

K-均值算法应用程序说明书

一、 程序

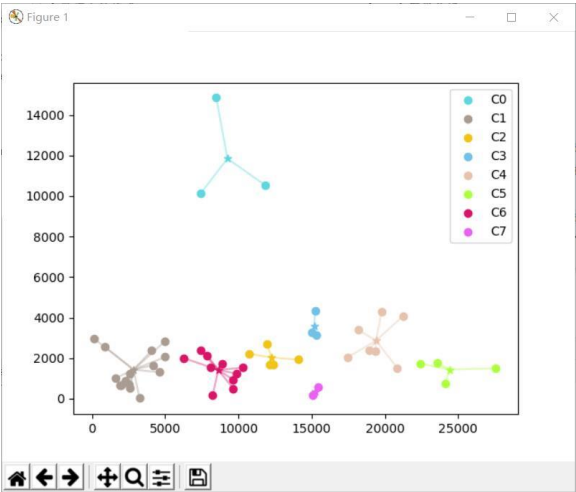
该程序为使用 Python 语言编写的 k-均值聚类算法的应用程序。

二、 使用说明

程序界面如下：



点击“导入”，导入数据路径输入框中对应的数据文件；点击“示例数据”，导入 fkmeansData.txt，其中有我准备好的一组数据，导入之后的数据显示在左侧框中，其格式为<标号: [属性 0, 属性 1, ...]>。迭代次数下面的滑动框可以调整迭代次数，范围为 1~1000，距离度量下拉框可以选择几何距离或者棋盘距离。在“聚类分析”按钮右侧输入 k 值，即簇的个数之后，点击“聚类分析”进行 k-均值聚类。聚类结果显示在右侧框中。在“二维可视化”按钮右侧输入要观察的两个维度（维度范围与数据维度相匹配）之后，点击“二维可视化”，可以观察到如下图像：



显示各个簇和各个簇中心点，并将簇中点和各个中心点连接起来。

### 三、 数据格式

输入数据需要满足的格式为（以下为示例数据前四项）：

---

4583, 268.93, 1305.55

853.4, 402.15, 2580.4

15230, 682.3, 4343.1

23600.0, 346.0, 1747.8

---

其中每一行是一个样本（即样本以'\r\n'分隔），由","分隔不同属性的值，文件中“fkmeansData.txt”是示例数据，可以不用自己准备数据直接运行。

### 四、 编译环境

使用的 Python 库有：

---

Python 标准库	os,threading,tkinter,math,random
------------	----------------------------------

第三方库	matplotlib,numpy
------	------------------

---

Python 编译器版本为 3.7 32 位。

如果已经装好了 Python,可以在命令行中输入以下语句安装外部库：

---

pip install numpy

pip install matplotlib

---

运行以下语句运行源代码：

---

cd [源代码路径]

python FinalKmeans.py

---

可执行 exe 程序由 pyinstaller 生成,将所有的依赖的库都打包在了一起，所以文件非常大,因此打开它也要花一段时间，但是打开之后运行还算流畅。