**游戏异常检测——模型**

模型分为两部分，分别是（一）基于XGBoost的异常行为检测和（二）基于GRU的恶意言论检测。（三）后续工作总结了现有模型的问题和改进点。

**（一）基于XGBoost的异常行为检测**

**1. 特征工程**

考虑6种特征：结算特征、匹配相关特征、出装相关特征、打野相关特征、消极异常和表现异常。不直观的特征说明如下。提取特征后，进行简单的处理（StandardScaler）并作为模型的输入。

**1.1结算特征**

'游戏时长(min)'

'胜败'：0或1

'英雄'：0-5，分别表示法师、刺客、战士、坦克、辅助、射手

'评分'

'击杀数', '死亡数', '助攻数'

'输出', '承伤', '治疗'

'参团率'

'总经济', '野怪经济'

'装备数'

'对塔伤害'

'投降次数', '健康回城次数', '被几个人举报'

'输出占比', '承伤占比', '总经济占比', '对塔伤害占比', '投降次数占比', '输出占比/经济占比', '承伤占比/经济占比'

**1.2匹配相关特征**

'是否匹配成功'：0或1

'匹配时间(s)'

'匹配成功率'

'平均匹配时间'

**1.3出装相关特征**

'10min\_wearing\_0'：10min时物理攻击装备数量

'10min\_wearing\_1'：10min时法术攻击装备数量

'10min\_wearing\_2'：10min时防御装备数量

'10min\_wearing\_3'：10min时辅助装备数量

'10min\_经济'

'10min\_装备数'

'15min\_wearing\_0'：15min时物理攻击装备数量

'15min\_wearing\_1'：15min时法术攻击装备数量

'15min\_wearing\_2'：15min时防御装备数量

'15min\_wearing\_3'：15min时辅助装备数量

'15min\_经济'

'15min\_装备数'

'wearing\_similar'：实际出装和推荐出装重合的件数

(recommended\_wearing = {0: [1, 1, 1], 1: [0, 0, 0], 2: [2, 2, 2], 3: [2, 2, 2], 4: [3, 2, 2], 5: [0, 0, 0]})

**1.4打野相关特征**

'敌人反野/骚扰'：敌人反野个数

'队友抢野'：抢野队友id或0

**1.5消极异常（挂机、转圈）**

'mean\_window\_distance'：两分钟内每次打点的平均位移

'mean\_window\_out'：两分钟内每次打点的平均输出

**1.6表现异常**

'history\_score'：历史平均评分

'history\_score\_std'：历史评分std

'history\_winrate'：历史胜率

'history\_winrate\_std'：历史胜率std

'recent\_winrate'：最近15场胜率

**2. 模型及对比**

XGBoost效果最优。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 模型/性能 | XGBoost | DeepFM | XGBoost+LR | XGBoost+FM | FM |
| 准确率 | 99.8% | 97.8% | 92% | 88.2% | 79.3% |
| 召回率 | 95.7% | 97.7% |  |  |  |

**2.1 XGBoost**

准确率99.8%，召回率95.7%

混淆矩阵

19336   10

28     626

**2.2 DeepFM**

准确率97.8%，召回率97.7%

混淆矩阵

19249 97

367 287

**2.3 XGBoost + Logistic Regression**

先用已有特征训练XGBoost，再用XGBoost学到的树构造新特征，最后用新特征和原有特征一起训练模型。构造的新特征是one-hot向量，向量的每个元素对应XGBoost的叶子结点。样本点落在某个叶子结点上，则新特征向量中该叶子结点对应的元素值为1。

准确率92%

**2.4 XGBoost + Factorization Machine**

准确率88.2%

**2.5 FM**

准确率 79.3%

**（二）基于GRU的恶意言论检测**

**1. 预处理**

分词、去标点、转换小写（text\_to\_word\_sequence()）

停用词

缩写转换（I’m->I am）

词干提取

**2.模型**

构造词表->embedding->GRU->全连接(relu)->dropout->全连接(relu)->全连接(sigmoid)。

GRU在Twitter网络暴力数据集上准确率96.4%，但在其他数据集上测试准确率很差，泛化性能受词表覆盖范围限制。

LSTM训练较慢，效果也远差于GRU。

**（三）后续工作**

1. 异常行为检测部分，由于模拟数据中没有重大团战，死亡相关特征目前无法加入

2. 恶意言论检测部分，需要找多个数据集训练，提高泛化性能（<http://hatespeechdata.com/>）

3. 将恶意言论检测融入异常行为检测模型中