

# 複数センサを用いた欠損データ復元

○豊川康平 山本大地 大谷雅之 (近畿大学)

## Recovery of Missing Data Using Plural Sensors

\* K. Toyokawa, D. Yamamoto and M. Otani (University of Kindai)

**Abstract**— In this research, we aimed to realize an IoT system that operates even if the sensor data is missing, and studied a method of restoring data using other sensor data located at the same location as the sensor where the data was missing. Specifically, we tried two methods using deep learning. 1. Restoration of missing data using time series data of the same kind of sensors in the same place. 2. Restoration of missing data using time series data of heterogeneous sensors located at the same place. In fact, we conducted experiments to restore sensor values using illuminance, sound, temperature sensor, and found out that it can be restored when there is a relation to the physical quantity acquired by the sensor.

**Key Words:** Deep Learning, Recovering sensor data, heterogeneous sensor

## 1. はじめに

近年、Internet of Things (IoT) 化が進み、家電やスマートフォンなどの多くの IoT デバイスが密接に関わりあうようになった。今後も自動運転や医療分野など多様な用途で用いられ、より IoT デバイスの普及は増加すると予測されている。IoT デバイスを用いるシステムの多くはセンサから取得したデータを、サーバや他の IoT デバイスに送信し、そのデータをもとにアクチュエータの制御などを行う。つまり、センサデータが確実に送られることがシステムの信頼性に繋がると言える。しかし、センサデータは故障や不具合、センサ自体の精度など様々な要因により欠損することがある。その結果、サーバや他の IoT デバイスにセンサデータが不到達となるという問題が起きる。

実際の例として、2018 年 8 月に、近畿大学水産研究所が飼育している近大マグロの実験場 (和歌山県東牟婁郡串本町大島実験場) にて、マグロの稚魚の生態を調査する実験が行われた。実験では、照度が稚魚に与える影響を調査するため、光センサを水槽周辺に 3 台設置し、照度データの測定を 3 日間行った。その実験期間中に実験場に台風が接近した影響で、瞬電などによるデータが欠損が度々発生した。災害以外でも、故障やネットワークの不具合は様々な要因によって起こりうる。さらに、海辺などの辺境に設置されたセンサは故障やネットワークの不具合が起こった際、その原因の調査や修理がすぐに行えない場合があり、センサデータの途絶が長期化することもある。

この問題に対し、センサデータの欠損が起こった際でも、センサデータを途絶させないシステムが求められている。実際の現場にセンサを設置する際は、近畿大学水産研究所の実験の例と同様に、複数のセンサが近傍に設置されることが予測される。本研究はこの状況を前提として、その中のあるセンサデータで欠損が起こった際、欠損が起こったセンサの近傍に設置されたセンサのデータから欠損したデータを予測し、補完することで欠損が起こったセンサデータの途絶を防ぐ方法を考案する。また、近畿大学水産救助大島分室で行った実験では、設置された 3 台のセンサは全て照度を測定するセンサ、つまり同種のデータを扱うものであった。しかし、これはたまたまであり、実験設定によっては、温度、音、照度などの異なる種類のデータを

測定するセンサが複数設置されることも予測される。そこで本研究では、予測するセンサデータと同種のセンサデータを用いた予測、予測するセンサデータと異種のセンサデータを用いた予測をそれぞれ行い、その結果を分析する。

## 2. 現場でのセンサデータ取得実験

2018 年 8 月に近畿大学水産研究所大島実験場でマグロのバースト現象の解析を目的として 3 日間の光センサデータの取得実験を行った。光センサは実験場内のマグロの稚魚を養殖している水槽付近に 3 箇所設置した。以下の Fig 1 にはセンサの設置環境の図である。この水槽は室内ではあるが光を完全に遮断しないビニールハウスのような環境であり、場所によって明照度には大きな差がある。そのため近傍に設置された同種のセンサではあるが、それぞれのセンサデータには大きな差があるため、センサデータの予測に他のセンサのデータを直接用いて補完することはできない。またセンサデータの測定は光センサの BH1750 を Raspberry Pi Zero W に接続して測定した。

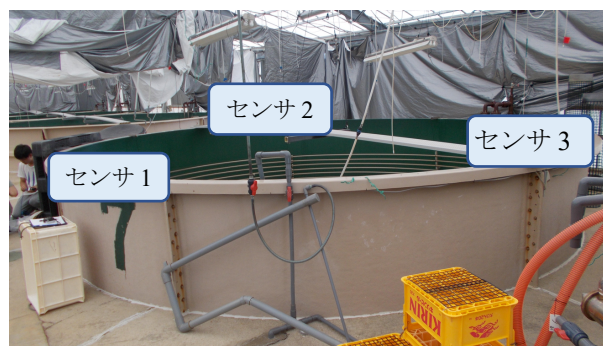


Fig. 1: Scene of the site

## 3. 同種のセンサデータによる予測

### 3.1. 復元方法

本研究では、扱うセンサデータを時系列データとみなし、一定時間に得られた複数のセンサデータから、欠損したセンサの期待される出力を予測する。時系列データの予測には、Long Short Term Memory (LSTM) [1]

を用いる。秒毎に記録された約30000個のセンサデータを使用し、9割を学習データ、残りをテストデータとして使用する。バッチサイズは100、エポック数は3000と設定するが、コールバック関数にEarlyStopping2)メソッドを適用し、lossの減少が収束したと判断した時に、自動で学習を停止させる。損失関数には平均二乗誤差、最適化手法にはRMSPropを使用した。以下のFig.2は使用した学習モデルのイメージ図である。

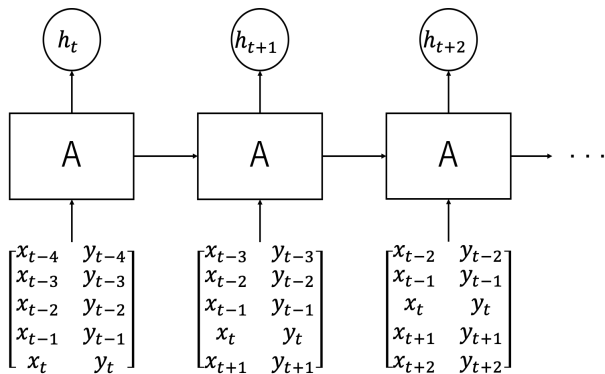


Fig. 2: Learning model

使用した3つのセンサをそれぞれh71, h72, h73とし、h71のデータとh72のデータを用いて、h73のデータを予測を行うとする。その時、図中のxはh71のデータ、yはh72のデータに対応し、Aは中間層(LSTM)、hは予測されたh73のデータとなる。入力には最大で2つのデータを使用し、1つのデータの予測を行う。

### 3.2. 実験結果

3.1節の手法を用いて、どの程度の予測精度が得られるかの確認を行う。予測精度の指標として、予測したセンサデータと実際のセンサデータの相対誤差を求め、より相対誤差の小さい予測データが予測データの多くを占めるほど予測精度が高いものとする。以下に示すグラフは相対誤差何%の範囲の予想データが予想データ全体の何%を占めるかを示すグラフである。Fig. 3は最も予測精度が高かったh71とh73のデータを用いて、h72のデータの予測を行ったものであり、相対誤差5%未満のものだけで約9割を占めている。また、Fig. 4は最も予測精度が低かったh71とh72のデータを用いて、h73のデータの予測を行ったものであり、相対誤差10%以上のものだけで約7割を占めている。Fig. 5は中間の予測精度となったh72とh73のデータを用いて、h71のデータの予測を行ったものである。

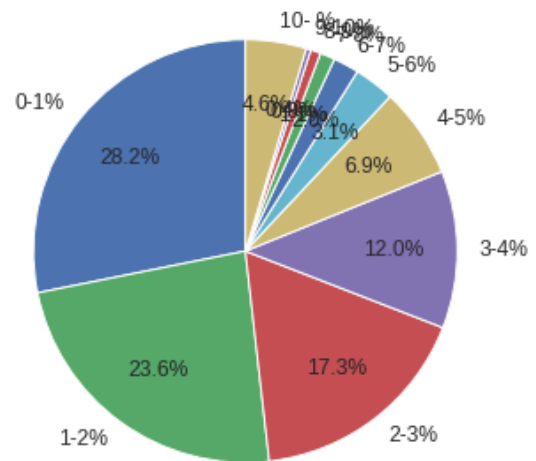


Fig. 3: Relative error (h72)

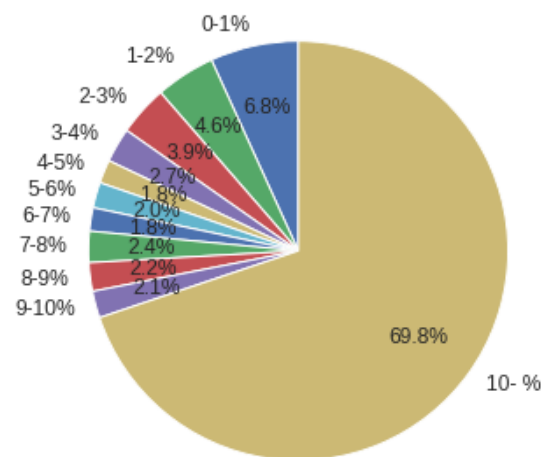


Fig. 4: Relative error (h73)

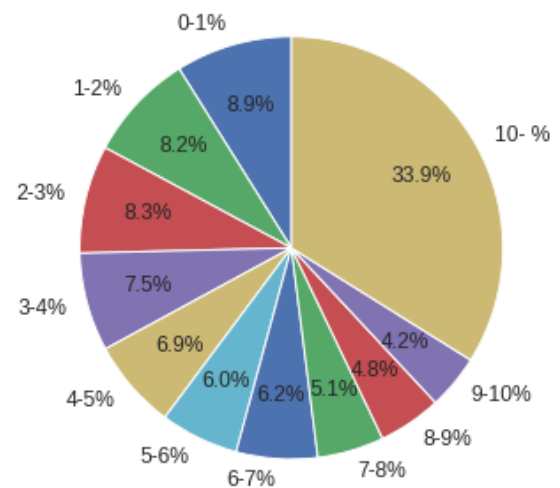


Fig. 5: Relative error (h71)

### 3.3. 考察

3.2節に示したような予測精度となった原因を使用したセンサデータより考察する。まず各使用データの時間毎の変化を確認する。Fig. 6は使用したデータを

100 個毎に取得, 最小値 0, 最大値 1 で正規化し, 各データの時間毎の変化を示したグラフである.

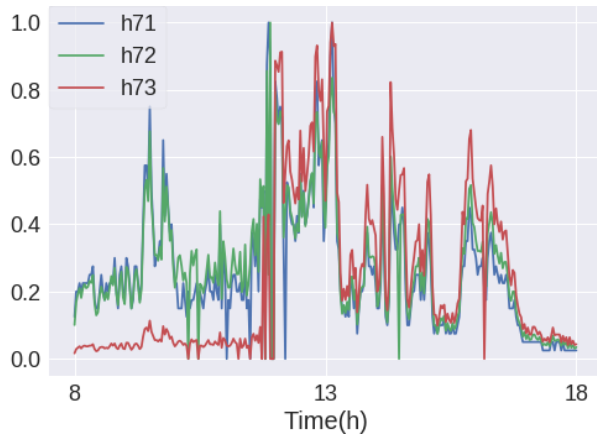


Fig. 6: Relation between datas (homogeneous)

3つのセンサは近傍に設置された同種のものであるため, 変化量に差はあれど同じように増減していることがわかる. しかし, それぞれの予測精度には大きな差が出た. その原因の一つとして使用データの分散が挙げられる. 使用した各データの分散を求めたところ, Table 1の表が得られた. この表は, 実際に使用した各データの分散を示している.

Table. 1: Variance (homogeneous datas)

Sensor	Variance
h71	0.019629635627761974
h72	0.003883186849847706
h73	0.041700260617692914

この表から, 分散の小さいデータを予測する場合に精度が高くなっていることが明らかになった.

また, 予測に使用するセンサデータを 1 つに減らした場合, 2 つのセンサデータを用いて予測を行った場合に比べ, 総じて予測精度は低くなった. これより予測に使用するセンサ数が増えることで, データを説明する変数が増え, 予測精度に影響を与えることが明らかになった.

## 4. 異種のセンサデータによる予測

### 4.1. 使用するセンサデータ

使用するセンサデータには 2019 年 1 月に大阪の民家で測定された温度, 音, 照度の 3 つのセンサデータを使用する. 実験地の温度は空調設備の影響を大きく受け, 音は主に室内の人間による生活音の影響を受ける. センサは窓際の日当たりの良い場所に設置したため, 照度は主に日光の影響を受ける. データの測定には GROVE-温度センサ, GROVE-音センサ, GROVE-光センサを Arduino Uno に接続したものをを使用した. 以下の Fig. 7 は実際に使用した機材である.

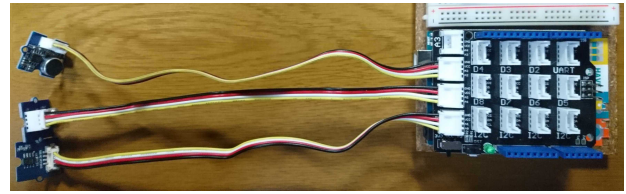


Fig. 7: Temperature, sound and Light sensor

### 4.2. 同種センサデータとの比較

0 章で述べた通り, 異種センサデータを用いた予測の目的は, 複数の環境データを測定するために異種のセンサデータを複数設置されるような環境において, 欠損したセンサデータの値を予測し, 補完するためである. 異種センサデータを用いての予測が可能となることで, 予測に用いるセンサデータの選択幅が大きく広がり, 実用性の向上を図ることが可能である. また復元手法にはエラー! 参照元が見つかりません. 節で述べた手法と同じものを用いてセンサデータの予測を行う.

### 4.3. 実験結果

3.1 節の手法を用いて, どの程度の予測精度が得られるかの確認を行う. 予測精度の指標として, 予測したセンサデータと実際のセンサデータの相対誤差を求め, より相対誤差の小さい予測データが予測データの多くを占めるほど予測精度が高いとする. 以下に示すグラフは相対誤差何%の範囲の予想データが予想データ全体の何%を占めるかを示すグラフである. Fig. 8 は最も予測精度が高かった温度センサと音センサのデータを用いて, 光センサのデータの予測を行ったものであり, 相対誤差が 1% のもので 8 割を占めている. Fig. 9 は最も予測精度が低かった光センサと音センサのデータを用いて, 温度センサのデータの予測を行ったものであり, 相対誤差 10% 以上のものが 8 割を占めている. Fig. 10 は中間の予測精度となった温度センサと光センサのデータを用いて, 音センサのデータの予測を行ったものである.

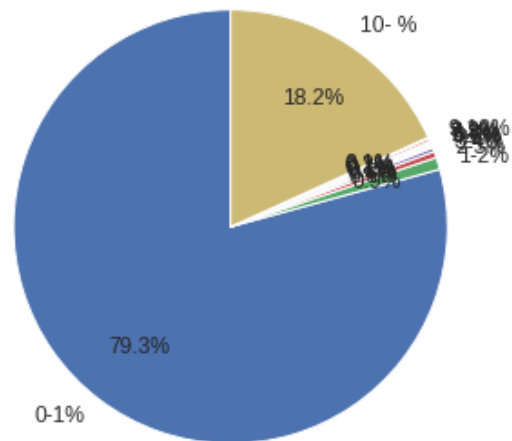


Fig. 2: Relative error (light)



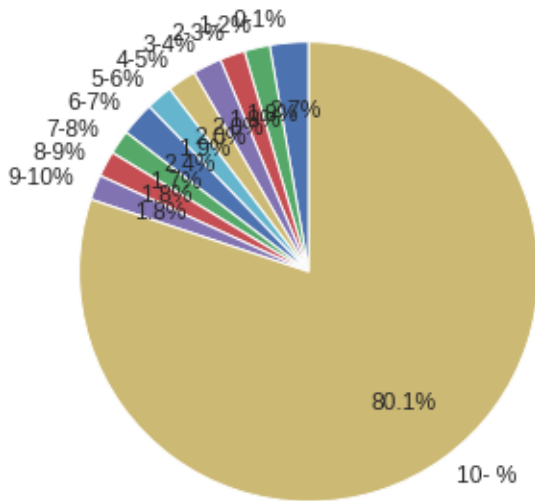


Fig. 9: Relative error (temperature)

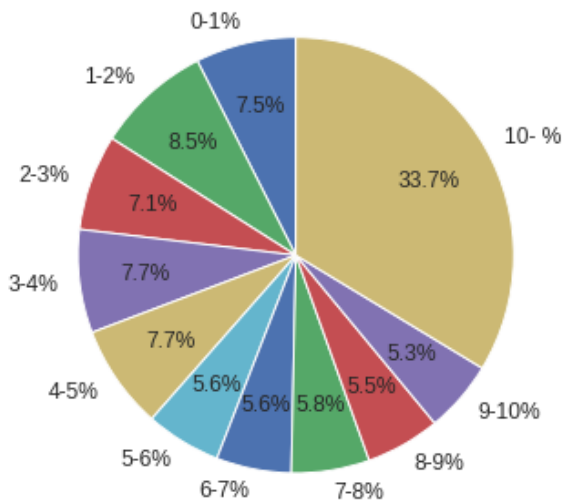


Fig. 10: Relative error (sound)

#### 4.4. 考察

4.3 節に示したような予測精度となった原因を使用したセンサデータより考察する。まず各使用データの時間毎の変化を確認する。以下の Fig. 11 は使用したデータを 100 個毎に取得、最小値 0、最大値 1 で正規化し、各データの時間毎の変化を示したグラフである。

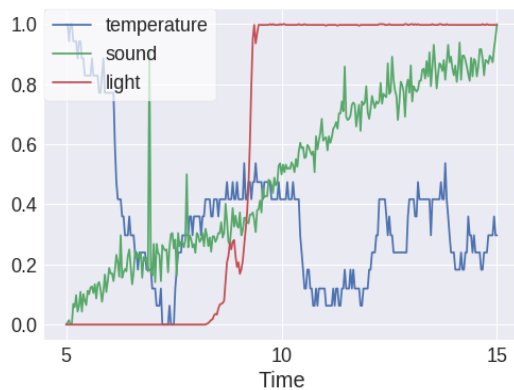


Fig. 11: Relation between datas (heterogeneous)

予測した際に一番予測精度が高くなった光センサのデータに注目すると、他センサのデータに比べ、大きく変化している時間が非常に短いことがわかる。このほぼ変化していない部分の予測が容易であったため予測精度が高くなったと考えられる。次に一番予測精度が低くなった温度センサのデータに注目すると、常にデータは変化し、またその変化も不規則であることがわかる。このことから予測が困難となり、予測精度が低くなったと考えられる。音センサのデータは常に変化はしているが、全体としては上昇傾向にあることから、ある程度の規則性はあるといえる。そのため温度センサのデータ予測の結果より予測精度が高くなったと考えられる。

また、予測するセンサを 1 つに減らした場合、音センサのデータのみを用いて光センサのデータを予測した場合と温度センサのデータを予測した場合のみ、2 つのセンサデータを用いて予測を行った場合の予測精度とほぼ変わらない予測精度が得られた。このことから、予測するセンサデータに対して使用する際、予測精度を向上させる特徴を持つセンサデータとそうではないセンサデータが存在することがわかった。

#### 5. まとめ

今回は 3 つのセンサを用いて実験を行った。同種センサのデータを用いた予測では、センサの数が増えれば予測の精度は上がったため、さらに予測に用いるセンサを増やし、センサ数と精度の関係を調査する必要がある。異種センサのデータを用いた予測では、センサデータ同士に相性があり、精度を向上させるデータとそうでないデータがあることがわかった。さらに多種のセンサを用いて予測を行い、予測するデータの特徴と精度を向上させるデータの特徴を比較、調査し、あるデータを予測するときどのようなデータを用いれば、高い精度の予測が可能となるのか明らかにする必要がある。

#### 参考文献

- 1) S.Hochreiter, J.Schmidhuber. Long short-term memory. Neural computation, 9(8):1735-1780, 1997
- 2) Caruana, S. Lawrence, and C.Lee Giles. Overfitting in neural networks: Backpropagation, conjugate gradient, and early stopping. In Advances in Neural Information Processing System 13:402-408, 2001