Formation2Vec: Exploring a Representation for Formation Segmentation and Detection in Soccer Games

Sarit Kraus
Department of Computer Science, Bar-Ilan University, Israel
pcchair@ijcai19.org

Abstract

In the soccer game, formation depicts the role of the soccer players on the pitch, revealing the basic tactic of the teams. During a game, a team will switch between offensive and defensive formations as they get or lose the ball. Although the analysis of formation changing is important in soccer, there are few methods to detect and classify the formations automatically. In this paper, we analogize the formation detection to the human action segmentation and classification problem. Referring to the general framework to solve such a problem, we firstly absorb a neighborhood awareness method to model the positional data and obtain the topological structure of the players, which can help represent the position. Our experiments prove that the representation outperforms the alternative representations, e.g. player positions and parameterized pitch. Then by a convolutional network, the temporal patterns will be captured, enabling us to segment the game into the periods with specific formations. The accuracy of the method achieves a satisfactory level in action segmentation and classification area. Compared with the existing methods in formation detection, our method supports fine-grained analysis of formation change and shows a wider application prospect.

1 Introduction

Soccer is the most popular sports in the world with huge commercial value. This 11-a-side competition shows a high degree of confrontation and dynamic. The players on the field (except goalkeeper) are organized by the term formation, which is arranged by the coach considering the characteristics of the players and the situation of the game. The formation of a team is described using the form n_b-n_m-n_f, which means that n_b backwards, n_m midfielders and n_f forwards (or n_b-n_{dm}-n_{om}-n_f, which means n_b backwards, n_{dm} defensive midfielders, n_{om} offensive midfielders and n_f forwards). For example, Barcelona from Spanish La Liga usually uses 4-3-3 formation this season and Liverpool from English Premier League prefers 4-2-3-1. Actually during a game, the formation of a team is not fixed. As the team obtains the ball, the players will move forward and the number of people in the frontcourt will increase; when the opponents attack and approach the goal, the team will adopt a deep defensive formation, like 6-3-1, to enhance the defense.

In this paper, we try to solve the problem of formation detection with positional data of the players. To be more specifically, we explored a new method to automatically segment the game into periods, each with a unique label indicating the formation of the team (including no formation). This problem is somehow similar to human action segmentation and detection, which generally extract the features from the video frame by frame and then learn the temporal patterns by training classifiers[Lea et al., 2017]. Different from the human action segmentation and detection problem, the formtaiton detection poses several challenges. First, the representation of positional data that best fit for formation detection is not clear. The original data format is the x, y coordinates of the players projected to a 1050×680 pixel 2D soccer field. To improve the accuracy of the classification, a more high-level feature representation that captures the topological characteristics of the positional data is needed. Second, the context information is crucial to detect a formation. At the time of formation transformation, the distribution of the players in a single frame may look like some other formation. So the model is supposed to capture the context feature of the data and segment the periods of time with stable formation. Third, the data labelling needs to be done by professional soccer analysts. Unlike human action, it's difficult to determine when a formation is formed even manually, because a team usually takes 5 to 8 seconds to transform from one formation to another formation when they get or lose the ball.

To address the aforementioned challenges, we propose a new representation of the positional data, which can capture the topological feature of the unstructured data. We then use temporal convolutional networks to learn the temporal patterns of the features.

2 Related Works

With the great development of computer vision and portable sensor, the experts in soccer are able to get access to the positional data of the players and dig deep into the tactics of the soccer game. Most of the works focused on the passing pattern of the teams. Given the passing events and positional data of the players, [Wang et al., 2015] developed a specific topic model and found 10 main passing patterns of the players of Barcelona. With a same data type, [Decroos et al., 2018] defined the phases of events and used hierarchical agglomerative clustering to extract the attcking patterns. To facilitate the analysis, [Wei et al., 2013] firstly developed a decision forest to segment the game into inplays and stoppages based on the player positions, team centroid and ball positions. Then they used formation as a prior knowledge to explore the offensive and conceding patterns of a team.

Few works focus on formation analysis. [Wu et al., 2019] developed a visual analytic system with a tailored sankey diagram to visualize the formation change. The formations are extracted by a two-step clustering developed by [Bialkowski et al., 2014]. The extraction consists of a k-means-like clustering and a hierarchical agglomerative clustering. However, they considered the formation as a stable factor, at least for a half game, which is not suitable for a fine-grained analysis of formation.

TCN

3 Data Preprocessing and Description

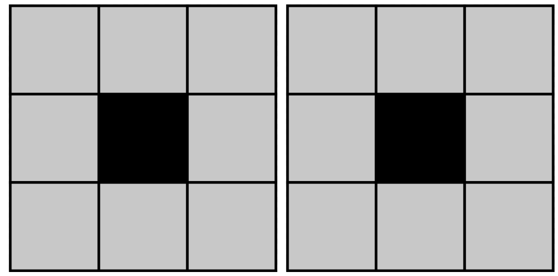
Our data is collected from the panoramic videos of soccer games on Pixellot[1]. We used a semi-automatical collecting method to label the position of the players in the videos. To make sure the efficiency and preserve the precision of the labelling, a particle filter based tracking algorithm,developed by [Dearden et al., 2006], is ran under the supervision of the annotators. When occlusion happens and the tracker loses the target, the annotators have to mannually stop and correct the position of the tracker and make sure the tracking is accurate.

With the position of the players in the panoramic videos, we obtain a affine projection to map the position data to a two dimensional soccer pitch. With the position of the players in the panoramic videos, we used an affine projection to map the position data to a two dimensional soccer pitch. According to the latest Law of the Game 2018/19[2] by the International Football Association Board (IFAB), the length of the pitch (touchline) is between 100m and 110m and the width (goal line) is between 64 and 75m. But in many important international competitions, such as 2018 Russia World Cup, the pitch dimension is 105m by 68m. So we normalized the positional data into a 105m $\times$ 68m area.

The events on the pitch, such as foul, out of line and goal, are also very important for game analysis. The offensive or defensive state of a team may be interrupted by the events because the game will stop and kick off. With the events, we segment the game into small periods, and the time between the periods is stoppage time. Each period is independent to other periods. The position change within a period can be viewed as a sequential process.

[1] https://www.pixellot.tv/

[2] https://img.fifa.com/image/upload/khhloe2xoigyna8juxw3.pdf



(a) Moore Neighborhood (b) Player Neighborhood

Figure 1: An illustration of Moore neighborhood

4 Approch

Formation represents the role of the players on the pitch. A natural idea is regarding the players as the nodes in a network and convert the problem to a graph embedding problem. However, the definition of formation doesn't illustrate the connection between the players, which means that there is no edges between the nodes (players). According to the experience of the domain expert, the players in the same line, such as backward line, don't necessarily have connections during a game, such as passing. In practice, when a domain expert needs to determine the formation of a team, he will take a look at the evolution of the relative position of the players within a period. So the formaition detection is a context aware problem and the model have to take the previous positions of the players into consideration. Our task becomes finding a suitable representation of the relative position of the players and a model to learn the temporal pattern of the representations.

4.1 Topological Representation

To capture the topological features of the player positions, we borrow the idea of moore neighborhood in cellular automata theory. The original definition of the moore neighborhood is the cellular itself and the surrounding eight cells, as Figure 1a. The mathematical definition of the Moore neighborhood with a radius k is

$$\{c_{ij} | \|c_{ij} - c\|_{L_\infty} < k\}, \tag{1}$$

where c_{ij} is the cell on the i-th row and the j-th column ,c is the center of the Moore neighborhood and L_∞ is the Chebeshev distance.

We firstly parameterize the soccer pitch.

References

[Bialkowski et al., 2014] Alina Bialkowski, Patrick Lucey, Peter Carr, Yisong Yue, Sridha Sridharan, and Iain Matthews. Large-scale analysis of soccer matches using spatiotemporal tracking data. In 2014 IEEE International Conference on Data Mining (ICDM), pages 725–730. IEEE, 2014.

[Dearden et al., 2006] Anthony Dearden, Yiannis Demiris, and Oliver Grau. Tracking football player movement from a single moving camera using particle filters. 2006.

[Decroos et al., 2018] Tom Decroos, Jan Van Haaren, and Jesse Davis. Automatic discovery of tactics in spatio-temporal soccer match data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery on Data Mining, pages 223–232. ACM, 2018.

[Lea et al., 2017] Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. Temporal convolutional networks for action segmentation and detection. In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 156–165, 2017.

[Wang et al., 2015] Qing Wang, Hengshu Zhu, Wei Hu, Zhiyong Shen, and Yuan 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, pages 2197–2206. ACM, 2015.

[Wei et al., 2013] Xinyu Wei, Long Sha, Patrick Lucey, Stuart Morgan, and Sridha Sridharan. Large-scale analysis of formations in soccer. In Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on, pages 1–8. IEEE, 2013.

[Wu et al., 2019] Yingcai Wu, Xiao Xie, Jiachen Wang, Dazhen Deng, Hongye Liang, Hui Zhang, Shoubin Cheng, and Wei Chen. Forvizor: Visualizing spatio-temporal team formations in soccer. IEEE transactions on visualization and computer graphics, 25(1):65–75, 2019.