DotA 2: Predicting Matches

Petra Grutzik, Joe Higgins, Long Tran
pgrutzik@stanford.edu, joe7@stanford.edu, longtran@stanford.edu

Background

Motivation

Our project aims to uncover what distinguishes winners from losers in professional DotA 2 matches. Insights from our model about what strategies, teams, and players are more likely to win in a game of DotA 2 (and other similar electronic games) would transform how professional esports are both played and managed. So far in development, we have obtained accuracy comparable to human performance.
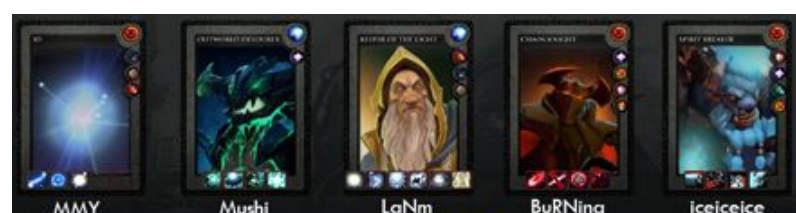
Problem Definition Predict the winner of a professional
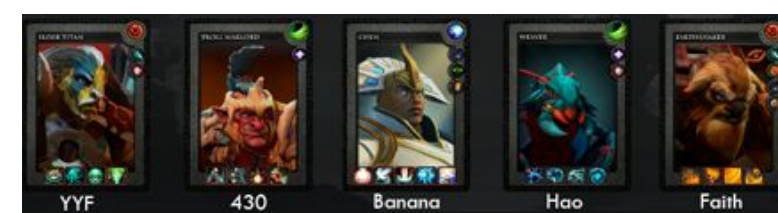
Dota2 match.

The Game

2 main phases

1. The Draft: 2 teams of 5 human players each take turns picking in-game characters ("heroes") that they will control during the gameplay phase

Radiant Team Picks   **Dire Team Picks**

2. The Gameplay: each team uses the heroes they picked in the draft phase to destroy the opposing team's base (in-game structures that belong to that team)

Data Set

- 47,440 professional Dota2 matches [1]
- Train & Dev data: 42,578 games before May 5, 2017
- Test data: 4,862 games after May 5, 2017

Feature Sets

1. **Team indicators:** 8587 binary indicator variables of which teams are playing in a given game. Ex: Team Secret is playing against Team EG, those two columns will have indicator variables set to 1 and the rest will be set to 0.
2. **Draft indicators:** 228 binary indicator variables of which in-game characters are being played and on which team (114 in-game characters x 2 teams to play on).
3. **Recent player statistics:** gold/min, xp/min, kills/min, lane efficiency, solo competitive rank (trailing 10 games)

Baseline prediction

Gosugamers.net bettors' collective prediction has a 62.8% accuracy rate [2]

Support Vector Machine

Configuration

$$Minimize\left(\left[\frac{1}{n} \sum_{i=1}^{n} \max\left(0, 1 - y_i(w \cdot x_i - b)\right) \right] + \lambda \|w\|^2 \right)$$

Results

Feature Set	Training Set Accuracy	Dev Set Accuracy
Team features only	72%	59%
Draft features only	58%	57%
Team & Draft features	72%	61%
Recent Player Stats & Draft Features	55%	55%

Neural Network

Configuration

- 2 hidden layers layers: (1) 4000 nodes (2) 2000 nodes Hidden layer activation function: ReLU

$$f(x) = x^+ = \max(0, x)$$

- Output layer activation function: Sigmoid function

$$S(t) = \frac{1}{1 + e^{-t}}$$

Results

Feature Set	Training Set Accuracy	Dev Set Accuracy
Team features only	67%	58%
Draft features only	68%	54%
Team & Draft features	69%	60%
Recent Player Stats & Draft Features	53%	53%

Analysis

The accuracy on the dev set for both the SVM and neural network model is much lower than the accuracy on the training set. The current iteration of our models may overfit our training set because we have ~50,000 matches but have ~8500+ possible combinations of variables.

Regularization

In an attempt to improve the dev set performance of our neural net model, we performed L2-regularization, early-stopping, and dropout.

Results

Team & Draft features	Training set accuracy	Dev set accuracy
Base model	68.6%	59.6%
L2 regularization (factor = 0.001)	68.5%	59.7%
Early stopping	68.6%	59.6%
Early stopping & L2 regularization (factor = 0.001)	68.3%	59.8%
Early stopping, L2 regularization (factor = 0.001), 20% hidden layer dropout	68.6%	58.8%
Early stopping, L2 regularization (factor = 0.001), 25% hidden layer dropout	68.6%	58.7%

Analysis

The improvement was limited, suggesting that the neural net setup remained prone to overfitting.

Test on SVM
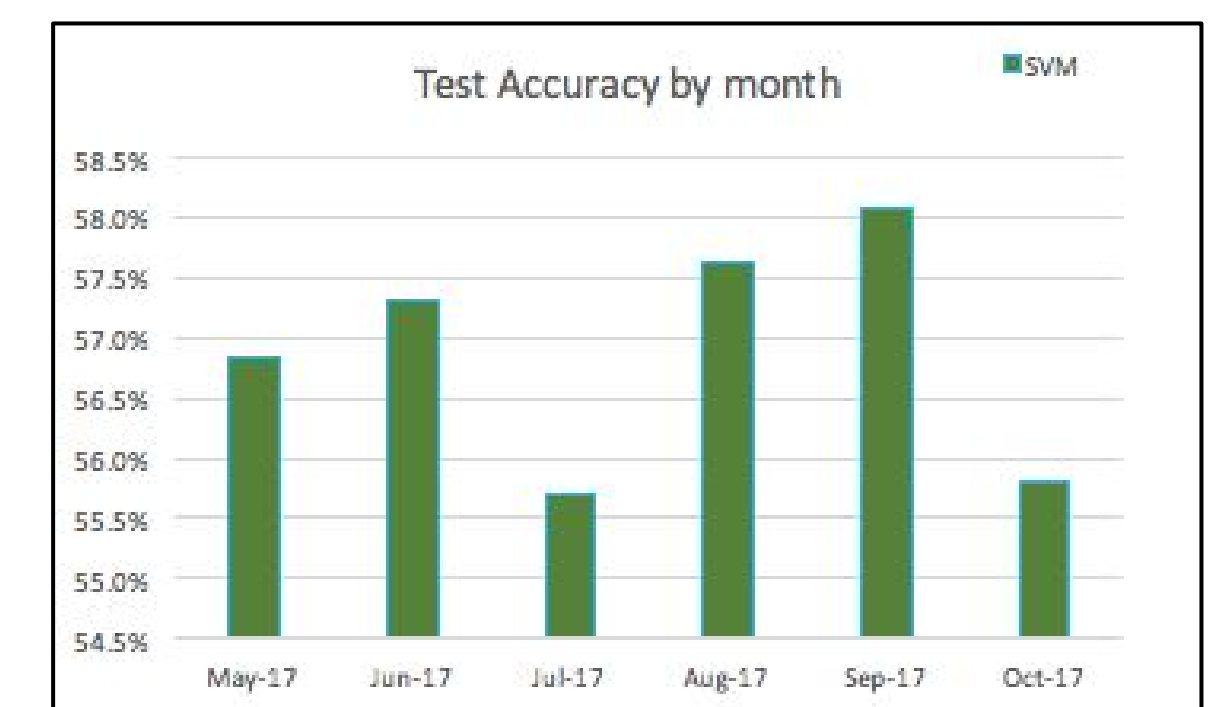
Next Steps

- Try different box score metrics and different trailing lengths
- Run a model to see which box score metrics were most predictive of victory in a game and use those as the trailing box score metrics
- Generate 2x more examples by viewing data from opposing side

1 OpenDota API (v17.5.0) https://docs.opendota.com/

2 Gosugamers.net has 20,100+ matches on record, 16,800+ of which have betting data.