



Predictive Modelling of World Golf Rankings

James Tyler, Computer Science and Mathematics, Stephen Slade, Computer Science, Yale University, Tamunonye Cheetham-West, Mathematics, Yale University

Introduction

Official World Golf Ranking (OWGR) uses a points based system, assigning events a point total, and awarding points based on finish position. This system fails when not all tours receive ranking points.

Another way the OWGR is limited is that it assigns points in a fixed proportion based off finish, not taking into account individual scoring differentials, or any other predictors.

There have been attempts to extend the statistical methods used to construct a world golf ranking, notably Data Golf, which uses a regression based framework to generate rankings.

This project seeks to apply even further statistical modelling techniques in order to test their effectiveness in predicting golfer performance.

Key Findings

- > Built 2 web scrapers to gather data from OWGR and LIV Golf websites, covering over 2 million tournament rounds
- > Backfilled missing data from OWGR to build baseline model to compare against.
- > Implemented regression to determine course and field independent adjusted score metric.
- > Trained models on 10 years of data, and tested out of sample performance on 4 most recent years of data.
- > Tested machine learning frameworks that previously had not been used on this problem.

Background / Related Work

The most well known result of advanced golf analytics is the development of Strokes Gained Analysis (Mark Broadie, 2014). Strokes gained evaluates golfer performance based on shot level data and constructing a player performance rating in units of strokes differential to the field average in the areas of driving, putting, chipping, and approach.

While shot level data was not available for all tournaments (only applies to PGA Tour), the concept is useful in constructing a rating of a player as a stroke differential against a baseline player (for most purposes, can define this as an average PGA tour player).

Work by Data Golf and a 2008 paper by Connolly and Rendleman shows that golfer performance can be autocorrelated (streaky), and depend on luck (high variance of score distribution). These effects are not fully encapsulated in the traditional OWGR approach.

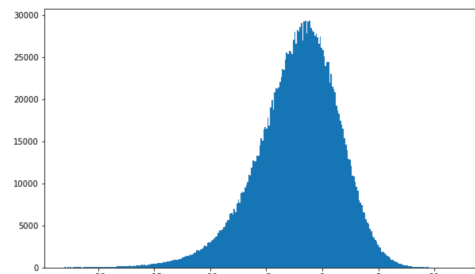
Methodology

Data split into train, test, and validation sets, spanning to 2010 (beginning of OWGR data including round score reporting).

Need to determine for each round in dataset the adjusted strokes gained of that round (depends on field strength, i.e., difficulty of outperforming the average score). Accomplished using rolling regression over player and round dummy matrix, which produces an estimate of golfer skill and an estimate of round difficulty.

Raw Strokes Gained - Round Difficulty becomes adjusted strokes gained, which is the main predictor used in model fitting, along with variations, such as EWMA of adjusted strokes gained with varying half lives, to capture resilient vs. streaky play.

Time series of adjusted strokes gained is the predicted y variable of all models we fit.



Distribution of round adjusted strokes gained is left skewed, where the mean is less than 0, since the baseline is set to PGA Tour player, which is better than an average player.

Results

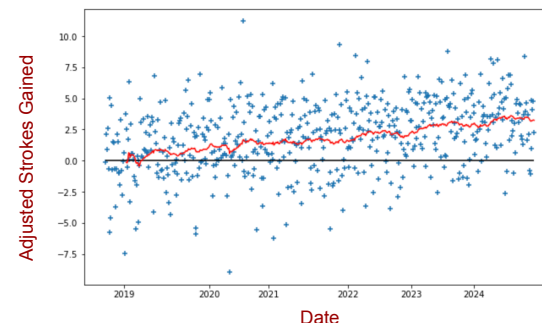
Calculated Performance Rating for every golfer in the dataset using multiple models, testing out of sample performance to predict round level adjusted strokes gained.

Used this analysis to construct a world ranking, based on Performance Rating produced by the model.

Gradient Boosted Trees achieved higher out of sample R^2 than regression model when predicting adjusted strokes gained.

Reconstructed OWGR to include LIV Golf events by simulating if LIV Golf events had earned ranking points when players had joined the new tour.

Performance Rating of Scottie Scheffler, World #1



Top 10 World Golf Ranking, according to model following 2024 US Open

	Player	Performance Rating
Ranking		
1	Scottie Scheffler	3.230968
2	Xander Schauffele	2.977069
3	Collin Morikawa	2.398464
4	Ludvig Aberg	2.199778
5	Rory McIlroy	2.176697
6	Tyrrell Hatton	2.101590
7	Hideki Matsuyama	2.055887
8	Bryson DeChambeau	1.985068
9	Sam Burns	1.934057
10	Tommy Fleetwood	1.775099

Conclusion / Future Work

Expand the number of models tested to include Neural Networks (validation set still reserved for this purpose).

Compare the performance of OWGR system versus model's predicted rankings. Question: is the OWGR field strength system accurately reflecting performance in alternate tours' events.

Using the results of these models, build a simulation framework to predict win probabilities of golfers in events.

Test for additional predictors, including isolated strokes gained data, course fit, and player health data, which are not specifically addressed by this model.