特集⊙データ圧縮

ハフマン符号木に関連した話題

Julia Abrahams(和訳: 山本 博資)

1 はじめに

ハフマン符号化[1]は最もよく知られた符号木問題であるが，それに関連した興味深い問題がいくつも存在することや，ハフマン符号の性質に関して多くの有用な結果が得られていることは，あまり知られていない．本論文では，ハフマン符号木問題，ハフマン符号木の特性およびそれに関連した問題(例えば，情報源アルファベットが可算無限個の場合，符号語が辞書順制約を受けるHu-Tucker符号木問題の場合，最大符号語長制約を受ける場合，平均符号語長以外の他の特性の最小化問題など)を紹介する．なお，符号の同期特性，符号語シンボルが不均一コストを有する場合の符号化，分解木(つまり可変長-固定長(V-F)符号木)，符号木の数え上げや表現方法などを含む符号木および分解木に関するより広範囲なサーベイが筆者[2]によりなされている．

ハフマン符号化問題では，既知で離散的な確率分布に従う定常無記憶な確率変数が情報源から出力されると仮定している．ハフマン符号は，固定長-可変長(F-V)の無損失(lossless)符号(つまり，情報源の1シンボル(あるいは固定長系列)を可変長の符号語シンボル列に符号化し，その符号語が誤りなく受信できるという仮定の下で，符号語から一意に元の情報源シンボル列が復号でき

[筆者紹介]

ジュリア アブラハムズ．1974年Yale大学B.S. degree，1975年Stanford大学M.S. degree，1979年Princeton大学Ph.D.(電気工学・計算機科学)．Rice大学電気・計算機工学科およびSpringer Veralagでの勤務を経て，1988年～1998年の間，海軍研究室(Office of Naval Research)で確率統計関係の基礎研究プログラムを統括．1998年から，Rutgers大学DIMACS(離散数学および理論計算機科学)センターの副長およびリサーチ教授．1994年以来IEEE情報理論ソサエティ理事，IEEE DCC 94およびDCC 95(Data Compression Conference)，ISITA 96(情報理論とその応用国際シンポジウム)のプログラム委員．情報源符号化，データ圧縮，組合せ論的検索問題などに興味を持つ．

[英文タイトル] Huffman Code Trees and Variants

[訳者] やまもと ひろすけ．東京大学大学院工学系研究科．

るような符号)の一つである．一意復号可能なためには，符号木の各内部節点から出ている各枝に異なる符号語シンボルを割り振った木のパスとして符号語が表現されていればよく，この符号構成で平均符号語長が最小となる符号を作ることができる．各情報源シンボルを葉に割り振り，根からその葉までのパスがその情報源シンボルに対応した符号語となる．復号は，根から符号語シンボル列にしたがって符号木をたどり，たどり着いた葉に対応した情報源シンボルを復号すればよい．引き続く符号語シンボルに対して再び根からたどりながら復号を繰り返して行くことにより，元の情報源系列を復号できる．このように，ハフマン問題を始めとするほとんどのF-V無損失符号化問題は，そのようなラベル付けされた木を求める問題である．ハフマン問題は，符号語長(根から葉までのパス長)を情報源の確率分布で平均したときの平均符号語長を最小化する符号木を構成する問題となる．生起確率の大きい(小さい)情報源シンボルに短い(長い)符号語を使うことにより，平均符号語長を最小にできる．

ハフマン問題は探索問題への応用において他の意味付けを与えることができる．木の各内部節点を1つの検査に対応させ，その節点から出ている枝をその検査で得られた異なる結果に対応させる．木の根から葉までのパスに対応した一連の検査結果で分類される項目とその葉とを1対1に対応付けることができる．起りやすい項目はより少ない検査で，また起りにくい項目は多くの検査を用いることにより，最小の平均検査回数で一つの識別項目を見つけることができる．ハフマン符号化[1]はあらゆる情報理論の教科書[3]~[6]で紹介されており，また多くの計算機科学の教科書[7]~[10]でも，探索問題へ応用を考慮しながら紹介されている．組合せ論的な探索問題としてのハフマン符号化は特にAhlswedeとWegener[11]およびAigner[12]の本で詳しく述べられている．

2値V-F情報源符号化では，符号木はK個の情報源シンボルの各々を，2値符号語シンボル列に対応付けるために使われる．符号木は，K個の葉と木の頂点に位置する根を含めてK-1個の内部節点を持つ完全木であり，各内部節点から1対の枝が出ている．左の枝に0，右の枝に1の符号語シンボルがラベル付けされている．また，全ての葉にはK個の情報源のシンボルの一つがラベル付けられている．根から葉までの符号木のパスが，葉の情報源シンボルに対応した符号語シンボル列を与える．符号化問題は，木構造に関する何らかの拘束条件下で，ある性能を最適化する符号木を見つけることである．$r>2$であるr値情報源符号化では，木の内部節点は2~r個の枝を持つことができ，その柔軟性によりr値符号木問題はより複雑なものとなる．しかしながら，ほとんどの文献では，r値符号化へは単純に拡張できるとして2値符号木問題しか取り扱われていない．

ハフマンアルゴリズムは，平均符号語長を最小化する符号木を見つけるという基本問題に対する解を与える．つまり，確率分布$P=\{p(i), i=1, \cdots, K\}$で

生起する情報源シンボルに対して，$\sum_{1\leq i<K} p(i)l(i)$ を最小にするパス長 $L=\{l(i), i=1, \cdots, K\}$ を持つ符号木を与える．この符号化問題では，他に拘束条件は存在しないが，符号語長の集合 L を持つ 2 値符号木が存在するための必要十分条件として，Kraft の不等式 $\sum 2^{-l(i)}\leq 1$ を満たさなければならない．その結果，最小の平均符号語長 l^* は，エントロピー $H=-\sum p(i)\log p(i)$ に対して，$H\leq l^*<H+1$ を満たす．ここで，確率分布 P が全ての i に対して $p(i)=2^{-l(i)}$ となる dyadic な分布のときに等号が成立する．最小の冗長度 l^*-H で考えると，$0\leq l^*-H<1$ となる．与えられた P に対して H が定数であることより，ハフマン符号化問題は，冗長度を最小にする L を求める問題ということもできる．また，符号木で定まる dyadic な確率分布 $Q=\{q(i)=2^{-l(i)}, i=1, \cdots, K\}$ は，最小識別 (minimum discrimination) の意味 (Kullback-Leibler 距離の意味) で P に最も近い dyadic 分布となる．したがって，ハフマン符号化は $\min_L \sum p(i)\log_2 p(i)/q(i)$ を最小にする dyadic 分布 Q を与える．

ハフマンアルゴリズムはボトムアップ型の結合アルゴリズムである．各段階において，最も小さな確率を持つ 2 つの節点 (または情報源シンボルに対応した葉) を兄弟節点として "結合 (merge)" して親の節点を作る．その親の節点に 2 つの子の確率の和を割り振り，子供の 2 つの節点は対象から取り除かれる．

なお，ハフマンアルゴリズムにより作られる符号木は一意であるとは限らない．また，符号語長の集合はハフマン符号と同じであるが，ハフマンアルゴリズムでは生成されない符号木も存在する．しかし，これらの区別はここでは重要ではない．

2 ハフマン符号木の特性

情報源の確率分布の一部 (あるいは全部) の情報から，そのハフマン符号木の形を特徴付ける問題を多くの研究者が取り扱っている．その結果，符号を実際に構成することなくハフマン符号木の形をある程度知ることができる．Katona と Nemetz[13] は，フィボナッチ数の比を含む特殊な情報源確率分布に対するハフマン符号木の特別な場合の特性に基づいて，確率 $p(i)$ を持つ情報源シンボルの符号語長の上界を求めた．彼らは，平均符号語長に対するエントロピー限界を導出するために，情報シンボルの自己情報量 $-\log_2 p(i)$ とその符号語長 $l(i)$ の間の関係を見つける目的でその研究を行った．ハフマン符号の個々の符号語長に関する研究が文献[14]～[21]で行われている．Schack [22] は，自己情報量と符号語長が最も離れている場合に対する情報源シンボルの生起確率の限界を求める Katona-Nemetz の研究を継続している．Cheng ら [23] は，ハフマン符号を固定長符号の代わりに用いる場合に，低確率のシンボルが高確率のシンボルより多くなって，一時的に膨張が生じる確率の限界を求めるために，Schack の手法を利用している．

最も確率が大きい情報源シンボルの生起確率 $p(1)$ が分かっているとき，その符号語長を1ビット以内で決定できる[14],[15],[24]．これらの結果は，与えられた $p(1)$ を用いてハフマン符号の冗長度を $[0,1)$ の範囲よりも正確に求めるために使うことができる．冗長度の限界に関する最も包括的な論文はManstetten[25]のものである．この冗長度の限界に関する研究はGallager[26]に始まっており，最大確率 $p(1)$ の代わりに最小確率 $p(K)$ が分かっている場合を含めて，この問題に多くの研究者が貢献している[27]~[33]．順序付けられた情報源確率のある部分集合が与えられているとき，その冗長度の限界を求める有用な一般的な手法がYeungの冗長度定理[34]であり，幾つかの冗長度限界は，彼の結果の特殊な場合となっている．

Golomb[35]は，複数の異なる符号木が同じ最小平均符号語長を持つような，有限情報源の確率分布について考察を行っている．

Geçkinli[36]は，最も長い符号語と最も短い符号語の差が0または1となるための情報源確率 P の必要十分条件を与えている．また，KatonaとLee[14]は，その差が与えられた整数値以下になるための十分条件を与えている．Geçkinliは，短い太った符号木がどのような時に最適になるかという問題を取り扱っているのに対して，Katona-Nemetz[13]の研究は，符号語長 $\{l(i)=i, i=1, \cdots, K-1, l(K)=l(K-1)\}$ を持つ細長い符号木がどのような場合に最適になるかという問題を取り扱っている．また，Vinokur[37]も独立にこの問題を取り扱っている．

ハフマン符号木のバランス特性，すなわち，符号木の内部節点に対応した重みの相対的なサイズに関する研究もなされている[38]~[40]．

情報源確率分布のあるパラメータ族に対しては，ハフマン符号木の形が明確に知られている．2値定常無記憶情報源出力系列をある固定長で区切り，そのブロックを1シンボルとみなすと，その拡張シンボルは2項分布を持つ．Jakobsson[41]はある条件を満たす2項分布に対して，ハフマン符号の形を明確に与えている．Stubley[42]は，全てのパラメータ値に対する2項分布を考え，この場合におけるハフマン符号化の冗長度を解析し，冗長度の限界を得ている．また，Stubley[43]は r 値定常無記憶情報源の固定長系列から得られる多項分布の場合も考察している．2項分布を取り扱った他の研究者は，平均符号語長がブロック長に対してどのようになるか，特にブロック長の増加につれて平均符号語長が単調に減少しないという問題に興味を持っている[24],[44]~[46]．

固定長の代わりにV-F符号を用いてシンボル拡張したものに対するハフマン符号化に関しても同様の問題が存在するが，それらはV-V(可変長-可変長)符号化あるいは双対木符号化の枠組みの中で考えることができる[2]．

ハフマン符号化と関連して研究されている他の分布は，有限の幾何分布である．この分布に対するハフマンの符号化はグループ分け検査で使われる．つま

り，一連の検査に基づいて，ある集合のすべての要素を2つのクラスのどちらかに分類する組合せ論的な探索問題において生じる．その各検査では，与えられた部分集合のすべての要素が一つのクラスに属しているかどうかが判断できるような検査が用いられる．Hwang[47]は，この分布に対して，個々の符号語長を求めることなく平均符号語長を得ている．一方，文献[48]では，符号語長があるパラメータ値に対してのみ明確に求められている．この問題は[49]，[50]などでも取り扱われている．

その他にもハフマン符号が利用できる有限情報源分布の他のパラメータ族が幾つか存在する．一様分布は，Geçkinli[36]で取り扱われているように，短い太ったハフマン符号木を常に与える．Günther と Schneider[51]および Campbell[52]は典型系列からなる K シンボルの典型情報源を考え，そのエントロピーと一様情報源のエントロピーの比較を行っている．この差は，大きい K に対して非常に小さくなるため，大きな典型情報源に対しては，短い太った木がほぼ最適となる．dyadic な確率分布に対しては，もちろん直ちにハフマン符号木が定まる．この場合のハフマン符号は冗長度がゼロであり，符号語シンボルの0と1の頻度が平均として等確率で生起するという興味深い性質を持っている．なおこの性質は，他の分布では一般には成立しない[53]．確率 $p(i)$ が $1/i$ に比例する有限ゼータ関数確率分布は，Gutman[54]により研究され，ある有限情報源アルファベットサイズに対しては，ハフマン符号木が明確に見つけられることが示されている．Tucker[55]は，$i(i=1, \cdots, K)$ に比例する確率を調べている．Hwang[56]は確率がただ2つの値のどちらかしか取らない場合を取り扱っている．あるデータ構造への応用において生じる特殊な形の分布が，文献[57]でハフマン符号化されている．その他にも，最小平均符号語長が明確に知られている可算無限情報源確率分布が幾つか存在するが，それらは次節で取り扱う．

Gallager が示した "sibling property"[26]は，符号木がハフマン符号木となるための必要十分条件を与えている．その条件は，木の各節点を確率の非増加順に並べたときに，兄弟節点が隣同士となるように並べることができることである．この性質は，与えられた符号木がある確率分布に対してハフマン符号木になっているかどうかを調べるのに使うことができる．

Longo と Galasso[58]はハフマン符号に対する最小識別問題の手法をとり，ある dyadic な分布と同じハフマン符号を持つ情報源確率分布の条件を見つけている．[59]でその拡張が取り扱われている．識別問題は，ある確率分布を持つ情報源が，異なる確率分布に基づいて過って符号化されるような確率分布のミスマッチを評価するときにも生じる[60]～[62]．

Hu と Tucker[63]および Hwang[64]は，実際に符号を構成することなく，異なった情報源確率分布に対してハフマン符号の平均符号語長を計算する問題

を考えている．彼らは，ハフマン符号の平均符号語長に関する不等式に対して十分条件となる情報源確率分布のある汎関数に関する不等式を与えている．彼らのアプローチは，不等式理論における majorization の概念と関係している．

Hwang の論文[64]は，他の点で興味深い．かれは，r 値ハフマン符号よりもさらに一般的なハフマン符号木を取り扱っている．その木では，各内部節点から出ている枝の数が，節点ごとで可変なある整数値になっている．Chu と Gill[65]も可変の内部節点次数を持つ符号木を取り扱っている．なお，Hwang と Chu-Gill は後で述べる Hu-Tucker 問題も議論している．

3 可算無限情報源アルファベットとユニバーサル符号化

ハフマンアルゴリズムは確率が最も小さい2つの情報源シンボルを結合していくボトムアップなものであるため，可算無限個のアルファベットにはハフマン符号化を直ぐには適用できない．しかし，Gallager と Van Voorhis[66]は，可算無限の分布から得られる有限分布に対するハフマン符号を構成し，その極限を取る手法で，可算無限の幾何分布に対する最小平均符号語長符号を得ている．彼らの研究はこの問題に対する Golomb[67]の初期の研究を拡張したものである．なお，幾何分布は定常無記憶な2値確率変数によって生成されるランレングスを記述する場合に生じる．同様の手法は，可算無限情報源分布の他のパラメータ族に適用されたり，あるいは特殊な構造を持つ無限木が最適となるための情報源確率分布に対する十分条件を求めるために用いられている[68]～[70]．Linder[71]は，この一般的なアプローチが情報源のエントロピーが有限なときはいつも最適な符号を与えることを証明したが，それは構成的な手法とはなっていない．Han, Kato, Nagaoka[72]は，無限エントロピー分布に対して証明を拡張している．

幾何分布のパラメータに依存した整数によってパラメータ化されている Gallager と Van Voorhis の無限符号木は Elias[73]の意味でユニバーサルではない．無限アルファベットに対するユニバーサル符号は次のような性質を持っている．より短い符号語がより確率の大きなシンボルを表現するために使われているとき，任意の可算無限情報源分布に対して，情報源エントロピーの定数倍でその平均符号語長が上からおさえられる．

多くの著者[74]～[85]がユニバーサル符号に関連して可算無限符号木を議論している．文献で取り扱われている多くのユニバーサル符号木は，[86]の3.3節で記述されているように命数法(Systems of numeration)[87]に基づいた共通の枠組みにはいる．命数法に基づくユニバーサル符号の具体的な例が[88]～[90]あるいは[91]に示されている．その一般的なアイデアは，次のようなものである．各正整数 x を2値符号木がそれぞれ必ず存在する2つの要素からなる2値の符号語に対応させる．無限系列 $V=\{V_i, i=1,2,\cdots\}$ を定め，$\sum_{1\le i\le j}$

$V_i < x \leq \sum_{1\leq i\leq j+1} V_i$ と $r = x - \sum_{1\leq i\leq j} V_i - 1$ を満たすペア (j, r) で $\sum_{1\leq i\leq j} V_i$ に対する相対位置を用いて x を表現する．j を表わすための j ビットの単進符号と r を示す $\lceil \log_2 V_i \rceil$ ビットの系列を連接して符号語を作る．もちろん，非ユニバーサルな可算無限アルファベット符号木も，特に r の表現に可変長系列を許すことにより，この命数法の2つの要素を用いて説明できる．無限符号木の観点に入らない(したがってここで述べていない)ようなユニバーサル符号化に対する他の様々なアプローチが存在する．

4 ハフマン符号化問題における辞書順制約: Hu-Tucker 問題

符号化される情報源シンボルが，符号木上で固定された左から右への順序で，葉に割り振らなければならない場合を考える．つまり，2値符号語がある特別な辞書順(あるいはアルファベット順，線形順序)を満たさなければならないとする．ある情報源の確率分布とそのシンボルの線形順序制約に対して，ハフマン符号木(あるいは，最小平均符号語長の木が一意でない場合は，ハフマン符号木と同じ平均符号語長を持った木)が希望した線形順序を満たすことがあるかもしれない．例えば，情報源シンボルがそれらの生起確率の単調増加あるいは単調減少順に線形順序をなす場合などである．しかし，一般にはハフマン符号木は希望する線形順序を満たさず，情報源シンボルに対する線形順序制約の下で平均符号語長を最小にする符号木は，ハフマン符号木の最小値より大きい平均符号語長を持つ．そのような線形順序を満たす符号木は，Hu と Tucker のアルゴリズム[92]～[96]により得ることができ，その平均符号語長は $[H, H+2)$ の間に入る．なお，辞書順符号に関する最初の研究は，Gilbert と Moore [97]まで遡る．

Hu-Tucker アルゴリズムの中心的な処理は，ハフマンアルゴリズムのように2つの節点を1つに結合する処理であるが，結合するシンボル対は，ある一般的な意味での線形順序で隣接するものに限られている．この結合操作を繰り返し適用することにより符号木が作られるが，その木を再配置することにより，元の木と同じ符号語長をもち線形順序を満たす木を作ることができる．単純なハフマンアルゴリズムと比べて理解するのが難しいアルゴリズムである．

Hu-Tucker 符号木に対して，Kraft の不等式に相当するもの(つまり，辞書順制約された符号木が存在できる符号語長の集合の特徴付け)を幾人かの著者が証明している[98]～[101]．この結果を基に，情報源確率とその線形順序について部分的に情報が与えられたとき，$[0, 2)$ の間にある Hu-Tucker 符号の冗長度をより精密に求めたり，あるいは，Hu-Tucker の平均符号語長の限界を同じ確率分布に対するハフマン平均符号語長で表現するような多くの研究結果が得られている[98]～[101]．Hu-Tucker 問題に対する情報源分布の最悪順序は，鋸歯状の順序であるという Kleitman と Saks[102]の結果は，Hu-Tucker 平均

符号語長の上界も与えている．Hu と Tan[103]は，Hu-Tucker 問題を特殊な場合として含むより一般的な順序付き探索木において，その性能の上界を導くために同様の手法を用いている．Hu-Tucker 符号木を特徴付ける他の一連の結果は Ramanan[104]に因っているが，彼は特殊な Hu-Tucker 符号木が最適となるための十分条件である情報源確率に関する不等式を与えている．なお Ramanan の研究の動機は，ある定められた情報源確率に対して与えられた木が Hu-Tucker 木になっているかどうかを，最適な木自身を構成するよりも計算量的に簡単にテストする方法を見つけ出すことであった．

Hu-Tucker の線形順序よりもより厳しい制約をハフマン問題に課しているものとして Van Voorhis 問題[105]があり，そこではさらに符号語長が単調に増加しなければならない．その制約の結果，Van Voorhis アルゴリズムによって得られる最小平均符号語長は，Hu-Tucker 符号の符号語長を用いて上界が与えられるような平均符号語長を持つ．

最小平均符号語長符号化問題以外でハフマンおよび Hu-Tucker 符号木が使われる組合せ論的な検索問題は，ある項目が与えられた集合に含まれているかどうかを区別できる一連の検査に基づいて，識別項目を検査する問題である．各項目が識別項目である可能性の程度を表す事前確率が各項目ごとに与えられているものとする．もし項目が，石油のパイプラインのように左から右に固定された順序で並んでいるあるシステムの構成要素であり，検査が検査点より左に識別項目(つまり欠陥)があるかどうかを検査するものである場合，その識別項目を同定するための検査の平均回数を最小にするテスト木は，Hu-Tucker 木となる．半順序内に単一の識別要素を持つような完全連結された半順序構造にしたがって各要素が配列されている拡張問題も研究されているが[107]，まだ完全には解かれていない．情報源シンボル上のこのような半順序制約に適用できる Hu-Tucker アルゴリズムの一般化は興味深いかもしれない．この場合やあるいは他の検索問題における拘束された木に対して適当な Kraft 型の不等式を導くことも(ほとんどの場合には無理かもしれないが)興味深い．

線形順序に並んだ項目を並列検索するため，あるいは順序付けのないハフマン問題の並列問題に対して，最適な森(複数の木)は，Hu-Tucker やハフマン結合アルゴリズムを K-m 回の結合の後ストップさせせることにより求められる[92]．森の中の m 個の木のおのおのを同時に検索することができ，それらの検索者のだれかが 1 つの識別項目を探し出すまでにかかる平均時間が最小になる．これらの問題に対して，上記で得られる最小平均パス長に関するエントロピー限界が[108]で与えられている．

なお，2 値でない Hu-Tucker 問題は，本質的にまだ未解決のままである．

5 ハフマン符号化と Hu-Tucker 符号化における最大符号語長制約

ハフマンと Hu-Tukcer 問題は，全ての符号語がある最大値より長くならないという制約をさらに加えて考えることができる．許容最大値が増加するにつれて，制約のないハフマンおよび Hu-Tucker 符号木に明らかに近づいていく．また，一般的には，最大長制約がない場合に比べて制約がある場合には，平均符号語長が長くなる．

最大符号語長制約下でのハフマン符号化に対する現時点での最もよいアルゴリズム的なアプローチは，Larmore と Hirschberg[113]のアルゴリズムに基づいた Moffat らのもの[110]～[112]である．ただし，同じ問題に対する多くのアプローチ[114]～[121]が初期のころになされている．最大符号語長制約下での Hu-Tukcer アルゴリズムは[122]で取り扱われている．

最大符号語長制約符号化問題に対していくつかの性能解析がなされているが，Capocelli と De Santis[123]は $p(1)$ または $p(K)$ が与えられている場合を含めて，この場合の冗長度の限界を得ている．符号語長制約された Hu-Tukcer 問題に対する同様の問題は，まだ研究されていない．最大長制約された Hu-Tukcer 問題に対する最悪の情報源確率の順序は制約のない場合と同様，のこぎり歯形の順序である[124]．

Anily と Hassin[125]の研究は，最大長制約のような制約に関係している．彼らは平均符号語長に関して最もよい木，2 番目によい木，…，s 番目によい木を計算し，それらの候補の木が制約を満たすかを順次調べることにより，制約を満たす最良の木を見つけている．一般に s 番目によい木の平均符号語長に対するエントロピー限界はまだ求まっておらず，興味深い問題である．

6 他の符号語長汎関数値の最小化

ハフマンと Hu-Tucker の両方の問題は，平均符号語長以外の符号語長の汎関数値を最小化するという一般化の下で考えることができる．特に，最適な木を構成するためにハフマンや Hu-Tucker アルゴリズムのような結合型アルゴリズムが使える場合が興味深い．

Parker[126]は，一般化されたハフマン問題に関するたくさんの研究を共通の包括的な枠組みに統合した．特に Parker は，X と Y が結合される節点の確率重みであるとき，結合された親節点の重みが“準線形(quasilinear)”な結合関数 $F(X, Y)$ で与えられる場合について研究を行っている．オリジナルのハフマンアルゴリズムでは，$F(X, Y)=X+Y$ であるが，$F_m(X, Y)=\max\{X, Y\}+c$ の場合も Parker の枠組みに入っている．なお，Paker の一般化結合アルゴリズムでは，最も小さい重み X と Y を持つ 2 つの節点を結合して重み $F_m(X, Y)$ を持つ親節点を作る．もし，G が木の内部節点上の“Schur の凹な”コスト汎関数の場合は，一般化結合アルゴリズムは G が最小化され

る木を作る．通常のハフマンアルゴリズムの場合 G は総和であり，内部節点の重みの総和を最小にする木は，平均符号語長 $\sum p(i)l(i)$ を最小にする木と完全に一致する．Parker の枠組みに入るコスト汎関数の他の例は $G_m=\max$ {内部節点の重み} の場合である．$F_m(X, Y)$ に対して一般化結合アルゴリズムで作った内部節点の最大重みを最小にする木は，$\max(p(i)+cl(i))$ を最小化する木と完全に一致する．その結果得られる最小符号語長汎関数値に対する一般化されたエントロピー限界は，[126]で示されている．$p(1)$ が与えられている場合に対しては，[127]で取り扱われている．Parker の枠組みに含まれる特殊な場合が，[128], [129]および[126]の文献で取り扱われている．Knuth[130]は，ハフマン問題を第二の最適化基準と統合するエレガントな抽象的な視点から，Parker の問題を再検討した．第二の最適化基準としては，与えられた情報源確率の下で，符号語長の総和を最小にするという Schwartz[131]の符号化問題や，符号語長の分散を最小にする Kou[132]の符号化問題などがある．なお，Schwartz と Kou の符号化問題に関しては文献[133]でも取り扱われている．Markowsky[134]もまた第二の最適化基準を伴うハフマン問題を取り扱っている．Chang と Thomas[135]は，符号木に関する Knuth の代数的な視点をさらに深め，情報源確率が確率変数で与えられている場合を取り扱っている．

一般的な結合関数を用いる場合に対する Hu-Tucker 問題は，Hu, Kleitman, Tamaki[109]により研究され，幾つかのある性質を満たす結合関数とコスト関数が取り扱われている．Parker の一般化ハフマンアルゴリズムが成り立つ F と G のクラスと，Hu-Kleitman-Tamaki の一般化 Hu-Tucker アルゴリズムが成立する F と G のクラスの間にどのような対応関係が存在するかは，すぐには明らかにはならない．これら 2 つの独立に一般化された定式化を一つのものにまとめることは，有益な研究課題である．Hu-Kleitman-Tamaki アルゴリズムで得られる辞書順符号に対する最小符号語長汎関数値に一般化したエントロピー型の限界を導出することも興味深い問題である．一般結合 Hu-Tucker 型のアルゴリズムが，$c=1$ の場合の F_m と "G_s=総和" という特殊な G, F のペアに対して最適な木を見つけることを，Zhang[136]が証明した．なおこの関数のペアは，Hu-Kleitman-Tamaki によって取り扱われた F と G の元のクラスには含まれていないものである．Kirkpatrick と Klawe[137]は，[109]で部分的に取り扱われた，2 値でない場合の辞書順符号に対して $c=1$ の場合の F_m と G_m の問題を詳しく取り扱っている．彼らは，この F と G のペアに対する一般化されたエントロピー型の上界も与えている．

一般化結合アルゴリズムで最適木が見つけられる Parker や Knuth の最適化基準以外の他の最適基準の下で，ハフマン問題が取り扱われている場合もある．そのような問題の一つは，Cover[138]の競合的最適符号(competitively optimal code)である．ハフマン木は，任意の $L'=\{l'(i)\}$ に対して，$\sum p(i)(l(i)$

$-l'(i))\leq 0$ を満たす $L=\{l(i)\}$ を与える．Cover は，$\sum p(i)\,\mathrm{sgn}\,(l(i)-l'(i))\leq 0$ となる木を見つける問題を考察した．Feder[139]と Yamamoto と Itoh[140]も競合的最適符号を取り扱い，Yamamoto-Itoh は，もし与えられた情報源確率に対して競合的最適符号が存在するときには，それはハフマン符号木であることを示し，競合的最適符号が存在するための条件を与えた．非線型関数 f のある一般的なクラスに対して，$\sum p(i)f(l(i)-l'(i))\leq 0$ となる木を見つけるのは興味深い問題となるかもしれない．

Larmore[141]は，符号化された系列の通信時における遅延をモデル化するときに生じる，平均符号語長と符号語長の分散の特殊な非線型関数を最小化する問題に興味を持った．彼は，ハフマン型のこの問題を解いているが，Hu-Tucker 問題は未解決である．その問題へのアプローチは，まず平均符号語長と符号語長の分散のすべての線形な組合せ関数を最小化し，次に問題としている特殊な非線型関数に関して，最初の結果得られた集合内のすべての候補木上で探索するという 2 段階のステップからなっている．そのため，その一般的なアプローチはもっと広い応用が期待できるように思われる．しかし，アルゴリズム自身は，結合型のハフマンアルゴリズムほどエレガントではない．

Cohen と Fredman[142]は，m 個の独立なプロセッサーが単一の検査木を並列に使う場合の性能の最小化問題を取り扱った．各プロセッサーは，それぞれ K 個の項目から独自の識別項目を同定するものとし，特に，m 個の識別項目は確率 P にしたがって独立に選択されるものとする．各プロセッサーの検索は計算の次のフェーズが始まる前に全て完了していなければならないという仮定の下で，木上の m 個のパスの最大値の平均長を最小化する問題である．$m=1$ の場合は，通常のハフマン問題となり，辞書順制約がある場合は Hu-Tucker 問題に一致する．彼らの準最適アルゴリズムでは，$m=1$ のハフマンアルゴリズムが適用できるある変換された確率の集合を導出するようなアプローチを用いている．この問題を最適に解くことは，明らかに興味深い問題である．

7 おわりに

性能評価や木の特徴付けの観点やあるいはアルゴリズム的な観点から，さらに研究する価値のあるハフマン符号化問題の興味深い関連問題が数多く残されている．本論文が従来の研究に対する有益なガイドとなれば幸いである．

[参考文献]

[1] Huffman, D. A., A method for the construction of minimum-redundancy codes, Proc. of the IRE, Vol. 40, 1098-1011, Sep. 1952.
[2] Abrahams, J., Code and parse trees for lossless source encoding, Proc. Sequences 97, Positano, Italy, 1997.
[3] Gallager, R. G., Information Theory and Reliable Communication, Wiley, New York, 1968.

[4] McEliece, R. J., The Theory of Information and Coding, Addison-Wesley, Reading, MA, 1977, second printing with revisions, 1979.

[5] Cover, T. M., and Thomas, J. A., Elements of Information Theory, Wiley, New York, 1991.

[6] Csiszár I., and Körner, J., Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic, New York, 1981.

[7] Aho, A. V., Hopcroft, J. E., and Ullman, J. D., Data Structures and Algorithms, Addison-Wesley, Reading, MA, 1983, reprinted with corrections 1987.

[8] Sedgewick, R., Algorithms, Addison-Wesley, Reading, MA, 1983.

[9] Knuth, D. E., The Art of Computer Programming, Vol. 3, Sorting and Searching, Addison Wesley, Reading, MA, 1973.

[10] Hu, T. C., Combinatorial Algorithms, Addison Wesley, Reading, MA, 1982.

[11] Ahlswede, R., and Wegener, I., Search Problems, Wiley, New York, 1987.

[12] Aigner, M., Combinatorial Search, B. G. Teubner, Stuttgart, Wiley, New York, 1988.

[13] Katona, G. O. H., and Nemetz, T. O. H., Huffman codes and self-information, IEEE Trans. Inform. Th., Vol. IT-22, No. 3(1976), 337-340.

[14] Katona, G. O. H., and Lee, M. A., Some remarks on the construction of optimal codes, Acta Math. Acad. Scientiarum Hungaricae, Vol. 23, No. 3-4(1972), 439-442.

[15] Nemetz, T., On the word-lengths of Huffman codes, Problems of Contr. Inform. Th., Vol. 9, No. 4(1980), 231-242.

[16] Longo, G., and Nemetz, T., Once more on the word-length of optimal source codes, Trans. of the Ninth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes, Prague, 1982, Vol. B. 63-69, D. Reidel, Boston, MA, 1983.

[17] Nemetz, T., and Simon, J., Self-information and optimal codes, Colloquia Mathematica Societatis János Bolyai, 16, Topics in Information Theory, Keszthely, Hungary, 1975, I. Csiszár and P. Elias, eds., 457-468, North Holland, New York, 1977.

[18] Buro, M., On the maximum length of Huffman codes, Inform. Proc. Lett., Vol. 45 (1993), 219-223.

[19] Birmiwal, K., The minimum codeword length and redundancy in the binary Huffman code for uncertain sources, IEEE Trans. Inform. Th., Vol. 36, No. 2(1990), 439-443.

[20] Capocelli, R. M., and De Santis, A., A note on D-ary Huffman codes, IEEE Trans. Inform. Th., Vol. 37, No. 1(1991), 174-179.

[21] Chu, K. C., and Gill, J., Upper bounds on Huffman codeword lengths, Proc. of the IEEE Intl. Symp. on Inform. Th., Budapest, Hungary, 1991, 111.

[22] Schack, R., The length of a typical Huffman codeword, IEEE Trans. Inform. Th., Vol. IT-40, No. 4(1994), 1246-1247.

[23] Cheng, J. F., Dolinar, S., Effros, M., and McEliece, R., Data expansion with Huffman codes, Proc. IEEE Intl. Symp. on Inform. Th., Whistler, BC, 1995, 325.

[24] Montgomery, B. L., and Vijaya Kumar, B. V. K., On the average codeword length of optimal binary codes for extended sources, IEEE Trans. Inform Th., Vol. IT-33, No. 2 (1987), 293-296.

[25] Manstetten, D., Tight bounds on the redundancy of Huffman codes, IEEE, Trans. Inform. Th., Vol. 38, No. 1(1992), 144-151.

[26] Gallager, R. G., Variations on a theme by Huffman, IEEE Trans. Inform. Th., Vol. IT-24, No. 6(1978), 668-674.

[27] Capocelli R. M., and De Santis, A., Tight upper bounds on the redundancy of Huffman codes, IEEE Trans. Inform. Th., Vol. 35, No. 5(1989), 1084-1091.

[28] Capocelli, R. M., and De Santis, A., Variations on a theme by Gallager, Image and Text Compression, Storer, J. A., ed., 181-213, Kluwer, Boston, MA, 1992.

[29] Capocelli, R. M., and De Santis, A., New bounds on the redundancy of Huffman codes, IEEE Trans. Inform. Th., Vol. 37, No. 4(1991), 1095-1104

[30] Johnsen, O., On the redundancy of binary Huffman codes, IEEE Trans. Inform. Th., Vol. IT-26, No. 2(1980), 220-222.

[31] Capocelli, R. M., Giancarlo, R., and Taneja, I. J., Bounds on the redundancy of Huffman codes, IEEE Trans. Inform. Th., Vol. IT-32, No. 6(1986), 854-857.

[32] Montgomery, B. L., and Abrahams, J., On the redundancy of optimal binary prefix-condition codes for finite and infinite sources, IEEE Trans. Inform. Th., Vol. IT-33, No. 1 (1987), 156-160.

[33] De Prisco, R., and De Santis, A., On the redundancy achieved by Huffman codes, Inform. Sci., Vol. 88(1996), 131-148.

[34] Yeung, R. W., Local redundancy and progressive bounds on the redundancy of a Huffman code, IEEE Trans. Inform. Th., Vol. 37, No. 3(1991), 687-691.

[35] Golomb, S. W., Sources which maximize the choice of a Huffman coding tree, Inform. Contr., Vol. 45(1980), 263-272.

[36] Geçkinli, N. C., Two corollaries to the Huffman coding procedures, IEEE Trans. Inform. Th., Vol. IT-21, No. 3(1975), 342-344.

[37] Vinokur, A. B., "Huffman codes and maximizing properties of Fibonacci numbers," Cybernet. Systems Anal., Vol. 28, No. 3(1993), 329-334, translated from Kibernetikai Systemnyi Analiz, Vol. 187, No. 3(1992), 10-15.

[38] Horibe, Y., Balance properties of optimal binary trees, Proc. of the IEEE Intl. Symp. on Inform. Th., Budapest, Hungary, 1991, 110.

[39] Hirschberg, D. S., Larmore, L. L., and Molodowitch, M., Subtree weight ratios for optimal binary search trees, UC Irvine Technical Report, No. 86-02(1986).

[40] Kirrinnis, P., An optimal bound for path weights in Huffman trees, Inform. Proc. Lett., Vol. 51(1994), 107-110.

[41] Jakobsson, M., Huffman coding in bit-vector compression, Inform. Proc. Lett., Vol. 7, No. 6(1978), 304-307.

[42] Stubley, P. R., On the redundancy of optimum fixed-to-variable length codes, Proc. Data Compression Conf., Snowbird, UT, 1994, Storer, J. A., and Cohn, M., eds., 90-97, IEEE Computer Society Press, Los Alamitos, CA, 1994.

[43] Stubley, P. R., The redundancy of optimum codes of multinomially distributed sources, preprint, 1994.

[44] Fenwick, P. M., Huffman code efficiencies for extensions of sources, IEEE Trans. Comm., Vol. 43, NO. 2/3/4(1995), 163-165.

[45] Berger, T., and Zhang, X., On Huffman codes for extension alphabets, Proc. IEEE Inform. Th. Workshop, Israel, 1996.

[46] Kricevskii, R. E., The block length necessary to obtain a given redundancy, Soviet Math. Dokl., Vol. 7, No. 6(1966), 1416-1420, translated from Dokl. Akad. Nauk SSSR, Vol. 171, No. 1(1966).

[47] Hwang, F. K., On finding a single defective in binomial group testing, JASA, Vol. 69, No. 345(1974), 146-150.

[48] Yao, Y. C., and Hwang, F. K., On optimal nested group testing algorithms, J. Stat. Planning and Inference, Vol. 24(1990), 167-175

[49] Hassin, R., A dichotomous search for a geometric random variable, Operations Research, Vol. 32, No. 2(1984), 423-439.

[50] Hinderer, K., and Stieglitz, M., On polychotomous search problems, European J. of Operations Research, Vol. 73(1994), 279-294.

[51] Günther, C. G., and Schneider, W. R., Entropy mean and variance, IEEE Trans. Inform. Th., submitted, and Entropy as a function of alphabet size, Proceedings IEEE Intl. Symp. on Inform. Th., San Antonio, TX, 1993, 70.

[52] Campbell, L. L., Averaging entropy, IEEE Trans. Inform. Th., Vol. 41, No. 1(1995), 338-339.

[53] Montgomery, B. L., Diamond, H., and Vijaya Kumar, B. V. K., Bit probalilities of optimal binary source codes, IEEE Trans. Inform. Th., Vol. 36, No. 6(1990), 1446-1450.

[54] Gutman, M., Fixed-prefix encoding of the integers can be Huffman-optimal, IEEE Trans. Inform. Th., Vol. 36, No. 4(1990), 936-938.

[55] Tucker, A. C., The cost of a class of optimal binary trees, J. Combinatorial Theory (A), Vol. 16(1974), 259-263.

[56] Hwang, F. K., An explicit expression for the cost of a class of Huffman trees, Discrete Mathematics, Vol. 32(1980), 163-165.

[57] Hwang, H. -K., Optimal algorithms for inserting a random element into a random heap, IEEE Trans. Inform. Th., Vol. 43, No. 2(1997), 784-787

[58] Longo, G., and Galasso, G., An application of informational divergence to Huffman codes, IEEE Trans. Inform. Th., Vol. IT-28, No. 1(1982), 36-43.

[59] Fabris, F., The Longo-Galasso criterion for Huffman codes: an extended version, Alta Frequenza, Vol. LVIII, No. 1(1989), 35-38.

[60] Blumer, A. C., Minimax universal noiseless coding for unifilar and Markov sources, IEEE Trans. Inform. Th., Vol. IT-33, No. 6(1987), 925-930.

[61] Pursley, M. B., and Davisson, L. D., Mismatch bounds for variable rate source codes with applications to universal data compression, Proc. AFOSR Workshop in Communica-

tions Theory and Applications, Provincetown, MA, 1978, 33-37.

[62] Nemetz, T., Information type measures and the applications to finite decision problems, Carlton Math. Lecture Notes. No. 17, Carleton Univ., Ottawa 1977.

[63] Hu, T. C., and Tucker, A. C., Optimum binary search trees, Proc. Second Chapel Hill Conf. on Combinatorial Math. and Its Appl., Chapel Hill, NC, 1970, 285-305.

[64] Hwang, F. K., Generalized Huffman trees, SIAM J. Appl. Math., Vol. 37, No. 1(1979), 124-127.

[65] Chu, K. -C., and Gill, J., Mixed-radix Huffman codes, Proc. 11 th Intl. Conf. of the Chilean Computer Sciences Society, Santiago, Chile, 1991, Baeza-Yates, R., and Manber, U., eds., Plenum, New York, 1992, 209-218.

[66] Gallager, R. G., and Van Voorhis, D. C., Optimal source codes for geometrically distributed alphabets, IEEE Trans. Inform. Th., Vol. IT-21, No. 2, 228-230.

[67] Golomb, S. W., Run-length encodings, IEEE Trans. Inform. Th., Vol. IT-12(1966), 399-401.

[68] Humblet, P., Optimal source coding for a class of integer alphabets, IEEE Trans. Inform. Th., Vol. IT-24, No. 1(1978), 110-112.

[69] Abrahams, J., Huffman-type codes for infinite source distributions, J. Franklin Institute, Vol. 331B, No. 3(1994), 265-271.

[70] Kato, A., Han, T. S., and Nagaoka, H., Huffman coding with an infinite alphabet, IEEE Trans. Inform. Th., Vol. 42, No. 3(1996), 977-984.

[71] Linder, T., Tarokh, V., and Zeger, K., Existence of optimal prefix codes for infinite source alphabets, IEEE Trans. Inform. Th., submitted.

[72] Kato, A., Han, T. S., and Nagaoka, H., Huffman coding with an infinite alphabet, IEEE Trans. Inform. Th. Vol. 42, No. 3(1996), 977-983.

[73] Elias, P., Universal codeword sets and representations of the integers, IEEE Trans. Inform. Th., Vol. IT-21, No. 2(1975), 194-203.

[74] Stout, Q. F., Searching and encoding for infinite ordered sets, Intl. J. of Computer and Inform. Sci., Vol. 11, No. 1(1982), 55-72.

[75] Knuth, D. E., Supernatural numbers, The Mathematical Gardner, Klarner, D. A., ed., Wadsworth, Belmont, CA, 1981, 310-325.

[76] Ahlswede, R., Han, T. S., and Kobayashi, K., Universal coding of integers and unbounded search trees, IEEE Trans. Inform. Th., Vol. 43, No. 2(1997), 669-682.

[77] Capocelli, R. M., and De Santis, A., Regular universal codeword sets, IEEE Trans. Inform. Th., Vol. IT-32, No. 1(1986), 129-133.

[78] Apostolico, A., and Fraenkel, A. S., Robert transmission of unbounded strings using Fibonacci representations, IEEE Trans. Inform. Th., Vol. IT-33, No. 2(1987), 238-245.

[79] Capocelli, R. M., Comments and additions to 'Robust transmission of unbounded strings using Fibonacci representations, IEEE Trans. Inform. Th., Vol. 35, No. 1(1989), 191-193.

[80] Lakshmanan, K. B., On univeral codeword sets, IEEE Trans, Inform. Th., Vol. IT-27, No. 5(1981), 659-662.

[81] Wang, M., Almost asymptotically optimal flag encoding of the integers, IEEE Trans. Inform. Th., Vol. 34, No. 2(1988), 324-326.

[82] Stout, Q., Improved prefix encodings of the natural numbers, IEEE Trans. Inform. Th., Vol. IT-26, No. 5(1980), 607-609.

[83] Yamamoto, H., and Ochi, H., A new asymptotically optimal code for the positive integers, IEEE Trans. Inform. Th., Vol. 37, No. 5(1991), 1421-1429.

[84] Reingold, E. M., and Shen, X., More nearly optimal algorithms for unbounded searching, part I: the finite case [and] part II: the transfinite case, SIAM J, Comput., Vol. 20, No. 1(1991), 156-208.

[85] Bentley, J. L., and Yao, A. C. -C., An almost optimal algorithm for unbounded searching, Inform. Proc. Lett., Vol. 5, No. 3(1976), 82-87.

[86] Witten, I. H., Moffat, A., and Bell, T. C., Managing Gigabytes, Van Nostrand Reinhold, NY, 1994.

[87] Fraenkel, A. S., Systems of numeration, Am. Math. Monthly, Vol. 92, No. 2(1985), 105-114.

[88] Moffat, A., and Zobel, J., Parameterised compression for sparse bitmaps, Proc. of 15th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval, Copenhagen, Denmark, 1992, SIGIR Forum, Belkin, N., Inguersen, P., and Pejtersen, A. M., eds., 274-285.

[89] Linoff G., and Stanfill, C., Compression of indexes with full positional information in very large text databases, Proc. of 16th Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval, Pittsburgh, PA, 1993, SIGIR Forum, 88-95.

[90] Fraenkel, A. S., and Klein, S. T., Novel compression of sparse bit-strings, preliminary report, Combinatorial Algorithms on Words, Nato ASI Series, Vol. F12, Apostolico A., and Galil, Z., eds., Springer Verlag, New York, 1985, 170-183.

[91] Amemiya, T., and Yamamoto, H., A new class of the universal representation for the positive integers, IEICE Trans. Fundamentals, Vol. E76-A, No. 3(1993), 447-452.

[92] Hu, T. C., and Tucker, A. C., Optimal computer search trees and variable-length alphabetical codes, SIAM J. Appl. Math., Vol. 21, No. 4(1971), 514-523.

[93] Garsia, A. M., and Wachs, M. L., A new algorithm for minimum cost binary trees, SIAM J. Comput., Vol. 6, No. 4(1977), 622-642.

[94] Klawe, M., and Mumey, B., Upper and lower bounds on constructing alphabetic binary trees, Proc. of the Fourth Annual ACM-SIAM Symp. on Discrete Algorithms, Austin, TX, 1993, 185-193.

[95] Hu, T. C., and Morgenthaler, J. D., Optimum alphabetic binary trees, Combinatorics and Computer Science, CCC'95, 8th Franco-Japanese and 4th Franco-Chinese Conf., Brest, France, 1995, Lecture Notes in Computer Science, Vol. 1120, Deza, M., Euler, R., and Manoussakis, I., eds., Springer Verlag, New York, 1996, 234-243.

[96] Pradhan, J., and Sastry, C. V., A new version of Hu-Tucker algorithm, Proc. of the 29th Annual Convention of the Computer Society of India, CSI-94, Calcutta, India, 1994, 27-31.

[97] Gilbert, E. N., and Moore, E. F., Variable-length binary encodings, Bell System Technical Journal, Vol. 38(1959), 933-967.

[98] Yeung, R. W., Alphabetic codes revisited, IEEE Trans. Inform. Th., Vol. 37, No. 3 (1991), 564-572.

[99] Nakatsu, N., Bounds on the redundancy of binary alphabetical codes, IEEE Trans. Inform. Th., Vol. 37, No. 4(1991), 1225-1229.

[100] Sheinwald, D., On binary alphabetical codes, Proc. Data Compression Conf., Snowbird, UT, 1992, 112-121, IEEE Computer Society Press, Los Alamitos, CA, 1992.

[101] De Prisco, R., and De Santis, A., On binary search trees, Inform. Proc. Lett., Vol. 45, No. 5(1993), 249-253.

[102] Kleitman, D. J., and Saks, M. E., Set orderings requiring costliest alphabetic binary trees, SIAM J. Alg. Disc. Meth., Vol. 2, No. 2(1981), 142-146.

[103] Hu, T. C., and Tan, K. C., Least upperbound on the cost of optimum binary search trees, Acta Informatica, Vol. 1(1972), 307-310.

[104] Ramanan, P., Testing the optimality of alphabetic trees, Theoretical Computer Science, Vol. 93(1992), 279-301.

[105] Van Voorhis, D. C., Constructing codes with ordered codeword lengths, IEEE Trans. Inform. Th., Vol. 21, No. 1(1975), 105-106.

[106] Abrahams, J., Codes with monotonic codeword lengths, Inform. Proc. and Management, Vol. 30, No. 6(1994), 759-764.

[107] Lipman, M. J., and Abrahams, J., Minimum average cost testing for partially ordered components, IEEE Trans. Inform. Th., Vol. 41, No. 1(1995), 287-291.

[108] Abrahams, J., Parallelized Huffman and Hu-Tucker searching, IEEE Trans. Inform. Th., Vol. 40, No. 2(1994), 508-510.

[109] Hu, T. C., Kleitman, D. J., and Tamaki, J. K., Binary trees optimum under various criteria, SIAM J. Appl. Math., Vol. 37, No. 2(1979), 246-256.

[110] Moffat, A., Turpin, A., and Katajainen, J., Space-efficient construction of optimal prefix codes, Proc. Data Compression Conf., Snowbird, UT, 1995, Storer, J. A., and Cohn, M., eds., IEEE Computer Society Press, Los Alamitos, CA, 1995, 192-201.

[111] Katajainen, J., Moffat, A., and Turpin, A., A fast and space-economical algorithm for length-limited coding, Proc. Intl. Symp. on Algorithms and Computation, Cairns, Australia, 1995, Lecture Notes in Computer Science, Vol. 1004, Staples, J., Eades, P., Katoh, N., and Moffat, A., eds., Springer Verlag, New York, 1995, 12-21.

[112] Turpin, A., and Moffat, A., Practical length-limited coding for large alphabets, Computer J., Vol. 38, No. 5(1995), 339-347.

[113] Larmore, L. L., and Hirschberg, D. S., A fast algorithm for optimal length-limited Huffman codes, JACM, Vol. 37, No. 3(1990), 464-473.

[114] Garey, M. R., Optimal binary search trees with restricted maximal depth, SIAM J.

Comput., Vol. 3, No. 2(1974), 101-110.
[115] Gilbert, E. N., Codes based on inaccurate source probabilities, IEEE Trans. Inform. Th., Vol. IT-17, No. 3(1971), 304-314.
[116] Hu, T. C., and Tan, K. C., Path length of binary search trees, SIAM J. Appl. Math., Vol. 22, No. 2(1972), 225-234.
[117] Van Voorhis, D. C., Constructing codes with bounded codeword lengths, IEEE Trans. Inform. Th., Vol. IT-20, No. 2(1974), 288-290.
[118] Murakami, H., Matsumoto, S., and Yamamoto, H., Algomithm for construction of variable length code with limited maximum word length, IEEE Trans. Comm., Vol. COM-32, No. 10(1984), 1157-1159.
[119] De Lameillieure, J. L. P., A heuristic algorithm for the construction of a code with limited word length, IEEE Trans. Inform. Th., Vol. IT-33, No. 3(1987), 438-443; Correction, IEEE Trans. Inform. Th., Vol. 34, No. 4(1988), 893-894.
[120] Pradhan, J., and Sastry, C. V., On optimal weighted binary trees, Intl. J. of High Speed Computing, Vol. 7, No. 3(1995), 445-464.
[121] Itai, A., Optimal alphabetic trees, SIAM J. Comput., Vol. 5, No. 1(1976), 9-18.
[122] Larmore, L. L., and Przytycka, T. M., A fast algorithm for optimum height-limited alphabetic binary trees, SIAM J. Comput., Vol. 23, No. 6(1994), 1283-1312.
[123] Capocelli, R. M., and De Santis, A., On the redundancy of optimal codes with limited word length, IEEE Trans. Inform. Th., Vol. 38, No. 2(1992), 439-445.
[124] Chu, Y., An extended result of Kleitman and Saks concerning binary trees, Discrete Appl. Math., Vol. 10(1985), 255-259.
[125] Anily, S., and Hassin, R., Ranking the best binary trees, SIAM J. Comput, Vol. 18, No. 5(1989), 882-892.
[126] Parker, Jr., D. S., Conditions for optimality of the Huffman algorithm, SIAM J. Comput., Vol. 9, No. 3(1980), 470-489
[127] Taneja, I. J., A short note on the redundancy of degree α, Inform. Sci., Vol. 39(1986), 211-216.
[128] Humblet, P. A., Generalization of Huffman coding to minimize the probability of buffer overflow, IEEE Trans. Inform. Th., Vol. IT-27, No. 2(1981), 230-232.
[129] Blumer, A. C., and McEliece, R. J., The Rényi redundancy of generalized Huffman codes, IEEE Trans. Inform. Th., Vol. 34, No. 5(1988), 1242-1249.
[130] Knuth, D. E., Huffman's algorithm via algebra, J. of Combinatorial Theory, Series A, Vol. 32(1982), 216-224.
[131] Schwartz, E. S., An optimum encoding with minimum longest code and total number of digits, Inform. Contr., Vol. 7(1964), 37-44.
[132] Kou, L. T., Minimum variance Huffman codes, SIAM J. Comput., Vol. 11, No. 1 (1982), 138-148.
[133] Horibe, Y., Remarks on 'compact' Huffman trees. J. of Combinatorics. Information and System Sciences, Vol. 9, No. 2(1984), 117-120.
[134] Markowsky, G., Best Huffman trees, Acta Informatica, Vol. 16(1981), 363-370.
[135] Chang, C. -S., and Thomas, J. A., Huffman algebras for independent random variables, Discrete Event Dynamic Systems: Theory and Applications, Vol. 4, No. 1(1994), 23-40, and Proc. IEEE Intl. Symp. on Information Theory, San Antonio, TX, 1993, 215.
[136] Zhang, C. -Q., Optimal alphabetic binary tree for a nonregular cost function, Discrete Appl. Math., Vol. 8(1984), 307-312.
[137] Kirkpatrick, D. G., and Klawe, M. M., Alphabetic minimax trees, SIAM J. Comput., Vol. 14, No. 3(1985), 514-526.
[138] Cover, T. M., On the competitive optimality of Huffman codes, IEEE Trans. Inform Th., Vol. 37, No. 1(1991), 172-174.
[139] Feder, M., A note on the competitive optimality of the Huffman code, IEEE Trans. Inform. Th., Vol. 38, No. 2(1992), 436-439.
[140] Yamamoto, H., and Itoh, T., Competitive optimality of source codes, IEEE Trans. Inform. Th., Vol. 41, No. 6(1995), 2015-2019.
[141] Larmore, L. L., Minimum delay codes, SIAM J. Comput., Vol. 18, No. 1(1989), 82-94.
[142] Cohen, D., and Fredman, M. L., Weighted binary trees for concurrent searching, J. of Algorithms, Vol. 20(1996), 87-112.

[**Abstract**]

Huffman coding is the most well known code tree problem, but there are a number of interesting variations of the problem which are less well known, and there are many results available concerning the properties of Huffman codes. This paper addresses the Huffman code tree problem and characteristics of the Huffman code tree along with several problem variations, specifically, when the source alphabet is infinite, when a lexicographic constrain is imposed on the codewords, or, that is, the Hu-Tucker code tree problem, when a maximum length constraint is imposed on the codewords, and when other tree characteristics besides average codeword length are minimized.