CPSC 490

Stephen Slade

Math Reader: TBD (to be appointed by Math DUS)

12 September 2024

James Tyler

<div align="center">Project Proposal: Predictive Modelling of World Golf Rankings</div>

Introduction

Men's professional golf developed its first ranking system in 1986, called the Official

World Golf Rankings (OWGR). As of the past few years, the ranking has been the

unquestionable source of who are the best professional golfers in the world. The ranking works

by assigning points values to tournaments, based on some heuristics (such as the four major

championships - The Masters Tournament, The US Open, The British Open, and The PGA

Championship - assign 100 points to their winners). The number of points a player earns is

divided by the number of events they play (minimum of 40 divisor), with a half-life applied to

points earned in older tournaments. For non major championships, the points total would be

determined by the strength of the field. Each player participating in the tournament would

contribute points to the strength of the field (SOF) based on their OWGR ranking. The top

ranked player would contribute more than the second ranked player, etc.

In recent years, almost all of the top players have played primarily on the PGA Tour. This

means that given this SOF calculation, the PGA Tour events award more points to the top

finishers than any other tour. This had the effect that in order to be a top ranked player, you must

play on the PGA Tour. One effect of this fact is that top players from other tours, in particular the

European Tour, started to leave for the US to play on the PGA Tour, as there were more opportunities to earn ranking points, which are used as the primary metric for determining who qualifies for the major championships. As the top players left the European Tour, it had fewer financial resources, and has decreased in relevance in recent years. However, in 2022, a new tour started that began to rival the financial resources of the PGA Tour, called LIV Golf. LIV attracted many top PGA Tour players under guaranteed contracts to play on their tour, including top players Jon Rahm, Cam Smith, and Bryson DeChambeau, all of whom were ranked in the top 10 in the OWGR when they left for LIV. The OWGR chose not to award ranking points to LIV tournaments. Over the past two seasons, many of the top LIV players have remained some of the best golfers in the world, however their OWGR rankings have plummeted since they can only receive ranking pionts in the majors.

Modelling Challenge

The challenge presented is that it is necessary to compare the skill level of players who mostly do not compete against each other, but rather in two separate settings, with a limited overlap (specifically the four major championships). One common approach to this problem, as described in a blog post on datagolf.com, is to estimate the strokes gained by a given player. Strokes gained is defined as the performance in a given round (or even a given shot, as detailed by Mark Broadie in his seminal work on the subject in 2011) against a baseline player (normally defined as an average PGA Tour player. This approach allows us to compare rounds on different days and in different locations. This amounts to estimating a regression based on shot or round level data from PGA Tour and other tour events. For this approach, I am interested in testing

whether more advanced models, such as a Gradient Boosted Tree or a Neural Network, will outperform a linear model in this setting. I am also interested in developing a version of the points based model that better handles the issue of separate tours. I also want to test a version of the model that merges a strokes gained approach with a points based approach, as both approaches seem to have merit.

Timeline

Week 3 - Project Proposal Due

Week 4/5 - 1. Data Acquisition: Write Code for a Webscraper to gather results from tournaments. Get this data into a usable format. 2. Read and take notes on past work done on this problem.

Week 6/7 - Build Pipeline for recreating Linear Model based Approach and Points based approach. Define baseline model and evaluation function for future models. Prepare Midterm report.

Week 8/9 - Test GBT and Neural Net. Learn about the math of these approaches. Brainstorm any novel approaches to test.

Week 10+ - Flesh out any ideas that emerge from the exploratory part of project. Aggregate results. Clean code and write up project for Final Report

Math Component of Project

The mathematical component of the project will consist of a precise formulation and derivation of statistical methods used in the research. For instance, I will revise and present motivation behind linear regression, gradient boosted trees, and neural networks.