Introduction

In tennis, there have been generally few studies conducted on ranking players and historical match data. One study that did this was published by ESPN Senior Writer Peter Keating in “Determining Tennis’ Greatest Player” from August 14th, 2011. It was determined from his research that Jimmy Connors was the “...greatest tennis player ever.” More than a decade later, we want to see how Jimmy’s position still holds up. Furthermore, we want to see how alternative ranking methods as used in this research compare to the official rankings provided by the Association of Tennis Players.

Dataset

The data used in the ranking methods were collected from Tennis Abstract. Tennis Abstract provides historical singles match data from 1968 to 2023 from the Association of Tennis Professionals (ATP). The dataset included details such as tournament ID, name, level, and date; surface of the court; draw size; match number; winners’ and losers’ ID, seed, entry, name, hand, height, country, and age; match score and “best of”; winners’ and losers’ ace, double faults, total serve points, points won on 1st and 2nd serves, serve games, break point saved and face counts; and winners’ and losers’ rank and rank points. Overall, the dataset includes comprehensive data about all the matches that occurred historically. The data was separated into CSVs by year on Tennis Abstract’s GitHub repository. For the methods to work, the data was concatenated together using a script written in Python. On top of merging the data, a step was added to sanitize the data by removing matches with no score value or invalid or unwanted values, such as “W/O” or “nan.” This action cleared 1,173 “invalid” matches out of 190,276 in total. Once the data was sanitized, all the players were reassigned new IDs starting from zero to one minus the total number of players ever in the Tennis Abstract ATP singles matches dataset. This assignment made it easier for the code to compute the matrices for each of the methods. The matches were then saved in one master spreadsheet to be used for analysis by the ranking methods.

Methods

Elo Method

The Elo ranking used in this project is based on the standard Elo method. In the Elo method, ratings for each player is calculated based on the player’s rating, K constant, actual and expected score. On the other hand, the Elo method also predicts the score using two player ratings.

$$E_{\mathsf{A}} = \frac{1}{1 + 10^{(R_{\mathsf{B}} - R_{\mathsf{A}})/400}}$$

The formula above calculates the expected score for player A with its rating (R_A) and player B’s rating (R_B). The order of R_A and R_B can be flipped to calculate the expected score for player B.

$$R'_{\mathsf{A}} = R_{\mathsf{A}} + K \cdot (S_{\mathsf{A}} - E_{\mathsf{A}})$$

The formula above calculates a new score for player A with its existing rating, K constant, actual and expected score. The formula adjusts the player rating based on their performance relative to what was expected. For example, if a player's actual score was higher than expected, his or her rating is increased to reflect the truth. The same is done with actual scores lower than expected except by decreasing the rating ("Elo Rating System").

PageRank Algorithm

The PageRank algorithm, which was used, was adapted from the Markov method. The algorithm is a model of a random surfer moving through the network of webpages. Since PageRank was based off of the Markov process, the formula for the Markov method is shown below.

$$A_{ij}(n) = A(X_n = j | X_{n-1} = i)$$

The **A** matrix is a transition matrix where all the rows add up to 1 because all the values are probabilities. In the matrix, an **(i, j)** element of the **A** matrix is the probability of a random surfer moving from webpage **i** to webpage **j**. The **A** matrix is created from knowing the probability of moving from page **i** to page **j** under the Google PageRank model.

The model states that there is an 85% chance that a random surfer will follow one of the hyperlinks–if there are any–on a webpage. The other 15% chance is the probability that a surfer will teleport–or jump–to any page in the network of pages. Cases where a web page has no links–called a dangling node–a surfer will teleport to any webpage in the network with equal probability. This sets the probabilities of going between any two web pages and creates the transition matrix **G**–or Google matrix.

To find the PageRank vector, which gives the ratings, we use this formula

$$Gv_{n+1} = v_n$$

until v_{n+1} is close to v_n. The underlying theory guarantees convergence so any initial vector that sums up to 1 will work (Langville and Meyer).

Colley Method

The standard Colley method was also utilized for ranking the players. The formula for the colley method uses a matrix.

$$C\mathbf{r} = \mathbf{b}$$

The *C* in the formula contains the Colley matrix. This Colley matrix is first created by setting an order of the teams, players, or entities for the columns and rows. The order can be anything as long as it is consistently used for both axes. Then, the numbers are found by the number of games each team played against each other. An example Colley matrix is shown below.

$$\begin{bmatrix} 5 & -1 & -1 & -1 \\ -1 & 4 & -1 & 0 \\ -1 & -1 & 5 & -1 \\ -1 & 0 & -1 & 4 \end{bmatrix}$$

The numbers diagonal along the diagonal from the top left to the bottom right of the matrix are calculated by

$$2 + \text{total games the team played}$$

On the other hand, for the numbers outside the diagonal, they are calculated with

$$-\textbf{total games the teams played each other}$$

and these two equations result in the matrix shown below.

$$\begin{bmatrix} 5 & -1 & -1 & -1 \\ -1 & 4 & -1 & 0 \\ -1 & -1 & 5 & -1 \\ -1 & 0 & -1 & 4 \end{bmatrix}$$

Next, in order to find the matrix for **b**, the formula below is used

$$b = 1 + \frac{1}{2}\left(wins_{team} - losses_{team}\right)$$

and the resulting matrix for *b* looks like

$$\begin{bmatrix} \frac{3}{2} \\ 0 \\ \frac{3}{2} \\ 1 \end{bmatrix}.$$

Finally, the **r** matrix, which contains the actual ratings of each entity, is found through solving systems of linear equations. Usually, this operation is automated by a machine, like through Python's Numpy's **linalg** function (Chartier et al.). The resulting **r** matrix looks like

$$\begin{bmatrix} 0.583 \\ 0.292 \\ 0.583 \\ 0.542 \end{bmatrix}.$$

Analyzing Effectiveness

In order to gauge the effectiveness of the implemented methods, the computed matrices of each method were then applied over each of the matches in the dataset. A correct prediction would mean that the player with the better ranking was the actual winner. Then, these correct predictions would be counted and divided by the total number of games to yield the predictability percentage.

Elo Method

The Elo method scored a predictability of 65.66% with a k-value of 32.

Elo Method Top 10 Player Rankings		
Rank	Rating	Player
1	745.15423	Roger Federer
2	666.36422	Rod Laver
3	626.29927	Ken Rosewall
4	610.14331	Daniil Medvedev
5	604.42844	Boris Becker
6	599.53688	Rafael Nadal
7	598.36419	Tony Roche
8	574.32691	John Newcombe
9	551.53588	Cliff Richey

10	548.53557	Arthur Ashe

Table 1

PageRank

PageRank scored a predictability of 67.53% with an alpha value of 0.85.

PageRank Top 10 Player Rankings		
Rank	Rating	Player
1	0.00015917	Jimmy Connors
2	0.00015880	Roger Federer
3	0.00015582	Rafael Nadal
4	0.00015581	Ivan Lendl
5	0.00015580	Novak Djokovic
6	0.00015368	Guillermo Vilas
7	0.00015305	Ilie Nastase
8	0.00015272	John McEnroe
9	0.00015248	Andre Agassi
10	0.00015142	Stefan Edberg

Table 2

Colley Method

The Colley method scored a predictability of 69.54%.

Colley Method Top 10 Player Rankings		
Rank	Rating	Player

1	1.31145	Novak Djokovic
2	1.29864	Rafael Nadal
3	1.29024	Roger Federer
4	1.22395	Carlos Alcaraz
5	1.22377	Ivan Lendl
6	1.21881	John McEnroe
7	1.21261	Pete Sampras
8	1.20671	Andy Murray
9	1.20523	Bjorn Borg
10	1.19669	Jimmy Connors

Table 3

Notes

Some reasons for why the results turned out the way they were may be from the data sanitization step; the equal amount of importance given to each match–for example, a match from decades ago being regarded equally as a recent one; or the fact that some players may have played longer and accumulated more wins than others. During the data sanitization step, it was noted that 1173 matches were dropped because of them consisting of data in an unwanted or invalid format. These missing matches may contribute to the current predictability percentage in each of the methods. Secondly, these methods are not weighed: they do not factor in the age of the match and may worsen a player's ranking as players improve overtime, but the methods still regard the first match as equally important as the last. Lastly, the implementation of the methods assumed that each player consistently played since the ATP began. Of course, this assumption is not at all possible, and some players will accumulate more wins from competing in more matches than other players. On top of these lurking variables, there are more factors that may influence the effectiveness of the rankings, such as the lack of account for injuries, court surfaces, etc.

Comparing Results

It is good to note that all the methods treat a player as equally talented over time. For example, the methods used treats Jimmy Connor from when he first began competing as equally talented as when he retired.

Elo Method

In table 1, Roger Federer ranks first among all tennis players. This ranking comes higher than in PageRank, where Federer was second, and in the Colley method, where he was third. in the official ranking from tournament 2021-540, Roger Federer was ranked eighth.

Jimmy Connors, who was ranked number one in the ESPN article "Determining Tennis' Greatest Player," completely dropped from the top 10 rankings in 1994-409. This ranking was consistent with his last tournament from 1996-409, where he was ranked 413th.

PageRank

In PageRank, Jimmy Connors was ranked first, which was consistent with the outcome in the ESPN article, but still significantly higher than his official ranking from 1996 of 413th. For Roger Federer, he was ranked second: one spot higher than his official ranking of third.

Colley Method

In the Colley method, Jimmy Connors dropped to 10th. On the other hand, Roger Federer maintained the same rank as his official rank of 3rd place.

Novak Djokovic held the title of first place in this method, which was higher than his official ranking of 5th from his last tournament in 2023-580.

Aggregated Results

All the method's rankings were aggregated using this formula

$$\text{Player Rating} = \sum_{i=\text{number of ratings}}^{n} \text{Rating}_n \cdot \frac{\text{Method Predictability}_n}{\sum_{i=\text{number of methods}}^{m} \text{Predictability}_m}.$$

Essentially, the aggregated player rankings were the sum of the method's rankings. However, a weight is applied where, for example, the rating that comes from a method with a higher prediction rate has a higher weight. The result of the aggregation is below.

Aggregated Top 11 Player Rankings		
Rank	Rating	Player
1	2.01914	Roger Federer
2	3.62861	Rafael Nadal
3	7.58173	Ivan Lendl
4	11.39849	Boris Becker
5	13.28688	Andre Agassi
6	13.67742	Stefan Edberg
7	14.44923	Arthur Ashe
8	14.63960	Novak Djokovic
9	15.74688	Rod Laver

10	17.01656	Andy Roddick
11	17.04196	Jimmy Connors

Table 4

Conclusion

After processing over five decades of data and aggregating the results, Roger Federer came out number one. Amazingly, Jimmy Connors still holds his place near the top at rank 11 despite retiring decades ago in 1996. When these results were compared with official rankings, Federer ranking was roughly the same as rank 8th in the 2021-540 tournament. On the other hand, Jimmy Connors aggregated rank was significantly higher compared to his official ranking of 413th in tournament 1996-409.

Future Work

"Weighted Elo Rating For Tennis Match Predictions" compared the effectiveness of weighted Elo ratings over using standard Elo. In the beginning, the document described the potential of model-based outcome predictions in sports to gain an advantage over the less informed in gambling. On top of this potential, the writers mentioned that there are few academic studies on tennis. In the studies that do exist, they went over "...regression-based, point-based and paired comparison approaches…" using probit and logit estimators, Bradley-Terry models, and the Elo method. Next, the paper went over the Elo method and its origin in chess. The Elo method takes into account a player's whole history and recent match outcomes to calculate a rating, but this study aimed to be more holistic by incorporating the scores of the matches to provide a more accurate prediction. In the end, it was determined that the weighted Elo method was more effective than the standard Elo (Angelini et al.). This write up can be relevant for future work on tennis rankings as the methods used in this study treats each player as equally gifted over time, which is not realistic. A future study could weigh the relevance or age of a match on the rankings.

In “A Bradley-Terry Type Model for Forecasting Tennis Match Results,” this report documented the usage of the Bradley-Terry model for predicting tennis match outcomes. In the beginning, the paper acknowledged the lack of published models that enable positive returns in sports betting, most likely to keep others from replicating the successes. Then, to develop a successful model, the researchers utilized the Bradley-Terry model with historical game data, the surface of the current match and past ones, and the time since the past matches. It was found that the new model was more successful than other models that used official ranking data. The researchers attributed the official ranking data’s deficiency to a lack of factors considered and a different purpose for the rankings. For example, the official rankings did not evaluate the difficulty of the opponents faced and preferred players that played consistently well throughout the season: a decision that decreased the accuracy of measuring player quality (McHale and Morton). A future study could expand the findings of the results found in this study by incorporating more data attributes to see its influence on the rankings and the accuracy.

Bibliography

Angelini, Giovanni, et al. “Weighted Elo Rating for Tennis Match Predictions.” *European Journal of Operational Research*, vol. 297, no. 1, Feb. 2022, pp. 120–32, https://doi.org/10.1016/j.ejor.2021.04.011. Accessed 13 Aug. 2023.

Chartier, Tim, et al. “Mathematics and Sports: Bracketology: How Can Math Help?” *Www.semanticscholar.org*, 2010, www.semanticscholar.org/paper/Mathematics-and-Sports%3A-Bracketology%3A-How-can-math-Chartier-Kreutzer/889825aa9b0de3f68182dd674cce5682ecbf045f. Accessed 20 Aug. 2023.

“Elo Rating System.” *Wikipedia*, 16 Aug. 2023, en.wikipedia.org/wiki/Elo_rating_system#Mathematical_details. Accessed 27 Aug. 2023.

Keating, Peter. “Keating: Determining Tennis’ Greatest Player.” *ESPN*, 14 Aug. 2011, www.espn.com/sports/tennis/insider/story/_/id/6861344/how-network-analysis-revolutionize-athlete-rankings. Accessed 20 Aug. 2023.

Langville, Amy, and Carl Meyer. *Who’s #1? The Science of Rating and Ranking*. Princeton University Press, 2012, pp. 29–52, 67–78.

McHale, Ian, and Alex Morton. “A Bradley-Terry Type Model for Forecasting Tennis Match Results.” *International Journal of Forecasting*, vol. 27, no. 2, Apr. 2011, pp. 619–30, https://doi.org/10.1016/j.ijforecast.2010.04.004. Accessed 14 Mar. 2019.