

多時間スケールを持つリカレントニューラルネットワークの学習則

Learning Rules for Multi-timescale Recurrent Neural Networks

松元 叡一
Eiichi MATSUMOTO

東京大学大学院総合文化研究科
Graduate School of Arts and Science, The University of Tokyo

Time series data, such as natural language text and music data, often have multiple time-scales. The conventional recurrent neural networks can deal in time series but it is difficult to capture long term patterns because stochastic gradient descent algorithms cannot propagate errors properly through long time gaps. In this study, we propose a hybrid method for time series analysis using the reservoir computing and deep learning algorithms. We conducted experiments in a learning time series which has multiple time-scales and showed our layered architecture can improve performances for these data.

1. はじめに

リカレントニューラルネットワーク(RNN)は入力を受け取りながら内部状態を更新していくことで、時系列データの処理を可能にしたニューラルネットワークのアーキテクチャである。長時間に渡る時間パターンをRNNで学習することは、vanishing gradient problem[Pascanu 13-1]により難しいとされてきたが、この問題を克服するいくつかのアルゴリズムが提案されている。時間カーネルRNN[Sutskever 10]では、変化のタイムスケールの異なるニューロンを複数用意することで長時間の相関を遅いニューロンに覚えさせることができる。また、Reservoir Computing[Lukoševičius 09]のアルゴリズムではRNNのリカレント結合の部分の学習を行わず、出力部の結合のみを学習することでこの問題を回避し、効率的に学習を行うことが出来る。

一方でフィードフォワードニューラルネットワークの文脈では、近年Deep Learning[Hinton 06]が大きな成功を取っている。Deep Learningでは教師なし学習による前処理で多層のNNを適切な初期値から学習することを可能にするが、そのようにして得られた多層NNでは、層が進むにつれて、より抽象的な、高次の特徴が抽出されていくことが知られている[Le 13]。

RNNとDeep Learningを組み合わせて、多層RNNを学習させようという試みは、Deep Learningで主流のアルゴリズムであるRBMを拡張したもの[Pascanu 13-2]や、Reservoir Computingに階層構造をもたせたもの[Triefenbach 10]などがある。本研究では、Deep LearningのアルゴリズムのひとつであるStacked Auto-encodersとReservoir Computingを合わせたモデルを提案し、階層構造がRNNに長時間のパターンを学習させる上で重要であることを示す。

2. モデル

2.1 Reservoir Computing

本研究では、図1のような構造を持ったレイヤーを多層に重ね、デコーダ層を加えた図2のようなアーキテクチャを用いる。各レイヤーはリカレントネットワークからなるReservoir部と、それにフィードフォワードに繋がるEncoder部からなり、式(1)から(3)のように時間発展する。

連絡先：松元叡一, 東京大学大学院総合文化研究科,
東京都目黒区駒場3-8-1 16号館807
matsumoto@complex.c.u-tokyo.ac.jp

ここで、 τ はレイヤーのタイムスケールを、 α は入力の強度を表すパラメタであり、 $\sigma(\cdot)$ は成分ごとにかかるシグモイド関数を表す。また、 $\langle \cdot \rangle$ は期待値をとる操作を表す。

レイヤーの学習は、Reservoir Computingの考え方を活かして、Reservoir部は重みを固定し、Encoder部への重みのみを学習する。この学習はDeep Learningでよく用いられるオートエンコーダのアルゴリズムを用い、式(4)で表されるLを最小化するように確率的勾配法で学習を行った。ここでの λ は正則化パラメタであり、L1正則化を行った。

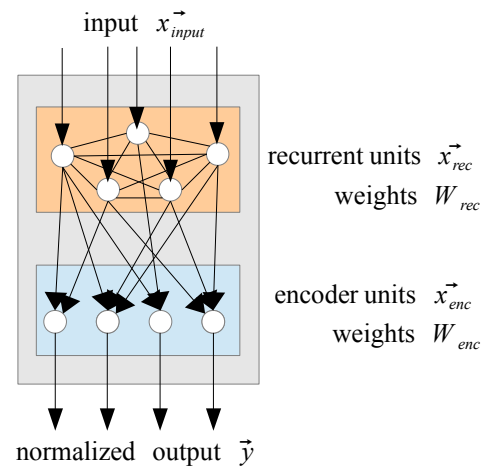


図1: レイヤーの構造

$$\begin{aligned} \vec{x}_{rec}(t+1) &= \left(1 - \frac{1}{\tau}\right) \vec{x}_{rec}(t) \\ &+ \tau \sigma(W_{rec} \vec{x}_{rec}(t) + \alpha \vec{x}_{input}(t+1)) \end{aligned} \quad (1)$$

$$\vec{x}_{enc}(t+1) = W_{enc} \vec{x}_{rec}(t+1) \quad (2)$$

$$\vec{y}(t+1) = \frac{\vec{x}_{enc}(t+1) - \langle \vec{x}_{enc} \rangle}{\langle (\vec{x}_{enc} - \langle \vec{x}_{enc} \rangle)^2 \rangle} \quad (3)$$

$$L = \langle \|W_{enc}^T W_{enc} \vec{x}_{rec} - \vec{x}_{rec}\|^2 \rangle + \lambda \|W_{enc}\| \quad (4)$$

2.2 Reservoir の多層化

図2のように、2.1で説明したレイヤーを多層に重ね、各レイヤーの出力をデコーダーレイヤーに通し、最終的な出力を生成するのがモデルの全体像である。デコーダーレイヤーの学習は、標準のパーセプトロンを用いた。

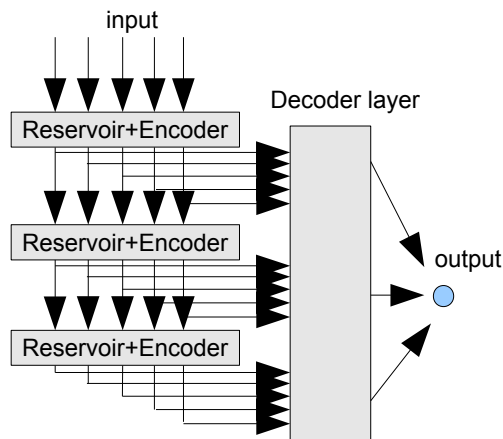


図2: モデルの全体図

3. 実験

3.1 テストデータ

本研究のモデルは、多層化により長時間のパターンを扱うことを目指している。そこで、複数の時間スケールのパターンを持つカオス時系列をテストデータとして用い、性能を評価する。

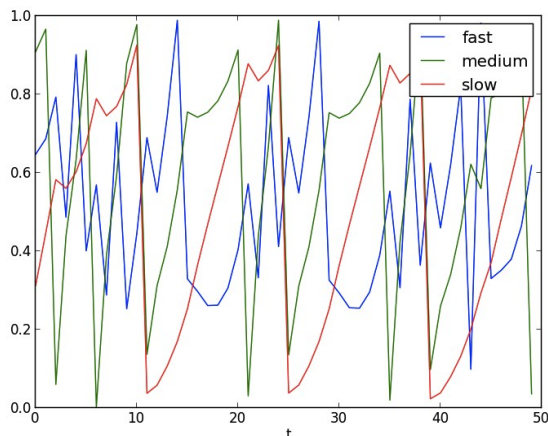


図3: テストデータの時系列 (一部)

ここでは、時間スケールの異なる3つのカオス振動子を結合して得られた図3のような時系列のうち、変動の速い方(図3の青線)を入力し、ステップtまでの入力からt+1の値を予測するというタスクで学習を行った。速い変数は残りの2つの変数の影響をうけるため、正確な予測を行うためには、速い方の変数の長時間パターンを見て遅い方の変数を推測できる必要がある。

3.2 結果

次の3つのネットワーク構造で学習を行い、誤差の減少の様子を調べた。デコーダ部はいずれの場合も中間層30ニューロンの3層パーセプトロンとした。以下のニューロン数、パラメタ数はデコーダ部を含めていない。

1. レイヤー数2, ニューロン数40, パラメタ数400
2. レイヤー数3, ニューロン数39, パラメタ数338
3. レイヤー数4, ニューロン数40, パラメタ数300

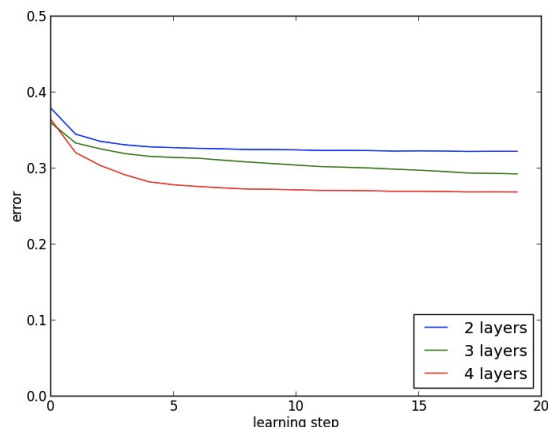


図4: レイヤー数を変えたときの学習曲線

学習の1ステップ(図4の横軸)は、時系列100,000ステップ分に対応する。図4より、ニューロン数がほぼ同数、パラメタ数は少ないにもかかわらず、レイヤーが多くなるほど学習能力が上がるのが分かった。このことから、ネットワークの階層構造が、長時間のパターンの認識において重要であることが分かる。

参考文献

- [Pascanu 13-1] Pascanu, R., Mikolov, T., Bengio, Y.: On the difficulty of training Recurrent Neural Networks, arXiv, 2013.
- [Sutskever 10] Sutskever, I., Hinton, G.: Temporal-kernel recurrent neural networks, Neural Networks, The official journal of the international neural network society, 2010.
- [Lukoševičius 09] Lukoševičius, Mantas, and Herbert Jaeger.: Reservoir computing approaches to recurrent neural network training. Computer Science Review, 2009.
- [Hinton 06] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh.: A fast learning algorithm for deep belief nets. Neural computation 18.7, 2006.
- [Le 13] Le, Quoc V.: Building high-level features using large scale unsupervised learning. Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, 2013.
- [Pascanu 13-2] Pascanu, Razvan, et al.: How to Construct Deep Recurrent Neural Networks., arXiv, 2013.
- [Triefenbach 10] Triefenbach, Fabian, et al.: Phoneme Recognition with Large Hierarchical Reservoirs., NIPS. 2010.