Tennis Prediction

April 14th 2020

1 Introduction

In this report we improve upon the work done by Jacob Gollub by introducing three additional features to the modeling of K-factor in the calculation of players' elo ratings. In section 3 we demonstrate that in pre-match prediction, our method has increased Jacob's surface elo 538 baseline accuracy by approximately 1%.

2 Features

The features we adopt are from the Ultimate Tennis Statistics Blog[1]. From the nine methods of scaling K-factor introduced by the website, we pick the three features that are supported by our ATP dataset and at the same time show best performance on the dataset. Specifically, we pick players' current elo ratings (S_{rating}), tournament levels ($S_{tny-level}$), and tournament rounds ($S_{tny-round}$). On top of the original formula for K_{it}, we scale K_{it} with the average of S_{rating}, $S_{tny-level}$, and $S_{tny-round}$.

$$E_i(t+1) = E_i(t) + K_{it} * (W_i(t) - \hat{\pi}_{ij}(t))$$

$$\hat{\pi}_{ij}(t) = (1 + 10^{\frac{E_j(t)-E_i(t)}{400}})^{-1}$$

$$K_{it} = \frac{250}{(5 + m_i(t))^{.4}} * avg(S_{rating}, S_{tny-level}, S_{tny-round})$$

The calculation of these three features are specified in the following subsections. Overall, the new elo k-factor formula reflects better predictability and portrait of players' abilities across eras.

[1] https://www.ultimatetennisstatistics.com/blog?post=eloKfactorTweaks#

2.1 Player's Current Elo Ratings

Adjustments on K based on the player's current elo ratings allow lower ranked players to improve their ratings more rapidly to escape the unstable zone of the ranking and alleviate the unfair disadvantage of having little records for newcomers, and at the same time stabilizes the higher ranked players.

$$S_{rating} = 1 + \frac{18}{1 + 2^{\frac{rating-1500}{63}}}$$

2.2 Tournament Level

The tournament levels we considered are: Grand Slam, Tour Finals, Masters, Olympics, and ATP 500. The motivation of this scaler is that a player's performance is different in different levels of tournaments because some are more competitive. Note that in the ATP dataset, Grand Slam includes Australian Open, US Open, Wimbledon, and Roland Garro. Adjustments based on tournament levels is:

$$S_{tny-level} = \begin{cases} 1 & \text{Grand Slam} \\ 0.9 & \text{Tour Finals} \\ 0.85 & \text{Masters} \\ 0.8 & \text{Olympics} \\ 0.75 & \text{ATP 500} \\ 0.7 & \text{others} \end{cases}$$

2.3 Tournament Rounds

The tournament rounds we considered are: Final, Semi-Final, Quarter-Final and Round-Robin, Rounds of 16 and 32, Rounds of 64 and 128, and For Bronze Medal. The motivation is straight forward. Playing well in the Final should weigh more than playing well in Rounds of 16 and 32.

$$S_{match-level} = \begin{cases} 1 & \text{Final} \\ 0.9 & \text{Semi-Final} \\ 0.85 & \text{Quarter-Final Round-Robin} \\ 0.8 & \text{Rounds of 16 and 32} \\ 0.75 & \text{Rounds of 64 and 128} \\ 0.95 & \text{Bronze Medal} \end{cases}$$

3 Results

In this section we present our experimental results. The dataset we used is the ATP dataset from year 2001 to 2019. During the experiment, we recognize that the prediction accuracy on matches with no surface information is significantly higher than the prediction accuracy on matches with surface information. To demonstrate our findings, in figure 2 we separately present the accuracy of all matches (labeled as w/ NaN) and surface-specific matches (labeled as w/o NaN).

Year	Elo 538 Baseline	Surface Elo 538 Baseline	Surface Elo 538 Baseline Hard	Surface Elo 538 Baseline Grass	Surface Elo 538 Baseline Clay	Surface Elo 538 Baseline NaN	Bookmaker
2010	0.656	0.689	0.689	0.684	0.711	0.821	0.725
2011	0.672	0.705	0.724	0.712	0.691	0.823	0.734
2012	0.684	0.705	0.711	0.716	0.703	0.806	0.728
2013	0.675	0.684	0.693	0.747	0.682	0.805	0.720
2014	0.675	0.689	0.720	0.702	0.671	0.814	0.719
2015	0.684	0.702	0.704	0.701	0.690	0.815	0.726
2016	0.685	0.688	0.701	0.694	0.685	0.776	0.722
2017	0.665	0.675	0.689	0.690	0.666	0.780	0.690
2018	0.650	0.653	0.670	0.663	0.660	0.761	0.691
2019	0.640	0.657	0.655	0.688	0.664	0.749	0.684

Figure 1: Surface-Specific Elo Ratings

Year	Elo Baseline	Elo 538 Baseline	Surface Elo 538 Baseline	Surface Elo 538 Baseline	Surface Elo 538 + 3 scalers	Surface Elo 538 + 3 scalers	Bookmaker
Surface	All	All	w/o NaN	All	w/o NaN	All	All
2010	0.676	0.669	0.689	0.720	0.695	0.755	0.725
2011	0.691	0.691	0.705	0.754	0.713	0.764	0.734
2012	0.695	0.695	0.705	0.746	0.709	0.755	0.728
2013	0.687	0.683	0.684	0.731	0.695	0.746	0.720
2014	0.684	0.687	0.689	0.742	0.700	0.753	0.719
2015	0.691	0.693	0.702	0.745	0.699	0.750	0.726
2016	0.687	0.686	0.688	0.717	0.695	0.724	0.722
2017	0.668	0.668	0.675	0.706	0.682	0.717	0.690
2018	0.651	0.646	0.653	0.688	0.666	0.701	0.691
2019	0.643	0.640	0.657	0.689	0.662	0.695	0.684

Figure 2: Summary of prediction accuracy on the ATP dataset