ALG LAB6

王世炟 PB20151796 2022/11/19

Huffman 编码问题

实验内容

编程实现 Huffman 编码问题,并理解其核心思想。

对文件 original.txt 中所有的大小写字母、数字 (0-9) 以及标点符号 (即:除空格 换行等之外的所有字符)按照 Huffman 编码方式编码为 01序列,输出如下格式的 table.txt 文件,并在控制台打印压缩率。

算法设计思路

首先,将所有的字符以及它们出现的频率存在结点中,并将这些结点放入优先队列,形成森林。

每一次将优先队列的前两个结点(频率最小的两个结点)出队,并合并成为一个新的结点,之后再将此结点加入优先队列,并维护优先队列的性质。

当优先队列中只剩下一个结点的时候,树便建立成功。

之后从根节点开始向下遍历, 依次添加编码。

源码+注释

结点结构:

优先队列比较方法:

```
struct CMP
{
   bool operator()(const pHFMNode &p, const pHFMNode &q)
   {
      return p->freq > q->freq;
   }
};
```

建立哈夫曼树:

```
void Huffman(priority_queue<pHFMNode, vector<pHFMNode>, CMP> &vctNode)
{
    while (vctNode.size() > 1)
    {
        pHFMNode first = vctNode.top(); //取vctNode森林中頻度最小的树根
        vctNode.pop();
        pHFMNode second = vctNode.top(); //取vctNode森林中頻度第二小的树根
        vctNode.pop();
        pHFMNode merge = new HFMNode; //合并上面两个树
        merge->freq = first->freq + second->freq;
        merge->lchild = first; //小的放左子树
        merge->rchild = second; //大的放右子树
        vctNode.push(merge); //向vctNode森林中添加合并后的merge树
    }
    return;
}
```

输出编码:

```
void PrintHuffman(pHFMNode node, vector<int> &vctCode, ofstream &outfile)
{
   if ((node->lchild == NULL) && (node->rchild == NULL))
   {
       node->code.assign(vctCode.begin(), vctCode.end()); //复制vctCode到code
       outfile << node->value << " " << left << setw(8) << node->freq;
       for (vector<int>::iterator iter = node->code.begin(); iter != node->code.end(); iter++)
           outfile << *iter;
       }
       outfile << endl;</pre>
       code_length += vctCode.size() * node->freq;
   }
   else
    {
       vctCode.push_back(0); //遇到左子树时给vctCode中加一个0
       PrintHuffman(node->lchild, vctCode, outfile);
       vctCode.pop_back(); //回溯,删除刚刚加进去的1
       vctCode.push_back(1); //遇到右子树时给vctCode中加一个1
       PrintHuffman(node->rchild, vctCode, outfile);
       vctCode.pop_back(); //回溯, 删除刚刚加进去的0
   }
   return;
}
```

算法测试结果

可得到编码如下:

2008 2008	字符	出现频率	
	F		
	Е	95	0011000010
	R	49	00110000110
	Р	50	00110000111
	"	391	00110001
	М	197	001100100
	1	5	00110010100000
3	2	1	0011001010000100
	K	1	0011001010000101
	3	3	001100101000011
	Χ	3	001100101000100
	7	1	0011001010001010
	5	2	0011001010001011
	4	4	
	Q		
	U	15	
	9	4	
1			
1 12	V		
	;		
1	!		
	T		
N			
j 1/20 0011100011 I 261 00111001 I 265 0011101 I 269 0011101 I 2630 0011101 I 7433 0100 I 7436 0110 I 7886 0110 I 8891 0111 I 1873 1810 I 2733 1810 I 1814 18110 I 141 18111000 I 141 181110000 I 141 1811100100 I 141 1811100100 I 141 18111001000 I 141 18111001000 I 141 181110010000 I 141 18111001001000 I 181 18111001001000 I 181 1811100100000 I 181 18111000000000000000000000000000000000			
N 245 000110010 k 985 0011101 k 985 0011101 c 7640 0010 c 7650 010 k 891 0110 k 891 0111 c 1714 180 k 9253 101 k 1911 000 k 1100 000 k 11100 k			
I 6361 001110011 8 937 0011101 8 2029 001111 7 7643 0100 1 7886 0110 1 011 100 1 102 1011 2 1721 100 1 123 101100 2 273 101100 2 274 10110000 3 280 10110010 4 121 101100100 5 18 1011001010 6 18 10111001010 7 12 10111001010 8 1011100101011 1001001010 9 12 1011100101011 10 12 101100101011 10 100 100 10 1100 11000 10 1100 110000 10 100 1100000 10 11000000 10 <td></td> <td></td> <td></td>			
k 985 0911101 r 763 000 r 765 0101 r 785 010 r 785 010 r 785 011 r 785 011 r 785 011 r 191 000 r 192 111 r 191 100 r 191 101 r 214 10110000 r 191 10110000 r 191 101100000 r 101 101100000 r 101 1011000100 r 21 10110001010 r 21 10110001010 r 21 10110001010 r 22 10110001010 r 100 100 r 100 100 r 1100 100 r 110000000000000000000			
b. 2029 001111 c 7633 0100 s 7685 0110 l 7896 0110 e 17214 100 p 2143 101100 p 2143 10111000 p 280 101110000 p 280 101110010 p 141 101110010 p 15 10111001010 p 18 101110010110 p 18 1011100101111 p 10 111110010111 p 18 101110010111 p 10 1110010111 p 10 1110000 p 100 110000 p 10 11100000 p 10 111000000 p 10 11000000 p 10 1110000000 p 10 1110000000 p 11100000000 <t< td=""><td></td><td></td><td></td></t<>			
r 7643 0100 s 7685 0101 r 7886 0110 s 7849 0110 c 17214 100 s 17214 1010 p 2149 10110 c 174 10110 x 274 1011000 8 280 10110001 9 10 101100100 1 141	k		
5 7685 0101 1 7886 0110 8 1911 8 17214 100 9 2133 1010 y 2173 101101 8 277 10110800 8 280 1011108100 6 18 101108010 6 18 10111081011 7 2 10 1011080111 8 101110801110 1 9 18 101110801111 v 192 101110801111 v 192 101110801111 v 193 110108000 3 100 1100000 3 11000000 3 11000000 4 31 11000000 5 39 11000000 6 11 1100000 7 2 11 11 8 11 11 9 11 <td>b</td> <td></td> <td></td>	b		
i. 886 8191 0111 b 8491 0111 c 1721 100 c 1723 10100 c 2533 101100 c 2149 1011000 c 101110001 c 274 1011100100 d 141 1011100100 d 141 10111001010 c 67 101110010110 d 18 101110010110 d 18 101110010111 v 192 101100111 d 193 1100 a 103 1100 d 1160 1100 a 100 1100 a 1100 1100 a 11100 1100 a 11100 </td <td>r</td> <td></td> <td></td>	r		
h 8491 0111 c 17214 100 c 2523 1010 p 2492 1011001 g 2773 10110000 B 2804 1011100100 B 2804 1011100100 G 181 10111001010 G 182 101110010101 g 182 101110010101 g 183 101110010101 g 183 10111001101 g 184 1011100101101 g 189 111001001101 g 180 1111001001 g 180 111001001 g 180	S		
e 17214 100 n 9253 1010 p 2149 101100 y 2173 10111000 x 274 10110000 b 280 101110010 c 141 101110010 c 18 1011100110 c 18 10111001110 c 29 10110010111 r 192 10111001 c 193 1001 d 193 11001 c 26 101111 c 27 1100101 d 11001 1100101 d 11000 1100101 g 275 111001001 c 10 1110010010	i		
n 9253 1010 p 2149 101100 y 2173 101100 x 274 10110000 8 280 10110010 7 67 101100101 6 18 1011001010 7 21 1011001011 8 10111001011 9 21 1011001011 10 10 1011001 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10<	h		
p 2149 101100 y 2173 101101 x 274 10111000 B 280 10111001 H 141 1011100101 G 18 101110010110 (18 101110010110 J 28 101110010110 1 18 101110010111 2 292 1011101 4 192 1011101 5 263 10111 6 1883 1011 7 153 11000 8 11001000 10 11001000 10 110010000 10 110010010 10 1110010110 10 11100101100 10 11100101100 10 11100101100 10 11100101100 10 11100101100 10 11100101100 10 111001011100 10 1110010	е		
y 2173 101100 x 274 10111000 a 280 101110010 H 411 1011100100 c 67 10111001010 (18 101110010110 (18 101110010111 - 22 10111001111 - 29 10111001 c 263 101111 c 263 101110 a 192 1011001 a 193 11001 a 193 11000 a 100 11001 a 100 11001000 a 110010000 11001000 a 11100101100 a 11100101100 a 11100101100 a 11100101101 a 11100101101 a 11100101101 a 11100101101 a 111001011101 a 111001011010	n		
x 274 101110000 8 280 101110010 * 67 10111001010 6 18 1011100101100 (18 101110010111 7 21 101110010111 8 22 10111001 9 1011101 0 283 101111 0 9890 1000 10 11001000 13 111001000 13 1110010010 16 111001001 17 110010000 10 1110010100 10 20 11100101100 10 40 11100101100 10 42 111001011101 1 45 111001011101 1 587 11100101111 1 584 1110011111	р		
B 280 101100100 H 141 101100100 G 18 10110010110 (18 1011100101110) 18 101110010111 - 29 10111001111 - 29 1011101 v 192 1011101 c 2663 101111 g 1980 1100 g 2761 11100100 g 2761 111001000 g 133 111001001 g 153 111001001 g 39 110010101 g 40 1110010101 g 45 11100101101 g 45 11100111010 g 45 11100111010 g 45 11100111010 g 45 11100111010 g 45 11100101100 g 11100111010 g 11100111010	У		
H			
? 67 10111001010 6 18 101110010110 (18 101110010111 7 21 1011100101111 - 292 10111001 v 192 1011101 c 2663 101111 o 930 1100 a 1834 1101 g 2761 11100100 d 312 111001001 C 160 111001001 S 339 111001001 S 339 111001010 Q 42 11100101101 D 42 111001011010 D 45 111001011101 2 45 111001011101 3 111001011101 4 111001011101 3 111001011101 4 111001011101 3 111001011101 4 111001011101 3 111000111101 <t< td=""><td></td><td></td><td></td></t<>			
G 18 1011100101100 (18 1011100101101) 18 1011100101110 - 21 10111001111 - 29 101110011 v 1192 1011101 c 2663 101111 o 9890 1100 a 1834 1101 b 2761 111000 c 153 111001000 c 160 1110010010 c 40 11100101100 c 40 111001011010 c 45 111001011010 d 45 11100101110 c 45 111001011101 d 45 111001011101 c 45			
(18 1011100101101) 18 1011100101110 Y 21 101110010111 v 192 10111001 v 192 1011101 c 2663 101111 o 9890 1100 a 10834 1100 A 312 11100100 C 160 111001001 C 160 1110010110 S 339 11100101100 C 40 111001011010 D 42 111001011010 z 45 111001011101 z 45 11100111101 z 45 11100111101 z 1 1 z 1 1 z 1 1 z 2 1 z 1 1			
) 18 1011100101111 Y 21 101110011111 - 292 101110011 V 1192 1011101 0 2663 101111 0 989 1100 a 10834 1101 g 2761 1110010 A 312 111001001 C 160 1110010011 C 160 1110010010 C 80 1110010110 D 42 11100101101 Z 45 111001011101 Z 45 111001011101 J 93 111001011101 L 45 111001011101 L 5847 111001011101	G		
Y 21 1011100101111 - 292 101110011 v 1192 1011101 c 2663 101111 a 1089 1100 a 10834 1101 g 2761 11100100 A 312 111001000 G 153 111001001 C 160 111001001 C 160 111001010 C 80 1110010110 c 40 11100101101 z 45 11100101101 z 45 11100101110 1 45 11100101111 z 584 1100101111 584 1512 1110011			
- 292 101110011 v 1192 1011101 c 2663 101111 o 9890 1100 a 10834 1101 g 2761 11100100 q 153 111001000 Q 153 1110010010 S 339 111001010 O 80 11100101100 I 40 111001011010 Z 45 111001011101 J 93 111001011101 J 93 111001011111 I 5847 111001			
v 1192 1011101 c 2663 101111 o 9890 1100 a 10834 1101 g 2761 1110010 A 312 110010010 C 160 1110010011 S 339 1110010010 O 80 11100101100 D 40 111001011010 z 45 111001011100 L 45 111001011101 J 93 111001011101 J 93 111001011101 J 93 111001011101 J 93 111001011111 J 5847 111001	Υ		
c 2663 101111 o 9890 1100 a 10834 1101 g 2761 111001 A 312 111001000 q 153 111001001 C 160 111001001 S 339 111001010 c 40 11100101100 z 42 111001011011 z 45 111001011100 L 45 111001011101 J 93 111001011101 J 93 111001011101 J 1512 111001011101 J 5847 111001			
o 9890 1100 a 10834 1101 g 2761 111000 A 312 111001000 q 153 1110010010 C 160 111001011 S 339 111001010 c 40 11100101101 D 42 11100101101 z 45 11100101101 J 45 11100101101 J 93 11100101111 1 5847 1110011			
a 10834 1101 g 2761 111000 A 312 111001000 q 153 1110010010 C 160 111001011 S 339 111001010 c 40 11100101101 D 42 11100101101 z 45 11100101101 J 93 11100101111 c 1512 1110011 1 5847 11101			
g 2761 111000 A 312 111001000 q 153 1110010010 C 160 1110010011 S 339 111001010 C 80 11100101100 E 40 111001011011 Z 45 111001011100 L 45 111001011101 J 93 111001011111 E 1512 1110011 1 5847 11101			
A 312 111001000 q 153 1110010010 C 160 1110010011 S 339 1110010100 C 40 11100101100 D 42 11100101101 z 45 111001011100 L 45 111001011101 J 93 11100101111 . 1512 111001 1 5847 11101			
q 153 1110010010 C 160 1110010011 S 339 111001010 0 80 11100101100 : 40 111001011010 D 42 11100101110 z 45 11100101110 J 93 11100101111 : 1512 1110011 1 5847 11101			
C 160 1110010011 S 339 111001010 0 80 11100101100 : 40 111001011010 D 42 111001011011 z 45 111001011101 J 93 11100101111 : 1512 1110010 1 5847 11101			
S 339 111001010 0 80 11100101100 : 40 11100101101 D 42 11100101101 z 45 11100101110 L 45 111001011101 J 93 11100101111 . 1512 1110011 1 5847 11101			
0 80 11100101100 : 40 111001011010 D 42 111001011011 z 45 111001011101 J 93 11100101111 . 1512 1110011 1 5847 11101			
: 40 111001011010 D 42 111001011011 z 45 111001011100 L 45 111001011101 J 93 11100101111 . 1512 1110011 1 5847 11101			
D 42 111001011011 z 45 111001011100 L 45 111001011101 J 93 11100101111 1 5847 11101			
z 45 111001011100 L 45 111001011101 J 93 11100101111 . 1512 1110011 1 5847 11101			
L 45 111001011101 J 93 11100101111 . 1512 1110011 1 5847 11101			
J 93 11100101111 . 1512 1110011 l 5847 11101			
. 1512 1110011 1 5847 11101			
1 5847 11101			
	C	11313	

压缩率: 0.641293