

实验五 PYTHON 外部库应用和开发

一、目的和要求

1. 熟悉 Python 外部库应用与开发;
2. 掌握 Python 语言基本语法;

二、实验环境

1. Win 7 操作系统;
2. Python 2.7.X, IDLE、PyCharm 等开发环境;

三、实验内容

(一) 验证实验 (每个同学完成)

1. Matplotlib 验证: 为实验三中第 6 小题绘制 P 随 M 变化的曲线;
2. GUI 验证: 为实验四中第 6 小题设计和编制 GUI;
3. 文件读写验证: 运行调试第七章各小节例示代码;
4. 异常处理验证: 运行调试第八章各小节例示代码。

(二) 设计实验 (二选一, 小组完成, 代码提交, 算法测试写入实验报告)

5. 数据分析综合应用:

(1) 以文本文件格式读入文件夹\dataanalysis\label\下的 MTL_*.dat, CMTL_*.dat, CEMTL_*.dat(*表示 White 或者 Male, 选择其中一种处理即可)中数据, 并且分别读入 numpy 数组 MTLLabel、CMTLLabel 或者 CEMTLLabel 中, 对各个数组取绝对值后按照降序排序, 并且记录数据元素排序前的下标号;

(2) 以文本文件格式读入文件夹\dataanalysis\train\下的 MTL_*_train.dat(*表示 White 或者 Male, 选择其中一种处理即可)中的数据, 并且读入 numpy 矩阵 TrainSample 中, 计算矩阵的行列数 (该矩阵包含了 1000 个维数为 3304 的样本, 1000 个样本分属两类, 每类含 500 个样本顺序保存在文件中), 分别根据 (1) 中三个数组的排序结果, 从 TrainSample 中选择 1000 个样本对应的维度子集 (可以选择 200, 400, 600, 800, 1000 维), 保存到新的矩阵中 TrainSub;

(3) 对于\dataanalysis\test\下文件作和(2)相同的处理(其中数据矩阵包含了800个维数为3304的样本,800个样本分属两类,每类含400个样本顺序保存在文件中);

(4) 阅读和学习\knnexample\下面关于最近邻分类算法 Knn 的实现,用(2)中数据训练分类模型,用(3)中数据测试分类结果,统计错误率。

6. 文本分析综合应用:

(1) 编写模块实现中文文本中给定字或词的频率统计功能;

(2) 运用(1)中功能模块分析文件“dreamofredmaison.txt”中前80回和后40回中常见文言虚实词的词频,分析结果存入文本文件;

(3) 采用 Matplotlib 可视化(2)中的分析结果;

(4) 运用 GUI 编制用户界面,为用户提供选择文言虚实字词的交互界面,按照用户选择采用(1)中功能实现频率统计,并且把(3)中实现的分析结果动态呈现给用户。

(三) 综合实验(个人完成,选做,验收,代码提交)

7. 综合实验自由选题:

以 Python 开发 1-2 个外部库为主要工具,自由选择开发任务,设计和完成算法、应用程序或系统。

例如:实验5第(4)步换成 SVM 分类器(可以从 <http://www.csie.ntu.edu.tw/~cljlin/libsvm> 获取 SVM 的 python 实现和样例)使用可作为本题的一种完成形式。