

# “The Thin Edge of the Wedge”: Accurately Predicting Shot Outcomes in Tennis using Style and Context Priors

Xinyu Wei<sup>1</sup>, Patrick Lucey<sup>2</sup>, Stuart Morgan<sup>3</sup>, Machar Reid<sup>4</sup> and Sridha Sridharan<sup>1</sup>

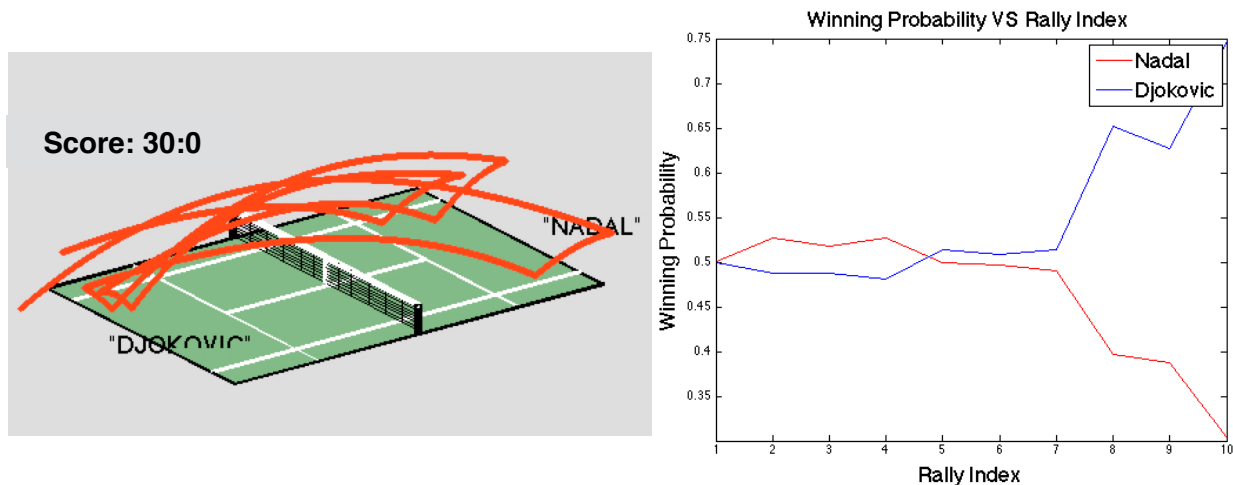
<sup>1</sup>Queensland University of Technology, <sup>2</sup>STATS LLC, <sup>3</sup>Australian Institute of Sport, <sup>4</sup>Tennis Australia  
Email: xinyu.wei@connect.qut.edu.au, plucey@stats.com, stuart.morgan@ausport.gov.au, mreid@tennis.com.au, s.sridharan@qut.edu.au

## Abstract

The aim of this paper is to discover patterns of player movement and ball striking (short-and long-term shots, and shot combinations) in tennis using HawkEye data which are indicative of changing the probability of winning a point. This is a challenging task because: i) behavior can be unpredictable, ii) the environment is dynamic and the output state-space is large and iii) examples of specific interactions between agents may be limited or non-existent (player A may not have interacted with player B). However, by using a dictionary of discriminative patterns of player behavior, we can form a representation of a player’s *style*, which is interpretable latent factors that allows us to *personalize* interactions between players based on the match *context* (opponent, match-score). This approach allows us to perform superior point predictions, and to understand how points are won by systematically creating and exploiting spatiotemporal dominance.

## 1. Introduction

Tennis is a game of strategy, where top players systematically attempt to manoeuvre their opponents into positions of weakness, before capitalising with a final winning stroke. While earlier work has often focussed on the characteristics of “winners” [5, 6], we recognize that the winning shot is often the least strategically-important shot in the rally - rather it is the strategy over the preceding strokes to move an opponent out of contention in a rally that matters. Therefore the aim of this paper is to discover the underpinning strategy in patterns of tennis stroke play used by top players to systematically win points.



**Figure 1:** (Left) We visualize the shot trajectories between Nadal and Djokovic for a complete rally (no serves). (Right) We show the probability of a player winning the point during the rally - something happens at shot 7 which changes sees Djokovic more likely to win the point. In this paper we show how we can calculate these probabilities.

2016 Research Papers Competition  
Presented by:

Understanding how points are won requires a deeper understanding of the way players manipulate their opponents to establish dominance. We present a method that discovers players' systematic shot combinations that establish spatiotemporal dominance, and demonstrate that players can be grouped by similar strategies. We refer to these strategies as "*style priors*". We also recognise that the context of the rally is important in predicting shots and understanding style priors, and thus we refer to these as "*context priors*".

Our core idea is illustrated in Figure 1, where we show a rally between Rafael Nadal and Novak Djokovic. On the left we show the shot trajectories between the two players, and on the right we show that the estimated likelihood for each player of winning the point varies throughout the point. For the majority of the rally it appears to be a stalemate, until the 7th stroke where something happens allowing Djokovic to take control and eventually win the point. We call this change in winning probability *the thin end of the wedge* - the winning player ekes out a small advantage, then systematically capitalizes on that advantage to secure the point. Using three tournaments worth of HawkEye data from recent Australian Open Tennis tournaments, we demonstrate that these insights can be used to achieve accurate prediction. Such fine-grained analysis, which is currently missing from tennis analytics, would be a extremely useful knowledge discovery tool as well as a valuable story-telling tool for broadcasters alike.

While this approach gives us insight into the systematic performance strategies by specific players, it also has an additional advantage. Even the top players only meet occasionally, and this is a problem for sports analysts since there is rarely sufficient shot and point data available to perform meaningful predictive analyses about a specific contest between two players. By using our style and context priors, however, we demonstrate that we can leverage the predictive power of much larger datasets by grouping players with similar style.

In sports such as tennis, players are often grouped according to their own style, using generic labels such as "serve-and-volley" or "baseliner". This begs the question, why are semantics such as style important? In general terms, it provides a common language which people who are familiar with the domain can use to describe the higher level aspects of play. Additionally, these labels are useful priors for coaches and players when preparing for opponents that they have not encountered previously. For example, even though a player may not have had a specific experience against an opponent, having knowledge of their style will give them a good indication of what to expect when they face them.

In this paper, we present a method which captures the "style" of a tennis player by learning a dictionary of shot trajectories directly from data. A dictionary can be thought of as a compact representation of many possible actions and contexts, that allows for efficient information retrieval. Our dictionary not only include elements of single shots but also shot combinations. Our shot dictionary is learnt by jointly optimizing prediction performance and reconstruction error. Experimental results show that our approach outperforms other dictionary learning methods for predicting shot likelihood.

In addition to style, we propose context priors that describe the specifics features of a given point scenario, such as the score-line, elapsed match time, environmental measures (wind, temperature), and the court surface. While previous work in basketball [7, 8], and soccer [9, 10, 11] have discovered the style (i.e., latent factors) of behavior using tracking data, they have treated this as a static high-level feature, rather than dynamic which changed depending on the "context". This work also follows the initial work which looked how a player's serving style varies depending on context [12]. We show that both style and context allow for better in-point prediction, which allows us to achieve our goal of determining dominant behaviors that enable players to win points.

## 2. Tennis Dataset

We used 3 years of Hawk-Eye data [1] drawn from matches that featured in the Australian Open Men's singles draw. In total, the dataset consisted 37727 shots. There were a total of 2292 winners and 2646 errors. As Grand-Slam tournaments in tennis are a *knock-out* format (i.e., if a player loses, they do not play any more matches), we focussed our analysis on the top 10 players who played the most matches in the tournament. Details of the shots of each player used in the dataset are shown in Table 1. The HawkEye system measures the  $(x, y, z)$  position of the ball as a function of  $t$ , as well as the position of each player at 20 Hz. Metadata relating to the contextual features for each point including current score, point duration, server and receiver identity are also provided which allows us to drill down to the specifics of player behavior.

Player	Total # of Shots	Winners	Errors
Djokovic	3325	167	277
Nadal	2716	176	237
Federer	1439	118	138
Murray	2277	145	163
Berdych	1164	84	107
Ferrer	856	44	70
Nishikori	698	40	62
Tomic	1974	89	188
Tsonga	1094	99	103
Hewitt	1581	67	119

**Table 1:** The players utilized in the data set.

## 3. Predicting Likelihood of Winning Point

Our task was to accurately estimate the probability that a player will win the point at any moment, given the current and previous shots in the rally. This probability will vary as the point progresses. Formally, given observations,  $\mathbf{X}$ , from past and current shots, our goal is to predict a continuous probability  $y$  where  $y$  is between 0 and 1. In training, we have the ground truth of each shot. In testing, we obtain the probability using the confidence of the classifier. To train the classifier, the first task is to form a suitable representation of the raw trajectories of the ball flight, and movement of the players. We propose two sets of features: *raw* features and *dominance* features.

### 3.1 Raw Features

Given the temporal boundaries of a shot derived by HawkEye, we estimated the ball flight trajectories for each shot. Using the ball flight information, we derived the angle, maximum height, average speed and instantaneous speed of the shot. Additionally we included the court location for each player at the start and end of each stroke. Table 2 presents a summary of these raw shot features. We only extracted features from the most recent previous shot in these experiments. Shot features prior to the most recent shot were not considered. A visualization of the shot trajectories are shown in Figure 2.

### 3.2 Dominance Features

In addition to the raw shot features, we also propose a set of dominance features that are predictive the shot outcome.

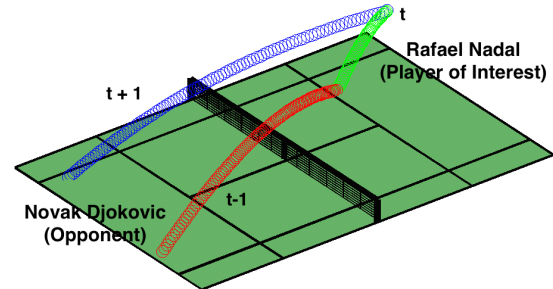
**Ground Stroke Speed Ratio:** The ground stroke speed ratio is the ratio between the ball speed of the incoming shot compared to the outgoing speed of the subsequent shot. This ratio encodes the inherent dominance achieved by the player who is striking ball more heavily than his/her opponent. This is further illustrated by the natural advantage for the

Feature	Description
Speed	Shot average speed
Angle	Angle between shot & center line
Court Position	Player and opponent court position when shot starts
Shot-Start Loc.	Location where shot starts
Shot-End Loc.	Location where shot impacts the court

**Table 2:** Descriptions of Raw Features.

2016 Research Papers Competition  
Presented by:

server (where the serve may be greater than 200km/hr) compared to the receiver who generally returns the ball at a lower relative speed. The server counters with a faster ground stroke to maintain dominance. A well timed/placed return of serve might also put the server on the back foot and reduce or reverse the dominance of the point. Using Fig. 2 as an example, the speed ratio of Djokovic would be the speed of the red arc (t-1) over the speed of the blue arc (t+1).



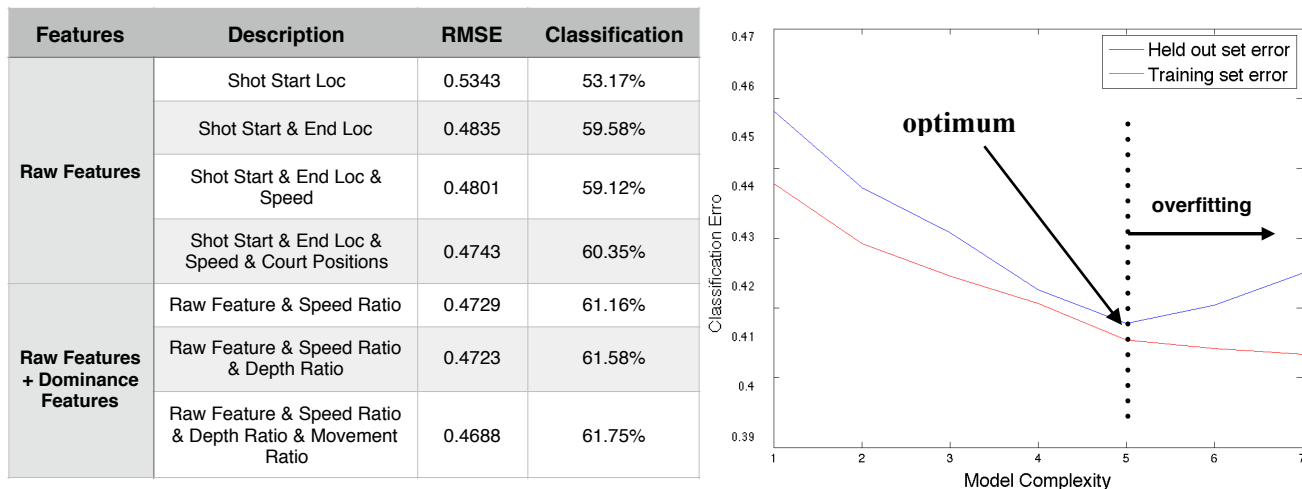
**Figure 2:** Figure shows an example of how we calculate the ground stroke speed ratio of Djokovic. It is the speed of the red arc over the speed of the blue arc.

**Ground Stroke Weight Ratio:** The ground stroke weight ratio is the ratio of the relative distance of the player to the baseline on the current stroke compared to the distance of the opponent to their baseline on the previous ground stroke. There is a strategic advantage in pressing the opponent further behind the baseline as it reduces their range of potential stroke angles, increases the difficulty of the subsequent stroke, and increases the time available to the dominant player to react to a subsequent stroke. Drop shots are also more difficult to play from a negative ground stroke depth ratio.

**Lateral Player Movement Ratio:** Lateral player movement ratio is the ratio between the lateral distance covered by the player between successive strokes. Typically the player in a dominant position in the point will move less, and maintain a more dominant central position. The less-dominant player is then forced to expend more energy in returning the ball, and may have fewer attacking options upon rushing to reach a wide stroke.

### 3.3 Baseline Experiments

Our initial experimental task was to predict the outcome of the point, and in particular the probability that the next stroke is likely to be a winner. This is essentially a classification task, and we conducted a series of experiments to report the baseline performance of our classification model. Our classifier takes the form of a Random Decision Forest, which is a non-linear classifier robust to the overfitting that might occur via bootstrapping. It also has good local-feature space adaptivity by randomly splitting the feature space at multiple levels of each tree.



**Figure 3:** (Left) Table showing the prediction performance and (Right) Figure shows the classification error against model complexity.

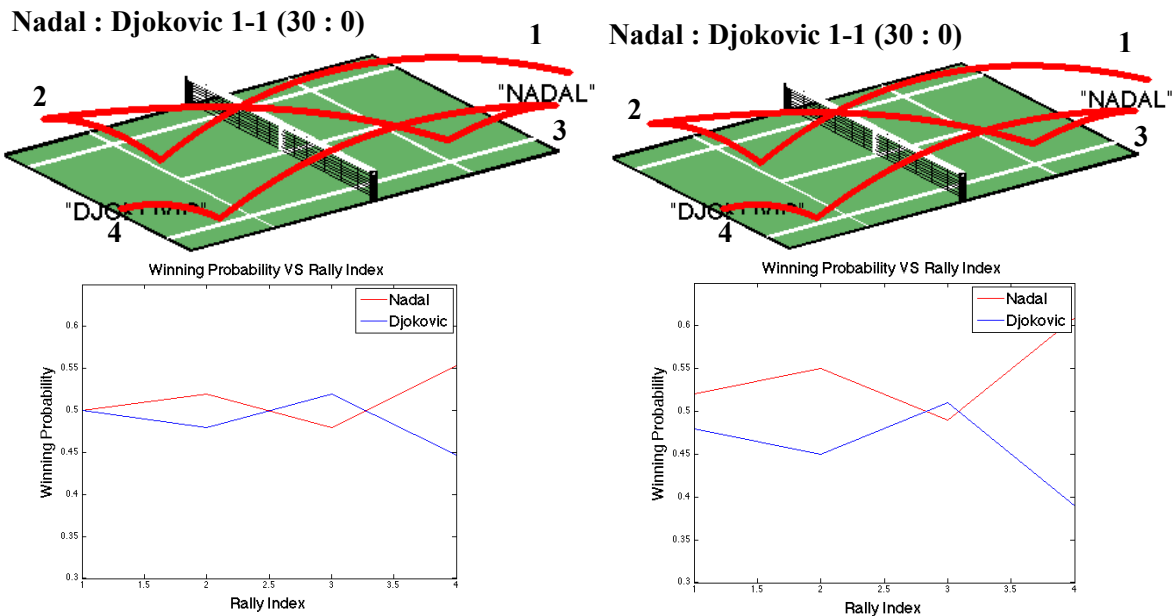
2016 Research Papers Competition  
Presented by:

We randomly split our dataset (3 years of Australian Open competition) into three sets: a training set, a validation set, and a test set. Each set included 9349 shots. We used the training set to train our model; the validation set to find the optimal hyper-parameters of our classifier; and the test set to report performance. In a Random Forest, the parameters are 1) the number of trees, and 2) the number of nodes in each min-leaf. To find the optimal parameters, we plotted the classification rate against the model complexity in Figure 3 (right). Optimal performance can be found when the number of trees is 40 and min nodes is 5.

To report the baseline performance, we use both the Root Mean Squared Error (RMSE) and classification rate. RMSE is the root of mean squared error between the predicted value and the actual value of a point outcome. For example, if player A wins a given point, and the prediction gives 70% confidence that player A will win the point, the error of this example will be  $1.0 - 0.7 = 0.3$ . In Figure 3 (left), we can see that combining both the raw and dominance features gives us best prediction performance.

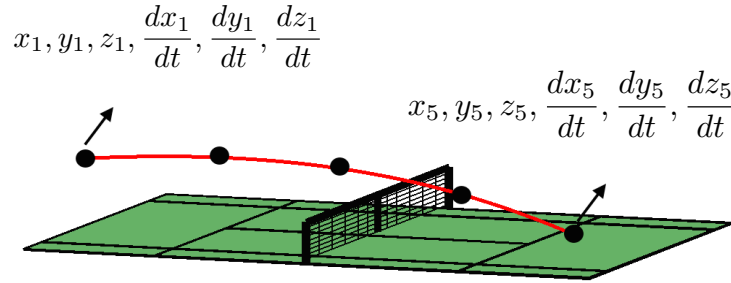
## 4. Personalizing to Specific Players and Match-Contexts

In the previous section, we trained a global model for all tennis players, underpinned by an assumption that every player performs in generically the same way - obviously a false assumption. Players tend to have their own unique style and strengths, which will change the probability depending on who they are playing. To highlight this, we show the following example in Figure 4 between Nadal and Djokovic. On the left we show our generalized model, but on the right we show our player specific model. It is important to note that the initial probabilities are different and that even though shape of the curves are roughly the same, the amplitudes are quite different. In this section, we wish to learn a more granular model by incorporating the observed styles of players. A style descriptor is a normalized frequency count of elements in the dictionary **D**. It characterizes a player's behavior and allows us to compare similarity/difference between players.



**Figure 4:** (Left) We show an example of a point using our generalized model. (Right) We have the same point where we model the specific player identities, as well as the match-context. As you can see, the probabilities are biased towards Nadal in this example and the initial probability is no longer equal. To learn these specific tendencies, we have to use a player style representation.





**Figure 5:** Figure shows an example of a ground stroke - represented as  $s$ . Here we sample 5 points from the shot ( $p = 5$ ).

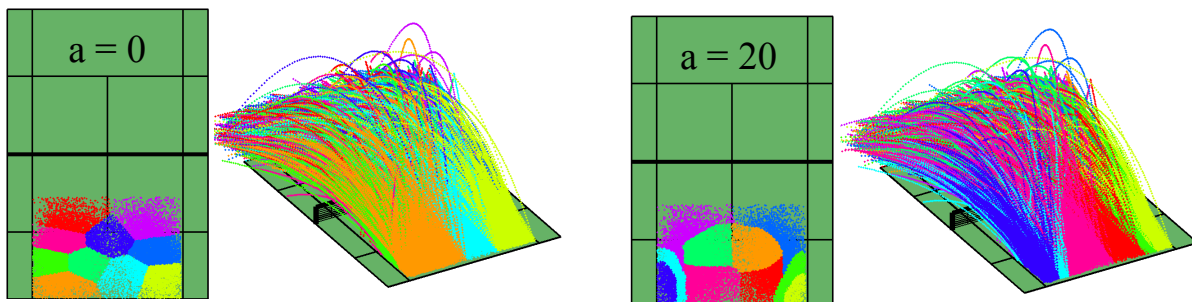
The learnt style descriptor will be used directly as an additional input feature for training a new classifier. Our dictionary included elements of both single shots as well as combinations of shots. Since our prediction performance relies on the quality of  $\mathbf{D}$ , this begs the question, what is the best way of learning the dictionary? As we want to use the dictionary to predict future behavior, it is ideal to include the prediction loss into the cost function and jointly learn the dictionary.

## 4.1 Personalizing for Player Style

### 4.1.1 Dictionary Learning for Single Shots

Let  $\mathbf{S}$  be a collection of shots. It consists of a set of  $m$ -dimensional  $N$  input signals (i.e.  $\mathbf{S} = [s_1, \dots, s_N]$ ). Each  $s_i$  is the spatiotemporal signal of a particular shot of a player. To compute  $s_i$ , we linearly sample  $p$  points from the start to the end of a shot. A point is represented by a six dimension vector which includes not only its spatial location but also its dynamics in real-world coordinates (Figure 5). We then concatenate all points of a shot into a one-column feature vector to make  $s_i$ . The most common objective for learning a dictionary is to minimize the reconstruction error which enforces the spatial consistency - Figure 6(left) shows an example of the spatial clustering result of shot bounce marks. To do this, we can learn a dictionary with  $K$  items in terms of reconstruction error can be formulated as:

$$\langle \mathbf{D}, \mathbf{U} \rangle = \underset{\mathbf{D}, \mathbf{U}}{\operatorname{argmin}} \|\mathbf{S} - \mathbf{D}\mathbf{U}\|_2^2 \quad \text{s.t. } \|u_i\|_0 = 1, \|u_i\|_1 = 1$$



**Figure 6:** Figure shows the clustering result for different alpha value when  $K = 10$ . When alpha = 0 (left), the clustering is based purely on the reconstruction error. When alpha is large (alpha = 20), the clustering is based more on the classification loss.

2016 Research Papers Competition  
Presented by:

where  $\mathbf{D} = [d_1, d_2, \dots, d_K]$  is the learned dictionary and  $\mathbf{U}$  is the assignment of shots to the dictionary and  $\|\mathbf{S} - \mathbf{DU}\|_2^2$  is the reconstruction error. The constraint sets the L0 norm and L1 norm of  $u_i$  to 1, since a shot can only be assigned to one existing item in the dictionary (exemplar based) and it can only be assigned once. This way we can maintain the semantic meaning of the dictionary. The optimum  $\mathbf{D}$  and  $\mathbf{U}$  can be found by iteratively minimizing the energy function. Each  $u_i$  is the assignment of a particular shot from a player. Essentially this is K-means clustering. However, as you can see in Figure 6(left), all shot locations are grouped into equally sized spherical groups, not taking into consideration the boundaries of the court which would be predictive of winning a point (i.e., hitting a shot near the lines of the court would correlate higher with winning the point then hitting a shot short).

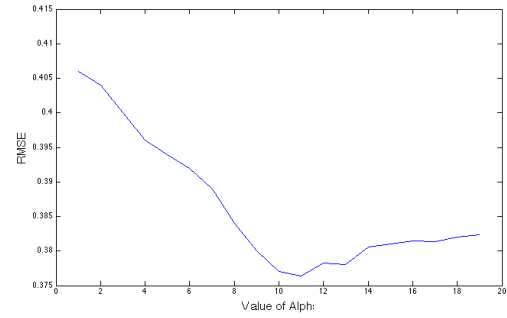


Figure 7: The RMSE against different alpha values.

Although grouping shots together in spatially coherent clusters is important, we also need to factor in the likelihood of winning a point. To do this we can include another parameter into our dictionary learning function where we also want to minimize the prediction loss which gives us a “discriminative dictionary”. We can do that as follows:

$$\langle D, U, W \rangle = \underset{D, U, W}{\operatorname{argmin}} \|\mathbf{S} - \mathbf{DU}\|_2^2 + \alpha \sum_i \mathcal{L}\{h_i, f(u_i, W)\}$$

$$\text{s.t. } \forall i, \|u_i\|_0 = 1, \|u_i\|_1 = 1$$

where  $\mathbf{L}$  is the classification loss function,  $h_i$  is the label of  $s_i$ .  $\mathbf{W}$  is the model parameters. Alpha controls the relative contribution between reconstruction and classification loss. This cost function is similar as in [3] but with a different constraint. The reconstruction error can be considered as a smoothing factor. The prediction task here is to predict whether the next shot is a winner. We use a linear predictive classifier  $F(u, W) = Wu$ . To find the optimal solution for all parameters, we can use the optimization method in [4].

Once we learn the dictionary, we can cluster shots by assigning each shot to the nearest dictionary item. Figure 6 shows the clustering results using different alpha values. When alpha = 0 (left), the clustering is purely based on the reconstruction error. When alpha is large (20), the clustering is based more on the classification loss (right) and captures the more dangerous shots. We used alpha=10 for this work (see Figure 7).

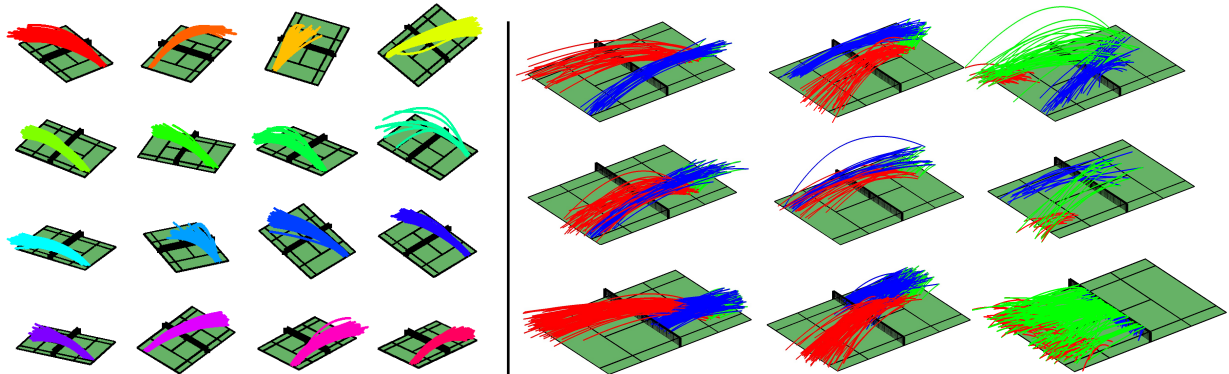


Figure 8: (Left) Single shot dictionary. (Right) Shot combos Dictionary. We use the centroid of each cluster as the dictionary element.

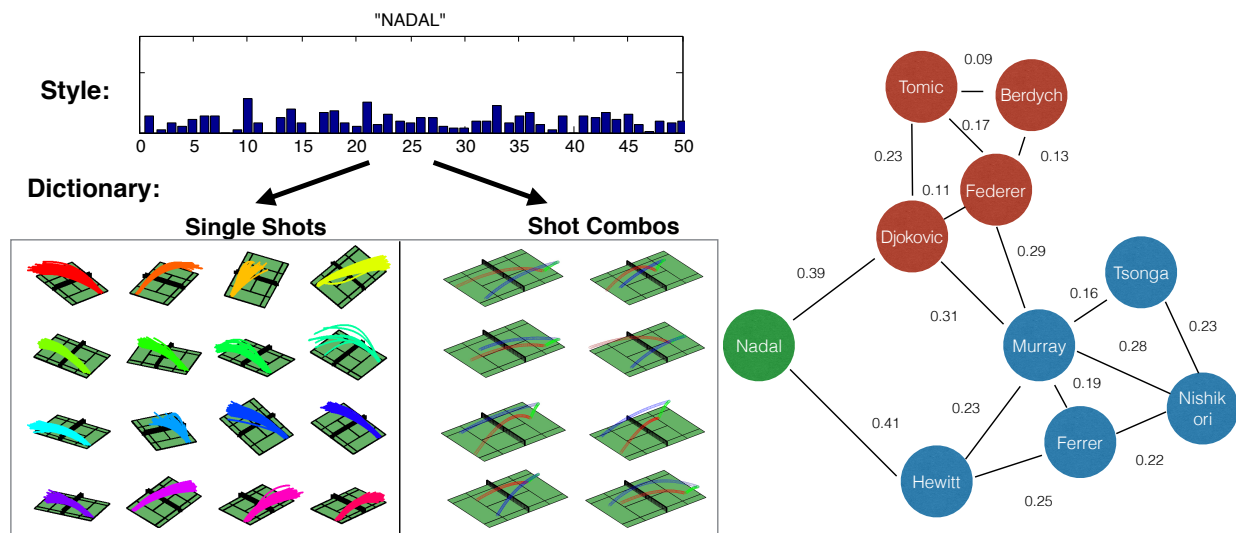
2016 Research Papers Competition  
Presented by:

#### 4.1.2 Dictionary Learning for Shot Combos

Representing the style of a player using only single shots only is not enough to encode the strategic concepts underpinning high level game play. Shot combinations better characterize player behavior as they also incorporate the temporal aspects of a player's style. We wish to learn a dictionary which can include all interesting/critical shot combinations from players. Also, since the dictionary should to be discriminative, combinations should only appear once in the dictionary. To achieve this we first detected all shot combinations that created significant change in the estimate point winning probability. We applied a hard threshold when selecting shot combinations. Once all interesting shot combos were selected, we further clustered them based on their spatial characteristics. The distance between two shot types can be measured by the mean of pairwise distances between sampled points. We employ a standard K-means algorithm to cluster all shot combinations. The center of each cluster was used to estimate the discrete elements of the dictionary. The total shot dictionary of both single shots and combinations of shots are shown in Figure 8 (left=single-shots and right=combination shots).

#### 4.1.3 Interpreting and Comparing Style Descriptors

The *style* of a player can be interpreted as the normalized frequency count of dictionary elements. In Figure 9 (left) we show the style descriptor of Nadal and can see that shots 10 and 21 are the most common shot types that he utilizes. In Figure 10, we show the style descriptors of four players - we can see that Murray's most frequent shot is shot-type 18, while Djokovic's is shot-type 3. It is also visible that Djokovic and Federer's distribution are roughly similar. Using these style descriptors allows us to quantify the similarity between players. A simple method of comparing two style descriptors is to find the sum of differences between them. A graph showing the similarity between the top 10 players is shown in Figure 9 (right). From this it can be seen that Murray and Ferrer exhibit similar styles with a raw difference of 0.19, while Nadal and Djokovic exhibit comparatively dissimilar styles with a raw difference of 0.39. A visualization of player similarities is shown in Figure 9(right). Djokovic and Federer belong to one group while Murray, Ferrer and Hewitt belong to one group. Nadal is quite different from other players.



**Figure 9:** (Left) Nadal can be described via a 50 element style feature. (Right) These descriptors allow us to compare similarity between players. We can see that Djokovic, Federer, Tomic and Berdych seem to have similar styles.



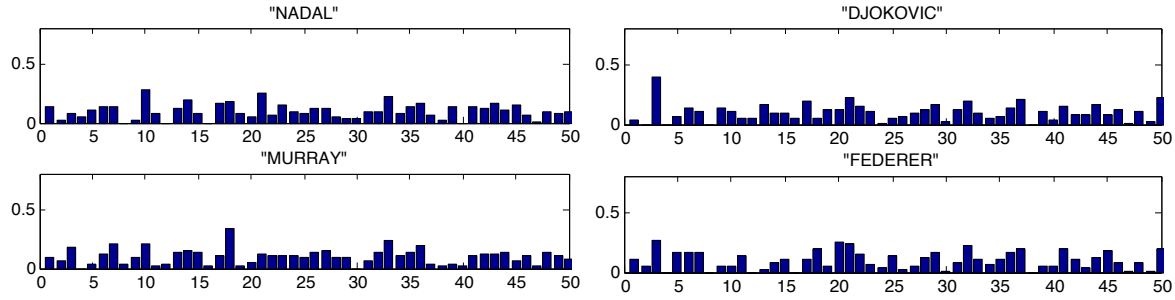


Figure 10: Style descriptors of the top 4 players

#### 4.1.4 Evaluating Style

We conducted a series of experiments to evaluate and compare different dictionary learning methods. The prediction task is the same as in Section 3, and in all experiments we included both raw features and dominance features. The additional style features are learnt with three different methods: i) Method 1 (Reconstruction error: single-shot dictionary), ii) Method 2 (Reconstruction error and prediction loss: single shot dictionary), and iii) Method 3 (Reconstruction error and prediction loss: single shot and shot combo). The results of these methods are shown in Table 3, which shows that Method 3 yields the best performance.

## 4.2 Incorporating Context

Context is another important factor which can influence the likelihood of a player winning a point. To incorporate context into our model, we proposed a set of context descriptors such as set score, point score and number of shots in the rally. These descriptors can be directly extracted from the meta data. We use features from Method 3 and add context descriptors into our model. The performance is reported in Table 4. A visualization of the winning probability after incorporating context can be found in Figure 11. Again, it can be seen that the shape of the curves are the same but they are offset by a couple of percent based on the prior context. In Figure 12, we show more examples of our in-point prediction across various rallies and contexts.

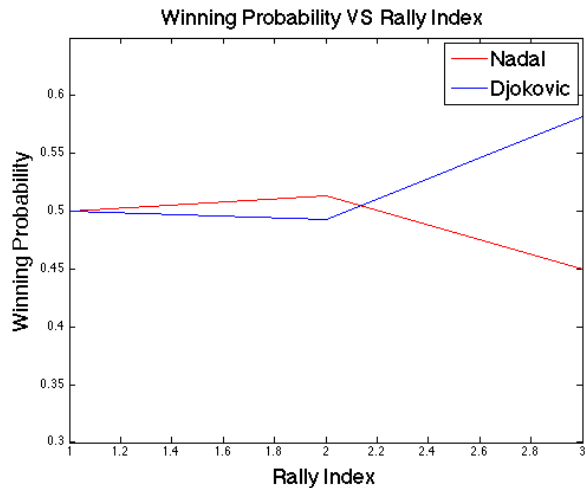
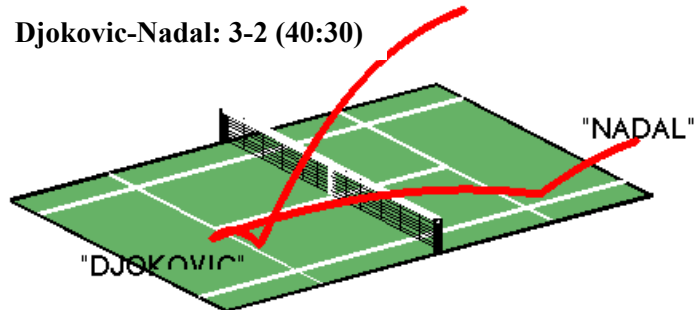
	RMSE	Classification
<b>Without Style (same as sec 3.3)</b>	0.4688	61.75%
<b>With Style (Method 1)</b>	0.4296	62.02%
<b>With Style (Method 2)</b>	0.3969	64.06%
<b>With Style (Method 3)</b>	0.3847	67.88%

Table 3: Experimental Results of various methods.

	RMSE	Classification
<b>Previous Features (Method 3)</b>	0.3847	67.88%
<b>Previous Features + Shot Index</b>	0.3796	68.65%
<b>Previous Features + Shot Index + Set Score</b>	0.3784	69.08%
<b>Previous Features + Shot Index + Set Score + Point Score</b>	0.3755	69.63%

Table 4: Experimental Results of various context descriptors.  
2016 Research Papers Competition  
Presented by:

Djokovic-Nadal: 3-2 (40:30)



Djokovic-Nadal: 3-2 (40:30)

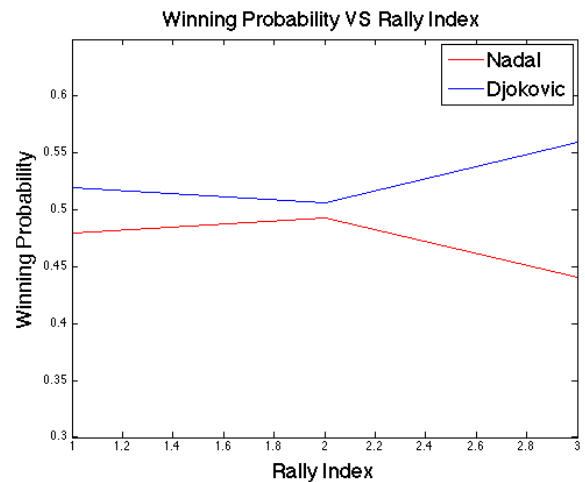
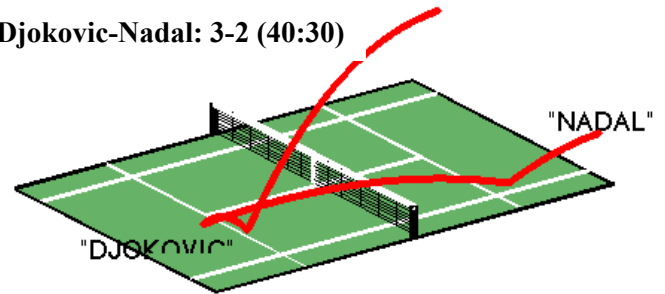
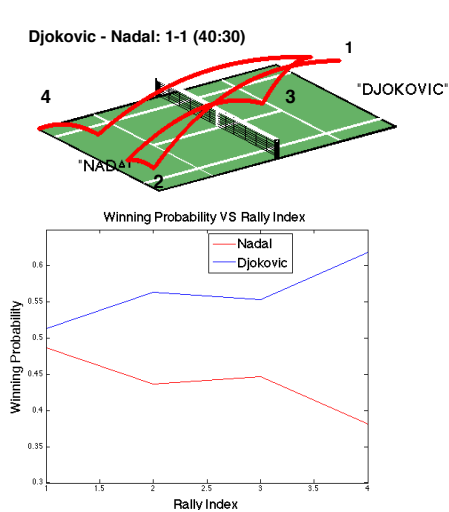
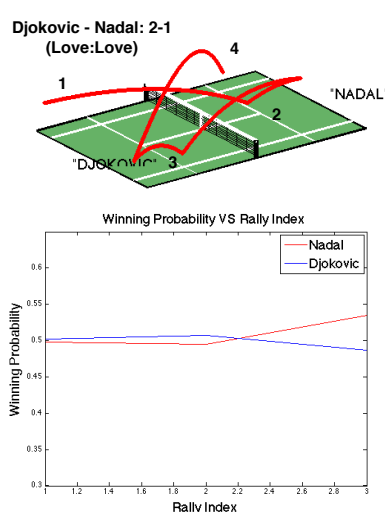


Figure 11: (Left) Winning probability without context descriptors. (Right) Winning probability after adding context descriptors.

Djokovic - Nadal: 1-1 (40:30)



Djokovic - Nadal: 2-1 (Love:Love)



Djokovic - Nadal: 2-1 (15:40)

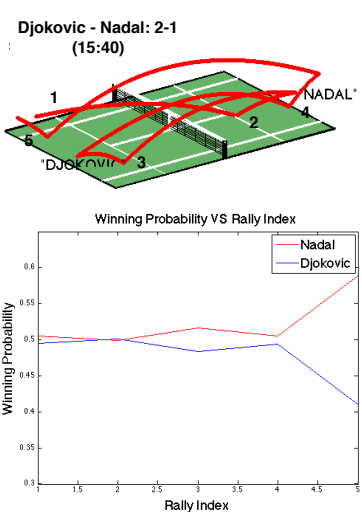


Figure 12: More examples showing our in-point prediction using style and context priors — only the rally portion of the point is shown and not the initial serve.

## 5. Summary

We proposed a method to better model adversarial behavior to predict the outcome of a point in tennis. This is a challenging problem since there is not enough data to learn a specific model between exact two agents. We proposed a “style” descriptor which allows us measure the similarity/difference between a player's opponents via dictionary learning. We then used the style descriptor with the data of the player against other opponents to help learning the prediction model. Our style descriptor include both single shots as well as shot combos. Experimental results show that our approach outperforms other methods and “style” is an important prior for learning adversarial behavior. In future, we aim to explore our approach on multi-agent adversarial domain such as soccer and basketball.

## References

- [1] Hawkey Innovations. <http://www.hawkeyeinnovations.co.uk>.
- [2] IBM SlamTracker, 2012. [www.australianopen.com/en\\_AU/ibmrealtime/index.html](http://www.australianopen.com/en_AU/ibmrealtime/index.html).
- [3] Z. Jiang, Z. Lin, and L. Davis, “Learning a discriminative dictionary for sparse coding via label consistent K-SVD”. in *CVPR*, 2011.
- [4] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation”, *IEEE Trans. on Signal Processing*, 54(1):4311 – 4322, 2006.
- [5] X. Wei, P. Lucey, S. Morgan and S. Sridharan, “Sweet-Spot: Using Spatiotemporal Data to Discover and Predict Shots in Tennis”, in *MITSSAC*, 2013.
- [6] X. Wei, P. Lucey, S. Morgan, S. Vidas, and S. Sridharan. “Forecasting Events using an Augmented Hidden Conditional Random Field”. In *ACCV*, 2014.
- [7] A. Miller, L. Bornn, R. Adams, and K. Goldsberry. “Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball”. In *ICML*, 2014.
- [8] Y. Yue, P. Lucey, P. Carr, A. Bialkowski, and S. Sridharan. “Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction”. In *ICDM*, 2014.
- [9] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews. “Identifying Team Style in Soccer using Formations Learned from Spatiotemporal Tracking Data”. In *SSTD at ICDM*, 2014.
- [10] A. Bialkowski, P. Lucey, P. Carr, Y. Yue, and I. Matthews. “Win at Home and Draw Away: Automatic Formation Analysis Highlighting the Differences in Home and Away Team Behaviors”. In *MITSSAC*, 2014.
- [11] A. Bialkowski, P. J. Lucey, P. Carr, Y. Yue, S. Sridharan, and I. Matthews. “Large-scale Analysis of Soccer Matches using Spatiotemporal Tracking Data”. In *ICDM*, 2014.
- [12] X. Wei, P. Lucey, S. Morgan, P. Carr, M. Reid, and S. Sridharan. “Predicting Serves in Tennis using Style Priors”. In *KDD*, 2015.