Huffman编码压缩

朱河勤 PB16030899

2017 年 11 月 11 日

目录

1 需求分析

1 正确创建huffman 树，并得到huffman编码

2 压缩编码文件，使一般的文本文件变小，二进制文件可能不会变小

3 解压文件，注意生成的文件要考虑同名，所以要考虑文件名的操作

4 压缩解压时显示进度条

5 解压完毕，统计文件数，压缩率，压缩时间等等

2 概要设计

2.1 结点node

ADT linkList{

数据对象 ：key , val, visited ,left,right

数据关系 ：属性集合

基本操作 ：

node() 新建一个结点

node(const node& nd 复制一个结点

operator¡ 重载操作符¡ }

2.2 huffman树

ADT linkList{

数据对象 ：root , encode_map, total

数据关系 ：属性集合

基本操作 ：

huffman(map) 传递字符的频率信息以map形式，然后建立huffman树

encode(string) 传递文件名，压缩

decode(string) 传递文件名，解压缩

preorder() 先序遍历，获得huffman编码

display 显示所有信息，包括结点信息，以及huffman编码 }

3 详细设计

3.1 编码，解码算法

```
template<typename ky,typename wt>
void huffman<ky,wt>::preOrder(node<ky, wt>* nd,string s)
{
    if(nd->left == NULL){
        encode_map[nd->key] =s;
        decode_map[s] = nd->key;
        delete nd;
        return ;
```

```
    }
    preOrder(nd->left ,s+'0');
    preOrder(nd->right ,s+'1');
    delete nd;
}
template<typename ky,typename wt>
string  huffman<ky,wt>::decode(string zipfile_name)
{
    string uniFileName(string);
    FILE * src = fopen(zipfile_name.c_str(),"rb");
    char file_name[nameLength];
    fgets(file_name ,nameLength,src);
    int ct=-1;
    while(file_name[++ct]!='\n');
    int pos = zipfile_name.find('.');
    if(pos==string::npos)pos=zipfile_name.size();
    string name(zipfile_name.substr(0,pos)) ,suffix(file_name ,file_name+ct),
    file=uniFileName(file);
    cout<<"extracting compressed file :"<<file<<endl;
    FILE * f = fopen(file.c_str(),"wb");
    if(!f){
        cerr<<"Error! Maybe the file doesn't exist!"<<endl;cerr<<"open error
        abort( );
    }
    char t[numDigit];
    fgets(t,numDigit,src);
    int sz=atoi(t);
    char code[sz];
    fread(code,sz,1,src);
    cout<<code<<endl;
    cout<<"SZ"<<sz<<endl;
//    return string();
    map<string,char>  mp;
```

```
    int idx=0;
    for(int i =0;i<sz;++i ){
        if(code[i]==' '){
            mp[string(code+idx,code+i)]=code[++i];
            idx=i+1;
        }
    }
    mapprint(mp);
    char buff[once],*p;
    string res;
    while(!feof(src)){
        p=buff;
        int cur=0;
        while(cur<once){
            ToEightBin(p,fgetc(src));
        }
        cout<<buff;
        return string();
        while(buff[++cur]!='\0');
        for(int i =0;i<cur;++i){
            res.append(1,buff[i]);
            if(mp.count(res)!=0){
                fputc(mp[res],f);
                res.clear();
            }
        }
    }
    fclose(src);
    fclose(f);
}

template<typename ky,typename wt>
string huffman<ky,wt>::encode(string file_name)
```

```
{
    string uniFileName(string);
    int pos =file_name.rfind('.');
    if(pos==string::npos)pos=file_name.size();
    string zipfile = file_name.substr(0,pos)+string(".zzip");
    zipfile = uniFileName(zipfile);
    cout<<"generating zip file :"<<zipfile<<endl;
    FILE * dst = fopen(zipfile.c_str(),"wb");
    FILE * f = fopen(file_name.c_str(),"rb");
    fputs(file_name.substr(pos).c_str(),dst);
    fputc('\n',dst);
    string data;
    for(class map<string,ky>::iterator i=decode_map.begin();i!=decode_map.en
        data.append((i->first));
        data.append(" ");
        data+=(i->second);
    }
    cout<<data<<endl<<"SZ"<<data.size()<<endl;;
//    mapprint( encode_map);
//    return string();
    int data_size = data.size();  // calculate the size of the code_data
    char sz[numDigit];
    itoa(data_size,sz,numDigit);
    int ct=0;
    for(;sz[ct];++ct)fputc(sz[ct],dst);
    fputc('\n',dst);
//    cout<<data<<data_size<<endl;
    fwrite(data.c_str(),data_size,1,dst);
    int  sum=0,digit=0,num;
    char src[once];
    while(!feof(f)){
        num = fread(src,sizeof(char),once,f);
        string tmp;
```

```
        for(int i =0;i<num;++i){
            tmp=encode_map[src[i]];
            for(int j =0;j<tmp.size();++j){
                if(tmp[j])sum += 1<<(digit%8);
                ++digit;
                if(digit==8){
                    fputc(sum,dst);
                    digit=sum=0;
                }
            }
        }
    }
    if(digit!=0){    //mark
        fputc(sum,dst);
        fputc(digit,dst);
    }
    else fputc(0,dst);
    fclose(f);
    fclose(dst);
    return zipfile;
}
```

3.2 输出进度条

通过输出@以显示当天进度，提示用户

3.3 改变文件名

由于当前工作目录可能有解压或压缩后的同名文件，要进行处理，我的方法是在后缀前，文件名后加上(n) 其中，n为数字

3.4 统计信息

处理后，统计信息，包括压缩的文件名目录，所用的时间，文件大小，压缩率等等

4 调试分析

4.1 编码信息的存储

我先计算了编码信息的大小，然后存在文件中，以键值对的形式,如"001 k"再存储编码信息，这样就能更快的读取，并能节省空间

4.2 最后一个字节

由于储存时一char八位一次地储存，所以最后一字节可能凑不齐，我通过补0地方式，并在最后再增加一字节记录补0的个数来解决

4.3 文件名的处理

文件名要要记录后缀以识别文件种类，以及文件是否存在，能否打开，生成的文件是否重名。比如，当发现重名时，你要生成（n）形式，但可能(n)也已经存在，所以要一直检查（1），（2）...检查到一个较大的数，或者是，第一次未出现时，才能确定N

5 用户手册

本程序需要在windows环境下运行，可以直接打开.exe文件，或者在命令行，支持命令行参数 下面为程序运行界面

```
C:\Users\mbinary\Desktop\dataStructrue\huffman
input file names separated by space
OR press enter to use the default file:初心未改.mp4


opening 初心未改.mp4...
the file 初心未改.zzip already exists! continue?  [Y/n]:generating zip file :初
心未改(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
compress 初心未改.mp4 successfully
```

```
continue to uncompress? [Y/n]

the file 初心未改(2).mp4 already exists! continue?  [Y/n]:
extracting compressed file :初心未改(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
get 初心未改(3).mp4 successfully

dealt file number  1
初心未改.mp4  18891ms  8274.84KB/8296.13KB :99.74%
```

6 测试结果

6.1 文本文件

输入了几个不存在的文件，检查强壮性。同时发现文本文件压缩率还是很令人满意的

```
input file names separated by space
OR press enter to use the default file:初心未改.mp4
em cvem ven   em.cc

opening em...
Path Error! Opening em Failed
opening cvem...
Path Error! Opening cvem Failed
opening ven...
Path Error! Opening ven Failed
opening em.cc...
the file em.zzip already exists! continue?  [Y/n]:
generating zip file :em(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
compress em.cc successfully
```

```
continue to uncompress? [Y/n]

the file em(2).cc already exists! continue?  [Y/n]:
extracting compressed file :em(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
get em(3).cc successfully

dealt file number  4
em.cc  2266ms  4.69KB/7.57KB :61.87%
```

6.2 二进制文件1

几乎没有压缩，原因应该是.exe文件本来就压缩得很好了，几乎已经是随机的编码了，所以huffman没有效果，这也是文件不能重复压缩直至没有大小的原因。

```
input file names separated by space
OR press enter to use the default file:初心未改.mp4
GitHubDesktopSetup.exe

opening GitHubDesktopSetup.exe...
the file GitHubDesktopSetup.zzip already exists! continue?  [Y/n]:
generating zip file :GitHubDesktopSetup(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
compress GitHubDesktopSetup.exe successfully

continue to uncompress? [Y/n]

the file GitHubDesktopSetup(2).exe already exists! continue?  [Y/n]:
extracting compressed file :GitHubDesktopSetup(2).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
get GitHubDesktopSetup(3).exe successfully
```

```
dealt file number  1
GitHubDesktopSetup.exe  309610ms  81718.46KB/81718.46KB :100.00%
```

6.3 二进制文件2

```
input file names separated by space
OR press enter to use the default file:初心未改.mp4


opening 初心未改.mp4...
the file 初心未改.zzip already exists! continue?  [Y/n]:generating zip file :初
心未改(3).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
compress 初心未改.mp4 successfully

continue to uncompress? [Y/n]

the file 初心未改(3).mp4 already exists! continue?  [Y/n]:
extracting compressed file :初心未改(3).zzip
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
get 初心未改(4).mp4 successfully

dealt file number  1
初心未改.mp4  13688ms  8274.84KB/8296.13KB :99.74%
```