Gymnastic Outcomes

Abstract

When a gymnast wants to progress into the NCAA or the Olympics, she must pass through ten levels to make it to the elite. However identifying who will succeed, or which gyms promote success is not currently possible. We want to discover factors that indicate the likelihood that a gymnast will be in a pool of top candidates based on gymnastic competition scores throughout the levels.

Introduction

Many competitive sports have similar characteristics to gymnastics as an athlete builds a career. One important thing for all competitive athletes to know is which location is the best for advancement, as well as which places have lower injury rates. If the NCAA or the Olympic organization can help collect data on gyms or other training locations about injuries and competitive progress, it will build the sports as well as further athletes' longevity. Having data to rank training facilities will help guide the growth of new facilities.

Procedure and data

We will be looking at competition scores throughout the careers of female artistic gymnasts. What this data can tell us information about how long they took on each level, what average score they got per competition, and how they compared to other gymnasts on each apparatus. We will also determine which gyms produce the highest scoring gymnasts and those most likely to proceed to elite level. What we will use is http://www.meetscoresonline.com/ and as well as http://ec2-54-243-30-70.compute-1.amazonaws.com/pages/women/events/program_events.html which list competitions and also gymnasts with one or many competitions. We can see each gymnasium, though there will be a bit of a challenge as it is not a simple search to find which ones they attend. Using news articles and facebook pages, we should be able to figure that out.

Another big data source will be video, either live streamed or as a vlog. We can do text analysis of the transcripts of these videos to see if information about injuries is reported, including whether the athlete lost time, and how much. We can look at known gymnast injuries to get a baseline. It will be difficult going from voice/video to text to do some analysis. We will try a couple different approaches.

One approach is to find a online converter such as http://www.360converter.com/ among others to analyze for accuracy. If one proves sufficiently accurate we will use that, otherwise we will find some alternate

path. One idea is to see how good free speech to text programs are.

Another approach would be to find an algorithm either already created or create our own to do text analysis on the video.

Timeline

Milestone one:

Sub Part 1:

Create a database of gymnasts with name, level progression across time, average score per apparatus per year and level, and highest and lowest scores. We will track frequency of competition in weeks, days, months and years.

Sub Part 2:

Identify gymnasiums, and whether they participate in NCAA or Olympic training.

Milestone two:

Sub part 1:

Do analysis on online converters for videos. Look at past gymnast blogs like Shawn Johnson as well as at current gymnasts. Look at live streamed competitions.

Sub part 2:

Look at analysis from speech to text software.

Determine if this approach will be useful.

Milestone three:

Work on correlating outcomes of level , score and gym to to NCAA or Olympic performance.

Milestone four:

Work on some visualization on how gymnasts move through the levels and to different gyms.