An approach for the ATP players' strength model

1 Introduction

Tennis ...

2 Problem Formulation

The objective is to build a model that takes as input an ATP tennis player P_i^t at a given time t and returns a value S_i^t representing the player's strength at time t. S_i^t should be updated periodically. Let T be the period (assumed constant, e.g., one week). For simplicity of notation, we denote $S_i^0, S_i^1, S_i^2, ..., S_i^{N_T}$ to represent $S_i^0, S_i^T, S_i^{2T}, ..., S_i^{NT}$, which are the strengths of player P_i respectively after $1, 2, ..., N$ periods.

For $1 \leq t+1 \leq N_T$, the strength of player P_i at time $t+1$, S_i^{t+1}, may depend on past values $S_i^0, ..., S_i^t$, his tournament results during the period between t and $t+1$ (score, match statistics, strength of his opponents, etc.), and other potential factors.

Let N_p be the total number of players and M the maximum number of matches a player P_i can play in a period T. We denote

$$\mathcal{O}_i^t = \{O_{i1}^t, ..., O_{iM}^t\}$$

as the list of matchups of player P_i during the period between t and $t+1$. Each matchup O_{ij}^t is represented by a triplet

$$(P_i^t, P_j^t, \chi_{ij}^t)$$

where P_j^t is the opponent of P_i^t during match j, and χ_{ij}^t is an indicator that equals 1 if player P_i played match j and 0 otherwise (note that not all players necessarily play all M matches in a given period, for instance, a player eliminated in the first round of a two-week Grand Slam tournament).

Two questions then arise:

- How to mathematically represent a player i at a given time t, $P_i^t, 1 \leq i \leq N_p, 0 \leq t \leq N_T - 1$?
- Given $\mathcal{S}_i^t = \{S_i^0, ..., S_i^t\}$ and $\mathcal{O}_i^t$ and assuming there exists an update function $\mathcal{F}$ such as

$$S_i^{t+1} = \mathcal{F}(\mathcal{S}_i^t, \mathcal{O}_i^t) \quad \forall 0 \leq t \leq N_T - 1,$$

how to approximate $\mathcal{F}$?

3 Evaluation Criteria

If the system is good, it should accurately predict the winner of a match. More rigorously, a good player strength model is one that, given two players P_i^t and P_j^t at a time t who compete in a match m during the period between t and $t+1$, ensures that if the winner of m is i, then $S_i^t > S_j^t$.

Although ATP Points are not a ground truth for player strength, they remain a robust, transparent, and fairly reliable system that is widely adopted by the tennis community. Thus, the player ranking obtained from the constructed player strength model should have a reasonably strong Spearman correlation with the ATP ranking.

This would also help ensure that the model's accuracy in match prediction is not merely due to chance. Indeed, note that a model that randomly predicts the winner of a tennis match already has an accuracy of 50%; it's a very poor model while a 50% of accuracy is not that bad.

4 Existing Systems

- ATP Points
- Elo Ratings

4.1 ATP Points

4.1.1 A brief overview

The ATP Points system is the reference ranking system for tennis players. It is updated every week (the maximum duration of an ATP tour tournament) throughout the year, except during the four Grand Slam tournaments, where it is updated every two weeks (the duration of a Grand Slam).

Each week of a given year Y_n, players lose all the points they earned in the same week of the previous year Y_{n-1} and gain new points by participating in tournaments held during the same period in year Y_n. The points earned depend on the tournament category and the stage reached by the player in the competition. Winning an ATP tournament can award between 250 points for the lowest-level tournaments and up to 2000 points for a Grand Slam. The point allocation rule is simple: each competition level grants a fixed number of points depending on whether the player wins or loses the match. There is no accumulation of points over multiple rounds in a tournament; points are only awarded when the player is eliminated. At that point, the player receives the points corresponding to the round reached, which are naturally higher than the points awarded to players eliminated in earlier rounds. The tournament winner earns the total number of points assigned to that tournament category (e.g., 2000 for a Grand Slam). For each tournament, the points awarded at each round are predefined, fixed, and known to all players. Players accumulate points by participating in ATP tournaments throughout the year.

Let's take the example of Carlos Alcaraz: Before the two weeks of Roland Garros 2024, he had accumulated 7300 ATP points. In 2023, he was eliminated in the semifinals, earning him 720 ATP points. In 2024, he won Roland Garros, which awarded him 2000 points. Thus, his total number of points after these two weeks of Roland Garros was $7300 - 720 + 2000 = 8580$ ATP points.

Note that this description of the ATP points system does not mention the rules for initializing points for new players entering the circuit.

Table 1 shows the number of ATP points awarded at each round in a Grand Slam tournament. A player earns a fixed number of points based on the round they reach. If a player loses in a given round, they receive the corresponding "L" (Loss) points. The tournament winner receives the maximum 2000 points.

Round	**Win (W)**	**Loss (L)**
Final (F)	2000	1200
Semifinal (SF)	1200	720
Quarterfinal (QF)	720	360
Round of 16 (R16)	360	180
Round of 32 (R32)	180	90
Round of 64 (R64)	90	45
Round of 128 (R128)	10	0

Table 1: ATP Points Distribution in a Grand Slam

4.1.2 ATP Points as a proxy for player strength

ATP points can be seen as a way to quantify a player's strength. Indeed, it is reasonable to assume that the stronger a player is (respectively weaker), the more matches they win (respectively lose) and the more points they gain (respectively lose). Thus, the ATP points system can be considered a good approximation of $\mathcal{F}$ 2. In this system, $T = 1$ week (2 for Grand Slams), $\mathcal{O}^t$ represents the player's opponents in the tournament they participate in during the week, and $\mathcal{F}$ depends only on the match outcomes (win or loss).

4.2 Elo Ratings

The Elo rating system is a widely used method to calculate the relative skill levels of players in two-player games such as chess, and it has also been adapted to rank tennis players. The system is based on the premise that the difference in ratings between two players gives a strong indication of the expected outcome of a match. The system uses match results and a player's current rating to update their rating after each game.

4.2.1 Mathematical Description of the Elo Rating System

The Elo rating system works by comparing the ratings of two players before a match and adjusting them according to the result. Let R_i^t represent the Elo rating of player P_i at time t, and let R_j^t represent the Elo rating of their opponent P_j at the same time. The expected score E_{ij} for player P_i against player P_j is given by the following formula:

$$E_{ij} = \frac{1}{1 + 10^{(R_j^t - R_i^t)/s}}$$

Where:

- E_{ij} is the expected probability of player P_i winning against P_j.
- R_i^t and R_j^t are the Elo ratings of players P_i and P_j at time t.
- s is a scaling factor.

After the match, the actual outcome is used to adjust the ratings. If player P_i wins, their Elo rating increases, and if they lose, it decreases. The rating update is calculated using the following equation:

$$R_i^{t+1} = R_i^t + K \cdot (S_{ij} - E_{ij})$$

Where:

- R_i^{t+1} is the new rating of player P_i at time $t + 1$.
- K is a constant that controls the maximum possible change in rating.
- S_{ij} is the actual outcome of the match: $S_{ij} = 1$ if P_i wins, $S_{ij} = 0$ if P_i loses, and $S_{ij} = 0.5$ in the case of a draw.

The value of K is a hyperparameter that controls how sensitive the rating system is to the results of matches. A larger value of K results in larger changes to a player's rating after each match, while a smaller value of K results in smaller changes.

4.2.2 Elo Ratings as an Update Function for Tennis Players Strenght

The Elo rating system can be directly applied as an update function $\mathcal{F}$ in the context of tennis. In this case, we treat the Elo rating of each player as a representation of their strength, and we update it after each match based on the result. The system assumes that player strengths are reflected by their Elo ratings, and it calculates the expected outcome based on the difference in ratings between the players involved.

Let S_i^t represent the strength (rating) of player P_i at time t. The goal is to update S_i^t based on the results of the matches played by player P_i between time t and $t + 1$. This update function is as follows:

$$S_i^{t+1} = \mathcal{F}(S_i^t, \mathcal{O}_i^t)$$

Where:

- S_i^t is the strength (Elo rating) of player P_i after the match t.
- $\mathcal{O}_i^t$ represents the only one opponent of the player P_i during the match $t + 1$.

Each match result affects the rating of player P_i according to the Elo update rule. For example, after a win, the player's rating increases, and after a loss, it decreases, based on the difference in ratings between the two players involved. The update process is straightforward: the stronger player is expected to win, and the weaker player's rating will adjust accordingly.

The primary factors influencing the strength in this system are:

1. **Period Choice**: The time period can be chosen to represent different frequencies of updates. For instance, a single match, multiple matches, or a longer period such as a week or month can be used to update the player's Elo rating. Typically, updates occur after each match, but they can also be aggregated over a period for a more smoothed estimate of a player's strength.

2. **Match Results**: The Elo rating system updates the player's strength based on their results in matches between t and $t+1$. If a player wins a match, their Elo rating increases, while a loss results in a decrease.

3. **Opponent Strength**: The previous strength of the opponent is factored into the rating adjustment. If a player defeats a highly-rated opponent, their strength will increase more than if they beat a lower-rated opponent. This dynamic ensures that the Elo system appropriately rewards players for defeating strong competitors and penalizes them for losses to weaker players.

5 Tennis Strength Model: A Hybrid System Combining Elo and ATP Points

5.1 Limitations of ATP Points and Elo Ratings

The ATP Points system, widely used to rank professional tennis players, is based on players' performance in individual tournaments. While it provides a solid foundation for ranking players based on tournament results through it simple yet interpretable points updating system, it lacks the ability to fully reflect the relative strength of players beyond tournament victories. Specifically, it not consider factors such as match dominance or players level gap, which are important to represent a player'strength.

On the other hand, the Elo rating system overcomes some of these limitations by incorporating the concept of expected scores, which are based on the difference in ratings between two players. The Elo system gives a more dynamic representation of a player's strength, allowing for changes after each match based on the result and the gap relative to the opponent. However, despite its advantages, Elo does not account for the specifics of match statistics, such as the number of unforced errors or winners, which can provide crucial insight into a player's true strength during a match. Furthermore, Elo fails to consider the dynamic nature of a player's form over a given tournament or season.

To address the shortcomings of both the ATP Points and Elo systems in representing a player strenght, we propose a hybrid model, combining the best features of both systems while overcoming their limitations. The goal of this model is to provide a more accurate and dynamic representation of a player's strength, taking into account not only the match results but also the match statistics (the way a player wins), and the momentum within a given tournament. This model also integrates the ATP Points system to preserve alignment with existing ATP ranking frameworks, ensuring consistency across the sport.

5.2 Towards a Hybrid System for Tennis Players Strenght

The following changes were made to the Elo system to enhance its ability to model a tennis player strenght.

5.2.1 Dominance Factor

Replacing the Binary Actual Score with a Dominance Factor In the traditional Elo system, the actual score is binary: 1 if a player wins, and 0 if a player loses. However, this simplification neglects crucial information about how a player wins a match. For example, a 6-0 6-1 victory is not comparable to a 6-7 7-6 7-6 win, even though both are technically wins. The dominance factor introduced in this model quantifies the extent to which a player dominates their opponent during a match. This allows the model to reflect more accurately the relative strength of the players, as a win by a wide margin generally indicates a stronger performance than a narrow win.

Formula for Dominance Factor and Estimation via Logistic Regression The dominance factor for each player is defined as x_i for player i, where x_i represents the dominance of player i over player j during the match. The sum of the dominance factors for both players equals 1, i.e., $x_i + x_j = 1$. The dominance factor is computed using a softmax function applied to a feature vector V_i, which represents various match statistics such as the number of first serves, winners, unforced errors, etc. The formula for calculating the dominance factor is:

$$x_i = \frac{1}{1 + e^{V_i \theta}}$$

Here, θ is the parameter vector of weights learned using a logistic regression model trained on post-match winner predictions. This logistic regression model estimates the weights of each stat on the dominance of a player, effectively capturing how match statistics influence the outcome.

5.2.2 A New Expected Score for Better Player Strength Modeling

A key aspect of Elo is the concept of expected scores, which are based on the difference in ratings between two players. In this hybrid model, the expected score is replaced with a more nuanced measure that combines two different sources of information: (1) the historical ratings of the players, and (2) the recent momentum of the players in the ongoing tournament. This approach ensures that both the players' relative strength and their recent performance are factored into the expected score, resulting in a more accurate reflection of their true strength at any given point in time.

We compute a new expected score with the geometric mean of two components:

- $E_{ratings}$: A score based on the players' historical ratings, similar to the standard Elo formula. In the traditional Elo system, the expected score between two players is determined by the difference in their ratings using the formula:

$$E_{\text{Elo}} = \frac{1}{1 + 10^{\frac{R_j - R_i}{s}}}$$

where R_i and R_j represent the ratings of players A and B, respectively. The use of s is because the formulation relies on an exponential function, which can lead to numerical instability or overflow issues when the ratings become very large (or very small), especially in cases of highly skewed ratings between two players. This can result in impractical rating updates or computational difficulties. To mitigate these issues, we replace the exponential function with a square root function. This adjustment provides a more stable and manageable way of computing the expected score between players without needing an arbitrary scaling factor s, while still well-refectling the gap between the two players. The updated expected score formula is given by:

$$E_{\text{ratings}} = \frac{1}{1 + \sqrt{\frac{R_j}{R_i}}}$$

This change has also the advantages of providing better interpretability as it eliminates arbitrary scaling factors, and it creates a smoother, more transparent and more controlled relationship between players' ratings and expected scores. But it assumes the ratings should always be positive.

- $E_{momentum}$: A score based on the players' momentum, which is the average of their actual scores in the ongoing tournament. For each match, the values for the two players are normalized with a softmax function to sum to 1.

The geometric mean is used instead of the arithmetic mean to prevent scores from becoming biased or too centered. The idea is that the strength of a player should be elevated or diminished simultaneously based on both factors (ratings and momentum), rather than one factor dominating. The geometric mean is computed as follows:

$$E_{\text{expected}} = \sqrt{E_{\text{ratings}} \cdot E_{\text{momentum}}}$$

This formulation ensures that a player's form during a tournament (momentum) is properly integrated with their historical strength, leading to a more meaningful strength value.

5.2.3 Use of ATP Updating Points Instead of a Custom K

A significant feature of this hybrid model is the use of ATP Points as the K-value in the Elo update. The K-value in the Elo system controls the amount by which a player's rating changes after a match. However, choosing an appropriate K-value is not trivial and can introduce arbitrariness into the model. By using the ATP Points system, which already reflects the relative importance of various tournaments, the model benefits from a more interpretable and consistent updating factor. This approach ensures that the ratings are aligned with the actual structure of the ATP tour, taking into account the varying significance of different tournaments.

Hence, the maximum number of strength points a player can gain in a match is set equal to the number of ATP Points awarded for the specific round of the tournament. This ensures that the system remains in synchronization with the ATP ranking framework, avoiding the need for arbitrary K-factors and non interpretable grid search techniques. The use of ATP Points also allows for a more meaningful comparison of player strengths across different tournaments and could facilitates the adoption of the model by the tennis community.

5.2.4 Points Update Rule

Several update rules have been tested in this model, each with its own advantages and limitations. Below, we describe and compare these different approaches.

Absolute Difference A variation of the original Elo method is to use the absolute difference:

$$R_i^{t+1} = R_i^t + K \cdot |\text{actual\_score} - \text{expected\_score}|. \tag{1}$$

Advantages:

- Ensures that rating values always increase, avoiding unexpected rating losses.

Limitations:

- Ignores the direction of the rating adjustment, potentially over-rewarding small upsets.

Arithmetic Mean Another tested approach involves the arithmetic mean of actual and expected scores:

$$R_i^{t+1} = R_i^t + K \cdot \frac{\text{actual\_score} + \text{expected\_score}}{2}. \tag{2}$$

Advantages:

- Smoothens rating updates, avoiding abrupt changes.
- Still accounts for both expected performance and actual match results.

Limitations:

- Tends to center values around 0.5, potentially underestimating high performers.

Geometric Mean The geometric mean of the expected and actual score:

$$R_i^{t+1} = R_i^t + K \cdot \sqrt{\text{expected\_score} \cdot \text{actual\_score}}. \tag{3}$$

Advantages:

- Ensures that both expected and actual scores contribute meaningfully to updates.
- Reduces the impact of extreme variations, providing a more stable progression.

Limitations:

- Can lead to conservative updates when one of the two terms (expected score or actual score) is very low, potentially slowing down the rating adjustments for emerging players or in cases of significant upsets.

Unlike the classical Elo system, these three update methods do not inherently include rating losses over time, which is a common limitation. This means that once a player accumulates points, they cannot lose them anymore. To address this, one could imagine alternative formulations of the update that enables point losses. Another approach could be inspired by the ATP Points System, which applies a rolling window mechanism where points earned in year Y_{n-1} gradually decay and are removed from a player's total during year Y_n.

5.3 Conclusion

The hybrid Tennis Strength Model presented in this section combines the strengths of both the ATP Points and Elo rating systems while addressing their limitations. By incorporating a match dominance factor, a more strength-aware expected score formula, and the ATP Points update mechanism, this model provides a more accurate and dynamic representation of a player's true strength. It takes into account not only match results but also match statistics, player momentum, and the varying significance of different tournaments, resulting in a ranking system that is both interpretable and effective.

5.4 Implementation details and Results

[ongoing, ..., github_link soon]

6 End-to-End Strength Model

Another approach to developing a player strength model would be to directly train a machine learning model on next-match winner prediction using past match data. Such a model should capture information about player strength, given that it predicts the winner of a match based on historical player data. Indeed, if X represents the feature matrix of historical data for a set of matches and Y is the binary target vector for the next-match results, then such a model $\mathcal{M}$ would predict, for any match m_i represented by a feature vector x_i, $y_i = \mathcal{M}(x_i)$, the probability that Player 1 wins the match. The probability that Player 2 wins would therefore be $1 - y_i$.

One would then need to find a way to extract the information about player strength captured by the model, if such information is indeed present. In fact, the assumption that such a model captures player strength is purely hypothetical, as a machine learning model, by nature, has the freedom to manipulate the provided features in any way necessary to optimize its cost function. In this case, the cost function is the likelihood of the Bernoulli distribution of match outcomes (win or loss).

Thus, in such an approach, one would need to either:

- Design the model a priori in a way that forces it to capture player strength information and use it for prediction, then retrieve the player strength values a posteriori.
- Build a model solely to maximize its accuracy in next-match winner prediction and then find a way to extract, a posteriori, something resembling a measure of player strength.

7 On the Importance of a Simple and an Interpretable System

The main goal of building a player strength model would be to have it adopted by the tennis community as a ranking system for players. Therefore, it is crucial to design a system that is fair, transparent, interpretable, and as simple yet accurate as possible, so that it is understandable to players, coaches, fans, and other tennis stakeholders.

Although the ATP Points system has limitations, as mentioned earlier, its simplicity and transparency make it accessible to all tennis stakeholders, even those with minimal technical background. Any player can easily calculate their points and ranking from match to match. This makes it a robust system that is difficult to replace.
One can draw an analogy with the point allocation system in football leagues, where a win grants +3 points, a loss gives 0 points, and a draw awards +1 point: simple and easy to understand for everyone.