

2022 年《数字信号处理》课程大作业

【规则】

1. 本次大作业于 2023 年 1 月 8 日 24 时开始，2023 年 1 月 21 日 24 时结束；
2. 提交时，请在网络学堂上提交你的作业报告（PDF 版）和源代码；
3. 请在报告中清晰的写出你的思路、方法和结果。对于参考前人或其他人工作成果的内容，要用明显的方式标记出来，并指明被参考对象。否则将被认为是抄袭。

【背景】

随着传感器网络与大数据技术的迅速发展，数据源间的相互关系被越来越多地在引入到信号处理技术中，为获取更丰富和更高质量的信息提供了有效途径。图（Graph）被认为是描述这种相互关系的一种有效方法（具体参考文献[1]）。

【问题】

以气温传感器网络为题。附件 1 文件 data.mat 中给出了 2015 年 3 月共 31 天的分布在全美 50 个州、共 150 个气象站的日平均气温测量数据，以及气象站的地理位置（经、纬度）。假设某天有 1 个气象站传感器异常（记录的平均气温比真实值升高了若干度）。请你设计算法，以尽可能高的成功率检测出存在异常的气象站编号。

算法的性能用平均成功率评价。即遍历所有 31 天、每天 150 个气象站，分别假设该气象站出错（在数据上人为地增加或减少一定的数值，如 20° ），采用同样的算法对修改后的数据进行检测。若在每次实验中，算法只给出了 1 个出错气象站位置，并该位置正确，则计成功 1 次。那么成功率等于：

$$\text{平均成功率} = \text{成功次数} / \text{实验次数} = \text{成功次数} / 4650$$

你还可以尝试分析你的方法灵敏度，即出错带来不同的温度误差时的检测性能。多少度

【说明】

1. 数据格式

data.mat 共 34 列, 150 行。每行是一个气象站的信息:

0	1	2	3	31	32	33
气象站 编号	第 j 列: 第 j 日测量气温						纬度值	经度值

表 1: 数据格式示意图

气象站的经纬度信息用于计算各结点之间的地理距离:

$$d_{A,B} = R \cdot a \cos(\cos(a_A - a_B) \cos \beta_A \cos \beta_B + \sin \beta_A \sin \beta_B)$$

R: 地球半径

a : 经度

β : 纬度

原始数据来源于美国国家气象数据库网站: <http://www.ncdc.noaa.gov>。若你认为温度数据不够好, 或者需要更多数据进行训练, 可以从该数据库中进行下载。但要说明数据的来源, 含义。

参考文献:

1. Shuman D, Narang S K, Frossard P, et al. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains[J]. Signal Processing Magazine, IEEE, 2013, 30(3): 83-98.