

Mid-term Project

April 21, 2021

多标签用户人格分类

MBTI理论认为人的个性可以从四个角度进行分析，用字母代表如下：

- 驱动力的来源：外向E---内向I
- 接受信息的方式：感觉S---直觉N
- 决策的方式：思维T---情感F
- 对待不确定性的态度：判断J---知觉P

其中两两组合，可以组合成16种人格类型；

本次大作业要求利用机器学习方法，通过用户的发言记录对用户的人格类型进行分类

数据链接: <https://www.kaggle.com/datasnaek/mbti-type>

1. 使用SVM进行人格分类:

- 对数据进行预处理
- 使用SVM进行人格分类
- 提交报告及代码

2. 使用集成学习方法完成人格分类：

- 对数据进行预处理
- 使用Random Forest进行人格分类
- 提交报告及代码

□ SVM方法：8分（可以调用外部库）

1. 分别选用linear、RBF、poly等核函数
2. 超参数的选取依据
3. 多折交叉验证

□ Random Forest：12分（自行实现算法）

1. 分别用Information Gain、Gain Ratio、Gini Index选择属性
2. 详细介绍连续值属性的划分方法
3. 超参数的选取依据

- 5月16日：提交SVM方法报告和代码
- 6月13日：提交Random Forest方法报告和代码

注意事项

- ❑ 可以参考网上资料，但不得抄袭他人或同学，抄袭0分；
- ❑ 加分项：使用英文撰写报告，设计合理实验（对比不同模型表现/不同超参数对性能影响/考虑一些数据集的其它的问题），自己撰写模型代码，尽可能少的调用工具。

Thank you!