

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.

GOAM Algorithm

Second, based on the *earth move distance*, we calculate the outlying degree.



Testing figcolor

where G_q is the query group, n is the number of compare groups, and h_{k_s} is the histogram representation of G_k in the subspace s .

Outlying Aspects Identification In this step, based on the value of outlying degree we will identify the group outlying aspects. If a feature’s outlying degree is greater than a threshold, the more likely the feature is group outlying aspect. When the dimensionality of features is high, we adopt a stage-wise candidate subspace construction strategy to alleviate the exponential explosion.

Experiment

Synthetic Dataset contains 10 groups and 8 features. Each group consists of 10 members, and each member has 8 features.

Method	Truth Outlying Aspects	Identified Aspects	Accuracy
GOAM	$\{F_1\}, \{F_2F_4\}$	$\{F_1\}, \{F_2F_4\}$	100%
Arithmetic Mean based OAM	$\{F_1\}, \{F_2F_4\}$	$\{F_4\}, \{F_2\}$	0%
Median based OAM	$\{F_1\}, \{F_2F_4\}$	$\{F_2\}, \{F_4\}$	0%

It can be observed that the GOAM method can identify the trivial outlying features and non-trivial outlying subspaces correctly and is obvious from the table that the accuracy of GOAM is the best, which is (100%).

NBA Dataset was collected from Yahoo Sports website (<http://sports.yahoo.com.cn/nba>). The data include all teams from the six divisions, and each player in the team has 12 features.

Teams	Trivial Outlying Aspects	NonTrivial Outlying Aspects
Cleveland Cavaliers	$\{3FA\}$	$\{FGA, FT\% \}, \{FGA, FG\% \}$
Orlando Magic	$\{Stl\}$	None
Milwaukee Bucks	$\{To\}, \{FTA\}$	$\{FGA, FTA\}, \{3FA, FTA\}$
New Orleans Pelicans	$\{FT\% \}, \{FTA\}$	$\{FTA, Stl\}, \{FTA, To\}$



Testing figcolor



Testing figcolor

New Orleans Pelicans on FT%

New Orleans Pelicans on FTA

New Orleans Pelicans has more players with lower {free throw percentage}, {free throws attempted}.

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