

時系列データの異常検出を目的とした 深層学習における再構築誤差の利用可能性に関する検討

A Study on Usage of Reconstruction Error of Deep Learning
for Anomaly Detection on Time-Series Data

河崎光毅 吉川大弘 古橋武
Koki Kawasaki Tomohiro Yoshikawa Takeshi Furuhashi

名古屋大学大学院工学研究科
Graduate School of Engineering, Nagoya University

Deep learning, a type of machine learning, is recently attracting attention for its use of extracting features from data automatically. Deep learning can extract features robustly even in data with noises. However, in data with momentary anomaly, it will smoothen the anomaly as if it were noises, which makes it harder to detect anomaly in practical use. In this paper, the usage of reconstruction error of deep learning for anomaly detection on time-series data is considered. As a result, we found that the reconstruction error can be used as threshold to detect anomalies, and the reconstruction data can eliminate artifacts of EEG data.

1. はじめに

情報技術の進歩により昨今は、大量の時系列データが容易に収集可能になっている。また同時に、これら収集した大量のデータにおける、データマイニングによる異常検出の必要性が高まっている [Chandola 09]。例えば生体センサーデータや機械システムデータに適用することで、体の病気や機械の故障の兆候を発見することが可能となる。大量のデータに対して、異常検出を目視で行うには限界があること、またデータによっては専門家の知識を必要とすることから、データマイニング等によって、人手を介さずに自動的に異常を検出する手法が求められている。

異常検出手法の一つとして教師あり学習がある。教師あり学習とは、正常データと異常データの特徴を学習し、未知のデータに対して正常または異常の識別を行う方法である。しかし一般に、異常データを大量に収集することや、収集した全てのデータに正常か異常かのラベルを付与することは難しい。正常データのみを学習するという方法も存在するが、正常データの収集時には異常データが一部混ざってしまうことも多く、それらの異常データが検出されない危険性がある。さらに正常データには複数の種類が存在するため、それら全てを学習する必要がある。そこで本研究では、上述の想定に対して、精度の良い学習が期待できるディープラーニングに着目する。

ディープラーニングは、近年注目を集めている機械学習の一つである。これはニューラルネットワークを多層にしたものであり、データセットの特徴量を自動で抽出することで、高い識別性能を発揮する。ディープラーニングの層の多さが表現力の高さを実現しているが、層が多い分、学習にかかる時間が長く、設定すべきパラメータが多いことが欠点として挙げられる。画像識別の分野では、人のおばあさん細胞（人の抽象表現をしている細胞）の作成に成功して一層注目を集めた [Le 13]。画像の分野以外でも、音声データの識別や時系列データの予測への適用が報告されている [Lee 09][Dalto 14]。本稿では、ディープラーニングにおけるプレトレーニングの再構築誤差を用いた異常データ検出手法を提案し、その性能を評価する。

2. 深層学習

2.1 学習方法

ディープラーニングでは、プレトレーニングとファインチューニングの2段階の学習が行われる。

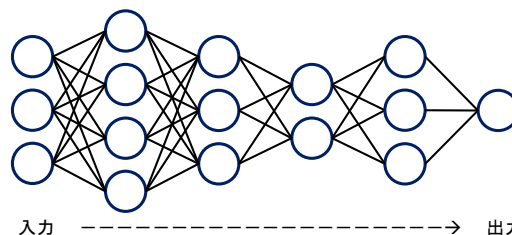


図 1: ディープラーニング

プレトレーニングでは、教師なし学習によって、出力層以外の全ての層を学習することで、データセットの特徴量を抽出する。例えば、図 1 のネットワークにおけるプレトレーニングでは、入力層（第 1 層目）から 5 層目までの 5 層全てを学習する。具体的には、入力層から 2 層ずつ順に学習していき、その結果を階層化させることで、全ての層を学習する。ある 2 層において、入力側の層を入力層、出力側の層を中間層、入力層と同じユニット数の層を出力層にした 3 層のネットワークを構築し、学習する。例えば、図 1 のネットワークの学習では、まず図 2 のようなネットワーク（図 1 の第 1 層目と第 2 層目）を考える。このネットワークの入力層と中間層の間は重み行列 \mathbf{W} 、中間層と出力層の間は重み行列 \mathbf{V} でつながっている。ただし、 $\mathbf{V} = \mathbf{W}^T$ と簡略化することが多く、本稿もこれにならう。中間層、出力層のそれぞれのバイアスを \mathbf{b} 、 \mathbf{c} とする。これらのパラメータ \mathbf{W} 、 \mathbf{b} 、 \mathbf{c} を用いて、入力 \mathbf{x} に対する出力 \mathbf{z} は、 $f(x)$ をシグモイド関数として以下のように求められる。

$$\mathbf{z} = f(\mathbf{W}^T \cdot f(\mathbf{W} \cdot \mathbf{x} + \mathbf{b}) + \mathbf{c}) \quad (1)$$

こうして計算した出力 \mathbf{z} と入力 \mathbf{x} との誤差の最小化問題として、パラメータ \mathbf{W} 、 \mathbf{b} 、 \mathbf{c} を学習する。この最小化問題の解法としては、RBM (Restricted Boltzmann Machines)[Bengio 07,

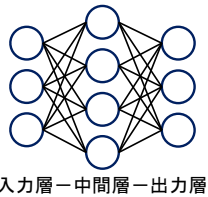


図 2: プレトレーニングに用いる 3 層のネットワーク

Hinton 10] を用いたものと, Autoencoder[Vincent 08] を用いたものの二つが有名であり, これらは誤差関数の定式化の方法が異なる. 学習した重み行列 W は元のディープラーニングの重み行列に, 中間層のバイアス b は元のディープラーニングの出力側 (第 2 層目) のバイアスに, 出力層のバイアス c は元のディープラーニングの入力側 (第 1 層目) のバイアスにすることで, 2 層の学習を終える. これを繰り返すことで, 全ての層の学習を行う.

ファインチューニングでは, 教師あり学習によって, プレトレーニングで得られた特徴量を入力とする識別器を作る. この識別器はニューラルネットワークが用いられることが多いが, SVM など他の識別器でも代用可能である [Tang 13].

2.2 異常の平滑化

プレトレーニングにおける特徴量の抽出では, 学習データの中の主要な特徴が捉えられる. 一方で, 一部の学習データにしか含まれない特徴については平滑化され, 抽出された特徴量には現れないという性質がある. これにより, ディープラーニングを用いて正常データと少数の異常データを学習する際に, 異常データに含まれる異常箇所が平滑化され, 正常データと同じ特徴量が抽出される. そうして得られた特徴量を用いてファインチューニングを行った際には, 誤った識別をする可能性がある.

3. 提案手法

2.2 の特徴に対して, 本稿では, ファインチューニングを行わない異常検出手法として, 再構築誤差による異常データ検出手法を提案する.

再構築誤差は, 入力データ x を, プレトレーニングで学習されたネットワークの最終層まで順伝播させ, そこからさらに順伝播と同じ重み行列とバイアスを用いて逆伝播させたデータ (再構築データ) を x' とすると, 下式で算出される.

$$\text{再構築誤差} = |x - x'|^2 \quad (2)$$

再構築データは, プレトレーニングの最終層から抽出された特徴量を基に構築されるため, ディープラーニングが元データをどのようなデータと認識したかを確認できる. すなわち, 再構築誤差が大きいほど, 特徴抽出の際に大きく平滑化が行われたことを意味する.

提案手法では, 異常データほど特徴抽出の際に平滑化が大きく行われることに着目し, 再構築誤差の値が閾値を超えているデータを異常と判定する. 閾値は学習データに対する再構築誤差の平均 μ , 標準偏差 σ , パラメータ k を用いて, $\mu + k\sigma$ と設定する. これは, 学習データの多くは正常データであり, μ と σ は概ね学習データの平均と標準偏差に等しいと考えられるためである.

4. 脳波データによる実験

4.1 実験に使用したデータ

実際にある一人の被験者から脳波を測定し, そのデータを用いて実験を行った. 被験者にはモニタを注視してもらい, そのモニタに一定間隔で図形を提示することで, その反応に対する脳波を計測した. 200Hz でサンプリングしたデータの, 一つの図形提示から次の図形提示までの区間を 1 データとして用いた.

学習データには, 300 回の図形提示に対する 8 電極のデータ (計 2400 データ) を用いた. テストデータには, 100 回の図形提示に対する 9 電極のデータ (計 900 データ) を用いた.

波形のうち, $\pm 50\mu V$ の範囲を超えている部分は眼球アーチファクトであると考えられる. 眼球アーチファクトは所望の波形とは異なるため, 検出し, 除去する必要がある. 本実験では, 眼球アーチファクトを含むデータを異常データとした時の, 異常検出の精度を評価した. テストデータには, 正常データが 873 データ, 異常データ (アーチファクトあり) が 27 データあった.

4.2 実験方法

4.1 で示したデータセットに対して, 提案手法を用いて異常データの検出を行った. ネットワーク構造は, 入力層から順に, ユニット数が 160, 200, 100, 100, 200 である 5 層ニューラルネットワークとした. プレトレーニングの学習率は 0.10, 学習回数は 2000 に設定した. テストでは, 各テストデータに対して再構築誤差を算出し, 識別を行った. 閾値に用いるパラメータ k は 0~2.5 まで 0.1 刻みで変化させ, k が異常検出に与える影響を調べた.

5. 実験結果と考察

閾値の設定に使う, 学習データに対する再構築誤差の平均 μ と標準偏差 σ はそれぞれ 6106.3, 10408.4 と求まった.

5.1 学習データの再構築誤差分布

学習の結果として, 全学習データに対する再構築誤差の分布を図 3 に示す. グラフから, 3000~5000 あたりをピークとした分布であることがわかる. しかし, 外れ値となるデータに引っ張られることで, 平均は 6106.3 と高い値となっている.

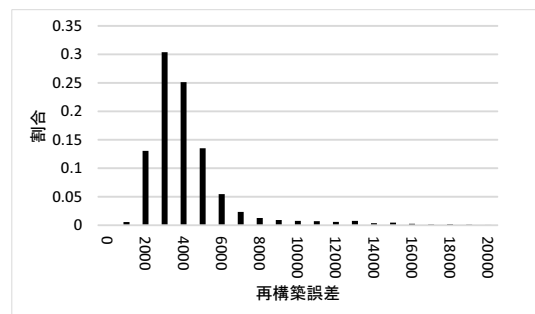


図 3: 学習データの再構築誤差分布

5.2 $k = 2$ としたときの異常検出精度

閾値のパラメータである k を 2 としたときの結果について述べる. テストデータに対する異常検出数を表 1 に示す. 表 1 において, 正常データに対する異常検出数が 0 であることから, 正常データは全て正常と識別されたことがわかる. 一方異

常データについては、14 データの異常が検出された。ここで、異常検出に対する各評価指標 (precision, sensitivity, F 値) を求める。precision は異常と判定したもののうち、異常の割合を示す。sensitivity は異常データのうち、異常と判定した割合を示す。この二つの指標はトレードオフの関係にあるため、これらを統合する値として F 値がある。F 値は以下のようにして求められる。

$$F \text{ 値} = \frac{2 \cdot \text{precision} \cdot \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (3)$$

これらの指標は [0, 1] の値をとり、高いほど精度がよいとされる。k = 2 としたときの各評価指標の値を表 2 に示す。k が低いときは、sensitivity が低いことから、F 値はあまり大きな値とならなかった。

表 1: テストデータに対する異常検出数

	正常	異常
異常検出数	0	14

表 2: 評価指標の値 (k = 2)

precision	1
sensitivity	0.51
F 値	0.68

5.3 k を変えたときの異常検出精度

k を変えたときの、各評価指標 (precision, sensitivity, F 値) の変化を図 4 に示す。k を大きくすることで precision は下がり、sensitivity は上がることがわかる。そして、F 値は山型のようなグラフとなっている。今回の実験では、k = 2 として異常検出の精度を確認したが、図 4 から、k = 1.8 として方が F 値が高くなることが確認された。k = 1.8 のとき、各評価指標の値は表 3 のようになった。

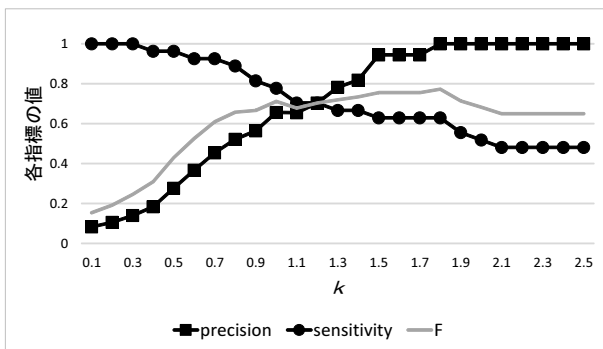


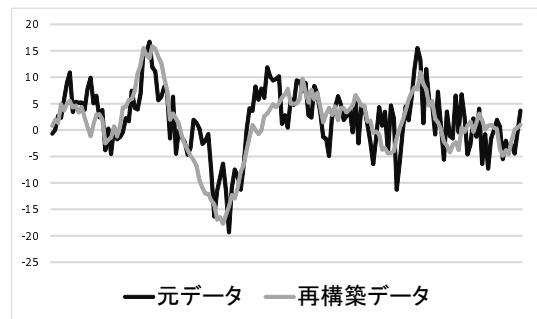
図 4: 閾値 k の値と各評価指標の関係

表 3: 評価指標の値 (k = 1.8)

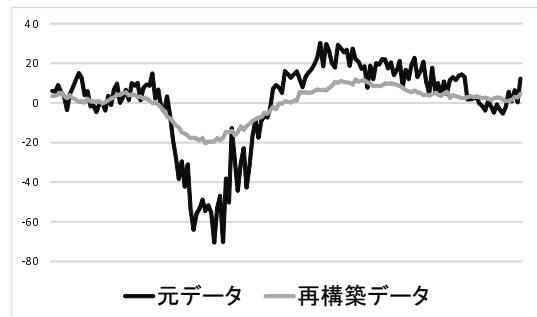
precision	1
sensitivity	0.62
F 値	0.77

5.4 元データと再構築データ

ここでは、様々な波形に対する元データと再構築データの関係について検討する。初めに、正しく判別できた例を図 5 に示す。正常データを正常と正しく判別できた図 5(a) では、再構築データが元データのノイズを除去した形になっており、元データと再構築データの概形が概ね一致している。異常データを異常と正しく判別できた図 5(b) では、異常部分を平滑化することで、再構築データを構築していることが確認できる。異常部分を平滑化している再構築データは、学習データに存在していた異常部分を除去していると解釈できる。すなわち、異常部分 (眼球アーチファクト) を含むデータを取り除くことなく、そのまま学習データとして使用できることを示唆している。次に、異常データを誤って正常と判別した例を図 6 に示す。図 6 から、異常の範囲が狭いこと、もしくは、異常の範囲以外の部分で振幅が極めて小さかったことから、再構築誤差が大きくならなかったことが原因と考えられる。これらの誤りに対しては、入力データに対するデータ長の設定方法を変えることで改善できると考えられる。



(a) 正常データを正常と判別した波形例



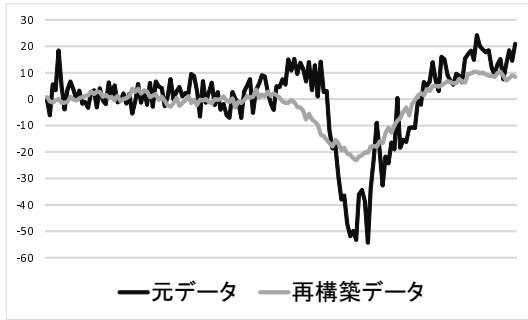
(b) 異常データを異常と判別した波形例

図 5: 正解判定例

上記では、判定が正しかった例と、異常を正常と判定した例について確認した。ここでは、k の値を変えたとき、正常ラベルのデータを異常とした例を図 7 に示す。図 7(a) では、眼球アーチファクトのような大きな振幅の部分があるが、その値が 50μV を超えていないため、異常ラベル (眼球アーチファクトラベル) が付与されなかったと考えられる。しかし、このようなデータについても、正解である異常データと考えても問題はないと思われる。図 7(b) では、眼球アーチファクトのように、一時的に振幅が大きくなるのではなく、大きなノイズが乗っているデータを検出した。これは眼球アーチファクトではないかもしれないが、アーチファクトの一種であり、何らかの異常データであると考えられる。これらの波形に対しても、アーチファクト部分を除去したような波形が再構築データでは得られ

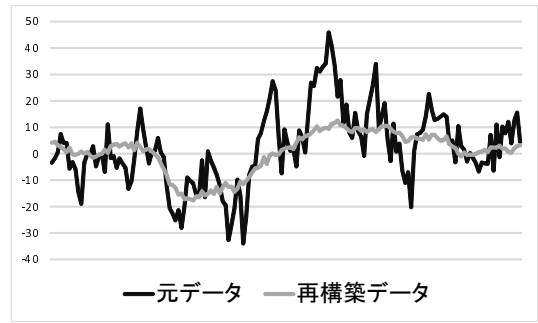


(a) 異常データを正常と判別した波形例①

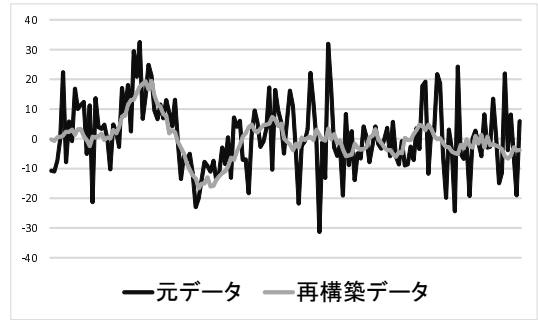


(b) 異常データを正常と判別した波形例②

図 6: 正常判定誤り例



(a) 正常データを異常と判別した波形例①



(b) 正常データを異常と判別した波形例②

図 7: 異常判定誤り例

ていることが確認できる。

6. おわりに

本稿では、ディープラーニングにおけるプレトレーニングの再構築誤差を用いた異常検出手法を提案した。実際の脳波データを用いて、眼球アーチファクトを異常として検出する実験を行った。再構築誤差を用いることで脳波のアーチファクトを検出でき、また、得られた再構築データは、それらアーチファクトを除去したデータと見ることができを確認した。このように平滑化されたデータは、フィルターにより除去することでも得られるが、その場合フィルターの特性上、一部のデータを捨てざるを得ない。再構築データは各入力に対して損失なく得られるため、脳波の解析にも役立つと考えられる。今後の課題としては、ディープラーニングの層の数、各層のユニットの数、学習率などのパラメータの設定方法や、入力データに対するデータ長の設定方法の検討が挙げられる。

参考文献

[Bengio 07] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al.: Greedy Layer-Wise Training of Deep Networks, *Advances in neural information processing systems*, Vol. 19, p. 153 (2007)

[Chandola 09] Chandola, V., Banerjee, A., and Kumar, V.: Anomaly Detection: A Survey, *ACM Computing Surveys (CSUR)*, Vol. 41, No. 3, p. 15 (2009)

[Dalto 14] Dalto, M.: Deep Neural Networks for Time Series Prediction with Applications in Ultra-Short-Term Wind Forecasting, *Rn ($\Theta 1$)*, Vol. 1, p. 2 (2014)

[Hinton 10] Hinton, G.: A Practical Guide to Training Restricted Boltzmann Machines, *Momentum*, Vol. 9, No. 1, p. 926 (2010)

[Le 13] Le, Q. V.: Building High-Level Features using Large Scale Unsupervised Learning, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013*, pp. 8595–8598 (2013)

[Lee 09] Lee, H., Pham, P., Largman, Y., and Ng, A. Y.: Unsupervised Feature Learning for Audio Classification using Convolutional Deep Belief Networks, in *Advances in neural information processing systems*, pp. 1096–1104 (2009)

[Tang 13] Tang, Y.: Deep Learning using Linear Support Vector Machines, *arXiv preprint arXiv:1306.0239* (2013)

[Vincent 08] Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A.: Extracting and Composing Robust Features with Denoising Autoencoders, in *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103 (2008)