

令和 5 年 6 月 6 日現在

機関番号：12301

研究種目：若手研究

研究期間：2019～2022

課題番号：19K14989

研究課題名（和文）有歪み情報源符号化における実用的符号設計のための歪み球を用いた理論限界の解明

研究課題名（英文）A Study on Fundamental Limits of Lossy Compression Based on Distortion Ball
Aiming for Practical Code Construction

研究代表者

齋藤 翔太 (Saito, Shota)

群馬大学・情報学部・准教授

研究者番号：60822145

交付決定額（研究期間全体）：（直接経費） 1,000,000円

研究成果の概要（和文）：本研究では、もとのデータと復元後のデータの間の違い（歪み）を許容した情報源符号化問題（非可逆圧縮問題）を扱った。歪みを許容した情報源符号化において、符号語長のキウムラント母関数という評価基準があり、これは平均符号語長や最大符号語長を一般化した概念である。本研究では、有限のデータ系列長を考え、歪みを超過する確率を一定値以下にしたもとの符号語長のキウムラント母関数の最小値をレニーエントロピーに基づいた情報量により特徴付けられることを示した。

研究成果の学術的意義や社会的意義

画像や音声のデータは、もとのデータと復元後のデータが多少異なっていても問題ない場合が多いため、これらのデータの圧縮には有歪み情報源符号化の研究成果が用いられる。そのため、有歪み情報源符号化の研究の進展により、データの記憶容量や通信コストが減少すれば、限られた資源・エネルギーの節約につながる。本研究では、理論的にはどこまでデータを圧縮できるかというアルゴリズムの設計において目指すべき指針を与えた研究である。また、圧縮対象のデータの長さが有限であるという仮定のもとで結果を導いており、実用的な仮定のもとでの理論成果を導出している。

研究成果の概要（英文）：In this research, we treated lossy source coding problem. For this problem, one of the criteria is cumulant generating function of codeword lengths, which can be seen as a generalization of mean codeword lengths and maximum codeword lengths. Under the constraint that the excess distortion probability is less than a given constant and for a finite codeword length, we characterized the minimum cumulant generating function of codeword length by using a quantity based on Renyi entropy.

研究分野：情報理論とその応用

キーワード：情報理論 シャノン理論 データ圧縮 有歪み情報源符号化 非可逆圧縮 歪み球 数学的基礎理論

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

情報理論は、通信の数学的基礎理論としてのみならず、情報を扱う科学の基礎理論として、現代の情報化社会の発展に寄与してきた。情報理論の中で、情報を圧縮し効率のよい通信を考えるのが、情報源符号化（あるいはデータ圧縮）の問題である。この問題は、次の数理モデルをもとに研究が行われてきた。

- 1) データが情報源から確率的に出力される。
- 2) 情報源から出力されたデータを別のデータ（例えば0と1から成る系列）に変換する。この操作を符号化と呼ぶ。
- 3) 符号化された系列を、もとのデータに復元する。この操作を復号化と呼ぶ。

このとき、もとのデータ x と復号後のデータ y との間に違いを許容することで、 x と y が完全に一致する場合に比べて、圧縮の効率を高めることが可能である。このような情報源符号化は有歪み情報源符号化（非可逆圧縮）と呼ばれる。例えば、画像や音声のデータは、もとのデータと復号後のデータが多少異なっても問題ない場合が多いため、これらのデータの圧縮には有歪み情報源符号化の研究成果が用いられる。有歪み情報源符号化の研究の進展により、データの記憶容量や通信コストが減少すれば、限られた資源・エネルギーの節約につながる。

有歪み情報源符号化の問題に対して、伝統的に次の2つのアプローチにより研究が行われてきた。

- ・理論アプローチ(a)：情報源出力データの長さが無限に大きくなるなどの理想的な仮定のもとで、データ圧縮の効率をどこまで高めることができるか、という理論限界を研究する。
- ・実用アプローチ：情報源出力データの長さに関する多項式オーダーの計算量で符号化、復号化を行うなどの制約のもとで、理論アプローチで解明された理論限界まで圧縮ができる符号化、復号化アルゴリズムを構築する。

理論アプローチ(a)では、実用とは離れた理想的な仮定（例えば、情報源出力データの長さが無限に大きくなる等）のもとで解析を行うため、実用アプローチとの間に大きな隔たりが存在した。これに対し、近年、以下の理論アプローチ(b)のような実用性を考慮に入れた理論限界の解明が行われるようになってきた。

- ・理論アプローチ(b)：情報源出力データの長さが有限であるなどの実用的な仮定のもとで、データ圧縮効率の理論限界を導出する。

理論的側面の研究が、理論アプローチ(a)から理論アプローチ(b)へと移り、理論アプローチが実用アプローチに近づいてきた。しかしながら、両者の間には、未だに隔たりが存在するということが研究開始当初の背景であった。

2. 研究の目的

前述の研究開始当初の背景において述べた問題を踏まえて、本研究では、理論アプローチをさらに発展させることで、実用的符号設計の指針となる理論限界を明らかにすることを目的とした。

3. 研究の方法

当初、本研究を進める上で重要な役割を果たすと考えていた概念が「歪み球」である。これは復号後のデータ y と正の実数 D が与えられたとき、 $\{x : d(x,y) \leq D\}$ と定義される。ここで、 $d(x,y)$ はもとのデータ x と復号後のデータ y の違いを測る関数である。すなわち、ある復号後のデータ y に対して、違い（歪み）が D 以下であるような情報源出力データ x の集合が、歪み球である。本研究では、この歪み球という概念を用いた明示的な符号の構成により理論限界を解明することができるか、ということを探った。

4. 研究成果

当初の予想とは異なり、歪み球という概念を用いた明示的な符号の構成は予想以上に困難であることが判明した。その代わりに、よりシンプルな方法で符号の構成を示すことができることがわかり、その研究成果は情報理論のトップジャーナルである IEEE Transactions on Information Theory に掲載された。さらに、歪み球という概念を深く研究することにより情報セキュリティ関係の研究にこの概念を応用できることに気が付き、情報公開の際の有用性と情報漏洩量のトレードオフ問題に対する研究成果を得ることができた。より詳細に述べると、本研究で得られた成果は以下の通りである。

(1) 歪みを許容した情報源符号化(非可逆圧縮)において、符号語長のキュムラント母関数という評価基準があり、これは平均符号語長や最大符号語長を一般化した概念である。歪みを超過する確率を一定値以下にしたもとでの符号語長のキュムラント母関数の最小値を導出する結果を得ていたが、証明中に一部不備があること(反例が存在すること)が査読者のコメントにより明らかになった。その誤りを修正したうえで、論文誌に投稿し、最終的には論文が受理された。この成果は、情報理論のトップジャーナルである IEEE Transactions on Information Theory に掲載された。なお、この成果は有限の長さの情報源系列に対して、符号語長のキュムラント母関数の最小値をレニーエントロピーに基づいた情報量により特徴付けた結果であり、本研究の目標であった「実用的符号設計の指針となり得る理論限界の解明」を行ったものであると言える。

(2) 情報公開の際の有用性と情報漏洩量のトレードオフ関係を理論的に評価することが、従来から情報理論の研究において行われている。本研究では、一定値以下の歪み(有用性の評価基準)を許したもとでの情報漏洩量に関する理論評価を行った。本研究課題のポイントである「歪み球」という概念を用いて理論的な評価式を導出した。この研究成果を論文としてまとめ、査読付き論文誌である IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences に掲載した。

(3) 歪みを許容した情報源符号化の特別な場合、すなわち、もとのデータと復元後のデータの間の違い(歪み)がゼロであるような情報源符号化の研究も行った。このような情報源符号化において、代表的な符号のひとつにベイズ符号がある。情報源系列の統計的性質が区間毎に異なるような情報源系列に対して効率的なベイズ符号のアルゴリズムを構築した。この結果は、2021 IEEE Information Theory Workshop にて発表した。また、このような効率的なベイズ符号のアルゴリズム構築の際に重要となる事前分布の性質についてまとめた。この事前分布の性質に関する研究成果を査読付きの論文誌 Entropy に掲載し、これを発展させた研究を情報理論のトップカンファレンスである 2022 IEEE International Symposium on Information Theory にて発表した。さらに、情報源系列が文脈木情報源から発生している状況に対する逐次型情報源符号化の最大冗長度の下界を、ベイズ符号の理論研究の成果を応用して導出し、研究成果を電子情報通信学会 情報理論研究会(群馬県前橋市)にて発表した。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Saito Shota, Matsushima Toshiyasu	4. 巻 69
2. 論文標題 Non-Asymptotic Bounds of Cumulant Generating Function of Codeword Lengths in Variable-Length Lossy Compression	5. 発行年 2023年
3. 雑誌名 IEEE Transactions on Information Theory	6. 最初と最後の頁 2113 ~ 2119
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/tit.2022.3229358	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 SAITO Shota, MATSUSHIMA Toshiyasu	4. 巻 E105.A
2. 論文標題 Upper Bound on Privacy-Utility Tradeoff Allowing Positive Excess Distortion Probability	5. 発行年 2022年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 425 ~ 427
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transfun.2021TAL0002	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 Nakahara Yuta, Saito Shota, Kamatsuka Akira, Matsushima Toshiyasu	4. 巻 24
2. 論文標題 Probability Distribution on Full Rooted Trees	5. 発行年 2022年
3. 雑誌名 Entropy	6. 最初と最後の頁 328 ~ 346
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e24030328	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 Murayama Haruka, Saito Shota, Iikubo Yuji, Nakahara Yuta, Matsushima Toshiyasu	4. 巻 20
2. 論文標題 Cluster's Number Free Bayes Prediction of General Framework on Mixture of Regression Models	5. 発行年 2021年
3. 雑誌名 Journal of Statistical Theory and Applications	6. 最初と最後の頁 425 ~ 449
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s44199-021-00001-5	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 Dobashi Nao, Saito Shota, Nakahara Yuta, Matsushima Toshiyasu	4. 巻 23
2. 論文標題 Meta-Tree Random Forest: Probabilistic Data-Generative Model and Bayes Optimal Prediction	5. 発行年 2021年
3. 雑誌名 Entropy	6. 最初と最後の頁 768 ~ 786
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/e23060768	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 山口純輝, 風間皐希, 鎌塚明, 齋藤翔太, 松嶋敏泰	4. 巻 J103-A
2. 論文標題 拡張直交配列を用いた混合水準の実験計画法に関する一考察	5. 発行年 2020年
3. 雑誌名 電子情報通信学会論文誌 A	6. 最初と最後の頁 17 - 24
掲載論文のDOI (デジタルオブジェクト識別子) なし	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39件 (うち招待講演 3件 / うち国際学会 8件)

1. 発表者名 Yuta Nakahara, Shota Saito, Akira Kamatsuka, Toshiyasu Matsushima
2. 発表標題 Probability Distribution on Rooted Trees
3. 学会等名 2022 IEEE International Symposium on Information Theory (国際学会)
4. 発表年 2022年

1. 発表者名 Shota Saito
2. 発表標題 On Meta-Bound for Lower Bounds of Bayes Risk
3. 学会等名 2022 IEEE International Symposium on Information Theory (国際学会)
4. 発表年 2022年

1. 発表者名 Taisuke Ishiwatari, Shota Saito, Yuta Nakahara, Yuji Iikubo, Toshiyasu Matsushima
2. 発表標題 Bayes Optimal Estimation and Its Approximation Algorithm for Difference with and without Treatment under URLC Model
3. 学会等名 2022 International Symposium on Information Theory and Its Applications (国際学会)
4. 発表年 2022年

1. 発表者名 Koshi Shimada, Shota Saito, Toshiyasu Matsushima
2. 発表標題 An Efficient Bayes Coding Algorithm for the Non-Stationary Source in Which Context Tree Model Varies from Interval to Interval
3. 学会等名 2021 IEEE Information Theory Workshop (国際学会)
4. 発表年 2021年

1. 発表者名 Shota Saito, Toshiyasu Matsushima
2. 発表標題 Evaluation of Error Probability of Classification Based on the Analysis of the Bayes Code: Extension and Example
3. 学会等名 2021 IEEE International Symposium on Information Theory (国際学会)
4. 発表年 2021年

1. 発表者名 齋藤翔太
2. 発表標題 セキュリティを考慮した2つの問題に対する情報理論的解析について ~ Local Differential Privacyの下でのパラメータ推定問題と、プライバシーと有用性のトレードオフ問題 ~
3. 学会等名 電子情報通信学会 情報理論研究会 (招待講演)
4. 発表年 2023年

1. 発表者名 齋藤翔太
2. 発表標題 文脈木情報源に対する逐次型情報源符号化の最大冗長さの下界について
3. 学会等名 電子情報通信学会 情報理論研究会
4. 発表年 2023年

1. 発表者名 Shota Saito
2. 発表標題 Lower Bound of Bayes Risk in Parameter Estimation under Local Differential Privacy
3. 学会等名 第45回情報理論とその応用シンポジウム
4. 発表年 2022年

1. 発表者名 中原 悠太, 齋藤 翔太, 一條 尚希, 風間 皇希, 松嶋 敏泰
2. 発表標題 決定木モデルにおけるメタツリーに対するマルコフ連鎖モンテカルロ法
3. 学会等名 第45回情報理論とその応用シンポジウム
4. 発表年 2022年

1. 発表者名 中原悠太, 齋藤翔太, 島田航志, 飯窪祐二, 風間皇希, 一條尚希, 松嶋敏泰, BayesML Developers
2. 発表標題 ベイズ決定理論に基づく機械学習ライブラリ: BayesML 0.2.0
3. 学会等名 第25回情報論的学習理論ワークショップ(IBIS2022)
4. 発表年 2022年

1. 発表者名 齋藤翔太
2. 発表標題 「情報」を測る：情報理論における情報の捉え方
3. 学会等名 第10回 早稲田大学データ科学センター・先端社会科学研究所ジョイントセミナー（招待講演）
4. 発表年 2022年

1. 発表者名 中原 悠太, 一條 尚希, 島田 航志, 飯窪 祐二, 齋藤 翔太, 風間 皐希, 松嶋 敏泰, BayesML Code Authors
2. 発表標題 ベイズ決定理論に基づく機械学習ライブラリ: BayesML 0.1.0
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 Shota Saito
2. 発表標題 On Bayesian Approach for Classification of Context Tree Model
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 一條尚希, 中原悠太, 齋藤翔太, 松嶋敏泰
2. 発表標題 潜在的分岐変数を持つ木構造からなる生成モデルと変分ベイズ法による近似推論
3. 学会等名 電子情報通信学会 情報論的学習理論と機械学習研究会
4. 発表年 2022年

1. 発表者名 齋藤翔太
2. 発表標題 ベイズリスクの下界に対するmeta-bound
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 于文斌, 風間阜希, 中原悠太, 一條尚希, 齋藤翔太, 松嶋敏泰
2. 発表標題 決定木モデルに対するベイズ最適な予測のメタツリーブースティング法による近似
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 Shota Saito
2. 発表標題 A Refined Analysis of Merhav and Ziv's Bayesian Approach for Classification of Markov Sources
3. 学会等名 第44回情報理論とその応用シンポジウム
4. 発表年 2021年

1. 発表者名 Yuta Nakahara, Shota Saito, Akira Kamatsuka, Toshiyasu Matsushima
2. 発表標題 Probability Distribution on Full Rooted Trees
3. 学会等名 第44回情報理論とその応用シンポジウム
4. 発表年 2021年

1. 発表者名 山田一翔, 中原悠太, 齋藤翔太, 飯窪祐二, 松嶋敏泰
2. 発表標題 URLモデルを仮定した下での新しい個体への施策の有無による反応の差に関するベイズ最適な決定とその近似アルゴリズム
3. 学会等名 第44回情報理論とその応用シンポジウム
4. 発表年 2021年

1. 発表者名 石渡泰祐, 齋藤翔太, 中原悠太, 飯窪祐二, 松嶋敏泰
2. 発表標題 URLモデルを仮定した下での母集団全体に対する施策の有無の差に関するベイズ最適な推定とその近似アルゴリズム
3. 学会等名 第44回情報理論とその応用シンポジウム
4. 発表年 2021年

1. 発表者名 中原悠太, 齋藤翔太, 風間皐希, 于文斌, 一條尚希, 松嶋敏泰
2. 発表標題 マルコフ連鎖モンテカルロメタツリー法
3. 学会等名 第24回情報論的学習理論ワークショップ(IBIS2021)
4. 発表年 2021年

1. 発表者名 香山渉, 齋藤翔太, 松嶋敏泰
2. 発表標題 クラスタ構造を説明する変数と回帰構造を説明する変数により表現された線形回帰モデルにおける計算量削減に関する一考察 ~ ベイズ最適な予測とその近似アルゴリズム ~
3. 学会等名 電子情報通信学会 情報理論研究会
4. 発表年 2021年

1. 発表者名 石渡泰祐, 齋藤翔太, 松嶋敏泰
2. 発表標題 潜在クラスタを仮定することで個人差を考慮した施策の有無による反応変数の差に関するベイズ最適な予測とその近似アルゴリズム
3. 学会等名 電子情報通信学会 情報理論研究会
4. 発表年 2021年

1. 発表者名 Shota Saito, Toshiyasu Matsushima
2. 発表標題 On Two Information Quantities Relating Two Distortion Balls
3. 学会等名 2020 International Symposium on Information Theory and Its Applications (国際学会)
4. 発表年 2020年

1. 発表者名 Shota Saito, Toshiyasu Matsushima
2. 発表標題 Evaluation of Error Probability of Classification Based on the Analysis of the Bayes Code
3. 学会等名 2020 IEEE International Symposium on Information Theory (国際学会)
4. 発表年 2020年

1. 発表者名 Shota Saito, Toshiyasu Matsushima
2. 発表標題 Non-asymptotic converse theorem on the overflow probability of variable-to-fixed length codes
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2021年

1. 発表者名 島田航志, 齋藤翔太, 松嶋敏泰
2. 発表標題 区間ごとに文脈木モデルが変化する情報源における効率的ベイズ符号化アルゴリズム
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2021年

1. 発表者名 山岡大志, 齋藤翔太, 松嶋敏泰
2. 発表標題 複数の隠れマルコフモデルの重み付けによるベイズ基準のもとでの最適な音素の予測
3. 学会等名 電子情報通信学会音声研究会
4. 発表年 2021年

1. 発表者名 齋藤翔太, 松嶋敏泰
2. 発表標題 ベイズ符号の解析に基づいた分類誤り確率の評価 ~ 拡張と具体例 ~
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2020年

1. 発表者名 山岡大志, 齋藤翔太, 松嶋敏泰
2. 発表標題 音素認識問題におけるベイズ規準の下最適な予測に対する近似手法
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2020年

1. 発表者名 村山 春香, 齋藤 翔太, 飯窪 祐二, 中原 悠太, 松嶋 敏泰
2. 発表標題 線形回帰モデルの混合の統一的なフレームワークにおけるベイズ最適な予測とその近似アルゴリズム
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 Nao Dobashi, Shota Saito, Yuta Nakahara, Toshiyasu Matsushima
2. 発表標題 Probabilistic Data Generating Process on Tree Structure Model: Bayes Optimal Prediction and Sub-Optimal Algorithm
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 山田一翔, 齋藤翔太, 松嶋敏泰
2. 発表標題 変分ベイズ法によって導出された近似事後分布を用いた予測分布とベイズ予測分布の α -ダイバージェンスの漸近評価
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2020年

1. 発表者名 Shota Saito, Toshiyasu Matsushima
2. 発表標題 Non-Asymptotic Fundamental Limits of Guessing Subject to Distortion
3. 学会等名 2019 IEEE International Symposium on Information Theory (国際学会)
4. 発表年 2019年

1. 発表者名 村山春香, 齋藤翔太, 中原悠太, 松嶋敏泰
2. 発表標題 クラスター説明変数と回帰説明変数により表現された線形回帰モデルにおけるベイズ最適な予測の近似手法
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2019年

1. 発表者名 土橋那央, 齋藤翔太, 松嶋敏泰
2. 発表標題 決定木モデルにおける集団学習法の考えを用いたベイズ最適な分類の近似アルゴリズム
3. 学会等名 電子情報通信学会情報理論研究会
4. 発表年 2019年

1. 発表者名 齋藤翔太, 松嶋敏泰
2. 発表標題 Evaluation of Error Probability of Classification Based on the Analysis of the Bayes Code
3. 学会等名 第42回情報理論とその応用シンポジウム
4. 発表年 2019年

1. 発表者名 齋藤翔太
2. 発表標題 Guessing ~ 基礎と広がり ~
3. 学会等名 第42回情報理論とその応用シンポジウム (招待講演)
4. 発表年 2019年

1. 発表者名 天野祐貴, 齋藤翔太, 松嶋敏泰
2. 発表標題 潜在変数に非正規分布を仮定した回帰モデルにおけるベイズ基準の下最適な予測に対する近似手法
3. 学会等名 第42回情報理論とその応用シンポジウム
4. 発表年 2019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
---------	---------