
河村，江口，重村氏の論文「ハフマンコード表の圧縮と
その応用」に対する意見

回答に対する質問

森 井 昌 克

**Comment for “Compression of Huffman Code Table
and Applications” by T. Kawamura, Y. Eguchi and
T. Shigemura**

Revised

MASAKATU MORII

ご回答にありましたように，貴論文 1) での CHT では，ハフマン符号の符号長をさらにハフマン圧縮し

た符号を元の文字コードに割り当てることによってハフマン符号表の記述量を削減しています。しかしながら次の2点について疑問があり、論文としての正確さに欠けると考えられます。

1. 貴論文¹⁾ 2章において従来法の説明を与えていますが、圧縮においてハフマン符号のパターンが本質ではなく、その長さが本質であることは従来から良く知られた結果です。また、それに要する記述量は n シンボルに対するハフマン圧縮では $n-2$ ビットです。さらに、回答を頂きましたように符号木の記述を用いる場合、元の文字コードとの対応表が必要ですが、その対応表は元の文字コードの順序付けで十分であり、したがって従来法としてのハフマン符号表の記述量は $\log_2(n!) + (n-2)$ 以下となり、より記述量を削減する方法²⁾ も知られております。すなわち、以上のような従来法についての正確な記述が必要ではないでしょうか。
2. 提案されているハフマンコード表の圧縮法の有効性が不明確です。すなわち比較対象がきわめて効率の悪い方法（論文¹⁾ で従来法として与えられている方法）であり、常に有効であるか否かが明確ではありません。特に記述量として $\log_2(n!) + (n-2)$ を要す

る方法と比較した際においてすら、必ずしも優れない場合が存在すると考えられます。また、先に述べましたようにハフマンコード表としてはプロパーワード (proper word; PW)^{3),4)} を表す符号木の情報と元の文字コードの順序付けで十分であり、この記述量²⁾ との比較も与えられていません。以上の疑問点から、論文¹⁾ での「画期的な圧縮率を得た」とは必ずしも主張できないのではないのでしょうか。

参 考 文 献

- 1) 河村知行, 江口賢和, 重村哲至: ハフマンコード表の圧縮とその応用, 情報処理学会論文誌, Vol. 35, No. 2, pp. 267-271 (1994).
- 2) 半田志郎, 田中初一: ハフマン符号の符号帳の記述に関する一考察, 電子情報通信学会論文誌 (A), Vol. J70-A, No. 10, pp. 1501-1503 (1987).
- 3) Even, S. and Lemple, A: Generation and Enumeration of All Solutions of the Characteristic Sum Condition, *Inf. Control*, Vol. 21, pp. 476-482 (1972).
- 4) 森井昌克, 今村恭己: 符号長の組の異なるすべての2元ハフマン符号の生成について, 信学技報, IT 82-42, pp. 1-4 (1983).

回答 その2

河 村 知 行 江 口 賢 和 重 村 哲 至

Reply (Part 2)

TOMOYUKI KAWAMURA, YOSHIKAZU EGUCHI and TETSUJI SHIGEMURA

ハフマンコードを表現するには、各コードのパターンを記録する必要はなくその長さだけ記録すればよいことは、本論文のまとめにもあるように以前に報告されたことです。CHTのその部分はそのことの再発見ということになると思います。しかし、本論文はその標題にもあるように、その応用まで含めた実用的な意味でのハフマンコード表の実現方法について論じたものです。

第一の特徴は、「多くの値をもつ文字」のハフマンコード表にCHTが適しているということです。2回

目のご意見の文献2)には2つのハフマンコード表の表現法がありますが、実用性の観点から2番目の方法(以後、方式Aと呼ぶ)がCHTと比較する対象になるのだと思います。方式Aではその表の大きさは約4Qビットです。これは、多くのデータに対する実験で $1\max - 1\min + 1$ の値が9以上になるからです。一方、CHTでは本論文の4.1節にあるようにその大きさは約2.7Qビットです。CHTの方が小さいのは2段階の符号化を行っているからです。

また、方式AにおけるQは出現しない文字も含め

ての文字数です。一方、CHT方式のQは出現した文字のみの文字数です。存在判定情報とHC表の合計が2.7Qビットです。このことは、出現文字数が少ないときにCHTが良い性能を発揮することを意味します。CHT方式を改造して、存在判定情報の表現範囲を255から2047に拡張し、その中に出現文字がまばらにしか存在しないような場合に、方式AとCHTの差は顕著なものとなります。

第2の特徴は、「EPEM」という適用の成功例を挙

げていることです。compressのようにCHTを適用するとかえってデータ量が増える場合もあるので、応用例を挙げることは意義があると思います。

最後に、「画期的」という言葉に対するご指摘ですが、本論文にあるように「画期的」なのはCHTではなくEPEMです。当時、最も良い圧縮率を示すとされていたcompressに比べて、EPEMは非常によい圧縮率を示したからです。最近使用されているgzipは、EPEMとほぼ同じ圧縮率を示します。