6C−5

ビットストリングによる2分木の表現

清水 道夫
長野県短期大学

1 はじめに

近年，組合せ論的な木の符号化問題，とくに2分木の符号化問題がいくつか研究されている．2分木の符号化問題は数学的に興味深いものであるが，応用面としては，木に関するアルゴリズムのふるまいを考えるときのランダムデータとしての使用を志向している．Ｚａｋｓ[1]は，2分木の2進数列への符号化を考察した．この2進数列は，2分木の内部節に1，葉に0とラベリングして，木を先行順（root-left-right）にたどったときのラベルをならべたものに対応している．このような数列をX-seqとよび，辞書式順序のもとでのジェネレーティング，ランキング，アンランキングを考えている．しかし，X-seqをランダムデータとして使おうとするとこのままでは非常に使いにくい．これは，木の構造についての情報が不足しているためと考えられる．そこで，2分木を虫が這うようにたどったときの2進数列を考える．虫は根から出発して左回りで根に戻って来るが，根から葉に向かう枝（ｕｐ）を1，その逆（ｄｏｗｎ）を0とする．これに対応する2進数列をＹ-seqとよぶ．Ｙ-seqは符号という意味では多少冗長であるが，木の構造を反映しており，いろいろな応用に適している．

2 定 義

2つの2分木間の辞書式順序（lexicographic order）をつぎのように再帰的に定義する．T_l，T_r はそれぞれ左部分木，右部分木を表す．

［定義1］[1] 2分木ＴとＴ’にたいして，次の条件を満たすとき，Ｔ＜Ｔ’であるという．

(1) Ｔが空（empty）でＴ’が空でない．または

(2) Ｔが空でないとき

(a) $T_l < T'_l$，または

(b) $T_l = T'_l$ かつ $T_r < T'_r$

Expressions for Binary Trees with Bit Strings
Michio SHIMIZU
Nagano Prefectural College

ところで，ｎ個の内部節を持つ2分木はカタラン数で表されるが，その集合内での上記辞書式順序をランクという．つぎに，2分木を表現する2種類の2進数列（bit string）を定義する．

［定義2］[1] X-seq ｘは次の条件を満足する．

(1) ｘにはｎ個の1とｎ＋1個の0を含む．

(2) ｘを左から走査して，任意のケタまでの0の個数は1の個数よりも多くない．

［定義3］Ｙ-seq ｙを次のように構成する．

$n=0$：$y=\phi$（空を表す）

$n=1$：ｙ＝”１０１０”

$n>1$：ｙ＝”１”＋α＋”０１”＋β＋”０”

ここに，αとβは｜α＋β｜＝４（ｎ－１）をみたすＹ-seqで，＋は文字列のつながりを示す．

（例）ｎ＝４にたいするランクＲ，Ｘ-seq，Ｙ-seq．

R	X-seq	Y-seq
1	101010100	1011011011010000
2	101011000	1011011101001000
3	101100100	1011101001101000
4	101101000	1011101101000100
5	101110000	1011110100100100
6	110010100	1101001101101000
7	110011000	1101001110100100
8	110100100	1101101000110100
9	110101000	1101101101000010
10	110110000	1101110100100010
11	111000100	1110100100110100
12	111001000	1110100110100010
13	111010000	1110110100010010
14	111100000	1111010010010010

3 系列の性質

X-seqが2分木に1対1に対応していることは，文献(1)で証明されている．また，Ｙ-seqが2分木に1対1に対応していることは，その構成法からも明かであるが，これを組合せ論における「括弧の問題」に置き換えるとより分かりやすい．X-seqとＹ-seqの長さは，

それぞれ２ｎ＋１，４ｎである．

２分木の左枝にたいするＹ-seqの１が，Ｘ-seqの１に相当することに注意すると，Ｘ-seqとＹ-seqが辞書式順序の意味で一致していることがわかる．とくに，２分木ＴのＸ-seqをｘ，Ｙ-seqをｙとすると，ｘとｙの左端から連続する１の個数は等しい．

ところで，２分木のランクを考えるとき，同じ特徴を持つ木ごとに分類して，その個数をあらかじめ調べておく必要がある．その特徴とは，Ｘ-seqでは左端路上の内部節の数であり，Ｙ-seqでは左端路上の枝の数である．この内部節の数と枝の数が等しいから，Ｚａｋｓの結果がそのままあてはまる．それによると，内部節の数がｎで，連続する１の個数（左端路上の内部節の数）がｊ＋１以上である数ａ（n,j）は，

$$a(n,j) = a(n,j+1) + a(n-1,j-1)$$
$$= \frac{j+2}{2n-j}\binom{2n-j}{n-j-1}$$

で表される．このｎについての再帰的な関係が，次節のアルゴリズムのポイントになっている．

4　ランキングアルゴリズム

内部節の数ｎとＹ-seqが与えられたとき，ランクＲを求めるランキングアルゴリズムを示す．逆に，ｎとＲが与えられたとき，対応するＹ-seqを求めることをアンランキングというが，ここでは省略する．

アルゴリズム ＲＡＮＫ

入力 ｎ，ｙ；出力 Ｒ

ステップ１：ｉ＝ｎ．

ステップ２：ｙを $y=\gamma+$"101"$+\delta$ のように３つの部分列に分ける．ここに，γ はすべて１からなる部分列で，その長さをｓとする．x(i)＝ｓ＋１．

ステップ３：部分列 δ をさらに $\delta=\alpha+$"0"$+\beta$ のように３つの部分に分ける．ここに，δ の左端のケタが０ならば α＝" " とし，そうでなければ，α は（$|\alpha| \bmod 4=0$）かつ（α の１の個数と０の個数が等しい）を満たす部分列とする．

ステップ４：$y=\gamma+\alpha+\beta$．

ステップ５：ｉ＝３ならばステップ６へ，そうでなければｉ＝ｉ－１としてステップ２へ．

ステップ６：ｙ＝"11010010" ならば Ｚ＝１とし，ｙ＝"10110100" ならば Ｚ＝２ とする．

ステップ７：ｉ＝３．

ステップ８：ｊ＝x(i)，Ｚ＝Ｚ＋ａ（i,j）

ステップ９：ｉ＝ｎならばステップ１０へ，そうでなければｉ＝ｉ＋１としてステップ８へ．

ステップ１０：Ｒ＝ａ（n,0）－Ｚ＋１

（例）ｎ＝８，ｙ＝"11011101110100100010001110100100" のとき，ランクＲを求める．

まず，ｎ＝８にたいして，ｊ＋１＝３以上のものは対象外であるから，ａ（8,2）＝５７２を除く．Ｙ-seq から"101"と"0"を除くと，ｎ＝７にたいしてｙ＝"11101110100100010011101001 00"となる．ここで，ｊ＋１＝４以上のものは対象外であるから，ａ（7,3）＝７５を除く．同様にして対象外のものを求めると，

Ｚ＝ａ（8,2）＋ａ（7,3）＋ａ（6,4）＋ａ（5,3）＋ａ（4,2）＋ａ（3,1）＋１

＝５７２＋７５＋６＋５＋４＋３＋１＝６６６

となる．最後の１は，ｙが最終的にｙ＝"11010010"となるからである．よってランクは

Ｒ＝１４３０－６６６＋１＝７６５

となる．参考までに，ｙの変化を書いておく．丸で囲まれたところが削除される部分である．

ｎ＝８：ｙ＝"11011101110100100010001110100100"

ｎ＝７：ｙ＝"1110111010010001001110100100"

ｎ＝６：ｙ＝"111101001001001110100100"

ｎ＝５：ｙ＝"11101001001110100100"

ｎ＝４：ｙ＝"1101001110100100"

ｎ＝３：ｙ＝"101110100100"

ｎ＝２：ｙ＝"11010010"

5　おわりに

Ｙ-seqに関する符号化の問題で残されていることはジェネレーティングである．いまのところ効率のよい方法がみつかっていないが，部分木のローテーション（rotation）による方法が可能と考えられる．これは分類（sort）や探索（search）における２分木の更新（update）を表現しており，ランダムデータとしてＹ-seqを用いる意味からも興味深い．

文　献

(1) Zaks S.:Lexicographic Generation of Ordered Trees, Theoretical Computer Science, Vol.10, pp.63-82(1980).