Masked Modeling-based Action Event Prediction Considering Bidirectional Time-series in Soccer

Ryota Goka*, Yuya Moroto*, Keisuke Maeda†, Takahiro Ogawa†, Huang-Chia Shih‡, and Miki Haseyama†
*Graduate School of Information Science and Technology, Hokkaido University, Japan
†Faculty of Information Science and Technology, Hokkaido University, Japan
‡Department of Electrical Engineering, Yuan Ze University, Taiwan
E-mail: {goka, moroto, maeda, ogawa, mhaseyama}@lmd.ist.hokudai.ac.jp
hcshih@saturn.yzu.edu.tw

Abstract—This paper presents an action event prediction based on masked modeling for bidirectional time-series analysis in soccer. Since optimal action events should be selected based on changes in match situations, it is important to consider bidirectional time-series changes in data. To predict the next action event with the consideration of the bidirectional time-series, the proposed method learns the contexts of action event sequences by predicting the masked action events from the preceding and following contexts. The prediction accuracy of our method is improved from that of the unidirectional method, which shows the effectiveness of taking the bidirectional time series into account in soccer.

Index Terms—Sports analytics, soccer, action event prediction, masked modeling, bidirectional time-series analysis.

I. INTRODUCTION

With the development of information processing and measurement technologies, various data have been collected in the sports field, and it has become common to perform tactical analysis and player evaluation from a scientific viewpoint with such data. Specifically in soccer, researchers have proposed the quantification method for the effectiveness of each player's decision-making and team strategy by predicting the occurrence probability of the next or subsequent important action event based on the current match situation [1]. Such event prediction-based quantification methods succeed in evaluating individual actions, even though conventional methods with match stats including ball possession and the number of shots on goal have difficulty with the individual evaluation. Since soccer players select action events based on subsequent changes in match situations, consideration of bidirectional time-series changes in data is significant. However, previous studies were limited to the consideration of only unidirectional time-series from the past to the present, and therefore there is room for improving prediction accuracy by introducing the bidirectional time-series relations.

In this paper, to consider of bidirectional time-series relations in soccer, we propose a novel action event prediction method based on masked modeling, inspired by BERT [2]. Concretely, by predicting the masked action events from the context before and after, our method enables to predict action events based on bidirectional time series relations.

II. PROPOSED METHOD

An overview of the proposed method is shown in Fig. 1. The proposed method treats players' action events as words, such as "Pass" and "Dribble," and acquires token embeddings by inputting these words into the embedding layer. In addition, by applying the full connected (FC) layer to the information, such as action event location coordinates and match time, that can identify the detailed match situation at the time when the action event occurs, we obtain continuous features and concatenate them to token embeddings. The concatenated embeddings are input to the transformer encoder along with the positional embeddings. Here, in the training phase, our method probabilistically replaces action events with masks when inputting such embeddings and predicts the replaced action events from the contexts. In the inference phase, by regarding the last action in the input sequence as a mask, our method enables the prediction
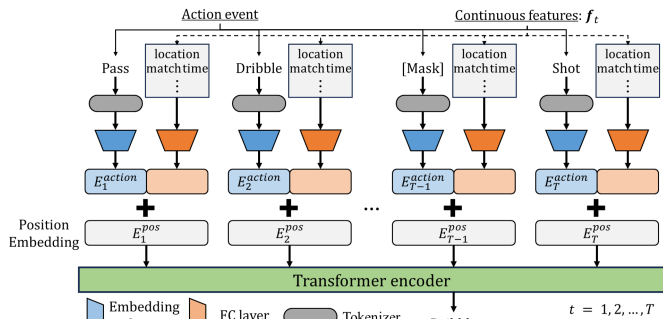
Fig. 1: The overview of our prediction method. Our method learns the contexts of action event sequences by predicting the masked action events from the context before and after.

TABLE I: Confusion matrix for action events predicted by our method.

Actual action event	Predicted action event			
	Pass	Dribble	Cross	Shot
Pass	973	107	7	5
Dribble	9	1189	0	6
Cross	30	4	32	0
Shot	17	7	0	16

TABLE II: Evaluation results for prediction performance.

Methods	Ours	Seq2Event [1]
Accuracy	**0.920**	0.914

of the next action event. In this way, our method learns the contexts of action event sequences based on the masked modeling framework to predict the next action event with the consideration of the bidirectional time-series analysis.

III. RESULTS AND CONCLUSIONS

In this experiment, we used 12,029 attack scenes consisting of 10 action event sequences taken from 146 matches in the StatsBomb Open Data[1]. Note that we limited to particularly important four action events ("Pass", "Dribble", "Cross", "Shot"). As a comparison method, we adopted transformer-based Seq2Event [1], which is a state-of-the-art action event prediction model in soccer. In addition, to measure the prediction performance, we used Accuracy as an evaluation metric.

Tables I and II show the confusion matrix for action events predicted by our method and the Micro-F1 for the prediction performance, respectively. From these results, we confirm that our method can predict the action event more accurately than that of Seq2Event, which shows the effectiveness of the bidirectional time-series analysis in soccer.

REFERENCES

[1] Ian Simpson, Ryan J Beal, Duncan Locke, and Timothy J Norman, "Seq2event: Learning the language of soccer using transformer-based match event prediction," in *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 2022, pp. 3898–3908.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171–4186.

[1] https://statsbomb.com/what-we-do/hub/free-data/#data