All for Goals: a Stylized Automated Analysis Framework in Football Matches

Abstract—**Automated analysis in football matches is meaningful for player and team evaluation. However, most related works ignore match style and team strength. In this paper, a novel stylized automated analysis framework termed All for Goals (AFG) is proposed for football matches, which considers match style and team strength to better quantify the relationship of all match states and player actions respectively with potential goals. AFG is composed of an automatic labeling module, a potential goal prediction module, and a state and player evaluation module. Specifically, in the automatic labeling module, relevant samples are given the same label to avoid manual labeling. In the potential goal prediction module, we introduce the Pretrain-Finetune paradigm. Based on labeled data, an average model learning to identify scoring difficulty is obtained in the first pre-training procedure, and tuned models learning specific styles are obtained in the second fine-tuning procedure. In the state and player evaluation module, the evaluation mechanisms of state, on-ball action, and off-ball running based on potential goal prediction result are designed for match review and tactics mining. Finally, we validate the rationality and validity of AFG on multiple tasks. On the goal prediction task, the models show high recall rates and remarkable difference in style. On real-time situation analysis, credit assignment for football events, and off-ball running analysis tasks, the evaluation mechanisms give the results consistent with football domain knowledge.**

Index Terms—**situation analysis, situation evaluation, deep learning, transfer learning, football**

I. Introduction

Automated analysis is a crucial task in sports. It can provide players with suggestions for improvement, give coaches objective player evaluation and situation analysis, and improve the viewing experience of fans. However, the complexity and dynamics of sport matches pose huge challenges to researchers. The football match is more difficult to analyze because of its larger spatio-temporal range and number of players than most sports. Furthermore, the low score character makes it difficult to quantify the relationship of state and action respectively with goals.

Facing the difficulty of football match analysis, some works [1], [2] propose the physical models to analyze the situation of football matches, the results of which are limited. Compared with the above methods, the methods based on machine learning algorithms that directly mine knowledge from data to avoid match modeling are more effective. In this paradigm, match data are of great significance. Available data of football matches are divided into two categories: event data and tracking data. Event data record the information of key events and the players directly related to the events. Tracking data record the position of all players and the ball at each moment.

Currently, event data have been widely used in automated analysis in football matches. Some works use specific event data such as shot and pass to predict the probability of scoring by logistic regression [3], predict the probability of interception by physics-based modeling [4], and discern the tactical patterns [5]. However, these works can not directly quantify the impact of player actions on goals. Different from the above approaches, some models based on all event data evaluate all kinds of actions with long-term vision [6]–[8], find spatio-temporal patterns that characterize attacking tactics [9], and analysis the performance of players to rank them [10]. The performance of action evaluation can be further improved by selecting more suitable algorithms and constructing richer features, such as combining contextual information [11]–[13]. However, the inherent flaws of event data, i.e., the lack of information of players unrelated to events and the long time interval between two consecutive events, make models based on event data unable to give fine-grained results to evaluate the running of off-ball players, for example.

Tracking data make up the inherent flaws of event data. With available tracking data, some works analyze the spatio-temporal patterns of a short window before a shot [14], decouple the probability of a goal into the goal opportunity brought by pass, shot, and dribble [15], and model the transfer of the ball among offensive team's players besides direct shot of the ball holder to calculate the scoring probability [16].

However, due to technical reasons, high quality large-scale tracking data are difficult to obtain. The above works address the small volume problem of tracking data by combining models with knowledge, and simplifying tasks. In the machine learning paradigm, transfer learning is also an efficient approach to solve this problem. The Pretrain-Finetune paradigm of transfer learning has achieved great success in the field of image and natural language processing [17]–[19]. As for automated analysis in football matches, there is no suitable pre-trained model, hence a source domain related to the football matches in the real world is necessary. Google Research Football Environment (GRF) [20] as a game environment that simulates regular football matches in the real world is a potential source domain.

GRF provides a platform for researchers to easily train agents, the player in football game, particularly using reinforcement learning techniques to play football matches. Different agents have their own styles, as well as the teams in the real world. In addition to studying the transferability from game to reality, we can also study the impacts of match style and team strength on automated analysis in football matches basing on GRF. Match style and team strength significantly affect the analysis results. For example, the Premier League is fast-paced and focuses on physical confrontation, while Serie A focuses more on tactics and defense. As for team strength, for a given state, different teams will adopt different strategies and lead to different results. However, most related works neglect these factors.

Based on the above discussions, current works on automated football match analysis fail to provide fine-grained analysis results and neglect the impacts of match style and team strength. To fill this gap, we propose a novel stylized automated analysis framework in football matches, termed All for Goals (AFG), to study the relationship of match states and player actions with potential goals. Firstly, we define a potential goal prediction task. Then we build a mixed dataset, most of which comes from GRF and a small portion from the real world. We implement an automatic labeling module that endows the subsequent modules with long-term vision. Furthermore, we train a potential goal prediction module through a two-stage training procedure. In the first pre-training stage, an average model learning to identify the difficulty of scoring is obtained. In the second fine-tuning stage, a set of tuned models learning specific styles are obtained. Based on the potential goal prediction result, we design a state and player evaluation module for state, on-ball action and off-ball running evaluation. Finally, we validate the rationality and validity of AFG on multiple tasks. On the goal prediction task, the models show high recall rates and remarkable difference in style. On real-time situation analysis, credit assignment for football events, and off-ball running analysis tasks, the evaluation mechanisms give the results consistent with football domain knowledge.

The contribution of this work concludes:

(1) providing fine-grained data of 9000 standard football matches among different teams from GRF;

(2) proposing a stylized automated analysis framework AFG that contains an automatic labeling module, a potential goal prediction module, and a state and player evaluation module;

(3) demonstrating valuable applications of AFG, including the real-time situation analysis, credit assignment for football events, and off-ball running analysis, results of which are consistent with the football domain knowledge.

II. METHOD

In this part, we describe our technical details. First, we define a task for predicting potential goals (section II.A). Then we build a mixed dataset for the task and process data according to the task definition (section II.B). By introducing the Pretrain-Finetune paradigm, we obtain a set of stylized potential goal prediction models to complete the task (section II.C). Finally, the potential goal prediction result is used to build several fine-grained evaluation mechanisms (section II.D).

A. Task definition

Potential goal prediction, one of the most classic automated analysis tasks in football matches, has undergone a series of developments. Researchers study the probability of a potential goal resulting from a shot in a given state when the task is first proposed [21]. To give a result with short-term vision, researchers predict the probability of a goal in a fixed time scale or fixed number of on-ball actions from the given state. We expect to predict the potential goal with a long-term vision. However it is still difficult to excellent actions that are far away from the moment of scoring A good state can be recognized even though it requires the cooperation of a long action sequence to score. Therefore, the potential goal prediction with a long-term vision task is defined as:

Given: s_t, a feature combination to represent a match state at time t.

Do: Estimate probabilities

- $p(h|s_t, \Theta)$: the probability of home team scoring before the change of ball possession from the given state s_t
- $p(a|s_t, \Theta)$: the probability of away team scoring before the change of ball possession from the given state s_t
- $p(b|s_t, \Theta)$: the probability of none team scoring before the change of ball possession from the given state s_t

where h, a, b represent the three possible results of an offensiveness: home team scoring, away team scoring and none team scoring. It is a typical classification task. We use a machine learning algorithm, deep learning, to solve this task. Θ represents model parameters. In this task, the sample imbalance is severe. The samples with the ball possession change label are much more than those with the other two labels, which means the accuracy is meaningless. Therefore the optimization objective is to minimize the average cross-entropy loss of all labels:

$$\Theta^* = \underset{\Theta}{argmin} - \frac{1}{|l|} \sum_{l} \frac{1}{N_l} \sum_{i} y_{il} l_n p(l|s_i, \Theta) \tag{1}$$

where l represents the label, $|l|$ the number of labels, i the index of sample, N_l the number of the samples with label l, y_{il} a symbolic function whether the label of sample I is l, $p(l|s_i, \Theta)$ the probability of predicting the label of sample I as l.

B. Automatic labeling module

1) Data collection and description: The game data in our mixed dataset comes from GRF. This environment can simulate complex standard football matches (such as 11 vs.11) and simple academic scenarios (such as 3 vs. 1). We appoint an agent (trained by the reinforcement learning algorithm PPO; more details are beyond this paper) to play with easy baseline, hard baseline, and self for the 11 vs.11 scenario. We respectively collect 3000 games for each team combination. The environment returns raw observation every 1.8 seconds in a game of 90 minutes. Hence, we finally collected 27 million samples. The raw observation is composed of rich features, including:

- `player information`: the two-dimensional position, two-dimensional speed, fatigue value, yellow card, and role of all players in the game.

- `ball information`: the three-dimensional position, three-dimensional speed, three-dimensional rotation speed, and ball possession which indicates the team and player controlling the ball.
- `game information`: the score, remaining time, and game mode which includes normal, kickoff, goal-kick, free-kick, corner, throw-in, and penalty.

The real-world data in the mixed dataset come from ten matches of the 2018 season of Chinese Super League. The tracking data record the two-dimensional positions of players, referees, and the ball on the pitch at a frequency of ten frames per second. Hence, we finally collected 540 thousand samples. The event data are used to supplement tracking data for extra information such as ball possession and information. The event data record the time, type, participant, result of events.

2) labeling approach: As shown in Figure 1, we divide a football match into sequential offensives. Each offensive is a ball-possession stream, in which a team creates goal opportunities through cooperation, mainly embodied in passing among team members, and dribbling by the ball holder. A stream ends and the next stream starts until a score or ball possession changing event occurs.

We give all frames in a ball-possession stream the same label, which depends on the result of the stream. If the result is a home team's score, the label will be 1. If the result is an away team's score, the label will be -1. Otherwise, the label will be 0. It seems that the labeling rule is equivalent to evaluating each match state in a stream as the same value. However, the value of a state lies in the distribution of its labels but not the label of a specific sample. In other words, a valuable state for the home team will frequently appear in the streams with label 1, where a bad state will only occasionally appear.

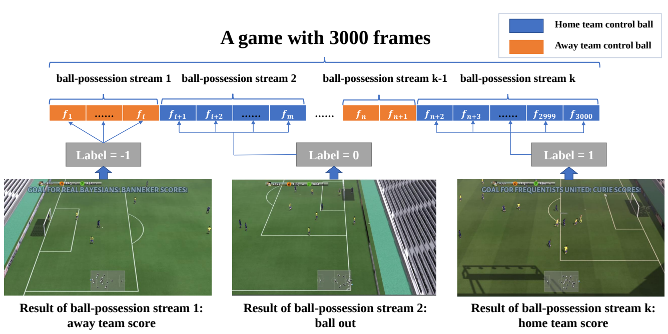


Fig. 1. A football match is divided into sequential ball-possession streams, and the labels of the frames in a ball-possession stream are the same.

C. Potential goal prediction module

Match style and team strength, as essential factors affecting the results of automated football match analysis, are always ignored by most existing works. We address this challenge by introducing the Pretrain-Finetune paradigm. In the pre-training procedure, a model learns to identify the difficulty of scoring in given state. Although the scoring difficulty for the same state differs from diverse stylized matches, there still exist similarities. For example, an unguarded shot is always easier to score than a guarded shot. In the fine-tuning procedure, a set of models emerge with different styles under the premise of basically keeping the scoring difficulty unchanged. In conclusion, the paradigm is reasonable and suitable for the similarity of the scoring difficulty in different stylized matches. In addition, the effect of the paradigm is to correct the difference of different stylized matches.

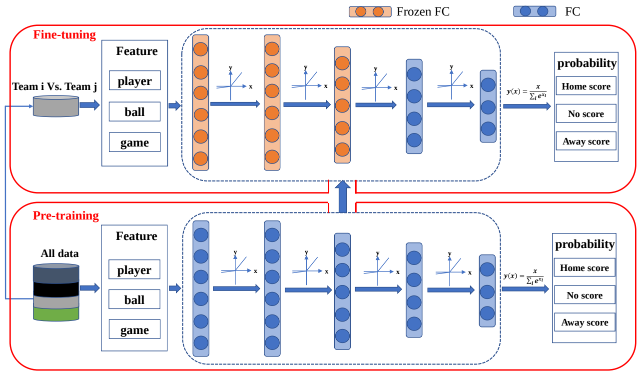


Fig. 2. The training of the potential goal prediction module is divided into two stages. In the pre-training stage, based on all data, an average model learning to identify the difficulty of scoring is obtained. In the fine-tuning stage, based on the specific style data, a set of stylized models learning specific styles are obtained.

Figure 2 shows the two-stage training procedures of the potential goal prediction module. The input and output of the models in this module have been introduced in section II.A. The network structure contains five fully connected layers, the activation functions of the first four layers are LeakyReLu, and the last is softmax. We force the models respectively to learn the scoring difficulty and match style by controlling the composition of the dataset. In the pre-training stage, the average model is driven by all stylized data to learn the commonalities of them. In the fine-tuning stage, the stylized models inherit the commonalities from the average model, i.e., the model parameters, and sequentially learn various styles of the specific data. Particularly, the parameters of low-level feature extraction layer are frozen when fine-tuning to reduce overfitting.

D. State and player evaluation module

In this part, potential goal prediction result is used to build a series of evaluation mechanisms, including evaluations of state value, on-ball action value, and off-ball running value for match review and tactics mining.

1) State evaluation: State evaluation can be represented by many indicators, such as the control degree of two teams on an area [22] or additionally accounting for the probability of the ball moving to the area [23]. For state evaluation, AFG converts match states to be directly related to potential goals, so in this paper, state value is the difference in the probability of two teams scoring a goal before the change of ball possession:

$$V_{state}^{i}(s_t) = p(i|s_t, \Theta_{ij}) - p(j|s_t, \Theta_{ij}) \tag{2}$$

where $V_{state}^{i}(s_t)$ represents the state value of team i in the given state s_t at time t, while $p(i|s_t, \Theta_{ij})$ and $p(j|s_t, \Theta_{ij})$ are the outputs given by Θ_{ij}(section II.C) as the input is s_t.

2) On-ball action evaluation: We define the on-ball action value of a player as:

$$V_{on-ball}^{i,k}(a_t) = V_{state}^{i}(s_{t+1}) - V_{state}^{i}(s'_{t+1}) \tag{3}$$

where $V_{on-ball}^{i,k}(a_t)$ represents the value of action a_t performed by the on-ball player k of team i at time t, s_{t+1} the match state at time $t+1$, and s'_{t+1} the match state assuming that the on-ball player k performs none action, while the other players still take the actions according to their strategies at time t. Equation (3) means that the value of player's action on the ball is equal to the change in state value whether the action is performed.

There are two advantages of the on-ball action evaluation mechanism in AFG: (1) Our model is so sensitive that it can quickly identify and significantly show the difference in the state value of two consecutive frames if any on-ball action is performed, such as pass, shot, and duel. Precisely because AFG only needs the state of two consecutive frames, it will not be affected by other on-ball actions when evaluating an on-ball action; (2) Strictly speaking, most models for action evaluation are not to evaluate actions but an event, giving just one value for the whole event. A special feature of AFG is that it can evaluate the actions of all participants in an event to analyze the contribution of everyone (more details in section IV.B).

3) Off-ball running evaluation: In football, team cooperation is essential, mainly reflected among the ball holder and off-ball players. The roles of off-ball players include (1) dragging the opponents to reduce the offensive pressure of the ball holder or pressure the opponents; (2) occupying a good position to receive the ball. Off-ball players play an indispensable role in most goals. However, at present, most researches devote to evaluating the value of the ball holder, ignoring the contribution of off-ball players.

We study the off-ball player's value from fatigue value and position. We find that the perspective of fatigue value is not practical for two reasons: (1) the fatigue value may not limit the player's continuous running and sprinting for the limitation of this game environment; (2) the agent's strategy that we use to collect data fails to learn to change its strategy according to its fatigue value. Fortunately, evaluating the value of off-ball players from the perspective of position performs well. The running value of an off-ball player's running action at time t is:

$$V_{off-ball}^{i,k}(p) = V_{state}^{i}(s_{t+1}) - V_{state}^{i}(s'_{t+1}) \tag{4}$$

where $V_{off-ball}^{i,k}(p)$ represents the state value of team i after the player k running to point p by perform action a_t relative to the state value that player k does not move.

It does not make sense to find the best position on the pitch for an off-ball player that makes the state value the most, owing to the dynamic character of football matches. Even if the best position is found, when the off-ball player runs to that position, the match state has changed a lot, as well as the best position. Therefore, the area near the player is more meaningful for consideration. We define a player's reachable region $R_{\Delta t}$ as the aggregation of points he can reach in Δt. Then we can find the maximum and minimum value in the $R_{\Delta t}$, $\underset{p\epsilon R_{\Delta t}}{max} V_{off-ball}^{i,k}(p)$ and $\underset{p\epsilon R_{\Delta t}}{min} V_{off-ball}^{i,k}(p)$. Furthermore, the range of the value of an off-ball player's running varies dramatically with his position. Hence, we normalize the value of the off-ball running action to calculate the relative value of every position within $R_{\Delta t}$:

$$\overline{V_{off-ball}^{i,k}(p)} = \frac{V_{off-ball}^{i,k}(p) - \underset{p\epsilon R_{\Delta t}}{min} V_{off-ball}^{i,k}(p)}{\underset{p\epsilon R_{\Delta t}}{max} V_{off-ball}^{i,k}(p) - \underset{p\epsilon R_{\Delta t}}{min} V_{off-ball}^{i,k}(p)} \tag{5}$$

III. EXPERIMENTAL ANALYSIS OF POTENTIAL GOAL PREDICTION TASK

We design multiple sets of experiments for answering the following questions by comparing and analyzing the experimental results to illustrate the rationality and validity of the models in AFG.

Q1: Whether the premise of using Pretrain-Finetune paradigm is satisfied in our task?

Q2: Whether Pretrain-Finetune paradigm helps models to complete potential goal prediction task better?

Q3: Whether Pretrain-Finetune paradigm additionally helps models stylized?

Q4: How to identify the strength of different teams?

A. Experiment descriptions and results

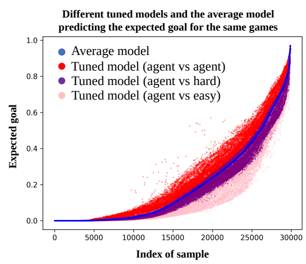
The experiment results are as shown in Table I and Figure 3. In Table I, "Easy (300)" represents 300 games of agent vs. easy, "Mixed (7200)" is 7200 games of agent vs. agent, agent vs.easy and agent vs. hard, "Real-world (8)" is 8 matches from the real world, while "Recall i" is the recall rate of label i.

B. Q1:Whether the premise of using Pretrain-Finetune paradigm is satisfied in our task?

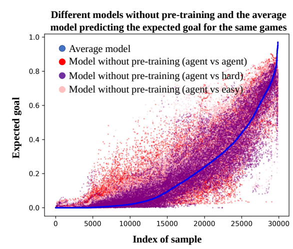
Pretrain-Finetune paradigm requires the correlation of source domain and target domain. In the potential goal prediction task, the matches played by different teams are different domains. In Figure 3(d), the predicted values of Model 6 and 7 basically increase as the values given by Model 5 increase. That means the different models trained on the data from different domains have a positive correlation in predicting potential goal task. In Figure 3(b), the prediction results of three models without pre-training are basically consistent with the results of Model 1, which means Model 1 has learned the correlation.

TABLE I
RECALL RATES ON TEST SET OF DIFFERENT MODELS FOR DIFFERENT STYLIZED MATCHES.

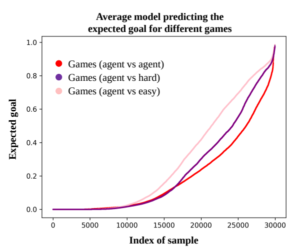
Model Number	Pre-trained Model	Train Set	Test Set	Recall 0	Recall 1	Recall 2
Model 1	None	Mixed (7200)	Mixed (900)	68.0%	71.9%	75.3%
Model 1	None	Mixed (7200)	Agent (300)	61.7%	83.0%	60.4%
Model 1	None	Mixed (7200)	Easy (300)	70.3%	51.6%	80.5%
Model 1	None	Mixed (7200)	Hard (300)	72.1%	59.4%	77.3%
Model 2	Model 1	Agent (50)	Agent (150)	68.9%	69.5%	66.6%
Model 3	Model 1	Easy (50)	Easy (150)	69.9%	64.0%	73.2%
Model 4	Model 1	Hard (50)	Hard (150)	67.4%	70.6%	74.6%
Model 5	None	Agent (2400)	Agent (300)	67.9%	70.0%	68.5%
Model 6	None	Easy (2400)	Easy (300)	66.6%	72.5%	66.8%
Model 7	None	Hard (2400)	Hard (300)	69.7%	68.7%	63.8%
Model 8	None	Real-world (8)	Real-world (1)	87.1%	46.7%	50.5%
Model 9	Model 5	Real-world (8)	Real-world (1)	82.1%	60.7%	62.9%



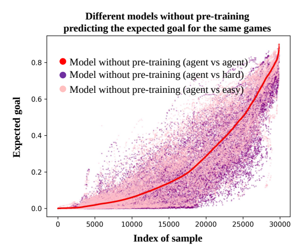
(a) We use Model 1, 2, 3 and 4 to predict potential goals for 30000 frames of 10 “agent v.s. agent” games, then sort the predicting results in size given by Model 1.



(b) We use Model 1, 5, 6 and 7 to predict potential goals for 30000 frames of 10 “agent v.s. agent” games, then sort the predicting results in size given by Model 1.



(c) We use Model 1 to predict potential goals respectively for 30000 frames of 10 “agent v.s. agent” games, 10 “agent v.s. easy” games and 10 “agent v.s. hard” games, then respectively sort the predicting results in size given by Model 1.



(d) We use Model 5, 6 and 7 to predict potential goals for 30000 frames of 10 “agent v.s. agent” games, then sort the predicting results in size given by Model 5.

Fig. 3. Potential goal prediction results of different models for games of different team combinations

C. Q2:Whether Pretrain-Finetune paradigm helps models to complete potential goal prediction task better?

For game environment, on the one hand, it can be seen in Table I that Model 1 is trained to be balanced among three recall rates for mixed data. However, for a specific team combination, the recall rates of Model 1 are very uneven before fine-tuning, which is caused by failing to take into account the strength of teams. The fine-tuned models 2, 3 and 4 are more balanced. On the other hand, compared with the directly trained models 5, 6 and 7, the tuned models with pre-training improve training efficiency that requires less data and training epoch to reach the same performance.

For real-world matches, the tuned model 9 performs better than directly trained model 8 on small sample real-world data by inheriting what it has learned from game data, which means similarity does exist between games and reality in our task and the similarity helps improve the performance of model on the real-world matches.

D. Q3:Whether Pretrain-Finetune paradigm additionally helps the models stylized?

Figure 3(a) and 3(b) visualize the results of potential goal prediction for ten games with different models. Compared with the latter, the former shows the styles of models by clustering their predicting results. Specifically, fine-tuned models obtain a nonlinear weight to fine-tune the original values given by the average model 1 while maintaining the typical consistency.

E. Q4:How to identify the strength of different teams?

Figure 3(c) shows the results of three team combinations: “agent vs. agent”, “agent vs. easy” and “agent vs. hard” respectively predicted the home team’s potential goal by the average model. It can be seen that when the agent is playing against the easy baseline, the potential goal values are the maximum, while the values are the minimum when playing against the agent. So the strongest team among the three teams is agent, and the weakest is the easy baseline.

IV. APPLICATIONS

In this section, we illustrate a series of valuable applications of the potential goal prediction models (sec 4.1) and the evaluation mechanisms (sec 4.2, sec 4.3) and test AFG in real-world (sec 4.4). In particular, the evaluation mechanisms are different from the potential goal prediction in lacking objective evaluation indicators. Therefore, we compare the results given by evaluation mechanisms with domain knowledge to prove their reasonableness and validity.

A. Real-time situation analysis

A typical real-time situation analysis is to predict the winner of the game in real-time.However, this prediction is coarse-grained. It is only affected by critical events, such as goals. We focus on a more fine-grained situation, i.e., goals. For a football match, AFG can give a real-time potential goal prediction diagram. We show the potential goal predictions of the home team and the away team changing from the start of a match to the 108^{th} second as an example in Figure 4. At the beginning of the game, the home team holds the ball. As players run on the pitch, the potential goals of both sides change slightly. However, once an on-ball action occurs, AFG will immediately recognize and reflect the significant change of the potential goals. The second pass of the home team is interrupted by the away team. At that time, the player closest to the ball of the home team tries to grab the ball. However, as the ball holder passes the ball, the home player loses the opportunity to obtain the ball possession. As a result, the potential goal of the home team is cleared. This interception event also moves the game from ball-possession stream 1 to ball-possession stream 2. A player of the away team then passes the ball with head.

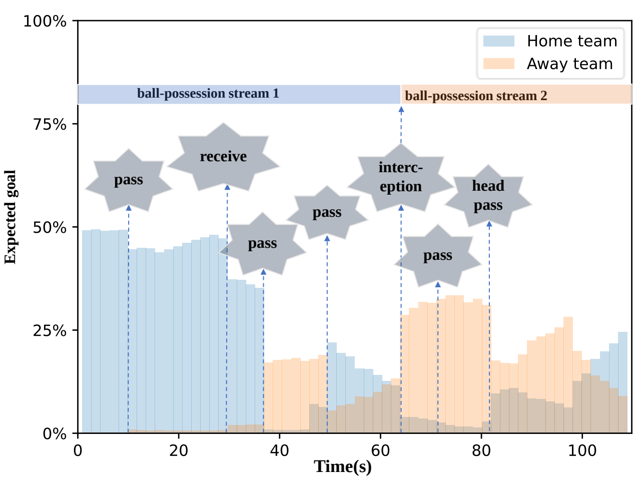


Fig. 4. The potential goals of the home team and the away team change with the events occurred in a football match.

B. Credit assignment for football events

Credit assignment for football events is to identify the contribution of every participant to events. In football matches, an event often has a long time span. AFG can complete the credit assignment, because an action has been evaluated at the moment when it is executed. We experiment on a pass event in a game, showing how actions during the event affect the state value of the away team in Figure 5. This passing event is divided into three stages. The first kicking stage mainly reflects the action value of the ball holder; the running stage reflects the action value of off-ball players; the final duel stage reflects the action value of several players close to the ball. The away team's state value changes from 0.01 to 0.26 caused by the ball holder kicking the ball out, which means the pass action value of this player is 0.25. AFG scored a high value for this action, which aligns with our intuition. If his teammate successfully receives the ball, the teammate could shoot with the pressure of only a back and a goalkeeper. In the running stage, the state value gradually changes from 0.26 to 0.47, which is the effect of the running action of all players on the pitch. In the figure, the color of the trajectory of the two receivers represents their running value. Although their running directions are both correct, the running value of the real receiver is relatively lower at the beginning of the event because he starts running slowly. In the second half of the running, both receivers have high running values. More analysis of running will be discussed in the next part. In the final stage, the duel action of the ball holder's teammate is evaluated as -0.54 because he loses to his opponent. In summary, although the real receiver intercepts this passing, AFG still gives a high positive evaluation to the ball sender and finds the target receiver of as the player responsible for this failure.

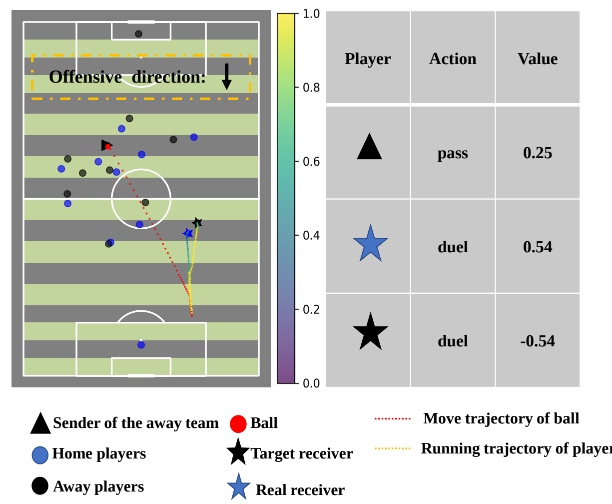


Fig. 5. The sender of the away team passes the ball to the target receiver, however the real receiver of the home team intercepts the passing.

C. Off-ball running analysis

In a football match, players are always in off-ball state, so off-ball running analysis is crucial. AGF not only evaluates off-ball running actions, but also gives the best running direction. We experiment on an event that an off-ball player of the home team intercepts a pass. Figure 6 shows the running analysis result of the off-ball home player. From the given state, the most likely strategy of the away team is: the player closest to the goal keeps running to get rid of the opponent, while the ball holder finds an opportunity to pass the ball to the player closest to the goal and then shoots. Hence, the best strategy for the analyzed home player is to intercept the pass of the away team. The color of each point of the reachable region in the bottom right subfigure in Figure 6 represents the normalized value of the player running to that point. Therefore, the gradient direction in the analysis graph is the best direction given by AFG. In addition, the value of off-ball running is not only related to the direction but also the speed of the player. It can be seen that the faster the player is, the larger the relative value of his running in the current state. The best direction

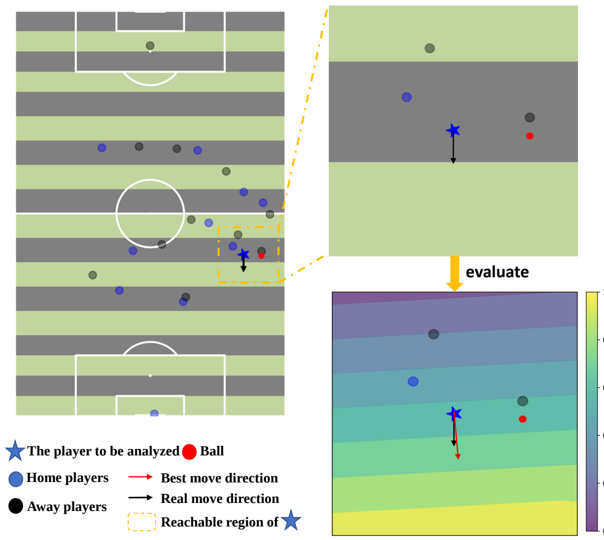


Fig. 6. A player of the home team blocked the ball holder of the away team from passing the ball to another player of the away team.

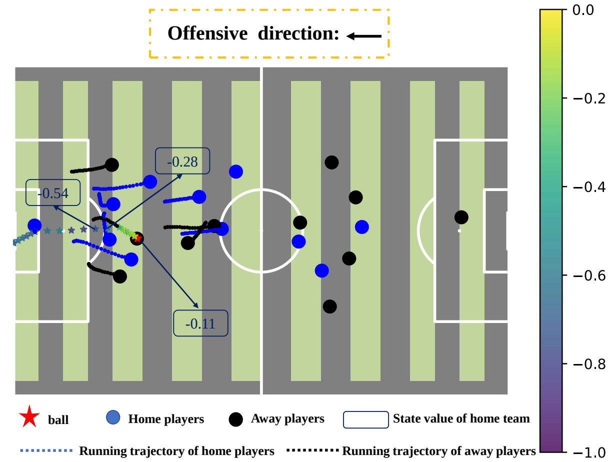


Fig. 7. A shot results in a goal. We visualize the trajectory of key players during the entire shot event (fixed color) and the trajectory of the ball (gradient color, which the color represents the state value while the ball moves to that point. We mark the specific state value of three key moments)

given by our framework is approximately perpendicular to the passing path between two away players. The reason that it is not precisely perpendicular is that the football game is dynamic, so the evaluation is made taking into account the actions of others.

D. Test on the real-world data

In this part, we verified the universality of AFG on a match of the 2018 season of Chinese Super League. Figure 7 shows the change in state value of the home team given by the state evaluation mechanism in AFG over a shot. Before the shot event, due to the blocking of players of the home team, the state value is -0.11. As the ball holder of the away team runs to the goal with the ball, the state value is reduced to -0.28. Then the ball holder performs a shot action reducing the state value to -0.54, which means it is a good action. Finally, although the goalkeeper touches the ball and changes the direction of the movement of the ball, he fails to prevent the scoring.

V. CONCLUSION AND FURTHER WORK

In this work, we provide a dataset of 9000 football games among different teams from GRF and propose a novel stylized automated analysis framework AFG for football matches, which firstly accounts for match style and team strength when quantifying the relationship of all match states and player actions with potential goals. We prove its rationality, validity, and application value. In addition, We believe that AFG is equally applicable to other sports because we do not introduce any knowledge in football as a constraint of AFG except the automatic labeling module.

Our framework still has some limitations for further optimization. In the future, we will further study transferring AFG from the game environment to the real world and more promising applications of AFG.

ACKNOWLEDGMENT

REFERENCES

[1] T. Taki and J.-i. Hasegawa, "Visualization of dominant region in team games and its application to teamwork analysis," in *Proceedings computer graphics international 2000*. IEEE, 2000, pp. 227–235.

[2] J. Fernandez and L. Bornn, "Wide open spaces: A statistical technique for measuring space creation in professional soccer," in *MIT Sloan Sports Analytics Conference 2018*, 2018.

[3] S. Gregory. (2019) Expected goals in context. [Online]. Available: https://www.statsperform.com/resource/expected-goals-in-context/

[4] W. Spearman, A. Basye, G. Dick, R. Hotovy, and P. Pop, "Physics-based modeling of pass probabilities in soccer," in *Proceeding of the 11th MIT Sloan Sports Analytics Conference*, 2017.

[5] Q. Wang, H. Zhu, W. Hu, Z. Shen, and Y. Yao, "Discerning tactical patterns for professional soccer teams: an enhanced topic model with applications," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 2197–2206.

[6] K. D. Routley, "A markov game model for valuing player actions in ice hockey," Ph.D. dissertation, Applied Sciences:, 2015.

[7] G. Liu, Y. Luo, O. Schulte, and T. Kharrat, "Deep soccer analytics: learning an action-value function for evaluating soccer players," *Data Mining and Knowledge Discovery*, vol. 34, no. 5, pp. 1531–1559, 2020.

[8] L. Bransen and J. Van Haaren, "Measuring football players' on-the-ball contributions from passes during games," in *International workshop on machine learning and data mining for sports analytics*. Springer, 2018, pp. 3–15.

[9] T. Decroos, J. Van Haaren, and J. Davis, "Automatic discovery of tactics in spatio-temporal soccer match data," in *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, 2018, pp. 223–232.

[10] L. Pappalardo, P. Cintia, P. Ferragina, E. Massucco, D. Pedreschi, and F. Giannotti, "Playerank: data-driven performance evaluation and player ranking in soccer via a machine learning approach," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 5, pp. 1–27, 2019.

[11] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis, "Vaep: An objective approach to valuing on-the-ball actions in soccer," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. International Joint Conferences on Artificial Intelligence Organization, 2020, pp. 4696–4700.

[12] ——, "Actions speak louder than goals: Valuing player actions in soccer," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1851–1861.

[13] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system in proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,(pp. 785–794)," *New York, NY, USA: ACM*, vol. 10, no. 2939672.2939785, 2016.
[14] P. Lucey, A. Bialkowski, M. Monfort, P. Carr, and I. Matthews, "quality vs quantity: Improved shot prediction in soccer using strategic features from spatiotemporal data," in *Proc. 8th annual mit sloan sports analytics conference*, 2014, pp. 1–9.
[15] J. Fernández, L. Bornn, and D. Cervone, "Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer," in *13th MIT Sloan Sports Analytics Conference*, 2019.
[16] W. Spearman, "Beyond expected goals," in *Proceedings of the 12th MIT sloan sports analytics conference*, 2018, pp. 1–17.
[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[20] K. Kurach, A. Raichuk, P. Stańczyk, M. Zajac, O. Bachem, L. Espeholt, C. Riquelme, D. Vincent, M. Michalski, O. Bousquet *et al.*, "Google research football: A novel reinforcement learning environment," *arXiv preprint arXiv:1907.11180*, 2019.
[21] S. Green. (2012) Accessing the performance of premier league goal scores. [Online]. Available: https://www.statsperform.com/resource/assessing-the-performance-of-premier-league-goalscorers/
[22] A. Fujimura and K. Sugihara, "Geometric analysis and quantitative evaluation of sport teamwork," *Systems and Computers in Japan*, vol. 36, no. 6, pp. 49–58, 2005.
[23] J. Fernández and L. Bornn, "Soccermap: A deep learning architecture for visually-interpretable analysis in soccer," *Lecture Notes in Computer Science*, p. 491–506, 2021. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-67670-4$_3$0