



数据分析及实践

Analysis and Practice of the Data

调研报告

刘 淇

Email: qiliuql@ustc.edu.cn

课程主页: <http://staff.ustc.edu.cn/~qiliuql/AD2021.html>



课程目标

2

□ 用科学的方法分析和应用数据

□ 实践（需要编程）5次（含报告4次，一般占40%）

■ 每人一份

■ 时间节点：大约每3周提交一次

□ 调研报告1份（一般占20%）

■ 每人一份

■ 时间节点：第12周（**2021.5.24**）

□ 以上+上课出勤情况即是考核方式



调研报告（2021.5.24）

3

- 结合本学期上课的内容，调研相关的书籍、课程、文献并撰写报告。可以从数据预处理、分析建模、应用或大数据平台等任一角度进行调研，调研的内容尽可能广，并将学到的内容和心得以报告形式提交。
- 建议：需要与本课程的教学内容相关
- 语言要求：中英文不限（包括调研的内容）
- 格式要求：专业、美观
- 报告内容：
 - 标题、摘要
 - 调研结果综述
 - 学习心得
 - 参考文献



调研报告（2021.5.24）

4

- 结合本学期上课的内容，调研相关的书籍、课程、文献并撰写报告。可以从数据预处理、分析建模、应用或大数据平台等任一角度进行调研，调研的内容尽可能广，并将学到的内容和心得以报告形式提交。
- 建议：需要与本课程的教学内容相关
 - 网络爬虫技术调研
 - NoSQL数据管理技术调研
 - 正则化、MAP等技术调研
 - 共轭先验分布及其应用调研
 - 常见参数估计方法的性质和应用
 - 熵在数据科学中的应用
 - 。 。 。



调研报告（2021.5.24）

5

- 想了解更多更前沿的研究领域，可以通过以下途径：
 - Nature, Science, IEEE/ACM Transactions等期刊及其子刊
 - KDD, AAAI, IJCAI, SIGIR, WWW, NIPS, ICML, WSDM, ICDM, CIKM, VLDB, SIGMOD, ICDE, ACL, EMNLP等国际会议官网（参考《清华大学计算机学科推荐学术会议和期刊列表》或者《中国计算机学会推荐国际学术会议和期刊目录》）
 - 各种科技论坛、公众号
 - 机器之心，新智元，专知、AI科技评论、
 - 在线课程



调研报告

6

□ 例如，从应用角度出发（推荐几篇）：

□ 社交网络

- An Influence Propagation View of PageRank
- From Footprint to Evidence: An Exploratory Study of Mining Social Data for Credit Scoring

□ 教育

- Question Difficulty Prediction for READING Problems in Standard Tests
- Collaborative Learning Team Formation: A Cognitive Modeling Perspective

□ 金融

- Product Supply Optimization for Crowdfunding Campaigns
- P2P Lending Survey: Platforms, Recent Advances and Prospects

□ 推荐系统

- Relevance meets Coverage: A Unified Framework to Generate Diversified Recommendations
- Personalized Travel Package Recommendation

□ 上述所有论文均可在实验室主页上可下载<http://staff.ustc.edu.cn/~cheneh/>



调研报告

7

- 例如，从模型角度出发：
 - 采样方法
 - MCMC - Metropolis-Hastings
 - MCMC - SLICE SAMPLING
 - 分类
 - 决策树
 - 支持向量机(SVM)
 - 贝叶斯分类器
 - 聚类
 - K均值算法(K-means)
 - 基于密度的聚类(DBScan)
 - 层次聚类



调研报告

8

- 例如，从模型角度出发：
 - 概率主题模型
 - 隐马尔科夫模型(HMM)
 - 隐狄利克雷分配模型(LDA)
 - 深度学习
 - 卷积神经网络(CNN)
 - 循环神经网络(RNN)



调研报告-评分标准

9

□ 多文献调研报告

- 独立写作程度如何、是否抄袭
- 是否有明确主题、主题是否符合课程内容、表述是否合适
- 是否有背景和现状综述
- 对各文献的单独表述是否清晰具体
- 是否对各文献涉及模型和方法做有意义的比较分析
- 逻辑表达是否通顺
- 参考文献是否完整
- 提交是否及时 (**2021.5.24**)



调研报告-评分标准

10

- 单文献调研报告
 - 独立写作程度如何、是否抄袭
 - 模型或方法部分表述是否清晰
 - 实验部分表述如何
 - 是否切中论文核心贡献
 - 逻辑表达是否通顺
 - 参考文献是否完整
 - 提交是否及时 (**2021.5.24**)



调研报告

□ 提交要求

- 将调研报告发送给助教: apdata2021@163.com
- 邮件标题: 姓名_学号_report
文件命名: 姓名_学号_report.pdf



课程信息

12

□ 课程主页

□ <http://staff.ustc.edu.cn/~qiliuql/AD2021.html>

□ 助教信息

□ 岳立楠（调研报告）lnyue@mail.ustc.edu.cn

□ 安彦青（调研报告）anyq@mail.ustc.edu.cn

□ 顾垠（实验）gy128@mail.ustc.edu.cn

课程QQ群



数据分析与实践-ustc

扫一扫二维码，加入群聊。