

科学研究費助成事業 研究成果報告書

令和 5 年 6 月 9 日現在

機関番号：34504

研究種目：基盤研究(C)（一般）

研究期間：2017～2022

課題番号：17K00023

研究課題名（和文）非一様情報源に対する高効率非可逆データ圧縮

研究課題名（英文）Efficient Lossy Compression for Non-Uniform Information Sources

研究代表者

井坂 元彦（ISAKA, Motohiko）

関西学院大学・工学部・教授

研究者番号：50351739

交付決定額（研究期間全体）：（直接経費） 2,300,000円

研究成果の概要（和文）：デジタル通信の効率化を目的として、原情報の忠実な復元を要求しない非可逆データ圧縮に関する研究を遂行した。研究期間の前半では、非一様分布に従う情報源に対するトレリス符号の構成と性能評価、および多元情報源の符号化に適した畳込み符号の構成を行った。研究期間の後半では、疎行列符号を用いたメッセージパッシング型符号化器の勾配降下法に基づく学習、および複数の観測ノードによる疎ベクトルの分散的な再構成に関して検討した。

研究成果の学術的意義や社会的意義

ネットワーク上で送受信されるデータ量の増大に伴い、情報伝送の効率化が重要な課題となっており、その解決に向けて非可逆データ圧縮の高度化が有用となる。原情報に対する復元データの忠実度と、圧縮率の間には原理的にトレードオフが存在し、その理論的性能限界に迫る圧縮法の具体化が求められる。本研究では、非可逆データ圧縮に適したトレリス符号の構成、多元アルファベット上の畳込み符号の設計、機械学習に基づく符号化器の効率化、複数の観測ノードによる原情報の分散再構成に関する成果を得ており、理論・実用の双方の観点で情報伝送の効率化に寄与している。

研究成果の概要（英文）： We studied lossy source compression in which original data is recovered from compressed data under certain fidelity criterion. In the first part of this project, we had emphasis on the design of trellis codes for non-uniformly distributed sources and non-binary convolutional codes with large free distance. Subsequently, we discussed the enhancement of algorithms for lossy compression in terms of the fidelity or computational and communication cost through the use of machine learning techniques, including the training of message-passing encoder for sparse graph source codes, and distributed recovery of sparse vectors observed by multiple sensing nodes.

研究分野：情報理論とその応用

キーワード：非可逆データ圧縮 符号化 非一様情報源 トレリス符号 確率伝搬法

1. 研究開始当初の背景

ネットワーク上で送受信されるデータ量の増大に伴って情報伝送の効率化が求められる中、送信データ量の削減を実現する非可逆データ圧縮の高度化は重要な課題となっている。原情報に対して圧縮のための符号化が行われ、その結果を基に原情報が復元される系において、復元データの忠実度と圧縮率の間には原理的にトレードオフが存在する。これは情報理論的な観点からはレート・歪み関数により特徴づけられるが、正確な導出が未解決である情報源のクラスも多く残されている。さらに、その理論的性能限界に迫る非可逆データ圧縮のための具体的な符号化法についても検討の余地が大きい。

本研究課題を開始する以前の10年程度の間、記憶のない2元一様情報源に対しては、線形符号をコードブックとする手法に顕著な進展が見られた。これは、低密度生成行列符号やポーラ符号など、通信路符号化分野における通信路容量に迫る符号化法を利用したものである。

また、2010年代中盤前後以降に機械学習分野において急速な研究の進展が見られたが、本研究課題の構想段階ではその方法論を情報理論や符号化、デジタル通信方式へ適用する動きは顕在化していなかった。

2. 研究の目的

本研究では、非一様情報源に対する非可逆データ圧縮のための符号化を主な検討課題とする。従来の符号理論では、雑音のある通信路における高信頼かつ高速の伝送を実現するための通信路符号化に対して、有限体上のベクトル空間として符号が構成されてきた。このような線形符号を非可逆データ圧縮に適用する場合、情報源が一様分布に従う場合や確率分布が対称性を有するならば、理論的性能限界に迫る特性を達成し得る。

しかし、本研究における研究対象のひとつである非一様情報源では、一般に線形符号を使用することでは理論的性能限界を達成できないことが示されている。これは、符号理論で蓄積されてきた成果の直接的な適用が難しいことを意味し、非線形符号の設計に関する基礎的事項から検討を行う必要性を示唆するものである。そこで、情報源の統計的性質に整合した符号の設計基準に関わる分析を行った上で、より広いクラスの非一様情報源に対して良好な性能を達成する情報源符号を具体的に構成することを目標とする。また、符号化やデジタル通信方式への機械学習の適用に関する研究が本研究課題の期間内に世界的に進展し、その有用性が明らかにされたことを踏まえ、非可逆データ圧縮への機械学習の活用を研究課題として加えることとなった。

3. 研究の方法

研究期間の前半では、情報源の統計的性質に整合した線形符号および非線形符号の設計を行った。特に、非一様情報源に対するトレリス符号化を対象とし、各種設計パラメータの非可逆データ圧縮の性能への影響の検討を行い、その結果を基にして具体的な符号の構成と性能評価を行った。また、より広範なアルファベット上での情報源符号化および通信路符号化に向けて、最小自由距離および重み分布に関して最適な多元畳込み符号を設計した。

研究期間の後半では、機械学習を取り込んだ非可逆データ圧縮法を扱った。低密度生成行列符号を用いて反復的な処理による符号化を行う系において、処理に内在するパラメータを勾配降下法に基づいて学習を行う手法を提案している。さらに、複数のノードが同一の疎ベクトルを観測した結果として得られた低次元ベクトルを基に、ノード群が協調的に原信号の再構成を行う分散的な圧縮センシングに関する検討を行った。

4. 研究成果

(1) 非一様情報源に対するトレリス符号化

一様分布に従う2元情報源に対して、線形符号の一種である低密度生成行列符号とメッセージ・パッシングアルゴリズムを用いることで理論的性能限界に迫るレート・歪み性能が達成される。一方、情報源の確率分布が一様でない場合には、線形符号ではこの限界は達成し得ないことが示されており、符号理論分野で培われた成果を直接的には適用できない。

この問題の解決策として、本研究ではトレリス線図と呼ばれるグラフ上で定義されるトレリス符号を用いた非可逆データ圧縮を検討した。原情報と復元データとの間の歪みを測る尺度をハミング距離とし、レート・歪み性能に関して最適な確率分布を特定した上で、それによって確

率的に符号を生成する手法を提案した。この符号は、情報源の統計的性質に整合した非線形符号を柔軟に与えられることに加えて、原情報に対する歪みを最小とする最尤符号化をトレリス線図上で効率的に実行できる利点を持つ。トレリス線図の状態数を増大することで、理論的性能限界に漸近する性能が達成されることを計算機実験により明らかにした。

(2) 重み分布に関して最適な多元畳込み符号の構成

畳込み符号は、トレリス符号の部分クラスをなす線形符号であり、符号理論分野では主に通信路符号化に向けた種々の研究が行われてきた。特に符号系列間のハミング距離の最小値として定義される最小自由距離は、多くの通信路での伝送において復号誤り確率に影響を及ぼす重要なパラメータである。このため、トレリス線図の各状態数について最小自由距離に関して最適な2元符号が構築されてきた。

一方、畳込み符号を非可逆データ圧縮に用いる場合、レート・歪み性能に関して最適な符号の設計基準は必ずしも明らかではないが、最小自由距離が大きい符号は最適ないしはそれに準ずる性能を与えることが経験的に知られている。

そこで本研究では、より多様なアルファベット上における通信路符号化および情報源符号化の両面で優れた多元畳込み符号の構成を行った。本研究の遂行中に、他研究グループにより最小自由距離に関して最適な多元畳込み符号に関する類似の結果が発表されたが、本研究課題による成果は最小自由距離およびそれに準ずる小さいハミング重みに対して、それらを与える符号系列対の個数を詳細に調査しており、より優れた符号を見出している。

(3) 低密度生成行列符号を用いた非可逆データ圧縮用符号化器の学習

低密度生成行列符号をコードブックとし、その局所的制約から記述される2部グラフ上でメッセージパッシング型のアルゴリズムを用いることで、2元一様情報源に対してレート・歪み関数に迫る符号化が行えることが知られている。この手法では、通信路符号化分野で知られるLDPC符号の復号法と異なり、反復処理の過程で出力されるメッセージの値を基に記号の逐次的な硬判定を施す操作が必要となる。従来研究では、このデシメーションと呼ばれる処理は経験的に定めた規則により実行されてきた。

本研究では、硬判定を施す操作に替えて、アナログ的なデシメーションを行う場合を対象とし、そこで現れるパラメータを学習により最適化する手法を提案した。符号の制約を表現する2部グラフを時間軸方向に展開したグラフを構成した上で、原情報と復元データとの間の歪みを損失関数とした勾配降下法により符号化器の訓練を行っている。2元一様情報源を対象として、実験的に導出されたパラメータを適用する場合と比して、より少ない計算量で優れたレート・歪み性能を達成できることを明らかにした。

(4) 分散圧縮センシングによる疎ベクトルの再構成

原情報が非零成分の比率が小さいベクトルであるとき、その次元より少ない回数の線形観測の下でも精度の良い原情報の再構成が可能であることが知られている。このような圧縮センシングでは、通常単一のノードにより原情報の観測、および再構成処理が行われるが、本研究ではその負荷を軽減するため、複数ノードで観測を行った上で、それらの間で協調的に原情報の再構成を行う系を想定する。

観測ノードが頂点により表され、通信リンクが辺に対応する木構造で表現されたネットワーク上において、各ノードが局所的な計算と通信を繰り返す分散的圧縮センシングの手順を導入した。本手法は直交近似確率伝搬法を基調とするため、従来研究より広いクラスの観測行列の下で精度のよい原信号の復元が実現される利点を有する。さらに、原信号の復元への寄与が小さいと考えられる成分の送信を抑制した計算および通信の手順を提案し、復元性能の劣化を招くことなくノード間の通信量が削減されることを計算機実験により明らかにした。

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件／うち国際共著 0件／うちオープンアクセス 1件）

| | |
|---|---------------------------|
| 1. 著者名 NOZAKI Takayuki、ISAKA Motohiko | 4. 巻 E105.B |
| 2. 論文標題 LDPC Codes for Communication Systems: Coding Theoretic Perspective | 5. 発行年 2022年 |
| 3. 雑誌名 IEICE Transactions on Communications | 6. 最初と最後の頁 894 ~ 905 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transcom.2021EBI0001 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |
| 1. 著者名 MIYAGI Shungo、ISAKA Motohiko | 4. 巻 E102-A |
| 2. 論文標題 Ternary Convolutional Codes with Optimum Distance Spectrum | 5. 発行年 2019年 |
| 3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences | 6. 最初と最後の頁 1688 ~ 1690 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.E102.A.1688 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 SOEN Hironori、ISAKA Motohiko | 4. 巻 E101.A |
| 2. 論文標題 Error Performance Analysis of Network Coded Cooperation for Gaussian Relay Networks | 5. 発行年 2018年 |
| 3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences | 6. 最初と最後の頁 2026 ~ 2036 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.E101.A.2026 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |
| 1. 著者名 Junya Hiramatsu, Motohiko Isaka | 4. 巻 101-A |
| 2. 論文標題 Lossy Source Coding for Non-Uniform Binary Source with Trellis Codes | 5. 発行年 2018年 |
| 3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science | 6. 最初と最後の頁 531 ~ 534 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.E101.A.531 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計9件（うち招待講演 1件 / うち国際学会 3件）

| |
|---|
| 1. 発表者名 Kazuki Ogawa, Hiromu Takata, Motohiko Isaka |
| 2. 発表標題 Lossy Source Coding of Binary Memoryless Source by Trained LDGM Codes |
| 3. 学会等名 2022 International Symposium on Information Theory and its Applications (ISITA2022) (国際学会) |
| 4. 発表年 2022年 |

| |
|---|
| 1. 発表者名 Ken Hisanaga, Motohiko Isaka |
| 2. 発表標題 Reduction of Communication Cost in Distributed Orthogonal Approximate Message Passing |
| 3. 学会等名 2022 International Symposium on Nonlinear Theory and its Applications (NOLTA2022) (国際学会) |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 久永健, 井坂元彦 |
| 2. 発表標題 分散的直交近似メッセージ伝搬法における通信コストの削減法 |
| 3. 学会等名 第44回情報理論とその応用シンポジウム(SITA2021) |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 西村拓実, 井坂元彦 |
| 2. 発表標題 分散台帳技術Tangleにおける承認対象取引の選択法 |
| 3. 学会等名 第44回情報理論とその応用シンポジウム(SITA2021) |
| 4. 発表年 2021年 |

| |
|-----------------------------------|
| 1. 発表者名 小川一樹, 井坂元彦 |
| 2. 発表標題 LDGM符号の学習に基づく有歪み情報源符号化 |
| 3. 学会等名 電子情報通信学会 情報理論研究会 |
| 4. 発表年 2022年 |

| |
|----------------------------------|
| 1. 発表者名 宮城俊吾, 井坂元彦 |
| 2. 発表標題 3元畳込み符号およびパンクチャド符号の構成 |
| 3. 学会等名 2019年電子情報通信学会ソサイエティ大会 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Motohiko Isaka |
| 2. 発表標題 Expected Error Rate of Probabilistic Network Codes over Gaussian Relay Network |
| 3. 学会等名 International Symposium on Information Theory and its Applications (国際学会) |
| 4. 発表年 2018年 |

| |
|----------------------------------|
| 1. 発表者名 宮城俊吾, 井坂元彦 |
| 2. 発表標題 重み分布に関して最適な3元畳込み符号の構成 |
| 3. 学会等名 平成30年電気関係学会関西連合大会 |
| 4. 発表年 2018年 |

| |
|-------------------------------|
| 1. 発表者名 井坂元彦 |
| 2. 発表標題 通信路符号化技術の動向 |
| 3. 学会等名 電気関係学会関西連合大会（招待講演） |
| 4. 発表年 2018年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|---------------------------|-----------------------|----|
|---------------------------|-----------------------|----|

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|---------|---------|