

# 程序及使用方法

## 运行的软硬件环境

1. 我们的程序处理数据时默认用 12 个 CPU 线程，训练时用全部的 CPU 核；如果 CPU 核心数目不到 12，可以在程序中改小
2. 训练时大约需要 92GB 内存，请保证内存充足 (测试时只需 30GB 内存，推断只需 1G 内存不到)
3. 如果要训练深度神经网络模型，需要显存容量大于 11G 的 NVIDIA GPU，请确保 NVIDIA 显卡驱动支持 CUDA 9.0 (如果只是用来推断，则可以在 CPU 上运行)
4. 互联网连接 (predict.py 会从 opendota.com 下载比赛记录并预测结果)
5. Linux x86-64 (在 windows 上未经测试)
6. python 3.7 (建议使用 conda 安装)

## 依赖库

1. numpy - 1.15.2
2. pandas - 0.23.4
3. scipy - 1.1.0
4. scikit-learn - 0.20.0
5. matplotlib - 3.0.0
6. requests - 2.19.1
7. tqdm - 4.26.0
8. pytorch - 1.0.0.dev20181003 (预览版)
9. lightgbm - 2.2.0

除 LightGBM 以外都能用 conda 安装; pytorch 安装方法参见 [pytorch.org](https://pytorch.org/); LightGBM 可以用 pip 安装, 参见 <https://lightgbm.readthedocs.io/en/latest/Python-Intro.html>

## 使用方法

### 推断某一局比赛的实时胜率

1. 在 opendota.com 搜索比赛编号，进入比赛页面
2. 如果该比赛未被 Parse, 要先在网页上请求 Parse，直到 Parse 成功
3. 执行 `python -O predict.py [比赛编号] model_lightgbm.txt model_deep.pt`
4. 输出会放在 [比赛编号].png 文件中

### 在测试集上测试

```
python -O test.py [数据目录] model_lightgbm.txt model_deep.pt
```

即可评估模型在测试集上的表现。注意，每个数据是按照比赛编号被随机 Hash 到训练集，验证集或者测试集上的，所以 data3 目录下并不是每个数据都会被用来测试。

## 训练 LightGBM

`python -O lightgbm_train.py [数据目录] [模型保存文件]`

## 训练深度神经网络

`python -O deep_train.py [数据目录]`

模型自动保存在当前目录下一个以时间戳命名的目录内

## 其它文件

`model.py` 定义了基于注意力的深度神经网络模型

`dota_data.py` 包含数据处理和特征提取的程序