Sports Handicapping ML

w207-5

Members:

Ji Zheng
Ye (Kevin) Pang
Jay Zuniga

Introduction

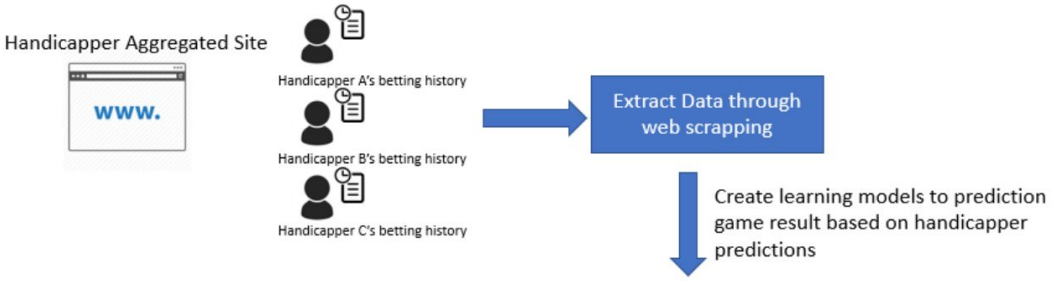
Sports betting is a big market in the USA as it combines two of America's passions, sports and gambling. Legalized sports betting in Nevada is "estimated at $4.5 billion" and that pales in comparison to the illegal betting market that is "estimated at $150 billion"[1]. As with any market, an ecosystem of 'experts' that provide their skill and opinions have emerged.

Handicapping refers to "the various methods by which spectators can predict and quantify the results of a sporting match. The term is applied to the practice of predicting the result of a competition, such as for the purpose of betting against the point spread"[2]. Handicappers are the people who predict results for various sports activities to assist people making bets. Sports handicapping has becoming a big business itself with many handicappers offering their predictions as a service to participants in the sports betting market. While there are a lot of solutions that try to predict the outcome of a sports match based on the qualities of the teams/players involved and other variables, we haven't yet seen anything that uses sports handicapper data. As sports handicappers already performed analytics on sports events, our project aims to use sports handicapper data to predict game outcomes by leveraging both their individual and swarm intelligence [3] and better understand how to interpret and act on handicapper's recommendations.

Methods and Data

Data

Our main input would be using historical predictions from known sports handicappers. This would provide data on what context he made his prediction in (details of the game). We would scrape this data from sports handicapper websites (e.g. https://www.sportschatexperts.com/).



	Home	Away	Juice	Handicapper A	Handicapper B	Handicapper C		Machine Learning		True Game Winner	ML Correct?
9/1/2017	Yankees	Boston	-100	Yankees	Yankees	Boston		Yankee		Yankee	Y
9/2/2017	Mets	Red	-120	Red	Red	Red		Red		Mets	N

We would then use handicapper data to make predictions on future games.

Methods

The group will consider multiple supervised learning algorithms to see what works best. However, at the top of our list, we will be looking at Decision trees since we think this works intuitively with finding the best handicapper prediction to use based on the details of a particular game. To improve our accuracy, we will be using Random Forests with our decision trees. We will also consider logistic regression and feedforward multilayer neural networks. We will use regularization to weed out weak handicappers.We will iterate and build multiple models with different combination of handcrafted features to test which one (or ensemble) gives the best performance.

References (or Works Cited)

1. SBNation, “The Supreme Court decision in New Jersey’s sports gambling case could be worth billions around the country”, SBNation link, Jun 28, 2017.
2. Wikipedia, “Handicapping”, Wikipedia Link, Mar 10, 2018
3. New Scientist, “How to beat bookies by turning their odds against them, New Scientist Link, Oct 17, 2017.