

# センサーデータの時系列特徴による間引きを用いた 送信量削減手法の提案

指吸 未来<sup>†</sup> 鈴木 智文<sup>†</sup> 岸田 隆祐<sup>†</sup> 塚田 晃司<sup>†</sup>

和歌山大学システム工学部<sup>†</sup>

和歌山大学大学院システム工学研究科<sup>‡</sup>

## 1 はじめに

近年、センサーが安価になったことにより、様々な場面でセンサーが利用されている。しかし、センサーデータの回収は十分な電力供給、良好な通信環境が前提とされている。そのため、既存の通信インフラが利用できない中山間地域などではセンサーデータを回収するために電波塔を建て専用のインフラ設備を整えて対応しているが、設備などの費用コストが大きいため普及は進んでいない。この問題を解決するには、効率よくデータを収集しなければならない。

そこで、本研究ではセンサーデータの変化量などから算出した時系列特徴によりデータを間引き、データの送信量を削減できるシステムを提案し、中山間地域にも対応できるデータ収集を実現する。

## 2 関連研究

送信量を減らすにはデータ圧縮が考えられる。[1]では相関のあるデータが観測されるセンサーに対し、ルーティング手法を組み合わせることで通信コストを最小化している。この方式では符号化に用いる行列式の情報をセンサーノードとシンクノードが共有している必要がある。

また、[2]では何らかの規則変動のある時系列データに対して、データの特性やタイミングに関わらず常に対象データと類似した学習データを選択できる手法である。この手法では予測モデルとして大量のデータを保存しておく必要がある。[1]では十分な通信環境、[2]では十分な設備を整えることのできない地域では対応できないという問題がある。

本研究ではこの問題に対応できるように中山間地域にも設置できるような設備を利用し、観測したセンサーデータを間引き送信量を減らすことで効率の良いデータ収集を実現する手法の提案を行う。

Proposal of communication traffic reduction method by decimating based on the time series variation characteristic of the sensor data.

<sup>†</sup>Miku Yubisui, <sup>†</sup>Tomoya Suzuki <sup>†</sup>Koji Tsukada

<sup>†</sup>Faculty of Systems Engineering Wakayama University

<sup>‡</sup>Ryusuke Kishida

<sup>‡</sup>Graduate School of Systems Engineering Wakayama University

## 3 提案手法

本研究は、回収対象となるセンサーデータの時系列特徴を用いることにより、データを間引き、データの大きな概形を取得できるように送信量を減らすことができ、山間部などの地域においても対応できるシステムを提案する。

### 3.1 システムの概要

本システムでは、観測されたデータを間引き、送信量を減らし、観測データと受信側のデータの誤差をできるだけ少なくすることを目的とする。システムの全体図を図1に示す。

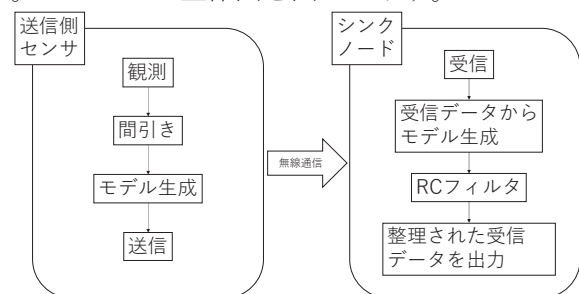


図1 システムの全体図

送受信双方で一日分のデータを観測モデル(以降、モデルとする)として保存し、送信側では送信しないデータの判断に使用し、受信側では送信されていないデータの補間に使用する。このように双方が観測モデルを保持していることで送信量を削減し、データを補間することが簡単に行うことが可能になる。

### 3.2 送信側処理

送信側は観測されたデータを間引き受信側に送信を行う。

観測日以前にセンサーから観測されたデータから算出したモデルを保存し、そのモデルとの時系列特徴の差を算出することでデータを送信すべきかどうかの判断を行う。その判断を行うにあたり以下の点に着目した。

- (1)日付ごとにおけるセンサーデータの変化量
- (2)時間ごとにおけるセンサーデータの変化量

また、今回はこれらの変化量を確認しやすい気温データを対象とする。

それぞれの変化量は日付を  $d$ 、ある時刻  $t$  の気温  $T$  とすると以下の式になる。

$$f_d(d,t) = T(d,t) - T(d-1,t) \quad (1)$$

$$f_t(d,t) = T(d,t) - T(d,t-\Delta t) \quad (2)$$

1日の気温は一定量ずつ変化するのではない。日が昇る時間帯では変化量は大きくなるが、深夜の時間帯では変化量は少なくなる。また、天候によっては、前日との気温変化がほとんどない日もあるため、日付ごとにおける傾きを使用することでデータの削減率を上げる。これらのことから、今回は傾きを使用する。

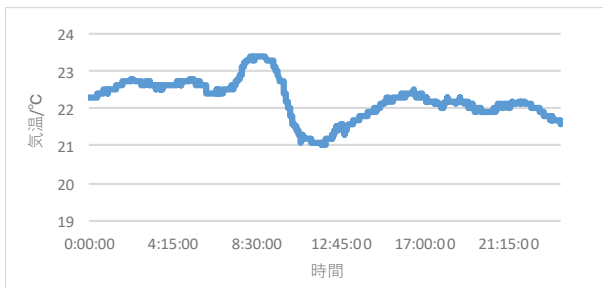
受信側も同じモデルを所持しているため、傾きの大きい部分は送信し、小さい部分はモデルの値を利用することでデータの補間を行う。

それぞれの変化量から傾きを求める。受信側ではそれぞれ閾値を設定し、その閾値よりも観測データとモデルの傾きの差が大きいときは送信するデータとする。それぞれのモデルを  $f_{dm}$ ,  $f_{tm}$  とし、閾値を  $a$ ,  $b$  とすると以下の式になる。以下の式が真のとき、データは送信される。

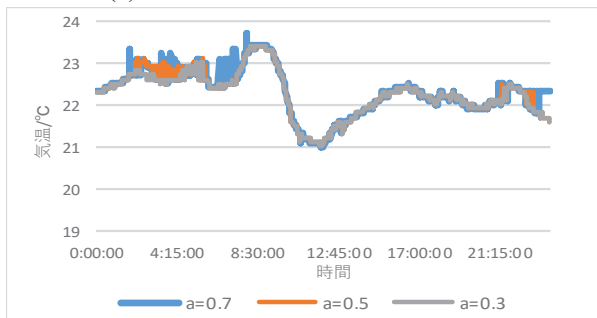
$$\left| \frac{\delta f_d(d,t)}{\delta d} - \frac{\delta f_{dm}(d,t)}{\delta d} \right| \geq a \quad (3)$$

$$\left| \frac{\delta f_t(d,t)}{\delta t} - \frac{\delta f_{tm}(d,t)}{\delta t} \right| \geq b \quad (4)$$

それぞれの閾値を大きくするとデータを少なく間引くことはできるが図2のように誤差も大きくなってしまふ部分がある。



(a) 元のデータ



(b)  $b=0.02$  のときの値の変化

図2 1日分のセンサーデータ

このように閾値を大きくすると誤差が大きくなっていくが、間引くデータを増やし送信量を減らすことが今回の目的であるので、データの大まかな概形を取得できる値まで閾値を大きくすることにする。

### 3.3 受信側処理

受信側は送信側からデータを受け取り、そのデータが観測されたデータとの誤差ができるだけ小さくなるように処理を行う。

受信側も前日までに受け取ったデータをモデルとして保存しておく。受信したデータに格納されているデータだけモデルを更新し、その更新されたモデルが当日の気温データとなる。

送信側で間引きが行われ、送信されたデータは観測されたデータとの誤差が閾値によっては大きくなる可能性がある。そのためローパスフィルタを使用し、誤差が小さくなるようにする。受信したデータを  $x$ 、フィルタをかけたデータを  $y$  とすると以下の式になる。 $c$  は係数で今回は  $0.1$  とする。

$$y[i] = c * y[i-1] + (1-c) * x[i] \quad (5)$$

### 3.4 誤差と削減率

それぞれ閾値  $a$  の値を変化させたときの観測値とフィルタを使用した後の整理された受信データの平均平方二乗誤差(RMSE)の値とデータの削減率を表1に示す。

	$a=0.7$	$a=0.5$	$a=0.3$
RMSE	0.159	0.128	0.095
削減率	23.5%	18.0%	16.6%

RMSE と削減率の値より気温データの削減とデータの大まかな概形は取得できていると考える。

## 4 おわりに

本研究では、データの時系列特徴による間引きを用いてデータを削減できるシステムを提案した。今回は気温データを対象として本手法を適応したが、山間部などに置かれるセンサーは、気温センサーの他に様々なデータがある。それらのデータに対し、本研究では周期的なデータであるならば、対応できると考えられる。

今後の課題としては、閾値  $a$ ,  $b$ 、フィルタ係数  $c$  の最適値の検証、評価、気温データ以外のデータへの対応、誤差の軽減などがある。

**謝辞** 本研究は、JSPS 科研費 15K00127 の助成を受けたものです。

### 参考文献

- [1] 柳沢 豊： 相関のあるデータを観測する無線センサネットワークのためのデータ集約手法, 情報処理学会研究報告, Vol.2011-UBI-32, No.4(2011-11)
- [2] 鈴木 雄也： 近傍データ収集法を用いた SW-SVR の改良, 情報処理学会研究報告, Vol-MBL-73 No.9(2014-11)