

平成 22 年 5 月 31 日現在

研究種目：基盤研究（C）
 研究期間：2006～2009
 課題番号：18560396
 研究課題名（和文）正整数のユニバーサル符号化における理論的符号語長関数に応じた具体的符号語割り当て
 研究課題名（英文）Codeword assignments to theoretical codeword length functions on the universal integer coding
 研究代表者
 中村 博文（NAKAMURA HIROFUMI）
 都城工業高等専門学校・機械工学科・教授
 研究者番号：40189056

研究成果の概要（和文）：情報を表す基本的記号の数に限定がない場合の符号化は、任意の正整数の符号化に帰着できる。正整数の符号化について、必要な 0 と 1 の数は、表したい正整数を変数とする関数、即ち、符号語長関数で表せる。特にオーダが小さい符号語長関数については、符号語割り当ての計算量の小さいアルゴリズムが未知のものが存在していた。その中の 2 つの符号語長関数について、符号語長と同じオーダで符号化できるアルゴリズムを提案した。

研究成果の概要（英文）：When the number of elemental symbols of information is unlimited, the coding of the information arrive at positive integer coding. The number of necessary 0 and 1 for positive integer coding can be expressed as a function of the positive integer, that is codeword length function. Especially, for small order codeword length functions, there exist no codeword assignment algorithm with small complexity. For two codeword length functions, we proposed codeword assignment algorithms. The order of the time complexity of each code is as small as its codeword length.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2006年度	900,000	0	900,000
2007年度	600,000	180,000	780,000
2008年度	500,000	150,000	650,000
2009年度	500,000	150,000	650,000
年度			
総計	2,500,000	480,000	2,980,000

研究分野：工学

科研費の分科・細目：電気電子工学・通信ネットワーク工学

キーワード：情報通信工学、正整数符号化、ユニバーサル性、符号語長関数、符号語割り当て

1. 研究開始当初の背景

本研究は、アルファベットサイズが巨大または無限大の情報源や、特に上限を定めない正整数を、情報の生起確率がどのようなものであっても、ある程度良い性能で 2 値符号化することに関するものである。これは、情報

理論の情報源符号化・データ圧縮の「正整数のユニバーサル符号化」と呼ばれる分野である。情報を表す基本的記号の数が多い場合や、その数に限定がない場合の符号化は、基本的記号に通し番号を付けて正整数の符号化に帰着できる。また、巨大な数のデータの中の

ひとつを指し示す道標は正整数符号の符号語と同一視できる。

任意の正整数（以下、 n と表わす）の符号語長（単位はビットとする）を n の関数として考えたとき、この符号語長関数のオーダをどこまで小さくできるかという問題に研究者たちのアプローチが進んでいた。その成果として、一定の定式化の下で、符号語長のオーダが小さい特徴的な理論的符号語長関数がいくつか発見されている。

しかし、理論的に数式による符号語長は知られているものの、 n の具体的な符号語を、時間計算量が符号語長と同じオーダである $O(\log n)$ になるようにして、更に例えば符号に無駄な符号語がないことを保証するクラフトの不等式を等式で満たすという性質も持たせながら、どのように割り当てるのかについては、未解決の理論的符号語長関数がいくつか存在していた。

2. 研究の目的

本研究は、具体的な符号語割り当てが未解決の符号語長関数への、より簡便な符号語割り当て方法の確立や、その符号の性質の解明や、正整数符号化の応用を目的とする。

この後の記述のために次のような記号や用語を用いる。底が2の対数を簡単に \log と表す。 $\log n + \log \log n + \log \log \log n + \dots$ のように、 n について対数を取り続けながら値が非負の間足した値を $\log^* n$ と表し、対数スター関数と呼ぶ。対数スター関数で足し合わせる項の数を $w(n)$ と表す。切り上げることを $\text{ceil}(\cdot)$ と表す。ラージオーダのことを $O(\cdot)$ と表し、定数オーダのことを $O(1)$ と表す。

(1) Leung-Yan-Cheong・Cover や Rissanen によって

$\log^* n + 1.5185 \dots = 1$
の関係が成り立つことが示されていた。もしシャノン符号化を用いると、この式を切り上げた

$\log^* n + 2.5185 \dots$
ビット未満で符号化できることになるが、実際には上界はこれよりも約 0.52 ビット短くてよく、 n を対数スター関数に対して高々2ビット長い

$\log^* n + 2 \dots \dots \dots$ 式(1)
ビット以下で符号化できることが示唆されていた。しかし、式(1)以下の符号語長による時間計算量 $O(\log n)$ での符号語割り当ての方法は知られていなかった。

式(1)以下の符号語長で符号語割り当てができるアルゴリズムを確立し、クラフトの不等式を等式で満たす改良を加えるとともに、その符号の性質を明らかにすることが、

目的の一つ目である。

(2) Ahlswede・Han・Kobayashi によって、正整数 n を、仮想的に実数符号語長で

$\log^* n - w(n) + \log w(n) + d^*$
ビットで符号化できることが示されていた。ここで、自然対数の底を e とすると $d^* = \log \log e$ 、 $d^* > 1$ 、 d^* は実数符号語長としてクラフトの不等式を等号で満たすような実定数である。整数符号語長にするためにシャノン符号化を用いると、この式を切り上げた

$\text{ceil}(\log^* n - w(n) + \log w(n) + d^*)$
 $\dots \dots \dots$ 式(2)

ビット未満で符号化可能である。即ち、

$\log^* n - w(n) + \log w(n) + d^* + 1$
ビット未満で符号化することが可能である。しかし、正攻法でシャノン符号化を用いた場合の時間計算量は、符号語長の n 倍のオーダであり、 $O(n \log n)$ である。符号語長と同じ時間計算量 $O(\log n)$ での符号語割り当てができるアルゴリズムを確立し、クラフトの不等式を等式で満たす改良を加えるとともに、その符号の性質を明らかにすることが、目的の二つ目である。

(3) 可能なら符号語長関数の性質についてより詳しく明らかにすることや、正整数符号化の応用を図ることが、3つ目の目的である。

3. 研究の方法

前掲の2つの理論的符号語長関数に対して、次のように具体的な符号語割り当てに取り組んだ。

(1) まず、式(1)掲載の理論的符号語長関数への符号語割り当てについて、次の2つの方向から取り組んだ。

まず、当初の研究計画通り取り組んだのは、既に知られていた「符号化したい正整数 n を基に再帰的に定義される桁数情報の列を、複数の桁数情報毎にまとめて2値符号化する符号」の改善についての探求である。この既存符号は符号語長関数から定数項を除いた部分のオーダが目的の式(1)を下回る符号化が可能であることまでは知られていた。しかし式(1)は定数項の要請が厳しいためオーダについて単純に大が小を兼ねるようには言い切れないことから、この既存符号が式(1)を実際に満たすかどうかを探求するというのが一つ目の取り組み内容である。具体的には、この既存符号には符号のパラメータが3つあるが、それらの組み合わせを必要な範囲で網羅して調べた。

既に知られていた「対数スター関数に対し

て高々3ビット長い符号」の改善によるというのが二つ目の取り組み内容である。この既存符号は、Levenshtein-Eliasの符号と呼ばれる符号語長 $\log_2 n + \tilde{w}(n)$ ビットの符号の符号語を基に、その符号語の中の特に冗長度の高いビットを集めて再符号化することによって $\log_2 n + 3$ ビット以下にするが、定数項の3を2にすることは厳しい要請であるため容易ではない。しかし、Levenshtein-Eliasの符号において一部のビット列の置き換えを付加するだけでオーダを改善するという比較的単純な構造の符号化という点を生かすことが期待された。

本研究では、この既存符号をどのように一般化したら、時間計算量を増さずに式(1)を満たす符号が作成可能かどうかについて調べた。そして、置き換えのアルゴリズムを探求し、符号の性質の解明を行った。

(2)2番目に、式(2)の理論的符号語長関数への符号語割り当てについて、次の2つの方向から取り組んだ。

まず、取り組んだのは、研究開始時に既に知られていた「数値計算によって符号語を決定する符号」を基にした改善である。

しかし、符号の各部の一般化を試みたが、符号語長の定数項の有意な改善はできなかった。また、クラフトの不等式を等式で満たせない点についても改善策は見つからなかった。

次に、 n から再帰的に定義できる長さ情報から、全体の符号語長に影響の大きい部分を取り出し、別途符号化する方法を試みた。

その過程で、長さ情報を対数スター関数の項の数 $\tilde{w}(n)$ でグループ化することで、数値計算などの時間計算量の大きい処理をしなくても符号語長を抑えた符号語割り当てができることが分かり、 $\tilde{w}(n)$ によるグループ化方式の符号化を用いる方向で取り組んだ。

(3)3番目に、正整数符号化の応用として、画像情報の符号化に取り組んだ。特に、画像のフラクタル表現は、可逆を狙うとき、コントラストやオフセットを実数データとして符号化することが数学的には正攻法といえる。これについて、画像の再生に係する領域の縦か横の画素数が1の場合は本申請者によって整数での扱いで可能になっていたが、それは言わば1次元の線領域であった。

画像の再生に係する領域が縦横とも複数画素になる通常の画像被覆の場合について、関連する画素の階調値の影響範囲と程度を調べる形で取り組んだ。

その過程で、実数データであるコントラストやオフセットを、整数番号を付けてグループ化しても可逆にできる条件が分かり、グル

ープ化方式の正整数符号を当てはめる方向で取り組んだ。

4. 研究成果

(1) 任意の正整数を、対数スター関数に対して高々2ビット長い符号語長(前掲の式(1))で符号語割り当てすることについて、符号化及び復号するアルゴリズムを完成した。

符号化の正当性は算術符号に帰着して明らかにしたが、状態遷移を用いた符号化操作の整理によって、一般的な算術符号化が計算コストの大きい乗除算を必要とするのに対して、加減算とシフト演算のみで符号化及び復号ができる。

また、符号語長が目標の式(1)以内であることや、この符号化がクラフトの不等式を等式で満たすことを、算術符号に帰着して証明した。

既存の「符号化したい正整数 n を基に再帰的に定義される桁数情報の列を、複数の桁数情報毎にまとめて2進符号化する符号」についてそのパラメータを網羅的に調べたところ、その既存符号でも $\log_2 n + 2$ ビット以下で符号化できるパラメータが存在することが分かった。その既存符号の性質を新たに一つ掘り起こしたと言える。しかし、簡便さの点では、その既存符号で準備すべき符号として、約3億個の符号語を確定して準備しておくかまたは符号化時に作成する必要があることも分かった。本研究で確立したアルゴリズムは、数値が150組程度入った表を用意すれば手作業でも符号化や復号が可能であり大変簡便である。

算術符号化において1回1回の符号化で十分な確率を分配することは当然視されているが、本研究の表への整理では、部分的に分配する確率が不足しても、複数回の符号化において十分な確率の分配にすることで、算術符号化の計算処理に必要なビット数を1だけ減らしている。同時に、表の大きさが半分になっている。これは、一般に算術符号化を表や状態遷移に整理する際に貴重な手法として使用できる。

(2) Ahlswede・Han・Kobayashi による修正対数スター関数を用いた理論的符号語長関数(前掲の式(2))への符号語割り当てについては、対数スター関数の項の数でグループ化するグループ化方式の符号化により、式(2)と定数オーダの差で符号語割り当てができるアルゴリズムを完成した。

この符号化がクラフトの不等式を等式で満たすことを示した。クラフトの不等式を等式で満たさない既存の符号よりも定数部分が小さいことを示した。同じオーダの符号の

中では、現在知られている中で最も定数項が小さい。

(3) Ahlswede・Han・Kobayashi は対数スター関数を用いた理論的符号語長関数をいくつか提案しているが、それらの正規化定数項の算出例はこれまで殆どなかった。

正規化定数項は符号語長関数のクラフト和が1になる符号語長関数の定数項の付加項と定義できるが、クラフト和を与える式中の多数の積和項について、比較的計算がしやすい上界と下界の組で扱い、その差を狭めていき、ある程度の精度で算出することを提案した。

この手法は、他の研究者が正規化定数項の算出が必要になった際に利用できる。

(4)正整数符号化の応用として、デジタル画像のフラクタル表現の分野で、グループ化方式の正整数符号化の考え方を用いて、従来実数値で扱われていたコントラストとオフセット値を整数で分類し符号語生成することで、可逆なフラクタル表現を可能にした。

画像のフラクタル表現の研究でも非可逆なフラクタル表現の研究は多数あるが、本研究開始まで、参照画像の画素数が縦横とも複数画素の可逆フラクタル表現を可能にした研究はなかった。本研究開始後に他の研究者によって可逆の手法例(“マルチドメイン手法を用いた静止画像のフラクタル可逆符号化”、小川真一、中川匡弘、日本シミュレーション学会大会発表論文集、2006年6月)が示されているが、それは複数の参照画素を使用している。本研究の方法は符号化対象画素中の参照画素が必ず一つで済み、確実に可逆になるためのパラメータの算出が単純である。

本研究が画像の可逆フラクタル表現で世界初とは言えなくなったが、本研究の成果は、始まったばかりの画像の可逆フラクタル表現の今後の研究において参照されるべき手法の一つであるといえる。

(5)正整数符号の符号語長の評価について、Eliasによって、任意の確率分布での正整数符号の使用における性能を、平均符号語長の関数とエントロピーとの比で評価する提案がなされている。Eliasはいくつかの符号について、上界をエントロピーの関数として与えている。

これに対して、多数の確率分布について比が極大になる場合の検証を実験的に行い、具体的な符号語を有する代表的な正整数符号4つについて、平均符号語長とエントロピーとの比の上界がEliasの解析結果よりも実際には小さいことを部分的に示した。

本研究で示した範囲は部分的ではあるが、

任意の確率分布での正整数符号の使用における性能について、従来考えられていたよりも実際の冗長性は小さいことを明らかにした。

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)

[学会発表](計7件)

中村博文、淵田孝康、加治佐清光、テキストデータの圧縮効率を高める前処理の可能性について、電子情報通信学会情報理論研究会、2010年3月5日、信州大学長野キャンパス

中村博文、正整数符号を用いた平均符号語長の対エントロピー比の上界について、情報理論とその応用シンポジウム、2009年12月4日、山口県山口市ホテルかめ福

中村博文、淵田孝康、加治佐清光、画像の可逆フラクタル表現の基礎的研究 2x2画素レンジの場合、情報理論とその応用シンポジウム、2009年9月17日、新潟大学

中村博文、正整数符号化におけるAhlswede-Han-Kobayashiの実数符号語長関数の定数項の値と再帰性の拡張について、シャノン理論ワークショップ、2008年11月7日、大分県別府市ホテル清風

中村博文、淵田孝康、加治佐清光、データ圧縮の効率を高める前処理の可能性について、情報理論とその応用シンポジウム、2008年10月9日、栃木県日光市あさやホテル

中村博文、算術符号に基づいた状態遷移を用いた符号語長 $\log_2 n + 2$ ビット以下の正整数符号化、情報理論とその応用シンポジウム、2008年10月9日、栃木県日光市あさやホテル

中村博文、村島定行、対数スター関数で足される項の数でグループ化する正整数符号、情報理論とその応用シンポジウム、2006年12月1日、北海道函館市花びしホテル

6. 研究組織

(1)研究代表者

中村 博文 (NAKAMURA HIROFUMI)
都城工業高等専門学校・機械工学科・教授
研究者番号：40189056