アルゴリズムとデータ構造，第5回 探索(1)の補足

有村 博紀, 北大情報
2017.4.25: created; 2019.5.07: modified

授業のスライドで示した定理を、きちんと証明してみよう。以下は，下記の参考書の証明の細部を補足して、書き下したものである。

- 1の二分探索木への挿入の平均コストでは、式(4)の導出が鍵である。
- 2のAVL木の高さでは、同じ定理でも、いろいろな証明の仕方があることに気がつくだろう。

参考文献（教科書）：「アルゴリズムとデータ構造」，茨城俊秀 著, 昭晃堂(第3章 順序つき集合の処理と整列, pp.66–71)
「アルゴリズム・イントロダクション」（コルメン・ライザーソン・リベスト・スタイン）等ほとんどのデータ構造とアルゴリズムの教科書でも良い．

---

1. 二分探索木への挿入の平均コスト

ランダム入力に対して，n 要素を二分探索木に挿入する際の平均比較数

任意の$n \geq 0$に対して，$C(n)$ で，ランダム入力に対して，n 要素を二分探索木に挿入する際の平均比較数の総和を表す．

定理1. $C(n) = 2(n+1)\log_e n - 4 + \frac{4}{n+1} = O(n \log n)$.

証明のスケッチ：そこで，以下，n 要素の $n!$ 個の順列が等確率でランダムに起こるとして，INSERTによる構成手間を評価してみよう．n 個の要素が $x_1, \dots, x_n$ の順に入った時，まず x_1 は根に位置し，そのあと，つづく要素 $x_2, \dots, x_n$ がそれぞれ x_1 と比較の後に個に進むから， x_1 が関与する比較の回数は $n-1$ 回である．また，仮定によって，x_1 は n 要素の中から等確率で選ばれる．それが i 番目の大きさならば，構成された二分探索木の左と右の部分木は，それぞれ， $i-1$ 個と $n-i$ 個の節点をもつ．よって，この観察から，次の再帰式をもつ．

$$C(0) = 0$$

$$C(n) = \frac{1}{n}\sum_{i=1}^{n}((n-1) + C(i-1) + C(n-i)) \quad (1)$$

なお，$C(0)$ の値は 0 とおいたが，定理1を証明する目的のためには、どんな定数でも良い．値 n に対して，上の式を整理すると，両辺に n をかけて，

$$nC(n) = n(n-1) + 2\sum_{i=0}^{n-1} C(i) \quad (2)$$

これを $n := n-1$ とおきかえて，一つ前の式を作ると，

$$(n-1)C(n-1) = (n-1)(n-2) + 2\sum_{i=0}^{n-2} C(i) \quad (3)$$

(2)と(3)の差をとって，$C(n–1)$ の項目を左辺へ移行し，右辺を連分数の形に整理すると，

$$\frac{C(n)}{n+1} - \frac{C(n-1)}{n} = \frac{4}{n+1} - \frac{2}{n} \quad (4)$$

この(4)式を，n の値を n から 1 までずらしながら両辺足して解くと（いわゆる望遠鏡論法），最終的に次の式を得る．

$$\frac{C(n)}{n+1} = 4\sum_{i=1}^{n}\frac{1}{i} - 2\sum_{i=1}^{n}\frac{1}{i} - 4 + \frac{4}{n+1}$$

$$= 2\log_2 n - 4 + \frac{4}{n+1}.$$

ここに，調和級数 $\sum_{i=1}^{n} ≒ \log_e n$ で近似される（上と下から挟まれる）ことを使った．（証明のスケッチの終わり）

上の定理から，n個の要素を挿入した際の平均比較数が$O(n \log n)$ 回なので，要素1個あたりの平均比較数（ならし平均比較数）は $O(\log n)$ 回になる．

$O(\log n)n$
したがって，ランダムな入力に対する，n 要素の二分探索木の葉の平均深さは であることがわかる．さらに，ランダムな入力を仮定したとき， 回のmemberと, insert, delete演算は$O(\log n)$平均時間で実行できることがわかる．

2. AVL木の高さ

まえおき）授業スライドに，平衡２分木の例として，AVL木の場合に，節点数nに対して，木の深さが$O(\log n)$になるという定理について，証明の概略を紹介している．一般に，定理の証明は一通りではない．同じ定理でも，その証明の方法は人によっていろいろあって良い．

ここでは，上記の定理の証明をふた通りに示す．一つは漸化式と級数を組み合わせた解析的（数式を使った）証明である．もう一つは，AVL木の定義にもとづいて，帰納法を用いた証明である．どちらもただし証明であり，どちらが良いということはない．自分の方法を工夫してやってみよう．

AVL木の定義

定義1： 2分探索木 T が**AVL木**であるとは、T のどの節点においても、その左部分木と右部分木の高さの差が 1 以下であることをいう。

これから

定理2： 任意の $n \geq 1$ に対して、n 節点のAVL木の高さは $O(\log n)$ である。

定義とは，新しいモノや言葉を定める文章である（数式でも良い），**定理**とは，定義とすでに示された定理（や補題，命題，成立することがわかった性質）を用いて，「正しい推論」方法のもとで示された性質をいう．証明とは，定理が正しいことを示す説明である．興味がある人は，コンピュータサイエンスや，離散数学の教科書を参照されたい．

証明1（解析的な証明）（スライド）

[定理2の証明]

f(h)を高さhのAVL木の最小節点数とすれば、以下の漸化式が成り立つ。

$$f(h) = f(h-1) + f(h-2) + 1, f(0) = 1, f(1) = 2$$

$F(h) = f(h) + 1$とすれば

$$F(h) = F(h-1) + F(h-2), F(0) = 2, F(1) = 3$$

F(h)はフィボナッチ数列だから

$$F(h) = (\varphi_1^{h+3} - \varphi_2^{h+3})/\sqrt{5}$$

ただし、 $\varphi_1 = (1+\sqrt{5})/2, \varphi_2 = (1-\sqrt{5})/2$ である.

高さhのAVL木の節点数をnとすれば、

$$F(h) \leqq \quad n$$

よって、

$$\rightarrow \varphi_1^{h+3} - \varphi_2^{h+3} \quad \leqq \quad \sqrt{5}(n+1)$$

$|\varphi_2| < 1$ だから

$$\rightarrow \varphi_1^{h+3} - 1 \quad \leqq \sqrt{5}(n+1)$$

両辺の対数をとって整理すると，次を得る.

$$\rightarrow h \leq \frac{(\log(\sqrt{5}(n+1)+1))}{\log \varphi_1 - 3}$$

$h = O(\log n)$
したがって が示された。（証明終わり）

証明2（帰納法による証明）

[定理2の証明]
パスの長さをその辺の数とする。

主張1: 任意のAVLにおいて、任意の二つの葉へのパスの長さは高々1しか違わない。（すべての葉の深さは高々1しかちがわない）

（主張1の証明）任意の二つの異なる葉をxとyとする。これらのxとyの共通の先祖で一番下にあるもの（最近共通先祖NCA）をzとする。すると、根からxとyへの二つのパスのうち、根からzまでの前半部分は共通である。さらに二つのパスの後半である zからxへのパス P_{zx} と、zからyへのパス P_{zy} は、それぞれ、zの左の部分木 T_L と右の部分木 T_R に別々に分かれて含まれる。
一方で，AVL木の仮定より，二つの部分木 T_L と T_R の高さの差は高々1以下なので、 P_{zx} と P_{zy} の長さは高々1しか違わない。よって示された。（主張1の証明の終わり）

主張2: 任意のAVLにおいて、ある正整数 $h \geq 1$ が存在して、根から任意の葉へのパスの長さは h または $h-1$ である。

（主張2の証明）主張1より直ちに示される。（主張2の証明の終わり）

上の主張2から、任意のAVL木は、ある h に対して、根から深さが $h-1$ までの部分木は平衡完全2分木 B になり、残りの部分である深さ h の頂点からなる部分 C は B の葉の子供だけからなることがわかる（図）。

すると、B はちょうど $|B| = \sum_{i=0}^{h-1} 2^i = 2^h - 1$ 頂点を含む。一方で、C は $1 \leq |C| \leq 2^h$ である。よって、深さ h のAVL木の頂点数 N は、 $2^h - 1 \leq N \leq 2^h - 1 + 2^h = 2^{h+1} - 1$ である。これより、左の不等式をとって、$2^h - 1 \leq N$ がいえる。よって、次の導出を得る：

$$\rightarrow 2^h - 1 \leq N.$$

$$\rightarrow 2^h \leq N + 1. \quad （\because 1を移項）$$

$$\rightarrow \log(2^h) \leq \log(N+1). \quad （\because 両辺のlogをとる）$$

ここで、 $\log 2^h = d$ なので、

$$h \leq \log(N+1) \leq \log(N) + 1 = O(\log N).$$

したがって $h = O(\log n)$ が示された。（定理2の証明終わり）

（有村 文責）

EOF