために、線形モデルには最小二乗法を、ニューラルネットワークにはバックプロパゲーションアルゴリズムを用いた。表情に関する特徴量のうち、主旋律である度合い(旋律度)と演奏情報であるヴェロシティを計算した結果について述べる。

2 人間が楽譜から得る情報

人間は、演奏するために楽譜からどのような情報を得るのであろうか。筆者は段階的に大きく分けて三つの種類の情報があると考えている。

1. 楽譜から直接得られる基本的な情報
これは楽譜上に記述された音高・音価・テンポ・楽語といった基本的な情報のことで、誰が見ても解釈が同じ値である。
2. 基本的な情報から推論することによって得られる情報
1の段階の情報から得られる情報である。フレーズ・旋律の音楽的役割などの音楽的解釈がこれに当たる。この段階以降の情報は、人によって違いがあり、必ずしも同一な値にはならない。
3. 演奏に必要な情報
1と2の段階の情報から得られる情報であり、人が実際に音を出すために必要な情報である。ヴェロシティ・微妙な揺らぎを含んだテンポや音高・楽譜上の音価と必ずしも一致しない音長・音色などである。

3 楽譜から直接得られる基本的な情報

本研究では、楽譜から直接得られる情報のうち、次に述べる項目を音符単位の情報としてまとめた。

1. 休符か休符でないか
その音符が休符であれば1、そうでなければ0という値である。
2. 音高
鍵盤楽器の中央の音 ハ の鍵盤の音高を60とし、半音あがるごとに+1、半音下がるごとに-1とした値である。
3. 音長
楽譜に記述された音価を数値で表す。例えば、四分音符は480、八分音符は240という値である。

4. 付属する強弱記号

楽譜に記述された強弱記号を、個々の音符に対応させる。 f, mf, mp, p がつく音符に対して、それぞれ2, 1, -1, -2と値を持たせる。付属する強弱記号がなければ0である。一度強弱記号がいたら、違う記号がつくまで前の強弱記号の値を持たせる。

5. パート番号

パート番号を評価要素値にする。

4 特徴量の計算モデル

楽譜から推論できる情報(特徴量)のうち、筆者が注目しているのは個々の音符の音楽的役割である。筆者は、特に音楽的役割のうち、主旋律に注目している。そこで、個々の音符(音符番号を i とする)が持つ旋律度という特徴量 y_i を定義し、楽譜から直接得られる情報から計算するモデルを提案する。特徴量 y_i は、楽譜の情報から計算できる p 個の重み $W(=w_1, w_2, \dots, w_p)$ つき評価要素 $x_i(=x_{i1}, x_{i2}, \dots, x_{ip})$ で計算できる。

$$y_i = f(x_i, W) \quad (1)$$

さらに、演奏に必要な情報であるヴェロシティについても、特徴量としてこの計算モデルを用いて計算できるのではないかと考えた。

5 評価要素

3節と、それらから計算できる情報を、評価要素(X)として定義した。

1. 楽譜情報

3節で述べた情報を、そのまま評価要素として利用する。

2. 音高の順位

楽譜上の同じ時刻で他のパートの音高を比較し、その順位を別の評価要素とする。

3. 同じ楽譜上の時刻に同じ音価の音符がいくつあるか

同じ楽譜上の時刻にある他のパートに、音価が等しい音符があるかどうかを調べ、その音符の個数を評価要素とする。

6 重みの計算方法

式(1)において、特徴量 Y を求めるためには、まず重み W を計算する必要がある。重み W を求める方法として、最小二乗法を用いる方法と、バックプロパゲーションを用いる方法の二つの方法で求めることを試みた。

6.1 最小二乗法

式(1)の重みは、次式によって計算できる。

$$W = (X^t X)^{-1} X^t Y \quad (2)$$

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, n: \text{音符数}$$

6.2 ニューラルネットワークにおけるバックプロパゲーションアルゴリズム

ニューラルネットワークとは、神経細胞を手本にした情報処理システムである。いくつかの外部の刺激を入力すると、その情報を神経細胞(ニューロユニット)でやりとりしながら、ある結果を出力として得ることができる。ニューロユニットの中では、

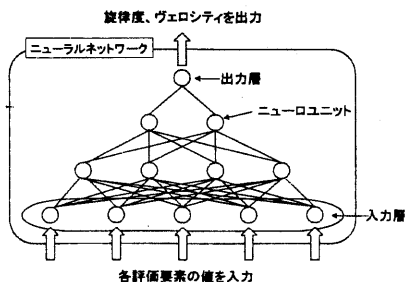


図1 ニューラルネットワークの概要

それぞれの入力に対して非線形変換を行い、次のユ

ニットに出力する。本研究の場合、図1の入力層には評価要素値が入力される。出力として個々の音符の旋律度と音符のヴェロシティを得る。

ニューラルネットワークにおいて、式(1)の重み W を決定するためには、バックプロパゲーションアルゴリズムを用いる。

本研究では、一個の音符の期待出力である教師値 d と、ニューラルネットワークの実際の出力値 y との二乗誤差

$$E = \frac{1}{2}(y - d)^2 \quad (3)$$

を評価関数として、音符の教師値を一つ入力するごとに一回重み修正を行う(逐次重み修正方法)。

7 実験

実際に曲の評価要素値 X とその重み W を計算し、特徴量 Y のうち旋律度とヴェロシティをそれぞれ計算する実験を行った。

7.1 実験に用いた曲

今回実験で使った曲は次の曲である。

- 曲名：亜麻色の髪の乙女
- 作曲者：ドビュッシー(編曲:伊藤辰雄)
- 編成：オーケストラ(14パート)
- 音符数：697個

7.2 重み W を計算するための教師データ

7.2.1 旋律度の教師データ

旋律度の教師データはアンケートで採取した。まず、楽譜を被験者に見せ、表情付きの SMF ファイルによる演奏を試聴してもらい、主旋律だと思っ部分に波線を引いてもらった。教師データとして、波線が引かれている音符を旋律度 1、そうでない音符を旋律度 0 として整理した。なお、被験者は一人である。

7.2.2 ヴェロシティの教師データ

文献[7]と対応して通信販売されている SMF データのヴェロシティ値を教師値として用いた。教師値

は、実際の生の演奏から採取するのが理想であるが、

- 本研究はオーケストラのような大きい編成の曲を対象にしている
- 机上で編集されたデータであっても、人間が表情のある演奏として良い方向に編集したデータである

ということから、今回の実験ではこれを表情がついた教師値データとして利用した。

7.3 ニューラルネットの設定

ニューラルネットの設定は次のとおりである。

- 入力層のユニット数：9
- 隠れ層の数：7
- 隠れ層のユニット数：それぞれ9
- 出力層のユニット数：1
- 学習回数：10000

8 実験結果

8.1 旋律度に関する実験結果

8.1.1 線形モデルによる旋律度の出力値

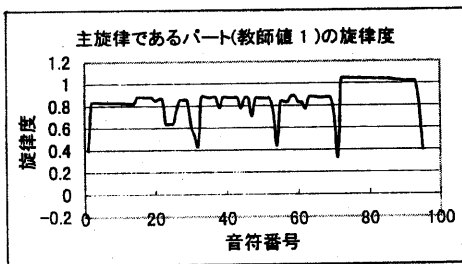


図 2

図 2 に主旋律であるパートの旋律度の出力値を示す。これは、1パート分を示している。横軸は、時間ではなく音符番号であり、曲の流れにしたがっている。

いくつかの音符は、旋律度の出力値が低くなっているが、これは

- もともと被験者が主旋律と判断しなかった長い休符

- 主旋律の合間の短い休符で、被験者が波線を引いた休符
- 同じ楽譜上の時刻に他のパートで同じ音価である音符が存在するために、旋律度が計算された音符である。

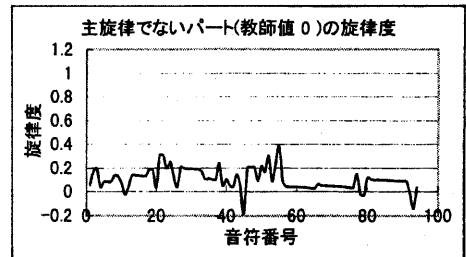


図 3

図 3 に主旋律でないパートの旋律度の出力値を示す。主旋律でない音符の教師値はすべて 0 であるが、図 2 と比較すると、低い値で出力値が計算されているものの、音符によるばらつきが多い。

8.2 ニューラルネットワークによる旋律度

ニューラルネットでは、旋律度の正確な出力値を得るまでの十分な重み W の学習が認められなかった。重みを学習した際の、誤差の収束の様子について述べる。誤差の計算式は式 (3) である。

8.2.1 旋律度の重み学習のための誤差の収束の様子

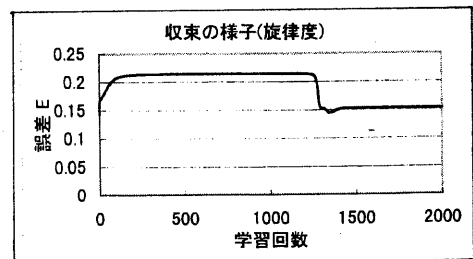


図 4

図 4 に旋律度の重み学習のための誤差の収束の様子 (2000 回まで) を示す。学習回数が 100 回まで

は、誤差が増大する方に学習が進んだが、1200回あたりから誤差が収束し、ほぼ一定の値に落ち着いた。しかし、この値は初期乱数を与えたはじめの誤差とほとんど変わらず、正確な学習とは言い難い。

8.3 ヴェロシティに関する実験結果

8.3.1 線形モデルに関する実験結果

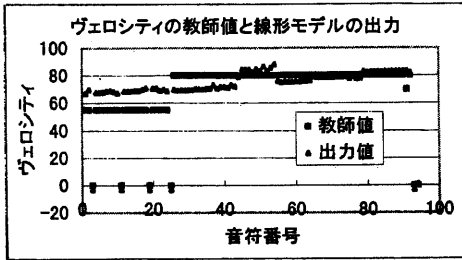


図 5

図 5 にヴェロシティの教師値と出力値を示す。線形モデルの出力値は、音符番号 40 の前後の音符が教師値と離れた値を取っているが、音符番号 20 までの部分と音符番号 50 以降の部分は、かなり教師値に近い値が得られた。

8.3.2 ニューラルネットワークに関する実験結果

誤差の収束の様子について述べる。

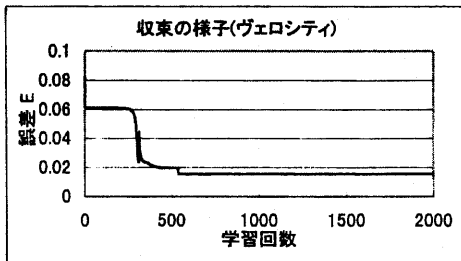


図 6

図 6 にヴェロシティの重み学習のための誤差の収束の様子を示す。学習回数が 250 回のあたりで誤差が大きくなる方法に学習が進んだが、500 回を超えたあたりで収束した。しかし、ニューラルネットによる出力値は教師値を反映しておらず、学習がうまくいっていない。

9 評価

最終的には教師値の特徴を重みで表現することが目標であるが、依然としてそれを達成するにはいたっていない。

9.1 最小二乗法に関する評価

9.1.1 自由度修正済みの決定係数

最小二乗法で計算した出力値がどの程度教師値に近いかという指標を表す数値として、自由度修正済みの決定係数を調べる。これは 0 から 1 の値を取り、1 に近いほど重みの精度が優れていることになる。自由度修正済み決定係数 $R^{2'}$ は次のように表せる。

$$R^{2'} = 1 - \frac{S_E / (n - p - 1)}{S_{yy} / (n - 1)} \quad (4)$$

$$S_E = \sum (d_i - y_i)^2$$

$$S_{yy} = \sum (y_i - \bar{y})^2$$

$$\bar{y} : \dots y_i \text{ の平均値}$$

本実験で得られたそれぞれの重みの $R^{2'}$ は次のとおりである。

- 旋律度に関する重みの $R^{2'}$: 0.631
- ヴェロシティに関する重みの $R^{2'}$: 0.870

この結果から、演奏情報に関する特徴量を計算するモデルとして、本研究で提案したモデルが有効だといえる。

9.1.2 重みの信頼性

最小二乗法は簡単で安定した手法であるが、旋律度やヴェロシティを正確に決定してはいない。特殊な規則を抽出するというより、統計的に有力である評価要素が重要になり、全体として無難な重みが計算されるのが特徴であり、数が少ない特徴が反映されない。

9.1.3 旋律度からどのように主旋律を判断するか

旋律度を求める目的は、個々の音符の音楽的重要性を数値で表し、主旋律を抽出することであるが、

このような実験結果からどのように決定すればよいのだろうか。簡単な考え方では、同じ楽譜上の時刻で音符の旋律度を比較し、一番値が大きい音符を主旋律と見なすという方法が取れる。しかし、この考え方では、次のような教師値が被験者から与えられた場合にうまく判断できない。

- 主旋律がわからない場合
- 主旋律が複数ある場合
- 主旋律がない場合

あるしきい値を超えたものを主旋律と考える方法がある。図2と図3を見比べると、旋律度の取りうる範囲に一定の隔たりを見出せそうである。これは今後の課題である。

9.2 ニューラルネットワークでの重み学習

最小二乗法より正確な重みを計算することを目指してニューラルネットワークを用いたのだが、実際には、ニューラルネットワークの設定が試行錯誤であるため、期待するような結果は得られていない。原因として考えられるのは、入力データとなる評価要素、出力データの値域がそろっていないことである。理論的には、どのような値域を持つ入力データであっても、その値域の違いはニューロンの重みに吸収されるといわれているが、実際には値域が異なると重みの学習は難しくなる。

しかし、最適な重みを学習するためのいくつかの指標が明らかになった。一つは入出力層につながるニューロユニットの数を増やすことである。本研究のような入力データの場合、9個の評価要素の特徴を出力データに吸収させるために、入出力層、特に出力層につながるユニットを増やすことが、正確な学習のために必要であることがわかった。

10 今後の課題

より有効な結果を得るためにさらに調査を進める必要がある。次の項目が今後の課題である。

- 評価要素の改良
- 多くの教師データによる実験
- ニューラルネットワークの設定

11 まとめ

本研究では、表情に関する特徴量を柔軟に計算するモデルを提案した。それは、このような特徴量が、楽譜から直接得られる情報をもとにして得られるいくつかの評価要素に重みがついた値で表せるというモデルである。モデルは、線形モデルと非線形モデル(ニューラルネットワーク)を仮定した。表情に関する特徴量のうち、旋律度とヴェロシティを計算した結果、線形モデルについては有効な結果が得られた。ニューラルネットについては依然として多くの課題が残されている。

参考文献

- [1] Gerhard Widmer, Learning Expression at Multiple Structural Levels, Proceedings of the 1994 International Computer Music Conference, 1994, pp.95-101.
- [2] Gerhard Widmer, Understanding and Learning Musical Expression, Proceedings of the 1993 International Computer Music Conference, 1993, pp.268-275.
- [3] G.U.Battel, R.Bresin, G.De Poli and A.Vidlin, NEURAL NETWORKS VS.RULES SYSTEM: EVALUATION TEST OF AUTOMATIC PERFORMANCE OF MUSICAL SCORES, Proceedings of the 1994 International Computer Music Conference, 1994, pp.109-113.
- [4] Howard Gardner, A Cognitivist Theory of Music Perception, LANGUAGE, MUSIC, and MIND, THE MIT Press, 1993, pp.11-35.
- [5] JUDY EDWORTHY, Melodic Contour and Musical Structure, MUSICAL STRUCTURE AND COGNITION, ACADEMIC PRESS, 1985, pp.169-188.
- [6] ROBERT WEST, PETER HOWELL, AND IAN CROSS, Modelling Perceived Musical Structure, MUSICAL STRUCTURE AND COGNITION, ACADEMIC PRESS, 1985, pp.21-52.
- [7] 伊藤辰雄, DTM スコア・クラシック 1, 全音楽譜出版社, 1996.
- [8] 菅民郎, 多変量解析の実践(上), 現代数学者, 1993.
- [9] 八名和夫, 鈴木義武, ニューロ情報処理技術, 海文堂, 1994.