CS229 Project

Predicting the winning side of DotA2

Kuangyan Song, Tianyi Zhang, Chao Ma

INTRODUCTION

- A DotA 2 game has ten players with five on each side. Each player decide for himself which hero to use.
- Other than the experience level of the player. Lineups on either side has an influence on the result of the game.
- Certain combination of heroes are complimentary while some others are restrictive which might have an impact on the game.

DATA COLLECTION

- Match and player data are retrieved via APIs provided by the game developer.
- We only consider those games modes that is “All Pick”, ”Ranked All Pick” and ”Random Draft”. Moreover, those games that has absent players or has zero kills throughout the game are considered to be invalid.


```
"gold_spent": 11811,
"gold": 1188,
"deaths": 14,
"item_0_name": "Force Staff",
"hero_damage": 13478,
"last_hits": 100,
"item_1_name": "Phase Boots",
"player_slot": 2,
"denies": 4,
"ability_upgrades": [
    {
        "level": 1,
        "ability": 5132,
        "time": 740
    },
```

```
"dire_team_id": 0,
"match_id": 1932820205,
"start_time": 1447422304,
"match_seq_num": 1707808251,
"radiant_team_id": 0,
"players": [
    {
        "hero_id": 20,
        "account_id": 4294967295,
        "player_slot": 0
    },
    {
        "hero_id": 17,
        "account_id": 4294967295,
        "player_slot": 1
```


METHODOLOGY

Baseline

- Without taking the experience level of the players into account, consider the lineups as the only feature. Using logistic regression to fit the data.

$$\phi(X) = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_{112} \\ X_{113} \\ X_{114} \\ \vdots \\ X_{224} \end{pmatrix}$$

$$X_i = \begin{cases} 1, & a\ player\ of\ radiant\ side\ plays\ hero\ i \\ 0, & no\ player\ of\ radiant\ side\ plays\ hero\ i \end{cases}$$

$$X_i = \begin{cases} 1, & a\ player\ of\ dire\ side\ plays\ hero\ i-112 \\ 0, & no\ player\ of\ dire\ side\ plays\ hero\ i-112 \end{cases}$$

$$Y_j = \begin{cases} 1, & Radiant\ side\ won \\ 0, & Dire\ side\ won \end{cases}$$

$$min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{m} log(exp(-Y_i(\phi(x)_i^T w + c)) + 1)$$

Fold	1	2	3	4	5	6	7	8	9	10
ε_{train}	0.28	0.29	0.30	0.27	0.29	0.25	0.23	0.29	0.33	0.31
ε_{test}	0.47	0.43	0.49	0.42	0.48	0.47	0.42	0.46	0.43	0.46

- Both the training error and test error are high. Most of the data points are very close to the decision boundary, we might have too many features. The assumption that all the players are equal is also questionable.

Adding features for hero combo

- Before selecting features, we took the interaction of heroes into account. To simplify the problem, we hand picked fifty known significant hero combination to observe as features, now we have 324 features.

Fold	1	2	3	4	5	6	7	8	9	10
ε_{train}	0.25	0.26	0.23	0.28	0.25	0.25	0.24	0.27	0.29	0.28
ε_{test}	0.40	0.41	0.45	0.43	0.41	0.39	0.41	0.43	0.40	0.42

- Slightly better result but still not good enough.

FEATURE SELECTION

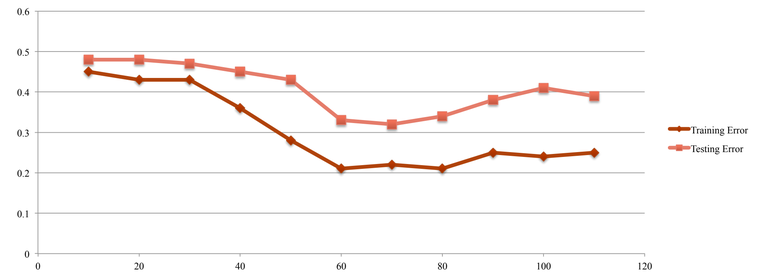
Since there are too many heroes, some of them might not have as significant effect as others on the result of matches. We decided to select features.

Algorithm 1 Stepwise regression

```
RequiredNumberOfFeatures ← some number
FeatureSet ← {1}
while NumOfFeatures ≠ RequiredNumberOfFeatures do
    Fit a model using FeatureSet
    Calculate ∂L(β)/∂β = Σ_{i=1}^{N} x_i(y_i − P(x_i; β))
    β_new ← arg max(|∂L(β)/∂β_i|)
    Add the feature that corresponding to β_new to feature set
    NumOfFeatures ← NumOfFeatures + 1
return FeatureSet
```

Two ways to decide whether we have the right number of features.

- Statistical threshold
- Cross Validation



Error rate v.s number of features

FUTURE WORK

According to our work, we can see that predicting the winning side based on hero selection is not enough. Future work should take the player into account to make the prediction more accurate. Moreover, the model might need to be changed to get a better result.