

Project Description

amenard

October 2024

1 Subject

This project intends to study the applications of machine learning in the field of League of Legends (LoL) esports predictions.

On one hand, LoL esports is a very young and volatile scene, with a heavy player turnover, a constantly changing game, and a rapidly evolving competitive structure, but as a video game it also provides us with a very complete and detailed amount of data, but is it enough for machine learning to grasp and tame LoL esports and accurately predict match results ?

With this project, I intend to learn how to create, design and tune different machine learning methods, as well as selecting and refining the relevant and meaningful data to said structures to reach the best possible prediction accuracy.

I propose to study various machine learning structures (neural network, random forest, SVM, KNN, GBM, and ensemble methods) with different game data (selecting which parts are relevant to the prediction of the outcome) to determine the best combination to predict the outcome of a game, and a comparative study of them. In a favorable case, I would also like to provide a trainable or trained model.

2 Plan

2.1 Data gathering

Several solutions are considerable :

- League of Legends API : Easily accessible, but requires a fair amount of data processing, and has limited request rates. Works for general games which may not be representative of esports games.
- Lolesport API : More complex to access, fair amount of data processing, direct esports game data, no request rate, but a lower total amount of data.
- ABIOS, Challenger mode *TBD*

2.2 Data refining

The selection of the relevant data will be at first based off of my game knowledge, then adjusted according to the models performances

I expect players' gold and KDA to have a predominant role in the prediction, followed by their playing stats (damage inflicted and received, all game or in the last minutes). I believe more detailed statistics such as armor, health, and attack stats can improve the prediction. I also expect the timestamp to bear a major role in the prediction.

2.3 Champion input

Champion selection is a major part of a game, and treating the team compositions might be tricky, as they are not numerical values.

A vectorization of the champion's space might be necessary which would be a highly complicated step. Ideally, I expect the in-game statistics of the champions to be enough a representation of the champion's style to avoid a vectorization. A vectorization using the games champion description could be considered, but is likely to be non-representative enough.

2.4 Models

Models to try :

- Neural network (Dense, LSTM, RNN ?)
- Random forest
- SVM
- KNN
- GBM

2.5 Evaluation

With an in-game data frame as input, the model will return a value between 0 and 1, with an output superior to 0.5 meaning that team A is more likely to win at this instant. I expect to have to normalise the value so that across all frames for which the model outputs X , team A ends up winning in $100X\%$ of the cases. This normalization should be easy and natural given the nature of the data and the prediction.

To assess the model's performance, the best fitting metric has yet to be determined. A few potential contenders are :

- A success rate (the model succeeds when outputting ≥ 0.5 and team A wins, or outputting ≤ 0.5 and team B wins)

- The average distance between the output and 1 (or 0) if team A wins (or team B wins). Eventually the average squared distance.
- The average distance between the output and 0.5 if the model succeeds, else 0. Eventually the average squared distance.

As the game sees major changes every year, depending on the amount of data necessary to train the models, I intend to :

- Train the model on any competition's regular season (or Group/Swiss stage) and validate it on the play-offs phase.
- Train the model on every competition (regional championships, MSI and world's play-ins) and validate it on the world's main event phase (Group/Swiss stage and play-offs).

This process will be done ***independently*** (to account for the major game changes each year) on the last 3 competitive years.

3 Project evaluation

The project evaluation will be based on a short report handed to Björn Thureson and Mads Dam, as well as a presentation to Björn Thureson.