**数据划分：**

考虑到数据集的规模，并且个人认为用户过去两周的游戏表现足够表征其游戏水平，故对数据集进行采样以及训练、验证数据构造。

采样：只使用大乱斗数据用于建模（因为测试集为大乱斗数据）

数据集的划分：使用前两周的数据作为用户特征提取，使用一周的数据用于训练，使用一天的数据作为验证，比如

* 3月4号-3月17号的数据用于构造user的统计特征
* 3月18号-3月24号的数据用于构造训练数据（只利用和test一致的特征）
* 3月25号的数据用于构造验证集（只利用和test一致的特征）
* 对于test（4月1号）的预测，采用3月18号-3月31号的数据用于构造user的统计特征

**建模思路：**

由于一个队伍五个人的位置可能不同，直接将一个队伍的user特征拼接后送入比如树模型或者神经网络我认为是不合理的（算力受限还未验证），因为直接拼接所有人的特征会出现特征错位的现象。

因此，我的建模思路为将每个队员看作一个节点，每个队可以构成一个由5个节点的全连通图，node embedding的学习能够利用到团队其他成员的特征，且每个队的graph embedding可以代表一个队伍的实力强弱。

由于这里考虑的是全连通图，可以用transformer encoder来替代graph attention操作，转化成一个文本分类问题，即每个队员可以看作是序列中的一个“字“，每个“字”对应的初始embedding特征就是user的统计特征。



代码文件：

EDA.py 用于划分数据集

feature\_engineering.py 用于构造user特征

training.py 训练代码

inter.py 预测结果