Forecasting Events using an Augmented Hidden Conditional Random Field

Xinyu Wei[1,2], Patrick Lucey[2], Stephen Vidas[1,3], Stuart Morgan[4] and Sridha Sridharan[1]

[1]Queensland University of Technology, [2]Disney Research Pittsburgh, [3]Nanyang Technological University, [4]Australian Institute of Sport
{felix.wei, patrick.lucey}@disneyresearch.com, svidas@ntu.edu.sg, s.sridharan@qut.edu.au, stuart.morgan@ausport.gov.au

Abstract. In highly dynamic and adversarial domains such as sports, short-term predictions are made by incorporating both local immediate as well global situational information. For forecasting complex events, higher-order models such as Hidden Conditional Random Field (HCRF) have been used to good effect as capture the long-term, high-level semantics of the signal. However, as the prediction is based solely on the hidden layer, fine-grained local information is not incorporated which reduces its predictive capability. In this paper, we propose an "augmented-Hidden Conditional Random Field" (a-HCRF) which incorporates the local observation within the HCRF which boosts it forecasting performance. Given an enormous amount of tracking data from vision-based systems, we show that our approach outperforms current state-of-the-art methods in forecasting short-term events in both soccer and tennis. Additionally, as the tracking data is long-term and continuous, we show our model can be adapted to recent data which improves performance.

1 Introduction

With the recent deployment of vision-based player and ball tracking systems in professional sports, researchers are looking at leveraging this data to forecast future events [1–3]. Application wise, this is useful as knowing the location of a

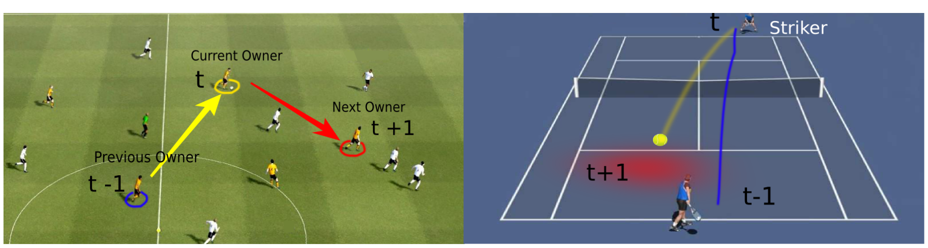


Fig. 1. In this paper, we use our a-HCRF method to: (left) predict the next pass in soccer, and (right) predict the location of the next shot in tennis.

future event would allow a camera to be intelligently positioned for automatic broadcasting [4]. An additional by-product is that higher-level analysis such as tactics and strategy can be gleaned from such systems which can aid in decision-making and story-telling aspects for coaches, broadcasters and viewers alike [5, 6]. Beyond sport, intelligent systems that can predict situations that can cause disruptions before they occur could be useful for logistics and surveillance and security domains.

An example of the problem we investigate in this paper is depicted in Figure 1, where given observations from the past n seconds, we wish to forecast/predict a future event. The event can vary from predicting the next ball owner in soccer (a), or predicting the location of the next shot in tennis (b). Even though the variance in decisions that can be made is extremely high, the number of feasible decisions can be greatly truncated by using recent contextual cues. For example in tennis, when the opponent is out of position, we can predict with high confidence that a player will hit the ball to the open side of the court in order to win the point. Our goal of this paper is to incorporate these factors into a statistical model to accurately predict these events.

Popular methods such as hidden Markov models (HMMs), dynamic Bayesian networks (DBNs), linear-chain Conditional Random Field (LCRF) obey the Markov assumption where the future state depends only upon the present state. However for complex systems where more temporal information is required, higher order models such as the HCRF (Hidden Conditional Random Field) [7] have been used to good effect. These models are effective as they decompose the input signal into a series of semantically meaningful sub-states which are hidden. However, the issue with such approaches is that the final prediction is based solely on the hidden-layer, meaning that no features directly influence the prediction.

In this paper, we propose an augmentation to HCRF which we call an *augmented-Hidden Conditional Random Field* (a-HCRF). By making the final prediction contingent on directly the hidden-layer as well as the observation, we show that we can improve prediction performance. This modification allows our model to not only capture a coarse summarization of what has happen so far through the hidden layer but also include fine-grained information of the current situation via the features. The advantages of using this configuration rather than the original HCRF or other models (e.g., DBNs, LCRF) are demonstrated by the model learning and evaluation. Additionally, as sports are adversarial and long-term, we show that our model can be adapted to match the prediction outputs. Experimental results show that our approach outperforms current state-of-the-art method in forecasting events short-term events both in soccer and tennis.

1.1 Related Work

In the computer vision literature, there has been recent work focussing both on early-event detection and event forecasting. Early event detection has the aim of detecting an event as soon as possible given that we know it has started (i.e., after it starts but before it ends) [8, 9]. Event forecasting is more complicated, as the goal is to predict what and when the event/action occurred and for how long. In terms of early-event detection, Hoai et al. [9] used a structured-output SVM to detect the length of emotions directly from faces. Ryoo [8], used a dynamic bag-of-words and maximum a posteriori (MAP) classifier to recognize human actions. In terms of event forecasting, Pellegrini et al. [10] and Mehran et al. [11] both modeled the social factors between pedestrians (i.e., avoid collisions) to improve their tracking performance. While Kitani et al., [12], utilized other factors such as nearby objects to improve the prediction of the most likely trajectory path using a partial observable Markov decision process (POMDP). In terms of predicting future crowd behavior, Zhou et al., [13] learned a mixture model of dynamic pedestrian-agents to estimate and simulate the flow of people at the Grand Central Station in New York.

In terms of adversarial behaviors, Kim et al. [1] used motion fields to predict where the play will evolve in soccer based on a region of convergence. In the subsequent work, they then used stochastic fields for predicting important future regions of interest as the scene evolves dynamically for a variety of team sports [14]. In [2, 3], a Dynamic Bayesian Network (DBN) to predict the type and location of the future shot in tennis was used. To circumvent the issue of prediction, Carr et al. [4] used an alternative approach by using virtual camera on a one-second delay and an L_1 filter to predict future behaviors.

2 Augmented Hidden Conditional Random Field

Given a period of past observations of an event, the goal of this paper is to forecast or predict what is going to happen in the short-term future (i.e., in the next 1 to 10 seconds). We do not assume past states of an event are given and only observations are available. Before we describe the method, we will first compare and contrast various models to explain the motivation of our approach.

2.1 Modeling Approaches

Linear-Chain Models: A popular way to perform prediction is to employ a HMMs or DBNs. In these models, a label of the future is dependent on its previous state as well as its observation. However, two assumptions are made here. First, it assumes each state y_i is independent of all its ancestors $y_1, y_2, ..., , y_{i-2}$ given its previous state y_{i-1} which is the Markov assumption. Secondly, to ensure the computational tractability, Bayesian models assumes features are independent. A linear conditional random field (LCRF) [15] relaxes this assumption by directly modeling the conditional distributions. However, for more complex tasks like predicting future behaviors in sport the Markovian assumption maybe limiting.

Higher-Order Models: Higher-order models, as the name suggest incorporates more than one previous state which means that the future label, y_i can

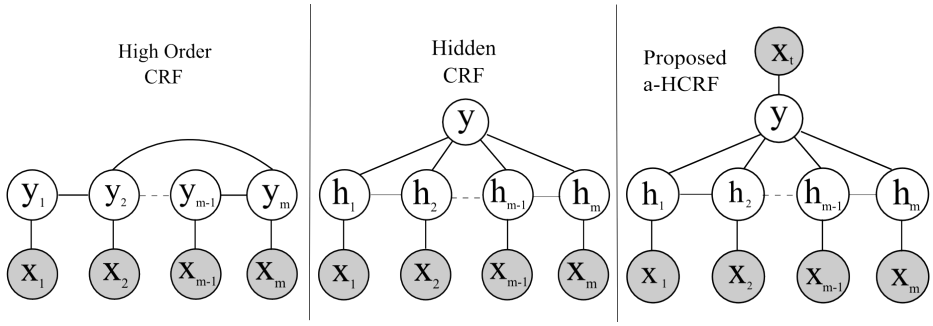


Fig. 2. (Left) Example of a higher order CRF. Output is a sequence. (Middle) Depiction of a hidden CRF (HCRF) where the sequence label y only depends on hidden states. (Right) Our proposed a-HCRF, where x_i is a past observation from $t_i - w$ to t_i, w is a feature window, h_i is a historical state at t_i, y is our prediction in the future. $\mathbf{x}$ is a global feature describes current game/player status. x_i is used for providing evidence for h_i. $\mathbf{x}$ is used for predicting the future event.

depend on any number of its ancestors $y_1, y_2, ..., y_{i-1}$. A popular example of a higher-order model is a higher-order CRF. Such models predict a *sequence* of labels - instead of making a single prediction (e.g., Figure 2(Left)). A hidden Conditional Random Field (HCRF) can circumvent this problem by making past states hidden. It only optimizes one label and it is a high order CRF. The idea is that it summarizes a temporal signal into a sequence of hidden sub-states and use these sub-states to predict the sequence label. The drawback, however, is that the prediction of y is solely based on the hidden layer. No features can directly influence y (Figure 2(Middle)).

Augmented-Hidden Conditional Random Field (a-HCRF): We modified the original HCRF by directly connecting the observation to y (Figure 2(Right)). This way, our model can not only capture fine grained information of the current situation via the top feature layer $\mathbf{x}$, but also incorporate a coarse summarization of what has happen so far (the context of the game) through the hidden layer $\mathbf{h}$. This modification is important since present information could be a strong cue for a future event. Here x_i is a feature extracted from $t_i - w$ to t_i to provide evidence for past state h_i. w is the feature window. $\mathbf{x}$ are features for predicting y. For example, $\mathbf{x}$ can be features which indicate the current game phase, player positions, player fatigue factor or any features which are predictive of the future event.

2.2 Formulation

The formulation of our a-HCRF is similar to the HCRF [7] with the key difference being the potential function. Given a set of observations $\mathcal{X} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_m\}$, we wish to learn a mapping to class labels $y \in \mathcal{Y}$. Each local observation $\mathbf{x}_j$ is represented by a feature vector $\phi(\mathbf{x}_j) \in \Re^d$. The posterior of a-HCRF is given by the following form,

$$P(y|\mathbf{x},\theta) = \sum_{\mathbf{h}} P(y,\mathbf{h}|\mathbf{x},\theta) = \frac{\sum_{\mathbf{h}} \mathrm{e}^{\Psi(y,\mathbf{h},\mathbf{x};\theta)}}{\sum_{y'\in\mathcal{Y},\mathbf{h}\in\mathcal{H}^m} \mathrm{e}^{\Psi(y',\mathbf{h},\mathbf{x};\theta)}}. \tag{1}$$

Each y is a member of a set $\mathcal{Y}$ of possible labels. For prediction, y refers to the label of a future event. The layer $\mathbf{h} = \{h_1, h_2, ..., h_m\}$, where each $h_i \in \mathcal{H}$ is a historical state of an event at time t_i. The term, θ is a set of parameters describing the feature functions. If the historical states are observed, then $\mathbf{x}$ will not influence $\mathbf{h}$. Therefore, this model can be simplified to just the top layer.

The potential function, $\Psi(y,\mathbf{h},\mathbf{x};\theta)$ measures the compatibility between a label, a set of observations and a configuration of the historical states,

$$\begin{aligned}\Psi(y,\mathbf{h},\mathbf{x};\theta) = &\sum_{j=1}^{n} \varphi(\mathbf{x},j,\omega)\cdot\theta_h[h_j] + \sum_{j=1}^{n}\theta_y[y,h_j] \\ &+ \sum_{(j,k)\in\mathcal{E}} \theta_e[y,h_j,h_k] \\ &+ \frac{\varphi(\mathbf{x},\omega)\cdot\theta_p[y]}{k},\end{aligned} \tag{2}$$

where n is the total number of historical states in the model, $\varphi(\mathbf{x},j,\omega)$ is a vector that can include any feature of the observation sequence for a specific time window ω, (i.e., each historical state can include features from $t-\omega$ to t).

The parameter vector, θ can be represented as $\theta = [\theta_{\mathbf{h}}\ \theta_y\ \theta_e\ \theta_p]$. In our work we use the same notation as [7] where $\theta_{\mathbf{h}}[h_j]$ is the parameters that correspond to state $h_j \in \mathcal{H}$. The function $\theta_y[y,h_j]$ indicates the parameters that correspond to class y and state h_j and $\theta_e[y,h_j,h_k]$ refers to parameters that between each edge h_j and h_i. Additionally, $\theta_p[y]$ defines the parameters for y given the features over the past.

The dot product $\varphi(\mathbf{x},j,\omega)\cdot\theta_h[h_j]$ measures the compatibility between the observation and the state at time j, while the dot product between $\varphi(\mathbf{x},\omega)\cdot\theta_p[y]$ measures the compatibility between the observation and the future event y. The total number of possible combinations of $\mathbf{h}$ is k and dividing by k avoids adding this term multiple times. This last term is added to capture the influence of features to a future event. Without it, a future event will only depend on past states.

2.3 Learning and Inference

Parameters can be learnt in many ways and use different objective functions. A common objective is to maximize the likelihood from labelled training data. Using the same definition in previous CRF work [16], the likelihood function is defined as follows

$$L(\theta) = \sum_{i=1}^{n} \log P(y_i|\mathbf{x_i},\theta) - \frac{1}{2\sigma^2}||\theta||^2, \tag{3}$$

where n is the total number of training examples. The first term is the log-likelihood and the second term refers to a Gaussian prior. Given a new input test sequence $\mathbf{x}$, and trained parameter θ^* we can obtain the estimated label y^* as

$$\mathrm{y}^* = \arg\max_{y \in \mathcal{Y}} P(y|\mathbf{x}, \omega, \theta^*). \tag{4}$$

In some situations, optimizing the likelihood on the training set may not generalize well to the test set. Alternatively, one can utilize a max margin criterion [17] or diverse M-best solutions [18] to learn these parameters. Other objective functions may also be used depending on the specific application (e.g., minimizing the distance between predicted location and estimated location). We used the maximum likelihood as the objective function in both of our experiments in sports.

Since the edge E in our model is a chain, exact methods for inference and parameter estimation are available. Gradient Ascent is used for each step of the tempered maximum likelihood learning. For labeling in test sequences, Maximum a Posteriori (MAP) inference is carried out using Belief Propagation.

3 Predicting Future Ball Location in Soccer

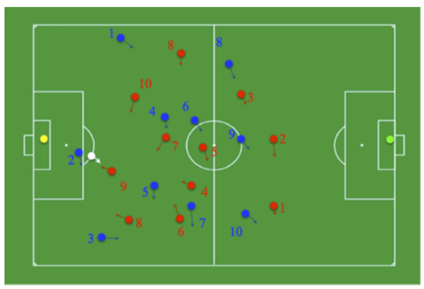
Given player tracking data over the past n seconds, our goal is to predict the owner of the ball in the future. Having the ability to predict the future ball owner has many foreseeable benefits across automatic sports broadcasting as well as improving real time ball tracking performance. This is a relatively unexplored area due to the lack of available data. Most current works are still centered on ball tracking. Recently, Wang et al. [19] formulated the ball tracking task in terms of deciding which player, if any, owns the ball at any given time. Our work extends this work, where instead of finding the ball owner at the present time, we are interested in predicting where the ball will be based in the short-term future.

3.1 Data

Spatiotemporal data has been used extensively in the visualization and officiating of sports action [20–22], but considerably fewer works [23–25] have used these large datasets to perform predictive analysis. In this experiment, we utilized the (x, y) positions of both players and the ball across 9 complete matches (over 13 hours) from a top-tier soccer league. Meta data such as the team label for each player, owner of the ball and event labels are also included. The granularity of the data is at the centimeter level, and was sampled at 10 fps. In each of these 9 matches, the team of interest was flipped to left in order to normalize team features.

3.2 Model Representation

For this experiment, h_i is a past state of the game at t_i which is hidden, y is the owner in the future, $\mathbf{x}_i$ is the observation of h_i. In each frame, we compute speed, position and moving direction for each player. The top $\mathbf{x}$ include features of current game phase (i.e. defence, attack, counter attack, corners, free kick), number of opponents currently near each player and team formation. The pairwise potential between h_i and h_{i+1} measures the transition of the game states. The unary potential between h_i and x_i measures the compatibility between a particular player and a set of features. Both potentials are automatically learned from data. A future owner is influenced by game states over the past as well as features of the current situation. Features extracted from each frame are illustrated in Figure 3 (Left).



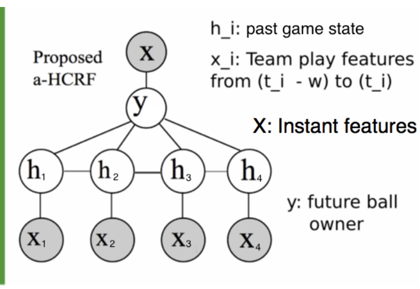


Fig. 3. (Left) In each frame, we extract speed, position and moving direction for each player. (Right) Model Representation for future ball owner prediction

3.3 Experimental Setup

Given 9 matches of soccer data, we first segment it into continuous plays and stoppages. In this research, we are only interested in predicting a future ball owner when the game is in the continuous state. For training, we have the event label which indicates the current state of the game. For testing, we employed a random forest to perform the segmentation. The idea is to break a continuous match into small chunks and assign labels to these chunks based on player features (i.e., player speed, location, etc). This task can be achieved at an average rate of 92.25% correct.

Once the segmentation is completed, the remaining frames are divided equally for training and testing. We extract data for the team of interest and its oppositions and train two models respectively. We use four nodes for the bottom chain structured CRF. Each node is 2 seconds later than the previous one. The historical state h_i can take one of 11 discrete values (i.e 11 players of this team). The future state y can take one of 12 values (representing the 11 players of the team + one for a turn over event). We only make a prediction if the same team

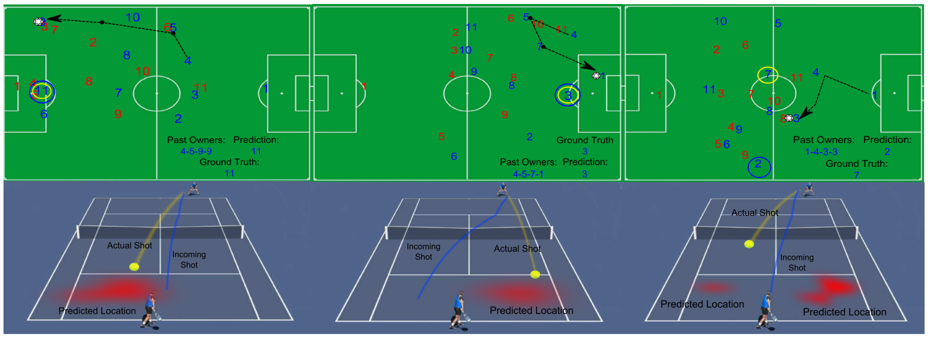


Fig. 4. Examples of our prediction results. Top: Examples of ball owner prediction in soccer. Black trajectories indicate the past passing patterns over the last four time steps, yellow circle shows the predicted ball owner while blue circle shows the ground truth. Bottom: Examples of shot prediction in tennis, yellow is the true shot trajectory while red area indicates the probability of the next shot location.

keeps the ball in the past 8 seconds. If there is a lot of turn over in the past 8 seconds, the prediction will be unreliable therefore it is not considered in this work. When testing, since our data provides the team label for each player, we can easily find out which team has the ball over the past 8 seconds and therefore apply the correct model. If the team label is not given (raw videos), one can use color features of player's jersey or optical flow combined with the ball evidence to find out which team has the ball over the last 8 seconds.

To the best of our knowledge, this is the first work on ball ownership prediction using spatiotemporal data. No existing work is available for comparison. In order to demonstrate the advantage of a-HCRF, we compare our result with other models, namely a Dynamic Bayesian Networks (DBNs), a linear chain CRF (CRF), and a Hidden CRF (HCRF). Each model has four past nodes and a future node. The last node (right most node) in DBNs or CRF is the future node and we only give past observations to that node. In HCRF, the sequence label is the future node while hidden nodes are past states. We also create two versions of our proposed model, a-HCRF-1 and a-HCRF-2. In a-HCRF-1, we set feature window ω as 1. That is, each past state can take features from the previous 2 seconds. In a-HCRF-2, feature window ω is set as 2. Thus, each state can take features from the last 4 seconds. We conduct experiments to answer three questions: i). Which model is the best? ii). How far in the future can we predict? iii). How many past features do we need?

3.4 Experimental Result and Discussion

In order to answer the above questions, we plot the prediction rate against the number of seconds in the future ranging from 1 second to 10 seconds (at 10 fps) which is shown in Figure 5. If we look in the immediate future (i.e., 1 second), the same player is more likely to have the ball which makes sense as

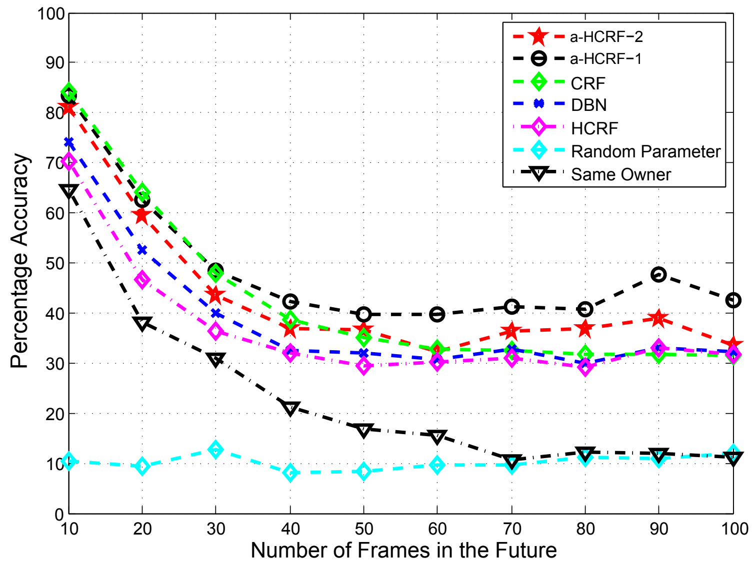


Fig. 5. Plot shows the ball owner prediction accuracy of different models at different time of the future. CRF and a-HCRF have similar performance within 2 seconds. After 2 seconds, proposed a-HCRF outperforms all other models. The black triangle curve at the bottom shows the result if we always assign the previous owner as the future owner.

the player needs time to control the ball and then execute their next decision. The black triangle curve at the bottom in Figure 5 illustrates the result if we always assign the previous owner as the future owner. Since the output can take one of twelve values, the cyan curve at bottom indicates the result of random assignment which is approximately 9%. When the future state is less than 2 seconds from the current time, the a-HCRF-1, a-HCRF-2 and CRF models have similar performance. However after 2 seconds, the a-HCRF-1 outperforms other methods. The a-HCRF-2 model is the second best method after 5 seconds. The HCRF model performs worse than DBNs, which we think is due to the model not utilizing any of the current features. Another thing to note is that at 9 seconds, there is a peak for all three CRF methods. We think this is a sweet-spot in soccer where it is more predictable.

3.5 Model Adaptation

In the previous section, our model is trained using all data from team of interest from all matches. This model assumes that a player/team will have the same behavior regardless of the opposition. This represents an area of improvement as the behavior or tactics of a player/team are heavily dependent on the opposition in adversarial activities (e.g., sports). However, to train a model between the exact two teams/players is problematic as obtaining enough data is difficult (players/teams may only play each other several times a year). A method to resolve this issue is to employ model adaptation. In this paper, we adapted a well trained generic behavior model (GBM) with a opposition-specific model (OSM) for a team/player to improve the predictive capability. Use tennis as an example, we can first use a-HCRF to train an GBM for Djokovic using data from all his matches. Then we train another model (OSM) using data just between Djokovic and a specific opponent (e.g., Nadal). Finally, we adaptively combine these two models. We expect this combined model will achieve a better prediction performance for Djokovic when he is against Nadal. The fusion can be implemented on several levels (e.g., feature, parameter, or output level.) While these combination schemes can all be explored for this task, output-level combination is of particular interest due to its simplicity (models can easily have over 1000 parameters). To do this task, we linearly combine the probability output from GBM and OSM as: $P_{comb} = \omega_1 P_{GBM} + \omega_2 P_{OSM}$ with weight ω. Here, $\omega_i \geq 0, i = 1, 2$. and $\omega_1 + \omega_2 = 1$. The optimum ω is found by maximizing the prediction rate.

Result We test the adaptation result on ball prediction in soccer in 4 seconds future. a-HCRF-1 is used to train both generic behavior model (GBM) and opposition specific model (OSM). The performance of adapted model is compared with each individual model in Table 1. The adaptive model achieves an improvement of 2.5%.

Table 1. Comparison of performance of generic behavior model (GBM), opposition specific model (OSM) and combine model (Comb).

	GBM	**OSM**	**Comb**
ω **(soccer)**	0.78	0.22	N/A
Prediction Rate (soccer)	42.6	42.4	**45.1**

4 Predicting Shot Location in Tennis

Given features of the past n shots in a rally in tennis, the goal here is to accurately predict the location of the next shot. This task is much more challenging than the previous work of Wei et.al. [2], where they predicted "what" type of shot (i.e. winner, error or continuation) but not "where" which is a potentially infinitely larger output state space. This experiment has potential value in high performance sport coaching.

4.1 Data

Using multiple fixed cameras, we used Hawk-Eye data which captured the(x, y, z) positions of the ball over time$t$ [22]. Player court positions are recorded as the

(x, y) positions of players on the court at 20 frames per second. For this work, we used the data from the 2012 Australian Open Men's draw which consisted of more than 10,000 points. We specifically modeled the behavior of Novak Djokovic at the tournament as he had the most data (winner of the tournament).

4.2 Model Representation

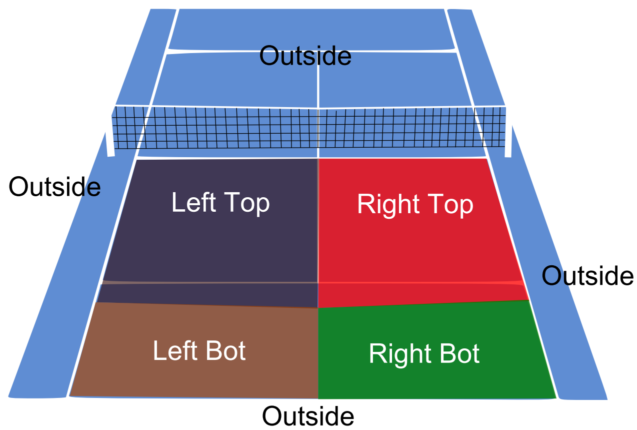


Fig. 6. An example of our court quantization scheme. Here quantization level is 2. There is 4 + 1 possible output locations for a future shot. (4 inner areas + outside)

Ideally, we want to predict the location of the shot at the most precise level (e.g., millimeter). However, as this essentially represents an infinite output state-space, we instead utilize a quantization scheme to make the problem more tractable. In order to find out the best quantization scheme, different levels of quantization are tested. The idea is we divide the receiving player's s side of the court into d areas where $d = n^2 + 1$. Here n is the quantization level, $n \in \{1, 2, 3, 4, 5, 6\}$. d is the number of areas under a particular quantization level. 1 is added to n^2 because there is a catch-all area which captured all shots that fell outside these d areas (an outside shot). For example, if we are currently using quantization level 3, then there will be $(3^2 + 1 = 10)$ possible output locations for an incoming shot (see Figure 6).

In this experiment, h_i is a past state of the game at t_i. y is the impact location of the next shot in the future. When $n = 3$, h_i can take 1 of 9 values representing the 9 inner areas of the court. (Previous shots can not be outside). y_i can take 1 of 10 values. The pairwise potential between h_i and h_{i+1} measures the transition of the game. Features used in this experiment can be found in Table 2.

Table 2. Description of the shot variables used in this paper.

Feature	Description
Speed	Shot average speed
Angle	Angle between shot & center line
Feet Location	Player and opponent court position when shot starts
Shot-Start Loc.	Location where shot starts
Shot-End Loc.	Location where shot impacts the court
No. of shots	Total number of shots in the point
Opponent Movement	Local speed & direction of the opponent before the player strikes the ball

4.3 Experimental Setup

We extract data for Novak Djokovic from an entire tournament of Australia Open 2012 Hawk-eye data. There are in total 1916 points played by him and 3410 shots. We divide this data equally for training and testing. We test our model with other models, namely DBNs, CRF and HCRF. Each model has four past nodes (the last four shots in this rally). We also create two versions of our model, a-HCRF-1 and a-HCRF-2 corresponds to different size of feature window ω. Conditional decoding is used for all models to find the optimum future label. Experiments are conducted to answer three questions: i). Which is the best model? ii). How many quantization level can we achieve while maintaining a reasonable accuracy? iii). How many history features are required?

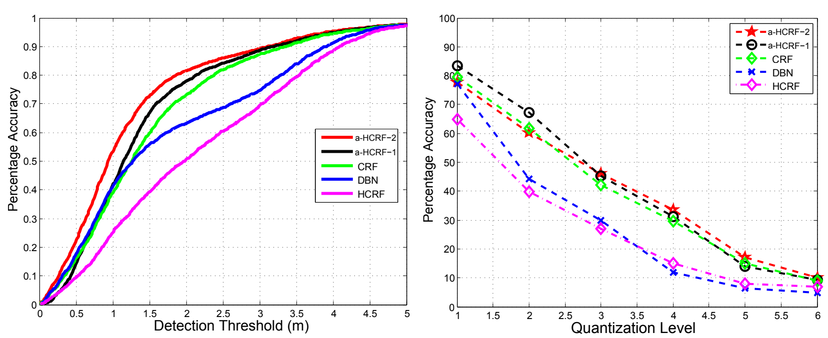


Fig. 7. (Left) Plot shows the prediction accuracy of each model against different detection threshold at quantization level 3. Proposed a-HCRF-2 (red curve) achieves the best result. (Right) Plot shows the prediction accuracy of each model at different quantization level. a-HCRF-1 achieves the best performance before level 3. a-HCRF-2 slightly outperforms other models after level 3.

4.4 Experimental Result

We first calculate the mean error (distance) between predicted location and actual location for each method at each quantization levels. We use the center of the predicted zone as the predicted location. Each method is tested 10 times and the average result is reported. Except DBN, all other models achieve the best result at quantization level 3. a-HCRF-2 gives the best result of 1.68 meter mean error[1]. Next, we plot the prediction rate against detection threshold for each model at level 3 (See Figure 7 Left). The red curve (a-HCRF-2) achieves the best result which indicates that features of two shots ago are still useful when predicting the next shot. Finally, we plot the prediction accuracy against different quantization levels (See Figure 7 Right). At level 1 (only two zones), a-HCRF-1 can predict whether a shot is inside or outside at 83% accuracy.

Table 3. Comparison of performance of generic behavior model (GBM), opposition specific model (OSM) and combine model (Comb).

	GBM	**OSM**	**Comb**
ω **(tennis)**	0.69	0.31	N/A
Prediction Rate (tennis)	48.1	39.8	**53.9**

In addition, we apply the same adaptation method on tennis. We test its performance on shot prediction in tennis at quantization level 3. a-HCRF-1 is used to train both generic behavior model (GBM) and opposition specific model (OSM). The performance of adapted model is compared with each individual model in Table 3. The adaptive model achieves an improvement of 5.8%.

5 Summary and Future Work

In this paper, we have proposed an augmented-Hidden Conditional Random Field (a-HCRF) which adds another feature layer to an HCRF to allow more effective prediction of a future event. The proposed model outperforms other models (CRF, HCRF, DBNs) across various spatiotemporal dataset for both ball ownership prediction in soccer and shot prediction in tennis. By adaptively combining a generic behavior model with an opposition-specific model of a team/player, we further improve its predictive capability. Future research will investigate other model training methods such as max-margin [17] or diverse M-best solutions. We will also explore the application of this modeling approach on other domains such as surveillance as well as trying it on datasets of larger magnitudes (e.g., seasons worth of sports data).

[1] Each side of the tennis court is 11 meters wide and 11.9 meters long

References

1. Kim, K., Grundmann, M., Shamir, A., Matthews, I., Hodgins, J., Essa, I.: Motion Fields to Predict Play Evolution in Dynamic Sport Scenes. In: CVPR. (2010)
2. Wei, X., Lucey, P., Morgan, S., Sridharan, S.: Sweet-Spot: Using Spatiotemporal Data to Discover and Predict Shots in Tennis. In: MIT Sloan Sports Analytics Conference. (2013)
3. Wei, X., Lucey, P., Morgan, S., Sridharan, S.: Predicting Shot Locations in Tennis using Spatiotemporal Data. In: DICTA. (2013)
4. Carr, P., Mistry, M., Matthews, I.: Hybrid Robotic/Virtual Pan-Tilt-Zoom Cameras for Autonomous Event Recording. In: ACM Multimedia. (2013)
5. Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Matthews, I.: "Win at Home and Draw Away": Automatic Formation Analysis Highlighting the Differences in Home and Away Team Behaviors. In: MIT Sloan Sports Analytics Conference. (2014)
6. Lucey, P., Bialkowski, A., Carr, P., Yue, Y., Matthews, I.: "How to Get an Open Shot": Analyzing Team Movement in Basketball using Tracking Data. In: MIT Sloan Sports Analytics Conference. (2014)
7. Quattoni, A., Wang, L., L., M., Collins, M., Darrell, T.: Hidden-State Conditional Random Fields. In: PAMI. (2007)
8. Ryoo, M.S.: Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos. In: ICCV. (2011)
9. Hoai, M., Torre, F.: Max-Margin Early Event Detectors. In: CVPR. (2012)
10. Pellegrini, S., Ess, A., Schindler, K., Van Gool, L.: You'll Never Walk Alone: Modeling Social Behavior for Multi-Target Tracking. In: CVPR. (2009)
11. Mehran, R., Oyama, A., Shah, M.: Abnormal Crowd Behavior Detection using a Social Force Model. In: CVPR. (2009)
12. Kitani, K., Ziebart, B.D., Bagnell, J.A., Hebert, M.: Activity Forecasting. In: ECCV. (2012)
13. Zhou, B., Wang, X., Tang, X.: Understanding Collective Crowd Behaviors: Learning a Mixture Model of Dynamic Pedestrian-Agents. In: CVPR. (2012)
14. Kim, K., Lee, D., Essa, I.: Detecting Regions of Interest in Dynamic Scenes with Camera Motions. In: CVPR. (2012)
15. Lafferty, J., McCallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models For Segmenting and Labeling Sequence Data. In: ICML. (2001)
16. Wang, S., Quattoni, A., L., M., Demirdjian, D.: Hidden Conditional Random Fields for Gesture Recognition. In: CVPR. (2006)
17. Wang, Y., Mori, G.: Max-Margin Hidden Conditional Random Fields for Human Action Recognition. In: CVPR. (2009)
18. Batra, D., Yadollahpour, P., Guzman-Rivera, A., Shakhnarovich, G.: Diverse M-Best Solutions in Markov Random Fields. In: ECCV. (2012)
19. Wang, X., Ablavsky, V., Shitrit, H.B., Fua, P.: Take your Eyes off the Ball: Improving Ball-Tracking by Focusing on Team Play. In: CVIU. (2013)
20. SportsVision. (http://www.sportsvision.com.au)
21. Stats. (http://www.stats.com/)
22. Hawk-Eye. (www.hawkeyeinnovations.co.uk)
23. Lucey, P., Bialkowski, A., Carr, P., Foote, E., Matthews, I.: Characterizing Multi-Agent Team Behavior from Partial Team Tracings: Evidence from the English Premier League. In: AAAI. (2012)
24. Lucey, P., Oliver, D., Carr, P., Roth, J., Matthews, I.: Assessing Team Strategy using Spatiotemporal Data. In: ACM SIGKDD Conference on Knowledge, Discovery and Data Mining (KDD). (2013)

25. Wei, X., Sha, L., Lucey, P., Morgan, S., Sridharan, S.: Large-Scale Analysis of Formations in Soccer. In: DICTA. (2013)