浙江水利水电学院

实验报告

（ 2018-2019 学年 2 学期）

|  |  |
| --- | --- |
| 课 程 名 称： | Python程序设计 |
| 班 级： | 软件工程S18-1 |
| 学 号： | 2018b31007/   2018b31006/2018b31003 |
| 姓 名： | 曹愉/祝潇佟/周应雷 |
| 实验室（中心）名 称： | 软件工程实验室 |
| 教学单位： | 信息工程与艺术设计学院 |

2019年 3 月 21 日

实 验 名 称：使用Python分析文本情感 指导教师：张鹏亮

实 验 日 期：2019.3.21 地 点：现南411

同组学生姓名：曹愉、祝潇佟、周应雷

|  |
| --- |
| 实验内容及要求：  **1.*实验内容：***  本次实验做简单的中文文本情感分析，即分析一段文本表达的正面（褒义）还是负面（贬义）情绪。  本次任务分为两阶段。  第一阶段为预处理阶段，是必做任务。主要训练Python基本语法、函数、组合数据类型的使用。  第二阶段为词性统计阶段。本阶段选做。主要训练开放性思维。  详细内容参见《实验二指导书》  **2.*实验要求：***  （1）完成各阶段任务，完成各阶段算法设计。  （2）分析算法（代码）的可靠性和准确性 |
| 主要仪器名称及型号：   1. 计算机：Windows，i5 3Ghz,4GB内存   2. 开发工具：Python |
| 实验过程：（可附页）   1. 通过open函数读取文件，并将文本中的语句存入textlist，   同时将中文褒贬义词典中定义的情感词分别存入polist与nelist中。  源代码：  fo = open("Text.txt", "r")  textlist = fo.readlines()  duanlist = []  for duan in textlist: # 依次读取每行  duan = duan.strip() # 去掉每行头尾空白  duanlist.append(duan)  positive = open(  r".\sentiment.dict.v1.0\tsinghua.positive.gb.txt", "r").readlines()  negative = open(  r".\sentiment.dict.v1.0\tsinghua.negative.gb.txt", "r").readlines()  polist = []  nelist = []  for word in positive:  word = word.strip()  polist.append(word)  for word in negative:  word = word.strip()  nelist.append(word)  fo.close() # 关闭文件   1. 实现第一阶段任务，利用split方法对标点进行语句的分割处理。   导入jieba库对语句进行中文词语分割并返回wordslist以供统计情感度。  源代码：  def split(sentence): # 传入句子，返回分词列表wordslist  pattern = '。|，|？'  wordslist = []  sentence = sentence.strip()  juzi\_list = re.split(pattern, sentence)  while '' in juzi\_list:  juzi\_list.remove('')  for juzi in juzi\_list:  wordslist = wordslist+jieba.lcut(juzi)  return wordslist   1. 实现第二阶段任务定义褒义度、贬义度变量p、n。   根据语句内褒义词贬义词出现的词频进行统计，同时根据词语下标实现对情感词语前程度词和否定词的判断，从而对情感度进行不同的赋值加减。  统计语句中的情绪符号如感叹号，从而对整句的情感度进行增减。  源代码：  def count(wordslist): # 传入分词列表，返回褒义值p和贬义值n  p, n = 0, 0  chengdu = ['很', '非常', '极度', '无比', '太']  fouding = ['不', '没有']  for index in range(len(wordslist)):  for poword in polist:  if wordslist[index] == poword:  # 否定反转  if fouding.count(wordslist[index-1]) != 0:  # 否定反转加程度  if chengdu.count(wordslist[index-2]) != 0:  n = n+2  else:  n = n+1  # 程度判断  elif chengdu.count(wordslist[index-1]) != 0:  p = p+2  else:  p = p+1  for neword in nelist:  if wordslist[index] == neword:  if fouding.count(wordslist[index-1]) != 0:  # 否定反转  if chengdu.count(wordslist[index-2]) != 0:  # 否定反转加程度  p = p+2  else:  p = p+1  # 程度判断  elif chengdu.count(wordslist[index-1]) != 0:  n = n+2  else:  n = n+1  # 感叹语气  t = wordslist.count("！")+1  p = p\*t  n = n\*t  return p, n   1. 将文本文件中的语句进行分段，通过遍历list实现分句输出语句，并显示每一句中的褒义度和贬义度。   源代码：  for sentence in duanlist:  wordslist = split(sentence)  p, n = count(wordslist)  while '！' in wordslist:  wordslist.remove('！')  print('{}\n褒义度：{} 贬义度：{}'.format(sentence, p, n)) |
| 实验数据记录或图片：（可附页）  实验第一阶段分词结果：    实验第二阶段统计语句情感度结果： |
| 实验结论：  曹愉 2018b31007  这次实验通过使用python实现了文件读入和语句分词以及情感词统计等实验目标。在实现的过程中使用了python第三方的库，这次实验的成果对语句中的不同情感词以及他们的程度和否定词都有了精确的判别，同时通过词频统计对句子情感也进行了准确的判断。总的来说很好的实现了这次试验所需的功能。但是因为算法有些单一，褒义和贬义词库也并不完善，需要在以后的改进中完善语义分析以及扩大词库的统计范围。  这次实验过程中也出现了几次问题，最开始的文件读入的时候因为回车换行以及部分空白的文本出现，因此需要使用strip()函数去掉头尾的空白，这样对之后的分词就不会造成混乱的影响。  之后的词频统计的时候发现有语句中会出现重复的词语，因此不能通过index()方法来获取词语下标来遍历词语list，最后还有一部分是对复杂的形容词进行分析，最后通过分析解决了类似“很不快乐”这种既有程度也有否定词的双重形容句子的情感分析。  这次实验项目的完成让我更加熟悉了python的编程操作，从一开始的慢慢摸索到后面的循序渐进体会了很多也收获了很多，慢慢也想出了一些问题的不同解决方法。在不断地查bug和细节修改中我也发现了一些细小的问题，所以以后在编写代码的时候也要注意这些细节，避免这些小问题的重复出现。  最后，这次实验成功地完成了需要解决的问题，实现了问题预先设计的各项功能。在以后的学习与工作中我们也要借鉴这次实验中带给我们的经验和开发过程，努力完善每一个代码细节，通过学习来充实自己的知识储备，用实践来发挥和磨砺自身的技术实力。  祝潇佟 2018b31006  这次的实验是做一个简单的中文文本情感分析。分两个阶段完成，第一阶段在课内完成，首先将txt文本导入，并将文中字典里定义的情感词分别用for 循环分别存入两个list，之后定义的这个split方法参考、询问了同学之后才弄懂怎么操作，后来又经同学指导添加了strip()函数用于去掉头尾的空白，使其整齐有序。  第一阶段花了大部分的课上时间才基本弄懂并完成，第二阶段的词性统计对于我来说显得尤为复杂，首先理解老师指导书中的内容，我认为应该用多重循环来实现，先定义一个褒义值和贬义值得初始量，再利用循环来进行程度的判断，以表达出对情感度的量化，但是自己写的代码还是很有问题，不清楚该怎么写for语句，上网查找了相关例子并经过同学的补充还是能基本完成。  这次实验使我对Python的理解与如何构思编写更加深刻，了解到Python的第三方库的实用性。通过本次实验发现还是不怎么熟悉这门语言，只能编写一些简单的程序，之后的实验还是需要更加地努力。  周应雷 2018b31003  这次实验有一定的难度，刚刚接触到Python，书本上的知识点还不足以完成实验。  第一阶段是进行分句和中文分词，课本上对于文件的输入输出部分讲的比较少，前期出现了很多问题，不会去空格，输出出现表示换行的“\n”等等。通过查阅了一些资料，才慢慢把一些问题解决了。  第二阶段需要做词性统计，这一部分需要导入词典，过程更加复杂一些。最终还是完成了，虽然最后的效果只是统计了褒义词和褒义词出现的次数，离机器学习的情感分析还是差了很多。不过也是初步了解到了机器学习的一些基本方法。  指导书的最后老师提到了自然语言分析（NLP）中的情感分析（sentiment）技术，需要牵涉到一些机器学习和数据挖掘的算法，如朴素贝叶斯，MaxEnt分类器，SVM，神经网络（深度学习）等等。这些内容在接下来的大数据基础课程中会涉及到一些。机器学习需要一些数学方面的基础，之前大数据基础课程中，学习了线性回归和逻辑回归，其中涉及到高等数学中的构建函数，和涉及到线性代数的矩阵，我还是能理解的，但是一些涉及到概率论与数理统计中的内容，由于没有学过，所以不太看得懂，接下来有时间还是需要粗略地看一下。 |
| 教师评语：（学生请勿填写和更改）  完成情况：优  设计分析能力：优  拓展能力：优  成绩：选择一项。 批阅教师：张鹏亮  2019年4月6日 |