Improving the Strokes Gained Statistics:
A new ranking system to access performance on the PGA Tour

Adam Levin

Recent advances in the quality and quantity of the data collected by the PGA Tour has led to an explosion of new research in the field of Golf Science. Mark Broadie is credited with inventing the Strokes Gained statistic, which measures the quality of a given golf shot by estimating the difficulty of the starting and ending position. This paper improves both the Strokes Gained measurement for individual shots and the statistics used to rank golfers based on their performances in various aspects of the game.

Data was obtained through the PGA Tour ShotLink Intelligence Program, which provides their data to researchers upon request. The data provides detailed data on approximately 15 million shots from all tour events starting in 2003. The data is first cleaned and prepared for analysis. Then a model for difficulty of a shot is built from the data. Difficulty is taken to be the number of shots recorded from a given location by an anonymous golfer. The model is a pipeline of two models; it takes into account distance, turf (fairway, bunker, etc.), angle of approach to the green, and slope. Superior accuracy of the model is demonstrated by comparing the currently used Strokes Gained baseline to the improved model.

The problem of accessing the performance of the golfers using these statistics is discussed with regard to the bias-variance trade off that exists. The current method of calculating the Strokes Gained To The Field is acknowledged and critiqued. This critique motivates a new approach to the problem of accessing the skill of a player at a given time. This approach involves accessing the golfers' performance relative to one another: the improved Strokes Gained statistics for shots by different players on the same hole, on the same day, and on the same turf are taken to be observations comparing the two golfers' skills. These observations are loaded into an Adjacency Matrix, with which a Network Ranking System first proposed by Park and Newman (2005) is utilized to rate the golfers skill. The original system is adapted to handle arbitrary strengths of comparisons between two players, and changing skills over time.

Finally, the new statistics are utilized in a forward-facing prediction problem with the dependent variable being the future performance of a golfer in tournaments. The relative importance of the various aspects of the game generally confirms the results of Broadie (2008), finding that the long game is relatively more important in predicting the future performance of the players than the short game. The proposed method of accessing the skill of golfers is compared to the existing method with regards to predictive accuracy and better accuracy is demonstrated.

This work can be seen as a more accurate barometer of golfers' skill levels on the PGA Tour. Python was used for the analysis and all code is available on Github.

490 words.

Category	Correlation with Future Success New Method*	Lower 95% C.I.	Upper 95% C.I.	Correlation with Future Success Strokes Gained to Field Rolling Weighted Average	Lower 95% C.I.	Upper 95% C.I.
Tee Shots Par 3s	**0.17**	0.18	0.16	**0.13**	0.14	0.12
Tee Shots Par 4 and 5s	**0.18**	0.19	0.17	**0.14**	0.15	0.13
Putts 0-5 ft.	**0.08**	0.08	0.07	**0.07**	0.07	0.06
Putts 5-10 ft.	**0.05**	0.05	0.04	**0.04**	0.04	0.03
Putts 10-20 ft.	**0.05**	0.06	0.05	**0.05**	0.05	0.04
Putts 20+ ft.	**0.07**	0.08	0.07	**0.07**	0.08	0.06
Rough 0-30 yds.	**0.08**	0.08	0.07	**0.06**	0.07	0.06
Rough 30-125 yds.	**0.06**	0.07	0.06	**0.06**	0.07	0.05
Rough 125+ yds.	**0.13**	0.13	0.12	**0.11**	0.12	0.10
Fairway 0-100 yds.	**0.09**	0.10	0.08	**0.08**	0.08	0.07
Fairway 100-180 yds.	**0.16**	0.17	0.15	**0.13**	0.14	0.12
Fairway 180+ yds.	**0.15**	0.15	0.14	**0.13**	0.14	0.12
Bunker play	**0.08**	0.09	0.08	**0.07**	0.08	0.06

*Pearson Correlation Coefficients for 67,080 Player-Tournament Observations. More tuning is underway to find the best parameters for the new method.