

H-13 Recurrent Neural Networkを用いたベクトル系列の生成

Vector Sequence Generation using Recurrent Neural Networks

川端 俊司

木村 英志

阿江 忠

1 はじめに

人間特有の知的情報処理として、認識、理解、創作が挙げられる。特に、創作は主観とコンテキストをもった知識により行われ、その知識は時間に伴い変化する。静的なルールを記述して総当たりに計算させるといった、従来の人工知能的手法でこれらの機能を工学的に実現するのは困難である。

本研究では、ある時刻における状態をベクトルで表し、その時系列から知識獲得・生成を行うモデルとしてニューラルネットに系列のコンテキストを捉えるニューロンを採り入れた Recurrent Neural Network(以下 RNN) を構築する。ベクトル系列の応用範囲は広いが、ここでは例として音楽における創作活動、作曲をとりあげて、提案モデルを用いたシミュレーションを行う。

2 ベクトル系列

本研究での知識獲得の対象は主観とコンテキストをもち、スカラー量で表現するのは困難である。また時間に伴って変化するため、知識獲得の対象をベクトル系列で表現する。

2.1 ベクトル系列の定義

実数の集合を \mathbf{R} とするとき、 n 次元の実数ベクトルを

$$\mathbf{V} \stackrel{\text{def}}{=} [v_1, v_2, \dots, v_n] \quad (v_i \in \mathbf{R}, i = 1, 2, \dots, n)$$

と表し、長さ m のベクトル系列を

$$\mathcal{V}_m \stackrel{\text{def}}{=} \mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_m$$

と表す。

2.2 知識獲得と生成

状態遷移の知識表現を行うために、ベクトル系列からの知識獲得と生成を行うシステム S について考える。システム S は入力ベクトル集合 \mathcal{I} と出力ベクトル集合 \mathcal{O} 、状態集合 \mathcal{Q} をもつ。さらに2つの動作モードとして、生成モード S_G と学習モード S_L をもつ。

さて、時刻 t での、ベクトルの入出力をもつシステム S の出力 $O(t)$ は、

$$O(t) = S_G(I(t), Q(t))$$

で表す。ただし、 $O(t)$ 、 $I(t)$ はそれぞれ、 $O(t) \in \mathcal{O}$ 、 $I(t) \in \mathcal{I}$ となるベクトルとし、 $Q(t) \in \mathcal{Q}$ とする。これは、システムが時刻 t での状態 $Q(t)$ と入力ベクトル $I(t)$ から、時刻 t での出力ベクトルを出力することを意味している。

また、

$$Q(t+1) = S_L(I(t+1), O(t), Q(t))$$

により、時刻 t での出力、時刻 $t+1$ での入力ベクトルと時刻 t でのシステムの状態 $Q(t)$ から、時刻 $t+1$ でのシステムの状態が決定される。すなわち、システムの状態遷移はこれによって行う。

3 一般的な Recurrent Neural Network

RNN は出力層あるいは隠れ層から入力層への Feed-Back となる層を持ち、システムが過去の出力や状態に依存する形になっており、Feed-Forward の階層型ニューラルネットの時空間への一般化と考えられる。これにより通時的な学習、すなわち時系列入力に対する学習が可能となる。

RNN の学習は信号伝達荷重を固定して出力の伝播を時間方向に展開することで階層化すれば、Back-Propagation 則を自然に拡張することができる (Rumelhart ら, 1986)。これは通時的 Back-Propagation(BPTT) と呼ばれ、初期時刻 t_0 から時刻 t までネットワークの計算を行いながら、入出力と信号伝達荷重を保持しておき、 t から t_0 までの誤差計算および信号伝達荷重の修正を行う。このとき用いる教師信号は時刻 t のもののみで、それまでの間のものはいない。したがって離散時間においては、RNN を Feed-Forward 型のニューラルネットとみなして Back-Propagation 則を適用することと等価である。

4 提案モデル

BPTT では時刻 t での学習において時刻 t_0 から $t-1$ までのネットワークの状態を保持しておかなくてはならず、長い系列に対して学習・生成を行うには、膨大なメモリと計算時間が必要となる。

そこで、Feed-Forward 型のニューラルネットの各ニューロンに、知識獲得の対象となるベクトル系列のコンテキストを保持する有限個のコンテキストニューロンを加える。ネットワーク内にコンテキストを保持することで計算も容易になり、かつ十分な個数のコンテキストニューロンを用意すれば系列のコンテキストをとらえることができると考えられる。

4.1 モデルの構成

入力層、出力層については Feed-Forward 型のニューラルネットと同様に、個々のニューロンは以下のような入出力をもつ。

W_{ji} はニューロン i, j 間の信号伝達荷重、 O_i, O_j はニューロン i, j からの出力信号とすると、

$$O_j = f_j \left(\sum_i W_{ji} O_i \right)$$

とし、 $f_j(x)$ はシグモイド関数とする。

隠れ層では,

$$O_{Hj} = f_j \left(\sum_i \left(W_{Hji} O_{Hi} + \sum_k W_{Cki} O_{Cki} \right) \right)$$

とする。ただし、 W_{Hji} は隠れ層のニューロン i, j 間の信号伝達荷重、 W_{Cki} はコンテキストニューロン k と隠れ層のニューロン i 間の信号伝達荷重、 O_{Hi} 、 O_{Hj} はそれぞれ隠れ層のニューロン i, j からの出力信号、 O_{Cki} はコンテキストニューロン k の隠れ層のニューロン i への出力である。また、 W_{Cki} は初期パラメータとして与え、学習による修正は行わない。

コンテキストニューロンは、 n 個前までの隠れ層のニューロンの複製を保持する。 $Q_{Hi}(t)$ が時刻 t における隠れ層のニューロン i の状態、 $Q_{Ck,i}(t)$ が k 時刻前の時刻 t のコンテキストニューロン i の状態を表すとすると、

$$Q_{Ck,i}(t) = Q_{Ck-1,i}(t-1), \quad (k = n, n-1, \dots, 2)$$

$$Q_{C1,i}(t) = Q_{Hi}(t), \quad (i \in H)$$

となる。ただし H は隠れ層の添字集合である。

Neurons of Hidden Layer

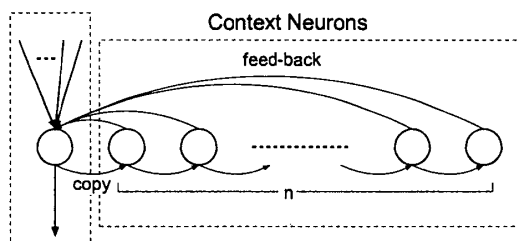


図 1: ニューロンモデル

4.2 系列生成の手続き

このような構成では、過去の入力から 1 時刻だけ先への写像を出力するに過ぎない。したがって、系列を学習・生成するためには、以下のように RNN を手続的に操作する必要がある。

1. 初期化 $V_t = V_0$
2. V_t を入力
3. ネットワークの計算を行って出力
4. V_{t+1} を教師信号として出力ベクトルと比較
5. Back-Propagation 則によりネットワークを修正
6. $V_t = V_{t+1}$ として 2. へ

5 シミュレーション

音楽は、楽譜で表されるように、作曲者の主観性をもった時系列多変量値とみなすことができる。本節では、前章において提案した RNN を例として作曲に適用し、メロディーの生成を試みる。

5.1 システムの入出力

楽譜を入力し、楽譜を出力することにする。提案モデルを用いたシステムはベクトルを扱うので、音長の最小単位を決定し、音高をビットベクトルにコーディングすることで、楽譜をベクトル系列に変換する。

5.2 シミュレーション条件

学習に用いた楽曲：J. S. Bach, “Aria con Variazioni”, 隠れ層の数：3 層, コンテキストニューロン：隠れ層の各ニューロンに 3 個, コンテキストニューロンから隠れ層の各ニューロンへの伝達荷重：0.6, 0.2, 0.1, 系列の学習回数：100000 回, 生成系列の長さ：96.

素の状態ではシミュレーションを行った結果、すぐに同じベクトルを生成し続け安定する状態や、学習したベクトル系列のなかに含まれるベクトルを生成して、そのまま学習したベクトル系列そのものに収束する状態に陥ることがわかった。そこで、こういった望ましくない状態を避けるため、5%程度の割合で信号伝達荷重に乱数を割り振った。96 の長さまで安定、収束しないものは 100000 個生成させた系列のなかで 3 系列であった。図 2 はその 1 例である。



図 2: 生成された楽譜の例

6 おわりに

RNN は決定性モデルであるので、安定状態や収束状態に陥る。それを回避するため、メロディー生成のシミュレーションでは乱数を用いて、そのような状態から強制的に抜け出させるようにした。今後の課題として、より有効な自動化の手法を実装することが挙げられる。

参考文献

- [1] 大槻和史, 中井満, 下平博, 嵯峨山茂樹, “HMM と音符 n-gram を用いた音楽リズム認識”, 音楽情報科学学会誌, Vol.42, No.10, pp.61-68, 2001.
- [2] Iain B. Collings, John B. Moore, “Identification of hidden markov models with grouped state values”, *IFAC YAC'95*, pp.194-199, 1995.
- [3] Jeffery L. Elman, “Finding Structure in Time”, *Cognitive Science*, Vol.14, pp.179-211, 1990.
- [4] John R. Koza, “Architecture-Altering Operations for Evolving the Architecture of a Multi-Part Program in Genetic Programming”, *Stanford University, STAN-CS-TR-94-1528*, 1994.
- [5] Tadashi Ae, Hiroyuki Araki, Keiichi Sakai, “Structured Vector Addition System –A Simulated Brain Model for Creative Activity–”, *Abstract Book CASYS'2000, Symposium7*, p.6, Liege, Belgium, 2000.