Data Mining on Dota2 Matches

Your Name
Date

Dota 2 data analysis using machine learning method

Dota (Defense of the Ancient) 2 is a multiplayer online battle arena video game published by Valve Corporation. I am planning to implement a machine learning on Dota 2 matches. In whole Dota field, it lacks a deeper data mining with professional statistic methods applied in economy or other serious subject. In the project, I am planning to use Principal Component Analysis (PCA), Linear Regression, Unsupervised learning, Naive Bayes, Neural networks and some other methods to make the data analysis.

Background

Dota 2 is a multiplayer online battle arena video game published by Valve Corporation. Every year Valve hosts a great electronic sports Dota 2 championship tournament: The International. It attracts many different players from all over the world, with a large prize pool (In 2015, it was $18 million). Also, millions of players played the game online every day. It is a team fight between two five-player teams, each of which occupies a stronghold in a corner of the map. Who destroy the Ancient on the other side wins the game. It has more than 100 heroes and complicated item system. So, for this big game, players need a data analysis report to guide them to play it well. All the teams need data mining process to see how to play against their opponent including which hero should be banned, which should be picked and how to make a strategy to make a hero group.

Currently, teams has their data analyzer help to make a basic statistic on which hero has higher win rate and how to combine a strategy with hero picked in a game. They may also try to study their opponent's preference on hero selecting. But in whole Dota field, it lacks a deeper data mining with professional statistic methods applied in economy or other serious subject. Players always use their own feeling to pick hero in a match or just make a simple ban/pick strategy. So, in this project, we will use machine learning method to find out those important factors and make a prediction for the match. In this way, if the result is satisfactory, it can generate a model for teams to build up their strategy when facing different opponents.

Data sources

I will use the data from valve server and other game statistic website:

Valve server queries: https://developer.valvesoftware.com/wiki/Server_queries

Format: https://partner.steamgames.com/documentation/webapi

Dotabuff: http://www.dotabuff.com/matches

Gosugamer: http://www.gosugamers.net/dota2/gosubet

The data collection may need web crawler or internet package script to acquire the information of each match.

Algorithm

In the project, I am planning to use Principal Component Analysis (PCA), Linear Regression, Unsupervised learning, Naive Bayes, Neural networks and some other methods to make the data analysis. Then it will generate a general report for common players to get more competitive in the game. The next target is establishing a predictive model to make a further analysis for professional team players through this machine learning result.

Principal Component Analysis

http://www.r-bloggers.com/principal-component-analysis-using-r/

Linear Regression

http://www.r-bloggers.com/using-linear-regression-to-predict-energy-output-of-a-power-plant/

Naive Bayes

http://www.r-bloggers.com/understanding-empirical-bayes-estimation-using-baseball-statistics/

Neural Network

http://www.r-bloggers.com/fitting-a-neural-network-in-r-neuralnet-package/

References

War and Picks http://www.datdota.com/blog/?p=1323

Analysis of the Time Aspect of the Matches at The International 3 (Dota 2 Tournament)

http://www.r-bloggers.com/analysis-of-the-time-aspect-of-the-matches-at-the-international-3-dota-2-tournament/

Result Prediction by Mining Replays in Dota 2

https://www.diva-portal.org/smash/get/diva2:829556/FULLTEXT01.pdf

Identifying Patterns in Combat that are Predictive of Success in MOBA Games
http://ciigar.csc.ncsu.edu/files/bib/Yang2014-MOBASucessPatterns.pdf