

多数のセンサーによる時空間センシングデータの 効率的な集約送信技術

宵 憲治¹ 山口 弘純¹ 廣森 聡仁¹ 内山 彰¹ 東野 輝夫¹ 柳谷 尚寿² 中谷 俊和² 立花 篤男³
長谷川 輝之³

概要: 本研究では、空間的に多数配置された多種多様なセンサーからのセンサーデータを時空間的に集約・圧縮し、リアルタイム遠隔センシングとモニタリングを効率よく実現するためのデータ集約圧縮技術を提案する。提案手法では、時間的類似性、空間的類似性が高いメタデータを有する複数のセンサーからのデータを時空間ブロックで集約することでメタデータ情報量の削減を実現し、かつ圧縮センシング技術を用いて圧縮率を適応的に制御する。あわせて家庭や集合住宅、オフィスビルや住宅地域などの拠点に設置された複数のセンサーからのセンシングデータを同様に拠点内に設置されたゲートウェイに集約し、データセンターなどにあるサーバへ送信するセンサデータ集約圧縮システムの設計開発を行っている。豊田市に実在する 29 施設に開発したシステムを設置し、データ集約評価を行った結果、センサーから発生した XML データ 10MB に対してシステムによる集約・圧縮を行い、送信データを 10 数 KB 程度に削減でき、さらに時空間ブロックによる集約により約 50%程度の削減を達成できている。

Aggregation and Transmission of Spatiotemporal Data from Heterogeneous Sensors

KENJI YOI¹ HIROZUMI YAMAGUCHI¹ AKIHITO HIROMORI¹ AKIRA UCHIYAMA¹ TERUO HIGASHINO¹
NAOHISA YANAGIYA² TOSHIKAZU NAKATANI² ATSUO TACHIBANA³ TERUYUKI HASEGAWA³

1. はじめに

スマートグリッドの本格的な普及・展開が期待されている。直近では、大手電力 10 社が 2024 年度までに管内の全家庭において、通信機能が付いた次世代電力計であるスマートメーターに完全転換するとの計画を経済産業省の制度検討会に提出した。政府が昨年閣議決定した日本再興戦略では「2020 年代早期に全世帯・全工場にスマートメーターを導入する」との目標を掲げており、それに備えて普及を急いでいる背景がある。こういった状況のもとでは、スマートグリッドに接続された個々の機器から、電力使用情報等に関する小規模データが多量に通信ネットワークに流通することとなるため、通信ネットワーク全体に大きな負荷がかかることが想定される。

スマートグリッドの発展はセンシング技術の発展と歩調を合わせている。家電製品、自動車、住宅、オフィスビル、そしてエネルギーや交通などの社会インフラに組み込む情報センサーの技術の発達によって、人やモノ、環境など実世界の情報を集めて活用する、実世界とサイバー世界の連携・融合が現実のものとなり、スマートコミュニティとよばれる社会を実現可能な時代へと移行しつつある。多種多様なセンサーから発生する大量のデータを収集・解析することで、デマンドレスポンスなどの電力制御や、高齢者の見守りや生活改善の推薦、大型施設でのナビゲーションなどのヒューマンセントリックなサービスの実現も期待される。例えば BEMS (Building Energy Management System) や HEMS (Home Energy Management System) では、温度センサー (空調制御)・照度センサー (照明制御)・人感センサー (人感制御) を前提とし、受変電、空調、熱源、衛生、電気、給配水、消防、セキュリティ、電力使用量、太陽光発電量などの監視と把握を行い、最適エ

¹ 大阪大学大学院情報科学研究科

² KDDI 株式会社ソリューション推進本部

³ 株式会社 KDDI 研究所

エネルギー制御を行う。環境センシングでは振動、照度、重量、大気、雨量などをモニタリングし防災や建造物管理などに活用し、移動体トラッキング・センシングではカメラやレーザ測域スキャナ、スマートフォン位置情報、WiFi ログ、スマートフォン内蔵センサーなどの異種情報源を組み合わせ、交通制御やスマートモビリティに活用することなどが想定されている。

これらのサービスの実現には、家庭や集合住宅、オフィスビルや住宅地域などの様々な場所に設置された多数のセンサーから得られる大量の電力データやセンシングデータをホームゲートウェイや地域ゲートウェイ（以下、ゲートウェイ）を介してデータセンターやクラウドサーバ（以下、サーバ）に収集し、解析するためのセンサーデータ収集基盤技術の確立が必要不可欠である。ここで一般に、センシングデータは様々なセンサーから様々なタイミングで取得されるため、多数のメタデータ（取得したセンサー名やセンサー精度、データ取得座標や頻度、時刻、プライバシーなどの属性情報）が付与されている。したがって、例えば数バイト程度のセンサー計測データ（以下、センシングデータ）に対しメタデータは数百バイトにもなる場合もある。さらに、センシングデータは多数のセンサーから時々刻々と生成されるため、ゲートウェイにおいて時空間的なデータ集約を行うことによるデータサイズの増大が無視できず、センシングデータとメタデータをそのまま集約および送信することはサーバ側のデータ処理リソースおよびゲートウェイとサーバ間の通信回線の利用効率の観点から適切ではない。特に環境センシングやヒューマンセンシングに関しては、山間部や過疎地など3Gアクセス網すら十分に利用できない地域を対象とすることも多く、ゲートウェイの通信回線断など不安定な状況も想定される。そのような場合には回線断期間中に送信されずゲートウェイに蓄積されたデータが回復時に一括送信されることによる、通信回線の更なる圧迫も予想される。したがって、通信状況に応じて適切なデータ圧縮を行えることも望ましく、一方で可能な限り正確に原データを復元できることが望まれる。

本稿では、多種多様なセンサーからのセンサーデータを時空間的に集約・圧縮し、リアルタイム遠隔センシングとモニタリングを効率よく実現するためのデータ集約圧縮技術を開発する。時間的類似性および空間的類似性の高いメタデータを有する複数のセンサーからのデータを時空間ブロックで集約することでメタデータ情報量の削減を実現し、かつ圧縮センシング技術 [1][2] を用いて圧縮率を適応的に制御することで、前述の要求条件を満足する。多くのセンシングデータは時刻に関するいくつかの情報（センシング時刻やデータ処理時刻など）に関する情報を有するため、許容できる精度範囲でメタデータ値を丸めて統一し、メタデータ数の削減を図る。同様の概念は空間座標に対しても

適用できる。これらを適用し時空間的に整列化したセンシングデータマトリクスから N 個の要素を取り出し、圧縮センシングにおける N 次元センシングデータ原信号とみなして事前に構築した圧縮用 $M \times N$ 行列により線形変換した M 次元データを送信する処理を繰り返す。圧縮センシングは、特定要素の出現頻度が高い（スパース性が高い） N 次元センシングデータに対して、高い圧縮率や復元率を実現できる特性を有している。これに対し、開発した技術では、時間的かつ空間的な類似度が高い部分データマトリクスごとに N 次元センシングデータとする仕組みを採用しているため、最頻値や平均値を特定要素とした圧縮センシングが有効に動作すると期待される。また次元数 M を動的に調整することで、圧縮率と復元率のトレードオフのもとで適応的に送信データ量を調整でき、通信容量の変化や劣悪な通信環境においても対応可能である。

家庭や集合住宅、オフィスビルや住宅地域などの拠点に設置された市販の温度、湿度、CO₂ センサーからの XML データを同様に拠点内に設置されたゲートウェイに集約し、データセンターなどにあるサーバへ圧縮送信する集約圧縮システムの設計開発を行った。豊田市に実在する 29 施設に開発したシステムを設置し、データ集約評価を行った結果、センサーから発生した XML データ 10MB に対してシステムによる集約・圧縮を行い、送信データを 10 数 KB 程度に削減でき、さらに時空間ブロックによる集約により約 50% 程度の削減を達成できている。

2. 関連研究

2.1 モバイルセンシングにおけるデータ集約

携帯端末等を用いてセンシングした様々なデータをクラウドサーバに集約し、環境センシングやモニタリングなどを行うモバイルセンシングアプリケーションが近年注目を集めている。これらは例えばスマートフォンや車載器といったモバイル端末がセンシングデータのゲートウェイもしくはセンサーそのものとなり、様々な地点でのデータを収集して解析する。これらの手法やシステムの多くは通信コストが無視できない 3G や LTE などの広域通信網の利用を前提としているため、データを集約・圧縮して通信コストの削減や端末の消費電力の削減等を実現する手法が提案されている [3]。このような環境では時々刻々と転送されるデータがメタデータを持っており、そのまま転送することは非効率である。これに対し [3] では、準同型暗号化等を用いて動的に時系列データの最小集合を得ることで通信コストを削減する方法を採用している。開発した手法は時々刻々と発生するデータを集約して送信する点はこれらの手法と類似するものの、メタデータそのものを削減し、できるだけ少ないメタデータを付与することでデータの削減を行う点が異なる。またこれらの方法論はセンシングデータ特性を仮定していないが、開発した手法ではデータの時間

的あるいは空間的類似性からデータのスパース性を導き、圧縮に有利なデータを構成するなどの工夫をしている。

2.2 多次元データベース集約

実時間データストリームを多次元データベースに蓄積し、必要なデータを抽出し、分析する研究が注目されている。データストリームはモバイルセンサーやプローブカーなどの情報生成源からの実時間データの流れであり、送信されるデータは複数のデータの組であることが多い。これは特にタプルストリームと呼ばれている。文献[4]では、大規模かつ継続的に送信されるタプルストリームデータを処理する方式として、圧縮チャンクをページバッファにまとめるコンテナ化の方式を提案している。チャンクとは、ストリームを部分配列に分割したもので、タプルストリームをチャンクに分割した後、それぞれのチャンクで圧縮処理を行い、ページバッファにまとめることで、効率的にストリームの送受信を行うことができる。提案手法では、温度、湿度、CO₂濃度と言った環境センサーのデータを扱っているが、これらのセンサーデータは、取得したセンサー名やセンサー精度、データ取得座標や頻度、時刻、プライバシーなどのメタデータを持っており、このメタデータを次元情報とすることでセンサーデータをタプルストリームとみなせるが、文献[4]をはじめとする多次元データベースでは基本的に集約済のデータの圧縮のみを扱うスキームを提案しており、本研究で行なっているような様々なセンサーからのデータ集約は主体的に考慮されていない。

2.3 圧縮センシング

圧縮センシングは Compressed sensing, Compressive sampling と呼ばれ、観測対象信号に対するスパース性や圧縮可能性のもとで、その信号に対する観測データができる限り少なく圧縮し、その少数のデータだけから観測信号を復元するための技術であり、センサー技術の急激な発展とともに注目を集めている [1], [2]。圧縮センシングは圧縮、復元の問題の理論的な側面のみならず、様々な応用への展開可能性を秘めている。一般的には、画像、音声、動画などのように冗長性を利用して圧縮されるマルチメディアデータに対して適用されることが多い。ここで、圧縮センシングにおけるスパース性とは、信号を適切な基底で表現した際、その表現形数のほとんどが0であることであり、圧縮可能性とは、画像など、DCT変換や離散ウェーブレット変換 [5] を適用した場合に大きな値をもつ少数の係数と、0と見なせるような非常に小さな値を持つ大多数の係数になる信号（すなわち周波数成分に偏りがあるデータ）のことを示している。

圧縮センシングは基本的に、圧縮過程および復元手法を共に設計する立場で議論されるが、復元手法を重点的に議論することも多い。その最もよく知られた手法が l_1 再構成

(l_1 ノルム最小化) であり、この原理が有効に機能するための圧縮過程に対する条件が示されている [6]。

圧縮センシングの理論的成果を実現した具体的な応用例として、単画素カメラと呼ばれるものがあげられる [7]。このカメラの最も重要な技術要素は、 N 画素の画像を撮像するために、 M 個 ($M < N$) のデータのみを検出すればよい点である。これを利用することで、4096 画素の画像をわずか 1600 個のデータから復元することができ、画素の 40% 以下のデータから元の画素が得られる。この他にも、圧縮センシングを天体観測画像に適用した例 [8] やネットワークへの活用 [9] など、様々な応用がなされている。

本実証システムでは、圧縮センシングを用いて、温度や湿度などのリアルタイムセンシングデータに対する圧縮および復元を行う。温度や湿度等のセンシングデータは画像等とは違い、時間および空間に対する変化量がそれほど大きくないため、マルチメディアデータに近い特性を有すると考えられ、最頻値などを基準にすればスパース性を有するデータとして扱うことができる。この特性を活用すれば少ないデータ欠損での圧縮が実現できると期待できる。また、通信環境に合わせて圧縮率を明示的に制御できるため、本実証が想定するような、必ずしも通信環境が整っていない遠隔地からのリモートセンシングに適した方法であるといえる。

3. 開発したセンシングデータ集約圧縮技術

本節では、家庭や店舗、集合住宅やオフィスなどの拠点に設置された多数のセンサーから、定期的に（例えば数秒から数十秒毎に）発生するセンシングデータを拠点のゲートウェイに集約し、ゲートウェイから無線または有線アクセス網を介して他のサーバ（クラウドサーバなど）に送信するような環境を想定し、集約されたセンシングデータを効率よく送信するための集約圧縮方式について述べる。各センサーはゲートウェイと常に有線または無線により通信可能であり、センシングデータは取得後速やかにゲートウェイに集約されているものとする。

3.1 メタデータ情報量削減手法

まず、多数のセンサーから発生するセンシングデータのメタデータの削減手法について述べる。各センシングデータは温度、湿度といったセンシングデータ値の他、そのデータを生成したセンサー名やシリアル番号、データ発生時刻やデータ処理時刻、データ生成座標やデータ種別といったメタデータが付随しており、センサーは数秒から数十秒程度のインターバルでデータを送信するものとするが、そのインターバルは一定である必要はない。現状の多くのセンサーは高頻度センシングや集約時のスケラビリティ、ゲートウェイにおけるデータバッファリングなどを考慮した設計になっておらず、センシングした値ごとに多くの付

随情報（メタデータ）が付与される。提案手法では、時空間的類似性および空間的類似性の高いメタデータを持つ複数のセンサーからのデータを時空間ブロックで集約することでメタデータ情報量の削減を目指す。

一定時間毎にあるセンサーから取得したセンシングデータは、そのセンサー名やシリアル番号などのセンサー情報とデータ種別により、それぞれにユニークな ID を付与し、ID とデータの時刻（タイムスタンプ）に基づき、データマトリクスと呼ばれるデータテーブルを構成する。なお、以下では議論の簡単のためデータマトリクスを 2 次元としているが、複数種のメタデータによる分類が必要であれば多次元データマトリクスにも容易に拡張可能である。例えば、センシング時刻のみならず、生成座標情報もメタデータとしては一般的であるため、これらを考慮した場合、データマトリクスは 3 次元となる。またデータマトリクスとは別に、各センサーの情報と ID との対応付けを記述したセンサー ID テーブルを圧縮側と復元側に持たせておくことで、センサー固有の属性情報を共有する。センサー ID テーブルに格納される情報はセンサー機器に関する情報であり、センシング期間を通じて変更等はほぼ生じないものとして

いる。以下にデータマトリクスの例を示す。表 1 は、午前 8 時 00 分 00 秒から午前 8 時 01 分 30 秒までの間に、ID.0 から ID.3 のセンサーから 30 秒毎にセンシングデータを取得した時の 2 次元データマトリクスの例である。ここで、“None” はセンサーエラーによるデータ損失やセンサー間のデータインターバルや取得タイミングの違い等によってその時刻およびセンサー ID に対応するデータが存在しないことを表す空要素である。

表 1 センシングデータのデータマトリクス例

時間 \ センサー ID	0	1	2	3
2013:07:03:08:00:00	125	125	250	252
2013:07:03:08:00:30	126	125	250	253
2013:07:03:08:01:00	126	126	251	None
2013:07:03:08:01:30	126	126	251	None

ゲートウェイでは各センサーからセンシングデータを集約し、表 1 で例示したようなデータマトリクスを構成した後、サーバへ送信するために時空間ブロックとよばれる単位でメタデータの集約を行ってメタデータ情報量の削減を行う。ここで時空間ブロックとは、上記のデータマトリクスの任意の部分マトリクスである。

まず、時空間ブロックの各次元において、精度に影響のない範囲でのメタデータインターバルを設定し、その固定インターバル値に適合するような丸めを行う。例えば 30 秒オーダー程度の精度で十分であるセンシングデータは、30 秒ごとの時刻への丸め処理を行う。これにより各次元

でのインターバルが一定でないデータマトリクスにおいても、一定インターバルの時空間ブロックに変換できる。一般的にデータマトリクスは各次元の精度が高いほど空要素（前例の“None”）数が増加するため、空データを適切に減少させ、かつデータの重なりがない丸め処理を行うことが適切である。この処理後に得られた n 個のセンサー ID と t 秒のデータ取得インターバルの時空間ブロックに対し、各インターバルにおいてセンサー ID. i のデータが存在するかどうかをビットで表したものを X_i 、初回インターバルの時刻を YYYYMMDDhhmmss（YYYY 年 MM 月 DD 日 hh 時 mm 分 ss 秒）とし、データ存在ビットが同じデータ列が初回インターバルの時刻から m 回繰り返されているとき、時空間ブロック全体のメタデータとして式 (1) のようなデータを生成する。これにより、各次元に対するメタデータ整列効果と、データ有無のビット列への圧縮効果によりメタデータの削減が可能となる。

$$(X_n \dots X_1 X_0)_{10}, \text{YYYYMMDDhhmmss}, t, m \quad (1)$$

例えば、表 1 の時空間ブロックに付加するメタ情報は以下のようになる。

15, 20131207080000, 30, 2,

14, 20131207080100, 30, 2,

提案手法ではこの時空間ブロックのメタデータ情報ならびに、この時空間ブロックのデータ要素列をデータ列として送信する動作を繰り返す。ここで、データ要素列に関しては、次節に述べる圧縮センシング手法を用いて更なるデータ圧縮を実施する。

3.2 圧縮センシングによるセンシングデータの圧縮と復元

圧縮センシング [1], [2], [10], [11], [12] は、その要素中にスパース性（零成分が多いがどこにそれが出現するかはわからないという性質）を持つ N 次元観測信号（センシングデータ）を、 N より次元数の低い信号に（一定の変換方法で）変換することで、次元圧縮によるデータ圧縮利得を得るとともに、スパース性を前提とした原値推定方法を採用することで高い確率で原値に復元する枠組みである。

本研究では、温度や湿度といったセンシングデータの時空間ブロック（以下では一貫性を失うことなくその元データであるデータマトリクスとも表現する）の要素データ列を信号とし、圧縮センシングによるデータ列の圧縮と復元を行う。データ数が N のデータマトリクスの要素データ列を、その列内の最頻値を 0 とした相対値で表現したものを N 次元の原信号ベクトル $\mathbf{x}_0 \in \mathbb{R}^N$ とする。最頻値からの差分を取ることで、最頻値部分を 0 とするために原信号ベクトル \mathbf{x}_0 は最大のスパース性を有することになる。以降、この原信号ベクトル \mathbf{x}_0 の非零要素の数が高々 K 個であるとしたとき、原信号ベクトル \mathbf{x}_0 は K -スパー

スであるということにする。圧縮側では、原信号ベクトルと $M \times N (M < N)$ の行列の積をとることで、より低次元 (M 次元) のベクトルを生成する。このときの $M \times N$ 行列を圧縮行列 A 、原信号ベクトルと圧縮行列の積によって生成された M 次元ベクトルを $\mathbf{y} \in \mathbb{R}^M$ としたとき

$$\mathbf{y} = A\mathbf{x}_0 \quad (2)$$

と表せる。復元側は、 \mathbf{y} と圧縮行列 A が既知である状態から未知の原信号ベクトル \mathbf{x}_0 を復元する。しかし、この問題を通常の連立方程式で解くと、条件式の個数 M よりも値を定めるべき変数の数 N の方が多く、解を一意に定めることが出来ないことは自明である。よって、条件式から一意な解を推定しなければならない。圧縮センシングでは、原信号ベクトル \mathbf{x}_0 が K -スパースであることが分かっていることを利用したいくつかの復元方式が提案されているが、提案手法では、 l_1 再構成 [1], [10], [13] を用いることで一意な解を推定する。以降、 l_1 再構成の詳細とその一例を述べ、また、後半で $M \times N$ の圧縮行列 A の決定方法についても述べる。

3.2.1 l_1 再構成による原信号復元

圧縮行列 A と圧縮後のベクトル $\mathbf{y} \in \mathbb{R}^M$ から、未知の原信号ベクトル $\mathbf{x}_0 \in \mathbb{R}^N$ の一意な解を復元するためには、線形制約 $\mathbf{y} = A\mathbf{x}$ を満たし、かつ K -スパースであるような \mathbf{x} を推定する問題を考える必要がある。本節では、 l_1 再構成を用いてこの問題を解決する手法について述べる。まず、 $p \leq 0$ に対する“ p -ノルム” ($\|\mathbf{x}\|_p$ で表す) を以下のように定義する。

$$\|\mathbf{x}\|_p = \begin{cases} \left(\sum_{i=1}^N |x_i|^p \right)^{1/p} & (p > 0) \\ \|\mathbf{x}\|_0 & (p = 0) \end{cases} \quad (3)$$

このとき、 $\|\mathbf{x}\|_0$ は \mathbf{x} の非零要素の個数である。そして、1-ノルム最適化問題

$$\hat{\mathbf{x}}^{(1)} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = A\mathbf{x} \quad (4)$$

を考える。すなわち、線形制約 $\mathbf{y} = A\mathbf{x}$ のもとで1-ノルムを最小化することによって、 \mathbf{x} の推定値を得る。この1-ノルム最適化問題が l_1 再構成である。また、補助関数として

$$\mathbf{t} = (t_1, t_2, \dots, t_N)^T \quad (5)$$

を導入し

$$\min \sum_{i=1}^N t_i \text{ subject to } -\mathbf{t} \leq \mathbf{x} \leq \mathbf{t}, \mathbf{y} = A\mathbf{x} \quad (6)$$

のように、線形計画問題として定式化することで、シンプレックス法や内点法といった一般的な解法で効率的に解くことが出来る。また、本研究では l_1 再構成により正しい推定結果を得る確率を向上させるための処理として、推定結果が原信号と異なる場合の推定値 \mathbf{x}' を用いて

$$\min \sum_{i=1}^N t_i \text{ subject to } -\mathbf{t} \leq \mathbf{x} \leq \mathbf{t}, \mathbf{y} = A\mathbf{x}, \mathbf{x} > \mathbf{x}' \quad (7)$$

という線形計画問題を値が一致するまで繰り返し解くことで、正しい推定値を得る確率を向上させる手法を採用した。このとき、推定結果の誤り検出用途として平均値もあわせて送信し、推定結果の平均値と比較することで、原信号との違いを判定するようにしている。

以上の l_1 再構成の一例として、 $N = 3, M = 2, K = 2$ の場合の原信号ベクトルの推定過程を以下に示す。まず、復元時に既知である \mathbf{y} と 2×3 の圧縮行列 A ($\text{rank} A = 2$) を

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad (8)$$

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{pmatrix} \quad (9)$$

とし、未知ベクトル

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (10)$$

を推定する問題を考える。このとき線形制約式 $\mathbf{y} = A\mathbf{x}$ は

$$\begin{cases} y_1 = a_1x_1 + b_1x_2 + c_1x_3 \\ y_2 = a_2x_1 + b_2x_2 + c_2x_3 \end{cases} \quad (11)$$

となり、三次元空間上の直線で表せる。また、 \mathbf{x} の1-ノルムは

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + |x_3| \quad (12)$$

となり、三次元空間上の原点を中心とした正八面体として表すことができる。正八面体の辺上は非零要素個数が2なので2-スパースであり、頂点は同様の理由から1-スパースであることがわかる。ここで原信号ベクトルが2-スパースであることを考えると、1-ノルムの最小化から推定される点は正八面体を徐々に大きくしていったときに、初めて式(11)の直線と接触したときの点となる。これを視覚的に表したものを図1に示す。推定値は黒点で、直線と交わっている正八面体の辺は太線でそれぞれ表される。また、この時の推定結果が違っていた時、線形計画問題により再度推定を行った時の図を図2に示す。違っていた推定値を赤点、更なる推定値を黒点で表している。このときの推定値は、図1の状態から正八面体をさらに拡大したときに、次に直線と正八面体の1辺が交わる点となる。

3.2.2 圧縮行列 A の決定

圧縮センシングにおける圧縮行列 A の決定は l_1 再構成が正しい推測結果を与えるかどうかに関係する。例えば、与えられた $M \times N$ の圧縮行列 A が、ある K -スパースな N

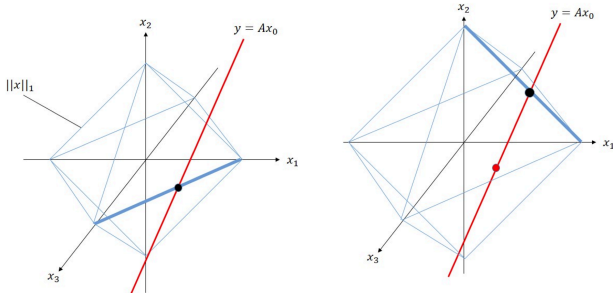


図 1 l_1 再構成の一例
($N = 3, M = 2, K = 2$)

図 2 図 1 の推定値が正しくない場合の 2 度目の l_1 再構成

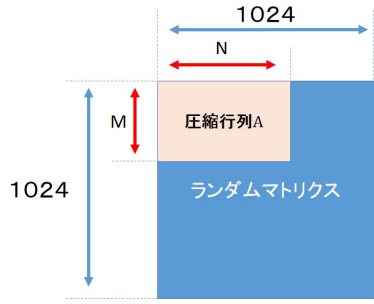


図 3 ランダムマトリクスからの圧縮行列 A の選出

次元ベクトル \mathbf{x} に対して $A\mathbf{x} = 0$ となってしまうと l_1 再構成は機能しない。そこで、適切な圧縮行列 A の性質を具体的に表現する概念として、制限等長性 (restricted isometry property; RIP) とよばれる性質が知られている [14]。制限等長性とは、 $1 \leq K \leq N$ に対して、不等式

$$(1 - \delta)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta)\|\mathbf{x}\|_2^2 \quad (13)$$

が任意の K -スパースなベクトル \mathbf{x} に対して成り立つような最小の δ を δ_K と定義した時、以下が成り立つことを指す。

定理 1. 圧縮行列 A に対して、 $\delta_{2K} < \sqrt{2} - 1$ が成り立つような $K \leq 1$ が存在するとき、任意の K -スパースな \mathbf{x}_0 に対して l_1 再構成は正しい推定結果を与える。

すなわち、任意の K -スパースな \mathbf{x} の大きさに対して $A\mathbf{x}$ の大きさがあまり変わらないような圧縮行列 A を選ぶことが出来れば、 l_1 再構成がうまく機能する。しかし、本研究で圧縮センシングを用いる環境では、原信号ベクトルの次元数 N も一様ではないため、最適な圧縮行列 A を選ぶのは難しい。そこで、本研究では A としてランダム行列を用いる。圧縮行列 A をランダム行列にした時の制限等長性について、文献 [15], [16] では以下の様な結果を得ている。

定理 2. $M \times N$ 行列 A の各要素を独立に平均 θ , 分散 $1/M$ の正規分布に従って定める。このとき、 $\alpha = M/N$ を有限として M, N を大きくする極限において、 $\rho = K/N$ がある値 $\rho_c(\alpha)$ より小さければ、 N に関して指数的に 1 に漸近する確率で $\delta_{2K} < \sqrt{2} - 1$ が成立する。

定理 1 と定理 2 から、圧縮行列にランダム行列を用いれば高い確率で正しい推定結果が得られることが結論されるため、提案手法でも高い復元率を規定できる。なお、提案手法では、様々な N, M の値に対応できるように、それらの値に対し十分大きい L を用いて予め生成しておいた $L \times L$ のランダムマトリクスを圧縮側および復元側で保持しておき、そこから図 3 のように $M \times N$ の圧縮行列 A を取り出すようにしている。

4. 実証構成概要

本研究では前章までに述べた手法をもとに、家庭や集合住宅、オフィスビルや住宅地域などの拠点に設置された複数のセンサーからのセンシングデータをゲートウェイに集約し、データセンターなどにあるサーバへ送信するセンサーデータ集約圧縮システムの設計開発を行っている。まず、拠点内の複数地点に温度、湿度、CO2 濃度を計測する小型のセンサー子機を設置し、無線通信により、数秒から数十秒のインターバルで計測データをセンサー親機に送信する。センサー親機は部屋やフロアに設置され、それぞれ子機と親機によるセンサーネットワークを形成する。親機に送信されたデータは、各拠点に設置されたゲートウェイに一定時間毎に送信される。ゲートウェイでは、集約されたセンシングデータに対して、3 節に示した技術によるデータ圧縮を行い、ネットワークを介してデータセンターへ送信する。データセンターへ送信されたデータは復元され、本研究で開発した温湿度・電力可視化ソフトウェアによる可視化、あるいは専用アプリケーションサーバによる可視化を行う。

開発したシステムは図 4 のように構成される。各モジュールの仕様と実現方法について以下に述べる。

4.1 集約・圧縮モジュール構成

センサー親機からゲートウェイサーバへ送信されるセンサーデータ XML ファイルを監視し、アップロードされる毎に前述の圧縮手法によりメタデータ情報量の削減と圧縮行列による実データの圧縮を行う。実行時にメタデータ削減のために構築するデータマトリクスのサイズと圧縮センシングによる圧縮率を指定し、原信号ベクトルの次元数 N と圧縮後のベクトルの次元数 M が決定され、ランダムマトリクスから $M \times N$ の圧縮行列を生成する。センサーから送信された XML データから、集約・圧縮を行った時の一例を図 5 に示す。

図 5 左がセンサーデータ XML ファイル、図 5 右が圧縮

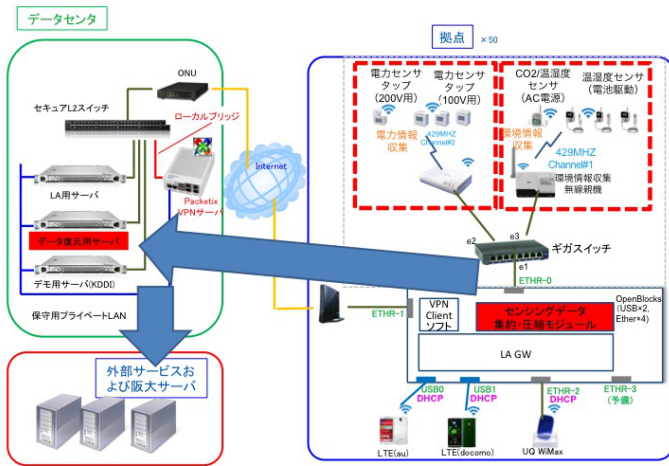


図 4 システムアーキテクチャの詳細

を待機し続ける。受け取ったデータは復元プログラムに渡される。クライアント側，サーバ側共に Python により実装している。復元モジュールはゲートウェイから送られてきたデータを l_1 再構成により復元し，削減したメタデータからデータマトリクスの各値に付加し直すことで，元の XML ファイルを生成する。 l_1 再構成の線形計画問題の解は Python のモジュールである OpenOpt と FuncDesigner を用いて得ている。復元プログラムは Ruby ベースで実装しており，Ruby 内で Python で実装した l_1 再構成を解くための関数を呼び出す。データマトリクスの情報から，原信号の要素数 N ，非零要素数 K ，最頻値および平均値を読み取り，提案手法による原信号の復元を行う。復元後に，実行時に指定したデータマトリクスのサイズからメタデータを復元し，各センサーデータに付加する。また，ゲートウェイ側と同様に，復元後のセンサーデータのログを記述した CSV ファイルを生成する。

可視化モジュールは収集されて CSV 化されたデータを現在から過去にわたって表示可能なソフトウェアであり，JavaScript で実装している。全拠点の電力のある時刻での利用状況を示すモードと，拠点毎に電力と温湿度 CO2 の時間変化を表示するモードがある。全拠点モードでは，複数のコンセントモニタの値を円グラフによる割合で示すことができ，特定の機器がどの程度電力を消費しているかをわかりやすく表示できる。拠点毎モードでは，表示時間幅を自由に指定することができ，詳細な値変化の確認や解析にも対応する。例えば始業時間と休憩時間の電力消費傾向の比較と，それにとまなう温湿度変化などを視覚的に解析し，特性を捉えることなども可能である。

各拠点に設置されたゲートウェイとデータセンター間通信には KDDI 研究所が有する CMT-SCTP (Concurrent Multipath Transfer Stream Control Transmission Protocol) を利用したリングアグリゲーション (LA) 通信システム [17], [18] が利用されている。LA は複数の物理的な回線を束ね，あたかも 1 本の回線であるかのように扱う技術であり，仮想的な利用帯域増加や冗長性向上を可能にする。実験環境では，複数のセンサーから大量の通信データが生成される状況を仮定しており，開発した集約圧縮技術においてそのデータ量を大幅に削減するものの，例えば回線断からの回線回復直後などではゲートウェイに蓄積したデータが一気にネットワークに流入する恐れもあり，アクセス回線等における輻輳を生じる可能性もある。開発した圧縮集約技術と LA を組み合わせることにより，FTTH 等の有線回線のみならず，LTE や WiFi 等の無線回線など利用可能な複数の回線リソースを活用し，通信容量を大きく占有することなく継続的に通信データをデータセンターに転送するスキームを実現することを目指している。

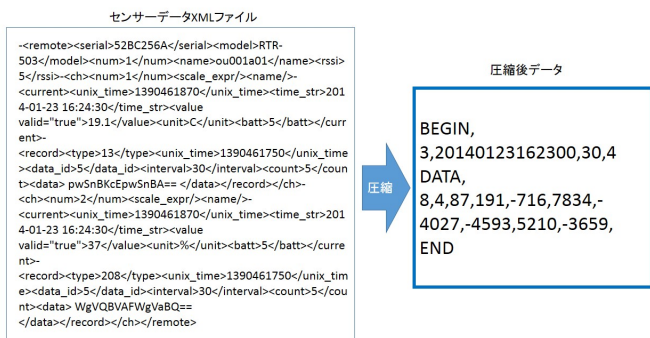


図 5 データ圧縮例

後のデータである。圧縮後のデータのフォーマットとして，BEGIN-DATA ブロック内にメタデータ情報量削減によってデータマトリクスに付加されたメタ情報，DATA-END ブロック内にデータマトリクスに対して圧縮を行った後のデータ列を記述している。復元側で必要な情報として，データ列の前に原信号のデータ数 N ，非零要素数 K ，最頻値，平均値を含めている。また，ゲートウェイ側で圧縮前の元データ列の確認等を行うため，XML タグを取り除き，センサー名やシリアル番号などの固有のメタ情報により ID を割り当てたものをセンサーデータログとした CSV ファイルを生成する。これらの実装はスクリプト言語 Python で行った。

同図中，クライアント側送信モジュールは圧縮後のデータをソケット通信によりデータセンターへ送信する。サーバ側の IP アドレスとポート番号を指定し，待ち受け状態のサーバ側受信プログラムへ接続し，データ送信を行う。通信断がある場合はそれを検知し，指定したインターバルでサーバへの再接続要求を継続する。サーバ側受信モジュールはクライアント側からの通信要求を待ち，各ゲートウェイから送信されたデータを受信する。クライアント側と同様，クライアントとの接続が切断されると接続要求

5. 実証結果

5.1 単体機器のセンシングデータを用いた集約圧縮効率の評価

設計開発したセンシングデータ集約圧縮システムの性能評価のための基礎的実験として、まず空調を常時起動させている室内に1台の温度・湿度計測センサーを設置し、3日間の計測を行ったXMLデータ約10MBに対し、システムによる圧縮と復元を行い圧縮率と復元率の評価を行った。各センサーは30秒毎の計測データを1分毎に親機からゲートウェイへ送信し、一度に送信するデータマトリクスの要素数 N は12, 15, 20, 30とした。このとき、 N の各値に対し t_1 再構成の圧縮率を0.9, 0.8, 0.7, 0.6倍と設定した場合の復元成功要素数ならびに圧縮後の容量を算出している。なお、復元成功率を確かめるため、原信号データは復元側でも記録した。

これに対し、まず開発技術によるメタデータ情報量削減量の評価を行う。各センサーデータはセンサー名などのセンサー毎に固有の特性、データ種別、データ発生時間をメタデータとして有するものとし、センサーから送信されたオリジナルのXMLデータから、XML表現による容量オーバーヘッド分を削減したデータテーブル(表2)と、そのデータを更に時空間ブロックで集約したもの(図6)の容量を比較した。

表2 XML表現を取り除いたデータテーブル例

年	月	日	時	分	秒	ID	計測値
2014	2	12	13	56	0	0	225
2014	2	12	13	56	0	1	190
2014	2	12	13	56	30	0	225
2014	2	12	13	56	30	1	190
2014	2	12	13	57	0	0	221
2014	2	12	13	57	0	1	200
2014	2	12	13	57	30	0	221
2014	2	12	13	57	30	1	200
...

```
BEGIN,
1111,20131213080100,30,2,
1110,20131213080200,30,1,
DATA,
250,260,250,265,250,260,
252,260,248,257,250,
END
```

図6 時空間集約後のデータ例

N の各値に対するそれらの値を比較したものを図7に示す。一度に送信するデータマトリクスの要素数の増加に伴い、メタデータ情報の削減量も増加するという結果が得られた。また、センサー数を増やした場合でも、データマト

リクスの要素数は増加するため、メタデータ情報の削減量が向上すると考えられる。全体的なデータの削減率に関しては、元のXMLデータ約10MBに対し送信データが10数KB程度に削減されており、時空間ブロックによる集約に限れば50%程度の削減ができています。これらのことから実用上十分な削減率を達成できているといえる。

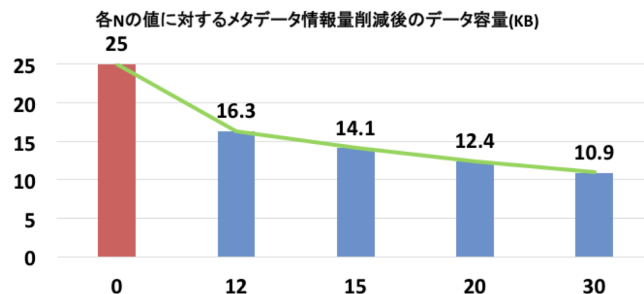


図7 時空間ブロックへの集約時のデータサイズ

次に、データマトリクスの要素に対して、開発技術による圧縮センシングを行い、実データ数の圧縮と復元を行った。まず、復元についての結果を表3に示す。

表3 圧縮センシングにおける復元率比較

N	M	原信号要素数	復元成功要素数	復元率
12	10	2112	1998	94.6%
12	9	2112	1996	94.5%
12	8	2112	1951	92.4%
12	7	2112	1923	91.1%
15	13	2100	1965	93.6%
15	12	2100	1947	92.7%
15	11	2100	1926	91.7%
15	9	2100	1908	90.0%
20	18	2120	1941	91.6%
20	16	2120	1872	88.3%
20	14	2120	1840	86.8%
20	12	2120	1839	86.7%
30	27	2100	1791	85.3%
30	24	2100	1786	85.0%
30	21	2100	1760	83.8%
30	18	2100	1786	85.0%

本評価で用いたセンサーデータの全要素数は2120個であり、センサーデータはデータマトリクス毎に送信され、指定したデータマトリクスの要素数に満たない場合は、次のセンサーデータの受信を待つ仕様となっている。したがって、原信号要素数は2120に一番近い N の倍数になっている。図の結果から、 N を大きくすれば、それに伴って復元率が落ちることがわかった。一方で、 N の値が小さいほどデータ送信のリアルタイム性は向上するものの、短い間隔でデータセンターへの送信を行うことになるためにメタデータ情報量が増加する。こういったトレードオフを

鑑み、本実験での N は 12 から 15 前後が適しているといえる。

最後に、表 4 に圧縮率と l_1 再構成についての結果を示す。完全復元回数は l_1 再構成回数に対して、要素が小さい誤差で全て復元された回数を示している。今回の評価実験で設定した誤差は ± 0.3 である。また、誤差が設定した範囲を超えていた場合でも、ほぼ全ての要素が ± 1.0 で復元できているような結果が得られた。元の温湿度データは 10 倍値を送信しているため、誤差値 1 が与える影響は例えば気温においては $\pm 0.1^\circ\text{C}$ と極めて小さい。復元時の僅かな誤差を許容することで、3/4 程度までのデータ圧縮が可能で、圧縮率の制御も可能である利点を楽しむことができる。

表 4 圧縮センシングにおける l_1 再構成と圧縮率比較

N	M	l_1 再構成回数	完全復元回数	容量 (KB)	圧縮率
12	10	176	142	13.4	82.2%
12	9	176	147	13.1	80.4%
12	8	176	142	12.8	78.5%
12	7	176	142	12.4	76.1%
15	13	140	113	12.0	85.1%
15	12	140	107	11.8	83.7%
15	11	140	103	11.4	80.9%
15	9	140	98	11.1	78.7%
20	18	106	79	11.2	90.3%
20	16	106	68	10.8	87.1%
20	14	106	63	10.3	83.1%
20	12	106	64	9.7	78.2%
30	27	70	40	9.5	87.2%
30	24	70	39	9.3	85.3%
30	21	70	37	8.5	78.0%
30	18	70	39	8.1	74.3%

5.2 実験実証データの解析と結果

実証実験を実施した 1 拠点の 2014 年 3 月 15 日から 21 日までの 6 日間に取得した温度、湿度、 CO_2 のデータを利用した集約圧縮効果を確認した。拠点には温度、湿度を計測するセンサー子機を 3 台、温度、湿度、 CO_2 濃度を計測するセンサー子機を 3 台の計 6 台を設置した。したがってセンサー総数は 15 となる。またそれらの機器が接続された親機を 1 台設置し、XML データを 5 分に 1 度、拠点のゲートウェイに出力する。センシング間隔は 30 秒とした。このもとでの評価結果を表 5 に示す。

表 5 実証実験における集約圧縮効率

メタデータ削減率 (XML 削減率も含む)	計測データ 圧縮率	完全復元率	復元時 誤差
0.49MB/11.58MB (4.2%)	108KB/150KB (72%)	34940/39639 (88.1%)	0.12

この結果はおおむね前節の結果と合致しており、実際の環境でも同様の傾向がみられることが確認できている。また、収集したデータの可視化をした結果を図 8 に示す。長時間で見るとばらつきはあるものの、比較的短時間、および空間的には概ね同じ値に偏向する傾向があり、これらに対し開発した集約圧縮手法を用いることでその効果が確認できることがわかる。

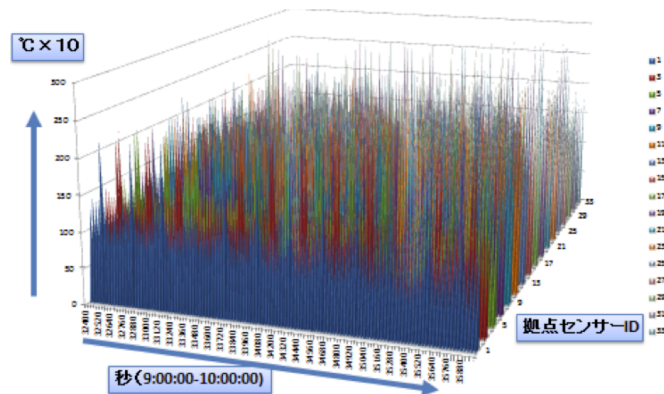


図 8 実証実験における時空間データブロックの可視化 (2014 年 3 月 24 日午前 9 時時点で稼働していた全拠点からの 1 時間分の温度データの集約効果)

最後に、大阪大学開発の可視化ソフトウェアを用いて、全拠点の電力量使用データと、拠点ごとの電力、温度、湿度、 CO_2 データの可視化を行った結果を示す。図 9 の各円グラフが拠点に対応し、拠点における各スマートエコワット (4 台) から得られた電力量の比を円グラフの比率で表している。ここでは拠点 1,2,3,6,8,9,10,12,13,18 を可視化した様子を示している。また、図 10 は拠点における上述の電力比率に加え、電力、温度、湿度の時間変化をグラフ化した様子を示している。これを用いることで電力と温湿度の関係なども容易に可視化でき、例えばエアコンの電力使用量と温湿度の相関関係などが直感的に把握でき、スマートグリッドやスマートハウスにおける電力把握やデマンドレスポンスへの活用が期待できる。

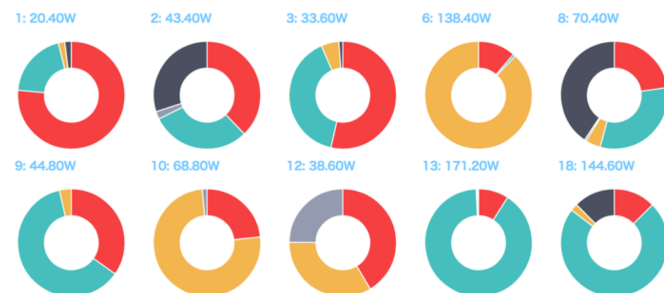


図 9 大阪大学開発の可視化ソフトウェアの実行状況。全拠点の電力表示モード

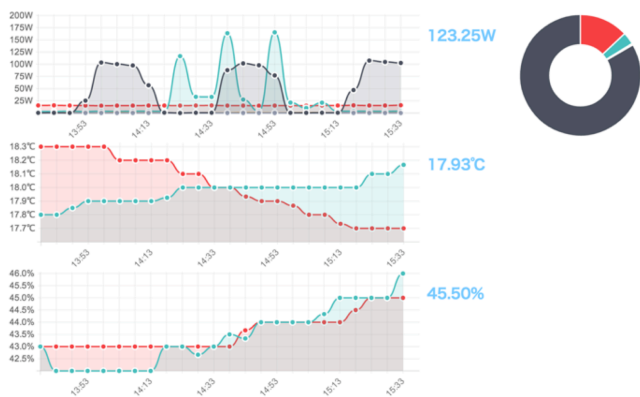


図 10 拠点毎の複数センサー値表示モード

5.3 多拠点のセンサーデータを用いた実証実験データの解析と結果

実証実験を実施した 29 拠点 201 センサー（拠点 1,2,3 には温度、湿度を計測するセンサー子機を 3 台と温度、湿度、CO2 濃度を計測するセンサー子機を 3 台の計 6 台、その他の拠点には温度、湿度を計測するセンサー子機を 3 台設置、総センサー数 201）の 2014 年 3 月 25 日午前 9 時から午前 10 時の 1 時間で計測した温度、湿度、CO2 のデータを利用した集約圧縮効果を確認した。拠点毎に、機器と接続された親機を 1 台ずつ設置し、XML データを 5 分に 1 度、拠点のゲートウェイに出力する。センシング間隔は 30 秒とした。このもとでの評価結果を表 6 に示す。また、1 度に送信するデータマトリクスの要素数 N は 20、圧縮センシングによる圧縮率は 0.6 倍とした。

表 6 多拠点のセンサーデータを用いた実証実験における集約圧縮効率

メタデータ削減率 (XML 削減率も含む)	計測データ 圧縮率	完全復元率	復元時 誤差
0.26MB/1.53MB (16.9%)	32KB/43KB (74%)	9616/10212 (94.1%)	0.12

この結果と 1 拠点での実証実験の結果（表 5）を比較すると、メタデータ削減率が若干低下していることが分かる。これは、センサー数増加に伴い、センサー ID の存在を表すビット列が増大した影響と考えられるが、現実的にはほぼ影響のない大きさであることがわかる。また、計測データ、完全圧縮率、復元時誤差はおおむね一致していることが確かめられており、多数の実センサーデータにおける効果が確認できている。

6. まとめ

空間的に多数配置された多種多様なセンサーからのセンサーデータを時空間的に集約・圧縮し、リアルタイム遠隔センシングとモニタリングを効率よく実現するためのデータ集約圧縮技術を提案した。提案手法では、時間的類似性、

空間的類似性が高いメタデータを有する複数のセンサーからのデータを時空間ブロックで集約することでメタデータ情報量の削減を実現し、かつ圧縮センシング技術を用いて圧縮率を適応的に制御する。提案手法に基づくセンサーデータ集約圧縮システムの設計開発を行い、豊田市に実在する 29 施設で実証実験を展開した。以下に本研究で得られた知見をまとめる。

- (1) メタデータ情報量削減量の評価を行った。図 3 に基づくと、一度に送信するデータマトリクスの要素数の増加に伴い、メタデータ情報の削減量も増加するという結果が得られた。また、センサー数を増やした場合でも、データマトリクスの要素数は増加するため、メタデータ情報の削減量が向上すると考えられる。全体的なデータの削減率に関しては、元の XML データ約 10MB に対し送信データが 10 数 KB までに削減されており、時空間ブロックによる集約に限れば 50%程度の削減ができています。これらより、メタデータの種類や性質にも依存するものの、実用上十分な削減率を達成できているといえる。
- (2) データマトリクスの要素に対して、開発技術による圧縮センシングを行い、実データ数の圧縮と復元を行った。復元についての結果（図 3）に基づく、 N を大きくすれば、それに伴って復元率が落ちることがわかった。一方で、 N の値が小さいほどデータ送信のリアルタイム性は向上するものの、短い間隔でデータセンターへの送信を行うことになるためにメタデータ情報量が増加する。こういったトレードオフを鑑み N は 12 から 15 前後が適しているといえる。
- (3) 図 4 に示した圧縮率と l_1 再構成についての結果から、ほぼ全ての要素が高々 ± 1.0 で復元できた。な結果が得られた。元の温湿度データは 10 倍値を送信しているため、誤差値 1 が与える影響は例えば気温においては ± 0.1 °C と極めて小さい。復元時の僅かな誤差を許容することで、3/4 程度までのデータ圧縮が可能で、圧縮率の制御も可能である利点を享受することができている。

今後は開発したシステムをもとに、回線切断・回復といった状況の発生時におけるシステムの性能評価、ならびに圧縮率を適応的に調整する機構の研究を行い、より柔軟性の高い方式へと拡張していきたいと考えている。

謝辞

本研究は、総務省の”スマートグリッドの通信ネットワーク技術高度化実証事業”の成果であり、関係者各位に深謝する。

参考文献

- [1] 田中 利幸, “圧縮センシングの数理,” *IEICE Fundamentals Review*, vol. 4, no. 1, pp. 39–47, 2010.
- [2] 三村 和史, “圧縮センシング: 疎情報の再構成とそのアルゴリズム (時間周波数解析の理論とその理工学的応用),” *Research Institute for Mathematical Sciences, Kyoto University*, vol. 1803, pp. 26–56, 2012.
- [3] Q. Li and G. Cao, “Efficient and privacy-preserving data aggregation in mobile sensing,” *Proc. of 2012 20th IEEE International Conference on Network Protocols (ICNP)*, pp. 1–10, 2012.
- [4] 嶋田 鉄兵, “多次元データの格納と処理方式に関する研究,” 福井大学 博士学位論文, 2011.
- [5] T.-Y. Lee and H.-W. Shen, “Efficient local statistical analysis via integral histograms with discrete wavelet transform,” *Proc. of 2013 IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2693–2702, 2013.
- [6] P. Indyk and M. Ruzic, “Near-optimal sparse recovery in the l_1 norm,” *Proc. of 2008 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 199–207, 2008.
- [7] D. Takhar, V. Bansal, M. Wakin, M. Duarte, D. Baron, K. Kelly, and R. Baraniuk, “A compressed sensing camera: New theory and an implementation using digital micromirrors,” *Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose*, 2006.
- [8] J. Bobin, J.-L. Starck, and R. Ottensamer, “Compressed sensing in astronomy,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 718–726, 2008.
- [9] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, “Compressed sensing for networked data,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 92–101, 2008.
- [10] 田中 利幸, “圧縮センシング: 基本原理とその応用,” *IEICE. WBS*, vol. 110, no. 222, pp. 25–30, 2010.
- [11] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [12] 芦野 隆一, “圧縮センシングの基礎とその研究動向,” *ISCIE*, vol. 55, no. 3, pp. 88–93, 2011.
- [13] 和田 山正, “圧縮センシングにおける完全再現十分条件について,” *The Brain & neural networks*, vol. 17, no. 2, pp. 63–69, 2010.
- [14] E. J. Candes, “The restricted isometry property and its implications for compressed sensing,” *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.
- [15] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [16] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Sbornik: Mathematics*, vol. 1, no. 4, pp. 457–483, 1967.
- [17] 立花 篤男, 長谷川 輝之, “汎用 android スマートフォンを用いた CMT-SCTP スキームの提案,” 電子情報通信学会 技術報告, vol. 112, no. 218, pp. 65–70, 2012.
- [18] R. R. Stewart and Q. Xie, “Stream control transmission protocol (sctp),” 2001.
- [19] 宵 憲治, 山口 弘純, 廣森 聡仁, 内山 彰, 東野 輝夫, 柳谷 尚寿, 中谷 俊和, 立花, 篤男, “多数のセンサーによる時空間センシングデータの集約圧縮方式の一検討,” 電子情報通信学会 2014 年総合大会予稿集, pp. B15–15, 2014.