

美赛经验分享



李可欣
2018.1.11

❖ 感谢我的队友们~

❖ 工业工程系 钟月漾

❖ 数学科学系 赵溢佳

赛前准备

❖ 技能点：

- ❖ 建模：经典模型及其实现、往年优秀论文
- ❖ 代码：MATLAB/Python/R/C++/SAS/Lingo...
- ❖ 写作：LaTeX、绘图、英语表达

❖ 资源搜索

❖ 团队合作：

- ❖ 良好沟通、安排分工；明确比赛要求

资源搜索

❖ 搜索引擎、图书馆电子资源、文献库、数据库

❖ 做笔记

UCI数据集: <http://kdd.ics.uci.edu/>

CMU数据集: <http://lib.stat.cmu.edu/datasets/>

<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/>

时序数据集: <http://www.stat.wisc.edu/~reinsel/bjr-data/>

金融数据集: <http://lisp.vse.cz/pkdd99/Challenge/chall.htm>

癌症基因数据集: <http://www.broadinstitute.org/cgi-bin/canc>

综合数据集: <http://www.cs.nyu.edu/~roweis/data.html>

数据集列表: <http://www.kdnuggets.com/datasets/index.html>

美国政府开放数据: <http://data.gov>

中国地方政府开放数据: 北京 <http://www.bjdata.gov.cn/>

上海 <http://datashanghai.gov.cn>

UCI机器学习网站

<http://archive.ics.uci.edu/ml/>

Weka官方网站

<http://www.cs.waikato.ac.nz/ml/weka/>

DBMiner官方网站

<http://ddm.cs.sfu.ca/>

SVM代码

<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

其它开源软件包: NB (朴素贝叶斯网络), NN (神经网络), DT (决策树)

相关软件: Matlab, StatSoft等商用软件; SQL Server 2008中也提供了相应的
Data Analysis数据分析工具

选题

MCM Problem A (continuous)

MCM Problem B (discrete)

MCM Problem C (data insights)

ICM Problem D (operations research/network science)

ICM Problem E (environmental science)

ICM Problem F (policy)

获奖比例、文献搜索难度 ~ 问题

时间分配

❖ Day 1:

❖ 确定选题、初步文献搜索

❖ Day 2:

❖ 确立论文初步结构、继续文献搜索

时间分配

- ❖ Day 3:
- ❖ 应用模型于具体问题&数据、修正模型
- ❖ Day 4:
- ❖ 修正模型、写文章、摘要（！）；最终检验

- ❖ 陈述问题、假设、思路方法、成果
- ❖ 表述清晰、整体把握
- ❖ 创新点