

Summary of the machine learning techniques applied:

1. Logistic Regression
2. Logistic Regression + taking hero-combo into consideration
3. Logistic Regression + 2nd order Principle Component Analysis
4. Logistic Regression + Sorted Principle Component Analysis
5. Random Forest
6. Augmented Logistic Regression with genetic algorithms
7. K-nearest neighbors

Challenge:

Predicting winning side of the game purely based on hero lineups is simple if we only consider the hero appears in each side. However, many people tried to do that and found the predicting results are not satisfying. This is due to the proc and cons of each hero and co-effect the hero brings when some heroes appear in the same side. Besides, the existing mutual restraint between heroes also brings uncertainty in the predicting process. Therefore, it is necessary to take those effects into consideration.

How are the previous predicting results?

1. "Pure Logistic Regression" gives unsatisfying results.
2. "Logistic Regression + taking hero-combo into consideration" gives better results but the "hero-combo" is chosen by reporters. Only 50 combos are taken into consideration. That "hero-combo" is not complete and subjective.
3. "Logistic Regression + 2nd order Principle Component Analysis" also improves the predicting accuracy as the PCA brings comparison between heroes in two teams.
4. "Logistic Regression + Sorted Principle Component Analysis" also improves the predicting accuracy as the PCA brings comparison between heroes in two teams.
5. "Random Forest" gives average accuracy
6. "Augmented Logistic Regression with genetic algorithms" improves predicting accuracy by introducing the successful set. Successful set actually gives a set of heroes with relatively high winning rate.
7. "K-nearest neighbors" must compare the query match to every match in the training set and compute weights and probabilities, this process was quite slow.

What I want to do

1. I want to take the mutual restraint between heroes through implementing hero winning rate against each other.
2. I want to determine the weight each heroes in the team. This is because hero can be classified as "Carry", "Solo", "Offlane" and "Support". I don't think each role contributes the same amount to a winning game. For example, "Support" are important but their contribution to the victory is smaller than "Carry" because "Carry" becomes so powerful in the late game and normally most matches last long enough to be regarded as full match.