Intuition

Like mentioned in the proposal, our method is mainly based on the timeline of a match, which means we use in-game data instead of pre-game data, so it’s more accurate and dynamic compared to those static predictions. Also, since we divide the whole game into stages of 5 minutes, and make a prediction every 5 minutes, we can dynamically give different results according to how players perform, and thus our prediction becomes more useful. And when our model is used by professional players to review their match, by viewing data provided by our model, they will know more about where should be improved, and thus our model become helpful to LoL players.

Experiments / Evaluation

Data source

Our data comes from Riot API, which includes millions of matches. We get several thousands of matches, each includes some pre-game data and a timeline. For these data, we divide them into two sets, one for training the model and another for testing the model.

Questions

Using our model, we can get data and get answers about these questions at each point (5 min, 10 min, etc.)

1. Which team has advantage over another?
2. What’s the possibility that our predicted winner really wins at last?
3. On what aspects the desired winner has advantage?
4. How much are the disadvantages between the two teams?

Experiments / Observation

First in the pre-training process for items, we tried two different ideas, one is to make each champion unique and for each of them, train a weight for each item, and the other one is to divide champions into several categories, like marksman, mage, assassin, fighter, support, tank, and then for each category we train a weight for each item.

We tried both ideas, and found that for the first idea, some champions are played much so we have much data about them, and the training works well. But for some champions, their data are rare, so the weight are sometimes not trained well. For the second idea, it works pretty well because champions belong to the same category often get similar items. So to make our model more stable, we decide to use the second idea to manage the item sets.

After selecting suitable features and pre-training the item model, we try to train our data on the several machine learning models mentioned above, and we get different results on different models.

For SVM, the training process is faster but when using testing data on it, the accuracy is lower than the other two models, especially models at an early time spot. When dealing with models with late data, the model is more accurate than before, but still not so accurate when faced with a one-wave attack. That’s because this model may result into overfitting when data are similar to each other, and cannot deal with extreme occasions.

For decision tree, the accurate is a little higher than SVM, but still not satisfying. On occasions when one team has advantage on some aspects like dragons or team score but disadvantages on other aspects like team gold or towers, the model will predict one team very probable to win, but actually they are in balance. In other worlds, decision tree algorithm makes some features too important, and affects the final result.

For logistic regression and neural network, their accuracy is much higher and they can deal with many different types of occasions. Between them neural network has a higher accuracy and a longer train time, while its structure is totally generated by the machine learning framework. Based on this, we are still trying different types of neural networks, like LSTM, and we will choose one model with the highest accuracy of prediction and keep others for comparison.

However, even with the best model listed above, the accuracy is still the earlier the lower, and the accuracy at last time spot is almost accurate. This thing happens because data in early spots are often similar to each other, and it’s almost impossible to predict a winner using such data. Some new types of models may solve this problem and we are still working on it.

Conclusion

Using intermediate timeline data from LoL, we can successfully develop a model to predict which team has a higher win rate at each time spot according to his in-game features, such as team score, team gold, etc. Based on machine learning algorithms and models and using reasonable features, the accuracy can be pretty accurate, and the in-game data are instructional to players.

Distribution

Yingzi Zhu and Rongrui Lin take charge of data collecting and cleaning.

Yiwei Li and Yidong Liu take charge of algorithm selecting.

Nan Li and Pei Lyu take charge of webpage developing.

All team members contribute similar to the project, and all reports are distributed evenly.