PUBG Game Data Analysis and prediction

Yang Cao

Deakin University, Australia

Introduction

In a PUBG game, up to 100 players start in each match (matchId). Players can be on teams (groupId) which get ranked at the end of the game (winPlacePerc) based on how many other teams are still alive when they are eliminated. In game, players can pick up different munitions, revive downed-but-not-out (knocked) teammates, drive vehicles, swim, run, shoot, and experience all of the consequences – such as falling too far or running themselves over and eliminating themselves. Different game behaviors will lead to different final rankings, so the main purpose is to build a model to predicts players' finishing placement based on their final stats, on a scale from 1 (first place) to 0 (last place).

PUBG Game Data Analysis aims to make A game team data know which game actions that make game teams get higher rank than others

PUBG win place prediction aims to help Players can also estimate their final ranking based on the current situation and make strategic decisions in advance (such as running away or fighting) Hence, the outlying aspects mining is also referred to *outlier interpretation* or *object explanation*.

Description and evaluation

- Use Mean Square Error to evaluate model (the average squared difference between the estimated values and the actual value)
- Train data MSE
- Test Data MSE

Data Visualization

We propose the use data visualization technique to show the game types proportion and the relationship between walking distance and win place.

Data Preprocess

We remove all the rows which has missing values or NaN values.

Train Dataset description There are 4446966 rows and 29 columns. 4446966 unique ID. 2026745 unique groupId

Test Dataset description There are 1934174 rows and 28 columns. 1934174 unique ID 886238 unique groupId

Feature and model selection

Linear regression, use grid search to look for best parameters.

Parameters	Values	CV
$\text{fit}_i ntercept$	True/False	3
normalize	True/False	3

Decision Tree, use grid search to look for best parameters.

Parameters	Values	CV
criterion	"mse", "$\text{friedman}_m se$", "mae"	3
$\min_s amples_l eaf$	1,2	3

Conclusion

Both training and testing data shows that linear model get lower mean square error value.

Most players choose to play squad-fpp and duo-fpp.

More walking distance always can bring high

Acknowledgement

- International Cooperation Project (Y7Z0511101) of IIE, Chinese Academy of Sciences