ISSN 2616-7433 Vol. 2, Issue 6: 36-39, DOI: 10.25236/FSST.2020.020609

Research on Team Strategy Based on Data Mining and Visualization

Yuhui Zhang, Xujie Fan, Wanhao Ma

School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan Hubei 430000, China

ABSTRACT. *Football is known as "the world's number one sport". The victory of a football game requires the tacit cooperation of the team in addition to the individual's outstanding performance. We analyze the passing data from the perspective of time and space. First, we set up a passing network, describing network parameters with network density, grid distance, and degree centrality. Second, we use the weights of edges in the network and the centrality weights of nodes to identify network patterns, and pass nodes that exceed the threshold to identify whether the team configuration is dyadic or triadic. From the perspective of time, the number of passes per minute represents the rhythm of the game, and the number of consecutive passes per minute represents the degree of interference from the opponent. We analyze the rhythm and opponent's influence to Huskies in games. From the perspective of space, we plot the passing trajectories of the first half and the second half in 2 games, and use clustering to obtain the habitual passing area of the Huskies. Finally, we analyze Pass flexibility to the team.*

KEYWORDS: *data mining*

1. Introduction

Football is known as "the world's number one sport". This world-famous sport is highly interesting and competitive. In this article, we need to build a model to solve the problem based on the data of 23, 429 passes and 59, 271 matches of 30 Huskies in 38 games: Create a network model for passing between players, and consider other structural indicators and network properties in the game. At the same time, explore the multiple scales.

2. Establishment of Passing Model

In this part, we carried out a visual analysis of the network, proposed a scheme for identifying the network pattern, and also analyzed the passing network from multiple dimensions such as time, and flexibility.

2.1 Visual Analysis of Pass Network Relationships

In order to increase the rationality and intuitiveness of the passing network, we will analyze the connection between the players in the game from the time dimension and the space dimension. We need to process the 23429 data in passevents. csv.

First, clean the data and filter out outliers. Here we choose the model detection method, that is, first establish a data model, and the outliers are the objects that cannot be fitted.

Define the passing data of a participating team as the adjacency matrix W in a complex network, with each participating player as a node, and the players'continuous edges based on the passing situation between the players. The weight of the consecutive edges is W in the adjacency matrix of the passing ball. The size of the value, the direction of the connected side indicates the direction of pass. Then define the matrix A, A represents the directed unauthorized network of passing situations int the match. If there is a passing record, it will be 1, otherwise it will be 0.

$$A = \begin{cases} a_{ij} = 0 & \text{if } \omega = 0, \\ a_{ij} = 1 & \text{if } \omega \neq 0 \end{cases} \qquad \begin{aligned} W &= (\omega_{ij})_{30*30} \\ A &= (a_i j)_{30*30} \end{aligned}$$

ISSN 2616-7433 Vol. 2, Issue 6: 36-39, DOI: 10.25236/FSST.2020.020609

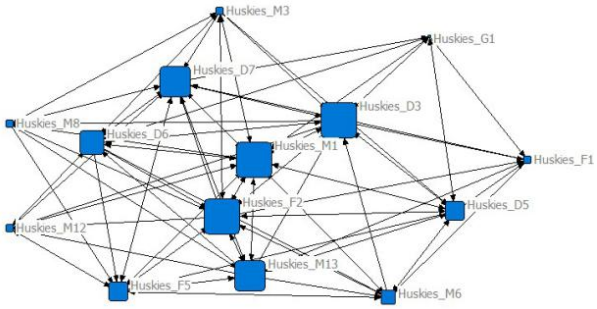

Fig. 1 Passing Network

Among them, ω_{ij} represents the number of times that player i passes to player j, and the passing direction is the network connection direction.

Using Ucinet, we can draw a diagram of the passing network relationship, as shown below.

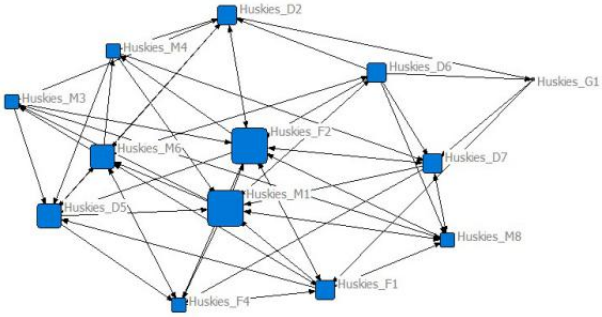

Fig. 2 Passing Network of Match 16 and Match 17

The passing network simulates the process of passing in a football match, reflecting the connection between players, and the centrality of players. In the 16th game, passes are mainly concentrated at the two nodes of center Huskies_M1 and forward Huskies_F2. It can be found from the positions of the forward and center that the passing area is mainly in the front midfield. In the 17th game, pass are mainly concentrated between defender Huskies_D3, center Huskies_M1 and forward Huskies_F2. The passing network diagram can intuitively reflect the passing relationship and passing trend of the entire team, and can also roughly show which nodes and connections in the team are more important. For a clearer understanding of the characteristics of the passing network, a quantitative analysis of the passing network is required.

2.2 Identification of Network Mode

In each match, the relationship between players is complicated. The personnel in the ternary network may also form a binary network with other players. A game may have both binary, ternary configurations and even multiple configurations. In order to simplify the problem, we only discuss the binary configuration and ternary configuration in the team, and use the configuration with the highest frequency as the match mode.

We randomly selected two games for testing, and the results are as follows.

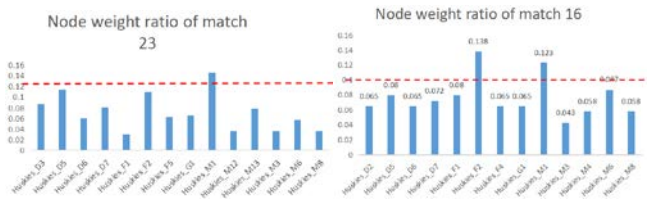

Fig. 3 Node Weight Ratio

ISSN 2616-7433 Vol. 2, Issue 6: 36-39, DOI: 10.25236/FSST.2020.020609

We set the weight ratio threshold to 0. 1, so it is not difficult to judge that the weight ratio of two players in match 16 exceeds the threshold, therefore, the match 16 network is a dyadic configuration. In match 23, there are 3 nodes whose weight ratio exceeds the threshold, therefore, the match 23 network is a triadic configuration.

2.3 Analyze Passes in the Time Dimension

From the perspective of time, we separately draw a line chart of the number of passes minute of Huskies in four matches. It can be seen from the line chart that match 1 and match 14 have more passes per minute, indicating that the rhythm of the game is faster. In match 11, the number of passes per minute is relatively low and the rhythm of the game is slow.

2.4 Analysis of Player's Passing Flexibility

Considering the passing flexibility of each player, we counted the passing methods of each player. When a player's passing methods are more diverse and the number of passes is more, which means the standard deviation is large, we think that the player has a high passing flexibility. In order to simplify the problem, we assume that the passing behavior will bring positive effects to the team. We use SPSS software for descriptive statistics and get results.

In the process of descriptive statistics, we found that almost every player performed Simple Pass many times, so we don't think project has research value, and don't consider it anymore. According to the table, we use the standard deviation of the data of each player's passing method to describe the flexibility of the player's passing. The greater the standard deviation, the more flexible the player, and the darker in the table. It can be seen that goalkeeper Huskies_G1 has the best flexibility, which means that the Huskies has a stronger defensive ability.

2. 5 Cluster Analysis of Passing Area

In order to prevent the opponent from making strategies based on the passing custom of Huskies, it is important to study Huskies'passing trajectory and passing area. To simplify this problem, we assume that the football field is square. We draw the trajectories of the first half and the second half of the first 5 games.

It is difficult to determine the distribution of the passing position quantitatively only based the trajectory chart. So we cluster the pass positions and use the clustering coefficient to help us judging.

We use systematic clustering method to calculate the clustering coefficient and draw a scatter diagram which use number of categories as x-axis. The diagram is as follows.

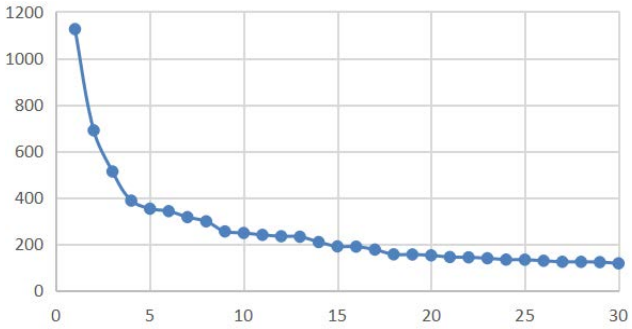

Fig. 4 Clustering Coefficient Scatter Diagram

According to the information in the figure, using the "elbow rule", we can easily find that the clustering coefficient does not change obviously after the number of classifications n=10, so we can determine that these passing positions can be grouped into 10 categories, which means the area of the ball is scattered, and it is less likely to be used by opponents.

ISSN 2616-7433 Vol. 2, Issue 6: 36-39, DOI: 10.25236/FSST.2020.020609

References

[1] Gu Shan (2019). Gephi-based visualization of the passing relationship network of football games[C]. Chinese Academy of Sports Sciences. Compilation of abstracts of the 11th National Sports Science Conference, pp.2907-2908.

[2] Wang Baoqing (2019). Research on CBA pass data network diagram based on GraphX[C]. Chinese Sports Science Society. Compilation of abstracts of the 11th National Sports Science Congress, pp.2987-2988.

[3] Cao Xuewei, Li Xiaotian, Fu Yingyao (2019). Study on the impact of home and away matches of Chinese Super League on team sports performance based on complex network method [J]. Sports Science, vol.40, no.4, pp.22-28.