Tennis Player Ranking using Quantitative Models

A. Demetris Spanias* and B. William Knottenbelt**

** Department of Computing, Imperial College London, South Kensington Campus, London, SW7 2AZ d.spanias10@imperial.ac.uk*

*** Department of Computing, Imperial College London, South Kensington Campus, London, SW7 2AZ wjk@doc.ic.ac.uk*

Abstract. The Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA) generate weekly rankings for professional tennis players by awarding points to each player depending on how far the player has advanced in a countable tournament. Since tournaments are designed such that top players face the lower-ranked players in the earlier rounds, a bias is introduced which favours the top players. In this paper we demonstrate two new algorithms, SortRank and LadderRank, which rank professional tennis players. Both ideas make use of a quantitative tennis model to assess the performance of individual players and then compare them with each other. SortRank uses traditional sorting algorithms to rank the players using the result of a simulated match between the two players as the comparison criterion. LadderRank ranks players using a "sports-ladder" style iterative algorithm, which also compares players based on the result of a simulated match between them. Both algorithms are flexible as they can be implemented using any underlying quantitative model. The ranking systems are demonstrated and assessed based on their ability to predict the outcome of matches played within the period used to rank the players.

1. Introduction

Professional tennis rankings are at the centre of attention of the tennis world. Both the Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA) rank professional tennis players and use their rankings to decide both the participation of players in tournaments, as well as the ultimate champion of the year. Being a top ranked player generates a great deal of prestige and popularity. In fact, most professional tennis players have, in one way or the other, mentioned their passion to reach the top of the rankings.

One may argue that any absolute ranking system is by definition flawed when applied to such a complex sport in which there is an unknown degree of transitivity and a multitude of parameters to take into account. Nonetheless, an overall ranking is a simplistic method of determining who is performing better at the sport and captivates both the public and media. Simplistic as it may be, there is a general desire for the overall ranking system to be "fair". Unfortunately rankings, as they are currently calculated, provide the top players with an unfair advantage as seeded tournaments make it increasingly difficult for lower ranked players to climb the rankings.

This bias which favours the top players of the rankings, does not only affect the lower ranked players but also researchers who have used these rankings as a tool for prediction of match outcomes. Clarke & Dyte (2000) propose an approach based on regression which uses ATP ranking points to simulate professional tennis matches. Addtionally, del Corral & Prieto-Rodriguez (2010) attempt to assess the degree to which the difference in ranking points are good indicators of the outcome of Grand Slam matches.

Some research has also been directed towards the invention of different ranking systems. Clarke (1994) proposed a ranking system which uses a player rating which is adjusted after each match played by the player. The adjustment is calculated using exponential smoothing on the difference between an expected result suggested from the previous ranking difference, and the actual match result. A more recent method was proposed by Radicchi (2011) which makes use of an algorithm similar to Google PageRank by Brin & Page (1998). Radicchi uses PageRank by assigning prestige values for all professional players and adjusts them relative to the number of victories they achieved against other players. This ranking system can be used to rank all players regardless of the time period they were active and thus contributes to an investigation on the best player of all time. A few years later, Dingle, et al. (2013) presented further evidence of PageRank's usefullness as a ranking tool for both female and male professional tennis players and also showed that a ranking generated using PageRank is a better predictor of match results than the official ATP rankings.

In this paper, we provide evidence towards the forementioned bias inherent in the current ATP ranking system and we attempt to introduce a new, flexible concept of ranking systems. We compare our ranking systems with PageRank for tennis and the official ATP Rankings by quantifying the extend to which they reflect the outcome of the set of matches used to generate them and show up-to-date (March 2013) results.

2. ATP Ranking System

The Emirates ATP Rankings is the official ranking system ATP used for 2013. It is "a historical objective merit-based method used for determining entry and seeding in all tournaments" as the official ATP World Tour website states. The ranking is generated using a summation of points players acquire while proceeding within seeded tournaments. Tournaments themselves are split into categories with some tournaments awarding more points than others (see Table 1).

Table 1 – ATP Ranking points awarded for main ATP tournament categories. (Additional points are awarded from the Barclays ATP World Tour Finals, Olympic Games, Challenger and Futures tournaments that are not included in this table. Numbers in brackets are dependent on the tournament draw size. Points for qualification are also dependent on draw size.)

	W	F	SF	QF	R16	R32	R64	R128	Qual.
Grand Slams	2000	1200	720	360	180	90	45	10	25
ATP World Tour Masters 1000	1000	600	360	180	90	45	10(25)	(10)	25
ATP 500	500	300	180	90	45	(20)	-	-	20
ATP 250	250	150	90	45	20	(5)	-	-	12

The summation of ranking points is over a maximum of 18 tournaments played within the previous 52 weeks, out of which four are the Grand Slam tournaments, eight are the compulsory ATP World Tour Masters 1000, and the rest are the best six results from the ATP 500, 250 and other tournaments (given a minimum of 4 ATP 500 tournament participations). Additionally, players who have finished within the top eight rankings at the end of the ATP tennis season, qualify to play at the Barclays World Tour Finals to earn points that count towards crowning the final champion of the year. In those years where the Olympics occur, the players also win extra points for the position they get in the Olympics.

While the Emirates ATP Rankings provide accurate rankings for the top 32 players, by design players ranked lower than the top 32 are at a disadvantage. The seeded tournament system makes it increasingly difficult for lower ranked players to proceed into the latter rounds of tournament and thus earn the necessary points to climb the rankings.

Almost all countable tournaments have seeded players, i.e. the top 16 or 32 players who are participating in the tournament have a seeded position. The tournament draw is set up in a way such that the seeded players do not face any other seeded players in the first round. The reasoning behind this is to avoid situations where top ranked players face off in the earlier rounds and get knocked out earning fewer points. This non-random selection of draws creates a bias towards the top 32 players as any players ranked lower than that have a much higher chance to face the top players in the early rounds of the tournament and therefore have much higher chance of being knocked out without earning the points they deserve. This can make it difficult to rank the true performance of these players especially when compared to one another. Evidence of this will be presented in the results section later on.

On the other hand, seeded tournaments together with the Emirates ATP Ranking system, create a much more accurate ranking of the top 32 players when compared to each other. The reasoning behind this is that these players get to face each other in higher frequency as they are more likely to proceed in the latter rounds and therefore there is a more data on which the rankings are based upon. Additionally, the difference in points earned for each victory is higher in the latter rounds. This higher difference in points boosts players who achieve victories against other high ranked players and thus enhances the subtle differences in their performance.

3. Background of Rankings and Tennis Models

In this section we briefly introduce some tennis models which have been presented in past literature and are used in combination with our ranking algorithms, SortRank and LadderRank. We first introduce how one can construct a hierarchical Markov model to estimate the probability of a player winning a match against another player using only the probabilities of the two players winning points while serving. We then describe a Markov model that can be used to calculate the probability of a player winning a point against another player while serving. Finally, we briefly present the PageRank ranking system for tennis that is used as a comparison system in the results section.

3.1 Hierarchical Markov Model for Tennis

A study performed by Klaassen & Magnus (2001), shows that even though points in tennis are not independent and identically distributed (i.i.d.), one may assume that they are for the purpose of modelling a tennis match because the deviation from independency is small. This means that one can estimate the probability of a player winning a game while serving or even the probability of a player winning a tiebreaker by constructing a Markov Chain of the game/tiebreaker which uses only two parameters, the probabilities of the two players winning a point while serving. This is turn allows one to calculate the probability of a player winning a set using only the probabilities of each player winning a service game and the probability of a player winning a tiebreaker. Finally, one can hierarchically calculate the probability of a player winning a match using only the probabilities of the two players winning sets in which they served first. Barnett & Clarke (2002) demonstrate this idea using a simple spreadsheet application which recursively calculates the probability of a player winning a match from every score-line.

3.2 Low-Level Point Markov Model

Having discussed how to hierarchically model a tennis match, all that remains is a method to estimate the probability of a player winning a point on serve. Spanias & Knottenbelt (2012) present a Markov chain in an attempt to model a tennis point and show two techniques of parameterising the model using historical player statistics. The first technique, named the "*Uncombined*" model, estimates the probability of a player winning a point while serving against the "average" professional player. The second technique, named the "*Combined*" model, estimates the probability of a player winning a point while serving against a specific player by combining serving statistics of the server with return statistics of the receiver. These two models will be used in conjunction with the hierarchical Markov model introduced earlier to generate rankings using SortRank and LadderRank.

3.3 PageRank Tennis Ranking

The PageRank tennis ranking system was first introduced by Radicchi (2011) and further investigated by Dingle, et al. (2013). It is an effective ranking system for tennis players which we use in this paper as a good comparison for the ranking systems introduced by this paper. The system is based on Google's PageRank algorithm summarized in Brin & Page (1998) which is used for ranking websites.

In order to explain how PageRank can be applied in tennis we need to define a few variables. Let w_{ji} be the amount of tennis matches player j has lost against player i, and s_j^{out} be the total defeats suffered by player j. Also let α be a weight factor between 0 and 1 and N be the total number of players being ranked. The prestige of player i is then described by the following equation.

$$P_i = (1-\alpha)\sum_j P_j \frac{w_{ji}}{s_j^{out}} + \frac{\alpha}{N} + \frac{(1-\alpha)}{N}\sum_j P_j \delta\left(s_j^{out}\right)$$

P_i, the prestige value assigned to player i, is calculated as the summation of three parts. The first part is the amount of prestige that is transferred from player j to player i, the second part is a constant redistribution of prestige and the third part is used as a constant value for players with no outward links (no defeats). An algorithm can be designed which will iteratively calculate the prestige of all players until they converge. Players are then ranked according to the amount of prestige they hold.

4. SortRank and LadderRank

SortRank and LadderRank are two similar approaches to ranking professional tennis players. They both use an underlying tennis model which estimates the outcome of a simulated match between players and rank players based on that outcome. SortRank is a faster algorithm but has the requirement that the underlying model is absolutely transitive. LadderRank is a slower algorithm which expands the idea of SortRank taking into account the non-transitive nature tennis models may have. This is done by sorting the same players over and over even after a regular sorting algorithm would have finalized their position. Also the algorithm has the ability to compare players with other players who are not immediately next them.

4.1 SortRank

The concept behind SortRank is very simple: take any tennis model, convert it into a binary model and then use it as the comparison criterion of a sorting algorithm. For example: let's assume that we have a list of players to rank. A sorting algorithm such as QuickSort, as described by Hoare (1961), can be used to sort this list of players by using a binary model which outputs a comparison criterion between players.

A limitation of any sorting algorithm is that it assumes absolute transitivity. This means that if Player A can beat Player B and Player B can beat Player C then it must hold that Player A can beat Player C. As a concequence, any model that is used as the comparison criterion should also be absolutely transitive.

An example of a fully transitive model is the "*Uncombined*" model mentioned in section 3.2. This model is transitive by definition as the opponent is not taken into consideration when estimating the parameter of a player winning a point while serving. Therefore, the output of any probability from the model is always compared against the constant "average" player. This "*Uncombined*" model can be converted into a binary model by using the resulting probability of Player A winning a match against Player B. If this probability is greater than 0.5 then the binary model returns "true", otherwise it returns "false".

This binary model can be joined with any sorting algorithm to generate a ranking. For this to happen, the sorting algorithm, when comparing two players, A and B, should use the binary model as the comparison criterion. That is, if the binary model returns "true" for Player A winning a match against Player B, the sorting algorithm places Player A above Player B in the rankings. By completing the algorithm for the entire list of players, the end result is a sorted list of players based on their performance, with the best player at the top of the list, thus a ranking.

4.2 LadderRank

To overcome the limitation of absolute transitivity, we constructed a new algorithm that does not assume the comparison criterion is absolutely transitive. This algorithm is inspired by normal "sports-ladders". In a "sports-ladder" there is an initial ranked list of players, and each of those players is allowed to challenge another player that is ranked up to X positions higher. If the challenger is victorious in the challenge, then he/she overtakes the player challenged and pushes everyone in-between one position down. The resulting algorithm is described by the pseudo-code below.

For this algorithm to function correctly it must be provided with these crucial variables: the *number_of_iterations*, the *positions_above_allowed_to_challenge* and the *ranking_list*. To ensure complete ranking of the players the *number_of_iterations* must always be larger than the number of players being ranked. The *positions_above_allowed_to_challenge* defines the number of positions in the ranking list that any player is allowed to jump after any challenge. Finally the *ranking_list* is the list of players ranked in an initial order.

```
for (int i =0; i <  number_of_iterations; i++) {
   foreach (current_player in ranking_list) {
      if (current_player.ranking > 0) {
         x = positions_above_allowed_to_challenge
         if (x > current_player.ranking) { x = current_player.ranking }
         for (int position = x; position > 0; position--) {
            PlayerA = PlayerWithRanking(current_player.ranking – position)
            PlayerB = current_player
            if (Compare(PlayerA, PlayerB) == false) {
               //if player A loses the match-up move player B
               //above A and push all players inbetween 1 spot down
               MovePlayerToRanking(PlayerB, PlayerA.ranking)
               position = 0 //stop challenging
            }
         }
      }
   }
}
```

The function PlayerWithRank(integer) which appears in the algorithm retrieves the player which has the ranking provided as the integer parameter. The function MovePlayerToRanking(player, integer) changes the ranking of the player to the integer value provided and shifts all rankings of players which were between the player and the new ranking by 1 position towards the direction of the player's current ranking. For example in a list of three players, A, B and C ranked as 1, 2 and 3 respectively, the function MovePlayerToRanking(C, 1) will change the rankings of A, B and C to 2, 3, 1 respectively.

5. Evaluation and Results

In this section we will present and discuss the results of our implementation of the LadderRank ranking system when using the "*Combined*" model as the comparison criterion. Figure 1 illustrates the top 100 players in the ATP Official Rankings on the 18th of March 2013 and corresponding LadderRank ranking generated over the same period for x=3. It can be observed that players ranked by the ATP in positions 1-32 are positioned very close to the y=x line. This means that the LadderRank system ranks them in a similar position to the ATP ranking system. The two systems start to deviate in the rankings a lot more for players ranked in positions greater than 32 by the ATP. This appears to support the theory that seeded tournaments deteriorate the accuracy of rankings of players ranked greater than 32 by the current ATP system. In fact similar results have been produced using the PageRank ranking system and are also evident in other periods (see Dingle, et al. (2013)).

In Figure 1, any players that appear above the y=x line are players which according to LadderRank are ranked higher than they should be by the ATP. Similarly players that appear below the line are players that are ranked lower than they should be.

A striking case is Mardy Fish who has dropped to ATP position 33 on the 18th of March 2013, from being number 9 in the world on the 19th of Match 2012. Mardy Fish on the other hand is still ranked in the top 10 players on the LadderRank system. The reason for this is the underlying model, which uses the average statistics of the player over the past year. Therefore the "*Combined*" model itself does not adapt fast to changes in performance of players and as such, the LadderRank ranking did not adapt quickly and is still showing Mardy Fish as one of the top players. This can be fixed by using a heavier weighting to more recent statistics when calculating the probabilities of winning the point on serve in the underlying model. Therefore this is not a problem of the LadderRank algorithm but a problem with the "*Combined*" model.

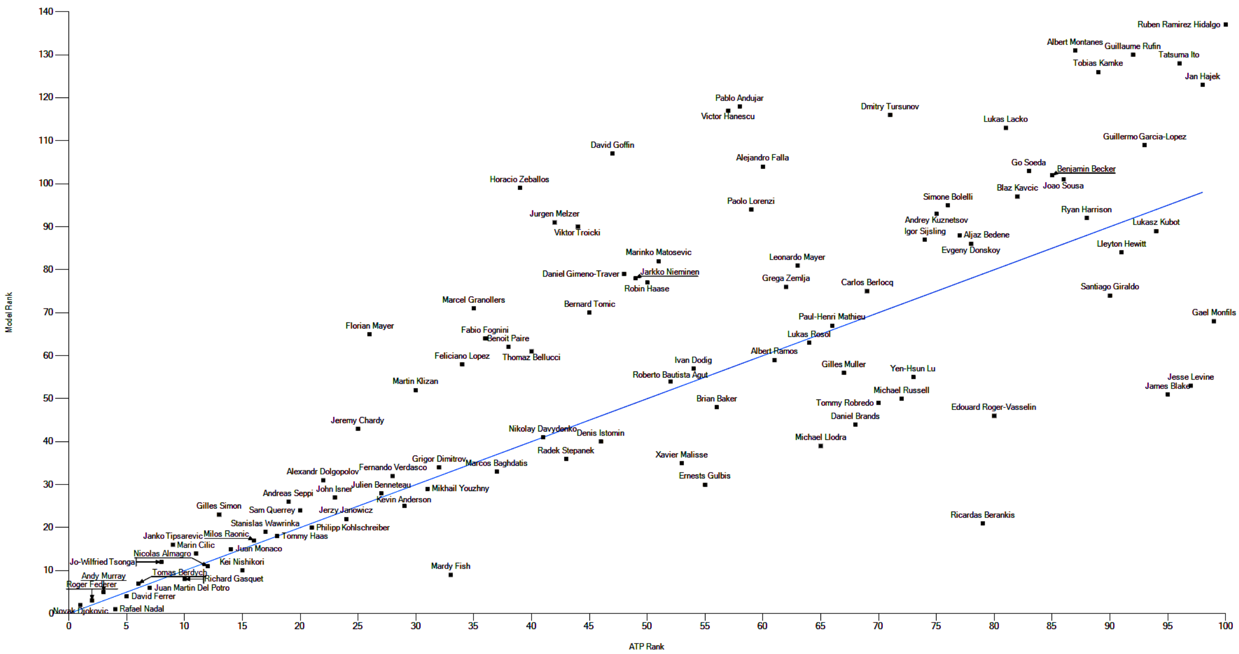

Figure 1 – Comparison between LadderRank-Combined model with x=3 and the ATP Rankings of the Top 100 players over the period 18/03/2012-18/03/2013.

To analyse the performance of our ranking systems further and get a metric to compare them with other systems such as the PageRank and the ATP, we used the rankings generated by each system and tested whether those rankings "predict" the outcomes of the matches within the period that was used to generate them. To put it simply, we calculate the percentage of matches in which the winner of the match is better ranked than the loser.

Using 2552 matches that were played in the period 18th of March 2012 to 18th of March 2013, we generated these percentages for 5 ranking systems – the ATP Official rankings, the PageRank system using Match Victories as weights, the SortRank-Uncombined system and the LadderRank-Combined system with x=1, 3 and 5). Table 2 shows these results in detail.

Table 2 – Comparison of predictive power of ranking models over matches played in the period used to generate them. These rankings were generated for 298 players who competed in 2552 matches over the period 18/03/2012 to 18/03/2013. In SortRank and LadderRank prediction results, 381 matches were not attempted as there were insufficient statistics (less than 10 matches) to model one or both the players who took part in those matches.

ATP	**Match PageRank (α=0.15)**	**SortRank Uncombined (381 Skipped)**	**LadderRank Combined (381 Skipped) (x=1)**	**LadderRank Combined (381 Skipped) (x=3)**	**LadderRank Combined (381 Skipped) (x=5)**
69.83354%	71.15987%	66.97374%	70.65868%	70.70474%	70.65868%

Observing the results presented in Table 2, it is evident that the PageRank system appears to describe the matches which were used to generate it better than the rest of the systems. Also the SortRank-Uncombined system seems to perform worse than the rest of the systems – something which is expected as the "*Uncombined*" model which used to generate the rankings also performs poorly. The LadderRank-Combined system appears to be in second place with marginally better performance when the allowed challenge positions, x=3. Both the LadderRank-Combined system and the PageRank system outperform the ATP Official Rankings.

In an attempt to provide further evidence that the ATP ranking system is inaccurate at ranking players with rankings greater than 32, we selected a subset of the matches played only in-between players ranked 32-80 according to the ATP Official Rankings on the 18th of March 2013. This subset was comprised of 275 matches that were played within the period 18/03/2012-18/03/2013. The small size of this subset also hints to the problem of the seeded tournament system as it is only 275 matches out of a total of 2552 that were played in the period. This means that this group of players play a much smaller number of matches between them and as a result there are not enough matches to compare the performance of these players against one another.

Table 3 presents how the ranking systems perform at predicting the outcomes of this subset of matches.

Table 3 – Comparison of predictive power of ranking models over 275 matches played between players ranked in positions 32 to 80 by ATP rankings in the period 18/03/2012 to 18/03/2013.

ATP	**Match PageRank**	**SortRank Uncombined**	**LadderRank Combined (x=1)**	**LadderRank Combined (x=3)**	**LadderRank Combined (x=5)**
55.27273%	58.54545%	54.54545%	56.00000%	56.3634%	56.00000%

The ATP Official rankings perform much more poorly in this subset of matches with a success rate as low as 55.27273%. The other models also perform much worse than when using the full range of players but still outperform the ATP Rankings. This generic drop in the success rate of the ranking systems to reflect the outcomes of matches played by players ranked in the range 32-80, could occur for a number a of reasons. It could be because the players of this range are more unstable in their performance which adds to the uncertainty of the outcome. Also, since we are comparing a group of players which are more similar to each other, the outcomes in the matches played between players in this group would also have increased uncertainty. Additionally, the small number of matches played between these players also affects the quality of the models: the PageRank model uses match victories to rank players and the "*Combined*" model uses average statistics from these matches. In other words, the seeded tournament system affects all these ranking systems as players ranked 32-80 face each other a lot less, thus reducing the quality of statistics available for these players.

6. Conclusion

We introduced a new, flexible idea for ranking professional tennis players by simulating a "*sports-ladder*" driven by a tennis model in the background. We demonstrated this idea by using existing models from the literature and comparing the rankings that they generate with the official ATP Rankings. We identified problems such as the slow adaptation of the LadderRank-Combined system and discussed how they could be solved by changing the underlying model to account for them.

Despite the slow-adapting underlying model that we used, comparing the LadderRank-Combined system's performance against the ATP rankings in terms of how well the rankings represent the set of matches used to generate them, the LadderRank algorithm outperformed the ATP rankings.

We also detected the bias created by seeded tournaments which is inherent in the official ranking systems and we provided evidence which support this. By simply comparing the differences in the rankings assigned to players by the various ranking systems we provided evidence towards the bias by detecting an explosion of disagreement with the ATP for players ranked greater than 32. By testing the performance of the ATP Ranking system on a subset of matches that were played between players ranked in the range 32-80, we found further evidence of the poor representation those same players have in the official rankings.

To sum up, even though the "*LadderRank*" ranking system joined with the "*Combined*" model does not perform as well as the PageRank ranking system, it still outperforms the ATP Official Rankings. This proves that it works as an idea. Also, since the quality of the rankings generated by the "*LadderRank*" system is directly dependent on the quality of the model that drives the comparisons of players, by using a more sophisticated model one can improve the performance further and this is exactly what makes "*LadderRank*" so flexible.

References

Barnett, T. J. & Clarke, S. R., (2002) Using Microsoft Excel to model a tennis match. *6th Australian Conference on Mathematics and Computers in Sport (G. Cohen ed.),* pp. 63-68.

Brin, S. & Page, L., (1998) The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems,* Volume 30, pp. 107-117.

Clarke, S. R., (1994) *An adjustive rating system for tennis and squash players.* Queensland, Australia, Bond University, pp. 43-50.

Clarke, S. R. & Dyte, D., (2000) Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research,* 7(6), p. 585.

del Corral, J. & Prieto-Rodriguez, J., (2010) Are differences in ranks good predictors for Grand Slam Tennis matches?. *International Journal of Forecasting,* Volume 26, pp. 551-563.

Dingle, N., Knottenbelt, W. & Spanias, D., (2013) On the (Page)Ranking of Professional Tennis Players. In: M. Tribastone & S. Gilmore, eds. *Computer Performance Engineering.* s.l.:Springer Berlin Heidelberg, pp. 237-247.

Hoare, C. A. R., (1961) Algorithm 64: Quicksort. *Commun. ACM,* 4(7), p. 321.

Klaassen, F. J. G. M. & Magnus, J. R., (2001) Are Points in Tennis Independent and Identically Distributed? Evidence from a Dynamic Binary Panel Data Model. *Journal of the American Statistical Association,* 96(454), pp. 500-509.

Radicchi, F., (2011) Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. *PLoS ONE,* Volume 6, p. e17249.

Spanias, D. & Knottenbelt, W., (2012) Predicting the outcomes of tennis matches using a low-level point model. *IMA Journal of Management Mathematics.*