

2017 “贝贝网·种子杯”初赛比赛报告

队伍名称

红鲤鱼与绿鲤鱼与驴

队伍成员

冼晋毅、王为红

使用语言以及运行环境

使用语言：python

运行环境：tensorflow

表格数据处理相关

数据处理相关

目录结构：

```
Src
├── 2017“贝贝网·种子杯”编程PK赛-初赛题.pdf
├── 红鲤鱼与绿鲤鱼与驴比赛报告.md
├── 红鲤鱼与绿鲤鱼与驴比赛报告.pdf
├── data
│   ├── check.py
│   ├── finalResult.csv
│   ├── predictPro_template.csv
│   ├── computeResult.csv
│   ├── finalTestResult.csv
│   └── removeBadData.py
```

```
├── computeScore.py
├── getTeamData.py
├── removeBadTestData.py
├── computeTestResult.csv
├── matchDataTest.csv
├── result.csv
├── computeTestScore.py
├── matchDataTrain.csv
├── teamData.csv
├── models
│   ├── ...
│   └── ...
├── nn
│   ├── ...
│   └── ...
```

文件介绍

teamData.csv

主办方给的每个篮球队的数据

getTeamData.py

读取_teamData.py_的内容，将每组的球员信息合并成向量，并使所有的队伍补齐队员人数存入_result.csv_中。

result.csv

存放每队的数据，一个队伍所有球员的数据存放在一行中。

check.py

检查result.py中每行是否是同样长度。

computeScore.py

读取result.csv，读取matchDataTrain.csv中的对局数据，
将队伍名使用result.csv中的数据替代，将"0胜1负"使用'0','1'代替，
最后的比分，若客场获胜即积1分，否则为0分。
将最后的数据写入computeResult.csv中。

computeResult.csv

存放训练集处理数据。

computeTestScore.py

读取result.csv，读取matchDataTest.csv中的对局数据，
处理方式与computeScore.py的方式相同。
将最后的数据写入computeTestResult.csv中。

computeTestResult.csv

存放测试集处理数据。

removeBadData.py

测试集中存在数据超出指定行数的数据行，对其进行删除。
数据存入finalResult.csv。

finalResult.csv

最终的训练集数据。

finalTestResult.csv

最终的测试集数据。

predictPro_template.csv

主办方给的提交数据格式

matchDataTest.csv

主办方给的测试数据集

matchDataTrain.csv

主办方给的训练数据集

代码相应的接口以及运行需要用到的变量含义

neural_network.py

初次训练神经网络，并保存计算图和相关变量

neural_network_continue.py

对神经网络继续进行迭代

在测试集上运算，给出计算结果

数据特征提取思路

尽量充分利用数据，
把客场和主场的队员信息特征（共使用15名队员的比赛数据）
以及两个队伍的战绩作为特征。

预测模型选取（包括对于规则的描述和最终模型的选择）

由于在将数据向量化后数据的维数过大，
故先用PCA对数据进行降维（降至207维）。
然后将降维后的数据输入神经网络当中进行训练。
神经网络为三层的全连接神经网络。神经网络的训练采用Adam算法进行训练。

对于模型参数的选择和优化思路

用较小的迭代次数，先对神经网络的隐藏层神经元数量进行选择。
初步选定几个神经网络隐藏层神经元数量后，
加大迭代次数，对模型进行进一步的训练。
之后在测试集当中分出的CV集中对不同模型的效果进行评估。
最后选取效果最优的模型来作最后的预测。