5Ｂ－4

ローテーションに基づく２分木間の距離

清水　道夫
（長野県短期大学）

1．はじめに

木の近さや類似性に関する問題は，パターン認識や誤り訂正構文解析などに関連して検討されてきた．この近さや類似性を表す尺度が距離であり，一般には重み付きレーベンシュタイン距離で定義されている（文献1）．本報告では，木のトポロジ（位相）を意識した距離として，ローテーション（回転）に基づく２分木の距離を考える．ローテーションは，木をデータ構造としてとらえたとき，見出し（キー）の挿入や削除にともなう更新のための基本的技法である．同じ大きさの２分木ＡとＢがあるとき，ＡからＢへ変換するローテーションの最小数を距離とする．とくに，木Ａと木Ｂの距離が１のとき，この２つの木は結合しているという．まず，２分木をコードワードとよばれる整数列で表現し，その辞書式順序に基づいて２分木のランク（番号）付けを行う．つぎに，大きさｎの２分木にたいして，ランクを要素とする結合表を作成する．２分木間の距離はこの結合表をもとにしたグラフから，単純な最短経路問題として求めることができる．

2．２分木の符号化

２分木をコンピュータ上で扱うために，それを整数列として符号化する．ここでは，D. Zerling（文献2）によって提案されたコードワード（ＣＷ）を生成し，それを２分木に対応づける．まず，２分木のローテーションを図１に示しておく．ここに，部分木α，β，γは空の場合もある．（ａ）から（ｂ）への変換を右ローテーション，（ｂ）から（ａ）への変換を左ローテーションとよぶ．右ローテーションでは，節Ａの右部分木βが節Ｂの左部分木になる．

ＣＷの生成はつぎのように行う．長さＬのＣＷをX_1，X_2，…，X_Lとおく．X_1からX_b（ｂ＝１，２，…，Ｌ）までの総和がｂ以下であるような整数列をすべて生成する．Ｌ＝１から３のとき

Ｌ＝１：０，１
Ｌ＝２：００，０１，０２，１０，１１
Ｌ＝３：０００，００１，００２，００３，０１０，０１１，０１２，０２０，０２１，１００，１０１，１０２，１１０，１１１

となる．ただし，要素間のカンマは省略している．なお，これは辞書式順序（lexicographic order）に並んでおり，この順番をＣＷの長さＬにたいするランクとよぶ．たとえば，０１０はランク５である．

ところで，２分木がＣＷにどのように対応しているかを示す．２分木の大きさは内部節の数ｎで表す．大きさｎの２分木の総数はカタラン数（Catalan number）${}_{2n}C_n/(n+1)$で表されるが，これは長さＬ＝ｎ－１のＣＷの総数に一致する．ＣＷの各要素は２分木の左端路上の節からみた右部分木に対応している．X_Lは根からみた右部分木の大きさで，左ローテーションによって右部分木がなくなるまでの回数を表している．つぎに，X_{L-1}は根における左ローテーションが終了した時点での，根の左子からみた右部分木の大きさである．以下同様であるが，ＣＷは左端路だけからなる線形の２分木（０…０）に変換されるまでの各レベルにおける左ローテーションの回数を表している．ｎ＝３の２分木とそのＣＷを図２に示す．

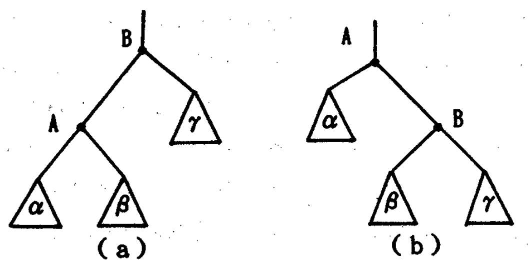

図１　ローテーション

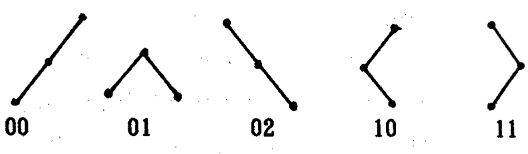

図２　ｎ＝３の２分木とＣＷ

A Distance between Two Binary Trees
based on the Rotations
Michio SHIMIZU
Nagano Prefectural College

3．結合表の作成

ここでは，表１のような結合表の作り方を説明する．ここに，ＣＷおよびそのランクＲは前節で生成したものである．つぎの欄のＢはブロック番号で，ＣＷの末尾のケタ以外が同じものの集合をブロックと呼ぶ．ブロック番号もＣＷの辞書式順序にしたがってつけられる．そのつぎのｕｖは，ＣＷ（２分木）を左，右部分木の大きさに従って分類したもので，左，右部分木の節の数をそれぞれｕ，ｖとすると，ｎ＝４のとき（ｕ，ｖ）＝（３，０），（２，１），（１，２），（０，３）の４通りに分けられる．ただし，表１のｕｖはカッコとカンマを省略している．ところで，ＣＷから左，右部分木の大きさを決めるにはつぎのようにする．ＣＷを右から走査して，Ｊ＋１ケタ目が０，かつ，１からＪ＋１ケタまでの数の総和がＪのとき，（ｕ，ｖ）＝（ｎ-１-Ｊ，Ｊ）とする．

表１のＲＡＮＫは結合する２分木のランクを示したもので，この求めかたを説明する．結合はブロック内同士とブロック外同士に分けられるが，双方の和はつねにｎ−１になる．ブロック内のＣＷの違いは末尾のケタだけで，これは根についてのローテーションによる変化を示しているから，隣合うもの同士が結合する（ただし，ブロックの両端では片側のみ）．たとえば，Ｒ＝２のブロック内結合はＲ＝１とＲ＝３である．一方，ブロック外結合は，表１のＢとｕｖさらにｎ＝３の結合表から決定する．ｎ＝３の結合表は省略するが，結合するランクだけを示すと（１：２，４），（２：１，３），（３：２，５），（４：１，５），（５：３，４）である．ここに，（１：２，４）はランク１がランク２とランク４に結合していることを示す．このｎ＝３の結合ランクに相当する表１のブロック番号のうち，ｕｖの等しいもの同士が結合する．たとえば，ランクＲ＝２はＢ＝１だからＢ＝２かＢ＝４を調べればよいが，Ｂ＝２のなかに相当するものはなく，Ｂ＝４のＲ＝１１が結合する．ところで，ｎ＝３の結合表は図２からも作れるが，ｎ＝２の結合表より再帰的に作成する．

表１　ｎ＝４の結合表

R	CW	B	UV	RANK
1	000	1	30	2, 5, 10
2	001	1	21	1, 3, 11
3	002	1	12	2, 4, 6
4	003	1	03	3, 7, 12
5	010	2	30	1, 6, 8
6	011	2	12	3, 5, 7
7	012	2	03	4, 6, 9
8	020	3	30	5, 9, 13
9	021	3	03	7, 8, 14
10	100	4	30	1, 11, 13
11	101	4	21	2, 10, 12
12	102	4	03	4, 11, 14
13	110	5	30	8, 10, 14
14	111	5	03	9, 12, 13

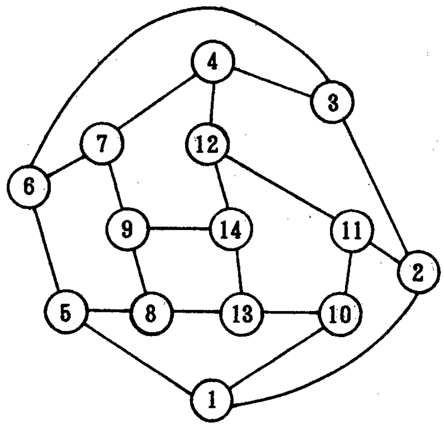

図３　ｎ＝４の結合グラフ

4．最短経路問題

表１からｎ＝４の結合グラフをかくと図３のようになる．節内の数はランクで，節から節に至る枝の数は２分木間の距離を表している．これは，単純な最短経路問題として求められる．たとえば，ランクＲ＝７とランクＲ＝１０の２分木間の距離は４である．まず，節７を０とし，それに結合する節４，６，９を１とする．以下，結合する節のラベル付けされたもののなかで，最小値＋１をその節のラベルとする．

文　献

(1) 田中栄一：構造をもつものの距離と類似度，情報処理，Vol.31, No.9, PP.1270-1279(1990).
(2) D.Zerling：Generating Binary Trees Using Rotations, J.ACM, Vol.32, No.3, pp.694-701 (July 1985).