

異常検知や予測の問題における時系列信号のノイズの影響や標準化の効果に関する基礎的検討

東海大学 正会員 ○三神 厚

1. はじめに

機械学習の方法の土木分野への適用にあたっては各種のツールが開発され、実装の便が図られるようになってきている。しかしながら、実装結果の成否は用いるデータの preprocessing や信号に含まれるノイズの影響を受ける。本研究では、時系列データに機械学習を適用する問題として、オートエンコーダを用いた異常検知と LSTM (Long Short-Term Memory) ²⁾ による予測の問題を取り上げ、取り扱う時系列データのノイズの影響や前処理の効果に関する基礎的な検討を行う。

2. 検討に用いる機械学習の方法

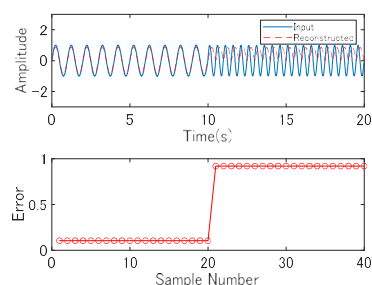
オートエンコーダは教師なし学習に分類され、正常データのみから異常を検知できる利点により異常検知などの分野でよく用いられる。エンコーダ部とデコーダ部からなり、入力したデータと同じデータが出力側で再構成されるように信号の特徴を学習させるネットワークである。LSTM は時系列データを長期、短期にわたり保存する機能を有するもので予測や分類などに用いられる。RNN (Recurrent Neural Network) の一種であるが、ニューロンが入力ゲート、出力ゲート、忘却ゲート等からなる LSTM ブロックに置き換わることで改善が図られている。

3. 異常検知と予測におけるノイズの影響と標準化の必要性

(1) オートエンコーダにおけるノイズの影響と標準化の必要性

図-1 は基礎的検討として、周波数が 1Hz から 2Hz に変化する正弦波を用意し、前半の 1Hz 区間を学習させた時、10 秒での周波数の変化を「異常」として検知できるかどうかの検討を行ったものである。その際、正規分布した乱数を発生させ正弦波にノイズとして付加したケースも用意し比較した。その結果、ノイズあり、なしの場合とも、図-1(a),(b)下段に示すように周波数変化点で再構成誤差が急増しており、異常を検知できていることがわかる。エンコーダによりノイズを除去しつつ信号の特徴を掴んでいるものと推察される。

次に標準化の効果について検討するため、図-2 のように右上がりの正弦波を用意し、その前半のみをオートエンコーダで学習させた後、信号全体を用いて異常検知を行った。この場合は特に異常が生じているわけではないが、時系列データの preprocessing



(a) ノイズなし

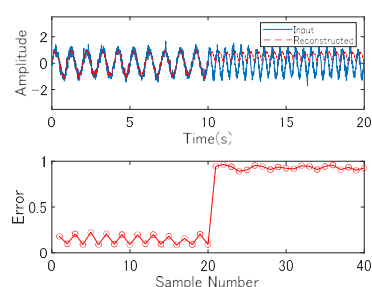
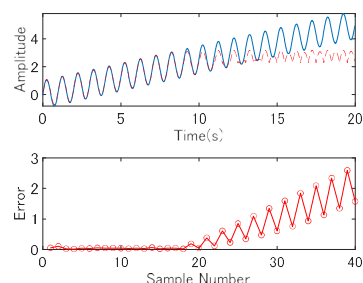
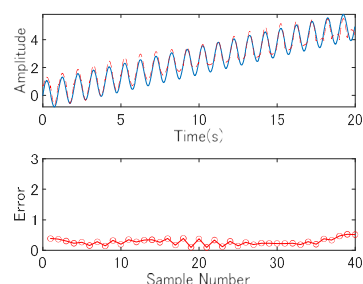
(b) ノイズあり ($\sigma = 0.2$)

図-1 オートエンコーダを用いた異常検知におけるノイズの影響



(a) 標準化なし



(b) 標準化あり

図-2 オートエンコーダを用いた異常検知における標準化の必要性

キーワード 機械学習, オートエンコーダ, LSTM, 標準化, ノイズ

連絡先 〒259-1292 神奈川県平塚市北金目 4-1-1 東海大学工学部土木工学科 TEL : 0463-63-4743 E-mail : atsushi.mikami@tokai.ac.jp

として標準化を行わない場合には、図-2(a)のように学習していない後半区間を異常と判定してしまう。一方、データの標準化を行った場合には、図-2(b)のように異常を検知せず概ね入力信号を再構成できていることがわかる。

(2) LSTM におけるノイズと関数の周期性の影響

次に LSTM の適用性について検討する。まずは周期的な関数として正弦波を用意し、それに正規分布した乱数を用いて異なるレベルのノイズを付加した 2 つの時系列データを用意する (図-3)。信号全体の 90% を学習させ、残り 10% 区間の予測を行った結果、ノイズの影響の度合いを変化させても、正弦波の周期的な特性を予測できており、さほどノイズの影響を受けないようである。

最後に周期的でない関数として、単調減少の指数関数を用意し LSTM の適用性を検討する (図-4)。先ほどと同様、異なるレベルのノイズを付加した上で、全体の 90% を学習に用い、残り 10% 区間の予測を行ったところ、ノイズが大きい場合は良好な予測結果が得られなかったが、ノイズのレベルを十分下げることによって、妥当な予測ができるようになった。検討の範囲ではあるが、LSTM はこのような単調増加、単調減少の関数の予測問題ではノイズの影響を受けやすく、あまり適していない可能性がある。

なお本研究で時系列信号に機械学習の方法を適用するにあたっては、MATLAB を用いた。

4. まとめ

オートエンコーダによる異常検知、LSTM による予測の問題を取り上げ、取り扱う時系列データのノイズの影響や前処理 (標準化) の効果に関する基礎的検討を行った。検討の範囲内ではあるが以下の結果を得た。

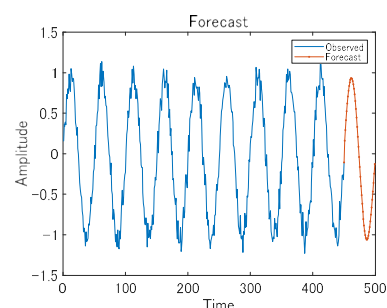
- ・取り扱う時系列が周期的な関数の場合、多少のノイズの影響があってもオートエンコーダによる異常検知、LSTM による予測が良好に行われた。
- ・時系列データの標準化を行うことによって、オートエンコーダによる異常検知が良好になる場合がある。
- ・時系列データが周期的な関数でない場合には、ノイズの影響を受けやすく LSTM による良好な予測ができない場合がある。

謝辞

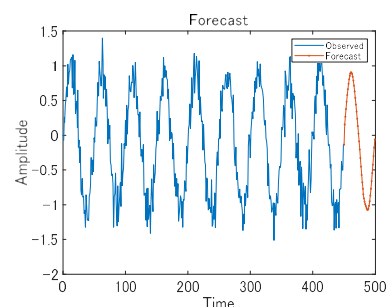
本研究は、JSPS 科研費 (課題番号: 19K04583) の助成を受けたものです。

参考文献

- 1) Hinton, G.E. and Salakhutdinov, R.R.: Reducing the Di-mensionality of Data with Neural Networks, Science, Vol.313 (5786), pp.504-507, 2006.
- 2) Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Computation, 9(8), pp.1735-1780, 1997.

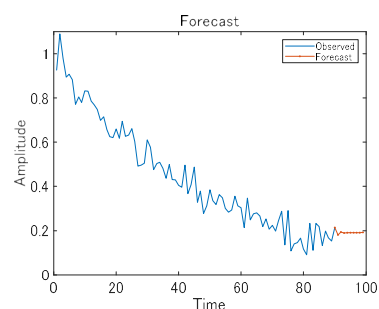


(a) ノイズ小 ($\sigma = 0.1$)

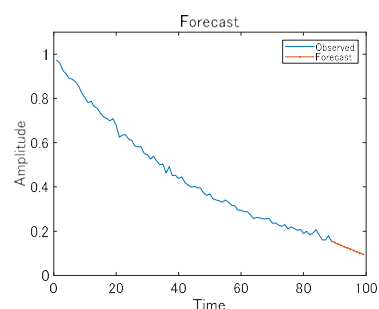


(b) ノイズ大 ($\sigma = 0.2$)

図-3 LSTM を用いた予測問題におけるノイズの影響



(a) ノイズ大 ($\sigma = 0.05$)



(b) ノイズ小 ($\sigma = 0.01$)

図-4 LSTM を用いた予測問題におけるノイズの影響