

模糊数学作业

244号队伍：陈逸飞 陈培林 王奕琳

题目描述

作业附件中给出了3类图片，请完成如下任务：

- (1) 构建指标，建立数据集中任意图片的特征样本；
- (2) 在(1)基础上进行任意两张图片的相似度度量，形成模糊相似关系；
- (3) 在(2)基础上利用传递闭包法，对所有图片进行聚类，并利用已有标签分类进行验证，给出聚类评估的相关指标。

解答

(1) 构建指标，建立数据集中任意图片的特征样本

一张.jpg图片实质是由许多个像素点组成的，每个像素点包含的信息有R、G、B三原色的数值，因此，一张 $m \times n$ 大小的.jpg图片的全部像素点信息就可以使用 `imread` 读入到Matlab。每张图片读入进来是一个 $m \times n \times 3$ 的三维矩阵。

我们需要把这个三维矩阵中的第三维消去，即将RGB三个数值综合为一个数值，我们这里简单采用三值相加：

$$\text{颜色值} = R + G + B$$

这样每张图片就是一个 $m \times n$ 的矩阵了，矩阵中每个元素的值就是这个像素点对应的颜色值。

然后对原始数据进行标准化，我们使用**平移极差变换**进行标准化：

$$x'_{ijk} = \frac{x_{ijk} - \min\{x_{ijk}\}}{\max\{x_{ijk}\} - \min\{x_{ijk}\}}$$

下标中的(i,j)代表像素点的坐标，下标中的k代表是第k张图的像素点

(2) 在(1)基础上进行任意两张图片的相似度度量，形成模糊相似关系

我们采用夹角余弦法对一个分类中的所有图片两两计算相似度，从而计算出一个含有各图片之间相似度的方阵

夹角余弦法公式：

$$FSM_{ij} = \frac{\sum_{p=1}^m \sum_{q=1}^n x_{pqi} \cdot x_{pqj}}{\sqrt{\sum_{p=1}^m \sum_{q=1}^n x_{pqi}^2} \sqrt{\sum_{p=1}^m \sum_{q=1}^n x_{pqj}^2}}$$

这样我们就得到了模糊相似矩阵FSM， $FSM(i,j)$ 的值表示图像i和图像j的相似度

(3) 在(2)基础上利用传递闭包法，对所有图片进行聚类，并利用已有标签分类进行验证，给出聚类评估的相关指标。

图片集中每一类有10张图，我们将他们放到一个文件夹all中，并将它们排序：

第1类图片：1~10号

第2类图片：11~20号

第3类图片：21~30号

我们采用平方法求传递闭包t(FSM)，并对传递闭包进行了截取， $\lambda = 0.98$ 得到30*30的截阵

将截阵中的分类提取出来得到如下分类结果：

第1簇：1 2 3 4 5 8 9 10 11 13 14 15 17 18 19 20 21 27

第2簇：6

第3簇：7

第4簇：12

第5簇：16

第6簇：22

第7簇：23

第8簇：24

第9簇：25

第10簇：26

第11簇：28

第12簇：29

第13簇：30

再对结果进行纯度分析，纯度计算公式为

$$Purity = (\Omega, \mathbb{C}) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j|$$

N ：总样本数

$\Omega = \{\omega_1, \omega_2, \dots, \omega_K\}$ ：一个个聚类后的簇

$\mathbb{C} = \{c_1, c_2, \dots, c_J\}$ ：表示正确的分类类别

ω_k ：表示聚类后第 k 个簇中的所有样本

c_j ：表示第 j 个类别中真实的样本

$Purity \in [0, 1]$ ， $Purity$ 越大表示聚类效果越好

使用Matlab计算得纯度为 $\frac{2}{3}$ ，聚类效果一般。