

Where Will Players Move Next? Dynamic Graphs and Hierarchical Fusion for Movement Forecasting in Badminton

Kai-Shiang Chang, Wei-Yao Wang, Wen-Chih Peng

National Yang Ming Chiao Tung University, Hsinchu, Taiwan
{kevin5260523.cs05, sf1638.cs05}@nctu.edu.tw, wcpeng@cs.nycu.edu.tw

Abstract

Sports analytics has captured increasing attention since analysis of the various data enables insights for training strategies, player evaluation, etc. In this paper, we focus on predicting what types of returning strokes will be made, and where players will move to based on previous strokes. As this problem has not been addressed to date, movement forecasting can be tackled through sequence-based and graph-based models by formulating as a sequence prediction task. However, existing sequence-based models neglect the effects of interactions between players, and graph-based models still suffer from multifaceted perspectives on the next movement. Moreover, there is no existing work on representing strategic relations among players' shot types and movements. To address these challenges, we first introduce the procedure of the **Player Movements (PM) graph** to exploit the structural movements of players with strategic relations. Based on the PM graph, we propose a novel **Dynamic Graphs and Hierarchical Fusion for Movement Forecasting model (DyMF)** with interaction style extractors to capture the mutual interactions of players themselves and between both players within a rally, and dynamic players' tactics across time. In addition, hierarchical fusion modules are designed to incorporate the style influence of both players and rally interactions. Extensive experiments show that our model empirically outperforms both sequence- and graph-based methods and demonstrate the practical usage of movement forecasting. Code is available at <https://github.com/wywyWang/CoachAI-Projects/tree/main/Movement%20Forecasting>.

1 Introduction

In recent years, the rapid advancement of technology has brought great convenience for data collection and has increased data diversity, which has revolutionized sports analytics for conducting various applications. For example, the National Basketball Association (NBA) has installed multiple cameras on every court to collect granular data on players' movements. These data are used to analyze the different views for recommending winning strategies, avoiding player injury, and scouting (Petra 2020). On the other hand, the size of the sports analytics market is expected to expand 21.3% from 2021 to 2028 (Research 2021). Evidently, effective analysis benefits the promising performance of teams or

individuals on the field (Wang et al. 2020; Melton and Mac-Charles 2021; Qian 2022), which has not only brought about the growing demand of sports analytics, but has also drawn researchers' attention to the flourishing and profitable sports market.

In this paper, we focus on turn-based sports and use badminton as the demonstration example. Various studies related to badminton have been carried out to collect badminton data, such as retrieving information from videos (Hsu et al. 2019; Cao et al. 2021; Yoshikawa et al. 2021) or sensors (Chiu et al. 2020; Steels et al. 2020). Based on the collected data, several applications have been developed to analyze tactic performance (Wang 2022), e.g., shot influence (Wang et al. 2021, 2022a), an analysis system (Chu et al. 2022), and stroke forecasting (Wang et al. 2022b). However, in badminton, as a fast-paced sport, in addition to returning strokes, players' movements on the court are also a vital factor of simulating tactics, which has not been addressed in prior works. Players with effective movements can reach the shuttle faster before it drops below the net, which enables them to return an aggressive stroke and take the dominant position (Decathlon 2022). Therefore, players and coaches can take advantage of movement forecasting since the movements on the court reflect the physical ability and the defensive position habit. Players can therefore be trained to be familiar with where to catch the shuttle and where to defend based on the specific situation¹.

Figure 1 illustrates an example of movement forecasting. Given the locations of two players and the shot type they performed in a rally, the player on the right has several options of shot type and location to return the shot and move back to defense, while the player on the left also has multiple choices of location to catch the shuttle based on the location the opponent is returning to. For example, when the left player returns a net shot to the right court, the right player can either also return a net shot and take a defense position closer to the net to prepare for another net shot, or can return a lob to the back court and move to the center of the half court to prepare for a smash. In general, movement forecasting can be viewed as the more challenging task since the goal consists of not only stroke forecasting (Wang et al. 2022b), but also the movements of players.

¹Related work is in Appendix A due to space limitation.

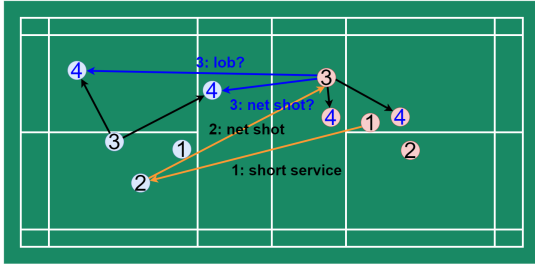


Figure 1: An example of movement forecasting. The blue nodes and red nodes represent the locations of two players at each time step, respectively. Nodes with blue numbers and blue lines are the possible choices of the moving locations of the two players and the shot type at the next time step.

To address this problem, we formulate movement forecasting as a sequence prediction task. One potential solution is the sequence-to-sequence models, which have been widely applied for trajectory prediction (Giuliani et al. 2020; Shi et al. 2020; Shafiee, Padir, and Elhamifar 2021). However, in our task, sequence-to-sequence models are ineffective in terms of modeling the interactions between different players and the relations between the same player at different time steps, which has a great impact on the distance traveled and the destination. For example, a net shot causes the opponent to move toward the net; consequently, the player has less time to go back to defense due to the short flying time. On the other hand, a clear move moves the opponent to the back court, which enables them to have sufficient time to move back to defense. In addition, the meanings of moving between returning the shuttle and defending are different: when returning the shuttle, the player tends to move to the corner, and when defending, the tendency is to move to the center (Valdecabres et al. 2020).

To mitigate the aforementioned limitations, a badminton rally can be represented as a graph with various relations to describe different purposes of the movements across different time steps. Nonetheless, there are three key challenges that hinder the direct use of existing graph methods in movement forecasting. 1) **Graph construction.** Since previous approaches have not adopted graphs for turn-based sports analysis, there is no existing method for constructing badminton rallies into graphs. Therefore, it is challenging to construct a graph to represent rally information, including the relations among players and their strokes. 2) **Dynamic tactics.** Players often change their tactics within a match or even a rally to avoid their strategies being seen through. In view of the ever-changing circumstances on the court, it is inadequate to fixedly model the styles and tactics of both players in different rallies and within the rally. 3) **Multi-faceted aspects.** To decide the next action, players consider not only their tactical preferences but also those of their opponents. Furthermore, players are confronted with different significance between styles and rally progress according to the current circumstance.

In light of the above challenges, we propose a novel **Dynamic graphs** and hierarchical fusion for the **Movement**

Forecasting model (DyMF) to capture player information through different relations for predicting the next move of both players and the corresponding shot type. For the first issue, we introduce a badminton graph called the **Player Movements (PM)** graph to construct strategic relations between two players and the players themselves in a rally across time. Specifically, DyMF is an encoder-decoder architecture with interaction style extractors and hierarchical fusions. For the second issue, the interaction style extractor is introduced to dynamically capture mutual interactions of players themselves and between players by aggregating information through different relations and pattern-generated weights. For the third issue, we design a hierarchical fusion to take the style influence of both players into account and combine player-player interactions with rally interactions for predicting the next move.

To summarize, the main contributions of this paper are four-fold:

1. We attempt to formalize the task of predicting the players' future movements and shot type based on the previous strokes. To the best of our knowledge, this is the first work for movement forecasting in badminton, which is also suitable for other turn-based sports such as tennis.
2. To represent a rally as a graph, we introduce a graph construction method by transforming rallies into PM graphs with strategic relations between players' locations.
3. Based on the PM graphs, we propose DyMF to extract different facets of interactions by interaction style extractors, and integrate them by hierarchical fusions.
4. We conduct extensive experiments to demonstrate the effectiveness of our proposed model by comparing with state-of-the-art sequence- and graph-based methods on a real-world badminton dataset. Furthermore, an application scenario is illustrated for movement forecasting.

2 Problem Formulation

A badminton rally is denoted as $R = \{[S^1, \dots, S^{|R|}], P\}$, where S^t represents the t -th stroke, $P = (p_a, p_b)$ denotes the two players who participate in the rally and take turns hitting the shuttle to form a rally (i.e., p_a is the player who serves and p_b is the other player in a rally), and $|R|$ is the length of the rally. At t -th time step (stroke), each stroke consists of two components $S^t = \{L^t, s^t\}$, where $L^t = \{L_a^t, L_b^t\}$ is the player locations of p_a and p_b when either player hits the shuttle, and s^t represents the shot type played by either player. Specifically, $L_a^t = (l_{ax}^t, l_{ay}^t) \in \mathbb{R}^2$ and $L_b^t = (l_{bx}^t, l_{by}^t) \in \mathbb{R}^2$ denote the 2D coordinates of the locations of two players on the court, respectively. The movement forecasting task is defined as follows: For each badminton rally, given the observed stroke sequence $[\{L^1, s^0\}, \dots, \{L^\tau, s^{\tau-1}\}]$ with length τ and pairs of players P , the objective is to predict future strokes in the rally $[\{L^{\tau+1}, s^\tau\}, \dots, \{L^{|R|}, s^{|R|-1}\}]$, including the locations of the two players and the shot types performed by either player. We note that s is offset by one index as the target is to predict the locations of the two players after each stroke; thus the first shot type s^0 is padded with a zero.

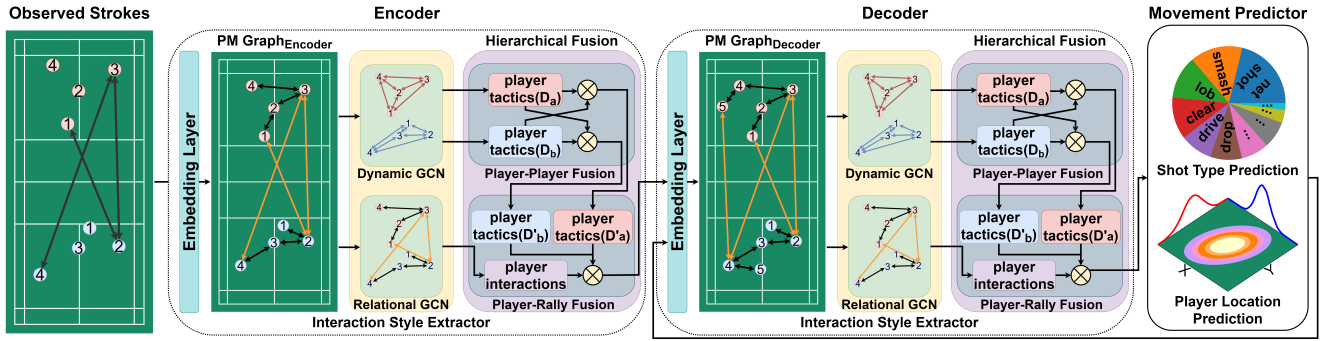


Figure 2: The architecture of our proposed DyMF. Interaction style extractors retrieve the context of mutual interactions on the PM graph and the dynamic tactics of a single player with previous patterns. Hierarchical fusions fuse the style influence of both players and rally interactions, and the movement predictor utilizes fused information to predict the locations of both players and the shot types they performed.

3 Our Approach

Figure 2 shows the overall architecture of our proposed DyMF. To represent the badminton rally as a graph, the PM graph is designed with the locations of players at each time step and the relations between them across time. Based on the PM graph, DyMF adopts an encoder-decoder architecture, where the encoder encodes the observed stroke sequence, while the decoder recurrently generates the locations of players and the shot types they performed based on the encoder contexts. Specifically, both the encoder and decoder are composed of two modules, respectively, an interaction style extractor, and a hierarchical fusion. The interaction style extractor takes players and corresponding locations as the information to capture mutual interactions of the players themselves and between both players by modeling the relations across time and the dynamic tactics of an individual player with past movement patterns. To adapt to tactics and styles changing as the rally progresses, a hierarchical fusion is proposed to consider the style influence of both players and rally interactions based on the current circumstance.

3.1 Graph Construction

We first introduce the components in the PM graph and then illustrate in detail the graph construction procedure of the PM graph in the encoder and decoder.

PM Graph. As a rally is comprised of players hitting the shuttle back and forth, the interactions between players within the rally can be described as structural representations in a graph. However, directly applying existing graph construction methods (e.g., a complete graph) to rally sequences fails to model strategic relations between players. To this end, a graph construction procedure is introduced by representing the rally progress as the evolution of a series of undirected multi-graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. The nodes in the multi-graphs are denoted as $(v_a^t, v_b^t) \in \mathcal{V}$, where v_a^t and v_b^t represent the locations of p_a and p_b at the t -th time step, respectively. The edges $(v_{a|b}^i, r, v_{a|b}^j) \in \mathcal{E}$ represent the relation between the nodes of different players or the nodes

of the same player between time steps i to j , $a|b$ represents that the term is related to either p_a or p_b , and the relations $r \in \mathcal{R}$ consist of 12 strategic types listed as follows: **10 shot types:** net shot, lob, defensive shot, smash, drop, push/rush, short service, clear, drive, and long service. **Defend:** Link the nodes of the same player at the time steps $t - 1$ and t when the purpose of the movement is to return to the defense position. **Return:** Link the nodes of the same player at the time steps $t - 1$ and t when the purpose of the movement is to return the shuttle.

Encoder Graph. Figure 3 (a) illustrates an example of the graph construction procedure with the observed sequence with $\tau = 4$ strokes. At the first time step, there are two nodes that represent the initial locations of the two players (p_a, p_b) before the rally starts. Afterwards, at each time step, two nodes and three edges are added to the graph to represent the movements and interactions of players on the court. Specifically, the new nodes are the locations of the two players after either of them returns the shuttle. The three new edges contain one shot type that links the node of the player who returns the shuttle with the time step $t - 1$ and the node of the player who will return the shuttle at the next time step with the time step t (red 3 to blue 4), one defend type that links the nodes of the player who returns the shuttle with the time steps $t - 1$ and t (red 3 to red 4), and one return type that links the nodes of the player who will return the shuttle at the next time step with the time steps $t - 1$ and t (blue 3 to blue 4). The encoder graph is passed to the proposed encoder to learn the embeddings of the nodes.

Decoder Graph. In the decoder stage, the PM graph is divided into five steps at each time step based on the encoder graph, which is depicted as an example at time step $t = 5$ in Figure 3 (b): 1) Add one new node and one new edge, where the new node represents the location of the player who will return the shuttle at the next time step and the new edge is a return type. 2) Compute the node embeddings through the model decoder and predict the shot type of the stroke at the current time step by using the node embedding of the player who returns the shuttle with time step $t - 1$ and the node embedding of the player who will return the shuttle at the next

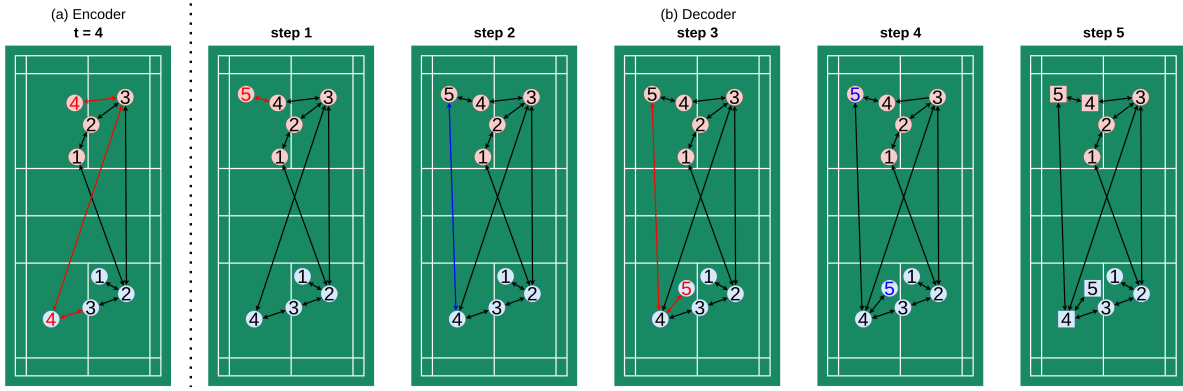


Figure 3: Illustrations of encoder and decoder PM graph constructions. The red nodes and blue nodes represent the locations of two players at the corresponding time steps, respectively. The newly added nodes and edges are decorated in red. The nodes with blue numbers and blue edges are the object to predict, and the square nodes represent the updated nodes at that time step.

time step with time step t . 3) Then, a new node and two new edges are added in the decoder graph, where the new node represents the location of the player who returns the shuttle and moves back to the defense position, and the two new edges are a defense type (blue 4 to blue 5) and a shot type (blue 4 to red 5) based on the prediction from the previous step. 4) Compute the node embeddings through the model decoder and predict the locations of the two players based on the node embeddings of the last nodes of the two players. 5) Update the node embeddings of the last four nodes that participate in the graph construction at this time step by the model. The decoder graph construction is then constructed recurrently until the end of the rally.

3.2 Embedding Layer

To represent the nodes with the locations of the players and corresponding players' information, the embedding layer takes $\{L_a^t, L_b^t\}$ and (p_a, p_b) as input to calculate the initial representation of the node e_a^t and e_b^t of v_a^t and v_b^t at the t -th time step, respectively:

$$\hat{e}_{a|b}^t = W^e (\text{ReLU}(W^L L_{a|b}^t) \parallel W^p p_{a|b}), \quad (1)$$

where $W^L \in \mathbb{R}^{d_l \times 2}$, $W^p \in \mathbb{R}^{d_p \times N_p}$, $W^e \in \mathbb{R}^{d_e \times (d_l + d_p)}$, are learnable matrices, and \parallel is the concatenation operator. d_l , d_p and d_e denote the dimension of location, player, and node representations, respectively, and N_p is the total number of players in the dataset.

It is noted that the initial representation in the decoder is employed with the locations of the two players at the last time step, since their locations will be predicted at the end of the time step. We also share the parameters W^L and W^p in the embedding layers on the encoder and decoder to reduce the number of parameters similar to (Wang et al. 2022b).

3.3 Interaction Style Extractor

The interactions and style confrontations between the players on the court are necessary for making a decision on the next move for both players. For example, a defensive player tends to return a clear or a lob, which gives more

time to move to a better position for defense and wait for the unforced error by the opponent. Therefore, the interaction style extractor is designed with relational GCN and dynamic GCN to capture the interactions between players and the individual style of both players, respectively.

Relational GCN. Both the shot types by either player and the mutual interactions (i.e., purposes of the movements) play a vital role in organizing tactics to influence the next locations of both players. For instance, a lob moves the player to the back court, while a net shot moves the player to the net. Moreover, the player tends to be in the center of the court when the purpose of the movement is to defend. When the purpose of the movement is to return the shuttle, the tendency is to move to the corner since the opponent would try to mobilize the player. Therefore, relational GCN is introduced to capture the interactions between different players and the relations between the same player at different time steps.

Inspired by R-GCN (Schlichtkrull et al. 2018), the architecture of relational GCN is a propagation model that learns the node embedding by aggregating the information from neighbors based on the relations between them. To preserve the information to the next layer, the message-passing process is added with a self-connection as:

$$z^{i(l+1)} = \sigma(\sum_{r \in \mathcal{R}} \sum_{j \in N_r^i} a_r^i W^{r(l)} z^{j(l)} + W^{o(l)} z^{i(l)}), \quad (2)$$

where $z^{i(l)}$ denotes the hidden state of the i -th node in \hat{v} in layer l , where $\hat{v} = [v_a^1, v_b^1, \dots, v_a^t, v_b^t]$, N_r^i is a set of neighbors of i -th node in \hat{v} for a corresponding relation r , σ is the activation function, $W^{r(l)}, W^{o(l)} \in \mathbb{R}^{d_e \times d_e}$ are the learnable matrices in layer l , and a_r^i is the normalized constant. The hidden states of the nodes at the first layer in the encoder are the initial node embeddings e_a^i and e_b^i . Besides, the propagated node embeddings \hat{e}_a^t and \hat{e}_b^t at previous time steps are used for the hidden states at the first layer in the decoder, except that the two newly added nodes use the initial node representation.

To stabilize the training of each relation, we adopt the regularization method of basis decomposition (Schlichtkrull

et al. 2018) for $W^{r(l)}$:

$$W^{r(l)} = \sum_{b=1}^B c^{rb(l)} M^{b(l)}, \quad (3)$$

where B is the number of basis matrices and $c^{rb(l)}$ denotes the learnable parameters of relation r for the basis matrix $M^{b(l)} \in \mathbb{R}^{d_e \times d_e}$ in layer l .

Dynamic GCN. Wang et al. (2022b) demonstrated the significance of considering player styles for predicting shot types and destination locations, but ignored the dynamic personal style of each player, which is frequently changed at each stroke for adjusting their tactics. To address this issue, we propose dynamic GCN by generating the weights based on the players' patterns to reflect dynamic personal style.

Specifically, the player information is first included in the outputs of the embedding layer:

$$n_{a|b}^t = W^n (e_{a|b}^t \parallel p_{a|b}), \quad (4)$$

where $W^n \in \mathbb{R}^{d_e \times (d_e + d_p)}$ are learnable matrices. In the decoder, $e_{a|b}^t$ is replaced by the propagated node representation $\hat{e}_{a|b}^t$. Then, we apply a 1D convolutional neural network (Conv1D) with kernel size K to the $E_{a|b}^t = [n_{a|b}^1, \dots, n_{a|b}^t]$ to extract the local movement patterns of each player:

$$C_{a|b}^t = \text{Conv1D}(E_{a|b}^t). \quad (5)$$

We note that $E_{a|b}^t$ is padded with zeros until the dimension is equal to d_e and the same length of input as the corresponding output.

Afterwards, LSTM (Sak, Senior, and Beaufays 2014) is applied to the pattern sequences $C_{a|b}^t$ to capture the long-term patterns:

$$h_{a|b}^{(s+1)} = \text{LSTM}(h_{a|b}^{(s)}, c_{a|b}^{(s)}, C_{a|b}^{t(s)}; W^{\text{LSTM}}), \quad (6)$$

where $h_{a|b}^{(s)}, c_{a|b}^{(s)} \in \mathbb{R}^{d_e}$ are the s -th hidden state and cell state. $C_{a|b}^{t(s)} \in \mathbb{R}^{d_e}$ are the s -th pattern of C_a^t and C_b^t , and W^{LSTM} is the learnable parameters. The initialization of the hidden state and cell state is filled with zeros following (Wang et al. 2021).

Finally, the pattern-generated weights $Q_{a|b} = [h_{a|b}^{(1)}, \dots, h_{a|b}^{(t)}]$ are regarded as the weights of dynamic GCN to calculate the context of player interactions for each node by replacing the weights in the graph convolutional network (GCN) (Kipf and Welling 2017) with $Q_{a|b}$:

$$d_{a|b}^i = \text{GCN}(n_{a|b}^i; Q_{a|b}), \quad (7)$$

where $d_{a|b}^i$ denotes the hidden states of nodes v_a^i and v_b^i .

3.4 Hierarchical Fusion

Although players move to the next location for returning the next stroke based on rally interactions and players' styles, it is expected that the importance of each varies under different circumstances. That is, promising players change their tactics according to their opponents in order to seize the winning opportunity. Hence, we propose a hierarchical fusion to decide the mutual importance of each style by modeling the style influence between the opponent and the current returning player, and to merge the contexts of rally interactions and player styles.

Player-Player Fusion. Parallel co-attention (Lu et al. 2016) is adopted in the sequence of the player's dynamic tactics from the output of dynamic GCN $D_a^t = [d_a^1, \dots, d_a^t]$ and $D_b^t = [d_b^1, \dots, d_b^t]$ to consider the mutual influence of the style \hat{a} and \hat{b} of each player:

$$\begin{aligned} G &= \text{tanh}((D_a^t)^T W^D D_b^t), \\ H_a &= \text{tanh}(W^a D_a^t + (W^b D_b^t) G^T), \\ H_b &= \text{tanh}(W^b D_b^t + (W^a D_a^t) G), \\ \text{att}_{a|b} &= \text{softmax}((w^{ha|hb})^T H_{a|b}), \\ \hat{a}|\hat{b} &= \sum_{i=1}^t \text{att}_{a|b}^i d_{a|b}^i, \end{aligned} \quad (8)$$

where $W^D, W^a, W^b \in \mathbb{R}^{d_e \times d_e}$ are learnable matrices and $w^{ha}, w^{hb} \in \mathbb{R}^{d_e}$ are learnable parameters. The outputs \hat{a} and \hat{b} are then transformed to the overall influence f_a and f_b :

$$f_a = \text{sigmoid}(w^{\hat{a}} \hat{a}), f_b = \text{sigmoid}(w^{\hat{b}} \hat{b}), \quad (9)$$

where $w^{\hat{a}}, w^{\hat{b}} \in \mathbb{R}^{d_e}$ are learnable parameters.

Afterwards, we derive the context of each player's dynamic player's style with the opponent's influence:

$$D_{a|b}^t = f_{b|a} \times D_{b|a}^t + D_a^t. \quad (10)$$

Player-Rally Fusion. The player-rally fusion aims at capturing the importance of player-player styles and rally interactions to predict each player's next move. To fuse these contexts, we first use the t -th player-player styles $d_{a|b}^t$ from $D_{a|b}^t$ and rally interactions in the last layer in the relational GCN $z_{a|b}^{t(l_{\text{rast}})}$ of both players to determine the corresponding dynamic weights $\alpha_a, \alpha_b, \beta_a, \beta_b$:

$$\alpha_{a|b} = \text{sigmoid}(w^S d_{a|b}^t), \beta_{a|b} = \text{sigmoid}(w^Z z_{a|b}^{t(l_{\text{rast}})}), \quad (11)$$

where $w^S, w^Z \in \mathbb{R}^{d_e}$ are learnable parameters.

Afterwards, the propagated node embeddings \hat{e}_a^t and \hat{e}_b^t are updated by:

$$\hat{e}_{a|b}^t = \alpha_{a|b} \times d_{a|b}^t + \beta_{a|b} \times z_{a|b}^{t(l_{\text{rast}})}. \quad (12)$$

3.5 Movement Predictor

To predict the shot type at the t -th time step (step 2 in the PM graph), the propagated node embeddings of both players at the t -th time step are calculated as:

$$\hat{s}^t = W^s (e_a^{t-1} \parallel \hat{e}_b^t), \quad (13)$$

where $W^s \in \mathbb{R}^{N_t \times 2d_e}$ are learnable matrices, and N_t is the number of shot types. We note that the former term becomes \hat{e}_b^{t-1} and the latter becomes \hat{e}_a^t when the number of the time step is odd, which switches the returning player and the defending player.

To predict the locations of the two players at the t -th time step (step 4 in the PM graph), we follow (Wang et al. 2022b) to learn a bivariate Gaussian distribution for the location of each player. Specifically, a bivariate Gaussian distribution consists of five parameters $\{\mu_x, \mu_y, \sigma_x, \sigma_y, \rho\}$, which represent the mean of the 2D coordinates, the variance of the 2D

coordinates, and the corresponding correlation coefficient. We apply a linear layer to the concatenation of the propagated node embeddings of two players to predict two bivariate Gaussian distributions $\hat{B}G_a^t$ and $\hat{B}G_b^t$ for both players:

$$\{\hat{B}G_a^t, \hat{B}G_b^t\} = W^{BG}(\hat{e}_a^t \parallel \hat{e}_b^t), \quad (14)$$

where $W^{BG} \in \mathbb{R}^{10 \times 2d_e}$. The dimension 10 is the total number of two five-parametric distributions. The predicted coordinates of the $t + 1$ time step for each player L_a^{t+1} and L_b^{t+1} are then sampled from $\mathcal{N}(\hat{B}G_a^t)$ and $\mathcal{N}(\hat{B}G_b^t)$, respectively.

The objective function is to minimize the total loss \mathcal{L} :

$$\mathcal{L} = \mathcal{L}_{shot.type} + (0.5 \times \mathcal{L}_{location.a} + 0.5 \times \mathcal{L}_{location.b}), \quad (15)$$

where $\mathcal{L}_{shot.type}$ is the cross-entropy loss:

$$\mathcal{L}_{type} = -\sum_{R \in \mathcal{D}} \sum_{t=\tau}^{|R|-1} s^t \log(\text{softmax}(\hat{s}^t)), \quad (16)$$

and $\mathcal{L}_{location.a}$ and $\mathcal{L}_{location.b}$ are the negative log-likelihood losses for each player:

$$\mathcal{L}_{location.a} = -\sum_{R \in \mathcal{D}} \sum_{t=\tau+1}^{|R|} \log(\mathcal{P}(l_{ax}^t, l_{ay}^t | \hat{B}G_a^t)), \quad (17)$$

$$\mathcal{L}_{location.b} = -\sum_{R \in \mathcal{D}} \sum_{t=\tau+1}^{|R|} \log(\mathcal{P}(l_{bx}^t, l_{by}^t | \hat{B}G_b^t)),$$

where \mathcal{D} is the set of rallies.

4 Experiments

4.1 Experimental Settings

Badminton Dataset. Since there is only one dataset of turn-based stroke records, we employ the badminton singles dataset (Wang et al. 2022b). After filtering the rallies with any missing value, the badminton dataset has 75 matches played by 31 high-ranking players remaining, which contain 180 sets, 4,325 rallies, and 43,191 strokes. Following (Wang et al. 2022b), we use the same 10 shot types, and the maximum rally lengths are truncated to 35.

For the task of movement forecasting, we utilize the rally id to separate rallies. Each rally contains the ball round of each stroke, which consists of the players, their locations, and the shot type either of them performed. The first 80% of rallies are split as training data and the last 20% as testing data for each match to have the model acquire each player’s previous information.

Baseline Methods. As this task has yet to be addressed, there are no existing baselines for direct comparison. Therefore, we examine our model with several state-of-the-art sequence prediction methods to verify its effectiveness. Specifically, these baselines are categorized into two groups: **Sequence-based models:** 1) Seq2Seq (Sutskever, Vinyals, and Le 2014), 2) TF (Giuliani et al. 2020), and 3) ShuttleNet (Wang et al. 2022b). **Graph-based models:** 1) dNRI (Graber and Schwing 2020), 2) GCN_{PM} (Kipf and Welling 2017), 3) R-GCN_{PM} (Schlichtkrull et al. 2018), and 4) E-GCN_{PM} (Pareja et al. 2020). It is worth noting that we introduce three new variants of the graph-based approaches (GCN_{PM}, R-GCN_{PM}, and E-GCN_{PM}) with encoder-decoder structures and the proposed PM graph for better exploring the performance of our proposed method. More details and setups of baselines are given in Appendix B.

Parameter Settings. The representation dimension (d_l, d_p, d_e) is set to 16 for all models. The dropout rate for each module is 0.1, and the number of heads in TF and ShuttleNet is set to 2. E-GCN_{PM}, R-GCN_{PM}, GCN_{PM} and our model have two GNN layers. The number of layers of TF is 1. The activation function σ in R-GCN_{PM}, GCN_{PM}, E-GCN_{PM} and our model applies ReLU at the first layer and Sigmoid at the second layer. The normalized constants (a_r^t) for R-GCN_{PM} and our model are set to 1 similar to (Schlichtkrull et al. 2018) and the number of basis matrices (B) is 3. The kernel size K is set to 3 and the layer of LSTM is set to 1 in dynamic GCN. We train all models for 100 epochs with the Adam optimizer, where batch size is set to 32 and learning rate is set to 0.001. All the training and evaluation phases were conducted on a machine with Ubuntu 20.04, Intel i7-9700K 3.6GHz CPU, and Nvidia GTX 2070 8GB GPU, while the methods are implemented in Python 3.7 with the PyTorch 1.9.0. 5-fold cross-validation was conducted for tuning the hyper-parameters.

Evaluation Metrics. Following (Wang et al. 2022b) for evaluating stochastic models, we generate 10 sequences for each rally and take the closest one as the final prediction for each rally, and utilize mean square error (MSE) and mean absolute error (MAE) for the prediction of the locations of the two players, and cross-entropy (CE) for the prediction of shot type. All the results are the average of 5 different random seeds.

4.2 Performance Comparison

To comprehensively evaluate the performance of each method, we perform experiments on three different observed stroke sequences $\tau = 2, 4, \text{ and } 8$. We note that extensive ablation experiments are discussed in Appendix C.1. Table 1 reports the results of our proposed model and the baselines, which demonstrates that DyMF consistently surpasses all baselines in terms of all metrics and different observed lengths. Quantitatively, the improvement in our model is up to 35.3% for MSE, 21.5% for MAE, and 24.3% for CE. We summarize the observations as follows:

- 1) Seq2Seq and TF degenerate location performance (MSE and MAE) compared to GCN_{PM} and R-GCN_{PM}, which not only indicates that naive sequence-to-sequence methods fail to model structural information between locations, but also demonstrates the importance of applying the PM graph to capture the influence on the intent of movement.
- 2) dNRI hinders all performance more substantially, although dynamically modeling the relation between players, which implies that two same nodes for representing two players separately across time is insufficient for the movement forecasting task. This again raises the need for using a PM graph with new location nodes to predict the next move of each player.
- 3) As ShuttleNet is designed for turn-based sequences and fusing different aspects according to different strokes, we can observe that it performs better predictions than both the sequence-based and graph-based baselines. Nonetheless, the comparison of DyMF and ShuttleNet reveals the importance of considering the purpose of players’ movements. Our proposed DyMF outperforms Shut-

Group	Model	$\tau = 2$			$\tau = 4$			$\tau = 8$			Avg. Rank
		MSE location	MAE location	CE shot type	MSE location	MAE location	CE shot type	MSE location	MAE location	CE shot type	
Sequence	Seq2Seq	1.3093	1.7552	2.0687	1.1880	1.6739	1.9905	1.2033	1.6920	1.9751	4.8
	TF	1.2233	1.7122	2.0915	1.2206	1.7094	2.0258	1.1569	1.6785	2.0134	5.0
	ShuttleNet	<u>1.1345</u>	<u>1.6179</u>	<u>1.9599</u>	<u>1.1539</u>	1.6364	<u>1.9575</u>	<u>1.1399</u>	<u>1.6360</u>	<u>1.9654</u>	<u>2.1</u>
Graph	dNRI	1.2803	1.7865	2.4269	1.2994	1.7921	2.4214	1.4651	1.9117	2.4318	7.2
	GCN _{PM}	1.2018	1.6878	1.9630	1.2033	1.6964	1.9854	1.2676	1.7508	1.9875	4.4
	R-GCN _{PM}	1.1724	1.6329	2.0620	1.1681	<u>1.6339</u>	2.0587	1.2006	1.6554	2.0470	3.8
	E-GCN _{PM}	2.8332	2.5003	2.2878	11.031	3.5865	2.2786	1314.0	13.929	2.1180	7.7
DyMF (Ours)		1.1006	1.5795	1.9543	1.1061	1.5888	1.9487	1.0827	1.5739	1.9570	1.0

Table 1: Performance of baseline models and our proposed model. In terms of MSE, MAE, and CE in different observed stroke sequence lengths (τ), the best results are highlighted in boldface and the second best are underlined.

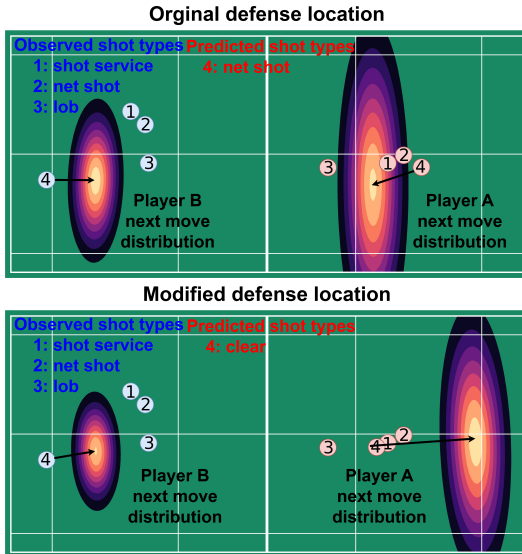


Figure 4: Illustrations of strategies in different defense locations. Red nodes are the served player (Player A) and blue nodes are the receiving player (Player B).

tleNet for all scenarios, which is attributed to our interaction style extractors for modeling mutual interactions across time as well as the dynamic tactics based on the previous movement patterns, and the player-player fusion for integrating the styles of each player. 4) We note that E-GCN_{PM} produces unfavorable performance, possibly due to the unstable states when there are only a few nodes in badminton rallies, and so is not suitable for this task.

4.3 Case Study

The analysis of the players' movements on the court can be viewed as understanding behaviors of players, where stroke forecasting (Wang et al. 2022b) only focuses on what types of strokes, and where to return to. One potential technique to investigate this is to use statistical methods on historical records. However, the occurrence of area coordinates is too sparse to be analyzed. To that end, a paradigm is illustrated

with DyMF to investigate the practicality of movement forecasting, and more cases are discussed in Appendix C.2. Figure 4 shows a scenario where Player A takes a defense location toward the net after returning the lob (red 3) and moving back to the middle court (red 4). In this case, we aim to change the defense location of the player to analyze the reaction of the opponent in different circumstances. In the original defense location, Player B returns a net shot, which leaves Player A in a passive situation, and thus has fewer choices for returning the shuttle. On the other hand, after changing the defense position, Player A is close to the net and makes Player B return a clear that enables Player A to launch an attack. This verifies that the opponent returns the shuttle based on the defense location of the player; therefore, standing at an effective position blocks the choices and forces the opponent to return a predictable stroke. With the help of DyMF providing likely strategies of returning shot types and moving towards a certain position on the court, badminton coaches and players are able to analyze opponents' implicit tactics at the stroke-level, and train for specific scenarios.

5 Conclusion and Future Works

In this paper, we introduce the movement forecasting task, which plays a vital role in investigating offense and defense behaviors in sports analytics. To address the problem, we represent badminton rallies with PM graphs to delve into the strategic structure based on the interrelations between players, and propose DyMF with interaction style extractors for extracting the mutual interactions across time and capturing the dynamic tactics based on the players' past movement patterns. Furthermore, the hierarchical fusions are designed to fuse mutual influence of players' styles and the importance of individual style and interactions. Extensive experiments on the real-world dataset empirically show the effectiveness of our proposed model compared to the baseline methods in different scenarios. Besides, an analysis scenario is illustrated to show the practicality of movement forecasting. For future research, we will extend DyMF to badminton doubles for movements investigation with greater concentration on the interactions between players.

Acknowledgments

This work was supported by the Ministry of Science and Technology of Taiwan under Grants MOST-110-2627-H-A49-001.

References

- Cao, Z.; Liao, T.; Song, W.; Chen, Z.; and Li, C. 2021. Detecting the shuttlecock for a badminton robot: A YOLO based approach. *Expert Syst. Appl.*, 164: 113833.
- Chiu, Y.; Tsai, C.; Sung, W.; and Tsai, Y. 2020. Feasibility of Smartphone-Based Badminton Footwork Performance Assessment System. *Sensors*, 20(21): 6035.
- Chu, X.; Xie, X.; Ye, S.; Lu, H.; Xiao, H.; Yuan, Z.; Chen, Z.; Zhang, H.; and Wu, Y. 2022. TIVIEE: Visual Exploration and Explanation of Badminton Tactics in Immersive Visualizations. *IEEE Trans. Vis. Comput. Graph.*, 28(1): 118–128.
- Decathlon. 2022. Badminton Footwork and Position on the Court. <https://blog.decathlon.in/articles/badminton-footwork-and-position-on-the-court>. Accessed: 2023-02-25.
- Giuliani, F.; Hasan, I.; Cristani, M.; and Galasso, F. 2020. Transformer Networks for Trajectory Forecasting. In *ICPR*, 10335–10342. IEEE.
- Graber, C.; and Schwing, A. G. 2020. Dynamic Neural Relational Inference. In *CVPR*, 8510–8519. Computer Vision Foundation / IEEE.
- Hsu, T.; Wang, C.; Lin, Y.; Chen, C.; Ju, N. P.; Ik, T.; Peng, W.; Wang, Y.; Tseng, Y.; Huang, J.; and Ching, Y. 2019. CoachAI: A Project for Microscopic Badminton Match Data Collection and Tactical Analysis. In *APNOMS*, 1–4. IEEE.
- Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR (Poster)*. OpenReview.net.
- Lu, J.; Yang, J.; Batra, D.; and Parikh, D. 2016. Hierarchical Question-Image Co-Attention for Visual Question Answering. In *NIPS*, 289–297.
- Melton, E. N.; and MacCharles, J. D. 2021. Examining sport marketing through a rainbow lens. *Sport Management Review*, 24(3): 421–438.
- Pareja, A.; Domeniconi, G.; Chen, J.; Ma, T.; Suzumura, T.; Kanezashi, H.; Kaler, T.; Schardl, T. B.; and Leiserson, C. E. 2020. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. In *AAAI*, 5363–5370. AAAI Press.
- Petra. 2020. How data analytics is revolutionizing the NBA. <https://digital.hbs.edu/platform-digit/submission/how-data-analytics-is-revolutionizing-the-nba/>. Accessed: 2023-02-25.
- Qian, T. Y. 2022. Watching sports on Twitch? A study of factors influencing continuance intentions to watch Thursday Night Football co-streaming. *Sport Management Review*, 25(1): 59–80.
- Research, G. V. 2021. Sports Analytics Market Size, Share & Trends Analysis Report By Component (Software, Service), By Analysis Type (On-field, Off-field), By Sports (Football, Cricket, Basketball, Baseball), And Segment Forecasts, 2021 - 2028. <https://www.grandviewresearch.com/industry-analysis/sports-analytics-market/>. Accessed: 2023-02-25.
- Sak, H.; Senior, A. W.; and Beaufays, F. 2014. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. *CoRR*, abs/1402.1128.
- Schlichtkrull, M. S.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling Relational Data with Graph Convolutional Networks. In *ESWC*, volume 10843 of *Lecture Notes in Computer Science*, 593–607. Springer.
- Shafiee, N.; Padir, T.; and Elhamifar, E. 2021. Introvert: Human Trajectory Prediction via Conditional 3D Attention. In *CVPR*, 16815–16825. Computer Vision Foundation / IEEE.
- Shi, X.; Shao, X.; Fan, Z.; Jiang, R.; Zhang, H.; Guo, Z.; Wu, G.; Yuan, W.; and Shibasaki, R. 2020. Multimodal Interaction-Aware Trajectory Prediction in Crowded Space. In *AAAI*, 11982–11989. AAAI Press.
- Steels, T.; Herbruggen, B. V.; Fontaine, J.; Pessemier, T. D.; Plets, D.; and Poorter, E. D. 2020. Badminton Activity Recognition Using Accelerometer Data. *Sensors*, 20(17): 4685.
- Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to Sequence Learning with Neural Networks. In *NIPS*, 3104–3112.
- Valdecabres, R.; Casal, C. A.; Chiminazzo, J. G. C.; and De Benito, A. M. 2020. Players' on-court movements and contextual variables in badminton world championship. *Frontiers in Psychology*, 11: 1567.
- Wang, W. 2022. Modeling Turn-Based Sequences for Player Tactic Applications in Badminton Matches. In *CIKM*, 5128–5131. ACM.
- Wang, W.; Chan, T.; Yang, H.; Wang, C.; Fan, Y.; and Peng, W. 2021. Exploring the Long Short-Term Dependencies to Infer Shot Influence in Badminton Matches. In *ICDM*, 1397–1402. IEEE.
- Wang, W.; Chan, T.; Yang, H.; Wang, C.; Fan, Y.; and Peng, W. 2022a. How Is the Stroke? Inferring Shot Influence in Badminton Matches via Long Short-Term Dependencies. *ACM Trans. Intell. Syst. Technol.*, 14(1).
- Wang, W.; Shuai, H.; Chang, K.; and Peng, W. 2022b. ShuttleNet: Position-Aware Fusion of Rally Progress and Player Styles for Stroke Forecasting in Badminton. In *AAAI*, 4219–4227. AAAI Press.
- Wang, W.-Y.; Chang, K.-S.; Chen, T.-F.; Wang, C.-C.; Peng, W.-C.; and Yi, C.-W. 2020. Badminton Coach AI: A Badminton Match Data Analysis Platform Based on Deep Learning. *Physical Education Journal*, 53(2): 201–213.
- Yoshikawa, Y.; Shishido, H.; Suita, M.; Kameda, Y.; and Kitahara, I. 2021. Shot detection using skeleton position in badminton videos. In *International Workshop on Advanced Imaging Technology 2021*, volume 11766, 312 – 317.