**第3章 树学习指导**

1、时间

2020年4月07日周二（1-4班3-4节，5-10班、数学、大数据专业5-6节）

2020年4月09日周四（1-4班3-4节，5-10班、数学、大数据专业5-6节）

2020年4月14日周二（1-4班3-4节，5-10班、数学、大数据专业5-6节）

2020年4月16日周四（1-4班3-4节，5-10班、数学、大数据专业5-6节）

4次课共8学时。

2、内容

课前观看录播的视频；

腾讯课堂讲解本章重点和难点；

QQ群交流练习题（中间根据需要采用QQ视频或腾讯视频讲解）；

投票方式答题（点名、检查上课人数）；

问答。

具体课程安排：

|  |  |  |
| --- | --- | --- |
| 1 | 4月7日 | (1) 二叉树的定义、术语、性质、存储结构 |
| 2 | 4月9日 | (2) 遍历及其他操作、线索化；堆、选择树 |
| 3 | 4月14日 | (3) 树的存储结构、遍历；树和森林与二叉树的转换； |
| 4 | 4月16日 | (4) 树（森林）与二叉树的应用（哈夫曼树及其应用、集合表示与等价分类、表达式求值）。二叉排序树。 |

**3、知识要点**

**3.1 树及术语**

**树**是n个结点的有限集合，非空时必须满足：只有一个称为根的结点；其余结点形成m个不相交的子集，并称根的子树。

**抽象数据型二叉树；**

**树结点：**树中一个独立单元。包含一个数据元素及若干指向其子树的分支。

**树根**：树中唯一没有前趋的结点。

**结点的度(Degree)：**结点拥有的子树数，称为结点的度。

**树的度：**树中各结点的度的最大值。

**树叶(Leaf)**：度为0的结点。也称叶结点。除根和叶子以外的其他结点称为中间结点。

**双亲(Parent)和孩子(Child)**：我们把一个树结点的直接前趋称之为该结点的双亲；反之把一个树结点的所有直接后趋称为该结点的孩子。

**兄弟(Sibling)**：同一双亲的孩子之间互称为兄弟。

**结点的层次(Level)：**从根算起，根为第一层，根的孩子为第二层，树中任一结点的层次等于它的双亲的层次加1。

**树的深度(Depth)** 树中各结点层次的最大值称为树的深度或高度。

**树的宽度(width)** 树中某一层具有最多的结点数，称为树的宽度。

**有序树和无序树** 如果树中结点的各子树可看成从左至右是有次序的（即不能互换），则称该树为有序树。否则称为无序树。在有序树的最左边的子树的根称为第一孩子。最右边称为最后一个孩子。

**森林(Forest)：**m (m≥0)棵互不相交的树的集合。对树中每个结点而言，其子树的集合即为森林。由此也可以以森林和树的相互递归定义来描述树。

树的四种不同表示方法：

* 树形表示法；
* 嵌套集合表示法；
* 凹入表示法
* 广义表表示法。

**3.2 二叉树**

二叉树的定义：是n≥0个结点的有限集，它是空集（n=0）或由一个根结点及两棵互不相交的分别称作这个根的左子树和右子树的二叉树组成。

\*二叉树不是树的特殊情形，与度数为2的有序树不同。

抽象数据型二叉树。

**【性质1】**二叉树上第i层上的结点数目最多为2i-1（i≥1）；

**【性质2】**深度为k的二叉树至多有2k-1个结点（k≥1）；

**【性质3】**在任意一棵二叉树中，若终端结点的个数为n0，度为2的结点数为n2，则n0=n2+1；

**【性质4】**具有n个结点的完全二叉树的深度为int(log2n)+1。

**满二叉树**是一棵深度为k，结点数为2k-1的二叉树；

**完全二叉树**深度为 k 的，有n个结点的二叉树，当且仅当其每个结点都与深度为 k 的满二叉树中编号从 1 至 n 的结点一一对应，称之为完全二叉树，或满二叉树在最下层自右向左去除部分结点。

**【性质5】**如果对一棵有n个结点的完全二叉树的结点按层序编号，则对任一结点i 有：

* + 1. 如果 i = 1，则结点i是二叉树的根，无双亲；如果i＞1，则其双亲结点是 i / 2；
    2. 如果 2 i ＞ n，则结点i无左孩子结点，否则其左孩子结点是 2 i ；
    3. 如果 2 i + 1＞n，则结点i无右孩子结点，否则其右孩子结点是 2 i + 1。

上述性质要会证明。

**（2）二叉树存储结构**

二叉树的顺序存储结构就是把二叉树的所有结点按照层次顺序存储到连续的存储单元中。（存储前先将其画成完全二叉树）

二叉树链式存储。BinTNode的结构为lchild|data|rchild，把所有BinTNode类型的结点，加上一个指向根结点的BinTree型头指针就构成了二叉树的链式存储结构，称为二叉链表。它就是由根指针root唯一确定的。

证明：n个结点的二叉链表共有2n个指针域，n+1个空指针。

**（3）二叉树的遍历**

二叉树的遍历：根据访问结点的次序不同可得三种遍历：

* 先序遍历（或先根遍历）
* 中序遍历（或中根遍历）
* 后序遍历（或后根遍历）

三种遍历算法的递归与非递归的程序实现；

三种遍历的时间复杂度为O（n）。

* 二叉树的层序遍历（补充）

**（4）线索二叉树**

利用二叉链表中的n+1个空指针域来存放指向某种遍历次序下的前趋结点和后继结点的指针，这些附加的指针就称为“线索”，加上线索的二叉链表就称为线索链表。线索使得查找中序前趋和中序后继变得简单有效。

算法：求给定结点p的中序前驱$p;

求给定结点p的中序前驱p$;

中序遍历线索二叉树；

**（5）堆**

如果一棵完全二叉树的任意一个非终端结点的元素都不小于其左儿子结点和右儿子结点（如果有的话）的元素，则称此完全二叉树为最大堆。

同样，如果一棵完全二叉树的任意一个非终端结点的元素都不大于其左儿子结点和右儿子结点（如果有的话）的元素，则称此完全二叉树为最小堆。

选择树与败者树

**3.3 树和森林**

树和森林及二叉树的转换是唯一对应的。

转换方法：

* 树变二叉树：兄弟相连，保留长子的连线。
* 二叉树变树：结点的右孩子与其双亲连。
* 森林变二叉树：树变二叉树，各个树的根相连。

树的存储结构：

* 双亲表示法：对于求指定结点的双亲或祖先十分方便，但不适于求指定结点的孩子及后代。
* 孩子表示法：为树中每个结点设置一个孩子链表，并将存放在一个向量中。
* 双亲孩子表示法：将双亲链表和孩子链表结合。
* 孩子兄弟表示法：附加两个分别指向该结点的最左孩子和右邻兄弟的指针域。

**3.4 树的应用**

**（1）哈夫曼树及应用**

假设有n个权值{w1，w2，…wn }，试构造一棵有n个叶子结点的二叉树，每个叶子结点带权为wi。显然，这样的二叉树可以构造出多棵，其中必存在一棵带权路径长度 WPL 取最小的二叉树，称该二叉树为最优二叉树也称哈夫曼树。

(注意：最优二叉树中只有度为0和度为2的结点）

　　在叶子的权值相同的二叉树中，完全二叉树的路径长度最短。

　　哈夫曼树有n个叶结点，共有2n-1个结点，没有度为1的结点，这类树又称为严格二叉树。

　　变长编码技术可以使频度高的字符编码短，而频度低的字符编码长，但是变长编码可能使解码产生二义性。如00、01、0001这三个码无法在解码时确定是哪一个，所以要求在字符编码时任一字符的编码都不是其他字符编码的

前缀，这种码称为前缀码（其实是非前缀码）。

哈夫曼树的应用最广泛地是在编码技术上，它能够容易地求出给定字符集及其概率分布的最优前缀码。哈夫曼编码的构造很容易，只要画好了哈夫曼树，按分支情况在左路径上写代码0，右路径上写代码1，然后从上到下到叶结点的相应路径上的代码的序列就是该结点的最优前缀码。

**（2）判定树**

**（3）集合表示与等价分类**

**（4）表达式求值**

**（5）二叉排序树**

**4、习题**

**（1）填空题**

1）树是n（n≥0）结点的有限集合，在一棵非空树中，有 个根结点，其余的结点分成m（m＞0）个 的集合，每个集合都是根结点的子树。

**2）**树中某结点的子树的个数称为该结点的 ，子树的根结点称为该结点的 ，该结点称为其子树根结点的 。 3）一棵二叉树的第i（i≥1）层最多有 个结点；一棵有n（n>0）个结点的满二叉树共有 个叶子结点和 个非终端结点。

4）设高度为h的二叉树上只有度为0和度为2的结点，该二叉树的结点数可能达到的最大值是 ，最小值是 。

5）深度为k的二叉树中，所含叶子的个数最多为 。

6）具有100个结点的完全二叉树的叶子结点数为 。

7）已知一棵度为3的树有2个度为1的结点，3个度为2的结点，4个度为3的结点。则该树中有 个叶子结点。

8）某二叉树的前序遍历序列是ABCDEFG，中序遍历序列是CBDAFGE，则其后序遍历序列是 。

9）在具有n个结点的二叉链表中，共有 个指针域，其中 个指针域用于指向其左右孩子，剩下的 个指针域则是空的。

10）在有n个叶子的哈夫曼树中，叶子结点总数为 ，分支结点总数为 。

**（2）选择题**

1）把一棵树转换为二叉树后，这棵二叉树的形态是（ ）。

A．唯一的 Ｂ．有多种

C．有多种，但根结点都没有左孩子 Ｄ．有多种，但根结点都没有右孩子

2）由3个结点可以构造出多少种不同的二叉树？（ ）

A．2 B．3 C．4 D．5

3）一棵完全二叉树上有1001个结点，其中叶子结点的个数是（ ）。

A．250 B． 500 C．254 D．501

答案：D

4）一个具有1025个结点的二叉树的高h为（ ）。

A．11 B．10 C．11至1025之间 D．10至1024之间

答案：C

5）深度为h的满m叉树的第k层有（ ）个结点。(1=<k=<h)

A．mk-1  B．mk-1 C．mh-1 D．mh-1

6）利用二叉链表存储树，则根结点的右指针是（ ）。

A．指向最左孩子 B．指向最右孩子 C．空 D．非空

7）对二叉树的结点从1开始进行连续编号，要求每个结点的编号大于其左、右孩子的编号，同一结点的左右孩子中，其左孩子的编号小于其右孩子的编号，可采用（ ）遍历实现编号。

A．先序 B. 中序 C. 后序 D. 从根开始按层次遍历

8）若二叉树采用二叉链表存储结构，要交换其所有分支结点左、右子树的位置，利用（ ）遍历方法最合适。

A．前序 B．中序 C．后序 D．按层次

9）在下列存储形式中，（ ）不是树的存储形式？

A双亲表示法 B孩子链表表示法 C孩子兄弟表示法 D顺序存储表示法

10）一棵非空的二叉树的先序遍历序列与后序遍历序列正好相反，则该二叉树一定满足（ ）。

A．所有的结点均无左孩子 B．所有的结点均无右孩子

C．只有一个叶子结点 D．是任意一棵二叉树

11）设哈夫曼树中有199个结点，则该哈夫曼树中有（ ）个叶子结点。

A．99 B．100

C．101 D．102

12）若X是二叉中序线索树中一个有左孩子的结点，且X不为根，则X的前驱为（ ）。

A．X的双亲 B．X的右子树中最左的结点

C．X的左子树中最右结点 D．X的左子树中最右叶结点

13）引入二叉线索树的目的是（ ）。

A．加快查找结点的前驱或后继的速度

B．为了能在二叉树中方便的进行插入与删除

C．为了能方便的找到双亲

D．使二叉树的遍历结果唯一

14）设F是一个森林，B是由F变换得的二叉树。若F中有n个非终端结点，则B中右指针域为空的结点有（ ）个。

A．n−1 B．n C．n + 1 D．n + 2

15）n（n≥2）个权值均不相同的字符构成哈夫曼树，关于该树的叙述中，错误的是（ ）。

A．该树一定是一棵完全二叉树

B．树中一定没有度为1的结点

C．树中两个权值最小的结点一定是兄弟结点

D．树中任一非叶结点的权值一定不小于下一层任一结点的权值

**（3）应用题**

1）试找出满足下列条件的二叉树

① 先序序列与后序序列相同 ②中序序列与后序序列相同

③ 先序序列与中序序列相同 ④中序序列与层次遍历序列相同

2）设一棵二叉树的先序序列：A B D F C E G H，中序序列：B F D A G E H C

①画出这棵二叉树。

②画出这棵二叉树的后序线索树。

③将这棵二叉树转换成对应的树（或森林）。

3） 假设用于通信的电文仅由8个字母组成，字母在电文中出现的频率分别为0.07，0.19，0.02，0.06，0.32，0.03，0.21，0.10。

① 试为这8个字母设计赫夫曼编码。

② 试设计另一种由二进制表示的等长编码方案。

③ 对于上述实例，比较两种方案的优缺点。

4）已知下列字符A、B、C、D、E、F、G的权值分别为3、12、7、4、2、8，11，试填写出其对应哈夫曼树HT的存储结构的初态和终态。

**（4）算法设计题**

以二叉链表作为二叉树的存储结构，编写以下算法：

1. 统计二叉树结点总数、度为1的结点个数，度为2的结点个数，度为0的结点个数。
2. 输出并统计二叉树的叶结点个数。
3. 设计二叉树的双序遍历算法（双序遍历是指对于二叉树的每一个结点来说，先访问这个结点，再按双序遍历它的左子树，然后再一次访问这个结点，接下来按双序遍历它的右子树）。
4. 用按层次顺序遍历二叉树的方法，统计树中具有度为1的结点数目。
5. 判别两棵树是否相等。
6. 交换二叉树每个结点的左孩子和右孩子。
7. 设二叉树T采用二叉链表结构存储，试设计算法求二叉树的深度。
8. 计算二叉树最大的宽度（二叉树的最大宽度是指二叉树所有层中结点个数的最大值）。
9. 求任意二叉树中第一条最长的路径长度，并输出此路径上各结点的值。
10. 求任意二叉树中第一条最短的路径长度，并输出此路径上各结点的值。
11. 输出二叉树中从每个叶子结点到根结点的路径。
12. 设二叉树T采用二叉链表结构存储，试设计算法查找后序遍历的第一个被访问的结点。
13. 设二叉树T采用二叉链表结构存储，试设计算法求给定两个结点的最近双亲结点。
14. 设二叉树T采用二叉链表结构存储，试设计算法求出二叉树中离根最近的第一个叶子结点。
15. 设计算法，以(Key，LT， RT)的形式打印二叉树，其中 Key为结点的值，LT和RT是以括 号形式表示的二叉树。