

SCoP におけるデータ圧縮方法

加々本 貴志[†] 竹内 雄亮[‡] 横山 聖[‡] 戸辺 義人[‡]
 東京電機大学未来科学研究科[†] 東京電機大学未来科学部[‡]

1. はじめに

近年センサデバイスの低価格化や小型化により、様々な分野でセンサが利用され身近なものとなってきた。それに伴い、環境モニタリングにおいてセンサネットワークへの活用の試みが広がった。従来のセンサネットワークにおけるセンシングデータの収集手法では、主にインターネットなどの広域な通信インフラを介してセンシングデータを収集している。しかし、このような手法では設置場所の制約や管理コストが非常に高くなるといった問題がある。そこで我々はセンサネットワークにおける新しいセンシングデータ収集機構として SCoP(Store, Compress & Pick-Up)を開発している^[1]。SCoP ではセンサデバイスで収集したセンシングデータを、無線を用いてクライアントが直接収集できる手法を採用する。センシングデータを効率よくクライアントへ送信するためにはデータの圧縮・分解を行う必要がある。本研究では SCoP におけるデータの圧縮・分解方法を確立する第一歩として Haar Wavelet 変換によるデータの分解を検証し SCoP における有用性を検証する。

2. SCoP システム

センサネットワークにおける問題を解決するために、我々は新しいセンシングデータ収集機構として SCoP を開発している。SCoP ではセンサが取得したデータをセンサデバイス内で圧縮・分解し、保持する。クライアントがセンサデバイス近傍通過時にそのデータをクライアントへ送信する。データを受け取ったクライアントは Wi-Fi エリアなどに移動をすると広域ネットワークを介してサーバへデータを送信しデータベースに格納される。SCoP ではセンサ自体に広域な通信インフラを必要としないため、既存の収集手法では困難であった設置場所の制約や管理コストの軽減を行うことができる。上記のとおり、本システムではデータの圧縮(Compress)、運搬(Pick-Up)、格納(Store)の 3 種類の機能で成り立っている。本稿ではデータの圧縮に焦点を当てて実験を行った。

3. 提案手法

本章では SCoP におけるデータ圧縮手法について提案する。SCoP システムにおいてデータの圧縮・分解を行う際に重要となることは、加工したデータを元のデータに復元する必要があるということである。また、センサデバイス付近に滞在する時間はクライアントによって異なる。そのため短時間の滞在で少量のデータしか取得でき

なかった場合でも特徴のあるデータを取得できることが望ましい。そこで、少ないデータでも特徴のあるデータを取得することを可能とするために Haar Wavelet 変換を用いてデータを分解する。Haar Wavelet 変換では連続したデータのうち前後のデータの平均のデータと差分のデータを取ることで、データ量を半分にすることが可能である。また、分解したデータと差分のデータを用いることによって元のデータに復元が可能である(図 1)。

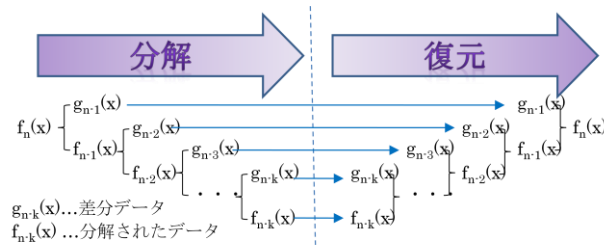


図 1. Haar Wavelet 変換によるデータ分解

4. 実験

提案手法のプロトタイプを実装し実験を行った。実験では圧縮したデータを復元し、元のデータとの差を取ることでデータの復元可能性の検証を行った。

4.1 実装

プロトタイプシステムでは Haar Wavelet 変換を用いて実装を行った。まず、元のデータ $f_n(x)$ を 1 つレベルを下げたデータ $f_{n-1}(x)$ と差のデータ $g_{n-1}(x)$ の 2 つに分解する(式 1)。 $f_{n-1}(x)$ は $f_n(x)$ の隣接したデータの平均から表すことができる(式 2)。さらに、 $g_{n-1}(x)$ は $f_n(x)$ の隣接したデータの差から求めることができる(式 4)。また、 $C_{2k}^{(n)}$ は $f_n(x)$ における $2k$ 番目のデータであり、 Φ_H と ψ_H はそれぞれ Scaling 関数と Wavelet 関数である。

$$f_n(x) = f_{n-1}(x) + g_{n-1}(x) \dots\dots (1)$$

$$f_{n-1}(x) = \sum_k \frac{1}{2} (C_{2k}^{(n)} + C_{2k+1}^{(n)}) \Phi_H(2^n x - k) \dots\dots (2)$$

$$\Phi_H(2^n x - k) = \begin{cases} 1 & \dots 0 \leq 2^n x - k \leq 1 \\ 0 & \dots \text{それ以外} \end{cases} \dots\dots (3)$$

$$g_{n-1}(x) = \sum_k \frac{1}{2} (C_{2k}^{(n)} - C_{2k+1}^{(n)}) \psi_H(2^n x - k) \dots\dots (4)$$

$$\psi_H(2^n x - k) = \begin{cases} 1 & \dots 0 \leq 2^n x - k < \frac{1}{2} \\ -1 & \dots \frac{1}{2} \leq 2^n x - k \leq 1 \dots\dots \\ 0 & \dots \text{それ以外} \end{cases} (5)$$

Evaluation of Data Compression for SCoP

[†] KAGAMOTO Takashi

[‡] TAKEUCHI Yuusuke, YOKOYAMA Takashi

[‡] TOBE Yoshito

Tokyo Denki University ([†], [‡])

この処理を繰り返すことにより、データ数が $1/2^k$ 倍にされたデータ $f_{n-k}(x)$ と k 個の $g_n(x)$ が生成される。さらに、これらのデータを用いることにより元のデータに復元することも可能である。分解されたデータのレベルを 1 つ上げるためには、分解されたデータに同レベルの差分データを加算減算することでデータ数を 2 倍にしてレベルを 1 つあげることができる(式 6)。ここにおける $d_k^{(n-1)}$ は $g_{n-1}(x)$ における k 番目のデータである。

$$f_n(x) = \sum_k [C_k^{(n-1)} + d_k^{(n-1)}] \Phi_H(2^n x - 2k) + \sum_k [C_k^{(n-1)} - d_k^{(n-1)}] \Phi_H(2^n x - 2k - 1) \dots (6)$$

以上の処理を繰り返し、すべての $g_n(x)$ を活用することで最終的に元のデータである $f_n(x)$ に復元することが可能となる。

4.2 実験方法

実験では電気式温度計によって得られた 2010 年 7 月 27 日 0 時 0 分から 18 時 27 分までの 1024 個の気温データを利用した。元のデータ数の $1/2^4$ 倍である 64 データになるまで計算を行い、分解されたデータのグラフと元のデータのグラフの特徴、誤差の平均値の比較を行った。式 7 は元のデータと分解されたデータとの差の平均値を出す式である。N は比較対象の分解されたデータが所有しているデータ数であり、 $a[i]$ と $b[i]$ はそれぞれ元のデータと比較対象のデータの i 番目のデータである。

$$F = \frac{\sqrt{\sum_{i=1}^N (a[i] - b[i])^2}}{N} \dots (7)$$

4.3 結果と考察

図 2 はデータ数 1024 個の元のデータとデータ数を 64 個まで減らしたデータの気温推移である。この図から、分解されたデータでも元のデータの気温推移の特徴をとらえていることがわかる。しかし分解を繰り返していくうちに、1 つ 1 つのデータ間における極めて小さな気温の変化がなくなっている。そのため、データをどこまで分解するべきであるかは今後検討していく必要がある。また、図 3 ではそれぞれ分解されたデータと元のデータとの差の平均値を出している。この図より、データ数を $1/2^4$ 倍にまで減少させた場合においても元のデータとの差の平均は 0.06°C 以下に抑えることができていることがわかる。気象庁が定めている気象測定検定^[2] において許容される電気式温度計の検定誤差が 0.5°C 以下であることから、 0.06°C は十分許容できる範囲の誤差であることがわかる。

5. 関連研究

今までにセンサネットワークの研究は多く行われてきた。南カリフォルニア大学で行われた Wisden^[3] では建造物に加速度センサを取り付け、振動で得られるデータをもとに建造物の健全性の評価を行っている。この研究では正確なデータ送信、データの圧縮、センサノード同士の同期に焦点が当てられており、データ圧縮には Wavelet 圧縮を行っている。

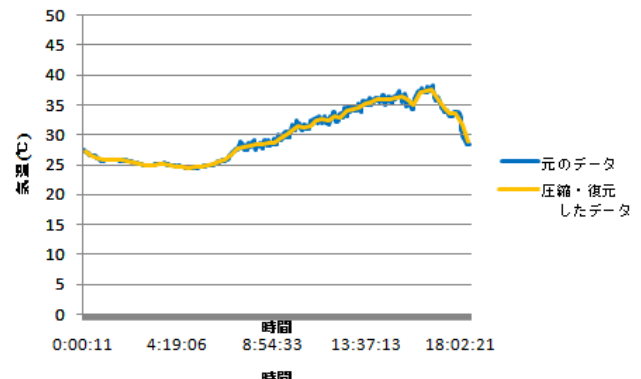


図 2. 気温推移図の比較

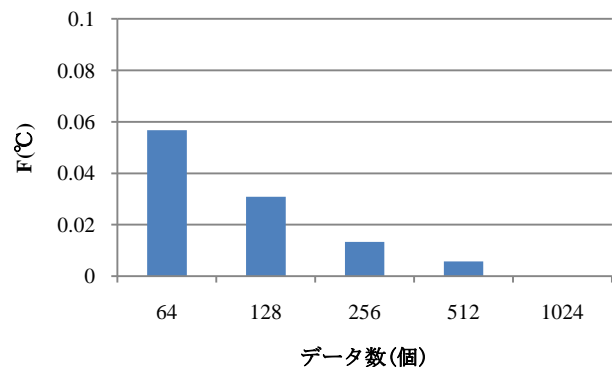


図 3. 元のデータとの差

6. まとめと今後の課題

本稿では、SCoP システムにおけるセンシングデータの分解・圧縮を行う手法として Haar Wavelet 変換を用いたデータ分解を提案し実装と実験を行った。その結果、本手法における分解方法であればデータ数を $1/2^4$ 倍まで分解を行っても元の気温データの特徴を残したままデータ数を減らすことができた。また、分解したデータを復元した場合にも元のデータとの誤差を 0.06°C 以下に抑えることができた。今回の実験ではデータ数は固定の数で実験を行ったが、実際に運用する場合には刻一刻とデータ数が増加していく。そのため、分解・圧縮するデータの範囲を指定する必要がある。そこで今後はデータ数が増加していく場合において利用するデータの範囲も考慮した実験を行っていく。

参考文献

- [1] 竹内雄亮, 横山聖, 加々本貴志, 高木篤大, 戸辺義人, “SCoP: 構築型センシングデータ収集機構”, 情報処理学会第 73 回全国大会, (2011)
- [2] 気象庁, “気象業務法第 9 条の検定の対象となる気象測器の検定の合格基準を定める告示”, http://www.jma.go.jp/jma/kishou/shinsei/kentei/goukakuki_jun.pdf (2001).
- [3] N. Xu, S. Rangwala, Krishna Kant Chintalapudi, D. Ganesan, A. Broad, R. Govindan and D. Estrin, “A Wireless Sensor Network For Structural Monitoring”, ACM SenSys’ 04, Baltimore, Maryland, pp13-24, (2004).