# 符号化におけるロバスト計算

山本博資　新領域創成科学研究科複雑理工学専攻
小川朋宏*　情報理工学系研究科数理情報学専攻
藤田八郎†　情報理工学系研究科 21 世紀 COE

## 概 要

情報を「効率良く，高品質で，安全に」伝送または記録するために,「データ圧縮，誤り訂正，暗号」の符号化技術が使われている．本サブプロジェクトでは，情報源や通信路の特性，不正者からの攻撃方法，計算困難性の仮定などによらず，上記の目標をロバストに達成できる符号化技術の開発を目的としている．

## 1　はじめに

符号化技術は大きく，次の 3 種類に分類される．
**(A) データ圧縮符号化**： データ系列を，より短いビット長で表現できるように符号化する．符号木を用いるエントロピー符号化法や，さまざまな情報源出力を同一のアルゴリズムでロバストに効率よく圧縮できる ユニバーサルデータ圧縮符号などがある．
**(B) 誤り訂正符号化**： 通信路の雑音あるいは記録媒体の傷やゴミなどで，正しく受信あるいは読み出しができない場合に，誤りを検出し自動的に訂正する．
**(C) 暗号化**： 盗聴や改ざんなどの攻撃から，情報を守る．

上記の (A)–(C) の符号化に関して，本年度得られた研究成果の概要を次節で紹介する．

## 2　平成 17 年度の成果の概要

**(A-1) 平均最適な FV 符号の一般情報源に対する性能解析** [1]

FV 符号 (fixed-to-variable length code) は，固定長のデータ系列を可変長の符号語に符号化するデータ圧縮符号のクラスであり，ハフマン符号を始めとする重要な符号をその特殊な場合として含んでいる．従来から FV 符号の性能は，無記情報源や定常エルドード情報源などに対して研究がなされているが，それらから外れる一般の情報源に対してはほとんど研究がされていなかった．情報理論の分野では，定常性やエルゴード性などの一切の仮定を置かない一般情報源を取り扱う情報スペクトル理論 (information spectrum theory) の研究が盛んに行われているが，本研究ではその情報スペクトル理論を用いて FV 符号の一般情報源に対するロバスト性を理論的に研究し，下記の事柄を明らかにした．

一般情報源 $\{X^n\}_{n=1}^{\infty}$ に対して，$X^n$ の 2 値 FV 符号の符号語を $\phi_n(X^n)$ とし，その符号語長を $l(\phi_n(X^n))$ とする．また，この符号の復号誤り率を $\varepsilon_n$ とし，漸近的に平均符号語長がエントロピーレート $H(\mathbf{X}) = \limsup_{n\to\infty}(1/n)H(X^n)$ を達成できる FV 符号を平均最適であるという．

このとき，$\{(1/n)\log_2(1/P_{X^n}(X^n))\}_{n=1}^{\infty}$ が一様可積分な情報源 $\{X^n\}_{n=1}^{\infty}$ に対して，$\varepsilon_n = 0$ を満たす任意の平均最適な FV 符号の符号語は，次式を満たすように確率収束する．

$$\frac{1}{n}l(\phi_n(X^n)) - \frac{1}{n}\log_2\frac{1}{P_{X^n}(X^n)} \to 0 \quad (1)$$

さらに，任意に与えられた $\varepsilon$ $(0 \le \varepsilon < 1)$ に対して，$\limsup_{n\to\infty}\varepsilon_n \le \varepsilon$ を満たす FV 符号で達成可能な平均符号語長を決定するとともに，その最適な平均符号語長を達成する FV 符号に対して，

*2005 年 9 月 30 日まで
†2005 年 12 月 16 日から

その符号語長 $\frac{1}{n}l(\phi_n(X^n))$ の漸近特性を明らかにした．

**(A-2) MTF データ圧縮符号の冗長性解析** [2]

MTF(move-to-front) データ圧縮法は，Recency-Rank 法，Block-Stack 法とも呼ばれ，高性能なデータ圧縮符号であるブロックソート法の中で使用されていることでよく知られている．MTF 法は，情報源アルファベットをリストとして記憶し，出現したシンボルをそのリスト内の順位として符号化し，出現したシンボルをリストの最初に移動する方式である．この MTF 法は，パソコンの日本語入力の変換候補リストの並べかえなど，データ圧縮以外でも，リストの更新手法としてさまざまなところで利用されている．

MTF 法の性能解析は，定常無記憶情報源に対してなされているが，出現シンボルをリストの最初に移すという非定常な動作のため，より広いクラスの情報源に対して解析することが困難であった．

本研究では，MTF 法の冗長度（符号語長とエトロピーレートの差）に着目してその漸近的な性能を理論的に詳細に解析した．その結果，MTF 法では，2 次マルコフ情報源に対しては，一般にエントロピーレートを達成できないことを示すとともに，MTF 法でエントロピーレートを達成できるのは，特殊な特性を持つ 1 次マルコフ情報源だけであることを明らかにした．さらに，定常エルゴード情報源，マルコフ情報源，2 値マルコフ情報源などに対して，MTF 法の漸近的な冗長度特性を解析した．

**(A-3) 反辞書法と算術符号を用いるデータ圧縮法** [3]

反辞書法 (anti-dictionary method) は，辞書法 (dictionary method) の特殊な場合として分類できるデータ圧縮符号化法であるが，通常の辞書法ではデータ系列に出現する系列を辞書に登録するのに対して，反辞書法では出現しない系列を辞書に登録することで圧縮を行う特徴を持っている．反辞書法はユニフィラな有限状態 2 値情報源で 0 と 1 の生起確率が各状態で $0, 0.5, 1$ のいずれかしか取らない（あるいはそれに近い確率分布を持つ）情報源に対して，エントロピーレート（あるいはそれに近いレート）まで圧縮できるが，一般の情報源に対しては圧縮率が悪い．しかし，反辞書法は，任意のデータ系列に対して，ユニフィラな有限状態情報源モデルを自動生成できる特長を持つ．一方，一般によく使用されている算術符号は，与えられた情報源モデルの下でエントロピーレートまでの圧縮率を達成できるが，データ系列に適した情報源モデルを求めることが一般には難しいため，通常は簡単なマルコフモデルしか用いられていない欠点がある．本研究は，反辞書法でデータ系列からユニフィラな有限状態情報源モデルを自動生成し，各状態におけるシンボルの生起頻度を用いて算術符号化を行うデータ圧縮法を提案し，従来の反辞書法や算術符号よりも高性能な圧縮を実現できることを明かにした．

**(B) 量子連接符号の構成法**

将来，量子通信が実用化したとき，量子状態の伝送時に発生する量子的な雑音を取り除く「量子誤り訂正」が必要となる．本研究は，古典符号理論の連接符号に相当する連接量子符号のあるクラスに対して，その構成法を示すと共に，その符号化率と最小距離を満たす限界式を導出している．（詳しい内容は，本報告集の「漸近的に良い連接量子符号の構成（藤田八郎）」で報告している．）

**(C-1) 一般アクセス構造に対する強いランプ型秘密分散法の構成法** [4][5]

秘密分散法 (secret sharing scheme, SSS) は，秘密情報を複数の分散情報に分散符号化し，破壊と漏洩の両方の脅威に対して，ロバストに安全な伝送や記録を実現するための符号化法である．

秘密情報を復元できる分散情報の部分集合と，秘密情報が全く漏洩しない分散情報の部分集合以外に，中間の特性を持つ部分集合を許す SSS をランプ型 SSS というが，ランプ型 SSS は符号化効率を大きく改善できる特長を持つ．

情報の一部を漏らすランプ特性において，秘密情報のどの部分も一様に曖昧さを残す「強い」ランプ型 SSS が，安全性から望ましいが，その一般的な構成法は知られていなかった．本研究では，複数の秘密情報を段階的に復号できる SSS を利用して，一般アクセス構造に対して強いランプ型

SSS を構成する具体的な手法を与えると共に，従来よく使われているランプ型 SSS は，強いランプ型 SSS の特性を持たないことを示した．

**(C-2) 量子秘密分散法の理論解析と構成法** [6]

将来，量子通信や量子計算が実用化すると，量子状態を安全に伝送したり記録したりする必要が生じる．古典 SSS と同様に，量子状態を複数の量子状態に分散符号化することにより，そのような安全性を実現するものに，量子秘密分散法 (Quantum SSS, QSSS) がある．従来の量子分散法に対する理論解析は非常に複雑であったが，本研究では，量子分散符号化および復号化の操作を量子通信路の可逆性の概念に対応させることにより，見通しのよい解析ができることを示した．その結果，QSSS の効率の限界を与える符号化定理のシンプルな別証明を与えた．さらに，ランプ型 QSSS の概念を提案し，ランプ型 QSSS の符号化効率が，古典 SSS と同様に，ランプ型でない場合に比べて符号化効率を大きく改善できることを示した．さらに，最適なランプ型 QSSS の具体的な構成方法を与えた．

**(C-3) 木構造を用いるグループ鍵管理システムの漸近特性** [7]

インターネットや放送などの公開通信路を通して，有資格者だけにコンテンツ配信を行う場合に，グループ鍵と呼ばれる秘密鍵を有資格者に配布することにより不正アクセスを防止できる．しかし，新規の加入者や脱会者がいると，その度にグループ鍵を更新しなければならない．その更新を効率よく行うために鍵配布に木構造が用いられるが，そのような木構造を用いる場合の鍵更新コストが，グループの退会確率に基づくエントロピーに漸近的に近づくことを証明した．

**(C-4) 盗聴通信路に対する多重符号化** [8]

雑音のある通信路を通して情報を伝送する場合，正規の受信者と盗聴の通信路で雑音特性が異なっていると，全く秘密鍵を用いなくても，盗聴者に 1 ビットも情報が漏れることなく正規の受信者に情報を送ることができる．そのときに送れる情報量は，秘匿通信路容量 (secrecy capacity) $C_S$ と言われるが，$C_S$ は通常の通信路容量 $C$ に比べて一般にかなり小さい値となる．本研究では，複数の情報を多重符号化して伝送することを提案し，その多重符号化を用いれば，各情報を盗聴者から完全に安全に保ったまま，トータルの伝送量として通信路容量 $C$ までの情報伝送が可能であることを証明した．

**(C-5) 相関情報源を伴うシャノン暗号システムの符号化定理** [9]

シャノンの暗号システムなど，情報セキュリティシステムの安全性は，従来条件付きエントロピーで評価されることが多かった．これに対して，暗号文 $C$ から秘密情報 $X$ を推測して当たるまでに必要な推測回数の期待値やその大偏差で評価する安全指標や，正しい $X$ を当てる確率で評価する安全性指標がある．また，秘密情報が送信情報 $X$ ではなく，$X$ と確率的な相関を持つ他の情報 $Y$ である場合も多い．本研究では，暗号文 $C$ から秘密情報 $Y$ を推測して当たるまでに必要な推測回数の期待値やその大偏差，あるいは $Y$ が当たる確率などで安全性を評価する場合の符号化定理を証明した．

**(C-6) 盗聴に対して安全なネットワーク符号化** [10]

多数の通信路からなるネットワークにおいて，通信路をつなぐ各節点が計算能力を持ち符号化が行えるとき，そのような符号化をネットワーク符号化 (network coding) という．

ネットワーク符号化を用いると，情報源点から複数の受信点に同じ情報を伝送するマルチキャスト通信を効率よく行えることが知られている．また，$n$ 本の通信路からなるネットワークにおいて，$k$ 本 ($k < n$) まで盗聴されても情報が漏れないネットワーク符号化法が知られている．これに対して，本研究では，$k$ 本以上盗聴されても，秘密情報の一部が明確に漏れてしまわない，より安全な符号化法を提案し，それが実現できる十分条件を与えた．

## 3 ロバスト符号化セミナー

本年度は，ロバストな符号化を実現するための新しい理論的および技術的進展をもたらすために，

ロバスト符号化セミナーを6回開催し，情報理論分野だけでなく関連する他分野との積極的な交流を実施した．以下に，その概要を示す．なお，会場は全て東京大学新領域創成科学研究科（柏キャンパス）複雑理工学専攻講義室で実施した．

**第 1 回** 2005 年 7 月 16 日

1. “暗号技術における擬似乱数生成,” 藤岡淳（NTT 情報流通プラットフォーム研究所）
2. “Relationship among Complexities of Individual Sequences over Countable Alphabet,” 葛岡成晃（東京工業大学）

**第 2 回** 2005 年 8 月 19 日

3. “Redundancy of Symbol Decomposition Algorithms for Memoryless Source,” 川端勉（電気通信大学）
4. “無ひずみデータ埋め込みのためのユニバーサル法,” 横尾英俊 (群馬大学)

**第 3 回** 2005 年 10 月 1 日

5. “Unconditionally Secure Steganography,” 四方順司（横浜国立大学）
6. “Web ページのアクセス数計数に対する暗号学的アプローチ,” 尾形わかは（東京工業大学）

**第 4 回** 2005 年 11 月 8 日

7. “On Gaussian Approximation, Asymptotically Optimal Linear and Nonlinear Detectors in CDMA and Multiuser Detection,” Marat Burnashev (Institute for Information Transmission Problems, Russian Academy of Sciences)

**第 5 回** 2005 年 12 月 10 日

8. “文脈自由文法を用いた効率的な圧縮アルゴリズム,” 坂本比呂志（九州工業大学 ）
9. “CFTP を用いた Perfect Sampling,” 松井知己（東京大学）

**第 6 回** 2006 年 3 月 4 日

10. “情報統計力学の方法,” 田中利幸（京都大学）

## 発表論文


[1] H. Koga and H. Yamamoto, “Asymptotic properties on the codeword lengths of optimal FV codes for general sources,” *IEEE Trans. on Inform. Theory*, vol. 51, no. 4, pp. 1546–1555, Apr. 2005

[2] M. Arimura and H. Yamamoto, “Asymptotic redundancy of the MTF scheme for stationary ergodic sources,” *IEEE Trans. on Inform. Theory*, vol. 51, no. 11, pp. 3742–3752, Nov. 2005

[3] 大川, 原田, 山本, “反辞書法に基づくモデル選択と算術符号を用いたデータ圧縮”, 電子情報通信学会情報理論研究会技術報告，IT2005-49, pp. 41–46, 2005

[4] M. Iwamoto and H. Yamamoto, “Strongly Secure Ramp Secret Sharing Schemes,” *Proc. 2005 IEEE Int. Sym. on Inform. Theory (ISIT2005)*, pp. 1221–1225, Sep. 4–9, 2005, Adelaide, Australia

[5] M. Iwamoto and H. Yamamoto, “Strongly Secure Ramp Secret Sharing Schemes for General Access Structures,” *Information Processing Letters*, vol. 97, issue 2, pp. 52–57, Jan. 2006

[6] T. Ogawa, A. Sasaki, M. Iwamoto, and H. Yamamoto, “Quantum Secret Sharing Schemes and Reversibility of Quantum Operations,” Physical Review A, vol. 72, no. 3, pp. 032318-1–7, Sep. 2005

[7] H. Sakai and H. Yamamoto, “Asymptotic Optimality of Tree-based Group Key Management Schemes,” *Proc. 2005 IEEE Int. Sym. on Inform. Theory (ISIT2005)*, pp. 2275–2279, Sep. 4–9, 2005, Adelaide, Australia

[8] D. Kobayashi, H. Yamamoto, and T. Ogawa, “How to attain the ordinary channel capacity securely in wiretap channels,” *Proc. 2005 IEEE Information Theory Workshop on Theory and Practice in Information-Theoretic Security*, pp. 13–18, Oct. 16–19, 2005, Awaji Island, Japan

[9] 林, 山本, “推測盗聴者と相関情報源を伴うシャノン暗号システムに対する大偏差理論,” 2005 シャノン理論ワークショップ (STW05) 予稿集, pp. 9–14, 2005

[10] 原田, 山本, “強いランプ型しきい値特性を持つ安全なネットワーク符号化法,” 第 28 回情報理論とその応用シンポジウム予稿集, pp. 741–744, 2005