2B-1

二項分布情報源に対してすべての2元ハフマン符号を生成する方法について

福岡　久雄† 　　　和田　雄次††
† 松江工業高等専門学校　　†† 東京電機大学

1　はじめに

二項分布情報源とは，2 値無記憶情報源の拡大情報源として得られる情報源である．元の 2 値無記憶情報源を情報源シンボル 0, 1 と各シンボルの生起確率 $P(0) = p$，$P(1) = 1 - p$ （但し，$1/2 \le p < 1$）により定義すると，その n 次拡大情報源の確率分布は，生起確率 $p^{n-i}(1-p)^i$ のシンボルの個数が $\binom{n}{i}$ 個であるような二項分布となる（但し，$i = 0, 1, \ldots, n$ ）．これを n 次の二項分布情報源と呼び，シンボル 0 を優勢シンボルと呼ぶ．

拡大次数を固定した状態でも，p の値に応じて（非可算個の）異なる二項分布情報源が得られ，それぞれに対して 2 元ハフマン符号が構成可能である．しかし，区間 $[1/2, 1)$ のある有限個の分割それぞれに対応して，有限種類のハフマン符号しか構成され得ないことが知られている．すなわち，同一分割内の p に対する二項分布情報源に関しては同じハフマン符号が構成される．ここで，同じハフマン符号とは，同じ符号長ベクトル（符号長を昇順に並べたベクトル）を有する符号という意味である．すなわち，符号語のビットパターンが異なっていても，符号長ベクトルが同一であれば，同じハフマン符号と見なす.

本研究では，ある次数の二項分布情報源に対して構成され得る相異なるハフマン符号をもれなく生成する方法を検討する．基本方針は，2^n 個のシンボルを持つ任意の情報源に対して構成可能なハフマン符号の中から，二項分布情報源に対応するもののみを抽出する手順を見出すことである．

2　関連研究

Fenwick は，2 次および 3 次の場合に限って，二項分布情報源に対して構成され得るハフマン符号と，各ハフマン符号に対応する上記分割を明らかにしている [1]．その方法は以下の通りである．まず，計算機シミュレーションによって，p の値を徐々に変化させながら，各 p に対する二項分布情報源を生成し，そのハフマン符号を構成する．この過程で得られる相異なったハフマン符号を列挙し，そこから上記分割を計算で求める．

この方法は,「p を徐々に変化させる」という計算精度に依存する処理が含まれていることが難点である．本研究では，これとは異なって，計算精度の問題を可能な限り排除した方法を目指している．

3　ハフマン符号長ベクトルの網羅的生成

Norwood[2]，Even and Lempel[3] は，シンボル数 N の情報源に対して構成可能なハフマン符号の数やその符号長ベクトルを求める方法を報告している．最近，Hoffman 等は，シンボル数 N のある情報源に対するハフマン符号の符号長ベクトルを基点として，シンボル数 N の情報源に対して他に構成可能なハフマン符号の符号長ベクトルを辞書順に順次生成する方法を報告している [4].

また，Jakobsson[5] や長谷川等 [6] の研究によれば，p がある値から 1 の間にある場合には，ハフマンの符号化手順を実行することなく，二項分布情報源に対するハフマン符号の符号長ベクトルを直接求めることができる．本稿では，これを JH ベクトルと呼ぶ.

提案する方法では，最初に n 次二項分布情報源の JH ベクトルを基点として，Hoffman 等のアルゴリズムを適用し，構成可能なハフマン符号長ベクトルを網羅的に生成する．それらから，以下の性質 1 に反する符号長ベクトルを削除した結果を，JH ベクトルを先頭に辞書順に整列させたリストを L と呼ぶ．L に対して，次項に述べるフィルタリング手順を適用する．

性質 1 符号長ベクトルの要素（符号長）において，同じ生起確率のシンボルに対応するものの値は高々 1 しか違わない.

4　フィルタリング手順

フィルタリングにおいて，二項分布情報源に対応する符号長ベクトル（二項ベクトル）に関する次の性質 2 が重要な役割を果たす.

性質 2 ある二項対応ベクトルと，辞書順でその次に位置する二項対応ベクトルとの差分ベクトルにおいては，その要素は －1, 0, 1 のいずれかである.

フィルタリング手順は以下の通りである.

A Method for Generating All Binary Huffman Codes for Binomial Sources

† Hisao FUKUOKA (fukuoka@matsue-ct.ac.jp)
†† Yuji WADA
Matsue National College of Technology (†)
Tokyo Denki University (††)

Step-1 $U = JH$ ベクトル とする．

Step-2 リスト L において U 以降に位置する符号長ベクトルの中から，性質 2 を用いて U の次に位置すべき二項対応ベクトルの候補を絞り込み、それらを V_1，V_2，…，V_k とする．

Step-3 U の平均符号長と V_1，V_2，…，V_k それぞれの平均符号長を等しいと置くことによって，p に関する k 個の多項方程式を得，それらの中で最大の解を区間 $[1/2, 1)$ に持つ方程式を与える V_i を U の次の二項対応ベクトルと決定し、出力する．

Step-4 V_i が，すべての要素が n であるような符号長ベクトルであれば終了する．そうでなければ，$U = V_i$ として，Step-2 へ行く．

5 具体例

3 次の二項分布情報源を用いて，提案する方法の概要を示す．この場合の JH ベクトルは，(1,3,3,3,5,5,5,5) であり，L は以下のような 9 個の符号長ベクトルからなる．第 1 回目の Step-2 終了段階で，以下の V_1〜V_6 が求められる．

$U =$	(1,3,3,3,5,5,5,5)	
$V_1 =$	(1,3,3,4,4,4,5,5)	
$V_2 =$	(1,3,4,4,4,4,4,4)	
$V_3 =$	(2,2,2,3,4,5,6,6)	
$V_4 =$	(2,2,2,3,5,5,5,5)	
$V_5 =$	(2,2,2,4,4,4,5,5)	
$V_6 =$	(2,2,3,3,4,4,4,4)	
	(2,3,3,3,3,3,4,4)	第 6 要素，性質 2NG
	(3,3,3,3,3,3,3,3)	第 1 要素，性質 2NG

Step-3 において得られる，p に関する多項方程式の解は次の通りである．

$$V_1 : p = 0.6666666667$$
$$V_2 : p = 0.6403882032$$
$$V_3 : p = 0.6180339888$$
$$V_4 : p = 0.6666666667$$
$$V_5 : p = 0.6666666667$$
$$V_6 : p = 0.7071067812$$

V_6 が最大の解を与えることから、JH ベクトルの次の二項対応ベクトルを V_6 と決定し、出力する．

次は，V_6 を U と設定して，次に位置すべき二項対応ベクトルを求める．第 2 回目の Step-2 終了段階では，以下の V_1 と V_2 に絞り込まれる．

$$U = (2, 2, 3, 3, 4, 4, 4, 4)$$
$$V_1 = (2, 3, 3, 3, 3, 3, 4, 4)$$
$$V_2 = (3, 3, 3, 3, 3, 3, 3, 3)$$

Step-3 において得られる，p に関する多項方程式の解は次の通りである．

$$V_1 : p = 0.6666666667$$
$$V_2 : p = 0.5969682832$$

V_1 が最大の解を与えることから、U（上述の V_6）の次の二項対応ベクトルを V_1 と決定し、出力する．

以下，同様の処理を繰り返す．

6 おわりに

ある次数の二項分布情報源に対するすべてのハフマン符号長ベクトルを生成する方法について報告した．現在，計算量の観点から、より高次の二項分布情報源に対する本方法の有効性を確認中である．

なお、今回報告した方法では最終段階で p に関する多項方程式を解くことが必要であり，その点では完全に組合せ論的な方法にはなっていない．この点を改良し，数値計算を必要としない方法の確立が今後の課題である．

本研究の一部は，日本学術振興会科学研究費補助金（基盤研究（C），課題番号 17560356）を受けて行った．

参考文献

[1] P.M.Fenwick, “Huffman code efficiencies for extensions of sources,” IEEE Trans. Commun., vol.43, no.2/3/4, pp163-165, Feb./March/April 1995.

[2] E.Norwood, ”The number of different possible compact codes,” IEEE Trans. Inf. Theory, vol.IT-13, no.5, Oct.1967.

[3] S.Even and A.Lempel, “Generation and enumeration of all solutions of the characteristic sum condition,” Information and Control, vol.21, pp.476-482, 1972.

[4] D.Hoffman,P.Johnson and N.Wilson, ”Generating Huffman sequences,” Journal of Algorithms, vol.54, pp.115-121, 2005.

[5] M.Jakobsson,”Huffman coding in bit-vector compression,” Information Processing Letters, vol.7, no.6, Oct.1978.

[6] 長谷川まどか，加藤茂夫，山田芳文，“低エントロピー 2 値無記憶拡大情報源に対するコンパクト符号構成法についての一考察，”信学論 (A)，Vol.J80-A，No.9，pp.1483-1489，Sept.1997.