

决赛事项

> 更新日期：2019-07-22 进入决赛阶段

题目

多分类问题，还涉及到多标签的问题，主要有正常心电图和八种异常心电图，包括如下。

编号	类别	英文缩写
0	正常	Normal
1	心房颤动	AF
2	一度房室传导阻滞	FDAVB
3	完全性右束支传导阻滞	CRBBB
4	左前分支阻滞	LAFB
5	室性早搏	PVC
6	房性早搏	PAC
7	早期复极图形改变	ER
8	T 波改变	TWC
9	其他心电图	Others

数据情况

分为训练、验证、测试三部分，前两个可见，测试集不可见用于排名。

- 12 导联，mV，500Hz，包含性别和年龄等信息

数据还没有发布，要到 13 号才能看到数据，另外这次的无法下载，只能远程桌面的形式访问，类似于当时脑卒中的数据

预说明

1. 此次参赛为线上的方式，京东云提供的云桌面，云桌面为断网状态。
2. 数据只能通过云桌面查看，不能下载，本地上传代码进行运行调试
3. 云桌面账号同一时段只能实现一台电脑的接入
4. 云桌面配置，Ubuntu 16.04, 12 核 48G, P40 GPU

时间线(重要)

重要时间点整理

- 7 月 26 日 10 点至 8 月 10 日 21 点	决赛整体时间
- 7 月 26 日 10 点至 8 月 7 日	S1 线上阶段
- 8 月 8 日	中止, 前往天津报到
- 8 月 9 日 9 点至 8 月 10 日 21 点	S2 天津 36 小时 hackathon
- 8 月 10 日 21 点后	S3 完成答辩 ppt 制作
- 8 月 11 日	S4 答辩

评价机制

多加了一个类别, 同时还有速度性能评分, 由于占比过小, 不予考虑

本次大赛采取基于多标签分类的评分方式[1], 从计算方式来看, 是基于标签来进行评价的。首先, 针对第 j 个类别定义如下四个变量, 其中 $0 \leq j \leq 8$,

$$TP_j = |\{x_i | y_i \in Y_j, y_i \in f(x_i), 1 \leq i \leq N\}|$$

$$FP_j = |\{x_i | y_i \notin Y_j, y_i \in f(x_i), 1 \leq i \leq N\}|$$

$$TN_j = |\{x_i | y_i \notin Y_j, y_i \notin f(x_i), 1 \leq i \leq N\}|$$

$$FN_j = |\{x_i | y_i \in Y_j, y_i \notin f(x_i), 1 \leq i \leq N\}|$$

由此计算每一类的 *Precision*、*Recall* 和 F_1 分数:

$$\begin{aligned} Precision_j &= \frac{TP_j}{TP_j + FP_j} \\ Recall_j &= \frac{TP_j}{TP_j + FN_j} \\ F_{1j} &= \frac{2 \cdot Precision_j \cdot Recall_j}{Precision_j + Recall_j} \end{aligned}$$

平均 F_1 分数为上述 9 个分数的算术平均值, 即:

$$F_1 = \frac{1}{9} \sum F_{1j}$$

[1] Zhang, Min-Ling, and Zhi-Hua Zhou. "A review on multi-label learning algorithms." IEEE transactions on knowledge and data engineering 26.8 (2014): 1819-1837.

数据集

训练集 /media/jdcloud/Train

验证集 /media/jdcloud/Val

验证集结果 ranking.jdworkspace.com

注意事项

- 1、不要更新系统，不要随意更新 python 版本（遇到了任何更新一律 cancel）
- 2、系统的磁盘情况
 - a) 磁盘 1: 150G, 可随意使用, 挂载点/media/uuser/data/ , 所有软件尽可能装到此处
 - b) 磁盘 2: 50G, 用于存放心电数据, 实际上是系统空间, 占用较多, 这一点与手册上略有不符, 不要随意占用其空间。
- 3、各队伍注意将代码保存至/media/uuser/data/final_codes, 方便拷贝代码文件以发送给各队伍（可以包含代码和参数文件, 但不能存放过大的模型和参数文件）;
- 4、同时将可运行脚本和参数等文件保存至/media/uuser/data/final_run, 便于算法测试。