Assessing Similarities in Soccer Team Tactics

Akitaka Moriyama Maxim Mozgovoy
The University of Aizu
Tsuruga, Ikki-machi,
Aizuwakamatsu, Fukushima, Japan
+81-242-37-2664
m5201103, mozgovoy@u-aizu.ac.jp

ABSTRACT

We describe similarity assessment of soccer game formations between teams by using two different approaches: (a) a sum of distances between each pair of assigned players obtained with Hungarian algorithm; (b) a cosine similarity between team heat maps. We test our algorithms separately for offense and defense strategies of teams using real soccer game data. As a result, by considering similarities we can reveal similar and distinct approaches to attack and defense used by different soccer teams.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Theory

Keywords

team similarity, classification, clustering

1. INTRODUCTION

The importance of data analysis on sports had been increasing year by year. Soccer game is not an exception. The outcome of a game depends on players and their behavior on the field. Skills and strategies are advanced as time goes on, so detailed analysis of team tactics becomes and important problem in sport analytics [1].

However, prediction of group games of player action and trends of the team such as soccer game is not easy. One of the problems is the difficulty of quantitative analysis of player abilities. Thus, research of soccer game data is scarce.

In this paper, we deal with formation analysis by comparing teams using real game data of soccer games. The data is taken from actual J1 League games (J1 League is the top division of Japan professional soccer). We attempt to analyze similarities and differences between team tactics, and evaluate consistency of tactics of the same teams across games.

2. METHODOLOGY

We use the data of five matches, provided by Data Stadium Inc [2]. These matches were played in J1 league in 2011 [3]. There are six teams in the dataset: Koube, Nagoya, Shimizu, Yamagata, Urawa, and Yokohama. Urawa and Yamagata teams played only one game, while other teams played two games.

Each game data file contains three chunks. The first chunk contains only the current frame number. The second chunk lists the status of each player (home team or away team, jersey number, coordinates, and speed). The last chunk contains the ball status, represented by ball coordinates, speed, owner, and information whether the game is currently suspended.

At the preprocessing stage we extract only game data that corresponds to non-suspended periods of the game. We assume that team formations in game pauses is not important. Next, team formation is different between defense phase and offense phase. We divide the data into defense phase and offense phase. In addition, we divide the data into the first half, and the second half. We compare team tactics across individual halves of the game, but never mix defense and offense phases. We use two different methods of comparison:

1. Sum of distances between each assigned player obtained with Hungarian algorithm.
2. Cosine similarity between team heat maps.

Hungarian method is one of the methods for solving the assignment problem [4]. We can optimally juxtapose the players of a team in two different situations to minimize a distance between them. In the current work, we consider assignment optimal, if the sum of distances between the assigned players is minimal. We calculate each of player 's average coordinates from team data, and apply Hungarian method to two teams, we then derive the distance. We repeat it for every team. We use Euclidean distance to measure a distance between the players.

Our second method needs a heat map of each team, obtained from coordinate data of each player. We divide a soccer field into a grid of 20 by 13 cells, and calculate a probability for each cell to be occupied by some team player from a team data, and make a heat map. We then connect map rows sequentially from top to bottom to obtain a vector. Thus a map converted into a vector. We repeat it for every team. These heat maps represented with vectors are compared using cosine similarity (a dot product between the vectors).

We calculate an average distance and similarity between each possible pair of teams separately for defense phase and offense phase.

3. EXPERIMENTAL RESULTS

Altogether, we compare 190 pairs of teams in each (offense / defense) phase. The average results are shown in the tables 1-4.

Table 1 shows the distances between team pairs obtained with Hungarian algorithm in the defense phase. It is clearly seen that the teams Koube, Urawa, and Yamagata are very consistent in their defense tactics, and show considerably smaller distances between their own games. However, Shimizu and Yokohama teams are sometimes closer to other teams in their defense style rather than to their own behavior in other games.

Table 1: Hungarian distances (defense)

	Koube	Nagoya	Shimizu	Urawa	Yamagata	Yokohama
Koube	5523.85					
Nagoya	8330.71	7905.68				
Shimizu	6322.05	7643.58	6997.47			
Urawa	7962.39	6338.40	6666.29	4436.14		
Yamagata	7509.70	6872.96	6172.73	5145.65	3953.71	
Yokohama	6909.35	6947.56	6516.39	5988.50	6091.26	6622.84

Table 2 shows the distances between team pairs obtained with Hungarian algorithm in the offense phase. In this experiment it is evident that the teams are more experimental in attack, and show higher variety of styles in different games. There is no consistency across different games of the same team (except Urawa team).

Table 2: Hungarian distances (offense)

	Koube	Nagoya	Shimizu	Urawa	Yamagata	Yokohama
Koube	5563.33					
Nagoya	7010.16	6274.00				
Shimizu	5914.80	8467.29	6466.92			
Urawa	5072.38	5318.69	6624.46	3971.05		
Yamagata	5695.01	6466.35	6171.26	4873.85	5918.03	
Yokohama	6291.78	7056.96	7189.49	6068.62	6471.06	7775.19

Table 3 shows the cosine similarities between team heat maps in the defense phase. Interestingly, heat maps of all teams are very similar, and the similarity score never drops below 0.88. Probably, a heat map better characterize the game of soccer itself rather than individual team tactics. However, this method also lists Koube and Urawa among the most consistent teams in the dataset.

Table 4 shows the cosine similarities between team heat maps in the offense phase. These values are again consistent with the values shown in Table 1: the highest similarities are exhibited by Koube, Urawa, and Yamagata teams.

4. DISCUSSION AND CONCLUSION

Table 3: Cosine similarities (defense)

	Koube	Nagoya	Shimizu	Urawa	Yamagata	Yokohama
Koube	0.95					
Nagoya	0.89	0.92				
Shimizu	0.92	0.91	0.94			
Urawa	0.92	0.92	0.93	0.95		
Yamagata	0.92	0.92	0.93	0.93	0.91	
Yokohama	0.88	0.91	0.91	0.92	0.91	0.90

Table 4: Cosine similarities (offense)

	Koube	Nagoya	Shimizu	Urawa	Yamagata	Yokohama
Koube	0.94					
Nagoya	0.88	0.90				
Shimizu	0.91	0.90	0.90			
Urawa	0.92	0.89	0.91	0.93		
Yamagata	0.90	0.91	0.91	0.91	0.94	
Yokohama	0.90	0.90	0.91	0.92	0.93	0.93

It is interesting to note that the features of individual team is more clearly visible in the defense phase of a game. Most teams adhere to similar defense strategies across different games. Offensive tactics vary highly, and teams experiment when playing against different opponents. This trend is seen both in Hungarian algorithm-based, and cosine similarity-based calculations.

In its turn, cosine similarity of heat maps cannot be considered a reliable indicator of team behavior similarity. All teams in our dataset produce very similar heat maps, probably, reflecting the nature of soccer rather than individual team tactical schemes.

The present work shows that most teams exhibit numerically detectable features that make them different from other teams. We hope that this method will contribute to data analysis and further development of soccer game research.

5. REFERENCES

[1] 加藤 健太. サッカーにおけるデータ分析とチーム強化 (data analysis and team strengthening on soccer game). **電子情報通信学会　通信ソサイエティマガジン** *(IEICE)*, 10(1):29–34, 2016.

[2] データスタジアム株式会社 (Data Stadium Inc.). `https://www.datastadium.co.jp`. Accessed: 2016-08-05.

[3] J リーグ (J League). `http://www.jleague.jp`. Accessed: 2016-08-05.

[4] James Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.