

## ニューラルネットワークを用いたアドリブ生成のためのリズム学習

薄 浩之 乾 伸雄 野瀬 隆 小谷 善行

(東京農工大学大学院 工学研究科 電子情報工学専攻)

アドリブの自動生成を目的とした、楽譜情報からのリズム感覚の獲得を目指した。人間のアドリブ演奏の、ある時点におけるリズムパターンと、その時点の直前のリズムパターンをそれぞれ教師値、入力値として与えたニューラルネットワークで学習を行う。アドリブ再生成時には、ある時点までの演奏情報をニューラルネットワークに与えることにより次の時刻の演奏情報が得られ、それを順次繰り返すことで曲の生成が行える。

筆者らは、学習のために演奏情報からリズムの決定要因と考えられる要素を抽出し、ニューラルネットワークの入出力とできるような形式に置き換えた。また、学習で用いるニューラルネットワークには、通常の階層型とリカレント型との二つを用意し、比較した。次に学習の評価のため、学習で得た知識を用いてアドリブの生成を行った。生成曲は新たな曲のコード進行を使用可能なスケールの列に変換し、そのスケール上にニューラルネットワークの出力を展開することで得られた。

## Rhythm Learning for generating ad-lib music by using Neural Network

Hiroyuki USUKI, Nobuo INUI, Takashi NOSE, Yoshiyuki KOTANI  
(Tokyo University of Agric. And Tech., Dept. of Computer Science)

We aim at generating automatic ad-lib music, and this paper proposes acquiring a rhythm sense from score information. A rhythm pattern at a certain point (A) and another rhythm pattern at just before another point (B) of the human ad-lib performance were used as an input value and teacher's data as an output to let a neural network perform its learning procedure. When re-generating the ad-lib, the performance information which is up to point A is given to the neural network. It then makes the system obtain the next performance information, and finally succeeds to generate a tune by repeating the same procedure.

For learning, an element which is thought to be a key factor of the rhythm is extracted from the performance information, and we changed the form of the factors so that the input and output of the neural network can be carried out. We also prepared two kinds of neural networks to compare with each other; feed-forward type and recurrent type.

To evaluate the learning procedure, we carried out the ad-lib generation by using the knowledge obtained by the neural network. First of all, we changed the chord progression of the new tune into a note scale progression, and then matched with the output of the neural network as a tune generated by the system.

# 1 はじめに

軽音楽等におけるアドリブ演奏の自動生成は、これまで数多くの研究がなされてきた。手法はさまざまであるが、その際、アドリブフレーズ中に現れるリズム感覚、たとえば呼吸とか間といった言葉で形容されるものの存在に触れた研究は少なかった。なぜならば、音高生成に関しては、(本質をあらわしてはいないものの)ある程度完成された和声理論があるのに対し、リズム生成に関してはそういった規則がなく、演奏者の感覚がそのほとんどを決定しているため、計算機での扱いが難しかったからではないだろうか。

しかし、人間のアドリブ演奏において、音高と音価の生成は相互に影響し合っており、両方が重要である。計算機での処理にも一方が欠けていては妥当なシステムであるとは言えないだろう。そこで、本論文ではアドリブ演奏の自動生成を行う際のリズム感覚の学習を、ニューラルネットワークを用いて試みた。

また、本論文では楽譜上の音符の組み合わせで表現される範囲のリズムを扱う。「ノリ」や「タメ」のような、演奏時の速度変化などについては、演奏表情生成の研究に譲る。

# 2 リズム感覚の学習

アドリブフレーズ中に現れるリズムには自然な「流れ」がある。したがって、人間がアドリブ演奏の際リアルタイムに生成する音価はそれまでの演奏によってある程度決定される、と言える。そこでニューラルネットワークに既存の人間の演奏から、ある範囲の演奏情報を入力、その直後の時刻に演奏されるリズムを教師値として与え学習を行う(図1)。そうすれば、生成時にはニューラルネットワークに演奏情報を与えることによって次の時刻の妥当なリズムを出力として得ることができる。さらに、その出力を次の時刻において入力として与えることを順次繰り返せば、次々とリズムが生成され、その結果、自然なリズムの流れが得られるだろう。

ここでニューラルネットワークを学習に用いる理由には次のようなものがある。

- ・ 次の時刻のリズムを決定する要因が、以前の演奏に含まれていることはわかっているが、具体的にどう応用されるかを予想することが難しく、リズム決定要因の抽出規則をあらかじめ用意することが難しい。
- ・ ニューラルネットワークが人間的な処理に有

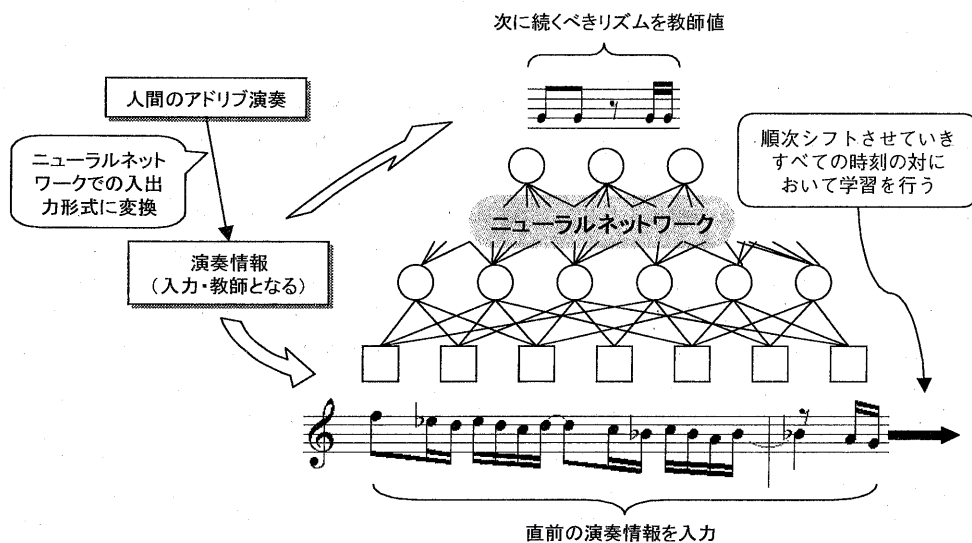


図1 既存の演奏情報からの学習

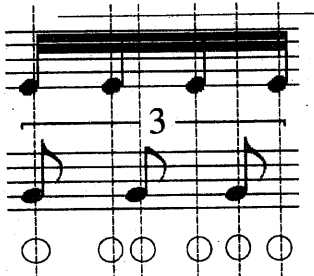


図2 一拍中でノートオンの可能性のある六つのタイミング

効であると考えられている。

- ・未知の入力パターンに対して、ある程度妥当な出力が期待できる。

また、このように本手法では演奏情報だけをリズム決定の要因として使うが、実際の人間の演奏では個々の曲の持つ特徴的なリズムパターンや、音楽ジャンルによるリズムの制約や指向なども大きなリズムの決定の要因であろう。しかし、本手法での学習はあくまでも人間の自然なリズム感覚の獲得に的を絞り、それらの要因についてはとりあえず考えない。ただ、学習の教師値においては、音楽ジャンルを絞り、演奏者や曲想、テンポなどが似ている曲を用いれば、(一般的な知識は獲得されないが)音楽形式等の要因も含めた学習が自ずと行えると考えた。また、そうすることにより、ニューラルネットワークでの学習の発散を避けた。

## 2.1 ニューラルネットワークの入出力要素の設計

演奏情報をニューラルネットワークの入出力要素とすると、学習がうまく行われるように適当な抽象化をしなくてはならない。つまり、リズム決定の要因となる要素だけを演奏情報から抽出し、コンパクトに記述する必要がある。

リズムはノートオンのタイミングの列で表されると考え、本手法では人間の演奏を記述した楽譜情報からノートオンのタイミングを抽出し、演奏をニューラルネットワークの入出力形式に置きかえる。その際、図2で示すように、一拍長さの中でノートオンの可能性があるタイミングが六

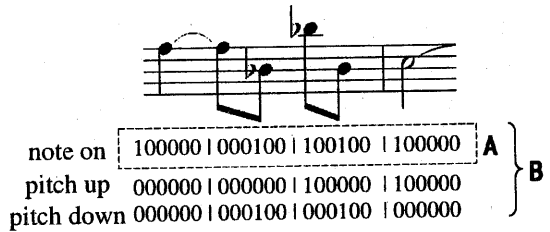


図3 ニューラルネットの入出力形式

つに限定できる<sup>1</sup>ことに着目し、一拍あたり6時刻の情報で表せるようにした。このそれぞれのタイミングに対し、単純にノートオン・オフで1と0を割り当て、それをニューラルネットワークの入出力形式とする(図3-A)。また、さらに音高の上下動がリズム決定に影響を与えているという予想も考慮し、先ほどの各時刻のノートオン・オフに加え、ノートオンの場合、その直前に発音された音高よりその音高が高いか低いかの情報も記述する入出力形式も用意する(図3-B)。以上の方法で記述された演奏情報を直接、ニューラルネットワークの入出力として扱い、学習、再生成を行う。

この方法では、ノート・オンのタイミングだけを扱い、音価は考慮されない。これはなるべく入出力要素を簡潔に記述するためと、リズムを表現する際にノート・オフのタイミングはさして重要ではないと考えたためである。

## 2.2 学習教師曲

学習教師曲として、次の曲を使用した。

- ・ Take the A train / クリフォード・ブラウン
  - ・ I'll close my eyes / ブルー・ミッチェル
- (どちらも [9] の譜面を使用した)

二つともトランペットによる演奏であり、単旋律である。また、演奏年代、曲想、テンポが似通っている。二つの曲を両方使用して学習を行った。

<sup>1</sup> ジャズのアドリブコピー譜を想定した。実際には、5連符など当てはまらない場合もあるが、その場合最も近いタイミングを選択した。また、装飾音符などは、それが付加される音符に含まれるとして扱う。

## 2.3 階層型ニューラルネットワーク

まず、一般的な階層型ニューラルネットワークで学習を行った。i 番目のニューロンの挙動は次の式で定義される。

$$x_i = f\left(\sum_j w_{ij} x_j\right)$$

f シグモイド関数

$w_{ij}$  j 番目のノードから i 番目のノードへの結合強度

階層型のニューラルネットワークでは、ニューロンを層状に配列し、最下層に入力値の組を与えると信号が上の定義式にそって伝播し、即座に最上層から出力値が得られる。したがって本手法に適用すると、ある時刻の直前の演奏情報までが出力（次のリズム）に影響を与えることになる。

学習には、バックプロパゲーションアルゴリズムを用いた。ただし通常の、重み修正ステップ幅が一定の場合だと、ほとんどの場合で学習がうまくいかない。そこで、通常のバックプロパゲーション法で重みを修正する式、

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$$

$\eta$  更新係数

E 出力誤差の平方の和

に、焼き鈍し法による学習回数  $t$  による温度項を加え、

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} + \xi(t)$$

$\xi(t)$  平均値 0 で分散  $\beta^2/\log(1+t)$  の独立な白色ガウス分布ノイズ

とし、重み修正にランダム性をもたせた。ここで  $\beta$  を 1 以上とすれば確率 1 で誤差の最小値が得られることが知られている。しかし実用的には 0.01 程度の値で十分な場合が多い。

## 2.4 入力する演奏情報の幅

入力層に、ある区間の演奏情報を与えるが、どの程度の幅の演奏情報を与えるかが問題となる。入力する情報が短すぎると、(同じ入力に対し多

数の異なる教師が存在するため) 次のリズムの予想は困難であるし、逆に長すぎても、効率の良い学習ができなくなる上、本来関係のないノイズ情報を拾ってしまい適切な学習は行われないであろう。

そこで、入力する演奏情報の幅を変えて学習を行い、学習が収束したときの誤差を観察する。入出力形式 A と入出力形式 B のそれぞれの場合を調べる。次の場合について学習を行った。

### ● 入出力形式 A

入力層に与える演奏情報の幅が 4 拍、6 拍、8 拍、12 拍のそれぞれについて。

### ● 入出力形式 B

入力層に与える演奏情報の幅が 4 拍、6 拍、8 拍のそれぞれについて。

また、出力層は一拍分の演奏情報を出力するノードを用意した（入出力形式 A では 6 個、形式 B では 18 個）。したがって学習は学習教師曲の一拍ごとに行われる。

学習回数と誤差のグラフを図 4、図 5 に示す。学習回数は一回につき教師曲全体をひとまわり学習したことを示す。

その結果、単純に入力する演奏情報の幅が長いほど、少ない誤差で学習が収束した。また、入出力形式 A よりも B の方が、つまり、音高の動きの情報も付加した方が少ない誤差に落ち着いた。やはり、演奏情報が多いほど、リズムの予測が可能になるといえる。

しかし、これはクロズドデータにおける誤差であるから単純に誤差が少ない方が良いとは言いきれないだろう。問題は未知のパターンに対してどれだけ妥当な出力をするかである。そのときには入力要素が多いほど、余分なノイズを拾う可能性が大きいのではないだろうか。

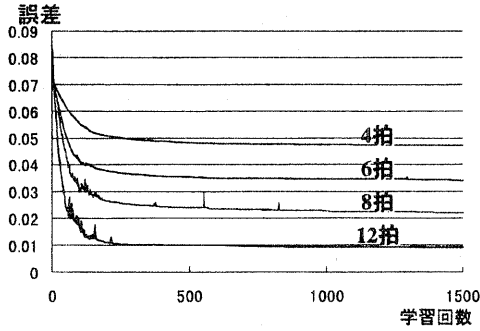


図4 入出力形式Aでの学習

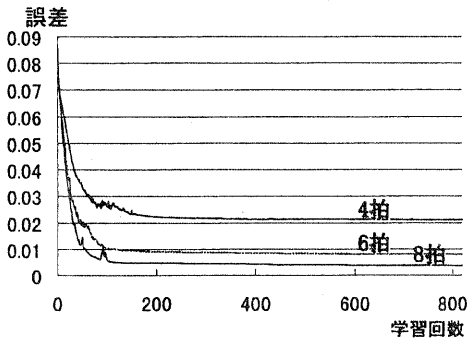


図5 入出力形式Bでの学習

## 2.5 リカレント型ニューラルネットワーク

次にリカレント型のニューラルネットワークで学習を行った。時刻  $t$  におけるリカレントニューラルネットワークのノード  $i$  の挙動は次の式で表せる。

$$x_i(t) = f\left(\sum_j w_{ij} x_j(t-1)\right)$$

$f$  シグモイド関数

リカレント型のニューラルネットワークでは、すべてのノードに自由な結合が許され、信号がネットワーク内部に残るため、時間概念を内包した処理ができる。

本手法に適用した場合、次々とその時刻ごとの演奏情報を与えることができ、その情報がネットワーク内部に留まるため、階層型ニューラルネットワークと違い、入力する演奏情報の幅を考慮の必要がない。また、入力が出力に影響を与えるまで、入力ノードと出力ノードの接続状況により時間がかかる。つまり、ある時刻の演奏情報が、出

力されるリズムに影響を与えるまで時間がかかる。

リカレント型ニューラルネットワークの学習には RTRL (Real-Time Recurrent Learning) 法を用いた。これは、 $w_{ij}$  の微小変化に対する出力信号の感度を計算し、1 時刻前の勾配情報を利用して  $w_{ij}$  を更新していく手法である。この学習においても、階層型のバックプロパゲーション法と同様に焼き鈍し法による乱数項を導入した。

入出力形式 A, B それぞれの場合での学習回数と誤差のグラフを図 6, 図 7 に示す。この誤差曲線の様子では、あまり学習はうまくいっていないように思われる。総誤差の平均も階層型ニューラルネットワークより大きい値に収束している。リカレント型ニューラルネットワークの学習がうまくいくかどうかはネットワークレイアウトの設計に因るところが大きいので、試行錯誤が必要であるが今回のレイアウトが失敗だった可能性もある。現在、さらに良い学習が可能なレイアウトを模索中である。

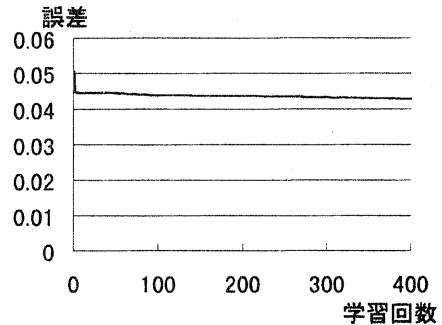


図6 入出力形式Aでの学習

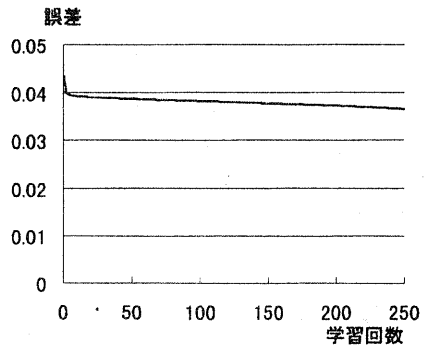


図7 入出力形式Bでの学習

### 3 アドリブの生成

学習で獲得した知識がオープン・データに対しても有効かどうかを見るために、新たなコード進行に対してアドリブを生成する。生成部における入力と出力は次のとおり。

- 入力
  - ◇ コード進行 A
  - ◇ コード進行 A 上での人間のアドリブ演奏
  - ◇ コード進行 B
  - ◇ リズムの学習で得た知識 (ニューラルネットワークの結合強度の組)
- 出力
  - ◇ コード進行 B に対するアドリブ

#### 3.1 アドリブ生成手法

ここで必要となる音高についての知識は人間のアドリブ演奏の楽譜情報とそのコード進行の対から学習する。まずコード進行 A を、あらかじめ用意された知識 (基本的なコード進行パターン, あるコード進行上で使用可能なスケールの知識) を用いて解析し、それをアヴェイラブル・ノート・スケールの列 A' に変換する。そして人間の演奏との対応を参照し、使用されているフレーズがどんなスケール上で使われているかを記憶し、フレーズ知識 P を得る。新たに与える曲のコード進行 B も同様に使用可能なスケール列 B' に書きかえる。

そして、学習されたニューラルネットワークが出力するリズムに適應するフレーズ知識 P から探し、スケール列 B' 上に展開する。

#### 3.2 生成アドリブ

生成されたアドリブが妥当なリズムの流れを持っているか、ということの一般的な評価を得るのは難しい。そこで、発表時に生成されたアドリブのデモを行うことにする。

### 4 おわりに

本論文では、人間のアドリブ演奏からの人間の

持つ自然なリズム感覚の獲得を目指し、ニューラルネットワークを用いて学習を行う手法を提案した。また、獲得した知識をアドリブの生成に適用した。

### 参考文献

- [1] 金森務ら, 即興演奏からのコード情報による曲想の抽出, 音楽情報科学研究会資料 Vol.11 No.3, 1992
- [2] 平井重行ら, ジャズの伴奏からの、調性を含めた感性情報抽出, 情報処理学会研究会報告, 95-MUS-8, 1994
- [3] 馬場則夫他, ニューラルネットの基礎と応用, 共立出版, pp.51-63, 1994
- [4] 甘利俊一, ニューラルネットの新展開, サイエンス社, pp.73-79, 1993
- [5] Johannes Feulner, Dominik Hornel, MELONET: Neural Networks that Learn Harmony Based Melodic Variations, ICMC Proceedings 1994, pp.236-239, 1994
- [6] Masao Nishijima, Kazuyuki Watanabe, Interactive music composer based on neural networks, ICMC proceedings 1992, pp.224-227, 1992
- [7] 田口友康他, 音楽情報処理の技術的基盤, 音楽情報科学に関する総合的研究・課題番号 04352030 調査報告, 1993,
- [8] 薄浩之, ニューラルネットワークを用いたコード譜とアドリブとの関連の学習, 東京農工大工学部電子情報工学科 1996 年度卒業論文, 1997
- [9] 佐藤貴洋, 小川隆夫, ジャズ・トランペット・アドリブ名演集, ドレミ楽譜出版社, 1994
- [10] 小谷教夫, ジャズ・ピアノ・アドリブ・マスター, リットー・ミュージック, 1994
- [11] 藤井貞泰, ジャズ・ピアノ・インプロビゼーション-1, リットー・ミュージック, pp.11-36, 出版年不詳