



DeepL

订阅DeepL Pro以翻译大型文件。

欲了解更多信息，请访问[www.DeepL.com/pro](https://www.DeepL.com/pro)。



University of Glasgow | School of Computing Science

## 评估课程作业

课程名称	网络科学（硕士） COMPSCI5078 & COMPSCI5107			
课程数量	地理定位			
截止日期	时间	下午 4:30	日期	2024 年 3 月 8 日
对最终航标	20%			
个人或团体 ✓ 个人	x	组别		
预计小时数	20			
提交说明	通过 Moodle 提交报告和 Python 代码（压缩包）			
请注意：本课程作业不能重新评估				

### 提交课程作业的评估规则

课程文件中将公布提交正式评估课程作业的截止日期，逾期提交作业将受到如下处罚。

对于在公布的截止日期后提交的课程作业，将按以下方式计算初级成绩和二级分数段：

- (i) 对于在截止日期后五个工作日内提交的作品
  - a. 作品将按常规方式进行评估；
  - b. 然后，将按迟交作业的每个工作日（或工作日的一部分）减少两个二级等级。
- (ii) 超过截止日期五个工作日后提交的作品将被评为 H 级。

如果有正当理由，则不会对迟交作业进行处罚。

您应通过 MyCampus 提交证明理由充分的文件。

**不遵守提交说明的处罚为 2 个等级**

您必须通过 <https://studentlrc.dcs.gla.ac.uk/> 填写所有课程作业的 "自己的作品" 表 - 如果您没有签署原创声明，您的分数将不予公布。

# 个人评估：地理定位

课程作业应于 2024 年 3 月 8 日（星期五）下午 430 时提交（须经 LTC 批准

），CW 满分为 100 分，占期末总分的 20%。

所有材料均通过 Moodle 提交

团队的数据目录中大约有 10-20K 条来自伦敦/英国的推文。数据为 json 格式。我建议您使用 Jupyter 笔记本，并提交代码和输出存档。

- (i) 我们将为您提供一个数据集（在数据文件夹团队中）。编写 python 代码，将推文整理成 1km x 1km 的网格。绘制图表和/或数字，分析数据的分布。

我们收集数据时使用的坐标系为伦敦 = [-0.563, 51.261318, 0.28036, 51.686031]。

[25]

在报告中

用伪代码描述算法（5 分）

提供数据统计（推文总数、单元格中的推文数量以及分布情况等），并解释统计结果--这意味着什么？

(5 分)

为网格数据的可视化提供图表和/或数字（10 分 - 编码和描述 5

分；显示输出 5 分）

描述你对数据（以及由此产生的可视化效果）的看法/解释--你可能想强调使用该数据的任何潜在地理定位技术的问题。

(5 分)

- (ii) 您将得到一组高质量、低质量和背景推文。基于该数据集开发新闻价值评分方法。通过经验调整阈值来修改新闻价值，并讨论结果。

[30]

在报告中

解释你的新闻价值计算方法，并附上算法/伪代码（15 分）。

进行数据分析并对各种阈值进行分析；数据分析可包括使用或不使用停顿词、调整阈值等。

(15 分)

- (iii) 使用上述新闻价值评分技术分析给出的地理标记数据集 (i)，并讨论结果

[25]

*在报告中*

将您的评分方法应用于任务 1 - 中的数据。

调查得分较低和得分较高的推文；找到合适的阈值将它们区分开来，并删除新闻价值得分较低的推文

通过实证分析提供支持性信息，证明所使用的阈值是合理的。

(10分)

提供数据统计（推文总数、具有一定新闻价值的推文数量、分布情况等。）

(5 分)

在有新闻价值的数据上应用您创建的可视化方法；绘制数字/图表，并将其与 (1) 中的结果进行比较。有什么不同？

(10分)

(iv) *[开放任务]*

*举例说明并讨论因推文或消息来源的性质而导致的地理定位问题 (10)*

(v) *报告 - 10 分*

*a. 结构和格式*

*b. 阐述观点*

*c. 创造性地完成任务*