Characterization and comparison of Japan professional football clubs based on attack patterns

Ryo Ishii† Michiharu Kitano‡ and Hiroshi Yadohisa†

Abstract — In recent years, studies using football data have proliferated. Most existing studies focus on the results of matches. In contrast, few studies have considered each event that is recorded successively during the matches. Therefore, in this paper, we characterize and compare football clubs by considering such information. We characterize the ball’s movement among players or different field areas into important events such as shoot and track the ball’s attack patterns. These important events are collectively defined as attack patterns. We analyze the attack patterns using social network analysis and build digraphs comprising players or areas in order to characterize and compare clubs.

Keyword: pattern extraction; digraph; social network analysis

1 Introduction

Some previous studies have conducted research on football using data collected during matches. (Sugimoto and Sakaori, 2009; Hirotsu and Miyaji, 2006; Kim et al., 2004). Sugimoto and Sakaori (2009) predicted the probability of football scores by considering the remaining time in matches using the Poisson distribution. Hirotsu and Miyaji (2006) evaluated the formation for J-League clubs by using game theory. Furthermore, Kim et al. (2004) predicted the results of matches by using a Bayesian network. However, few studies have been conducted using event data recorded by observing successive events during matches, because most existing studies are focused on the results of matches.

In this paper, extract patterns whose flow into events of the shoot has a sequence length of three and count each type among the patterns. In addition, relationships among players or field areas are represented as digraphs and analyzed using social network analysis. Using this analysis, different clubs that consider the recorded data are compared.

2 Football data

The data used in this paper was provided by Data Stadium Inc. and it describes a J-League Division 1 held in 2008. In this paper, we focus on attack patterns, which are defined based on the movement of the ball during the match. The recorded data captures the events that occurred over the ball in one transaction. The variables used in this analysis are series number, player names, club names, and coordinates the ball.

3 Methods of analysis

In this section, some definitions of the terms used in the study are provided and the methods of analyzing the attack patterns are explained.

3.1 Definition

Definition 3.1 Field and area
The field is defined as follows:

\[ F = \{(x, y) | -158 \leq x \leq 158, -113 \leq y \leq 113, x, y \in \mathbb{R}\}. \]

The field is divided into nine parts and each part is defined as an area. The manner in which the field is divided is shown in Figure 1.
Definition 3.2 Player
A person who takes actions such as trap, pass, and shoot against the ball is defined as $P_i$ ($i = 1, \cdots$), where $i$ indicates the uniform number.

Definition 3.3 Sequence of players
The sequence pattern of the ball's movement among players in the same team and in the match is defined as follows:

$$SP_\ell = (P_{i_1,n-\ell+1}, P_{i_2,n-\ell+2}, \cdots, P_{i_n,n}) \quad (\ell \leq n),$$

where $P_{i_1,n-\ell+1} \neq P_{i_2,n-\ell+2}$, $P_{i_1,n-\ell+2} \neq P_{i_3,n-\ell+3}$, $\cdots$ and $P_{i_{n-1},n-1} \neq P_{i_n,n}$. $\ell$ is the length of the sequence and $n$ is the total number of players in the series.

Definition 3.4 Sequence of areas
The sequence pattern of the ball's movement among areas in the same team and in the match is defined as follows:

$$SA_\ell = (A_{j_1,m-\ell+1}, A_{j_2,m-\ell+2}, \cdots, A_{j_m,m}) \quad (\ell \leq m),$$

where $\ell$ is the length of the sequence and $m$ is the total number of areas in the series.

3.2 Extracting attack patterns
For analyzing attack patterns, we focus on $SP$ and $SA$ sequences. These patterns with the length of three ($\ell = 3$) and $P_{3,n}$, $A_{3,m}$ correspond to the event of shooting, because we are interested in studying the characteristics of how each team attacks. For characterizing and comparing clubs on the basis of attack patterns, we count the types of attack patterns and extract the variability characteristics. In addition, we build digraphs among the players or field areas and extract these characteristics through social network analysis.

We extracted daily reports to produce detailed methods and results.

Acknowledgment
We would like to express our gratitude to Data Stadium Inc. for providing the football data.

References

