张伯阳 1551265

词法分析器

1. **实验内容**

1．给出某高级程序设计语言的单词子集及机内表示，试编写一个词法分析器，输入为源程序字符串，输出为单词的机内表示序列。

2．在完成以上基本要求的情况下，对程序功能扩充：

（1）增加单词（如保留字、运算符、分隔符等）的数量；

（2）将整常数扩充为实常数；

（3）增加出错处理功能；

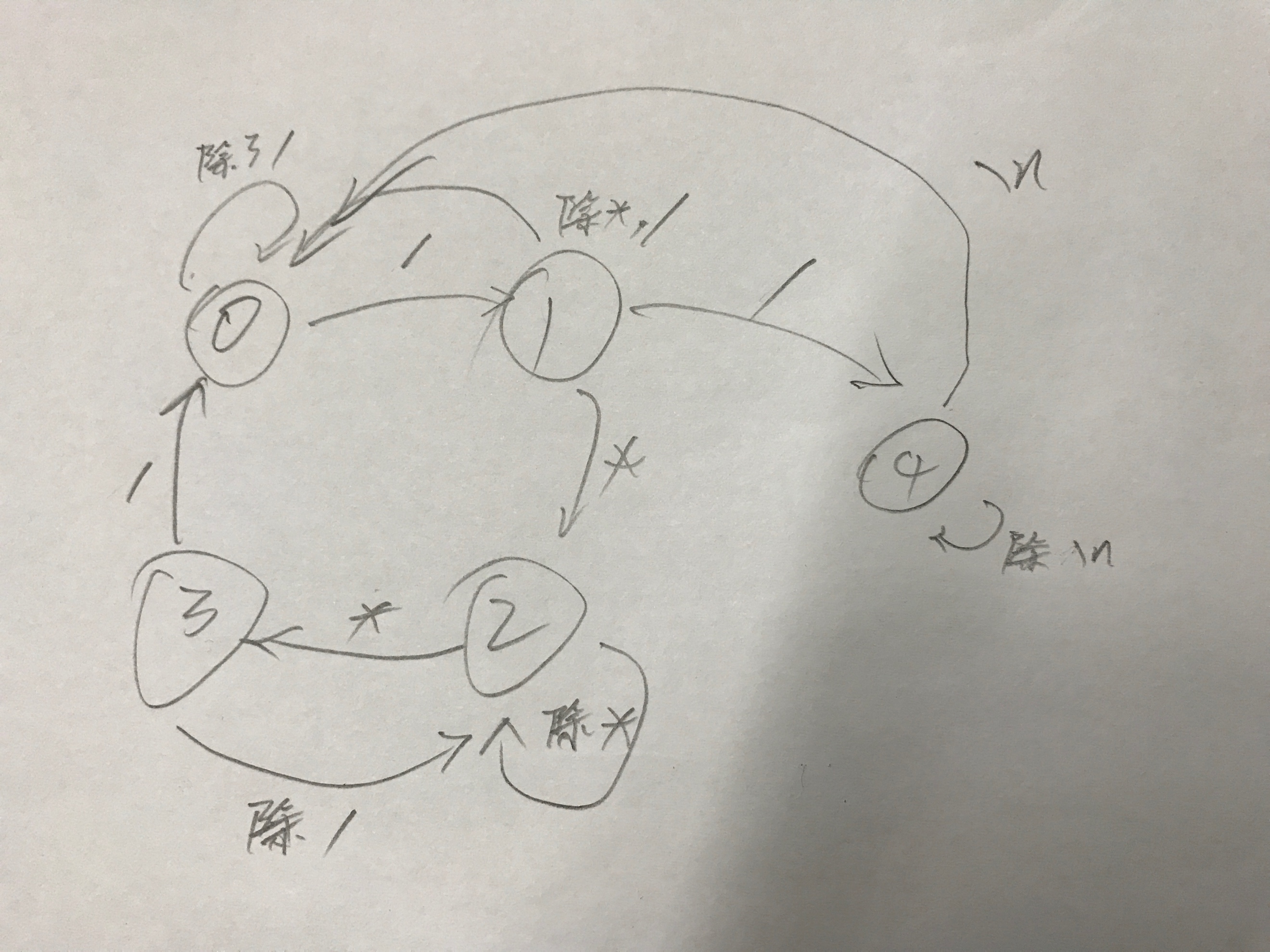
（4）增加预处理程序，每次调用时都将下一个完整的语句读入扫描缓冲区，去掉注解行，并能够对源程序列表打印。

1. **词法分析器的原理设计**

* **预处理：去除注释段**

去除/\*………\*/段注释 以及“//”行注释

实现该功能使用了自动机中的DFA



* **主要功能：**

1、基本功能：实现了基本的关键词，保留字，常数，运算符，界符的识别

* 关键字表包括（基本的c保留字和库函数保留字）："ifstream","const","short","int","char","bool","double","float","if","else","while","for","do","break","void","switch"可任意扩展
* 运算符： :=、=、<、<=、>、>=、+、-、\*、/
* 界符： {、}、（、）、；、,、.、空格、\t、\n
* 常数： 数字 或 数字串.数字串
* 标识符： 字母或“\_”打头的由字母、数字串或“\_”组成的任意长度的符号串

2、扩展功能：

* 用mfc实现了图形化的效果
* 可以将待处理的源码从文件中读取和输入文本的保存
* 扩展整数到实常数，能够识别带小数点的实数
* 识别““双引号引起的字符串
* 识别如file.input中的“.“符号的识别
* 识别“>>” 符号的识别（++，--等类似，因此实现了一种）
* 对空格，换行符等进行处理
* **出错处理：**

1、对不在字符集中的非法字符进行报错，如“@”“$”，并跳过继续识别

2、处理实常数后跟字母的错误，如“123asd”

3、对出错位置进行定位，显示出行号和错误类别信息

1. **主要函数说明(很多函数为MFC模版函数不再赘述):**

void pret(char buffer[])

预处理输入内容 使用五状态的状态机去除注释(可识别/\* \*/和//两种形式)

去除注释时要注意换行 否则行号会发生混乱

void CByDlg::analyse(FILE \*fpin, FILE \*fpout)

词法分析主函数 先逐字符将文件中内容读入字符串数组 再逐字符进行分析

分析时最外层使用一个大while循环 当读到\0时退出 内层主要使用else-if判断依次进行词法规则的匹配

void CByDlg::OnOpen()

打开文件

void CByDlg::OnFenxi()

文件读取和处理

void CByDlg::OnSave()

文件保存

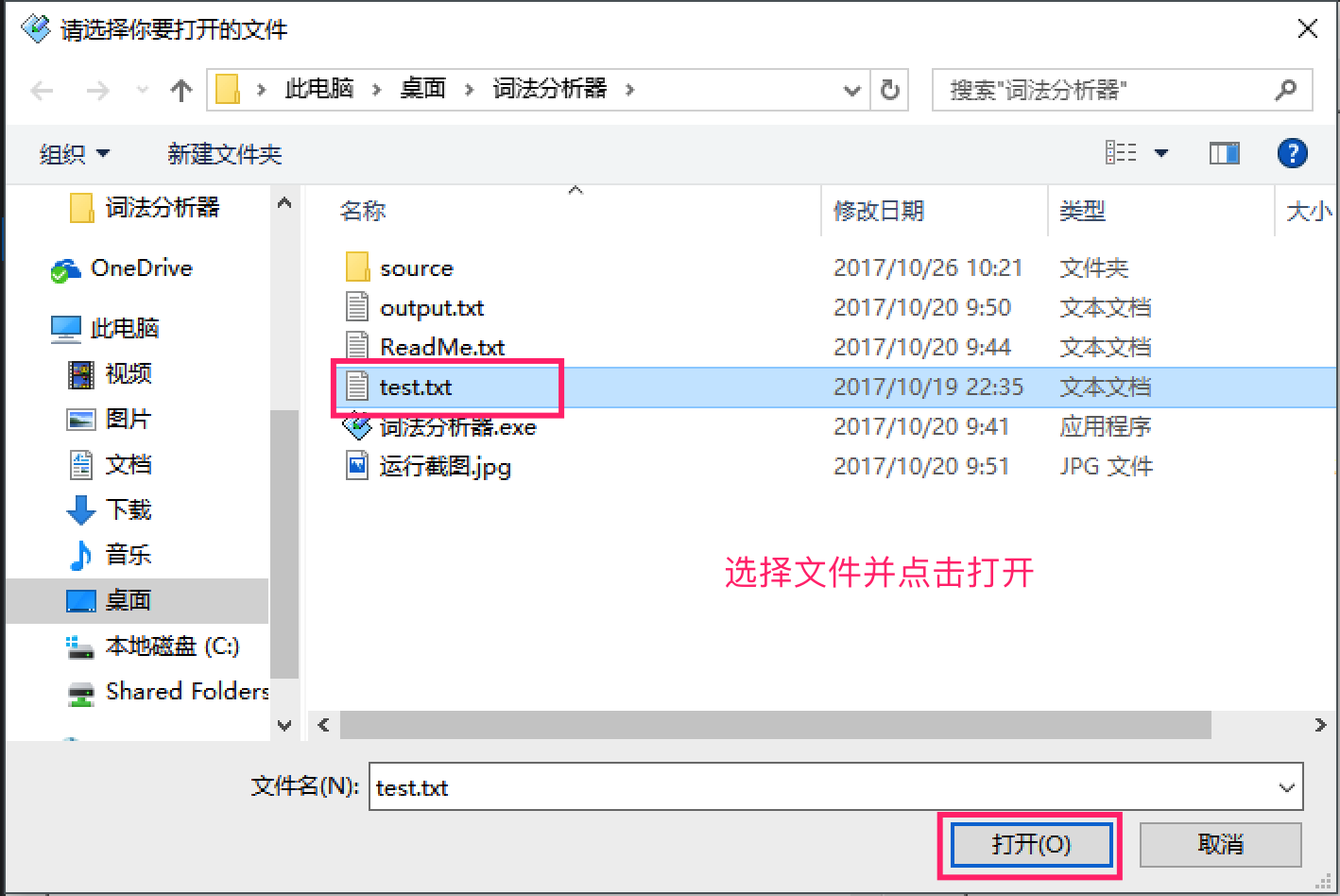
**4、词法分析器使用方法和结果展示**

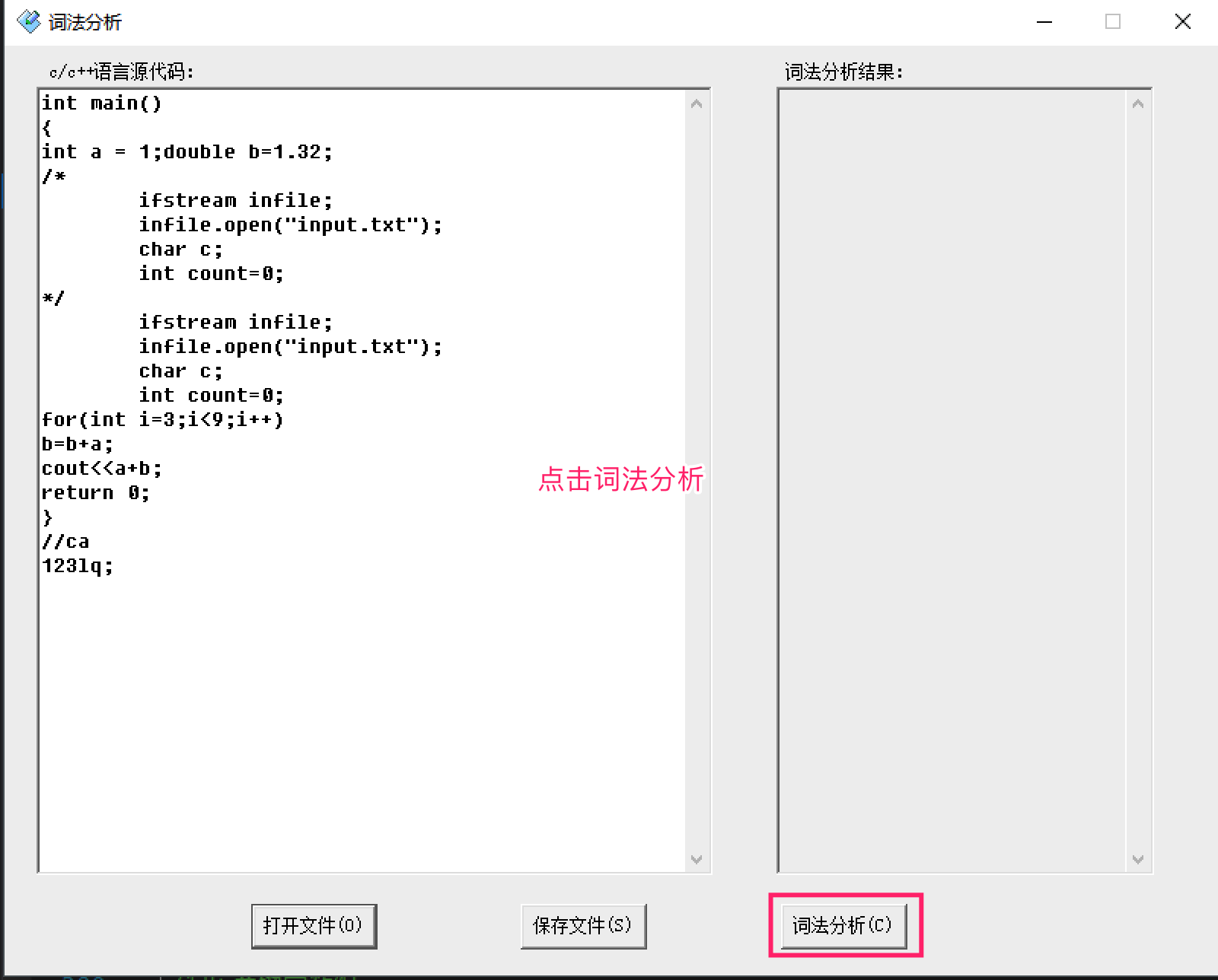
* **使用方法：**

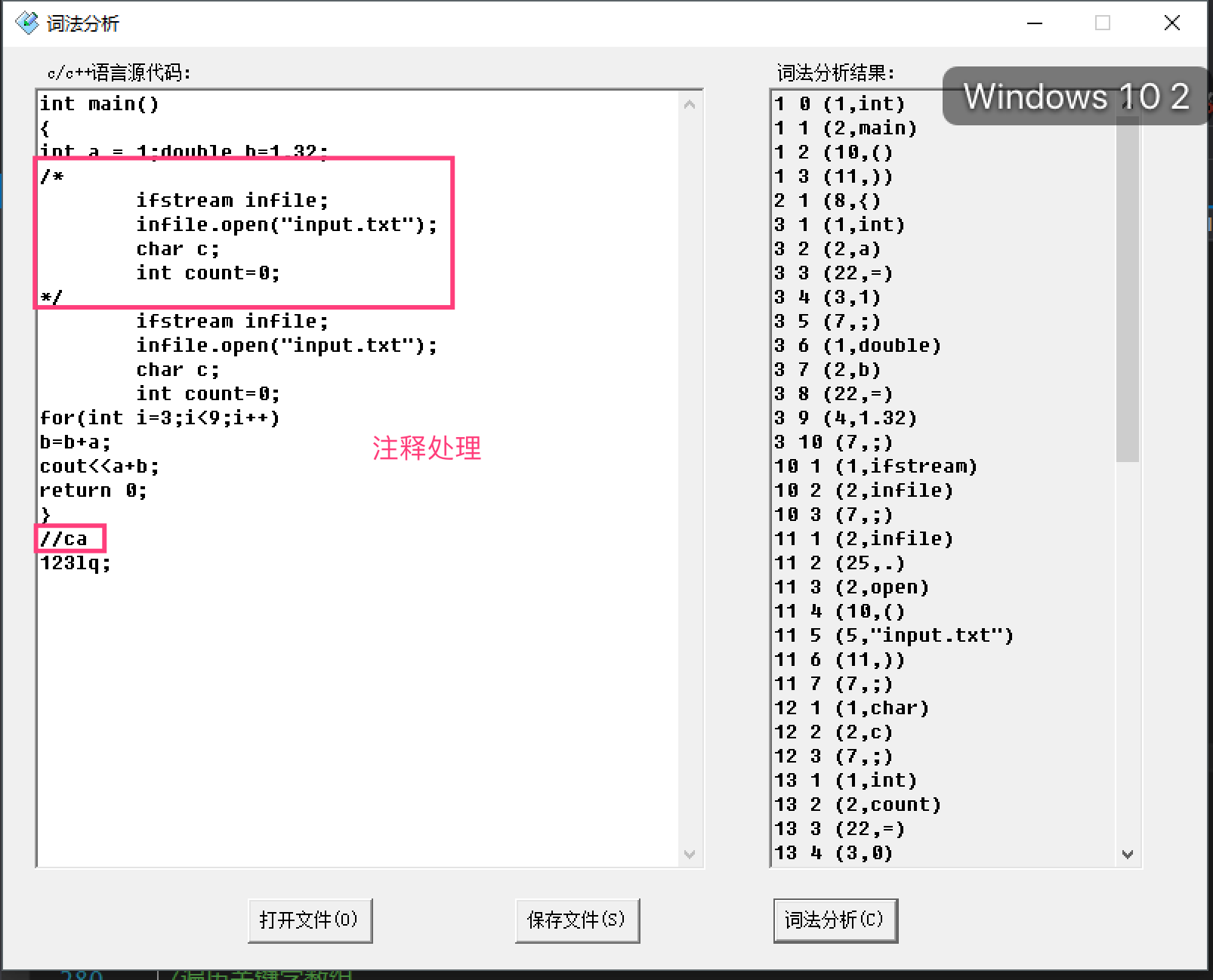
使用方式1:如果从文件中读取源码,点击"打开文件"按钮,选择电脑中储存的C/C++源代码文件,点击确定,即可将文件导入,点击"词法分析"即可实现词法分析.

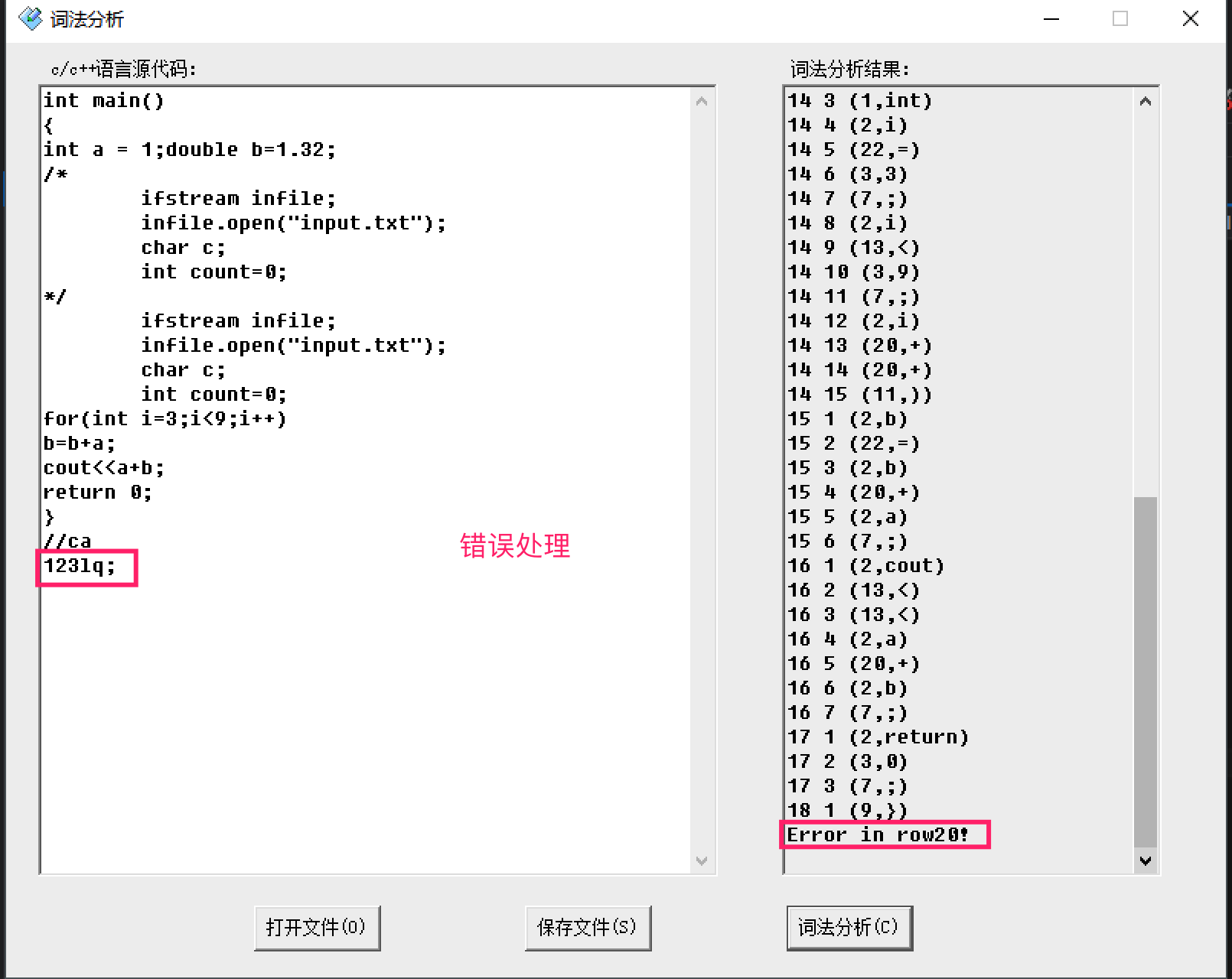
使用方式2:如果手动输入代码或复制代码进行分析,在源代码输入框中输入代码后,点击"保存文件"按钮,保存后点击"词法分析"即可实现词法分析.

* **程序结果展示：**





****

****

**bug提示:打开exe文件会弹窗遇到不适当的参数 点击确认后可以正常运行 查了一些资料也没有解决 暂时搁置在这里**