

楽曲評価におけるニューラルネットワークモデル

村上 英典
京平成大学

m721027@thu.ac.jp

あらまし “楽曲を聴いて評価する” 過程をニューラルネットワークで再現して、“楽曲のイメージ” に与える音楽要素の影響を結合値によって示そうと試みた。階層型ネットワークを採用し、バックプロパゲーションアルゴリズムで学習をおこなう。入力信号は、Standard MIDI File を元に作成した楽曲情報、教師信号は、“簡単なイメージ” に対しての5段階評価とした。楽曲中の音楽要素からの興奮と抑制の作用によってイメージは発生すると仮定し、その作用に反応するニューロンが、音楽要素として考えられるものに対応して中間層を構成する。楽曲の評価を行い学習して、その結果得られた結合について考察する。

キーワード ニューラルネットワーク、イメージ

The Neural Network model in case of music evaluation

Hidenori Murakami
Teikyoheisei University

m721027@thu.ac.jp

Abstract It reproduces The process of “hear and evaluated music” by the neural network and it tried when would show the influence of the musical element to give “the image of the music” by the combination value. It adopts a hierarchy -type network and it learns in back-propagation. The music information that the input signal made Standard MIDI File to the origin, The teacher signal is 5 steps evaluation to “the simple image”. It supposes that the image occurs by the operation of the excitement and the restraint from the musical element in the music, the neuron which react to the operation corresponds to the one which is thought of as the musical element and composes a middle class. It learns the evaluation of the music by the doing and as a result, it considers it about the gotten combination.

key words Neural network, Image

1. はじめに

コンピュータに個人の“音楽的感性”を理解させ、楽曲を入力したとき、個人と同じ“イメージ”を抱かせることが本研究の目的となっている。

音楽から受けるイメージと言っても「夏の暑い日の青い海のような」などと視覚、記憶、皮膚感覚等に表現され、楽曲分析だけではなく、人の五感であるとか、記憶、生体構造に至るまで追求し総合的に考える必要がある[1][2][3]。

楽曲分析1つとっても、楽曲がもつ“音楽要素”の認知・分析[4][5][6][7][8][9]、感情との関連[10][11][12]、等、の総合的なモデルが必要となり、楽曲を聴いている環境も考慮した上で、楽曲聴取からイメージの発生まで全ての過程（人間の仕組み）をモデル化するのは容易でない。

そこで「夏」＝「あつい」のように、発生する事象とその結果をニューラルネットワークで直接的に結びつけ、事象である楽曲を入力情報、結果であるイメージを教師信号（出力値）とした。リズム、旋律、和声といった音楽要素を中間層として学習を行うことにより、音楽要素とイメージの関係つまり、ネットワークの結合値に音楽的感性を見いだそうと考えた。

音楽の諸要素がもつ効果は聴取時の精神状態によって異なってくるが[13]、総合して精神・心理へ影響を与えており[14]、その反応はホメオスタシスに類似する[15][16]。また、聞こえる音だけが影響を与えるのではなく可聴範囲を超えた音にも注目しなければならない。[17]

これら考慮すべき点をふまえ、概念を考える必要がある。楽曲情報として Standard MIDI File（以下 SMF）を使用しているため聴取範囲は限られるが、楽曲から受ける心理への影響は、音楽諸要素の“中心値を境とした生理的な興奮と抑制の作用である。”と考え、基本概念としている。

2 ネットワークの概要

2. 1 入力情報

SMF から1音ずつに小節番号、小節内での発生音番号、テンポ、拍子、小節内での位置、直前音からの時間（秒）、発音時間（秒）、音符時間（秒）、上位・拍子・下位のリズム判定、拍節の判定、発生音数、ノート番号、ON ペロシティ、OFF ペロシティの情報を算出、抽出した情報ファイルを作成し入力情報とした。[18]

2. 2 教師信号と出力値

明るい、暗い、軽快、重々しい、落ち着き、不安の6つのイメージそれぞれに0から5段階での評価を行う。（0：評価しない、1：少し～5：すごく）

ネットワークによる評価の際は「出力範囲」に対応する「評価」を出力し、学習の際は「評価」に対応する「教師データ」を教師信号とする。

評価	出力範囲	教師データ
0	0	0.0
1	0より大 0.2以下	0.1
2	0.2より大 0.4以下	0.3
3	0.4より大 0.6以下	0.5
4	0.6より大 0.8以下	0.7
5	0.8より大	0.9

2. 3 ネットワークの構造

階層型ニューラルネットワークで学習に、バックプロパゲーションアルゴリズムを使用している[19]。

中間層の第4層まで横方向に、リズム、旋律、和声の3グループに分かれ、中間層第1層、第2層は、発生音、動機、フレーズ、センテンスの時間区間に分かれる。矢印の方向にニューロン出力が伝播され、入力層以外の円で表されるその内部には興奮作用と抑制作用における音楽要素と方向性のニューロンを持ち、計4ニューロンが次の層の円内部とそれぞれ結合している。図1

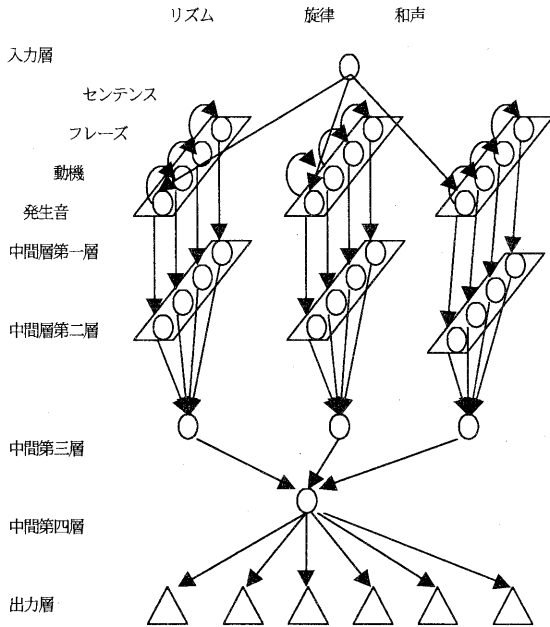


図1 ネットワークの構造

2. 3. 1 入力層

入力層のニューロンは、音のもつ要素のみで反応するグループと、直前の音と比較を行い、その方向性の結果によって反応を起こすグループに分かれる。また、入力情報から音楽理論[20][21]による判定で反応するものと（表1の#記号）、個人情報（図2）によってテンポ（60sec/テンポ）、アーティキュレーション（スタッカート0～100%レガート）、音高（低い0～127高い）、和音数（1～）、音量、の各要素の最大値、最小値、中心値を元に作成したシグモイド関数（図3、図4）によって、入力情報に反応するものがある。

前者である（表1）で#記号がついているニューロンは、入力情報の該当する要因が興奮であれば+1.0を出力、抑制であれば-1.0を出力し、該当しなければ0.0を出力する。

後者の#記号がついていないニューロンにおいて、音要素の反応は、中心値～最大値または最小値～中心値で作成したシグモイド関数（図3）で出力され入力情報の要因が中心値を境に、興奮要因か、抑制要因（表1）かで、興奮関数（図3左側）か抑制関数（図3右側）、

どちらかに決定する。興奮方向であれば+符号、抑制方向であれば-符号で出力される。

方向性グループは興奮、抑制方向の要因、どちらに変化したか（表1）で符号が変わり、変化の大きさは、個人情報の最小値以下を0、最小値～最大値により作成したシグモイド関数（図4）により出力される。興奮方向（図4右側）であれば+符号、抑制方向（図4左側）であれば-符号で出力される。

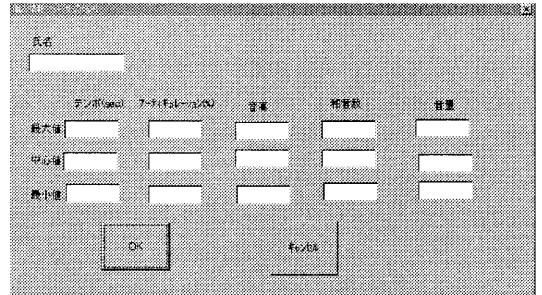


図2 個人情報の入力画面

表1 入力情報とその作用要因

	ニューロン	興奮要因	抑制要因
音要素	—R—		
	テンポ (音符時間)	早い (短い)	遅い (長い)
	リズム #	下位リズム	上位リズム
	拍 #	強拍	弱拍
	アーティキュレーション	短い (カット)	長い (スタッカート)
	—M—		
	音高		
	—H—		
	和音数	高い	低い
	和音協和 #	累積小協和音 長調	累積大不協和音 単調
方向性	—R—		
	テンポ	早くなる	遅くなる
	リズム #	細くなる	大きくなる
	拍 #	強拍になる	弱拍になる
	アーティキュレーション	短くなる	長くなる
	—M—		
	音高	高くなる	低くなる
	音程 #	協和音程	不協和音程
—H—			
和音数の変化	少なくなる	多くなる	
和音	短から長調	長から短調	
和音の解決 #	解決	未解決	

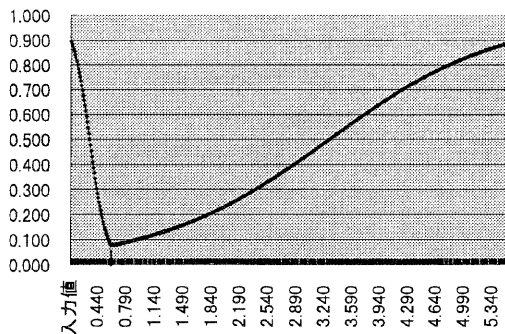


図3 テンポ（音符時間）音要素での関数。

最大値（6.0）、中心値（0.6）、最小値（0.01）

中心値～最小値：（興奮）、中心値～最大値：（抑制）

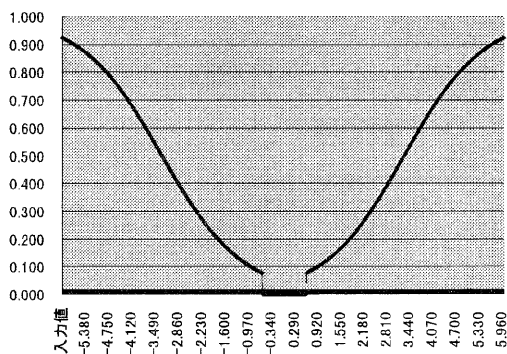


図4 テンポ（音符時間）の比較での関数。

最大値（6.0）、最小値（0.01）

（-）：遅くなる、（+）：早くなる

これら17個のニューロンは、入力情報からの1音ごとに反応を示し、図1のようにリズム、旋律、和声のグループに分れ次の層の第一列目に出力する。

ニューロンは音ごとに反応を示すため1音単位の時間軸に支配されている。ネットワークの中間層には、楽式論的時間区間による、動機単位やフレーズ単位といった特徴抽出が行われるため一定の時間単位と時間軸が存在しない。また出力層からは曲を聴き終わった時点での評価が教師信号として送られており、この入力層での時間軸、時間単位（1音単位）に当てはまらないので、この層のニューロン出力に対する結合の学習は行われない。

2.3.2 中間層第1層

第1列目

入力層からの信号が（表1）のリズム（R）、旋律（M）、和声（H）別に興奮・抑制と要素・方向で区別されている4つのブロックごとに値を合計する。

入力信号はどれも最大が1、0であり各ブロックの最大値はその入力信号の個数となる。それを最大値としたシグモイド関数にブロック合計を入力した値がこの列以降で基本となるニューロングループ（図1、円表示）内の各ニューロン出力となり、第1層第2列目と第2層第一列目へ伝播する（図1）。

ニューロングループ

興奮要素ニューロン	興奮方向ニューロン
抑制要素ニューロン	抑制方向ニューロン

第1列目以降

第2列：動機、第3列：フレーズ、第4列：センテンス（図1）は、（図5）で指定された時間区間で反応する。（図6）

各列は、時間区間内にある前列からの、ニューロングループの各ニューロン出力を合計する。

要素と方向は、時間区間内では1つの興奮か抑制反応を引き起こす要素にすぎないと考え、方向ニューロンを作用別の要素ニューロンに加え、要素ニューロンとする。（図6実線、図7上）

その列の方向ニューロン出力は、時間区間内での前列からの入力を比較しその作用別に合計して方向性ニューロンとする。（図6点線、図7下）

各ニューロンは入力回数を最大値としたシグモイド関数で各ニューロンの出力を決定する。

出力は真後ろの列とその真下の層へと伝播していく。入力層と同じく、時間単位が一定でないのでバックパケーションによる結合値の学習は行われない。

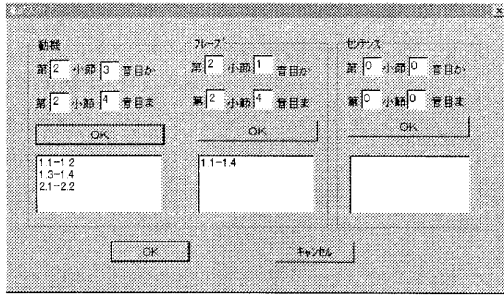


図5 時間区間の入力画面

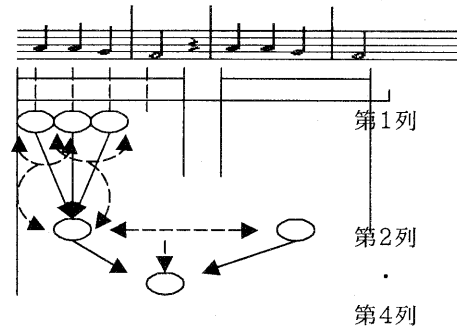


図6 時間区間(列)ごとの反応

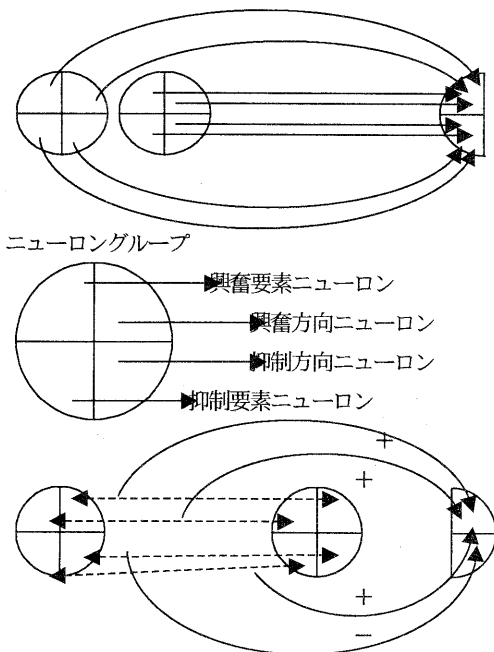


図7 第二層におけるニューロングループ構成

2. 3. 3 中間層第2層

中間層第1層と構造は同じく、横方向にリズム、旋律、和声の3グループに分かれ、縦方向には、発生した音全ての入力を受ける発生音と時間区間で分かれる動機、フレーズ、センテンス、それぞれに基本ニューロングループが存在する。(図1)

中間層第1層からの入力を受け、各ニューロンごとに合計された値は入力回数を最大値としたシグモイド関数でその出力を決定する。

時間区間で分れていたものが、次の第3層でリズム、旋律、和声のニューロングループに収束される。

この第2層、第3層間の結合からバックプロパゲーションによる学習を行う。

この第2層-第3層間の結合値に“楽曲のイメージ”への“各時間区間が及ぼす影響力”を示せると考えた。

2. 3. 4 中間層第3層

リズム、旋律、和声、の3ニューロングループで構成しており、中間層第2層から同種のニューロングループごとに入力を受ける。(図1)

各ニューロンごとに合計した値は入力して来るニューロングループの個数(4, 0)を最大値としたシグモイド関数でその出力を決定する。

音楽3要素で分れていたものが、次の第4層で、一つのニューロングループに収束される。

この第3層-第4層間の結合値に“楽曲のイメージ”への“音楽3要素が及ぼす影響力”を示せると考えた。

2. 3. 5 中間層第4層

1つニューロングループで構成しており、中間層第3層の各ニューロングループから、同種のニューロンごとに入力を受ける。

各ニューロンごとに合計した値は入力して来るニューロングループの個数(3, 0)を最大値としたシグモイド関数でその出力を決定する。

興奮・抑制作用の要素・方向性の4ニューロンが出力層へつながり、その結合にイメージへの影響力を示せると考えた。

2. 3. 6 出力層

各ニューロンにつながる中間層第四層のニューロンの個数(4. 0)を最大値としたシグモイド関数でその出力値を決定し、2. 2の“出力範囲”に当てはまる“評価”を出力する。

学習の際は“評価”に当てはまる“教師データ”を教師信号とする。

3. 実験

個人情報:

	最大値	中心値	最小値
テンポ	1.0	0.6	0.25
アーティキュレーション	100	60	10
音高	127	60	1
和音数	5	3	1

曲名:「かぜにそよぐ木」(図8) temp 1 2 0

評価: 明るい(4) 暗い(1) 軽快(3)
重々しい(1) 落ち着き(1) 不安(1)

動機: 1小節目第1音 ~ 2小節目第1音
3小節目第1音 ~ 4小節目第1音
5小節目第1音 ~ 8小節目第1音
9小節目第1音 ~ 10小節目第1音
11小節目第1音 ~ 12小節目第1音
13小節目第1音 ~ 16小節目第1音
リズム: 1小節目第1音 ~ 8小節目第1音
9小節目第1音 ~ 16小節目第1音

セテス: なし

で学習をおこなった。

ネットワーク全結合の初期値(1. 0)

学習回数n、出力層ニューロン出力と教師信号との誤差和を E_n (図9) としたとき

$E_{n-1} - E_n < |0.00001|$ (図10)

で学習を終了するようにプログラムした。

139回目の学習でその条件となり終了した。



図8 「かぜにそよぐ木」

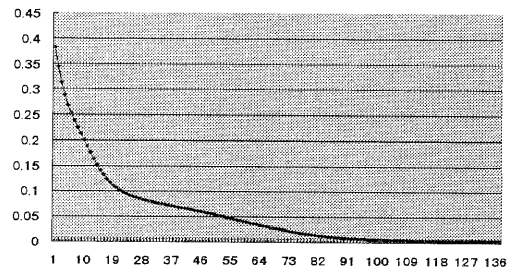


図9 学習回数nと誤差和 E_n

横軸: 学習回数n、縦軸: 誤差和 E_n

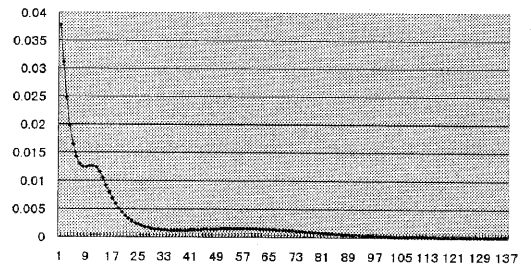


図10 学習回数と誤差和変動

横軸: 学習回数n、縦軸: $|E_{n-1} - E_n|$

表2 中間層第3層と第4層の結合

	興奮要素	興奮方向	抑制要素	抑制方向
リズム	1.24613	2.44628	1.97095	1.83308
旋律	1.22806	2.59178	1.97229	1.74894
和声	1.22526	2.58815	1.28884	1.74330

表3 中間層第2層と中間層第3層の結合

	興奮要素	興奮方向	抑制要素	抑制方向
———第1列———				
リズム	1.01009	1.04646	1.08663	1.09214
旋律	1.01070	1.43648	1.00311	1.05549
和声	1.25595	1.03821	1.00204	1.01311
———第2列———				
リズム	1.00956	1.04111	1.07788	1.13229
旋律	1.00994	1.04202	1.04407	1.05413
和声	1.15037	1.03780	1.00201	1.04209
———第3列———				
リズム	1.00944	1.04079	1.06369	1.19124
旋律	1.00960	1.04096	1.04407	1.04926
和声	1.03443	1.03787	1.00204	1.04220
———第4列———				

センテンスの時間区間入力を受けていないので反応しない。

表4 中間層第4層と出力層の結合

	興奮要素	興奮方向	抑制要素	抑制方向
明るい	3.94859	6.37086	8.17069	8.11484
暗い	0.35046	1.35046	2.05533	1.91569
軽快	2.82043	4.53507	5.72495	5.69045
重々	0.62406	0.96256	1.21201	1.13739
安定	0.62406	0.96256	1.21201	1.13739
不安定	0.62406	0.96256	1.21201	1.13739

4. 考察

表2は“音楽三要素の楽曲への影響”を示す部分である。興奮方向での旋律の結合が表の中でもっとも強く出ている。12小節まで、上昇、下降を同じように繰り返しているが、13小節目から最終小節にかけ常に上昇していることがその要因と考えられる図8。また、リズムが興奮要素と抑制方向で結合が強く出ているのは、4分音符がもっとも多く使われているこの楽曲がtemp 120であり4分音符=0.5秒となり、個人情報を中心値より、興奮作用であるためと、曲の最終部分で付点2部音符が使われたために抑制方向の反応が強く出たと考えられる。

表3は“時間区間の楽曲への影響”を示す部分であるが、楽曲が短くフレーズが2区間、センテンス0区間であるために、それらの特徴が現れていないと考えられる。その中でも第一列(発生音単位)の結合値が強く出しており、この曲は、動機、フレーズ単位ではなく、1音1音に対してイメージを受けたと考えられる。

表4の中間層第4層と出力層の結合においては、常に、抑制要素、抑制方向、興奮方向、興奮要素の順に結合値が大きい。その要因については今後調査をおこなう。

5. おわりに

1曲ごとでは、その楽曲の特徴を中間層の“音楽3要素”と“楽式論的時間区間”で示すことが出来ている。しかし、実験を繰り返すうちに、以前に行われた学習の結果がそれ以降の学習に生かされないことが判明した。“学習のしすぎ”で前の学習結果を完全にその時点の結果にしてしまう。この改善が当面の課題となる。

謝辞：

本研究の基本概念と音楽理論に関し御助言頂き、
日本大学芸術学部 大倉教授に感謝いたします。

参考文献

- [1] 井口 征士：「感性情報処理が目指すもの」
情報処理学会 Vol.35, No.9, pp792-798 (1994)
- [2] 波多野 諄余夫、三宅 芳雄：「認知科学の動向」
人工知能学会 Vol.7, No.5, pp746-753(1992)
- [3] 平賀 譲：「音楽認知への計算的アプローチとその課題」
情報処理学会 Vol.35, No.9, pp822-829(1994)
- [4] 吉野 巖、阿部 純一：「メロディの調を認定する過程の計算モデル」
音楽情報科学研究会報告、3-4、pp23-30 (1993.10.23)
- [5] 大矢 健一：「ニューラルネットのダイナミクスによるリズム認知モデル」
音楽情報科学研究会報告、7-1, pp1-6 (1994.8.6)
- [6] 矢向 正人、土屋 景一、荒木 敏規：「旋律パターンの分類」
音楽情報科学研究会報告、16-5、pp27-32 (1996.7.28)

- [7] 村尾 忠廣：「音長の原構造パターンとクロージャ
ーランキングの設定」、音楽情報科学研究会報告、12-3、
pp13-18 (1995.10.21)
- [8] 野瀬 隆：「楽曲の構造をとらえる」、コンピュータ
サイエンス誌「bit」別冊、pp156-162(1987)
- [9] 平賀 譲：「音楽の認知・理解」、コンピュータサイ
エンス誌「bit」別冊、pp171-179(1987)
- [10] 玉木 明和、姚鳳会、加藤 清史：「フィードバッ
ク機能を持つ自動演奏システムの試み」、音楽情報科学
研究会報告、15-13、pp73-78 (1996.5.26)
- [11] 金森 務、平井 宏、堤 喜代司、弓場 芳治、
新美 康永：「コード及びメロディー・パートからの感
情情報の抽出」、ヒューマンインターフェイス 45-10、
情報メディア 8-10、pp75-79(1992.11.13)
- [12] 日間賀 充寿、大西 昇、杉江 昇：「情動に相関
のある楽曲中のパラメータについて」、音楽情報科学研
究会報告、8-2、pp7-12 (1994.11.19)
- [13] 水野 和彦：「音楽効果」、情報センター出版局
(1991)
- [14] 大倉 康義：「脳と脳波の聴感覚」、管楽教育研究
所
- [15] 大倉 康義：「人間の聴覚に於ける中心音及び中心
テンポの実験報告」(1997)
- [16] 櫻林 仁：「音楽療法研究」、音楽之友社(1996)
- [17] 日本音響学会：「音の何でも小事典」、講談社(1996)
- [18] 中島 安貴彦：「MIDI バイブルⅠ」、リットーミ
ュージック出版(1997)
- [19] 合原 一幸：ニューロ・ファジィ・カオス、
オーム社 (1993)
- [20] 黒沢 隆朝：「楽典」、音楽之友社 (1947)
- [21] 門馬 直美：「音楽の理論」、音楽之友社 (1992)