

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

柳 沢 豊^{†1} 岸 野 泰 恵^{†1}
前 川 卓 也^{†1} 須 山 敬 之^{†1}

気象データのように、観測されるデータの間に関連性があるセンサネットワークでは、Slepian-Wolf Coding と Shortest-Path Tree を使った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できることが知られている。この、Slepian-Wolf Coding の実装法としては、LDPC 符号を用いる方法がある。しかし LDPC 符号を用いる場合、符号化に用いる行列式の情報をセンサノードとシンクノードで共有する必要があり、行列式の情報を伝送すると大きな通信コストがかかる。そこで本研究では、サイズの小さな基礎行列を対角線上にタイルパターンのように配置して行列を生成した行列を使う、LDPC-CC 符号を用いる。さらに、行列式の生成を擬似乱数を利用して少量のパラメータで生成できるようにすることで、センサノードとシンクノードの間で伝送する行列式に関する情報量を抑えた。

Data Gathering in Wireless Sensor Networks using Correlation of Sensor Data

, YUTAKA YANAGISAWA,^{†1} YASUE KISHINO,^{†1}
TAKUYA MAEKAWA^{†1} and TAKAYUKI SUYAMA^{†1}

The combination of Slepian-Wolf Coding and Shortest-Path Tree routing can minimize the communication cost of the correlated data gathering on wireless sensor networks. LDPC (Low Density Parity Check) code is one of the most popular encoding methods for Slepian-Wolf Coding. The sink node and a sensor node, however, must share large matrix data for encoding and decoding. To share the matrix data between these nodes, the sensor node must send large data to the sink node. To solve this problem, we introduce LDPC-CC whose matrix data can be generated with a small base matrix. Moreover, we propose the algorithm to generate the base matrix with several parameters.

1. はじめに

本稿では、温度や湿度などの環境データを観測する無線センサネットワークに適した、データの圧縮方式について提案する。対象とする無線センサネットワークは、多数の観測地点に配置された無線通信機能をもつセンサが、観測したデータをマルチホップ通信によって、特定のノード（シンクノード）に集約するネットワークのことである。こうしたセンサネットワークは、環境のモニタリングや、動植物の生態調査などに用いられている^{1),2)}。一般に、こうした無線センサネットワークを構成するセンサノードは、バッテリーの搭載量が少ない。各センサノードの消費電力を可能な限り抑え、ネットワークのライフタイムを伸ばすことは、無線ネットワークの研究分野における重要な課題の1つとなっている。

消費電力を抑える最も効果的な方法は、無線通信にかかる消費電力、すなわち通信コストを抑えることである³⁾。無線を用いる場合の通信コストは、一般に送受信データ量に比例し、ノード間の距離の二乗に比例する。つまり、通信コストを下げるためには、通信経路を設計するときに各センサノードがデータの送受信を行う距離を短くすることと、データの符号化技術や集約技術を用いて送受信データ量を削減するという2つのアプローチが必要になる。

この課題に対して、Cristescuら⁴⁾、Liuら⁵⁾は Networked Slepian-Wolf Coding (NSWC) と呼ばれる1つの解決法を提示している。NSWCは、Slepian-Wolf Coding (SWC)⁶⁾と Shortest Path Tree (SPT) を組み合わせた通信路の設計手法である。各センサノードが観測するデータに相関性があるときに、NSWCを用いることで、ネットワーク全体の通信コストを最小化できることがLiuらにより示されている。ここでいうデータの相関性とは、各ノードで観測するデータが一定の確率分布に従うとき、ある二つ以上のノード v_1, v_2 が観測するデータの確率分布 $p(v_1), p(v_2)$ と結合確率 $p(v_1, v_2)$ について $p(v_1) \cdot p(v_2) > p(v_1, v_2)$ が成り立つことを言う。本稿で扱う相関度は、相互情報量 $I(v_1; v_2)$ と同じ意味である。相関性があるということは、すなわち $I(v_1; v_2) > 0$ であることを意味する。例えば、屋外で雨量や温度、湿度といった気象データを収集する場合、各センサノードで観測されるデータ間には相関性が存在することが多い。NSWCを用いると、こうした相関性のあるデータを効率的に集約することができる。

NSWC で設計した通信路の通信レートを実現するデータの圧縮手法として、低密度パリティ

^{†1} 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

ディ検査 (LDPC) 符号^{7),8)}がある。LDPC 符号を使うことで、NSWC で与えられる理論的な下限に近い通信レートまでデータを圧縮できることが知られている。しかし、実際の無線センサネットワークで LDPC 符号を用いたデータ圧縮手法を利用するにあたり、解決すべき二つの問題がある。

ひとつめの問題は、符号化に使う行列式の情報をデータの情報量の合わせて作り直し、なおかつ行列式の情報をシンクノードとセンサノードで共有する必要がある点である。LDPC 符号でデータの圧縮率を最小化するためには、符号化に使用する行列式の縦横のサイズを、観測されたデータ系列の情報量に合わせて動的に変更する必要がある。また、復号時にも符号化時に使用した行列式の情報が必要となる。このため、センサノードが符号化時に使用した行列の情報量を、シンクノードとの間で共有する必要がある。このためには、センサノードとシンクノードの間で行列式の情報を伝送する必要がある。しかし、LDPC 符号の行列式のデータサイズは、圧縮するデータのサイズより必ず大きくなる。このため、行列式の情報量を圧縮し、シンクノードとセンサノードの間の伝送コストを大きく下げることが必要である。本稿では、この方法として LDPC-CC 符号⁹⁾の導入と、数個のパラメータから行列式を生成するアルゴリズムの導入を行う。

ふたつめの問題は、行列式のサイズを厳密に求めるためには、シンクノードがすべての観測データが必要である点である。つまり、センサで観測されたデータがシンクノードに集める前に、シンクノードがすべてのデータの内容を知っている必要があることを意味する。これは明らかに目的と矛盾しており、厳密にこれを実現することは不可能である。そこで、厳密な行列式のサイズを計算するのではなく、予測によってサイズを推定する方法を導入する。さらに、サイズの予測が実測値を下回り復号に失敗したときに、不足する通信レートに相当する符号語を作成しなおし、再送する手順を追加する。これらの手法の導入により、行列式の厳密なサイズに求めなくとも、データの圧縮性能を良好に保つことができる。

以上の方法を導入することで、センサネットワーク上で LDPC 符号を用いて、データの相関を利用したデータ圧縮を実現することができる。筆者らは、提案手法を実機上に実装し、少ないパラメータから生成した LDPC-CC 行列を用い、取得した温度データに対して符号化処理が行えることを確認した。本稿では、実機での動作の様子についても紹介する。

2. 相関性を利用したデータ集約

本章では、観測されるデータに相関性がある無線ネットワークにおいて、通信コストを最小化できるネットワークの設計方法である Networked Slepian-Wolf Coding について説

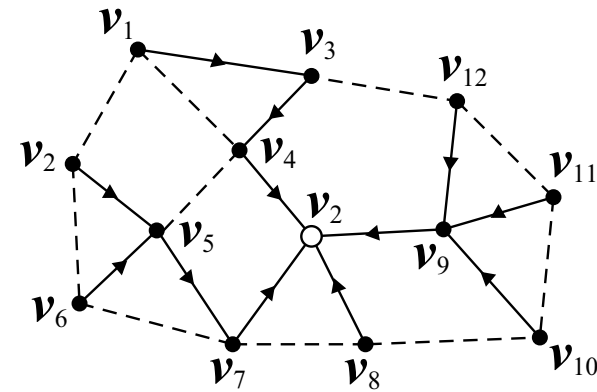


図 1 対象とする無線センサネットワーク
 Fig. 1 Wireless Sensor Network

明する。まず 2.1 節で本稿が対象とする無線センサネットワークについて説明する。次に 2.2 節でデータ集約コストを最小にできるネットワークの設計法である NSWC について述べる。そして 2.3 節では本稿に関する関連研究について触れる。

2.1 無線センサネットワーク

図 1 に示すに n 台のセンサノード (source node) v_1, \dots, v_n とセンサデータの集約拠点である 1 台のシンクノード (sink node) v_t という $n + 1$ 個のノードで構成される無線センサネットワークを考える。 n 個のセンサノードとシンクノードの集合を $V = \{v_1, \dots, v_n, v_t\}$ 、通信可能なノードの対 v_i, v_j を結ぶ枝 (通信経路) を $e(i, j) \in E$ とすると、この無線センサネットワークは $n + 1$ 個のノードをもつグラフ $G(V, E)$ とみなせる。シンクノードもセンサノードのひとつであるため、以下特に断らない限り、センサノードと称するときはシンクノードも含むものとする。各センサノードは、センサを用いて降水量や温度などのセンサデータを取得する機能、そのデータを符号化する機能、および近隣のノードと無線通信する機能を持ち、定期的にデータを観測する。各ノードの無線通信機能に性能差はなく、通信可能な距離の上限は同じであると想定する。各センサノードからシンクノードへ直接通信が行えない場合は、他のセンサノードを中継してデータを送る。このとき、無線センサネットワークは分断されておらず、シンクノードから全てのセンサノードに到達できる通信経路があるものとする。ノードの位置は移動しないものとし、通信にかかる時間 (遅延)

やスループット、データ送受信の失敗については本稿では考慮しない。

各ノード $v_i \in V$ で観測されるセンサデータは、一定時間毎にサンプリングされた離散的な数値列として表す。この数値列は、一定の確率分布に従う無記憶で離散的な情報源 X_i が発生させるものと見なす。このとき、 X_1, \dots, X_n の結合ベクトル $X = \{X_1, \dots, X_n\}$ は、同時確率分布 $p(X_1 = x_1, \dots, X_n = x_n) = p(x_1, \dots, x_n)$ と表せる。時刻を τ 、 $X(\tau)$ を時刻 τ にすべてのノードで同時に観測された観測値の結合ベクトルとすると、確率過程 $\{X(\tau)\}_{\tau=1}^{\infty}$ は定常過程である。なお、本稿ではデータの空間的な（ノード間の）相関性のみを議論の焦点を絞るため、 X_1, \dots, X_n は独立同分布 (i.i.d) であるとする。

次に、データ集約時の通信コストについて定義する。本稿で扱う通信コストとは、通信にかかる電力量のことである。一般に、無線通信によってデータを送受信するために必要となる電力量は、通信距離の2乗に比例し、通信するデータの長さに比例する。

二つのノード v_i, v_j を結ぶ通信経路 $e(i, j)$ を流れるデータの、1秒あたりの平均情報量 (bit) を $b(i, j)$ 、通信経路 $e(i, j)$ の距離を $len(e(i, j))$ 、通信経路 $e(i, j)$ で要する通信コストを $r(i, j)$ (以降、簡略化のため、 $e(i, j), b(i, j), r(i, j)$ をそれぞれ e, b, r と表す) と表記する。このとき、通信経路 e でデータの送受信に要する通信コスト r は $r = \alpha \times b \times len^2(e)$ で表せる。ここで α はノードの使用する通信手法および、ノードを構成するハードウェア等に依存する定数である。ネットワークを構成するノードのハードウェアおよび通信手法がすべて同種の場合、この定数 α はどのノード間においても同じ値であるとみなせる。このため、本稿では記述の簡素化のために通信コスト r は単に $r = b \times len^2(e)$ で与えられるものとする。

次に、ネットワーク全体の通信コストについて定義する。まず、グラフ $G(V, E)$ の各通信経路 $e \in E$ に関する通信コスト r の集合を W とする。シンクノード v_t との間に直接通信経路を持たないセンサノード v_i は、他のセンサノードを経由してデータを送信する必要がある。このとき、ノード v_i からシンクノードまで中継される経路の距離の二乗和を d_i と記述する。例えば、 v_i が v_t と直接通信する場合は、 $d_i = len^2(e(i, t))$ である。ノード v_j を経由する場合は、 $d_i = len^2(e(i, j)) + len^2(e(j, t))$ である。このとき、このネットワークにおける通信コストの総計 W_G は次式で表せる。

$$W_G = \sum_{r \in R} r = \sum_{i=1}^n b \cdot d_i$$

この W_G を最小化できるようにネットワークを設計することで、無線センサネットワー

ク全体の消費電力を最小化することができる。

2.2 Networked Slepian-Wolf Coding

Networked Slepian-Wolf Coding (NSWC)^{4),5)} は、相関性のあるデータを観測するネットワーク $G(V, E)$ において、データ集約時の通信コスト W_G を最小化できるネットワークの設計手法である。通信コスト W_G を小さくするには、次の2つのアプローチが必要となる。

- 観測されたデータを符号化してデータの長さを圧縮し、通信経路を流れるデータの平均ビット数 b を小さくする。
- 各センサノードからシンクノードまでの通信経路を決める際に、通信経路の距離 len の総和を可能な限り小さくする。

NSWC は、通信経路として Shortest Path Tree (SPT)、通信路として SWC を併用するネットワークの設計方式である。この方法を用いてネットワークを設計することで、通信コスト W_G を最小化できる。この証明は、Liu らによって示されている⁵⁾ ため、以下では NSWC の概略についてのみ述べる。

まず、Slepian-Wolf Coding (SWC)⁶⁾ について説明する。SWC とは、図2のように2つの無記憶情報源 X_1, X_2 に対して、それぞれ個別の符号器で符号化し、それら符号を1つの復号器で元の2つのデータ列に復元する伝送システムのことである。この伝送路では、情報源 X_1, X_2 に対して最適な符号化を行うことで、伝送路の帯域の下限は結合エントロピーと等しくなることが知られている。つまり、情報源 X_1 のエントロピー (平均情報量) を $H(X_1)$ 、 X_1, X_2 の結合エントロピーを $H(X_1, X_2)$ とし、 X_1, X_2 に対する伝送路の帯域を R_1, R_2 としたとき、 $R_1 + R_2 \geq H(X_1, X_2)$ となるように帯域を設計することで、理論的にはすべてのデータをロスなく伝送できる。情報源 X_1, X_2 の間に相関性がある場合、 $H(X_1) + H(X_2) > H(X_1, X_2)$ が成り立つので、 X_1, X_2 に対して独立した伝送路を用意する場合 (すなわち $R_1 \geq H(X_1), R_2 = H(X_2)$ とする場合) よりも、少ない帯域でデータ列を伝送できる。

Cristescu ら⁴⁾ は、情報源のデータ間の相関性があらかじめ既知であれば、他の情報源のデータを使用せずに各ノードで観測した情報だけを用いて符号化できる (符号化の際にノード間で通信する必要がない) こと、 n 個の無記憶情報源 X_1, X_2, \dots, X_n に対しても最適な符号化を行う場合、エントロピーの総和が結合エントロピー $H(X_1, X_2, \dots, X_n)$ と等しく可以利用したネットワークの設計法として、NSWC を提案した。さらに Liu らは NSWC を用いることで、通信コストの総和を最小化できることを示した。NSWC では、最

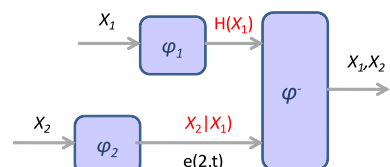


図2 ノードが2点の場合の Slepian-Wolf Coding

Fig. 2 Slepian-Wolf Coding for two nodes

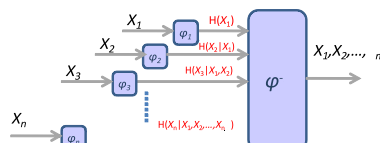


図3 ノードが n 点の場合の Slepian-Wolf Coding

Fig. 3 Slepian-Wolf Coding for n nodes

短経路木を通信経路とする無線通信ネットワーク上において、相関性を利用して各ノードで生成した符号をシンクノードまで送信する。つまり、グラフ G のシンクノード v_t を根とする SPT をなす部分グラフを $G_{SPT}(V, E_{SPT})$ とし、 n 個センサノード v_1, \dots, v_n が、 G_{SPT} 上で v_t への通信経路の距離が短い順にソートされている場合、各センサノード v_i において、 $R_i \geq H(X_i|X_{i-1}, \dots, X_1)$ (図3) となるように帯域を設計する。このように設計すると、通信コストの総和 W_G の下限は次式で与えられる。

$$W_G \geq \sum_{i=1}^n H(X_i|X_{i-1}, \dots, X_1) \times d_i \quad (1)$$

この式で与えられる W_G が、ネットワーク G における通信コストの理論的な最小値である。すなわち、通信経路として SPT を用い、情報源間の既知の相関性を利用して最適な符号化を行えば、通信コストは W_G まで低減できる。

2.3 関連研究

本稿では、NSWC に基づきデータ集約を行う手法について述べるが、NSWC 以外にもデータ集約時の通信コストを低減しようとする研究は、これまでに多数行われている。

例えば、Wavelet 変換等の信号処理に用いられる手法を利用してデータを圧縮する方式^{10),11)}、一定観測されるデータが疎であることを利用してデータの圧縮を行う Compressive Sensing Theory¹²⁾ に基づく集約方式^{13),14)} がある。これらは、単独のノードで観測したデータの時間的な相関性を利用したり、シンクノードまでの通信経路上にある他のノードとのデータの相関性を利用してデータを圧縮している。一方で、NSWC は時間的な相関性を考慮せず、ネットワーク全体で観測するデータの空間的な相関性を利用してデータの集約を行う。そのため、これら従来の方と併用することで、通信コストをさらに削減するこ

とが可能であると考えられる。

また、未来の観測データを予測し、その予測の範囲内でデータが観測されている間はデータを送信しないことで送信データ量を削減する方式^{1),2)} もある。これらの手法は、観測するデータが疎であったり、変化の予測が容易なデータであるときには、通信コストの削減に大きな効果がある。しかし、変化の頻度が高い場合には削減効果が薄くなる。NSWC は、こうしたデータの特性に依存することなく、常に低い通信コストでデータの集約を行うことができる。

さらに、ネットワーク全体の通信コストを平滑化することで、ネットワークのライフタイムを最大化することを目的とした研究もある¹⁵⁾。中でも Luo ら¹⁶⁾ は、データを集約する経路全体に通信コストを分散させることで、通信コストを可能な限り削減しながらコストの平滑化を行う方法を示している。これらの手法は通信コストを平滑化できる一方で、ネットワーク全体の通信コストの総和は NSWC より高くなる。このため、ネットワークの運用者が、用途に応じて NSWC とこれらの手法とを選択的に使うことが、実用上は想定される。

3. 課題

ここでは、NSWC によって与えられる通信コストを実現できる符号化方式である LDPC 符号と、これを用いて実際にデータを符号化するとき発生する問題点について述べる。

3.1 低密度パリティ検査符号 (LDPC)

NSWC は、理論的に通信レートを最小化できるネットワークの設計法を与えるものである。しかし、NSWC によって与えられる通信コスト W_G でデータを伝送するための、データの符号化手段については NSWC では与えられない。この通信コストでのデータ伝送実現できる符号化方式としては、低密度パリティ検査 (LDPC: Low Density Parity-Check) 符号が知られている。LDPC については文献7) が詳しい。ここでは、LDPC について以降の議論に必要な最小限の説明を行う。

以下ではまず、二元線形符号としての LDPC 符号を用いて、ひとつのセンサノード X が観測するデータの系列 X を符号化する場合について説明を行う。 A_X を、各要素が 1 もしくは 0 である $n \times m$ ($0 < m < n$) 行列とする*1。 x を長さ n のデータの系列とする。行列 A_X に含まれる 1 の数が、 A_X の要素数に比べて極めて少ないものとする。 X

*1 LDPC は一般にはエラー訂正符号として用いられるが、本稿ではデータ圧縮のための符号として用いる。このため、検査行列にはエラー訂正ビットを含まないものとして扱う

は $\mathbf{X} = (x_1, x_2, \dots, x_n)$ で与えられ、各 x_i の取りうる値は 0 または 1 である。すなわち $x_i \in \{0, 1\}$ である。このとき、データ系列 \mathbf{X} に対する LDPC 符号の符号語 $c(\mathbf{X})$ は次の式で与えられる。

$$c(\mathbf{X}) = A_X \mathbf{X}^T$$

符号語 $c(\mathbf{X})$ の長さは m であり、各要素の取りうる値は 0 もしくは 1 である。符号語 $c(\mathbf{X})$ の復号には、符号語と行列 A_X 、データ系列の各要素の出現確率 $p(x=0), p(x=1)$ の情報が必要になる。これらの情報を入力情報とし、sum-product アルゴリズム⁷⁾ を利用することで復号できる。このとき、LDPC 符号によるデータの圧縮比率は行列式のサイズ n, m により m/n で与えられる。

3.2 解決すべき課題

センサノードに実装することを考慮したとき、LDPC 符号を用いる上で課題となりうる二つの特徴がある。

- センサノードとシンクノード間での行列式の共有の必要性: 符号化、復号の過程で同じ行列 A_X を用いる。すなわち、符号化を行うセンサノードと復号を行うシンクノードで同じ行列 A_X の情報を共有する必要がある。しかし、復号器の復号性能を高める(理論的な通信レートに近づける)ためには、可能な限り大きな行列式を使う必要がある。既存の研究により、十分な性能を得るためには行列サイズを $n = 10^4 - 10^5$ 程度にする必要があることが知られている。このサイズの行列式を、センサノードとシンクノード間で共有するためには、データ量以上に大きな通信コストを必要とする。小さな通信コストで、行列式の情報を共有する仕組みが必要になる。
- 圧縮率の事前計算の必要性: LDPC 符号は、行列式のサイズの比率でデータの圧縮率が決まる。しかし、この与えられた圧縮率が対象とするデータ系列の情報量 $H(\mathbf{X})$ 以下であるときは、復号アルゴリズムによって正しく復号することができない。すなわち、復号したデータに歪みが生じる。歪みを生じさせないためには、予め情報量を計算してから検査行列のサイズを決める必要がある。そして、このサイズを求めるためには、復号前のデータに含まれる 1 と 0 の生起確率 $p(x=0), p(x=1)$ の情報が必要になる。これらを計算する過程の導入が必要になる。

次に Slepian-Wolf Theory に基づき、二つのセンサノードに対して LDPC 符号を用い、データの符号化を行う場合について説明する。センサノード X_1, X_2 が存在し、それぞれのノードで観測されるデータの系列をそれぞれ $\mathbf{X}_1, \mathbf{X}_2$ とする。 \mathbf{X}_1 に関しては、前述のノードがひとつの場合の方法を用いてデータの符号化および復号を行う。 \mathbf{X}_2 の符号化に関

しては、行列のサイズと復号時に用いる生起確率の情報について、センサノードが一つの場合とは異なる情報が必要となる。

データ系列 \mathbf{X}_2 の伝送に必要な通信レートの下限は、Slepian-Wolf Theory により $\mathbf{X}_1, \mathbf{X}_2$ の結合情報量である $H(\mathbf{X}_1, \mathbf{X}_2)$ で与えられる。すなわち、 \mathbf{X}_2 の符号化に用いる行列式 A_2 のサイズ m_2, n_2 は $m_2/n_2 > H(\mathbf{X}_1, \mathbf{X}_2)$ となるように決める。

この行列 A_2 を用いて符号化された符号語 $c(\mathbf{X}_2)$ を復号するときは、 $p(x=0), p(x=1)$ の代わりに、 $p(x_2=0|x_1=0), p(x_2=1|x_1=0), p(x_2=0|x_1=1), p(x_2=1|x_1=1)$ の情報が必要になる。実際には $p(x_2=0|x_1=0) = 1 - p(x_2=1|x_1=0)$ として求められるため、 $p(x_2=0|x_1=0), p(x_2=0|x_1=1)$ の情報があればよい。

ここで問題となるのは、符号化や復号時に必要となる $H(\mathbf{X}_1, \mathbf{X}_2), p(x_2=0|x_1=0), p(x_2=0|x_1=1)$ のそれぞれの情報は、 $\mathbf{X}_1, \mathbf{X}_2$ の両方のデータを受け取っているシンクノードでしか計算ができない点である。正確な値を計算するためには、符号化を行う前にシンクノードでこれらの値を計算しなければならないという矛盾が生じる。LDPC 符号を用いて NSWC で与えられる通信コストを実現するためには、この問題について解決する必要がある。

4. 提案手法

前章で述べたように、センサネットワークにおいて LDPC を用いてデータの相関を利用したデータ圧縮を行おうとしたときには、以下の二つの課題を

- (1) 圧縮率を可変にするためには、行列式の大きなデータをセンサとシンク間で伝送する必要がある。
- (2) データの最適な圧縮率を、データをシンクノードに集約する前に正確に知ることができない。

これらの問題に対して、次の二つの解決法を導入することを提案する。

- (1) LDPC-CC の導入とパラメータからの行列式の生成法
- (2) 過去のデータからの圧縮率の予測と復号失敗時の再符号化

以下、これらについて順に説明する。

4.1 畳込み LDPC 符号 (LDPC-CC)

復号性能のよい LDPC 検査行列の生成法については、数多くの研究がある。LDPC 符号を用いることで、理論的に与えられる通信レートに限りなく近い圧縮性能が得られることは知られている⁷⁾。しかし、現時点で理論どおりの性能が得られる LDPC 行列の生成法は知

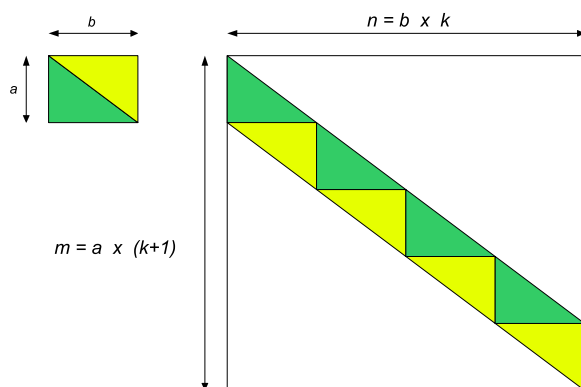


図4 LDPC-CC の行列式の生成
 Fig.4 Generating an LDPC-CC matrix

られていない。理論にちかい性能を得られる LDPC 検査行列の生成法は、これまでもいくつか提案されている。その中のひとつに LDPC-CC 検査行列がある⁹⁾。

LDPC-CC では、小さな行列式を大きな行列の対角線上にタイルパターン状に配置して生成した行列式を用いる。復号性能は LDPC と比べるとわずかながら低下するものの、LDPC-BC 符号のような他の符号と比較して、よい性能が得られることが知られている。その一方で、行列式を生成するために必要な情報量が LDPC の符号化器より少ないことから、符号化器と復号器との間で共有する必要のあるデータ量が削減され、なおかつ符号化時に必要なメモリ量が少なくて済むという利点がある。

LDPC-CC 検査行列の生成法を図 4 に示す。この例では、符号化するデータの長さを $n = bk$ とし、計算によって得られた通信レートが $R = n/m = b/a$ で与えられているものとする。まず、基礎行列としてサイズが $a \times b (0 < a < b)$ である LPPC 検査行列を予め生成しておく。これを、図中の左側に示したように対角線上で分割し、分割されたそれぞれの行列式を図の右側に示したように対角線上に交互に $k (0 < k)$ 個配置する。

このようにして生成された LDPC-CC 検査行列のサイズは $a(k+1) \times bk$ となる。LDPC 行列と比べると k だけ符号語の長さが増加する。そのため、圧縮率は $a(k+1)/bk$ となり、LDPC 検査行列よりは圧縮率が下がる。しかし、 k を十分大きく取ることで、性能の低下の影響を小さく抑えることができる。行列式のサイズは LDPC と比べると $1/k$ となることから、データサイズを $1/k$ に抑えることができる。

4.2 パラメータからの行列生成

基礎行列となる LDPC 検査行列は、sum-product 復号法を用いる場合、復号性能を高く保つ上で次の性質をもつことが望ましいことが知られている。

- (1) 各行に含まれる 1 の数が等しい。
- (2) 各列に含まれる 1 の数が等しい。
- (3) 任意に選んだふたつの行または列について、いずれの行または列の成分が 1 である位置が 1 箇所以下である。

このような性質をもつ行列式によって符号化された符号を、正則 LDPC 符号と呼ぶ⁷⁾。この行列式の生成法についても、さまざまな方法の提案がなされている。これらに対し、本稿では通信コストを抑制し、センサノードの少ない計算資源で行列式が生成できることに焦点をおいた、少ないパラメータから行列式を生成するシンプルなアルゴリズムを用いることを提案する。

アルゴリズムを以下に示す。

- (1) 基礎行列のサイズ a, b および各行に含まれる 1 の個数 i 、各列に含まれる 1 の個数 j 、擬似乱数の seed である数値 s 、ループ回数 c を入力とする。ただし i は a の約数、 j は b の約数である。
- (2) すべての要素が 0 である、サイズが $a \times b$ である基礎行列 E を生成する。
- (3) すべての要素が 1 であるサイズが $i \times j$ である行列 F を生成する。これを $a \times b$ 行列の対角線上に繰り返し配置する。すなわち、 E 上の $a' \bmod a = 0, b' \bmod b = 0$ であるような要素 $e_{a',b'}$ の位置が、 F の要素 $f_{0,0}$ の位置になるように F を配置する。
- (4) E 上の二つの要素をランダムに選ぶ。乱数の生成には s を seed とした擬似乱数を用いる。 $e_{u,v}, e_{x,y}$ とする。このとき $e_{u,v} = 1, e_{x,y} = 1$ かつ $e_{u,y} = 0, e_{x,v} = 0$ であるように要素を選ぶ。そして、 $e_{u,y} = 1, e_{x,v} = 1, e_{u,v} = 0, e_{x,y} = 0$ として要素を入れ替える。この操作を c 回繰り返す。
- (5) E を LDPC-CC 行列として出力する。

この結果得られる E は、1, 2 の条件は見たすが、3 の条件を満たす保証はない。そのため、予め複数の s と c の組み合わせで行列式を作り、正則 LDPC 符号を生成できる行列式を出力する組み合わせを見つけておく方法を用いる。なお、パラメータの組み合わせによっては 3 の条件を満たす行列式を実用的な時間で生成できないことがある。この場合は、生成した行列式の中から、各行または各列の 1 の位置の重複数が少ないものを行列式として使用する。

このアルゴリズムを用いることで、パラメータの組み合わせが同じである場合は、必ず同じ行列式が生成できる。シンクノードからセンサノードへ、これらのパラメータを送るだけで、符号化に使用する行列式の情報を送ることができる。

4.3 圧縮率の計算

前述のように、LDPC 符号をデータ圧縮に用いる場合、予め復号可能な圧縮率を計算しておく必要がある。しかし、そのためにはシンクノードにデータをすべて集める必要がある。これは原理的に不可能である。そのため、圧縮率については過去に収集したデータから予測する方法を用いる。

センサネットワークで収集対象となる温度や湿度などのセンサデータは、値の変化や緩やかで周期性がある。そのため、データの平均情報量の時間変化が少ない。このことから、過去のデータから計算した平均情報量（結合平均情報量）を、そのまま圧縮率の計算に用いることとした。

シンクノードでは、復号操作を行うたびに復号できたデータの情報量を計算し、その情報量を用いて圧縮率を計算する。そして、その圧縮率に相当する LDPC-CC 検査行列を生成し、その行列の生成に必要なパラメータをセンサノードに送る。各センサノードでは、シンクノードから得たパラメータを使用して行列式を生成し、データの符号化を行う。

4.4 復号失敗時の再符号化

LDPC 符号をデータの圧縮に用いる場合、仮にデータの平均情報量を正確に求めて行列式のサイズを決めたとしても、復号に失敗することがある。しかも、各センサノードでは条件付き確率の情報がなく、復号が可能かどうかを確認することができない。そのため、シンクノードに符号語がすべて集まった時点で、初めて復号に失敗することがわかる。復号に失敗する原因としては、圧縮率がデータの情報量に対して高すぎるか、符号化に使用した行列式の性能が十分でなかったかのいずれかである。これを解決するためには、通信レートをやや高めに設定しなおし、最初に設定した通信レートとの差分と等しいビット数の符号語を追加で生成して、シンクノードに送る必要がある。

追加する通信レートは、便宜上 a ビットとする。これから、サイズが $a \times bk$ である LDPC 検査行列 K を生成する。この行列式を使って符号化を行うと、長さ a の符号語が生成される。復号時には A の下部に K を結合した行列を用いて復号を行う。この操作を行ってもさらに復号に失敗した場合は、さらに通信レートを a ビットだけ拡張し、同じ操作を行う。なお、このようにして結合生成した行列式は LDPC-CC 検査行列ではない。そのため、理論的な性能の保証はない。性能の理論的な解析については別途行う必要がある。

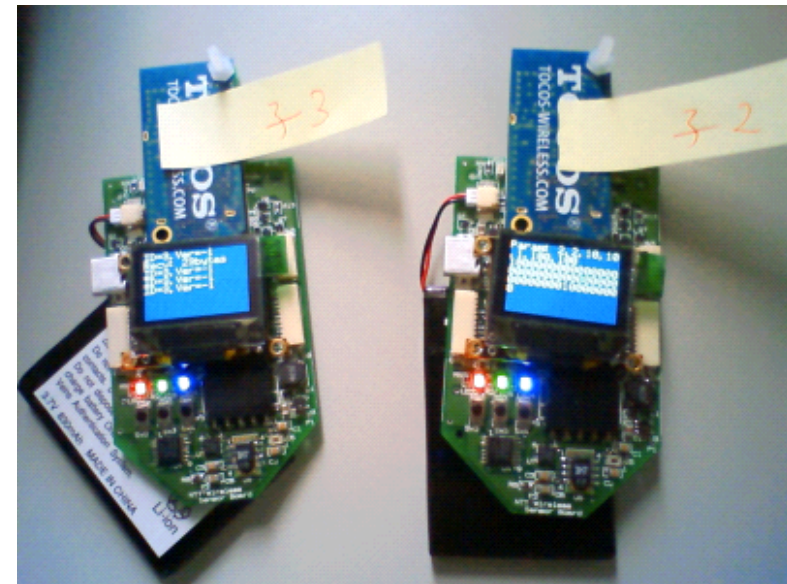


図 5 実機での動作の様子
Fig. 5 Sensor Nodes

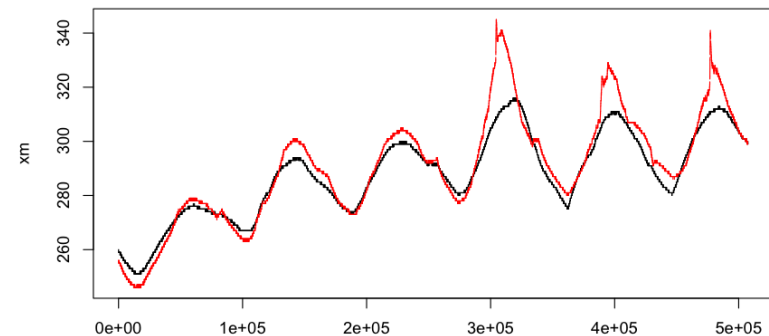


図 6 二台のセンサノードで取得した 6 日間の温度データ
Fig. 6 Temperature data obtained by two sensor node for 6 days

5. 実機の動作

筆者らは、LDPC-CC の行列生成アルゴリズムと符号化アルゴリズムを、筆者らが実装した 32bit CPU を搭載する無線センサノード上に実装した。実機での動作の様子を図 5 に示す。センサノードでは温度データを 1 秒毎に計測し、温度変化がなかったときは 0 を、変化があったときは 1 をデータとして取得する。これを 60 秒分蓄積し、60 秒経過した時点で符号化を行う。この例では、右側のノードでは符号化が行われた結果 (0,1 ビットの列) が画面に表示されている。符号化に用いる行列式は、シンクノードにデータを送信するときに、シンクノードから受け取る。左側の実機の画面には、シンクノードから受け取ったパラメータを使って行列式が生成した経過が表示されている。計測した温度データの変化の様子を図 6 に示す。図の縦軸の 1/10 の値が摂氏の温度に等しい。

LDPC-CC の圧縮性能の評価については現在進めているところである。圧縮率が適切に与えられるケースでは概ね想定通りの圧縮率が得られている。しかし、圧縮率が実データの情報量とあわず、再送を必要とする場合には性能が大きく劣化するケースがあることも確認されている。これらの点については、今後さらなる手法の改良が必要であると考えている。

6. ま と め

本稿では、複数のセンサノードから観測したデータをセンサノードに集約するセンサネットワークにおいて、データの相関性を利用してデータを集約するために、符号化手法である LDPC を実機のセンサノードに適用する方法について提案した。そして、提案手法が実機で動作することを実装によって確認した。今後は、実データに対する圧縮性能について詳しく調べる予定である。

参 考 文 献

- 1) Silberstein, A., Puggioni, G., Gelfand, A., Munagala, K. and Yang, J.: Suppression and failures in sensor networks: a Bayesian approach, *VLDB '07: Proceedings of the 33rd international conference on Very large data bases*, VLDB Endowment, pp.842–853 (2007).
- 2) Leonard, N., Paley, D., Lekien, F., Fratantoni, D., Hole, W., Sepulchre, R. and Davis, R.: Collective Motion and Sensor Networks and Ocean Sampling, *Proceedings of the IEEE, special issue on the emerging technology of networked control systems*, pp.48–74 (2007).
- 3) Akkaya, K. and Younis, M.: A survey on routing protocols for wireless sensor networks, *Ad Hoc Networks*, Vol.3, No.3, pp.325 – 349 (2005).
- 4) Baltasar, R.C., Cristescu, R., Beferull-lozano, B. and Vetterli, M.: On Network Correlated Data Gathering, *InfoCom '04: Proceedings of The 23rd Conference of the IEEE Communications Society*, pp.2571–2582 (2004).
- 5) Liu, J., Adler, M., Towsley, D. and Zhang, C.: On optimal communication cost for gathering correlated data through wireless sensor networks, *MobiCom '06: Proceedings of the 12th annual international conference on Mobile computing and networking*, New York, NY, USA, ACM, pp.310–321 (2006).
- 6) Slepian, D. and Wolf, J.K.: Noiseless Coding of Correlated Information Sources, *IEEE Transactions on Information Theory*, Vol.19, No.4, pp.471–480 (1973).
- 7) 和田山正：誤り訂正技術の基礎，森北出版株式会社 (2010)。
- 8) Liveris, A.D., Xiong, Z. and Georghiades, C.N.: Compression of binary sources with side information at the decoder using LDPC codes, *Communications Letters, IEEE*, Vol.6, No.10, pp.440–442 (2002).
- 9) Felström, A.J. and Zigangirov, K.S.: Time-varying periodic convolutional codes with low-density parity-check matrix, *IEEE Trans. Inf. Theory*, Vol.45, No.6, pp.2181–2191 (1999).
- 10) Gupta, H., Navda, V., Das, S. and Chowdhary, V.: Efficient gathering of correlated data in sensor networks, *ACM Transactions on Sensor Networks*, Vol.4, No.1, pp.1–31 (2008).
- 11) Yoon, S. and Shahabi, C.: The Clustered AGgregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks, *ACM Transactions on Sensor Networks*, Vol.3, No.1, p.3 (2007).
- 12) Donoho, D.: Compressed Sensing, *IEEE Transactions on Information Theory*, Vol.52, No.4, pp.1289–1306 (2006).
- 13) Haupt, J., Bajwa, W., Rabbat, M. and Nowak, R.: Compressive Sensing for Networked Data, *IEEE Signal Processing Magazine*, Vol.25, No.2, pp.92–101 (2008).
- 14) Rabbat, M., Haupt, J., Singh, A. and Nowak, R.: Decentralized Compression and Predistribution via Randomized Gossiping, *IPSN '06: Proceedings of International Conference on Information Processing in Sensor Networks*, pp.51–59 (2006).
- 15) Liang, W. and Liu, Y.: Online Data Gathering for Maximizing Network Lifetime in Sensor Networks, *IEEE Transactions on Mobile Computing*, Vol.6, pp.2–11 (2007).
- 16) Luo, C., Wu, F., Sun, J. and Chen, C.W.: Compressive data gathering for large-scale wireless sensor networks, *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*, New York, NY, USA, ACM, pp.145–156 (2009).