Olympic Climbing and Ranking

The 2020 Olympics was set to be the first Olympics featuring rock climbing, a sport that has been growing in popularity in recent years. Unfortunately, due to COVID, climbing will have to wait till Summer 2021 to make its debut in the Olympics. This give us ample time to evaluate the competition and create a model to predict which climbers will win the gold. The crux of the problem is to create a ranking of the competitors, for both men and women, in order to predict who will win the Gold, Silver and Bronze medals. We will take this opportunity to explore ranking metrics to help us build a model that predicts the climbing Olympic winners.

Climbing Context

This project focuses on building a model for the first Olympic qualifying competition the 2019 IFSC Climbing World Championships Combined in Hachioji Japan, which is the most recent event with the same format as the Olympics where all competitors were present. At this competition, and similarly in the Olympics, each climber participates in three event: lead, bouldering and speed climbing. Each discipline has a different set up and scoring methodology, lead climbing is on a tall wall with the climber using a rope to protect themselves from falling, bouldering is on shorter wall without ropes, and speed climbing is on a tall wall that is easier and that the climbers have been able to previously practice. Lead and bouldering are scored on how high a climber reaches on the wall while speed is scored by how fast the climbers can get up the wall.

It was decided that there would be one medal for climbing, so despite the disciplines being vastly different a combined score was devised to combine the events. Each climber is ranked in each event then the ranking is multiplied to get their combined score which is then ranked to find the winner. An example computation can be seen below. This means we want a model which will predict the ranking of our climbers and a metric which will evaluate this ranking.

Name	Lead Rank	Bouldering Rank	Speed Rank	Combined Score	Final Ranking
Janja Garnbret	2	1	2	4	1
Akiyo Noguchi	1	2	3	6	2
Shauna Coxey	3	3	1	9	3

Each discipline has been around for a while and there are frequent competitions where climbers compete in each discipline individually. This as well as past combined world championships give us the data needed to predict the winners. I used this [dataset from kaggle]() where a user had scraped the results data from the IFSC (International Federation of Sport Climbing) website.

Ranking Metrics

Many of the most popular metrics for ranking are geared towards binary relevance like search engine ranking. Which means, these metrics try to quantify out of the top n results how many are relevant. This problem requires a metric that quantifies how well the list was sorted, and doesn't have a binary relevance. This leads us to focus on two metrics: Normalized Discounted Cumulative Gain (NDCG) and Kendall's rank correlation coefficient.

Normalized Discounted Cumulative Gain (nDCG)

The nDCG metric is built off two simpler metrics, Cumulative Gain and Discounted Cumulative Gain. Cumulative Gain is the sum of the relevance (rel_i) for a given position p in the results list:

$$CG_p = \sum_{i=1}^{p} rel_i$$

The relevance value should be larger for more relevant results, this required some manipulations of our ranking which we will go into below. The issue with CG is that for the p results the ordering within the set does not matter. That is why a logarithmic reduction factor was add:

$$DCG_p = \sum_{i=1}^{p} \frac{rel_i}{log_2(i+1)},$$

where $log_2(i+1)$ increases as i increases causing later results to count for less in our metric. Now both of these can be unbounded values and will increase with p as long as there are no negative relevance values. In order to put this on a scale between 0 and 1 we can divide by the ideal discount cumulative gain at p ($IDCG_p$), which is the cumulative gain at p if our list was perfectly ordered:

$$nDCG_p = \frac{DCG_p}{IDCG_p}$$

This metric helps us quantify how close to the actual rank each result was.

Kendall Rank Correlation Coefficient

The other metrics we will consider is the Kendall rank correlation coefficient (τ), which looks to measure how well ordered a list is. For each pair in the results list, $(actual\ rank,\ predicted\ rank)$, we compare it to every other pair and determine if they concordant or discordant or if ranked relatively well.

For example if Janja Garnbret's actual rank was 1 and the predicted rank was 2 and Akiyo Noguchi actual rank was 2 and predicted rank was 3 then this pair is concordant because they respect the actual ordering. In this example if Janja's predicted ranking was 2 and Akiyo's was 1 that would be a discordant pair because they did not respect the actual ranking.

After comparing over all values in the list and counting the number of concordant and discordant pairs you can then calculate the coefficient:

$$\tau = 2\ \frac{(number\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{n(n-1)}$$

This value will be between -1 and 1, if the list is reverse ordered or correctly ordered respectively. This gives us a metric that helps us quantify how well the list was ordered not just how close each rank was predicted.

Scoring Methods

Base Scoring: Multiplied Average of Individual Discipline Events

For the first scoring attempt I didn't use a model at all but just an average of past discipline performance. What I mean by this is for all the past bouldering, lead and speed competitions I took the competitors average ranking, then, in the same way the combine event is scored, I multiplied these averages together. This will be the score to beat with our more complex models.

Below we can see the top 5 competitors per gender based on this score, called avg_rank_multi in the dataframe. For the women this small gut check looks pretty good, we correctly placed the first and second place climbers. However, for the men our top pick ended up placing 18th in the competition. How could this have happened?

Here it is important to look back at the competition, Adam Ondra is favorite to win the Olympics and was expected to do well in this competition. He was disqualified in the lead event for standing on the metal bolt on the wall, that is only supposed to be used to secure the climber to the wall with a rope. This was an anomalous event that we might expect our model to do a poor job predicting.

	first	last	rank	gender	lead_avg_rank	boulder_avg_rank	speed_avg_rank	avg_rank_multi	rank_score	avg_rank_multi_score
37	Adam	ONDRA	18	M	2.333333	6.833333	72.000000	1148.000000	0.10	0.991463
20	Tomoa	NARASAKI	1	M	12.166667	4.166667	37.500000	1901.041667	0.95	0.985863
21	Jakob	SCHUBERT	2	M	4.600000	14.000000	48.727273	3138.036364	0.90	0.976664
25	Kokoro	FUJII	6	M	12.142857	8.428571	38.083333	3897.712585	0.70	0.971015
39	Jongwon	CHON	20	M	27.000000	6.769231	53.714286	9817.318681	0.00	0.926994

	first	last	rank	gender	lead_avg_rank	boulder_avg_rank	speed_avg_rank	avg_rank_multi	rank_score	avg_rank_multi_score
0	Janja	GARNBRET	1	F	2.181818	1.100000	34.888889	83.733333	0.95	0.999377
1	Akiyo	NOGUCHI	2	F	5.375000	2.307692	35.461538	439.859467	0.90	0.996729
4	Miho	NONAKA	5	F	14.142857	2.727273	24.111111	930.000000	0.75	0.993084
9	Jessica	PILZ	10	F	3.000000	10.000000	41.777778	1253.333333	0.50	0.990680
5	Ai	MORI	6	F	3.500000	17.000000	63.666667	3788.166667	0.70	0.971829

In order for us to use the nDCG metric we need to create rank_score and avg_rank_multi_score which are high for the top ranking climbers. Here I took the max value of the column then subtracted the value and divided by the max. This value is what we will use to train our models.

$$\frac{max - rank}{max}$$

How did this average ranking do with respect to our metrics? As anticipated we did significantly better predicting the women's results given that we didn't have an outlier event.

```
Men's NDCG All:  0.8077067072045893      Womens's NDCG All:  0.9770159311401715
Men's NDCG Top 3:  0.59160052907239      Womens's NDCG Top 3:  0.974264436127282
Men's Kendall:  0.052631578947368425     Womens's Kendall:  0.6105263157894737
```

I have included the nDCG score for all competitors and for the top 3 as well as the Kendall coefficient. Since we are trying to predict the medal winners it is important that we get those top three predictions correct. We can see that this especially penalizes us in the men's competition where we can recall Adam Ondra was disqualified despite being predicted to win.
If we remove Adam Ondra from the men's results we do a bit better but still not better than the women's.

```
mens_results = np.asarray([list(pred_aggs[(pred_aggs.gender == 'M') &
                                          (pred_aggs['last'] != 'ONDRA')]['rank_score'].values)])
mens_avg_pred = np.asarray([list(pred_aggs[(pred_aggs.gender == 'M') &
                                           (pred_aggs['last'] != 'ONDRA')]['avg_rank_multi_score'].values)])

print("Men's NDCG All: ", ndcg_score(mens_results, mens_avg_pred))
print("Men's NDCG Top 3: ", ndcg_score(mens_results, mens_avg_pred, 3))
print("Men's Kendall: ", kendalltau(mens_results, mens_avg_pred)[0])
```

```
Men's NDCG All:  0.9329325814578928
Men's NDCG Top 3:  0.9613966541909231
Men's Kendall:  0.14619883040935672
```

We can also see for the women the Kendall rank correlation is significantly better, which means not only is each competitor close to their actual ranking but they are also well ordered with respect to the other competitors.

Basic Linear Regression Train/Test Split

The next question is can we create a machine learning model that predicts the results better than the average ranks multiplied together. My first attempt at a model was to do the typical train/test split on all the data, predict using linear regression with the individual discipline averages and calculate the metrics. However this gave us wildly fluctuating results and didn't take into account the genders. We can do better if we trained and tested with specific competitions.

```
feature_columns = ['lead_avg_rank', 'boulder_avg_rank', 'speed_avg_rank']
X = pred_aggs[feature_columns]
y = pred_aggs['rank_score']
X_train, X_test, y_train, y_test  = train_test_split(X, y, random_state = 1)

reg = LinearRegression().fit(X_train, y_train)

pred_LR = np.asarray([list(reg.predict(X_test))])

print("NDCG: ", ndcg_score(pred_LR , np.asarray([list(y_test.values)])))
print("Kendall: ", kendalltau(pred_LR, np.asarray([list(y_test.values)]))[0])
print("Number Test Examples: ", y_test.shape[0])
```

```
NDCG:  0.9789542166568036
Kendall:  0.4319297483312999
Number Test Examples:  10
```

```
X_train, X_test, y_train, y_test  = train_test_split(X, y, random_state = 10)

reg = LinearRegression().fit(X_train, y_train)

pred_LR = np.asarray([list(reg.predict(X_test))])

print("NDCG: ", ndcg_score(pred_LR , np.asarray([list(y_test.values)])))
print("Kendall: ", kendalltau(pred_LR, np.asarray([list(y_test.values)]))[0])
print("Number Test Examples: ", y_test.shape[0])
```

```
NDCG:  0.8378307274224305
Kendall:  0.089893314995098956
Number Test Examples:  10
```

Training and Testing on Specific Competitions

For the next set of models I re-processed the data so that I had a training set which aimed to predict the 2018 Combined World Championship then I tested those models on the 2019 Combined World Championship per gender. This yielded more stable promising results that were more reasonable to compare to my base scoring.

Results and Conclusion

The scoring that performed best for the women's competition was the simple multiplied average ranking of previous individual events. For the men's we were able to improve both the nDCG and Kendall's coefficient with a linear regression model trained on the average ranking of previous individual events.

```
Men's NDCG:  0.9422327714247841
Men's NDCG Top 3:  0.8422170657278147
Men's Kendall:  0.4631578947368421
```

There are many possible avenues to go that might better predict the Olympic climbing results, including predicting the actual scores of each event then accumulating the results for the final ranking predictions, or adding in more data beyond competition data that might better predict the climbers readiness for the competition including training schedule, outdoor climbing performance and previous competition experience. Any further extension would still require metrics to help quantify how close the model predicted each climbers rank and how well the model ordered the climbers. We have about a half year left to improve our models and test them on the actual climbing Olympic results.