Development of Defence and Offence Play Items for Deep Learning Model of Offence Play Analysis in Soccer Game

Hiroki Matsuoka*, Yasuhiro Tahara*, Kozue Ando and Takahiko Nishijima****

*Doctoral Program in Graduate School of Comprehensive Human Sciences, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574 Japan
s1730500@s.tsukuba.ac.jp
**Faculty of Health & Sport Sciences, University of Tsukuba
[Received September 9, 2019; Accepted June 22, 2020]

The purpose of this study was to develop offence and defence tactical play items for a deep learning model of tactical play analysis. The procedures used in this study included the following three steps. First, features of offence and defence tactical play items were developed. Then, long short-term memory (LSTM) was architected. Finally, tactical plays were predicted by the model. The ball touch data and the tracking data from two official soccer games in the J-League 2016 season were used. The ball touch data, recorded player actions in text such as passes and shots with time-series order, and the tracking data of players, were used to construct thirty-one tactical play items. For the deep learning model, LSTM was used. LSTM allows the analysis of time-series text data. 6,444 sequential plays were used. The highly accurate tactical play predicted from LSTM was the feed after tactical play which started at low press defence and finished at GK ball catch (8 correct predictions of 8 frequencies in the data). In conclusion, all 31 items measuring offence and defence tactical play in soccer games constructed from the ball touch data and the tracking data are the feature items used to analyse tactical plays using LSTM.

Keywords: machine learning, long-short term memory, ball touch data, tracking data

[Football Science Vol.17, 69-85, 2020]

1. Introduction

Big data in relation to soccer games, such as tracking data and ball-touch data, has been collected in the J.League in Japan and in the top leagues in Europe. The game performance analysis in soccer has expanded to include data-driven analysis by machine learning (ML) and big data. For applying ML to the analysis, features to represent offences or defences in soccer are needed.

To measure offence performance, Tenga et al. (2009; 2010a; 2010b) constructed variables such as "pass penetration" and "centre pass" based on the divided pitch zones or the goal position. However, those variables were not enough to be able to apply ML, because of limitations in data collection such as time-consuming tasks and reliability of data.

Luccey et al. (2014) analysed expected goal value in soccer using ML and positional information of players in tracking data. Hobbs et al. (2018) classified fast attack by ML and positional information of players in 10 second intervals in tracking data. Decroos et al. (2019) developed features of individual actions such as passes and shots from ball-event data from various sports companies to analyse action values leading to goals. Features or items of game performance have been developed to apply ML for performance analysis in previous research. They allowed the estimation of action values which permit the evaluation of individual player performance based on a ball related action other than the number of goals scored or the number of shots in the game.

However, soccer game analysis using ML has in the past used only technical element items such as ball-related actions like passes and shots; there is a need to measure tactical elements in order to analyse soccer tactical performance. Moreover, features of tactical performance need to be developed from tracking data and ball-touch data for a data-driven analysis of tactical performance using ML or deep learning (DL).

A DL model to analyse tactical play sequences is needed because tactical performance of the offence and defence consists of consecutive tactical actions with time series order. In natural language processing (NLP) research, researchers have been trying to get a computer to learn human language such as words or conversations. In the NLP language model (LM), a context is predicted and reproduced by estimating the probability of the next word appearance from previous sequences of words in a given context using DL models of recurrent neural network (RNN) or long short-term memory (LSTM) (Mikolov et al., 2010; 2011; Sundermeyer et al., 2012).

Assuming that consecutive actions in the defence and the offence correspond to a context in NLP, it is possible to predict and reproduce the next action appearances. For example, the probability of a shot can be predicted from past tactical play sequences consisting of consecutive tactical actions [pressing defence, ball gain, passing, dribbling, crossing]. RNN or LSTM can be applied to analyse tactical play sequences represented in the sequential text data of a pass or a dribble in ball-touch data. For the DL to analyse a tactical play sequence, starting from defence and finishing with offence, feature items to measure offence tactical actions such as forward pass are necessary. In addition, defence tactical items such as defence line position need to be developed from tracking data.

Thus, the purpose of this study was to develop feature items for analysis of tactical sequences from defence to offence in soccer games. The offence and defence tactical play items were developed using ball-touch data and tracking data. Then, LSTM was used as DL model. Finally, feature items of tactical sequences for the DL model were analysed.

2. Methods

2.1. Research procedures

The research procedures are shown in **Table 1**. There were three steps: the development of feature items, the development of the DL model, and examination of tactical play items for the DL model. In step 1, offensive tactical play items were developed from ball-touch data, and defensive tactical play items were developed from tracking data. In step 2, a neural network model and a LSTM language model (LSTM-LM), were developed. In step 3, the offence and defence tactical play items were examined using the results of the LSTM-LM. This study was approved by the Research Ethics Committee at the Faculty of Health and Sport Sciences, University of Tsukuba (Project No. 30-54).

2.2. Data for analysis

Most research on soccer game performance analysis using ML has used tracking data or ball-touch data (Herold et al., 2019). In the present study, tracking

Table 1 Steps of research procedures

Steps	Sub-Steps	Procedures
1. Feature processing	Analysis of play items	The causal-effect analysis
	Development of offence play items	Processing of plays from ball touch data
	Development of defence play items	Processing of play items from ball touch data and ball touch data Analysis of DFL press by decision tree analysis
2. Model development	Development of baseline model	Neural network
	Development of LSTM model	Long Short-Term Memory
3. Interpretation of predicted offence play	Analysis of descriptive statistics of final play items and previous play items	Frequency of final play items and previous play items
	Prediction of attacking play from LSTM	Characteristics of offence plays with high prediction rate Characteristics of offence plays leads to target offence tactical item

data and ball-touch data of two matches in the 2016 J.League Championship final (1st and 2nd leg) were used to develop the offence and defence tactical play items. Ball-touch data was collected by trained staff at a sports analytics company using software for soccer game data collection (Kato, 2016). In ball-touch data, game ID, halves, time, offence ID, possession team, player, technical play item (action), and location of ball (x, y) were recorded (**Appendix 1**). Tracking data was collected using TRACAB (ChyronHego, NY, USA) which has high data measurement accuracy (Linke et al., 2020). Location of all players on the pitch was recorded in every frame (1/25 fps, 0.04s). Tactical play items were constructed from ball-touch data and tracking data (**Appendix 2**).

The sample comprised 7,163 actions in 599 possessions in 2 games, recorded using ball-touch data. Possessions (n = 31) which did not continue for more than 2 actions were eliminated.

Then, 6,444 play sequences were extracted from 7,163 actions. First, offence play data, showing consecutive technical play items from the defence (ball gain) to the offence, were processed from technical play items (actions) in ball-touch data (**Appendix 3**). Offence play sequence data were then processed from the offence play data by dividing consecutive technical play items in each offence ID into play sequences of consecutive actions (**Table 2**).

An example of tactical play sequence in the offence play sequence data is as follows: An offence play [1. Tackle with HPD, 2. Ball control, 3. Through ball to central area, 4. Ball control, 5. Shot in PA] in **Appendix 1** consisted of 5 consecutive tactical play items (tactical actions) that started with 1. Tackle with high press defence (HPD) and finished with 5. Shot in penalty area (PA). The offence was divided into four tactical play sequences, as shown in **Table 2**, including the play sequence 1 [1. Tackle with HPD, 2. Ball control], the play sequence 2 [1. Tackle with HPD, 2. Ball control, 3. Through ball to central area], the play sequence 3 [1. Tackle with HPD, 2. Ball control, 3. Through ball to central area, 4. Ball control], and the play sequence 4 [1. Tackle with HPD, 2. Ball control, 3. Through ball to central area, 4. Ball control, 5. Shot in PA]. The play sequence 4 corresponds to the offence in **Appendix 1**. To predict the appearance of the last tactical play items in the play sequences, the tactical play items at the right side in the play sequences were defined as dependent variables (**Play$_t$**), and left-side consecutive tactical play items were used as independent variables. This means that the independent variable in the play sequence 1 is 1. Tackle with HPD and the dependent variable is 2. Ball control. In the play sequence 4, the independent variables are four consecutive tactical play items [1. Tackle with HPD, 2. Ball control, 3. Through ball to central area, 4. Ball control] and the dependent variable is 5. Shot in PA.

2.3. Construction of feature items

2.3.1. Fish-bone diagram of offence and defence tactical play items

The offence and defence tactical play items for soccer were constructed using ball-touch data and tracking data (**Appendix 2**).

This study constructed the offence tactical play items in terms of invasion in the longitudinal

Table 2 Description play sequence data

					Play sequence				
Game ID	Half	Offence ID	Team	Play sequence ID	Play$_{t-4}$	Play$_{t-3}$	Play$_{t-2}$	Play$_{t-1}$	Play$_t$
1	1st Half	1	Team 1	1				Tackle with HPD (start)	Ball control (+1)
1	1st Half	1	Team 1	2			Tackle with HPD (start)	Ball control (+1)	Through ball to central area (+2)
1	1st Half	1	Team 1	3		Tackle with HPD (start)	Ball control (+1)	Through ball to central area (+2)	Ball control (+3)
1	1st Half	1	Team 1	4	Tackle with HPD (start)	Ball control (+1)	Through ball to central area (+2)	Ball control (+3)	Shot in PA (+n)
1	1st Half	2	Team 1					MPD	MPD
		.		.					
		.		.					
		.		.					
2	2nd Half	599	Team 2	6,444					MPD

direction on the pitch followed by offence tactics, as in previous research (Yates et al., 2006; Tenga et al., 2010a; 2010b; Lucey et al., 2014; Fernandez-Navarro et al., 2016; Rathke 2017; Hughes & Lovell, 2019).

Features of tactical performance processed from ball-touch data are shown on a fish-bone diagram (**Figure 1**). The top side of the diagram shows the offence tactical play items and the bottom side shows the defence tactical play items. The box highlighted in bold shows the technical play items (actions) recorded in ball-touch data. The offence technical items were: shot, through ball, cross, dribble, and pass. The defence technical items were: intercept, tackle, block, clearance, and hand clearance. The definition of the items included the definition of the variables given in Liu et al. (2013). The definition of GK hand clearance used in this study includes ball punching and ball fisting. The box that is not highlighted in bold shows the offence and defence tactical play items which are newly processed in this study.

2.3.2. Construction of the offence tactical play items

The offence tactical play items are shown in **Table 3**, and the ball positions used as criteria are shown on **Figure 2**. Thirteen offence tactical play items were constructed based on the nine offence technical play items (shot, dribble, through-ball, cross, pass, flick-on, ball control, and ball touch).

Shot in PA and shot outside of PA were constructed separately because shot position affects shot success (Lucey et al., 2014; Rathke 2017). This study defined pass direction (forward, square, backward) following Fernandez-Navarro et al. (2016), because Yates et al. (2006) showed the importance of forward passes (Appendix 4). However, if ball location was not recorded on ball-touch data due to technical problems, then the item name was defined as pass. Through ball is a crucial action for increasing scoring opportunities (Tenga et al., 2010a; 2010b) by breaking the opponent's defence line (DFL), so through ball was defined as through ball to central area or through ball to wide area. The cross was defined as either an early cross to behind the opponents' DFL or as an ordinary cross. Dribble is an action to transfer the ball forward, but also dribbling that cuts in to the central area is a crucial play that leads to shots. This is why the criteria of dribbling were defined in terms of direction.

Sixteen items, including the four offence technical play items of flick-on, ball control, ball touch, and kick-off, the three GK related technical play items of goal kick, GK ball catch, and feed, the five set-piece related items of throw-in, PK, CK, direct FK, and indirect FK, the three foul related items of foul committed, foul suffered, and offside committed, and ball-out, were processed from the record of ball-touch data.

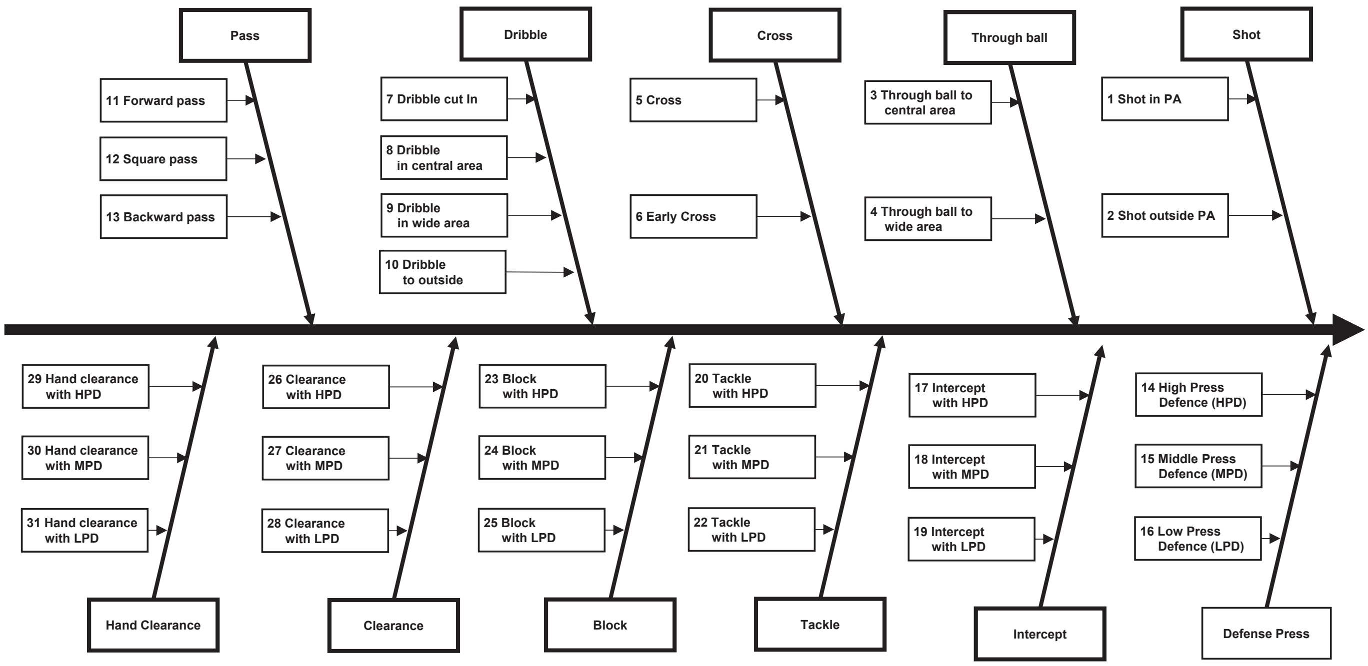


Figure 1 The causal-effect diagram of offence and defence tactical skill items to measure features

Table 3 Offence tactical skill items and criterion

No	Offence tactical skill item	Criterion
1	Shot in PA	Shot location: x>88.5, 52.5m =< y =>17.5m
2	Shot outside PA	Shot location: x<=88.5m, y>52.5m or <17.5m
3	Through ball to central area	next ball location: 52.5m =< y =>17.5m
4	Through ball to wide area	next ball location: y>52.5m or <17.5m
5	Cross	all crosses except early cross
6	Early cross	Cross location: x<=88.5m, y=>52.5m or =<17.5m
7	Dribbling Cut In	Dribbling start location: y<52.5m or >17.5m Dribbling end location: 52.5m =< y =>17.5m
8	Dribbling in central area	Dribbling start location: 52.5m =< y =>17.5m Dribbling end location: 52.5m =< y =>17.5m
9	Dribbling in wide area	Dribbling start location: y<52.5m or >17.5m Dribbling end location: y<52.5m or >17.5m
10	Dribbling to outside	Dribbling start location: 52.5m =< y =>17.5m Dribbling end location: y<52.5m or >17.5m
11	Forward pass	Degree : 0-45 or 315-359 degree
12	Square pass	Degree : 45-135 or 225-315 degree
13	Backward pass	Degree : 135-225 degree

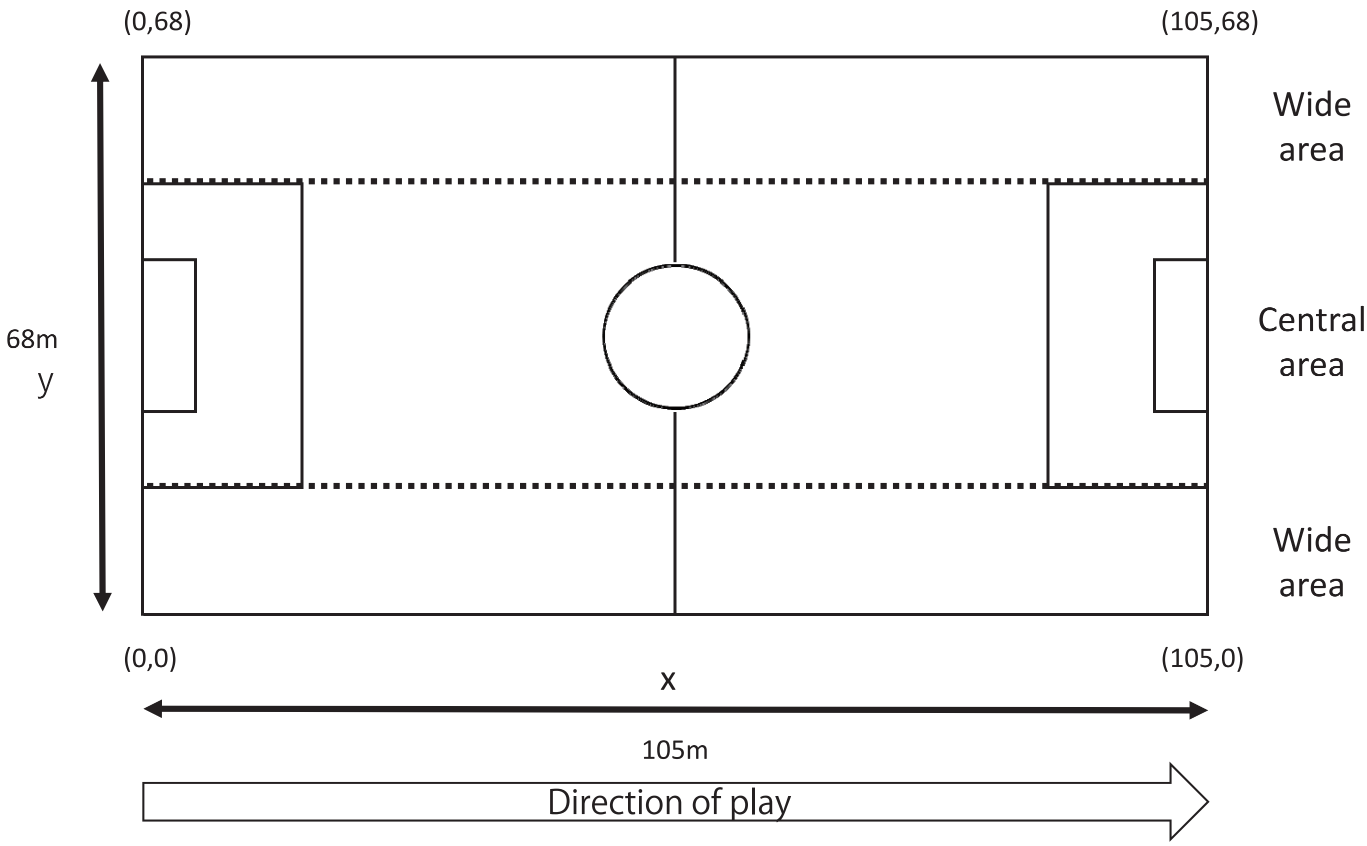


Figure 2 Ball position (x, y) and divisions of the pitch

2.3.3. Construction of the defence tactical play items

The objective of the pressing defence is to gain the ball and prevent the opponents' build up by decreasing space and time (Lucchesi, 2004). High pressing defence intends to press and push back the opponents to the opponents' goal (Fernandez-Navarro et al., 2016). The pressing defence was classified based on the assumption of different DFL speed in each DFL position. The defence classification was high press defence (HPD) to push up the DFL to decrease the opponents' space and time, middle press defence (MPD) to keep the DFL stable to prepare against opponents' invasion, and low press defence (LPD) to move the DFL back towards one's own goal to protect the goal.

Three DFL press items – HID, MID, and LPD – were defined. Fifteen defence tactical items were constructed by combination of the three DFL press items and five defence technical items: interception, tackling, blocking, clearance, and GK hand clearance. In total, eighteen defence tactical items were constructed (**Table 4**).

2.3.4. Classification of DFL press

DFL press criteria were classified by decision tree analysis using the DFL position and the DFL speed. The independent variable was the result of DFL press, and dependent variables were the DFL position and the DFL speed. The DFL position was measured by the position of the deepest player (apart from the goalkeeper) in tracking data. The DFL speed was measured by the vertical movement of the DFL divided by the time to movement. To take

Table 4 Defence tactical skill items and criterion

No	Defence tactical skill item	Criterion
14	High press defence (HPD)	(a) DFL speed: > 0.1 km/h (when DFL position: x > 38.7 m) DFL speed: > 3.5 km/h (when DFL position: x > 30.0 m) DFL speed: > 1.6 km/h (when DFL position: x > 10.2 m) DFL speed: > 3.5 km/h (when DFL position: x <= 10.2 m)
15	Middle press defence (MPD)	(b) DFL speed: > -4.6 km/h (when DFL position: x > 38.7 m) DFL speed: > -1.2 km/h (when DFL position: x > 30.0 m) DFL speed: > -1.2 km/h (when DFL position: x > 10.2 m) DFL speed: > -4.6 km/h (when DFL position: x <= 10.2 m)
16	Low press defence (LPD)	(c) DFL speed: <= -4.6 km/h (when DFL position: x > 38.7m) DFL speed: <= -1.2km/h (when DFL position: x > 30.0m) DFL speed: <= -1.2 km/h (when DFL position: x > 10.2m) DFL speed: <= -4.6 km/h (when DFL position: x <=10.2m)
17	Intercept with HPD	Intercept with condition (a)
18	Intercept with MPD	Intercept with condition (b)
19	Intercept with LPD	Intercept with condition (c)
20	Tackle with HPD	Tackle with condition (a)
21	Tackle with MPD	Tackle with condition (b)
22	Tackle with LPD	Tackle with condition (c)
23	Block with HPD	Block with condition (a)
24	Block with MPD	Block with condition (b)
25	Block with LPD	Block with condition (c)
26	Clearance with HPD	Clearance with condition (a)
27	Clearance with MPD	Clearance with condition (b)
28	Clearance with LPD	Clearance with condition (c)
29	Hand Clearance with HPD	Hand Clearance with condition (a)
30	Hand Clearance with MPD	Hand Clearance with condition (b)
31	Hand Clearance with LPD	Hand Clearance with condition (c)

into account the direction of pressing, the DFL speed was shown with a negative value when the DFL was moving towards the own goal direction, and it was shown with a positive value when the DFL was moving towards the opponents' goal. The DFL speed with a positive value can be interpreted as the pressing defence. In the result of DFL press was binary scale data. A successful DFL press was defined as when the opponents passed in the backward direction, and an unsuccessful DFL press was defined as when the opponents passed the ball in the forward direction.

In the decision tree analysis, the DFL position was designated as the first root node. Three DFL press items—HPD, MPD, and LPD—were classified by division of the DFL speed in each range of the DFL position affecting the result of the DFL press.

The CHAID algorithm was used for the classification (Kass, 1980). Chi-square statistics were used to classify the nodes. The defence play data showing the DFL position and the DFL speed corresponding to the opponents' offence action was processed. The defence play data (n = 3,460) were used for the analysis of classification accuracy. 70% (n = 2, 474) were used for training data, and 30% (n = 986) were used as validation data. IBM SPSS ver.23.0 was used.

2.4. Application of the long short-term memory language model to soccer tactical play analysis

Long short-term memory (LSTM), an improved RNN model, was used. RNN allows the analysis of time-series data and contexts which have sequential meaning. Hidden layers in RNN have a loop, and this loop allows information to be passed from the previous hidden layer to the current hidden layer (Elman, 1990), which outputs current h_t to the next (**Left side of Appendix 5**). Here, t means current time. The right side of **Appendix 5** shows a structure of expanded hidden layers. The RNN layer has the role of passing the previous information to the next RNN layer. The passed information and current input data determine the output.

LSTM allows the analysis of longer time series data (Hochreiter & Schmidhuber, 1997), and the LSTM layer is used in the hidden layer. The LSTM layer has a cell state (C_t) in memory and controls information in the cell state (**Appendix 6**). The amount of information in the cell state (C_t) is controlled by the forget gate (f), input gate (i), and tanh layer ($\tilde{C}_t$). These gates and layers affect the cell state, and information passes through the output gate (o) to the next LSTM layer. The forget gate (f) has a sigmoid function and decides how much information to delete from the cell state (C_t). The input gate has current inputted data and tanh layer, ($\tilde{C}_t$) and determines how much information to update in the cell state (C_t). Finally, the product of the input gate (i) and tanh layer ($\tilde{C}_t$) affects the cell state (C_t). The output gate (o) decides how much information in the cell state (C_t) to pass to the next step (h_{t+1}).

h_t is calculated by the equations shown below. Where, σshows a sigmoid function, tanh shows a hyperbolic function, X shows inputs data, W shows the weight of parameter, and b shows a constant.

$$h_t = o_t \times \tanh(C_t) \quad \cdots (1)$$

$$o_t = \sigma(X_t W_X^{(o)} + h_{t-1} W_h^{(o)} + b^{(o)}) \quad \cdots (2)$$

$$C_t = (C_{t-1} \times f_t) + (\tilde{C}_t \times i_t) \quad \cdots (3)$$

$$f_t = \sigma(X_t W_X^{(f)} + h_{t-1} W_h^{(f)} + b^{(f)}) \quad \cdots (4)$$

$$\tilde{C}_t = \tanh(X_t W_X^{(\tilde{C}_t)} + h_{t-1} W_h^{(\tilde{C}_t)} + b^{(\tilde{C}_t)}) \quad \cdots (5)$$

$$i_t = \sigma(X_t W_X^{(i)} + h_{t-1} W_h^{(i)} + b^{(i)}) \quad \cdots (6)$$

Output (y_t^n) in LSTM-LM is the probability distribution of the next play item ($plays_t^n$) appearance when the play sequence (X = [x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}]) is inputted (Mikolov et al., 2010). This probability distribution shows how much play sequence is natural order, and when play sequence is not unstrained order, a high probability is given. In NLP, the probability distribution of words is shown, but this study estimates probability distribution of tactical play items.

$$y_t^n = Probability(play_t^n | X = [x_{t-4}, x_{t-3}, x_{t-2}, x_{t-1}]) \quad \cdots (7)$$

Probability of each play item (y_t^n) is estimated through the softmax function. The softmax function is used when independent variables are n-classes which is over 3 classes, and the function converts the estimate of each class into a probability distribution representing between 0.0 to 1.0. The probability of each play item is estimated and the sum of the probability in all play items is 1.0.

$$y_t^k = \frac{\exp(h_t^k)}{\sum_{i=1}^{n} \exp(h_t^i)} \quad \cdots (8)$$

Here, n is a number of play items.

Tactical play items were represented in the text such as forward pass and dribbling in the central area, with chronological order. It is assumed that the last action appears depending on the previous play sequence. For example, take the situation where offence play started with a ball gain by tackling with high press defence, and the player controls the ball, and took the ball through. Finally, the player controlled the ball and took a shot. In this play sequence, it is assumed that the shot was produced through the influence of previous consecutive actions such as tackling with high press defence and through ball. The probability of the appearance in the next play item is estimated by learning the relationship between consecutive play items in the text. For example, if the tactical play sequence X consisting of four actions (X = x_{t-4}[Tackle with HPD], x_{t-3}[ball control], x_{t-2} [Through ball to central area], x_{t-1} [ball control]) is inputted to the model, the probability of the next play item (y_t) is estimated from chronological relationship of four plays ($X_{t-4} \sim X_{t-}$) (**Figure 3**).

2.5. Development of DL model

This study developed the neural network and LSTM-LM using Keras, a Python API. In the neural network model, a network with three layers consisted of an input layer, a hidden layer, and an output layer. The dimension of the hidden layer was set at 32. The activate function in the hidden layer was a sigmoid function; the loss function was categorical-cross entropy. Adaptive moment estimation (Adam) was used as the optimiser. In the LSTM-LM, the LSTM module in Keras was used. The dimension of the LSTM layer was set at 32. The activate function, loss function and optimizer were softmax function, categorical-cross entropy, and adaptive moment estimation (adam), respectively.

To learn the tactical play sequence of four items, the length of the input data was set as four. Hughes and Lovell (2019) noted that the initial two actions after ball gain are crucial for increasing scoring opportunity, but they did not measure actions of ball control or ball touch of the ball holding player. Thus, this study considered that input data consisted of four tactical plays, including one defence tactical play item for ball gain, and three offence tactical play items

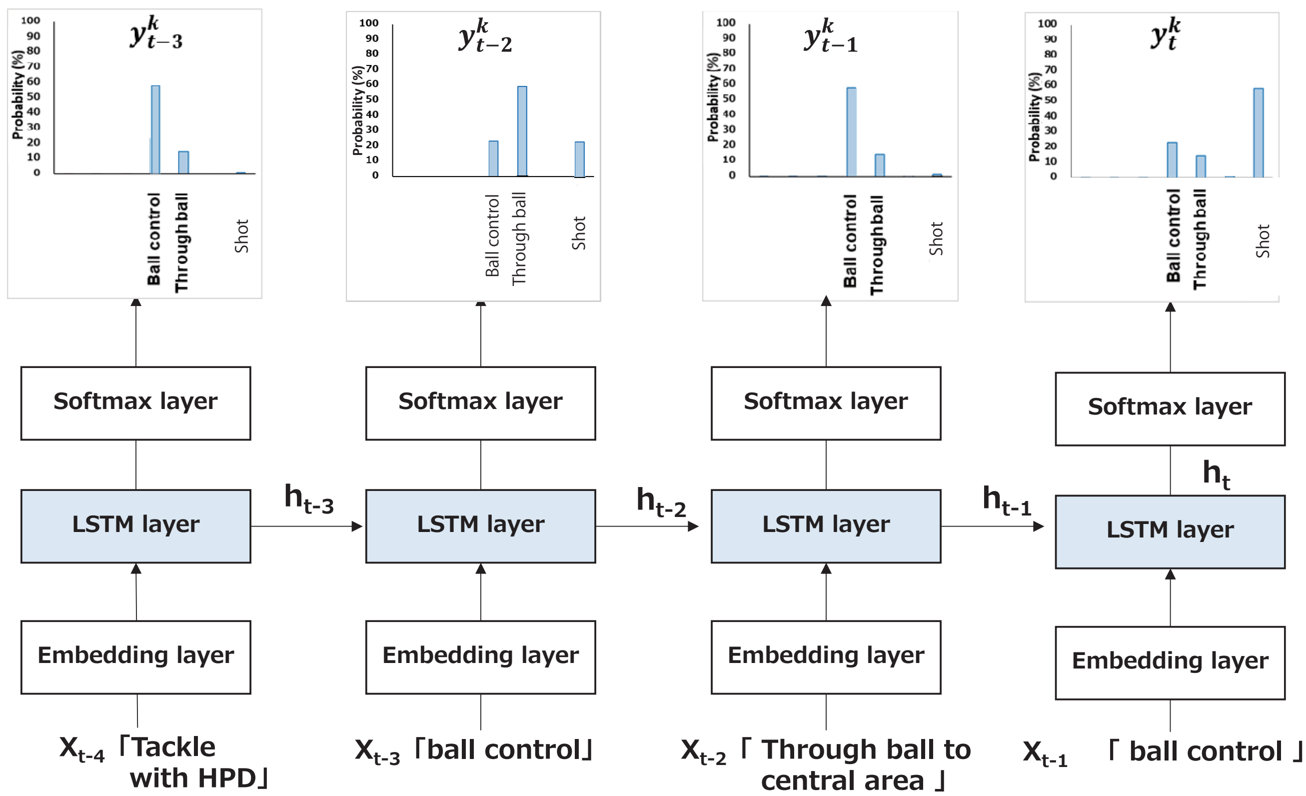


Figure 3 LSTM model for soccer tactical play analysis

including ball control or touches.

Play sequence data was divided into training data (5,155 sequences) and validation data (1,289 sequences). Epoch was set at 50 times and batch size was set at 32. The models were evaluated by loss and accuracy of validation data. Moreover, perplexity was used for evaluation of LM.

$$\text{Perplexity} = e^{loss} \quad \cdots (9)$$

2.5. Examination of learning results of consecutive tactical play

To examine the learning results of LSTM-LM using 48 tactical play items, comparison between a number of last play items and a previous play item in 599 possessions, characteristics of tactical play with high accuracy of learning results, and characteristics of tactical play leading to attacking results predicted by LSTM-LM.

This study counted the number of last play items in 599 possessions to compare with the learning results of LSTM-LM. In addition, the number of the last play items and the number of previous play items in 599 possession play instances was counted to show the relationship between the consecutive actions. Ten offence play items related to attacking results (1. shot in PA, 2. shot outside PA, 3. through ball to central area, 4. through pass to wide area, 5. cross, 6. early cross, 7. dribbling cut in, 8. dribbling in central area, 9. dribbling in wide area, 10. dribbling to outside) were shown.

Then, the characteristics of the tactical play sequence with high accuracy of prediction results by LSTM-LM were examined. The probability of the next tactical play items for all 48 items was estimated by the tactical play sequences of four items. The tactical play items with the highest probability were interpreted as the last tactical play items which were predicted by LSTM-LM. Play sequence X(X[$Play_{t-4}$, $Play_{t-3}$, $Play_{t-2}$, $Play_{t-1}$]), last play item ($play_t^k$) and its probability, the number of the last play items predicted by the play sequence with LSTM-LM (N of prediction), the number of the last play items that appeared after the same play sequence in the real data (N in data), and accuracy (N of prediction / N in data) were shown in the tables. Then, the characteristics of the tactical play sequences leading to the highest probability of offence play items were examined. Five last offence play items – shot in PA, shot outside of PA, through ball to central area, early cross, and cross – were chosen to be analysed. The tactical play sequence X([$Play_{t-4}$, $Play_{t-3}$, $Play_{t-2}$, $Play_{t-1}$]), offence play item ($play_t^k$) and its probability, the number of the last offence play items predicted from the play sequence by LSTM-LM (N of prediction), the number of the offence play items that appeared after the play sequence in the real data (N in data), and accuracy (N of prediction / N in data) were shown. In addition, the predicted tactical play items with a higher probability of appearance than the offence play items were shown.

3. Results

3.1. Construction of tactical play items

Thirty-one tactical play items were constructed from tracking data and ball-touch data (**Table 3**, **Table 4**). Defence tactical play items of DFL press (high press defence, middle press defence, and low press defence) were constructed. These items were classified by the DFL position and the DFL speed to determine the successful defence pressing (**Table 4**). For example, a criterion of HPD was the DFL speed greater than 0.1 km/h when the DFL position stayed 38.7m away from their goal. However, the criteria of HPD were different in each category of the DFL position, showing the DFL speed greater than 3.5km/h when the DFL position stayed more than 30m away from their goal, the DFL speed greater than 1.6km/h when the DFL stayed more than 10.2m away from the goal, and the DFL speed over 3.5km/h when the DFL was positioned closer than 10.2m to the goal. The accuracy of the classification showed as 72.1% (711/986) in validation data.

3.2. Development of the DL framework

The neural network and LSTM-LM models were developed. Accuracy, loss, and perplexity after learning the neural network model were 0.33, 2.27 and 9.68, respectively. Accuracy, loss, and perplexity after learning the LSTM-LM were 0.43, 1.92 and 6.82, respectively.

3.3. Examination of learning results tactical play sequence.

The last tactical play items in 599 possessions were counted in **Table 5**. The number of shots in PA, an offence result, was 13 times (2%) in all possessions. In **Table 6**, the relationship between the last tactical play items and a previous tactical play items was shown. Previous tactical play items, cross (3 times) and CK (3 times), leading to shot in PA in the next were the highest frequency in the possessions. A previous play that led to through ball to central area was ball control (4 times). Previous play that led to cross and early cross were dribble in wide area (6 times) and ball control (4 times), respectively.

Tactical play sequences with high accuracy of

Table 5 Number of the last play items of offence in two games

		Game I		Game II			
No	**Final play**	**team A**	**team B**	**team C**	**team D**	**N**	**(%)**
1	Shot in PA	4	1	3	5	13	2.17
2	Shot outside PA	2	3	1	2	8	1.34
3	Through pass to central area	1	3	3	1	8	1.34
4	Through pass to wide area	5	1	4	3	13	2.17
5	Cross	2	2	5	5	14	2.34
6	Early cross	2	1	3	1	7	1.17
7	Dribbling cut In	0	0	0	0	0	0.00
8	Dribbling in central area	0	2	0	1	3	0.50
9	Dribbling in wide area	2	2	1	5	10	1.67
10	Dribbling to outside	1	2	0	0	3	0.50
11	Forward pass	32	35	30	26	123	20.53
12	Square pass	15	15	17	22	69	11.52
13	Backward pass	0	6	4	4	14	2.34
14	High press defence	0	0	0	0	0	0.00
15	Middle press defence	0	0	0	0	0	0.00
16	Low press defence	0	0	0	0	0	0.00
17	Intercept from high press defence	0	0	0	0	0	0.00
18	Intercept from middle press defence	0	0	0	0	0	0.00
19	Intercept from low press defence	0	0	0	0	0	0.00
20	Tackle from high press defence	2	2	0	2	6	1.00
21	Tackle from middle press defence	2	1	3	1	7	1.17
22	Tackle from low press defence	3	6	4	1	14	2.34
23	Block from high press defence	0	0	3	1	4	0.67
24	Block from middle press defence	1	2	7	2	12	2.00
25	Block from low press defence	2	5	3	5	15	2.50
26	Clearance from high press defence	1	4	6	2	13	2.17
27	Clearance from middle press defence	2	3	4	6	15	2.50
28	Clearance from low press defence	4	5	3	2	14	2.34
29	Hand Clearance from high press defence	0	0	0	0	0	0.00
30	Hand Clearance from middle press defence	0	0	0	1	1	0.17
31	Hand Clearance from low press defence	0	0	0	1	1	0.17
32	Ball control	25	21	15	19	80	13.36
33	Ball touch	9	8	8	11	36	6.01
34	Frick on	6	3	2	1	12	2.00
35	Pass	0	3	3	1	7	1.17
36	Feed	3	1	0	1	5	0.83
37	GK ball catch	9	10	5	10	34	5.68
38	Through-In	2	1	7	6	16	2.67
39	Direct FK	5	1	1	4	11	1.84
40	Indirect FK	0	0	0	0	0	0.00
41	Corner kick	3	0	0	1	4	0.67
42	Goal kick	0	0	1	1	2	0.33
43	PK	0	1	1	0	2	0.33
44	Kick off	0	0	0	0	0	0.00
45	Ball out	0	0	0	0	0	0.00
46	Offside	0	0	0	0	0	0.00
47	Foul committed	0	0	0	1	1	0.33
48	Foul conceded	0	2	0	0	2	0.17
	Total	145	152	147	155	599	100

Table 6 Number of the last and previous tactical skill items in offence

No	Last tactical	N	Previous play (N)
1	Shot in PA	13	Cross (3), CK (3), Square pass (2), direct FK (2), Ball control (1), Frick on (1), No previous play (1)
2	Shot outside PA	8	Ball control (3), Dribbling Cut In (1), Dribbling in wide area (1), Forward pass (1),High press defence (1), No previous play (1)
3	Through pass to central area	8	Ball control (4), Dribbling Cut In (1), Square pass (1), Backward pass (1), Middle press defence (1)
4	Through pass to wide area	13	Ball control (6), Forward pass (2), Square pass (2), Middle press defence (2), Clearance from low press defence (1)
5	Cross	14	Dribbling in wide area (6), Ball control (3), Through pass to wide area (2), Through pass to central area (1), Square pass (1), Backward pass (1)
6	Early cross	7	Ball control (4), Through pass to wide area (1), Dribbling in wide area (1), Dribbling to outside (1)
7	Dribbling cut In	0	n/a
8	Dribbling in central area	3	Forward pass (1), Middle press defence (1), Tackle from middle press defence (1)
9	Dribbling in wide area	10	Square pass (3), Through pass to wide area (2), Forward pass (2), Tackle from low press defence (1), no previous play (2)
10	Dribbling to outside	3	Forward pass (1), Square pass (1), no previous play (1)

prediction by LSTM-LM were shown in **Table 7**. Feed (probability = 0.96) was predicted after tactical play sequence X [LPD→ LPD→ LPD→ GK ball catch] by LSTM-LM. The number of tactical play sequences X [LPD→ LPD→ LPD→ GK ball catch] in the data was 8 times (N in data), and the number of prediction of feed after the same tactical sequence was 8 times (N of prediction), meaning that the accuracy was 1.00.

The tactical play sequences leading to the highest probability of the offence tactical items was estimated by LSTM-LM (**Table 8**). The highest probability of shot in PA was 0.11, predicted from tactical play sequence X [Ball control→ Ball control→ HPD→ Block from HPD]. The number of appearances of tactical play sequence X [Ball control→ Ball control→ HPD→ Block from HPD] in the data was 1 time. However, the number of predictions of shot in PA after this tactical play sequence was 0 times, because the probability of HPD (probability = 0.22) was the highest. Accuracy of prediction of shot in PA was 0.00.

The highest probability of cross was 0.60, predicted from tactical play sequence X [Square pass→ Ball control→ Through ball to wide area→ Dribbling in wide area]. The number of appearances of this tactical play sequence in the data was 1 time, and the number of predictions of cross after this tactical play sequence was 1 time (N of prediction). Accuracy of prediction of cross after this tactical play sequence was 1.00.

Table 7 Accuracy of the top 5 tactical play items predicted by LSTM

$Play_{t-4}$	$Play_{t-3}$	$Play_{t-2}$	$Play_{t-1}$	Probability*		N of prediction	N in data	Accuracy
LPD	LPD	LPD	GK ball catch	0.96	Feed	8	8	1.00
Square Pass	Ball control	Square pass	Backward pass	0.96	Ball control	7	7	1.00
GK ball catch	Feed	Ball control	Square pass	0.71	Ball control	6	6	1.00
	Throw-In	Backward pass	Ball control	0.50	Square pass	4	4	1.00
Square pass	Ball control	Forward pass	Dribbling in the wide area	0.53	Cross	4	4	1.00

*: $Probability\left(Play_t^k \mid X[Play_{t-4}, Play_{t-3}, Play_{t-2}, Play_{t-1}]\right)$

Table 8 Accuracy of the offence tactical play items predicted by LSTM

$Play_{t-4}$	$Play_{t-3}$	$Play_{t-2}$	$Play_{t-1}$	Probability*	$Play_t^k$	N of prediction	N in data	Accuracy
Ball control	Ball control	HPD	Block from HPD	**0.11** (0.22	**Shot in PA** HPD)	0	1	0.00
			Dribbling in wide area	**0.02** (0.16	**Shot outside of PA** Cross)	0	3	0.00
HPD	Tackle from HPD	Square pass	Dribbling in wide area	**0.05** (0.23	**Through ball to central area** Cross)	0	1	0.00
Feed	Ball control	Backward pass	Dribbling in wide area	**0.15** (0.36	**Early cross** Cross)	0	1	0.00
Square pass	Ball control	Through ball to wide area	Dribbling in wide area	**0.60** (0.60	**Cross** Cross)	1	1	1.00

*: $Probability\left(Play_t^k \mid X[Play_{t-4}, Play_{t-3}, Play_{t-2}, Play_{t-1}]\right)$
Bold shows probability and offence tactical items estimated by LSTM
( ) shows the highest probability of offence tactical items estimated by LSTM

4. Discussions

4.1. Development of tactical play items

Traditional performance analysis used events count, and had investigated KPIs to discriminate game results or successful offences. Tenga et al. (2009) constructed match performance items based on the literature and collected the data from video-images. Tenga et al. (2010a; 2010b) showed the effectiveness of items for match performance analysis by investigating the effect of the items on offence results (goal scoring and score-box entry). Wright et al. (2012) noted the importance of performance analysis based on coaching philosophy, but also the importance of analysis for the improvement of team performance. Hughes and Lovell (2019) analysed offensive transitions and found that ball gain by tackling and subsequent actions of dribbling or long passes were crucial for increasing goal scoring opportunities. However, past performance analysis research was very limited in the use of big data because these studies only analysed event count data, and the feature of tactical sequences from defence to offence had not been analysed.

Evolution in data technology and science have necessitated the development of items for the data-driven analysis of soccer games from tracking data and ball-touch data. Decoroos et al. (2019) designed features of action items in ball-touch data from various sports companies and developed a technique to estimate values of technical action by ML. This approach allowed evaluation of the player based on the contributions of the technical actions that they performed, instead of on the basis of shots or goals events. However, although their research represented a data-driven strategy, tactical sequences had not been analysed. This is why it was necessary to develop features for measuring tactical play items to analyse sequences from defence to offence in tactical performance.

In the present study, the offence and defence tactical play items were developed from ball-touch data and tracking data. This allows the analysis of sequential tactical performance from defence to offensive success. Defence tactical items allow the measurement of the degree of DFL press, which relates to measurement of HPD in relation to pushing up to the opponents' offensive zone. When the criterion of HPD showed a positive value of the DFL speed (>0.1 km/h), this meant that the DFL had moved in the direction of the opponents' goal.

Offence tactical items were constructed based on the ball positions. The items measured ball-related actions to progress and bring the ball to the best shot position. For example, forward passes, through ball to central area, dribbling in central area, and early cross were the tactical play items used for measuring offence progression and penetration.

Thus, the tactical play items for defence and attack are clearly measuring tactical play sequences from defence to attack, which was difficult to measure using count data of technical play in traditional analysis.

4.2. Examination of features for the deep leaning model

Probability of the next tactical play item's appearance in 48 items by LSTM-LM was estimated, and the prediction and accuracy of the next tactical play items after the tactical play sequence were analysed for DL of tactical play sequence from defence to offence.

The tactical play sequence with high prediction accuracy by LSTM-LM was considered to be due to the large frequencies of the relevant items in the data. (**Table 7**). For example, feed was predicted after the tactical play sequence X [LPD→ LPD→ LPD→ GK ball catch]. This is a series of tactical performance, which consisted of three LPD, moving to backward direction, GK ball catch, and then leading to feed by GK. In another example of a tactical play sequence with high prediction accuracy, the tactical play sequence X[GK ball catch→ Feed→ Ball control→ Square pass] can be interpreted to mean that the tactical play started with GK ball catch after which the GK fed the ball, then a player controlled the ball and passed in the square direction, leading to ball control. In another example, the tactical play sequence X [Square pass→ Ball control→ Square pass→ Backward pass] can be interpreted as indicating that the tactical play started with square pass and ball control. Then, the player passed in the square direction and controlled the ball before making a backward pass. Ball control is predicted after the tactical play sequences in this example. The three examples above are interpreted as tactical play where the GK caught the ball after LPD and fed the ball forward, with failure to break through the opponents leading to backward pass or square pass to keep the possession in their defence line. The frequency of square pass and ball control appearing on these tactical play sequences is high in the data, and this is considered to reflect the higher accuracy of learning results by LSTM.

On the other hand, lower prediction accuracy of offence tactical items (the offence results) such as shot in PA after tactical play sequences by the LSTM-LM was considered to be due to the small number of appearances of the tactical play items in the data. Tenga et al. (2010b) pointed out that goal scoring and shots were not appropriate events because of their very rare appearance, only around 1%, in the game.

In addition, it is suggested there were common tactical play items influencing the prediction results after the different types of tactical play sequences by LSTM-LM. In **Table 8**, the probability of a cross was the highest, at 0.60, and the previous play item ($\boldsymbol{Play_{t-1}}$) in the tactical play sequence was dribbling in wide area. Moreover, the probability of several of the offence play items (the shots outside of PA, through ball to central area, and early cross) was the highest after dribbling in wide area ($\boldsymbol{Play_{t-1}}$) in the tactical play sequences (**Table 8**). However, the cross was the actual predicted tactical item in the above tactical play sequences. For example, the tactical play sequence leading to the highest probability of early cross was the tactical play sequence X [Feed→ Ball control→ Backward pass→ Dribbling in wide area]. In this tactical play sequence, the offence play started with GK feeding the ball to the forward, controlling the ball, then a player passing backward, and finally a player dribbling in wide area. The probability of a cross was estimated to be 0.36, which was higher than the probability of the early cross (0.15). In the count of the previous play item before cross, the number of dribbling in wide area was 6 times (**Table 6**). This suggested that the LSTM-LM had learned that dribbling in wide area is a key play for leading to a cross.

The present study examined the learning results of the tactical play sequence from defence to offence using the LSTM-LM with 48 tactical items to measure tactical play. The tactical play sequence leading to the maximum probability of shot in PA is the tactical play sequence X [Ball control→ Ball control→ HPD→ Block from HPD] (**Table 8**). This tactical play sequence is interpreted as successful tactical performance from defence to offence which starts with quick ball gain by HPD, and where this tactic of pressing defence leads to the shot in PA. Similarly, the tactical play sequence leading to the maximum probability of through ball to central area also starts with the ball gain with HPD.

Vogelbein et al. (2014) found that successful teams (top 5 in final league table) in the German Bundesliga showed shorter defence reaction time to ball gain compared with the unsuccessful teams, suggesting that successful teams pressed high to the ball. Hughes and Lovell (2019) found that ball gain closer to the opponent's goal increased goal scoring opportunities. The tactical play sequence leading to the maximum probability of shot in PA or through ball to central area were also the offences that started from HPD. This result supports the findings of the previous studies, and it is suggested that the LSTM-LM successfully learned the successful tactical performance from high pressing defence in offences.

Therefore, it is clear that the defence and offence tactical play items processed from ball-touch data and tracking data for DL consist of natural items without any discrepancy.

4.3. Limitations.

In this study, it is assumed that the maximum length of tactical play sequences was four plays. In addition, 6,444 tactical play sequences extracted from the data in the final (1st leg and 2nd leg) of the knock-out stage for both teams were used for the analysis. It is necessary to consider whether it would be possible to generalize the findings in this study within the limitations imposed on the research in terms of data analysis methods and samples.

5. Conclusions

The purpose of this study was to develop offence and defence tactical play items for tactical play analysis using a DL model. For that purpose, this study constructed tactical play items for offence and defence in soccer games from ball-touch data and tracking data, applied the LSTM-LM, and analysed the items to measure tactical play for DL. The following conclusions were obtained.

The forty-eight items measuring defence and offence tactical play in a soccer game processed from ball-touch data and tracking data can be used as the features for tactical play analysis with DL, using the LSTM-LM.

Acknowledgements

This research was a collaborative research project based on an industry-academia collaborative research agreement between the University of Tsukuba and DataStadium Inc. We appreciate the permission to use soccer game tracking data and ball touch data measured by DataStadium Inc.

References

Decroos, T., Bransen, L., Van Haaren, J., and Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 1851-

1861.
Elman, J. L. (1990). Finding structure in time. Cogn. Sci., 14: 179-211.
Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P.R., and McRobert, A.P. (2016).Attacking and defensive styles of play in soccer: Analysis of Spanish and English elite teams. J. Sports. Sci., 34: 2195-2204.
Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., and Meyer, T.(2019). Machine learningin men's professional football: Current applications and future directions for improving attacking play. Int. J. Sports. Sci. Coach., 0: 1-20.
Hobbs, J., Power, P., Sha, L., Ruiz, H., Lucey, P. (2018). Quantifying the value of transitions in soccer via spatiotemporal trajectory clustering. In Proceedings of the 12th Annual MIT Sloan Sports Analytics Conference, Boston, MA, USA.
Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural. Comput., 9: 1735-1780.
Hughes, M., and Lovell, T.(2019). Transition to attack in elite soccer. J. Hum. Sport Exerc., 14: 236-253.
Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. J. R. Stat. Soc. C. Apple., 29: 119-127.
Kato, K. (2016). Data analysis and team reinforcement in soccer. IEICE Communications Society Magazine, 10: 29-34. (in Japanese).
Linke, D., Link, D., and Lames, M. (2020). Football-specific validity of TRACAB's optical video tracking systems. PLOS ONE 15: e0230179.
Liu, H., Hopkins, W., Gómez, A.M., and Molinuevo, S.J. (2013). Inter-operator reliability of live football match statistics from OPTA Sports Data. Int. J. Perf. Anal. Spor., 13: 803-821.
Lucey, P., Bialkowski, A., Monfort, M., Carr, P., and Matthews, I. (2014). Quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data. In Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference, Boston, MA, USA.
Lucchesi, M. (2004). Pressing. Spring City: Reedswain Inc..
Mikolov, T., Karafiát, M., Burget, L., Černocký, J., and Khudanpur, S. (2010). Recurrent neural network based language model. In Eleventh Annual Conference of the International Speech Communication Association.
Mikolov, T., Kombrink, S., Burget, L., Černocký, J., and Khudanpur, S. (2011). Extensions of recurrent neural network language model. In 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 5528-5531.
Olah, C. (2015). Understanding of LSTM Networks. Retrieved from https://colah.github.io/posts/2015-08-Understanding-LSTMs/. (accessed 2020-01-15).
Rathke, A. (2017). An examination of expected goals and shot efficiency 1 in soccer. J. Hum. Sport Exerc., 12: 514-529.
Sundermeyer, M., Schlüter, R., and Ney, H. (2012). LSTM neural networks for language modeling. In Thirteenth Annual Conference of the International Speech Communication Association.
Tenga, A., Kanstad, D., Ronglan, L. T., and Bahr, R. (2009). Developing a new method for team match performance analysis in professional soccer and testing its reliability. Int. J. Perf. Anal. Spor., 9: 8-25.
Tenga, A., Holme, I., Ronglan, L.T., and Bahr, R. (2010a). Effect of playing tactics on goal scoring in Norwegian professional soccer. J. Sports Sci., 28: 237-244.
Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010b). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. J. Sports Sci., 28: 245-255.
Vogelbein, M., Nopp, S., and Hökelmann, A. (2014). Defensive transition in soccer–are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. J. Sports Sci., 32: 1076-1083.
Wright, C., Atkins, S., and Jones, B. (2012). An analysis of elite coaches' engagement with performance analysis services (match, notational analysis and technique analysis). Int. J. Perf. Anal. Spor., 12: 436-451.
Yates, I., North, J., Ford, P., and Williams, M. (2006). A quantitative analysis of Italy's world cup performances: A comparison of world cup winners. Insight, 54-59.

Name:
Hiroki Matsuoka, MS

Affiliation:
Doctoral Program in Physical Education, Health and Sport Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574 Japan

Brief Biographical History:
2017- Doctoral Program in Physical Education, Health and Sport Sciences, University of Tsukuba

Main Works:
- Matsuoka, H., Tahara, Y., Ando, K., and Nishijima., T. (2020). Development of criterion-referenced measurement items of soccer defensive tactical play from tracking data. Football Science, 17: 29-40.

Membership in Learned Societies:
- Japanese Society of Science and Football
- Japan Society of Physical Education, Health and Sport Sciences

Appendix 1 Description of ball touch data

Game ID	Half	Game time (s)	Offence ID	Team	Player	Technical skill item	Ball position (x, y)		
1	1st Half	1	1	Team 1	Player 4	Tackle	x	y	← start
1	1st Half	3	1	Team 1	Player 8	Ball control	x	y	← + 1
1	1st Half	5	1	Team 1	Player 6	Through ball	x	y	← + 2
1	1st Half	8	1	Team 1	Player 10	Ball control	x	y	← + 3
1	1st Half	10	1	Team 1	Player 10	Shot	x	y	← + n
1	1st Half	20	2	Team 2	Player 1	Goal Kick	x	y	
						.			
						.			
2	2nd Half	2700	599	Team 1	Player 5	Clearance	x	y	

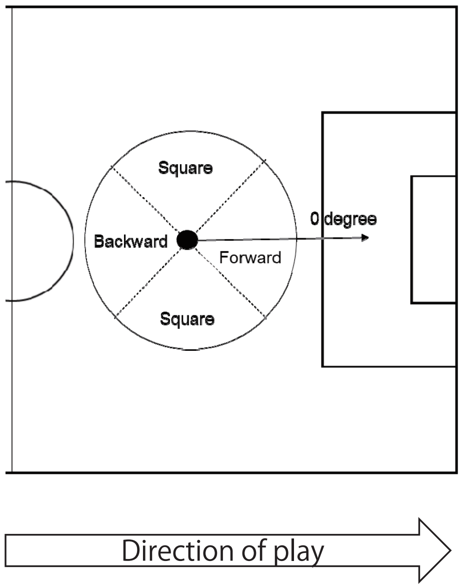
Appendix 2 Tactical skill items constructed from technical skill items

Game ID	Half	Game time (s)	Offence ID	Team	Player	Team 1 Technical skill item	Ball position (x, y)		DFL position (m)	DFL speed (km/h)	DFL press	Tactical skill item	
1	1st Half	1	1	Team 1	Player 4	Tackle	x	y	40.1	1.11	HPD	Tackle with HPD	← start
1	1st Half	3	1	Team 1	Player 8	Ball control	x	y	**n.a.	**n.a.	**n.a.	Ball control	← + 1
1	1st Half	5	1	Team 1	Player 6	Through ball	x	y	**n.a.	**n.a.	**n.a.	Through ball to central area	← + 2
1	1st Half	8	1	Team 1	Player 10	Ball control	x	y	**n.a.	**n.a.	**n.a.	Ball control	← + 3
1	1st Half	10	1	Team 1	Player 10	Shot	x	y	**n.a.	**n.a.	**n.a.	Shot in PA	← + n
1	1st Half	20	2	Team 1	*n.a.	*n.a.	*n.a.	*n.a.	38.0	-1.48	MPD	MPD	
						.							
						.							
2	2nd Half	2700	599	Team 1	Player 5	Clearance	x	y	25.5	-2. 6	LPD	**Clearance with LPD**	

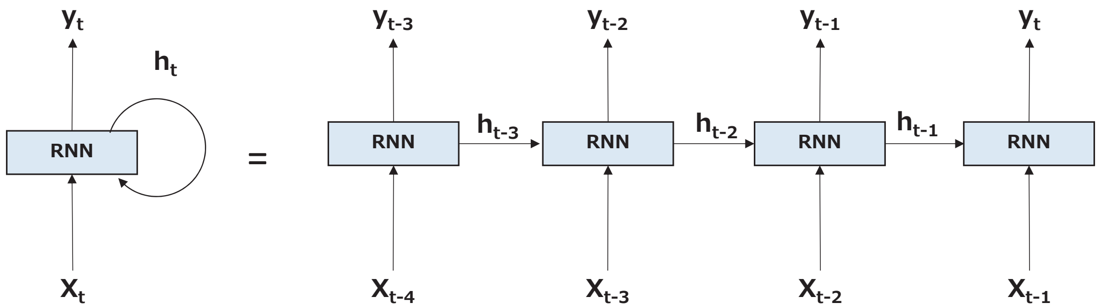
*n.a. shows data in the cells is not available for team 1 because team 2 plays offence technical action at that time points.
**n.a. shows data in the cells is not available for team 1 because team 1 plays offence technical action at that time points.

Appendix 3 Description of offence play data

Game ID	Half	Offence ID	Team	Offence play Start	+1	+2	+3	+n
1	1st Half	1	Team 1	Tackle with HPD	Ball control	Through ball to central area	Ball control	Shot in PA
1	1st Half	2	Team 1	MPD	MPD	n/a	n/a	n/a
		.						
		.						
		.						
2	2nd Half	599	Team 1	Clearance with LPD	n/a	n/a	n/a	n/a

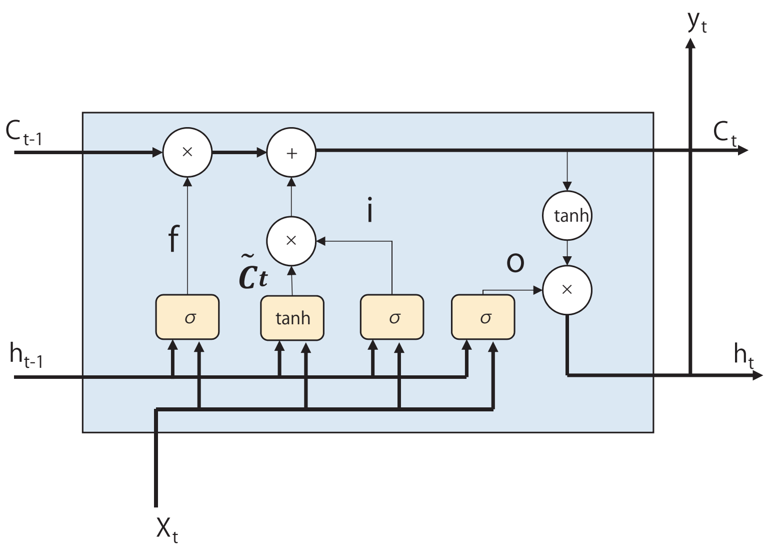


Appendix 4 Direction of pass



Appendix 5 Structure of Recurrent Neural Network (RNN)

RNN has a loop structure and passes the information to the next network (left side). In unrolled RNN (right side), when X is inputted to the network, y is outputted and, information is passed to the next step of RNN layer (ht). This structure allows the analysis of time-series data.



Appendix 6 Structure of LSTM block (Olah, 2015)

In LSTM block, several gates control information in Cell state (C). Cell state connects information to next LSTM block. Forget gate (f) decide to what information discard from cell state. Input gate layer (i) and tanh layer ($\tilde{C}_t$) are called input gate which decide to update information in cell state. Output gate (o) controls what information output from cell state.