

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 28 日現在

機関番号：12608

研究種目：若手研究(B)

研究期間：2013～2014

課題番号：25820154

研究課題名(和文)非同期型分散データ圧縮に対するデータ圧縮法の性能解析

研究課題名(英文)A Performance Analysis of Data Compression Methods for the Asynchronous Distributed Data Compression

研究代表者

松田 哲直(Matsuta, Tetsunao)

東京工業大学・理工学研究科・助教

研究者番号：00638984

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)：相関を有する複数の情報源から生じた固定長のデータ列を、非同期的に動作する分散した符号化器で独立に圧縮符号化し、復号器で同時に復号する非同期型分散データ圧縮に対して以下の成果を挙げた。

(1)非同期的な動作から生じる符号化遅延の最大値がデータ列長と共に変化する場合における、圧縮率の理論的な限界を明らかにした。(2)情報源の統計的性質が未知の場合でも対応できるデータ圧縮法の存在を示した。(3)データ列長が長くなるに連れて復号誤り率が指数関数的に減少するデータ圧縮法の存在を示し、その指数部分を閉じた式で与えた。

研究成果の概要(英文)：This study deals with the asynchronous distributed data compression in which encoders independently and asynchronously encode input sequences emitted from correlated sources into fixed-length codewords, and the decoder reconstructs all input sequences from the codewords. For this data compression, we obtained the following results:

(1) We deal with the case where the maximum of encoding delays to be changed with the data length, and clarified the theoretical limit of coding rates. (2) We showed the existence of a data compression method coping with the situation that statistical properties of the information sources are unknown. (3) We showed the existence of a data compression method such that the error probability vanishes exponentially, and gave a closed-form expression of the exponent of this probability.

研究分野：情報理論

キーワード：非同期 情報源符号化 ユニバーサル符号 誤り指数 多端子情報理論

1. 研究開始当初の背景

図1に示すように、相関を有する複数の情報源から生じた固定長のデータ列を、分散した符号化器で独立に圧縮符号化し、復号器で同時に復号する一連のシステムは、分散データ圧縮と呼ばれている。近年では、分散ストレージ、センサネットワークや多視点カメラのような、分散データ圧縮の実用的なシステムがよく用いられるようになり、分散データ圧縮に関する研究はますます重要になっている。

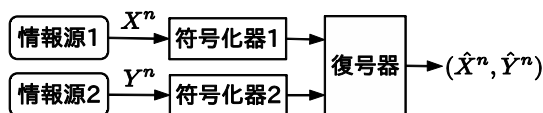


図1：分散データ圧縮

分散データ圧縮は大きく、分散無歪みデータ圧縮と分散有歪みデータ圧縮の2つに分けられる。分散無歪みデータ圧縮は、復号器において元のデータ列と同一のデータ列を復号する場合の分散データ圧縮である。他方、分散有歪みデータ圧縮は、復号器において一定範囲内の歪みを許容して元のデータ列を復号する場合の分散データ圧縮である。従って、分散有歪みデータ圧縮では必ずしも元のデータ列と同一のデータ列が復号されるわけではない。分散無歪みデータ圧縮はSlepianとWolf [1]によって、復号誤り率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在が示された。他方、分散有歪みデータ圧縮における圧縮率の限界は未だに解明されていないが、図2に示すようにどちらか片方の符号化器が元のデータ列をそのまま送ることができる場合には、復号の際に歪みが許容値を超える確率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在が示されている。この特別な場合はWynerとZiv [2]によって解析が行われたため、Wyner-Ziv符号化と呼ばれることが多い。

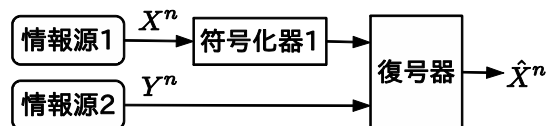


図2：Wyner-Ziv符号化

上記の成果において、符号化器間では完全に同期して圧縮することを仮定している。しかしながら、分散データ圧縮では、符号化器間で通信を行わないため、符号化器間で同期をとることが困難となる。従って、分散データ圧縮の実用上、符号化器が非同期である場合を検討することが重要となる。このような非同期型分散データ圧縮においては、符号化器はそれぞれ異なる時刻に情報源から生じたデータ列を処理することになる。すなわち、符号化器はそれぞれ基準となる時刻から

ある遅延を伴ったデータ列を処理することになる。非同期型分散無歪みデータ圧縮に対して、Willems [3]は符号化器において遅延は未知であるが、復号器において既知である場合を扱い、符号化器が完全に同期している場合と同一の圧縮率を任意に小さな復号誤り率で達成できるデータ圧縮法の存在を示した。しかしながら、復号器において遅延が既知であるという状況は現実的とは言えない。これに対して沖ら[4]は、遅延の最大値が有限であり、その最大値は復号器において既知であるが、遅延は符号化器と復号器において未知である場合を扱い、符号化器が完全に同期している場合と同一の圧縮率が任意に小さな復号誤り率で達成できるデータ圧縮法の存在を示した。他方、非同期型分散有歪みデータ圧縮に関する研究はあまりされていないが、Venkataramanaら[5]が非同期型Wyner-Ziv符号化において、変動しない遅延が符号化器と復号器において既知であり、2つの符号化器が同一のデータ列を圧縮するという非常に特別な場合を扱い、復号の際に歪みが許容値を超える確率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在を示している。

2. 研究の目的

上述したように、非同期型分散データ圧縮に対するデータ圧縮法の性能解析を行う研究がいくつか存在するが、十分な研究がされていないとは言えない。本研究の目的は、非同期型分散データ圧縮、特に非同期型分散無歪みデータ圧縮と非同期型Wyner-Ziv符号化に対するデータ圧縮法の性能解析を十分に行うことである。具体的には、実用上重要であると考えられる以下の課題に取り組む。課題は、(1)非同期型分散無歪みデータ圧縮と、(2)非同期型Wyner-Ziv符号化のそれぞれに分けて述べる。

(1) 非同期型分散無歪みデータ圧縮

① 沖らの成果の拡張：沖らの成果において、遅延の最大値は復号器において既知であったが、一般にはこの最大値も未知であることが考えられる。この場合に、符号化器が完全に同期している場合と同一の圧縮率を、任意に小さな復号誤り率で達成できるデータ圧縮法の存在を明らかにする。

② 情報源の統計的性質が未知である場合への拡張：従来の非同期型分散データ圧縮の研究では、符号化器と復号器において情報源の統計的性質が既知であることを仮定していた。しかしながら、現実的には情報源の統計的性質は未知であることが多い。従って、そのような場合でも対応できるデータ圧縮法の存在を示し、そのようなデータ圧縮法の性能解析を行う。

③ 復号における誤り確率の解析：非同期型

分散データ圧縮の研究は圧縮率に注目したものが多く、復号誤り率の評価が十分にされていない。これらの確率が圧縮するデータ列の長さに伴って、どのようなオーダで減少していくかはデータ圧縮法の性能を解析する上で重要であるため、これを解析する。

- (2) 非同期型 Wyner-Ziv 符号化
- ① Venkataramanan らの成果の拡張: 非同期型 Wyner-Ziv 符号化において、遅延の最大値と遅延が符号化器と復号器において未知である場合に、復号の際に歪みが許容値を超える確率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在を明らかにする。
- ② 情報源の統計的性質が未知である場合への拡張: 情報源の統計的性質が未知である場合でも対応できるデータ圧縮法の存在を示し、そのようなデータ圧縮法の性能解析を行う。
- ③ 復号の際に歪みが許容値を超える確率の解析: 復号の際に歪みが許容値を超える確率が、圧縮するデータ列の長さに伴ってどのようなオーダで減少していくかを解析する。

3. 研究の方法

非同期型分散データ圧縮に対するデータ圧縮法の性能解析について、以下の方法で研究を遂行した。

- (1) 非同期型分散無歪みデータ圧縮
- ① 沖らの成果の拡張: この課題を解析するために、まず圧縮するデータ列長に伴って遅延の最大値が分数指数関数のオーダで緩やかに増加する仮想的な状況を考え、この状況におけるデータ圧縮法の存在を示す。そのような圧縮法の存在は、申請者ら [6] の既存の手法を拡張して適用することで示すことができる。但し、ここでは増加する遅延の最大値は復号器において既知であると仮定する。次に、データ列長を十分に長くすることで、ここで示したデータ圧縮法が所望の圧縮法となることを示す。これはなぜなら、真の遅延の最大値が有限であれば、データ列長を十分に長くすることで仮想的な状況における遅延の最大値が真の遅延の最大値を上回ることになるためである。このとき、仮想的な状況に対するデータ圧縮法は真の遅延の最大値に依存していないため、真の遅延の最大値が未知であったとしても遅延に対応できるデータ圧縮法となる。
- ② 情報源の統計的性質が未知である場合への拡張: 起こり得る全ての情報源の中からランダムに選択された情報源からデータが生起する状況を考える。このような状況における情報源は混合情報源と呼ばれている。この混合情報源に対す

るデータ圧縮法が、統計的な性質が未知であっても対応できるデータ圧縮法であることを示す。これを示すために、申請者ら [6] の既存の手法を非同期の場合へ拡張して用いる。

- ③ 復号における誤り確率の解析: Gallager [7] が導入した、最尤復号器と和集合の確率に関する不等式を用いた解析手法を用いて、復号における誤り確率を解析する。
- (2) 非同期型 Wyner-Ziv 符号化
- ① Venkataramanan らの成果の拡張: 圧縮率の限界の解析には、完全に同期している場合の圧縮率の限界の解析手法が応用できるため、その手法を応用して限界を明らかにする。また、その限界を達成するデータ圧縮法の存在については、既存手法である典型系列を用いた復号法 [8] を用いて示す。
- ② 情報源の統計的性質が未知である場合への拡張: この課題については、申請者ら [6] の手法を有歪の場合にも拡張できるかを検討したが、うまく解析することができなかった。申請者らの方法以外では、Csiszar と Korner [9] による手法が存在するが、非同期の場合の解析とは相性が悪く、こちらを用いてもうまく解析することができなかった。
- ③ 復号の際に歪みが許容値を超える確率の解析: この課題については、Gallager [7] が導入した解析手法を用いることが出来ないかを検討したが、うまく解析することができなかった。

4. 研究成果

上記の研究方法のもとで研究を遂行することで、以下の成果を挙げた。

- (1) 非同期型分散無歪みデータ圧縮
- ① 沖らの成果の拡張: 研究方法で述べたように、この課題では遅延の最大値がデータ列長と共に変化する場合を考えている。そのため、副次的な成果として、遅延の最大値がデータ列長と共に変化する場合における、復号誤り率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在を示すことができた。その結果、一般には符号化器が完全に同期している場合と同一の圧縮率を達成することは不可能であることを明らかにした。また、遅延の最大値が有限であれば、その最大値が復号器において未知であったとしても、符号化器が完全に同期している場合と同一の圧縮率を任意に小さな復号誤り率で達成できるデータ圧縮法の存在を明らかにした。
- ② 情報源の統計的性質が未知である場合への拡張: 符号化器と復号器において情報源の統計的性質が未知の場合でも対応できるデータ圧縮法の存在を示し、そ

のような圧縮法の圧縮率の限界を明らかにした。

- ③ 復号における誤り確率の解析：圧縮率が理論的限界値を超えないある符号が存在し、その符号の復号誤り率はデータ列長が長くなるに連れて指数関数的に減少することを示した。また、その指数部分を閉じた式で与えることで、どの程度の速さで誤り率が減少するのかについても明らかにした。

上記①と②の成果の一部は情報理論研究会において発表している[10]。また、研究課題の年度内に発表することはできなかったが、上記①と②の成果を整理して③を加えた成果を、2015 IEEE Information Theory Workshop に国際会議論文として提出中である。

(2) 非同期型 Wyner-Ziv 符号化

- ① 復号の際に歪みが許容値を超える確率を任意に小さくできるデータ圧縮法の圧縮率の限界と、その限界を達成するデータ圧縮法の存在を明らかにした。この成果は、第 36 回情報理論とその応用シンポジウムにおいて発表し、これを拡張した成果について 2014 IEEE International Symposium on Information Theory において発表した。また、この成果は学術論文として掲載された。

研究目的において述べた課題の②と③については、課題の年度内にまとめた成果を出すことは出来なかった。これらの課題の解決が今後の課題である。

本研究の成果は、符号化器が非同期的に動作するという現実的な状況を考慮した場合の性能解析を行なっているため、従来の研究成果よりも実際の分散データ圧縮の設計指針として用いやすい。この理由から、本研究は分散データ圧縮の発展に貢献できたといえる。

<引用文献>

- [1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inform. Theory, vol. 19, pp. 471-480, July 1973.
- [2] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inform. Theory, vol. 22, pp. 1-10, Jan. 1976.
- [3] F. Willems, "Totally asynchronous slepian-wolf data compression," IEEE Trans. Inform. Theory, vol. 34, pp. 35-44, Jan. 1988.
- [4] N. Oki and Y. Oohama, "Coding for the asynchronous Slepian-Wolf data compression," Proc. of SITA' 97, pp. 89-92, Dec. 1997.
- [5] R. Venkataramanan and S. S. Pradhan,

"Source coding with feed-forward: Rate-distortion theorems and error exponents for a general source," IEEE Trans. Inform. Theory, vol. 53, pp. 2154-2179, June 2007.

- [6] T. Matsuta, T. Uyematsu, and R. Matsumoto, "Universal Slepian-Wolf source codes using lowdensity parity-check matrices," IEICE Trans. Fundamentals, vol. E93-A, no. 11, pp. 1878-1888, Nov. 2010.
- [7] R. G. Gallager, Information Theory and Reliable Communication. New York: John Wiley & Sons, Inc., 1968.
- [8] T. M. Cover and J. A. Thomas, Elements of Information Theory. Wiley-Interscience New York, 2006.
- [9] I. Csiszar and J. Korner, Information Theory: Coding Theorems for Discrete Memoryless Channels. New York: Academic, 1981.
- [10] T. Matsuta, and T. Uyematsu. "Universal coding for asynchronous Slepian-Wolf coding systems," 電子情報通信学会 情報理論研究会, 電子情報通信学会技術報告, Vol. 112, No. IT2012-62, pp. 1-6, Feb. 2013.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計1件)

- ① Tetsunao Matsuta, Uyematsu Tomohiko. On the Wyner-Ziv Source Coding Problem with Unknown Delay, IEICE Trans. Fundamentals, IEICE, Vol. E97-A, No. 12, pp. 2288-2299, Dec. 2014, 査読有り, DOI: 10.1587/transfun.E97.A.2288

[学会発表] (計2件)

- ① Tetsunao Matsuta, Uyematsu Tomohiko. Rate-Distortion Functions for Source Coding When Side Information with Unknown Delay May Be Present, 2014 IEEE International Symposium on Information Theory, Proc. of 2014 IEEE Inter. Symp. on Information Theory, Honolulu (Hawaii), pp. 936-940, 2014年7月1日.
- ② 松田哲直, 植松友彦. 未知の遅延を有する副情報を伴う歪み符号化におけるレート・歪み関数, 第36回情報理論とその応用シンポジウム, 伊東ホテル 聚楽(静岡県伊東市), pp. 491-496, 2013年11月28日.

6. 研究組織

(1) 研究代表者

松田 哲直 (MATSUTA, Tetsunao)

東京工業大学・大学院理工学研究科・助教
研究者番号: 00638984