|  |
| --- |
| 北京邮电大学 |
| Python大作业报告 |
| 贪吃蛇+AI诗人 |
|  |
| **组内成员：王美淇、曹睿妍、李思婷、杨慧茜** |
| **2019/6/8** |

目录

[1. 选题描述 2](#_Toc515536003)

[1.1 选题一 2](#_Toc515536004)

[1.2 选题二 2](#_Toc515536005)

[2. 成员分组和任务分工 3](#_Toc515536006)

[2.1 成员分组 3](#_Toc515536007)

[2.2 任务分工 3](#_Toc515536008)

[3. 设计思路 3](#_Toc515536009)

[4. 程序运行效果 4](#_Toc515536010)

[4.1 4](#_Toc515536011)

[4.2 4](#_Toc515536012)

[5. 遇到的问题和解决方法 4](#_Toc515536013)

[5.1 4](#_Toc515536014)

[5.2 4](#_Toc515536015)

## 选题描述

### 选题一 用python编写实现贪吃蛇程序

用python实现贪吃蛇游戏，通过上下左右键控制贪吃蛇移动吃东西，累计积分，碰壁或头尾相碰则game over

### 1.2 选题二 AI诗人

编写一段能够自动写诗词（题材为唐诗、宋词、元曲，现代诗或者歌词）的Python程序。该程序采用完全随机的算法，即不输入诗句，但输入限制性约束，如诗歌体裁（四言诗、五言诗、七言诗），诗歌类型（普通诗、藏头诗）来随机生成平仄相对的诗词。

## 成员分组和任务分工

### 成员分组

组内成员：王美淇 17班2018210473 李思婷 16班2018210450

曹睿妍 17班2018213325 杨慧茜 17班 2018210471

### 任务分工

**选题一 贪吃蛇：**

王美淇：负责图形界面的部分设计和贪吃蛇游戏功能的实现及部分实验报告

李思婷：负责调试及bug的解决及部分实验报告

曹睿妍：负责图形界面的部分设计及部分实验报告

杨慧茜：负责测试调试贪吃蛇移动及部分实验报告

**选题二 AI诗人:**

曹睿妍：负责机器学习的爬取、学习、检验、藏头代码及部分实验报告

杨慧茜：负责自动成诗的诗源爬取及部分实验报告

王美淇：负责自动成诗平仄、四言诗、五言诗及部分实验报告

李思婷：负责七言诗、藏头诗及部分实验报告

## 设计思路

### 3.1贪吃蛇设计思路

Python游戏设计，应用pygame游戏库

定义颜色变量,速度参量，定义画布宽高和上下左右操作

定义main函数（初始化pygame）

定义开始游戏界面

定义游戏结束页面

定义rungame函数：初始化贪吃蛇的起始坐标（随机），初始化长度，初始化目标的位置（随机），上下左右移动在代码上体现为x坐标和y坐标的增减，判断是否吃掉目标，score+，判断是否碰壁或首尾相碰，若则跳出含苏进入gameover页面

### 3.1AI诗人设计思路

一．数据分析

（1）名词（以名词为例）



首先，将诗词中常出现的名词整理到一个txt文本中，然后对诗词中常出现的名词进行分析，并通过词云的形式展示。我们可以看到，“万里”，“故人”，“烟霞”等意象是古诗词中最常出现的名词。这表明了古人写诗多以感慨世事，怀念过往经历与人事为主。

再如（2）动词（3）唐诗（4）宋词（5）诗经（不一一举例）

二．诗词生成程序的算法比较

1. 算法与原则

（1）诗词数据的文本格式处理



通过正则表达式，将所抓数据的卷名，题目，作者删除，只保留诗句部分，诗句中间用逗号或者句号隔开。也可以改变参数选取不同的保留内容，从而对诗歌进行多方位的分析。

（2）诗词数据的词性分析

提取高频名词

通过jieba，对名词、动词、形容词等词进行切片分析，以过滤掉

没用的部分，从而进行文本的分析。通过改掉不同的词性参数，来从

不同方面分析诗词。

（3）根据词频随机作诗

通过random，pinyin并自定义函数的方式，对（2）所提取

的不同词性的词进行分析进而随机作诗。并通过自定义函数的方式，

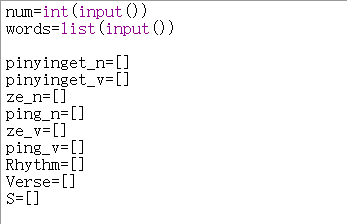
让诗句保持一定的对仗形式（如：名名动名名，名名动动名）。通过

改变爬取文本的txt，来使所做的诗更具随机性与诗歌化。

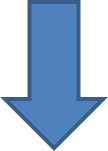
诗词数据读音分析

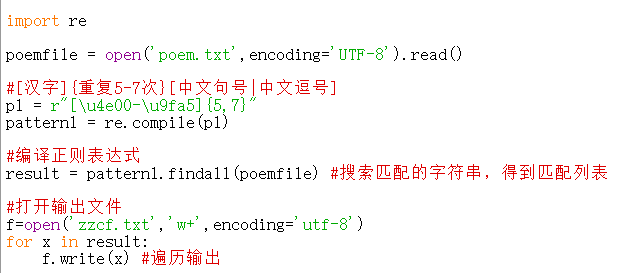
通过pinyin模块对诗词数据的读音进行分析，并由一声二声与三声四声分成平仄两类，对随机诗句进行格式上的规整。同时，pinyin模块可以对诗句最后一字的韵母进行规范，进而控制住韵脚，使随机生成的诗句更具诗歌色彩，读来更朗朗上口。

三．流程图

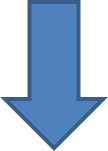


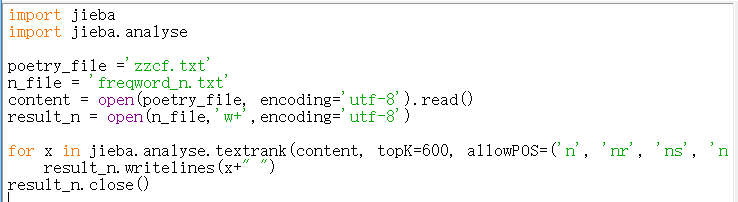
（输入有两行，第一行输入4或5或7来规定随机生成的诗句的言数；第二行输入四个字的词或直接回车来约束藏头诗和普通诗的输出。）

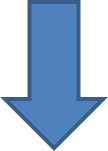




（诗词数据的文本格式处理）

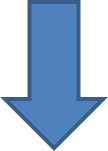


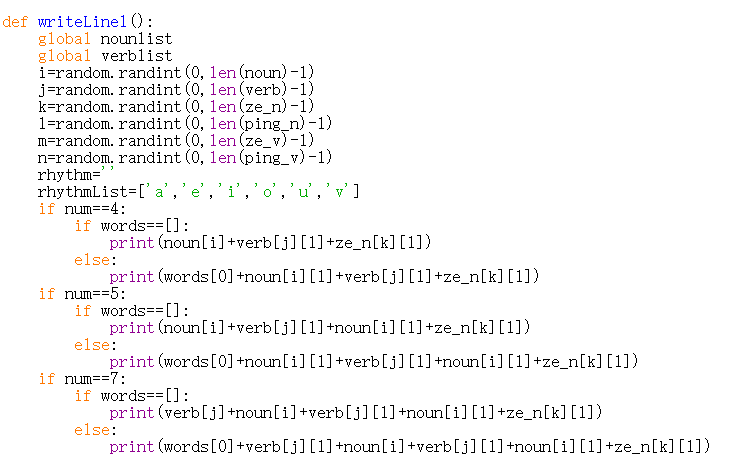
（jieba提取高频名词）



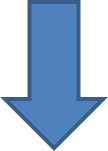


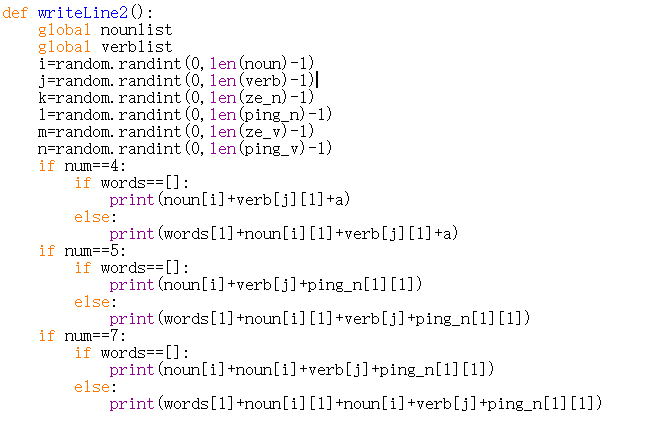
（把jiaba按词性不同切成的不同文本进行按平仄分类，分为名词平音，名词仄音，动词平音，动词仄音，为下面平仄诗做准备。）

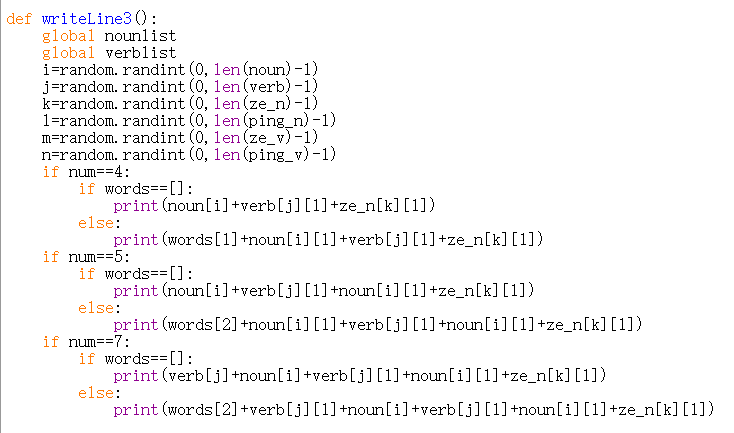


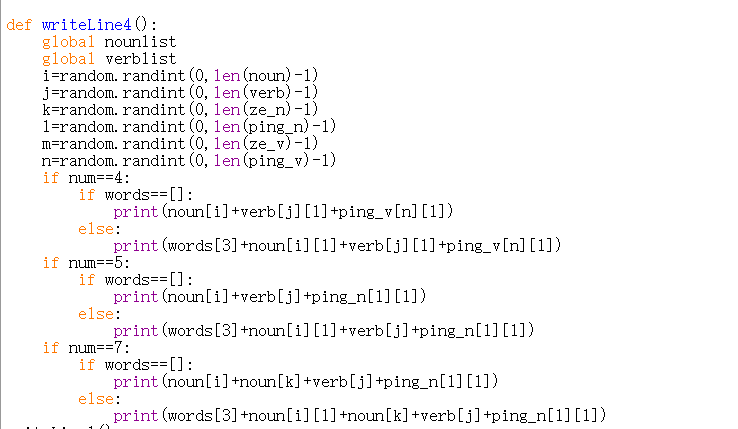


（以第一句为例，通过定义函数，和设置随机变量，以“名名动动名”和“名名动名名”为对仗格式，将不同词性的词随机输出，并以4，5，7为四言诗，五言诗，七言诗的约束）

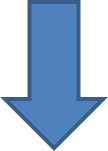


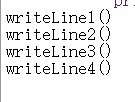




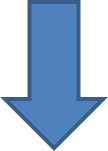


（此为第二句，第三句，第四句的代码，同第一句几乎相同。）





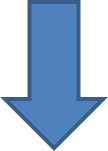
（输出约束的随机生成诗句，程序结束。）

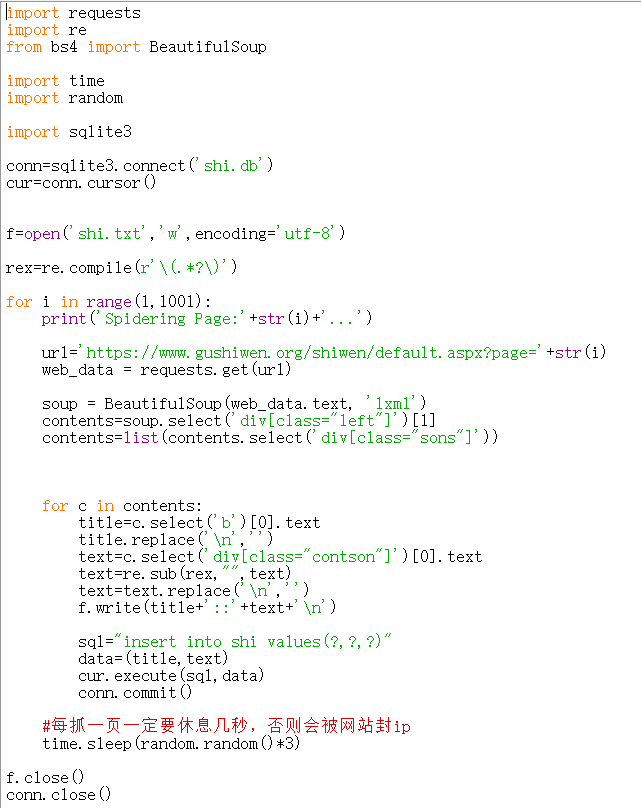




（4）机器学习

（首先先从github上下载tensorflow等与机器学习有关的程序及数据，并在cmd中创造python3.6及以下的虚拟环境{因为我的python本身就是3.6.8版本，所以没有设置虚拟环境}，其次在cmd中安装Beautiful Soup等使用工具）

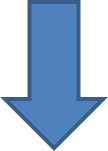


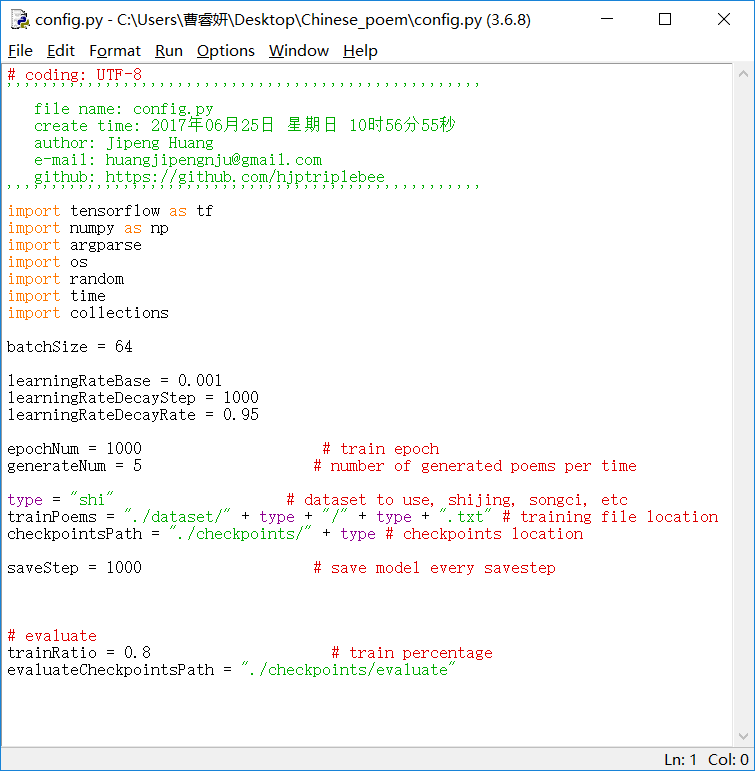


（然后利用spider.py代码，对<https://www.gushiwen.org/shiwen/default.aspx?page=>上的诗歌进行抓取，爬虫。并将爬取数据储存在txt中（shi.txt））

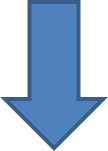


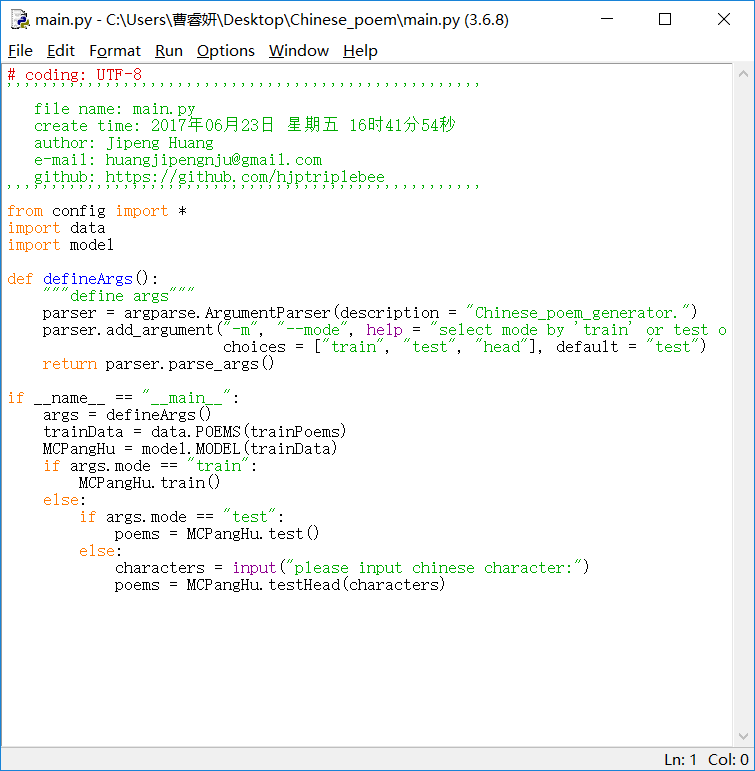
（此为爬取内容）



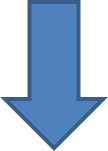


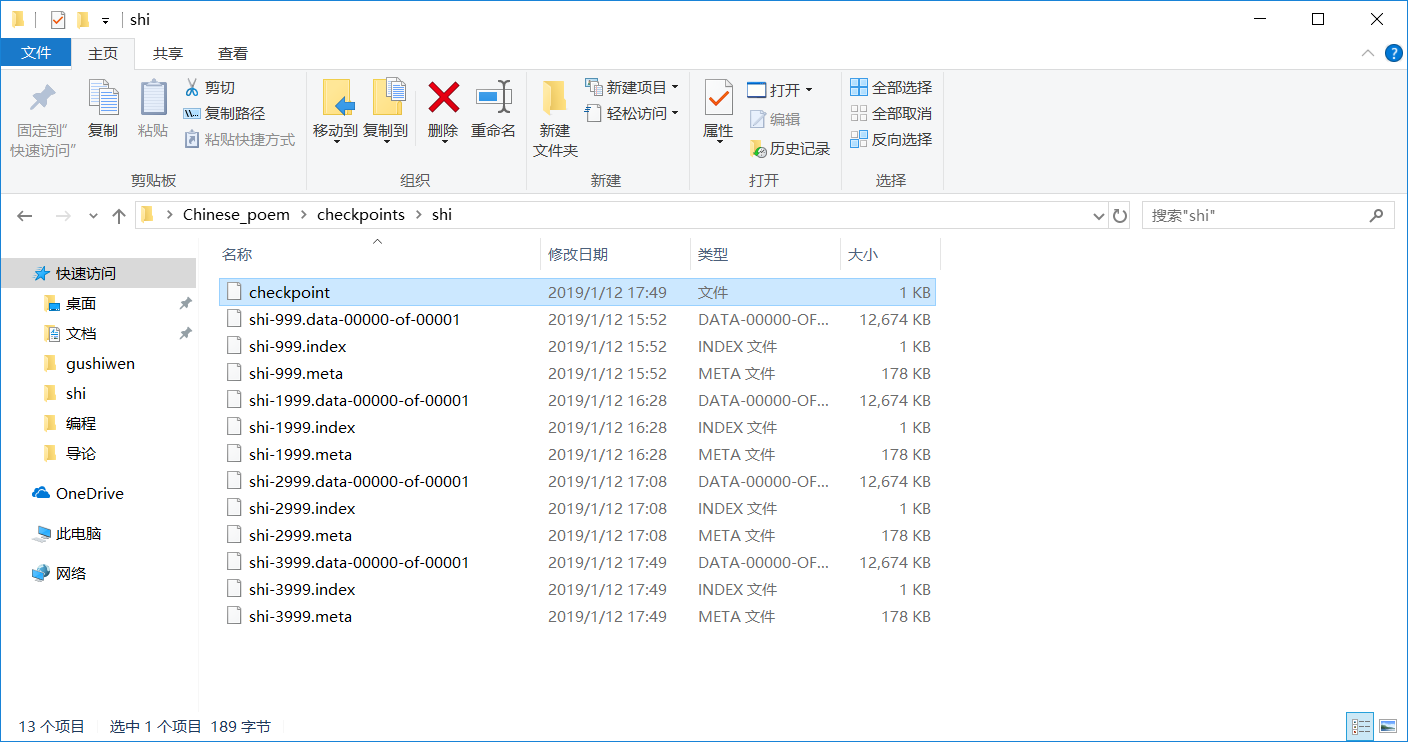
（在config.py上对所爬数据进行修改（将poerySong改为shi）即可运行训练不同的数据，使机器学习更客观可靠。）



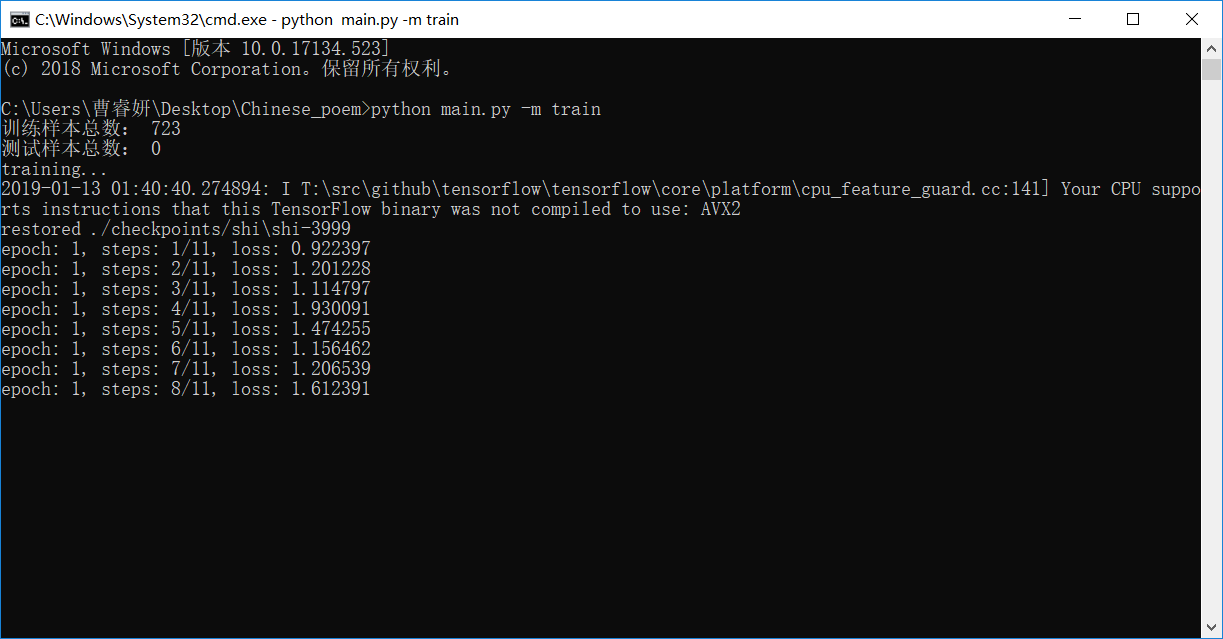


（用main.py对所爬内容进行训练已达到test与head的效果。并将所学存储在checkpoint.doc中。）





（图为训练结果）

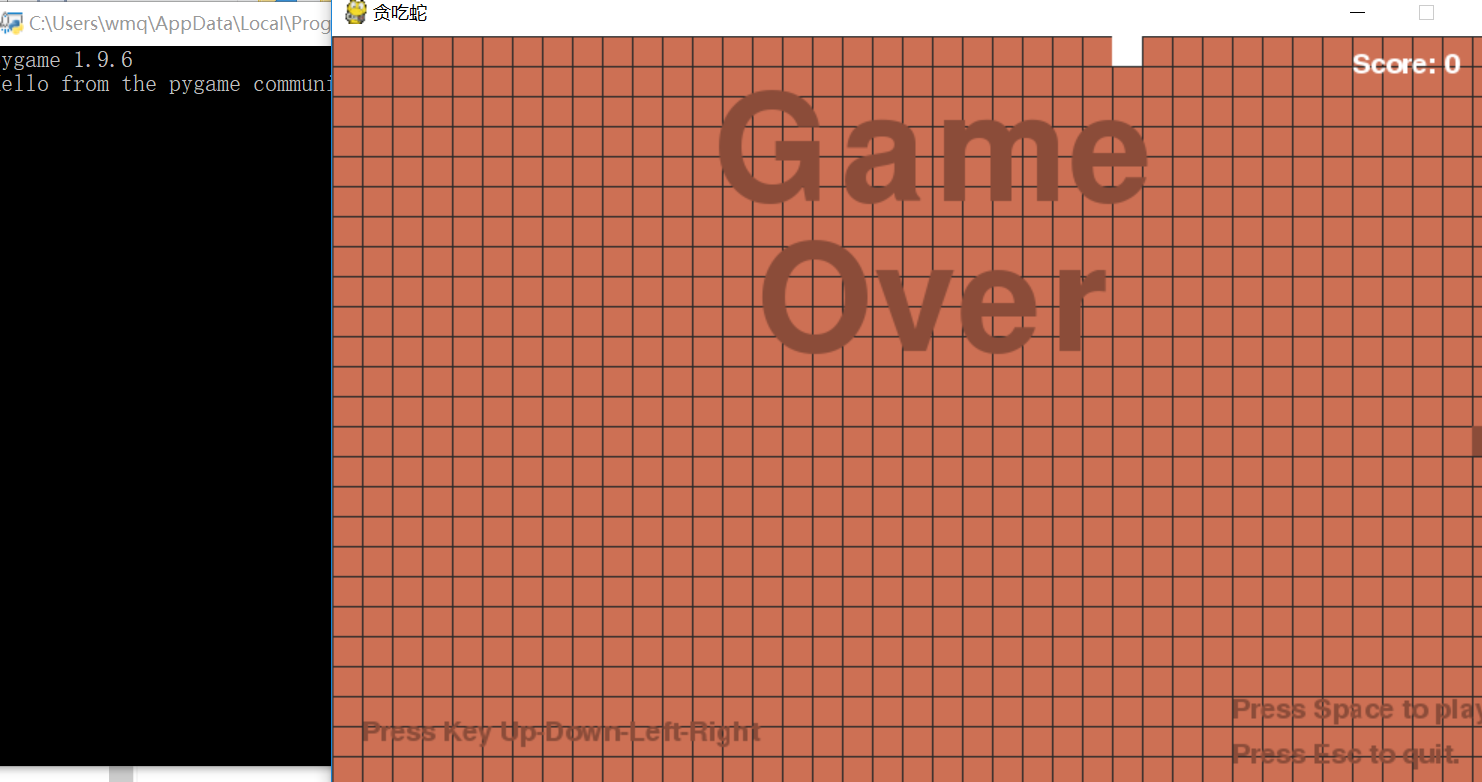
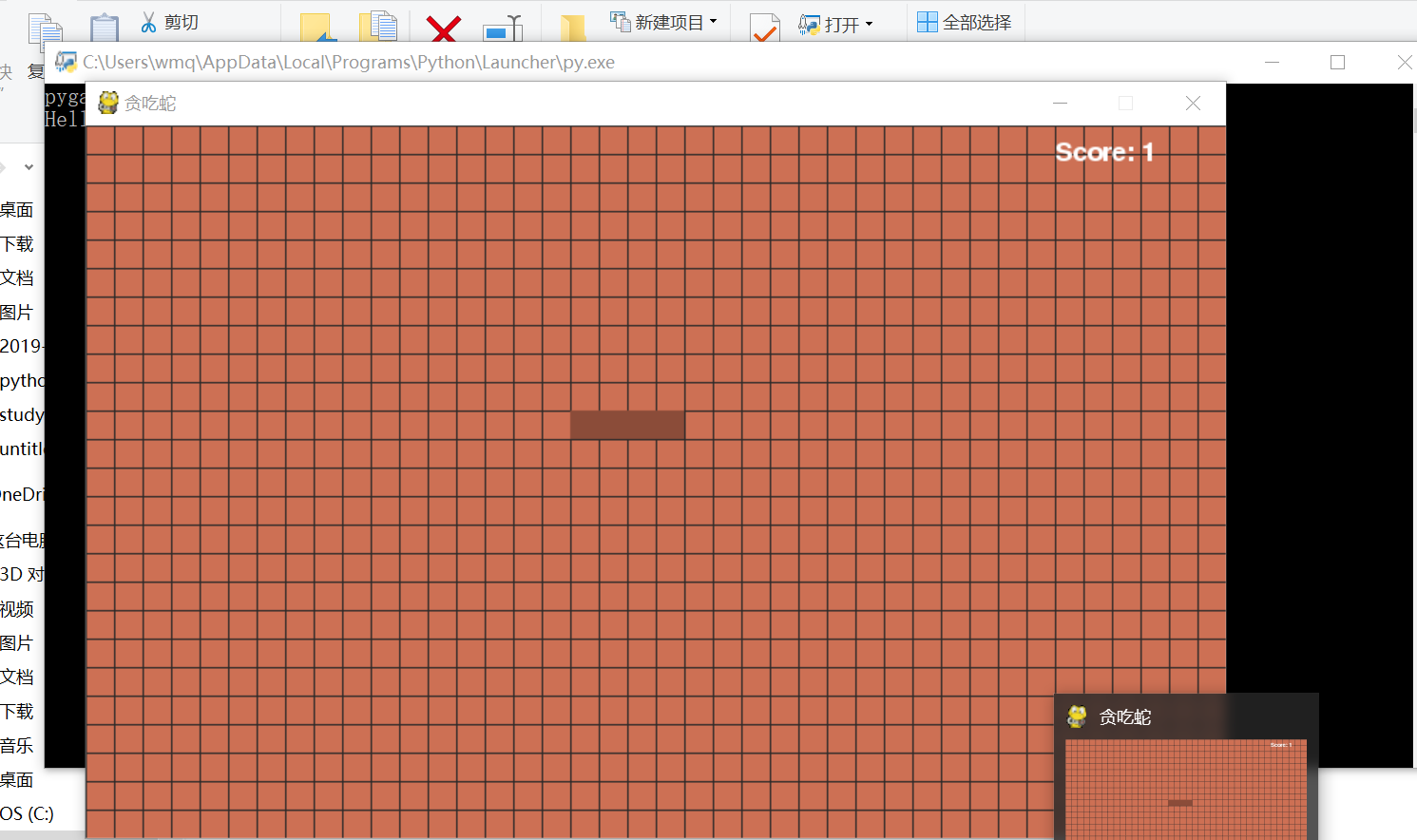


（此为训练过程）

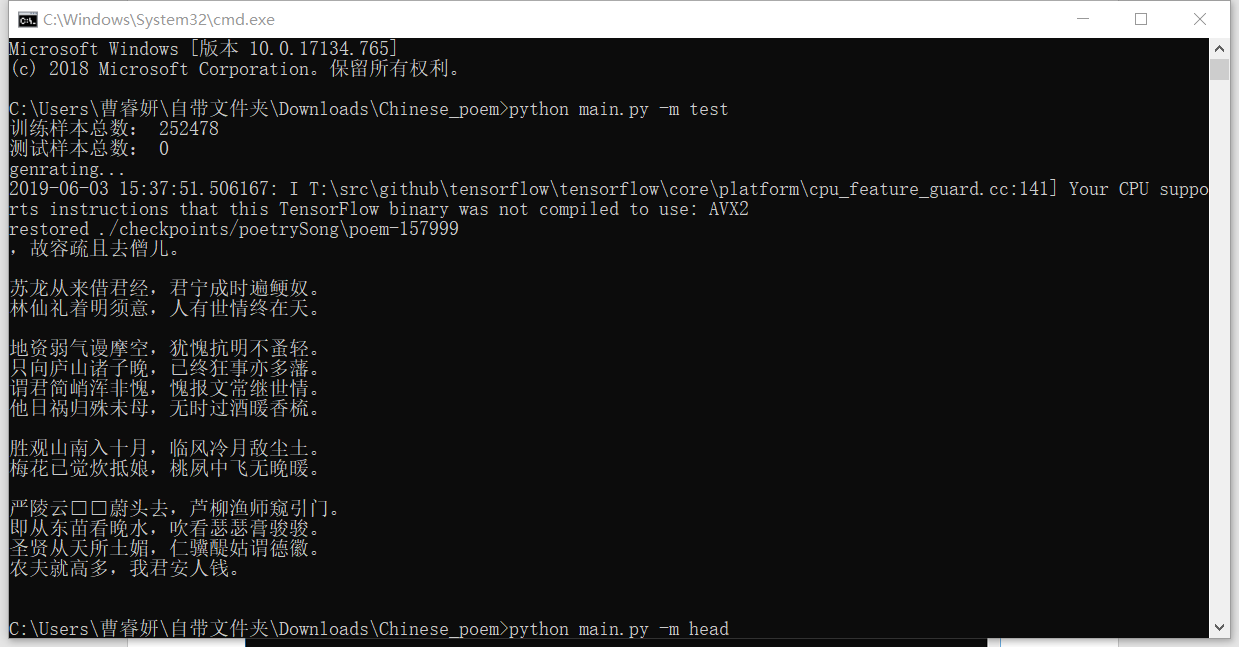
## 程序运行效果

### 4.1 贪吃蛇运行效果

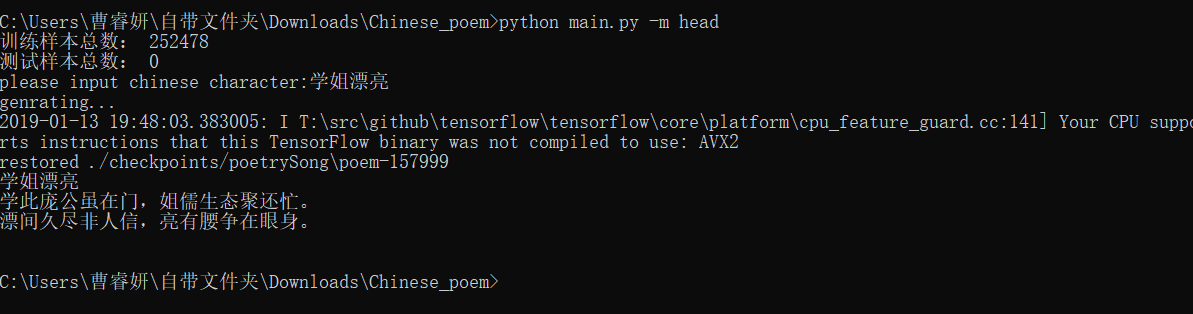


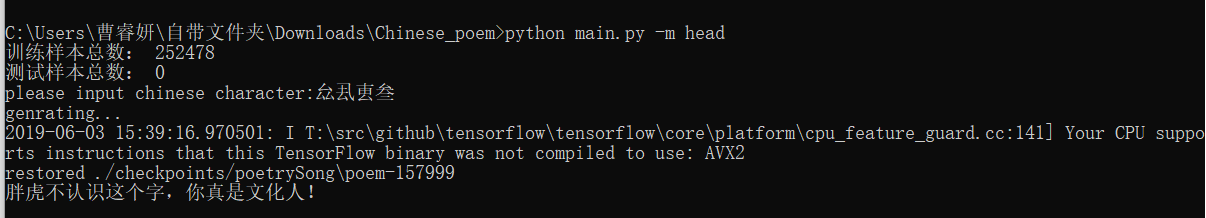


### 4.2 AI诗人 运行效果



（此为测试结果）





（此为藏头诗的结果）

## 遇到的问题和解决方法

### 贪吃蛇

1. pygame库的调用：需要pip安装pygame库然后在pycharm中
2. 中文字体：用系统pygame.font.SysFont(‘SimHei’)
3. 贪吃蛇移动：采用xy坐标加一的方法
4. Score计分出错：开始采用吃掉目标加1的方式，后改为蛇的长度-3，更好实现
5. 判别是否gameover：

if wormCoords[HEAD]['x'] == -1 or wormCoords[HEAD]['x'] == Cell\_W or wormCoords[HEAD]['y'] == -1 or \

wormCoords[HEAD]['y'] == Cell\_H:

return # game over

1. updownleftright键的设置：

if (event.key == pygame.K\_LEFT) and direction != RIGHT:

direction = LEFT

### AI诗人

1.爬虫时显示电脑中未装python;

解决办法：改变IDLE的环境变量

1. 使用pinyin模块时程序总报pip版本过低

解决办法：将IDLE降为3.7一下版本

1. 只用爬好的词语txt很难区分平仄

解决办法：在代码里建立4个列表将分好的词语再细分

1. 机器学习爬的诗数目较大，有时很难爬干净

解决办法：分组爬取，并通过txt删除功能再次过滤删选一遍