Data Analysis and Machine Learning on League of Legends

- Project Proposal -

Your Name: Yiyi Zhou
Chang Liu
Date: 10/05/2017

Introduction

This Project aims to address some specific issues like the degree of impact of some particular changes in a game patch, and how these variables affects the win rate, as well as other prediction issues based on machine learning methods and algorithms, including principal component analysis, supervised and unsupervised learning.

Background

League of Legends (abbreviated LoL) is a multiplayer online battle arena (MOBA) video game developed and published by Riot Games. Released in 2009, it has been a great popular MOBA game around the world. Professional eSport championship is held by Riot Games every year, with all teams selected and competed in each region professional league to take part in the tournament, this worldwide annual tournament has had $29,203,916 USD in prize money.

During a normal match, players will be involved in a 5v5 team game to work with teammates and compete with the enemy team. The ultimate goal is to destroy the main base of the enemy's team. For each match, players can select a different champion as the character to operate, and each champion has unique abilities and game roles.

For over the seven years of professional tournaments, there are huge numbers of match records demonstrating various details and information in a very particular time stage for this game. Like the game style in the early stage years ago would be a gank oriented strategy, hence the jungle would be the core role

during the entire game, while currently this game kernel idea has been team fighting and a late-stage game.

There are some data analysis for players when updates are released, so they can follow the tendency brought by those slight changes in every update, like which champion is stronger, and which has been nerfed. But in general, there lacks data mining analysis based on machine learning methods for specific data analysis, like the degree of the impacts of every slight change of particular equipments and champions' basic attributes. These data aggregated information could be helpful to better understanding for the game decision in a particular patch of game.

Data Sources

The datasets that will be used in this project are from kaggle and other game statistic website:
Kaggle: League of Legends
https://www.kaggle.com/chuckephron/leagueoflegends (Professional)
https://www.kaggle.com/lanls1/matchidv1 (Rank)
https://www.kaggle.com/xenogearcap/league2016 (Rank)

Official Match History site:
http://matchhistory.na.leagueoflegends.com/en/#page/landing-page

Official stats API:
https://developer.riotgames.com

Algorithm

This project will be addressing particular question with different, specific methods or models. Including Principal Component Analysis (PCA), Linear Regression, Unsupervised Learning, Supervised Learning, Naive Bayes, Neural Networks as well as some other algorithms and methods.

Principal Component Analysis

http://scikit-learn.org/stable/modules/generated/
sklearn.decomposition.PCA.html
https://en.wikipedia.org/wiki/Principal_component_analysis

Linear Regression
http://scikit-learn.org/stable/modules/generated/
sklearn.linear_model.LinearRegression.html
https://en.wikipedia.org/wiki/Linear_regression

Naive Bayes
http://scikit-learn.org/stable/modules/naive_bayes.html

Neural Network
http://scikit-learn.org/stable/modules/neural_networks_supervised.html
http://scikit-learn.org/stable/modules/neural_networks_unsupervised.html

References

Dataset Analysis of League of Legends
https://kwtrnka.wordpress.com/2015/09/21/bigger-league-of-legends-data-set/

Machine Learning to League of Legends
http://data-speaks.luca-d3.com/2017/05/machine-learning-to-analyze-league-of.html

League of Legends stats website
http://www.lolking.net/

Basic Data Exploration in LoL
https://www.kaggle.com/erenan/basic-data-exploration-matches-2014-2017