DEVELOPING STATE-BASED RECOMMENDATION SYSTEMS FOR GOLF TRAINING

Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Systems Engineering

By
Jacob Ziller

April 28, 2020

Technical Project Team Members
Alanna Flores
Orlando Jimenez
Christopher Kaylor
Kelly Rohrer

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR
William T. Scherer, Department of Engineering Systems and Environment

Developing State-Based Recommendation Systems for Golf Training

Kelly Rohrer, Christopher Kaylor, Jacob Ziller, Orlando Jimenez, Alanna Flores
Engineering Systems and Environment Department
School of Engineering and Applied Science at the University of Virginia

Abstract—**The NBA, MLB, NFL and other professional leagues utilize sports analytics, but the potential of professional golf analytics is largely untapped. Instead of using data-driven methods connecting practice to tournament performance, training regimens are often based on conventional wisdom. How can data be used to recommend training regimens for golfers to improve performance? We partnered with golf analytics company, GameForge, to develop tools and methods for golf analytics to capture these markets, including the development of a state-based training recommendation system. We used Gameforge, PGA, and LPGA data to build markov models using k-means clustering, and linear models. These two model types form the basis of our recommendation system. In the future, these methods can be used to inform training decisions, particularly as more data is collected.**

Keywords—golf, Markov, k-means, sports analytics, PGA, LPGA

I. Introduction

The analytics industry has completely changed the sports landscape in the past decade as the NBA, NFL, MLB and other professional leagues leverage analytics to increase revenue, improve player performance, enhance team quality, or prevent injury [1][2][3]. However, the potential of professional golf analytics has yet to be tapped. According to the National Golf Foundation, there are as many as 24.2 million golfers in the United States at both the recreational and professional levels [4]. While the PGA collects data, it is not analyzed and the statistics of elite amateurs and recreational golfers remain largely uncollected. These large sources of data present an opporrtunity to bring analytics to the forefront of the golf community, with a range of applications related to player performance available [5].

Alongside GameForge Golf, a golf analytics company, researchers created methods for golf analytics, including the development of a data-driven training recommendation system [6]. GameForge collects separate data from the PGA TOUR describing clients' tournament performance and performance on GameForge drills. By integrating analysis performed on GameForge and professional data, models suggesting specific drills allowing golfers to transition over a season to improved performance states given their current level of play can be built. The goal of this research is to provide GameForge with a methodology for improving their drill recommendation system.

Our first hypothesis, Phase 1, is that drill performance will correlate with on-course performance. The second, Phase 2, is that we can assess which skills cause changes in performance.

A. Phase 1

In order to build an effective training system, there must be a deeper understanding of the drills' influence on scoring and how players transition into stronger or weaker golfers. We expected drills would improve a player's score in a related skill category – for example, putting drills would improve a player's putting during a round of play. The longevity of a drill's effect on a skill was analyzed to understand the frequency and duration a drill should be performed to achieve performance goals. We expected players to be clustered based off of their performance and show clear transitions from one cluster to another. We expected skill improvements to be a common catalyst in the transition from one cluster of player type to another. Considering different types of players at different skill levels, it is expected that certain drills will prove more helpful for each circumstance. The next step was to integrate the findings between both data sets to recommend a drill that would improve a specific characteristic of a player's game, which would then transition them to a better state of player type.

B. Phase 2

Phase 2 involved understanding how golfers transition into better players and identifying which skills are most indicative of each transition. PGA and LPGA TOUR round data were used to analyze how players go from a lower-level player to an elite player. Describing the states of each player type, transitions from state to state, and the catalysts for these transitions provides GameForge the models and informative statistics to properly train their users. Using a larger amount of data from professional golfers allows us to dive deeper into analysis and find more concrete results. This phase answers questions like what kind of players can transition from the bottom tier of the TOUR into a tier of top-ten players, which skills are most important for that transition, and how likely that transition is for a particular golfer.

Moving forward, the methodology and findings provided by Phases 1 and 2 will give GameForge a robust framework as GameForge users input more data. More data will bridge the connection between GameForge drills and rounds played grows, increasing the accuracy and specificity of findings.

II. Data and Associated issues

Two unique data sources define Phases 1 and 2 of analysis.

A. Phase 1

The first source of data was user input into the GameForge site. Data values are proprietary metrics created by GameForge to capture performance on drills and rounds, stored in two separate datasets dating from 2017 to the present. The drill dataset includes classification of each drill by skill and game category. All 2,000+ drills can be placed in one of 73 skill categories which then map into one of 11 game categories.

Early entries in the round data set included fewer variables as GameForge began tracking new variables over time. To keep data consistent, researchers narrowed the round dataset from

Research sponsored by Ankura Consulting Group, LLC.

140,000 rows to 850 rows, only including entries with all variables. This was one of the primary limitations of the dataset; cleansing reduced the data to a fraction of its original size. A second limitation of this data was the lack of overlap between the drill and round datasets. In many cases, drill and round data would be recorded by players in alternating rather than concurrent intervals. This became problematic for the linear models developed in Phase 1 of analysis; there were few cases in which a group of drill entries were followed immediately by a round entry made by the same player [7]. This made it hard to use the drill data as predictors for round performance. The sparse nature of the round data also impacted analysis for the Markov models [8]. To create a more robust model, rather than creating a new state for each core score (a GameForge metric similar to Relative to Par Score (RTP Score)), core scores were binned so that the model had fewer states with more data in each state [9]. Data issues were a primary driver in model iteration during Phase 1 and necessitated the need for analysis during Phase 2.

B. *Phase 2*

The second source of data was ShotLink data collected and made publicly available by the PGA and LPGA TOUR's[10].

Analysis of PGA TOUR data used the event level dataset. The dataset spans from 1983 to 2019 and includes 201 total variables. After an initial cleaning, a new dataset was created containing 133 common variables and spanning from 1999 to 2019. The data was then pared to 16 relevant variables identified by GameForge. Before analysis, the data was aggregated so that each entry contains average values for a single player in a single year. To supplement this data the PGA TOUR website was scraped to find world ranking for each player each year, a variable not included originally. To maintain consistency, the dataset was shortened to the most recent 20 years. Another issue was the type of data stored. The GameForge site collects different metrics than the historical PGA TOUR data. These metrics are chosen based on analysis by the GameForge team. Lastly, historical PGA TOUR data only includes professional golfers, while GameForge data includes both professionals and amateurs. In order to apply conclusions formed from this part of the analysis to the GameForge users, researchers assumed professionals and amateurs are from the same population.

The LPGA dataset used in analysis spans from 1992 to 2019 and contains 19 distinct variables, including a money rank variable. The data contains entries for each player every year. The money rank category was binned to create more data in each category. The primary issues with this dataset are again the type of data stored and the assumption that the LPGA players are from the same population as GameForge's amateur players.

For both the PGA TOUR and LPGA data, a Markov model was used to represent player transitions across years. The data only had enough transitions to make a 2^{nd} order model [11].

III. Models

A. *Phase 1: Markov Models*

Four iterations of Markov models were used to describe how players either worsen or improve between rounds. State definition is critical for the success of these models [12]. These models used a subset of the GameForge round data, only including rounds with either RTP Scores or Core Scores of -5 to +5. RTP Score is a common metric in golf describing a player's score in relation to the expected score (or "par") on a course [13]. Core Score is calculated by subtracting a player's birdies from their bogeys during a round, reducing the noise from double bogeys or eagles. The first model categorized transitions between RTP Scores (used as states), analyzing the frequency with which players shoot specific scores following a given prior round. The second model was nearly identical, but used Core Score as the defined states instead of RTP Scores. Sparseness of data necessitated a third model where the 11 separate Core Score states were grouped into 5 Core Score bins, decreasing the number of possible transitions from 121 to 25, and increasing the observed instances of each possible transition. Finally, continuing to use Core Score bins as states, separate Markov models were created for instances of players who performed specific drills between two rounds and those who did not perform the specified drill, potentially revealing different frequencies of transition depending on drill performance.

B. *Phase 1: Linear Models*

Phase 1 also included five candidate linear models, summarized in Table I [14]. Linear models would ideally use specific drills as independent variables, but sparse data necessitated the use of skill or game categories instead. In candidate model 1, no lookback period was defined for independent variables, violating the independence property. To maintain this property, candidate model 2 used only a player's most recent round – each player's performance on drills was only associated with one data entry. Candidate model 3 used putting game stats as the dependent variable. RTP score and core score were used as independent variables. Candidate model 4, named the delta model, aimed to capture the decaying impact of a drill on a golfer's performance in a round by observing the change in players performance. Candidate model 5 was a delta model aggregating average drill score and frequency variables over a game category rather than a skill category.

TABLE I. Summary of Linear Models

Candidate Model	Independent Variable(s)	Dependent Variable(s)
Model 1	average score of drills performed in a skill category, frequency of drills performed in a skill category	core score, relative to par score
Model 2	average score of drills performed in a skill category, frequency of drills performed in a skill category	core score, relative to par score from most recent round
Model 3	average score of drills performed in a skill category, frequency of drills performed in a skill category, relative to par score, core score	putting game statistics
Model 4	difference in round variables for a golfer in a given two-week time period, the score & frequency of drills done in a given skill category during same two-week period	change in core score, change in relative to par score
Model 5	difference in round variables for a golfer in a given two-week time period, the score & frequency of drills done in a given game category during the same two-week period	change in core score, change in relative to par score

C. *Phase 1: Feature Selection Model*

Third, researchers used a feature selection model called the Extra Trees Classifier, a subset of the Random Forest Classifier [15]. The Extra Trees Classifier is an algorithm aggregating the results of multiple de-correlated decision trees in a random forest to output its classification result [16]. Using this ensemble method helped determine which features for each RTP Score cluster of players were most significant. Knowing how important each feature was in our model is vital to understanding how the model predicts, allowing easier explanation of the model. Researchers ranked features of most importance for each cluster and identified how features changed between clusters.

D. *Phase 2: LPGA Models*

Two Markov models were made to describe how individual LPGA professionals transition between money ranks from year to year. Money ranks were binned into 5 categories: Positions 1-10, 11-30, 31-50, 51-80, and Outside the Top 80. Using these bins as states, both a 1st order and 2nd order Markov model were created. The 1st order model only considered a player's previous season as a prior state, while the 2nd order model considered a player's previous two seasons as a prior state – increasing the number of prior states from 5 to 25 and creating thinner data in the 2nd order model. For example, the 1st order model would consider a player who went from 31-50, to 11-30, and into 1-10 the same as a player who transitioned from 51-80, to 11-30, and into 1-10, while the 2nd order model would have considered these two players to have been in different prior states.

E. *Phase 2: PGA Models*

A 1st order Markov model was used to describe player transitions from year to year. Using K-means clustering, initially with the Hartigan-Wong algorithm and Euclidean distance method, players were clustered into four unique groups based on their year-long performance [17]. The 1st order model was made up of 4 states, one for each cluster, but did not hold the Markovian property for most transitions, necessitating a 2nd order model. This model, like the 2nd order LPGA model, took into account a player's previous two seasons as a prior state. To create the 2nd order Markov model, both the Lloyd-Forgy and Hartigan-Wong algorithms, as well as both the Euclidean and Manhattan distance methods, were used. Three overall combinations were tested, each of which produced a model with roughly two-thirds of transitions maintaining the Markovian property [18]. Further analysis of the 2nd order model exclusively used results from the Hartigan-Wong algorithm and Euclidean distance method. Because the states were based on clusters rather than rankings, researchers analyzed players' performance in each cluster. Comparisons of cluster were made along both world-ranking and performance in input variables.

IV. Results

A. *Phase 1: Markov Models*

TABLE II. RTP Score Markov Transition Matrix

		AFTER										
		-5	-4	-3	-2	-1	0	1	2	3	4	5
BEFORE	-5	7.73%	13.81%	12.15%	15.47%	10.50%	13.26%	7.73%	9.39%	5.52%	2.21%	2.21%
	-4	5.94%	7.92%	13.86%	11.88%	11.88%	12.87%	14.52%	5.94%	6.60%	3.96%	4.62%
	-3	6.68%	8.91%	12.38%	15.10%	9.16%	13.37%	8.17%	10.40%	8.66%	5.20%	1.98%
	-2	4.55%	8.06%	9.98%	11.38%	12.61%	13.84%	12.26%	9.81%	8.23%	5.25%	4.03%
	-1	5.09%	6.40%	8.21%	11.33%	14.12%	11.00%	13.14%	9.20%	8.87%	8.54%	4.11%
	0	3.07%	6.42%	6.13%	10.51%	12.99%	11.97%	12.85%	10.22%	10.36%	7.88%	7.59%
	1	2.02%	4.18%	5.12%	9.97%	11.05%	14.15%	13.75%	11.73%	11.99%	9.84%	6.20%
	2	1.88%	3.61%	4.70%	8.62%	9.56%	11.91%	13.64%	11.60%	13.64%	12.23%	8.62%
	3	1.32%	2.31%	6.26%	7.74%	8.07%	12.19%	13.34%	14.17%	7.74%	15.98%	10.87%
	4	0.89%	2.86%	3.58%	7.16%	8.94%	9.30%	14.67%	16.28%	11.81%	12.70%	11.81%
	5	0.92%	1.38%	2.30%	5.29%	6.44%	9.89%	14.25%	14.02%	16.78%	16.09%	12.64%

Table II shows the Markov Transition Matrix using 11 RTP Scores as states. The matrix shows the likelihood of shooting a particular score in the "After" round given a specific score in the "Before" round. For example, if a player shot -1 "Before", he or she would have a 14.12% chance of repeating that score in the next round, but only an 11.33% chance of improving to -2.

TABLE III. Core Score Bins Markov Transition Matrix

		AFTER				
		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
BEFORE	Bin 1	17.35%	30.41%	37.69%	12.69%	1.87%
	Bin 2	13.33%	27.54%	40.06%	14.46%	4.60%
	Bin 3	8.33%	20.92%	43.58%	20.13%	7.04%
	Bin 4	4.15%	17.05%	44.88%	25.00%	8.92%
	Bin 5	5.07%	13.77%	39.61%	26.57%	14.98%

Legend	
Bin 1	-4 or less than
Bin 2	[-3, -2]
Bin 3	[-1, 0, +1]
Bin 4	[+2 ,+3]
Bin 5	+4 or greater than

Table III uses the 5 Core Score bins as states, which resulted in more robust findings because of the reduced dimensionality. The states in Table II are defined differently than in Table I, but the matrix reads the same. For example, a player who achieved a Core Score of -3 or -2 in the before round would have a 27.54% chance of repeating a Core Score of -3 or -2 in the after round.

TABLE IV. Core Score Bin Markov Transition Matrix: Drill in Game Category 5 Performed

		AFTER				
		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
BEFORE	Bin 1	42.86%	28.57%	14.29%	14.29%	0.00%
	Bin 2	0.00%	36.84%	31.58%	26.32%	5.26%
	Bin 3	11.29%	14.52%	50.00%	22.58%	1.61%
	Bin 4	3.45%	13.79%	51.72%	20.69%	10.34%
	Bin 5	7.14%	14.29%	28.57%	14.29%	35.71%

*Drill in *Category 5* **was** performed between before & after round

TABLE V. Core Score Bin Markov Transition Matrix: Drill in Game Category 5 Not Performed

		AFTER				
		Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
BEFORE	Bin 1	17.01%	30.43%	38.00%	12.67%	1.89%
	Bin 2	13.54%	27.40%	40.20%	14.27%	4.59%
	Bin 3	8.25%	21.09%	43.41%	20.07%	7.19%
	Bin 4	4.17%	17.14%	44.70%	25.11%	8.88%
	Bin 5	5.00%	13.75%	40.00%	27.00%	14.25%

*Drill in *Category 5* **was not** performed between before & after round

TABLE VI. Core Score Bin Markov Transition Matrix: Summary Table of Tables III & IV

	Game Category 5 Drills	
	Drill Performed	Drill Not Performed
Improved	34.35%	34.37%
Constant	39.69%	31.75%
Worsened	25.95%	33.87%

The models in Tables II and III were largely descriptive, while the models in Tables IV, V, and VI are more analytical. Separate transition matrices were created depending on if a drill from a specific game category was performed. Of the 6 game categories tested (as the 5 other categories did not have enough data), only drills in Game Category 5 showed statistical significance. Performing a drill in Game Category 5 between two consecutive rounds made players both more likely to stay in the same Core Score bin and less likely to move to a worse bin.

B. Phase 1: Linear Models

Candidate model 1 was a step-model predicting RTP Score from an interaction model with all variables. This model had an adjusted R-squared of 0.089, but violated the independence assumption of linear models. If a player had multiple rounds, the same drills were used as predictors for multiple rounds.

Candidate model 2 improved on model 1 by using each player in the data only once. The version of candidate model 2 predicting Core Score had an adjusted R-squared value of 0.059; the version predicting RTP Score had a value of 0.053.

To improve predictive value, candidate model 3 used game statistics as the independent variable. The version of candidate model 3 predicting total putts had an adjusted R-squared value of 0.036; the version predicting greens-in-regulation (GIR) had a value of 0.08. These values, and a lack of normality in candidate model 3, led to the creation of candidate model 4.

Candidate model 4, used the algorithm shown in Figure 1:

1. Separate drill and round data by Game Category
2. For each player, select a short time period over which he/she has performed many drills for a specific skill category.
3. Find the rounds performed immediately before and after drills.
4. Calculate:
a. The number of days between the first and last drills.
b. The number of drills performed in the skill category for the player in the date range.
c. The change in RTP and Core Score between the two rounds in Step 3.
d. The change in score for some of the round variables between the two rounds in Step 3.

Fig. 1. Algorithm used to construct Candidate Model 4.

This process was performed manually and produced too few data points for a model, so we abstracted further by applying the same method at the game category level for candidate model 5. This model had a higher adjusted r-squared value and a distribution that was closer to normal.

Candidate model 5 was our best linear model, as it produced both the highest adjusted R-squared value and met the assumptions for linear modeling best, despite having just 50 data points. This version of candidate model 5 was a step model predicting delta core score. The statistically significant predictors were time span, frequency, delta one putts, dbombs, and drills per day. Delta relative to par score was not included as a predictor. The adjusted R-squared value was 0.295, meaning almost 30 percent of variability in the change in core score of players was explained by our model.

When more data is available at the skill and drill level, this method could be applied to create a more specific delta model. At the drill level, this model would help determine which drills improve core scores, and better training regimens could be determined to better improve a specific area of a player's game.

C. Phase 1: Feature Selection Model

In three unique clusters – rounds with RTP Scores of -5, even par, and +5 –features of importance were identified. In the +5 cluster (denoting worst performance), the five features most important were Par-5 Scoring, Up & Downs, 1-putts, 3-putts and P6 percentage. In the even par cluster, the five features were 1-putts, GIR's, In-Positions, P6 percentage, and Effective-Green Conversions. In the -5 cluster, the most important features were Par-5 Scoring, 1-putts, Bombs, In-Position conversions, and Fairways. Identifying features relevant to specific levels of play provides a potential pathway for players to improve RTP Scores.

D. Phase 2: LPGA Models

The model in Table VII characterizes the frequency of transitions on the LPGA TOUR – 78.83% of players Outside the Top 80 do not move into the Top 80, and the best chance to become a Top 10 player is to have already been a successful Top 10 (55.85% chance), or 11-30 (29.81%), player. The nature of transition changes depending on which two states a player is transitioning between. Certain statistics, such as driving accuracy or putting average, significantly improve when moving from Outside the Top 80 into the 51-80 bracket, and significantly worsen when making the opposite transition. These same statistics do not significantly change when moving between more elite money brackets. Changes in GIR are significant regardless of which brackets players move between.

TABLE VII. 1ST ORDER LPGA MARKOV TRANSITION MATRIX

		Next Season				
		Top 10	11 to 30	31 to 50	51 to 80	Outside Top 80
Prior Season	Top 10	55.85%	29.81%	9.06%	4.15%	1.13%
	11 to 30	12.71%	39.63%	23.36%	16.82%	7.48%
	31 to 50	5.50%	23.15%	23.72%	27.70%	19.92%
	51 to 80	0.26%	7.99%	17.82%	30.28%	43.64%
	Outside Top 80	0.39%	2.07%	4.47%	14.24%	78.83%

While data in Table VIII was sparse, conclusions can still be drawn. For example, the matrix says a golfer is more likely to remain in the Top 10 if they had spent the previous two years there than if they had spent one prior year in the Top 10. This also holds for golfers Outside the Top 80. Preliminary chi-square results show this model could be a better fit to the GameForge data, but as more years of LPGA results accrue it will be easier to tell which model is optimal.

TABLE VIII. 2ND ORDER LPGA MARKOV TRANSITION MATRIX

		Next State				
		1	2	3	4	5
Prior Two States	11	66.67%	24.82%	5.67%	1.42%	1.42%
	21	43.75%	32.81%	12.50%	9.38%	1.56%
	31	34.48%	37.93%	20.69%	6.90%	0%
	41	0%	100.00%	0%	0%	0%
	51	16.67%	50.00%	16.67%	0%	16.67%
	12	20.27%	43.24%	18.92%	10.81%	6.76%
	22	15.84%	42.08%	23.76%	14.85%	3.47%
	32	8.47%	33.05%	27.12%	21.19%	10.17%
	42	3.39%	38.98%	27.12%	16.95%	13.56%
	52	0%	33.33%	20.00%	26.67%	20.00%
	13	16.67%	16.67%	12.50%	29.17%	25.00%
	23	8.33%	27.50%	24.17%	25.00%	15.00%
	33	4.24%	22.88%	27.97%	31.36%	13.56%
	43	2.31%	20.00%	24.62%	32.31%	20.77%
	53	0%	18.46%	18.46%	26.15%	36.92%
	14	0%	30.00%	10.00%	30.00%	30.00%
	24	1.16%	13.95%	23.26%	27.91%	33.72%
	34	0.75%	8.96%	16.42%	35.07%	38.81%
	44	0%	7.44%	16.74%	30.23%	45.58%
	54	0%	3.88%	16.50%	28.64%	50.97%
	15	33.33%	0%	33.33%	33.33%	0%
	25	0%	10.81%	5.41%	24.32%	59.46%
	35	0%	3.13%	11.46%	18.75%	66.67%
	45	0.38%	2.63%	6.77%	17.67%	72.56%
	55	0%	0.76%	2.77%	11.98%	84.49%

Legend	
Code	Place
1	Top 10
2	11-30
3	31-50
4	51-80
5	80 +

E. Phase 2: PGA Models

Aggregated data described in the data section was clustered into 4 groups based on 16 relevant variables. The average world-ranking of each cluster is shown in Table X. Table IX describes the standardized performance of each cluster in each of the 16 variables. On average, players in cluster 1 perform the best, while players in cluster 4 perform the worst. Players in clusters 2 and 3 perform better than cluster 4 and worse than cluster 1.

TABLE IX. STANDARDIZED PERFORMANCE OF 4 PGA TOUR CLUSTER IN 16 INPUT VARIABLES

Attributes (Average)	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Eagles	0.128	-0.097	-0.185	-0.373
Birdies	0.487	-0.675	-0.468	-1.420
Pars	0.346	-0.221	-0.111	-1.944
Bogeys	-0.545	0.582	0.370	2.229
Doubles	-0.365	0.321	0.225	1.662
Total Holes Over Par	-0.577	0.593	0.386	2.410
Conversion of Greens Hit (Birdie or Better)	0.516	0.205	-0.991	-1.998
Drives	0.160	0.074	-0.319	-0.613
Fairways Hit	0.259	-0.117	-0.408	-0.822
Total GIR	0.517	0.202	-0.991	-1.999
Scrambling	0.451	-1.044	0.052	-1.664
Sand Save	0.208	-0.649	0.108	-0.667
Overall Putting	-0.198	1.399	-0.726	0.660
Putting (GIR Putts)	0.423	0.552	-1.036	-1.805
One Putts	0.169	-1.323	0.711	-0.541
X3 Putt Avoid Total 3 Putts Round	-0.220	0.906	-0.308	0.753

TABLE X. AVERAGE WORLD RANKING OF 4 PGA TOUR CLUSTER

Cluster	Avg of Rank
1	84.27
2	104.15
3	104.16
4	118.37

Consistent with the world-ranking averages, players in Cluster 1 perform best in most categories while players in Cluster 4 perform worst in most categories. Comparing Clusters 2 and 3, players in Cluster 2 perform worse in recovery skills such as putting, short game, or sand play, while players in Cluster 3 excel in these skills but perform worse tee-to-green ball-striking metrics such as Total GIR or Fairways Hit.

TABLE XI. 2ND ORDER PGA TOUR MARKOV MODEL

		Third Cluster			
		1	2	3	4
First State	(1,1)	**0.866**	0.048	0.079	0.006
	(2,1)	**0.655**	0.153	0.153	0.039
	(3,1)	**0.618**	0.128	0.209	0.045
	(4,1)	**0.551**	0.188	0.188	0.072
	(1,2)	**0.512**	0.205	0.236	0.047
	(2,2)	**0.448**	0.256	0.176	0.120
	(3,2)	**0.366**	0.193	**0.345**	0.097
	(4,2)	**0.327**	0.265	0.163	0.245
	(1,3)	**0.497**	0.151	0.293	0.059
	(2,3)	**0.401**	0.232	0.232	0.134
	(3,3)	0.331	0.219	**0.351**	0.100
	(4,3)	**0.343**	0.149	**0.299**	0.209
	(1,4)	0.298	0.255	**0.319**	0.128
	(2,4)	**0.286**	0.222	0.254	0.238
	(3,4)	0.190	0.207	0.224	**0.379**
	(4,4)	0.100	0.183	0.217	**0.500**

Table XI uses the 4 Clusters as states in a 2nd order Markov model – only a third of transitions in a 1st order model met the Markovian property. The 2nd order model uses a player's prior two seasons as a prior state – a player transitioning from row (2,4) to column 1 will have moved, in consecutive years, from Cluster 2 to Cluster 4 and into Cluster 1. Cluster 4 players looking to improve might move directly into Cluster 1 but are more likely to first move into Cluster 2 or 3. This suggests there are two paths for improvement. However, conclusions should be weighed tentatively; only 60% of transitions held the Markovian property. Different algorithms and distance methods could be tried. As data grows, a 3rd order model might also be appropriate.

F. Phase 2: PGA Model Case Study

Markov models could be useful in determining when and how a player's game is changing by analyzing how a player moves through states. Once a specific part of a player's game is highlighted for improvement, the linear delta model approach from phase I could inform the player of drills to perform.

1) PGA TOUR player Jordan Spieth has been a cluster 1 player since 2013, despite slumping over the past few years. His clusters and probability of being in each cluster are in Figure 2.

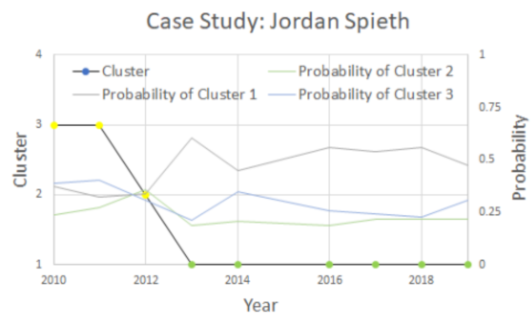

Fig. 2. Jordan Spieth's cluster classifications and probabilities (2010-2019)

Probabilities in Figure 2 show the most likely cluster is 1 for years 2013-2019; however, there is a drift towards cluster 3 in 2014 and 2019. To explore this question in depth, we graphed the standardized scores of Spieth's 2016-2019 performance metrics beside the centroids of Cluster 1 (shown in Figure 3).

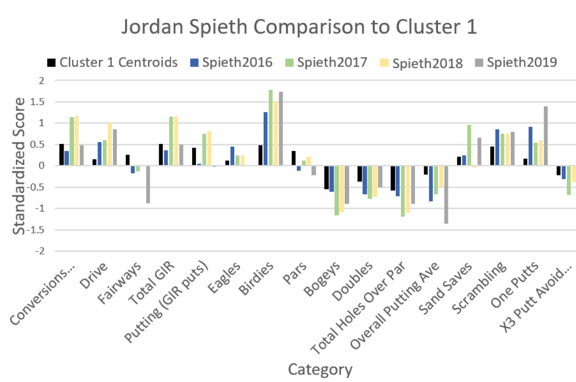

Fig. 3. Spieth's standardized performance metrics vs. Cluster 1 centroids

Based on Figure 3, Spieth's game still generally resembles the centroids of a cluster 1 player represented by the black bars. Two exceptions are Spieth's Fairways and Pars, as he performed worse in these categories than expected, particularly in 2019.

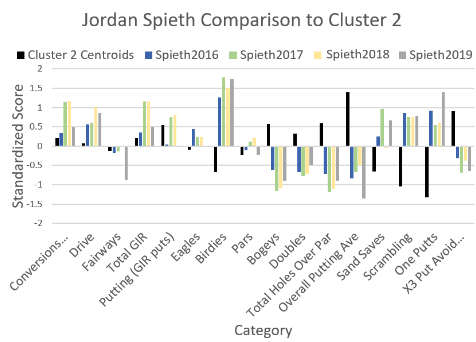

Fig. 4. Spieth's standardized performance metrics vs. Cluster 2 centroids

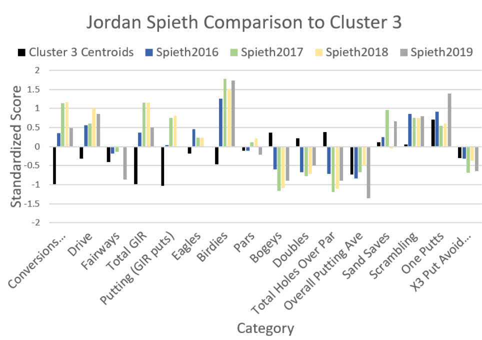

Fig. 5. Spieth's standardized performance metrics vs. Cluster 4 centroids

Figures 4 and 5 compare Spieth's 2016-2019 performance to Clusters 2 and 3. His statistics do not fit either cluster well, but seem to align closer with cluster 3 in terms of Overall Putting Average, One-putts, Fairways and Pars. He is also playing closer to a Cluster 2 player than a Cluster 1 player in Pars.

This information could be helpful in determining what types of drills might improve Spieth's fairways and pars. Spieth might consider performing drills aimed at improving these metrics specifically for a Cluster 3 player. This could also be used in conjunction with the delta model to determine which drills might be helpful in improving this part of a player's game.

V. Recommendations and Future Work

This research provides GameForge with framework for implementing an analytics-based training recommendation system for its clients. As the amount of data in the GameForge databases increases, insights from analysis will increase as well. Once data reaches an appropriate level, a system might be created that takes into account the type of player a golfer is and recommend drills tailored to specific clusters of players.

Future efforts might particularly concentrate on the possibility that 1st order, or even 2nd order Markov models, do not fit the data as well as higher-order models. Preliminary chi-squared tests suggested higher-order models may fit better, but current data is too sparse to test such models. As more rounds are input to GameForge, such models will become feasible. Another potential next step is an analysis of the impetus behind the observed state changes on both the PGA and LPGA TOURs. The performance-based clustering done on PGA TOUR players could be performed on LPGA players. Analyses can be performed to determine exactly which variables – either GameForge metrics or round statistics – changed to cause improvement. This will benefit coaches designing specific training regimens to help pupil's scoring. Specific drills could match to the metrics identified on the PGA and LPGA TOURs.

Acknowledgments

We would like to thank Matt Burkett of Ankura Consulting Group, LLC for sponsoring this research. We would also like to thank GameForge Golf for providing support and guidance throughout, specifically Mark Sweeney and Brian Bailie. Additionally, we would like to thank the Engineering Systems and Environment Department at the University of Virginia for support on this research.

References

[1] K. Peng *et al.*, "Predictive analytics for University of Virginia football recruiting," *2018 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, 2018, pp. 243-248.

[2] G. Beckwith *et al.*, "Systems Analysis for University of Virginia Football Recruiting and Performance," *2019 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA, 2019, pp. 1-6

[3] J. Corscadden et al., "Developing analytical tools to impact U.Va. football performance," 2018 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, 2018, pp. 249-254.

[4] Golf Industry Facts. (2018). Retrieved April 2, 2020, from https://www.ngf.org/golf-industry-research/

[5] J. E. Blanchfield *et al.*, "Developing Predictive Athletic Performance Models for Informative Training Regimens," *2019 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA, 2019, pp. 1-6.

[6] https://mygameforge.com/landing

[7] Seber, G. A., & Lee, A. J. (2012). *Linear regression analysis* (Vol. 329). John Wiley & Sons.

[8] Scherer, W. T., & Glagola, D. M. (1994). Markovian models for bridge maintenance management. *Journal of transportation engineering*, *120*(1), 37-51

[9] Fried, H. O., Lambrinos, J., & Tyner, J. (2004). Evaluating the performance of professional golfers on the PGA, LPGA and SPGA tours. *European Journal of Operational Research*, *154*(2), 548-561.

[10] Bryant, R. (2020). Retrieved 2019, from http://www.shotlink.com/about

[11] Adams, S., Scherer, W. T., & Beling, P. A. (2017). Data, insights, Models, and Decisions: Machine Learning in Context. *Intuition, Trust, and Analytics*, 213.

[12] James, N. (2007). The statistical analysis of golf performance. *International Journal of Sports Science & Coaching*, *2*(1_suppl), 231-249.

[13] Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). Application of neural and regression models in sports results prediction. *Procedia-Soci Behavio Sci*, *117*, 482-487.

[14] Zheng, A., & Casari, A. (2018). Feature engineering for machine learning: principles and techniques for data scientists. " O'Reilly Media, Inc.".

[15] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, *63*(1), 3-42.

[16] Kassambara, A. (2017). Practical guide to cluster analysis in R: Unsupervised machine learning (Vol. 1). STHDA.

[17] Carballido, J. A., Latini, M. A., Ponzoni, I., & Cecchini, R. L. (2018). An Evolutionary Algorithm for Automatic Recommendation of Clustering Methods and its Parameters. *Electronic Notes in Discrete Mathematics*, *69*, 229-236.

[18] Filar, J. A. (1981). Ordered field property for stochastic games when the player who controls transitions changes from state to state. *Journal of Optimization Theory and Applications*, *34*(4), 503-5