Research on the Team Processes in Competitive Team Sports Based on Network Structure Characteristics and Other Factors

Yuang Liu

School of Computer Science, Beijing University of Posts and Telecommunication, Beijing, China.

liuyuang@bupt.edu.cn

Keywords: Team performance, Social network analysis, Competition result prediction.

Abstract: This paper aims to use the nature of the network and the analysis methods of social networks to analyze the passing behavior of football players in the game, and give a set of quantitative methods to evaluate team performance, and finally hope to select some reasonable game process indicators and predict the outcome of the game through machine learning. The data set used is huskies team's real game data. We combined the network model and players' position to figure out the dyadic, triadic configurations, and the structure of the team. After that, we use Centrality and Closeness Centralization to evaluate players, use Weighted Directed Network Entropy and Clustering Coefficient to evaluate team. What's more, we have constructed several indicators from the data to measure the ability of players and the performance that reflect successful teamwork. Firstly, we have established passing networks for huskies, including a diagram of the passing network based on passing data from the entire season, and several passing networks of different matches. From the former, we can see the passing situation of the entire team throughout the season, and in the latter, we can see the changes in the network during the season. We draw a schematic diagram that both reflects the player's actual passing position and reflects the passing network between players. Such a graph is handy for analyzing the dyadic and triadic configurations and team formations. At the same time, we study the tactics of our struggle against the enemy, figuring out structures and critical attacking paths of the Huskies in several matches. This is followed by individual and team evaluations. We defined evaluation indexes and established evaluation models for them. For individuals, we selected several representative members in the Huskies team and analyzed their characteristics through data. As for the team, we analyzed the changes in the evaluation factors over the course of the season. We use BP Neural Network to forecast the matches outcome and find out the critical features of a team to win. We use degree centrality to measure the centralization of the team, use eigenvector centrality to measure the importance of players. For the entire passing network, we use closeness centralization to measure team-centricity and Weighted directed network entropy to measure team decentralized level. Then we find how the centralization of the team impacts the behavior during the match.

1. Introduction

The network model has been widely used in various aspects, such as biological networks, neural networks, network links in computer, and so on. Graph theory gives us much theoretical basis for research. For example, node centrality, degree center, network density, and so on. With these parameters, we can accurately and reasonably describe the nature of the network under study. Of course, network analysis also applies many other scientific theories, such as probability theory, random processes, optimization theory, genetic algorithms, and so on, and even some of the concepts of entropy, statistical mechanics, thermodynamics in physics have also been applied. It can be said that network analysis is a comprehensive discipline. We can see the different fields of disciplines here cross-integration of the collision of the wonderful spark.

As we all know, the NBA has adopted a very scientific quantitative system to evaluate players and teams. The system that covers all aspects of basketball can be very detailed, scientific, objective enough to evaluate the players' level and ability. However, how to choose these indicators, and how to allocate the weight of these indicators in the establishment of an evaluation system to be more reasonable is a very fundamental and challenging question. With such a clear indicator, everyone can know what they should improve in the next training.

Using some of the indicators of network structure in graph theory to analyze the team is the next very exploratory and meaningful field. The theory is very instructive for the group's recommendations.

2. Notations

Tab. 1 The Symbols of Definition and Description

Symbol	Meaning (Units)
P	Time in possession divided by the total time in the game
A	An indicator of the team's offensive ability
D	An indicator of a team's defensive capabilities
Ski	An indicator of team skills
Coo	An indicator of a team’s cooperation ability
$Core$	A metric defined to measure the degree of decentralization of the team's ability
Ssp	The indicator to measure the overall sports quality of the team (Sportsmanship)
$P_{control}$	The proportion of time our team control the ball over the total time of the game
P_{pass}	The rate of our successful pass
P_{foul}	Our foul action’s percentage of all events
P_{off}	Our offset action’s percentage of all events
P_{duel}	Our successful rate of dueling balls
P_{shot}	Natural Mortality
x	The number of goals we scored

3. Establishment and Characteristics of the Passing Network

For the Huskies, we treated each player as a node, and each pass constitutes a link between the players. And then we drew the passing networks of its teams, which includes all the games that the team has played throughout the season. In this way, we can clearly understand the passing strategies (dyadic and triadic configurations and team formations) adopted by players of huskies during the game, to analyze whether the team has a very important center role. A clear understanding of these characteristics is conducive to improving your team and understanding the opposite team, to attack the problematic part of the opponent.

3.1 Passing networks of the teams

For the convenience of drawing and discussion, we made this one-to-one correspondence table between the actual ID number of the team members.

Tab.2 the One-to-one Correspondence

real ID	D1	D2	D3	D4	D5	D6	D7	
node	1	2	3	4	5	6	7	
real ID	D8	D9	D10	F1	F2	F3	F4	F5
node	8	9	10	11	12	13	14	15
real ID	F6	G1	M1	M2	M3	M4	M5	
node	16	17	18	19	20	21	22	
real ID	M6	M7	M8	M9	M10	M11	M12	M13
node	23	24	25	26	27	28	29	30

The Huskies team has a total of 30 players, but only 11 players play in each game. We set up a 30 * 30 two-dimensional matrix to describe the situation of one-to-one passing between players. For better viewing effect, we draw a bubble chart according to the matrix. The size of the bubble represents the number of passes. More giant bubbles indicate more passes. We can see the effect below (fig1(a)):

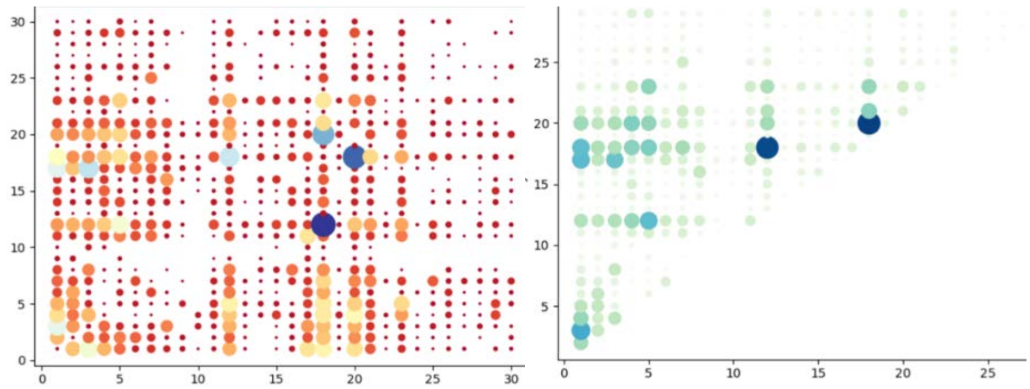


Fig. 1(a, b) Bubble Chart that Reflect One-to-one Passing

In order to make the primary passing behavior and cooperation more obvious, we added the times of A passes to B and B passes to A. After that, we ignored some small amounts, and used dark and more exaggerated sizes to indicate more passes. Finally, we have the fig1(b) as follows. In this picture, we can clearly see the passing characteristics between the players.

And the passing network (Attention: the player's position in the picture below is not the actual position on the court!)

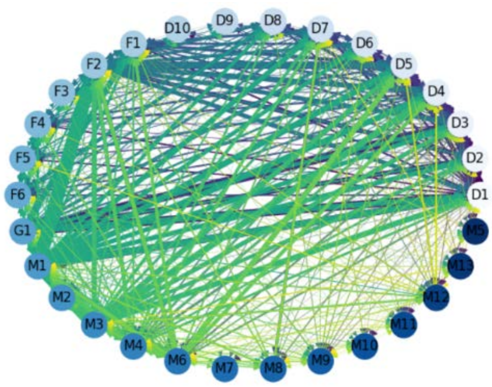


Fig. 2 The Passing Network

There are some problems with the picture above. For example, some players only took part in a few matches, which will inevitably cause the player to become an unimportant node in the network. But it seems more reasonable and informative if we take into account that the low number of appearances itself shows the unnecessity of the player.

3.2 Analysis of the Players' and Team's Characteristics Depend on Parameters of the Passing Network

3.2.1 Measurement of player centralization level: centrality

The indicators of nodes: degree centrality, betweenness centrality, closeness centrality, eigenvector centrality.

Two suitable indicators have been chosen to describe the centrality of the nodes: degree centrality and eigenvector centrality.

3.2.2.1 Degree Centrality

The node will be more important when it is directly connected to many other nodes in a network. In the football network, such nodes represent that the player often receives passes from other players, indicating that the player is very significant to the passing process. Therefore, we should protect this kind of players in our team and pay more attention to those who in the opposite. We use the standardized centrality value to measure degree centrality：

$$DC(v) = \frac{InDegree(v)}{V-1} \tag{1}$$

$InDegree(v)$ is the indegree of node v, and V is the number of nodes in the network

Via calculations, we get thirty players' degree centrality of huskies.

Tab. 3 Values of players' DC

real ID	D1	D2	D3	D4	D5	D6	D7	
node	1	2	3	4	5	6	7	
DC	23.38	16.1	20.07	21.2	20.03	11.3	15.48	
real ID	D8	D9	D10	F1	F2	F3	F4	F5
node	8	9	10	11	12	13	14	15
DC	9.9	1.47	1.07	16.31	31.97	2.52	9.97	7.69
real ID	F6	G1	M1	M2	M3	M4	M5	
node	16	17	18	19	20	21	22	
DC	26.48	17.1	35.07	2.38	24.72	17.96	1.41	
real ID	M6	M7	M8	M9	M10	M11	M12	M13
node	23	24	25	26	27	28	29	30
DC	18.52	0.62	6.1	4.9	1.38	1.72	7.4	1.96

3.2.1.2 Eigenvector centrality

The importance of a node depends on both the number of its neighbor nodes and the importance of the neighbor nodes. Define x_i as the importance of v_i.

$$EC_i = x_i = c\sum_{j=1}^{n} a_{ij}x_j \tag{2}$$

where c is a proportional constant, $x = [x_1, x_2, \dots, x_n]^T$, When the steady-state is reached several iterations, it can be written as the following matrix form:

$$x = CAx \tag{3}$$

This indicates that x is the eigenvector corresponding to the eigenvalue c of matrix A

The calculation steps are:

1.The characteristic decomposition of the pair of calculation sits adjacent to the matrix

2.Select a feature vector with the largest feature value

3.The centrality of the node i is equal to the element i in the feature vector

In 5.1, we discuss the evaluation indicators of individual ability and give an overall score. This full score can be calculated as a weight of importance here. Compared with degree centrality, eigenvector centrality can better reflect the importance of the node.

3.2.2 The measure of Team-centricity: closeness centralization

The closeness centralization is used to compare the centrality of edge points and the central point of the network.

If a network is concentrated, the centrality of the central point is necessarily high, and the centrality of the edge points is low.

If a network is sparse, there is not much difference in the centrality of the center point and edge points

In the passing network, if the closeness centralization of the network is very high, the behavior of the passing of the team is highly dependent on individual players. This also means that the team is relatively fragile and can hardly afford to lose after the player's problems. The calculation formula for the closeness centralization is as follows:

$$DC_{graph} = \frac{\sum_{i=1}^{V}(DC(n^*)-DC(v_i))}{(V-1)(V-2)} \tag{4}$$

n* is the point that has the maximum of degree centrality.

After bringing in the data of all passes of the Huskies team, we get the closeness centralization of Huskies team's passing network:

$$DC_{graph}= 0.8525140139289963 \tag{5}$$

With this value, it can be concluded that the team's centralization of the passing network is very high.

3.2.3 Team decentralized level: Weighted directed network entropy (WDNE)

3.2.3.1 Edge importance and weighted directed edge entropy

Dfinition1: The importance of the directed edge e_{ij} is shown in formula:

$$E_{ij} = \frac{W_{ij}}{\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} W_{ij}} \tag{6}$$

Among them, n represents the number of nodes in the network, and W_{ij}represents the weight of e_{ij}. $E_{ij}$is the ratio of the weight of e_{ij} to the sum of the weights of all edges.

Definition 2: The weighted directed edge entropy of e_{ij} is

$$H_{i,j} = -E_{i,j} ln E_{i,j} \tag{7}$$

3.2.3.2 Weighted Out (In) Entropy to Node

Definition 3: The weighted outbound node entropy of node i represents that as the sum of the weighted directed edges at the starting point, as in formula ():

$$V_{i+} = \sum_{j=1,i\neq j}^{n} H_{ij} \tag{8}$$

Table shows the weighted out(in) to network node entropy of each node.

Tab. 4 Nodes' Network Node Entropy

node	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
V(in)	0.3490	0.2603	0.3127	0.3229	0.3097	0.1897	0.2575	0.1664	0.0301	0.0228	0.2697	0.4687	0.0536	0.1812	0.1431
V(out)	0.4365	0.3140	0.3837	0.3107	0.3331	0.1931	0.2391	0.1571	0.0413	0.0275	0.1512	0.4466	0.0416	0.0861	0.1036
sum	0.7855	0.5743	0.6964	0.6335	0.6428	0.3828	0.4966	0.3236	0.0714	0.0503	0.4208	0.9153	0.0951	0.2674	0.2467

node	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
V(in)	0.1787	0.2534	0.5062	0.0503	0.3755	0.2984	0.0314	0.3005	0.0141	0.1111	0.0983	0.0304	0.0383	0.1372	0.0428
V(out)	0.1395	0.2665	0.6071	0.0549	0.4504	0.2856	0.0325	0.2868	0.0174	0.0897	0.0899	0.0279	0.0483	0.0890	0.0534
sum	0.3182	0.5199	1.1133	0.1051	0.8260	0.5840	0.0638	0.5873	0.0315	0.2008	0.1882	0.0583	0.0866	0.2262	0.0963

In order to show the Weighted Node Entropy of each node in a more intuitive form, we draw a pie chart.

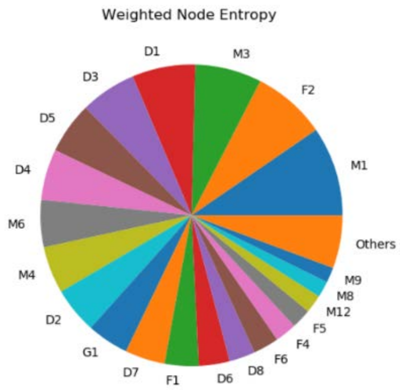


Fig. 3 Weighted Node Entropy

It can be proved that the sum of the weighted inward-to-node entropy of each node is equal to the sum of the weighted outward-to-node entropy. Moreover, we use one of them to represent the WDNE of the network.as in formula ():

$$G = \sum_{i=1}^{n} v_{i^-} = \sum_{i=1}^{n} v_{i^+} \tag{9}$$

Through calculation, we get the WDNE of the passing network of Huskies team.

$$G = 8.9893$$

In the same way, we calculated the WDNE of all opponents and their average

Huskies' passing network is more stable than the average of all teams. This may be why the Huskies' coaches often change the players: their team is not particularly centralized and therefore more robust.

match	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Huskies	10.4339	7.3934	10.0385	10.1370	10.6517	9.7167	10.4466	9.8291	6.9267	10.5175	7.0946	6.1010	8.4356	10.6628	9.0050	5.0146	9.9194	10.0483	8.4494
Opponent	7.4613	11.1268	12.3034	10.2403	10.9320	9.4896	7.6611	9.3215	12.5559	8.3736	10.4207	12.1932	12.8377	10.7022	8.4475	13.8506	10.8532	9.9334	13.2184
match	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Huskies	8.1928	10.3973	8.4860	8.3057	9.8344	9.1369	9.0575	9.2087	8.4042	7.8022	9.9327	10.0434	5.7884	7.5684	10.3480	10.2253	9.4183	9.4629	9.1610
Opponent	8.5329	7.5392	13.2190	12.1549	7.4209	9.0999	9.9971	8.6555	8.9104	8.4554	7.3427	8.2396	16.7124	10.4697	9.1938	8.1741	9.5931	8.6170	10.7481

Tab.5 WDNE of Huskies and opponents

3.2.4 The Clustering Coefficient of the Network

The clustering coefficient measures how connected a vertex’s neighbors are to one another. Specifically, the clustering coefficient is a measure of the density of the 1.5-degree egocentric network for each vertex. When these connections are dense, the clustering coefficient is high. In the passing network, now let’s assume there’s a player called A, if the players who often pass the ball to A also often pass the ball to each other, then we say A has a high clustering coefficient.

it is calculated as:

(the number of edges connecting a vertex’s neighbors) / (the total number of possible edges between the vertex’s neighbors)

In 38 matches, there have been 20 teams, the Huskies and 19 other teams. The Huskies played 38 matches while the other team each played twice. We compiled the pass data from both sides of the 38-game campaign, put the same team's data together, and then calculated the clustering coefficient value of the 20 teams' passing networks.

The calculation results：

Tab.6 20 Teams’ Clustering Coefficient Value

team	Clustering_Coefficient
Huskies	7.9213
opponent1	3.3807
opponent2	7.4006
opponent3	9.1566
opponent4	5.9841
opponent5	7.5326
opponent6	5.6713
opponent7	3.2565
opponent8	5.4918
opponent9	8.5313
opponent10	3.8878
opponent11	5.3119
opponent12	7.0328
opponent13	10.1352
opponent14	7.1299
opponent15	4.9988
opponent16	11.5321
opponent17	6.4496
opponent18	6.0279
opponent19	5.4027

3.4 Macro: Husky's passing network changes over the season

The following three pictures are the NO.1, NO.19 and NO.38 games’ passing network of Huskies team.

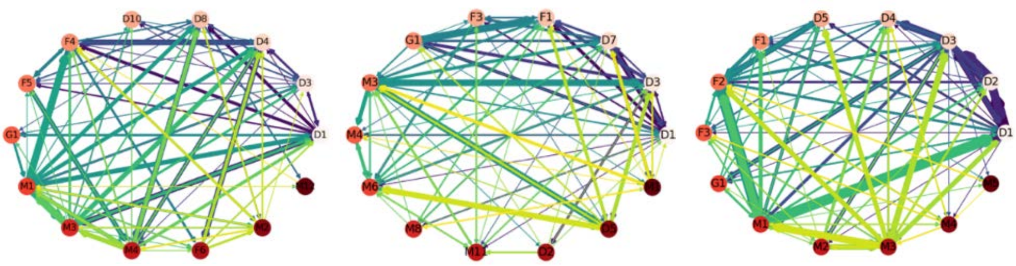


Figure 4(a, b, c) Huskies’ Passing Networks’ Changes

We can clearly see how the passing network changes with the number of matches. We saw the appearance of some central players in the first match:0 4 11 17 19（D1,D5,F2,M1,M3）, And compared with the latter two games, the frequency of passing is significantly higher. The 19th match shows a decentralized passing network, which indicates a very balanced passes between the players.

The first and the 19th game have one consistent feature, that is, the cooperation of No.19 with No.2 and No.4(D3,D5,M3), from this we can believe the passes between No.19 and No.2, No.19 and No.4 (M3 andD3,M3 and D5)is a stable mode

In order to reflect the actual position of the player on the field in a better way and to analyze the structural characteristics of the passes easier, we have this idea: Select a game, calculate the position of each player in the game when passing the ball, and analyze the position of each player to get the maximum likelihood position of the player. The player number is marked at this position in the figure. After determining the positions of all the players in this game, count the passes between them. As above, the more passes, the thicker the line between them. We have selected three best matches of the players: The No.14, No.30, No.36.

From the three passing networks below, we can see the technical and tactical characteristics of the Huskies team. First, the team has a strong sideline ability and prefers to attack from the left. Because most of the passes on the left and right sides are centered, and the left sideline, especially the left side defender D4 (No. 4), interacts with the centerline frequently, that is, not only passing the ball, but also there are a lot of incoming balls, which shows that the left sideline is the stronger part of the team's attack. Second, the team's lineup has a lot of changes, including changes in the starting lineup and the entire lineup. In this case, each player is required to have the super strong personal ability, instead of relying on a few star players.

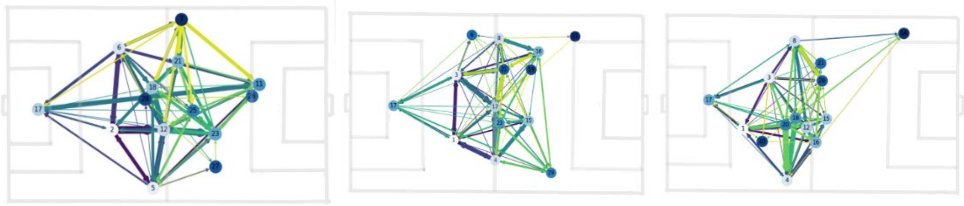


Fig. 5(a, b, c) Three Passing Networks With the Real Location

4. Performance Indicator

4.1 Personal Evaluation system

We have drawn the statistical chart of the times of player appearances. D1-D5, F1, F2, F4, G1, M1, M3, M4, M6 may have a higher contribution to the team.

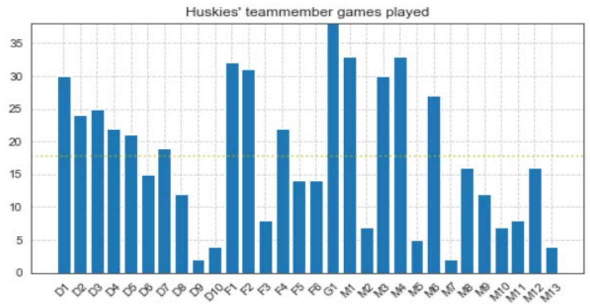


Fig. 6 Huskies' Players' Appearance Times

4.1.1 Using Classic Evaluation Factors to Establish Evaluation Model

We have drawn pie graphs of the team members about five kinds of events to see how much the player has contributed during this season.

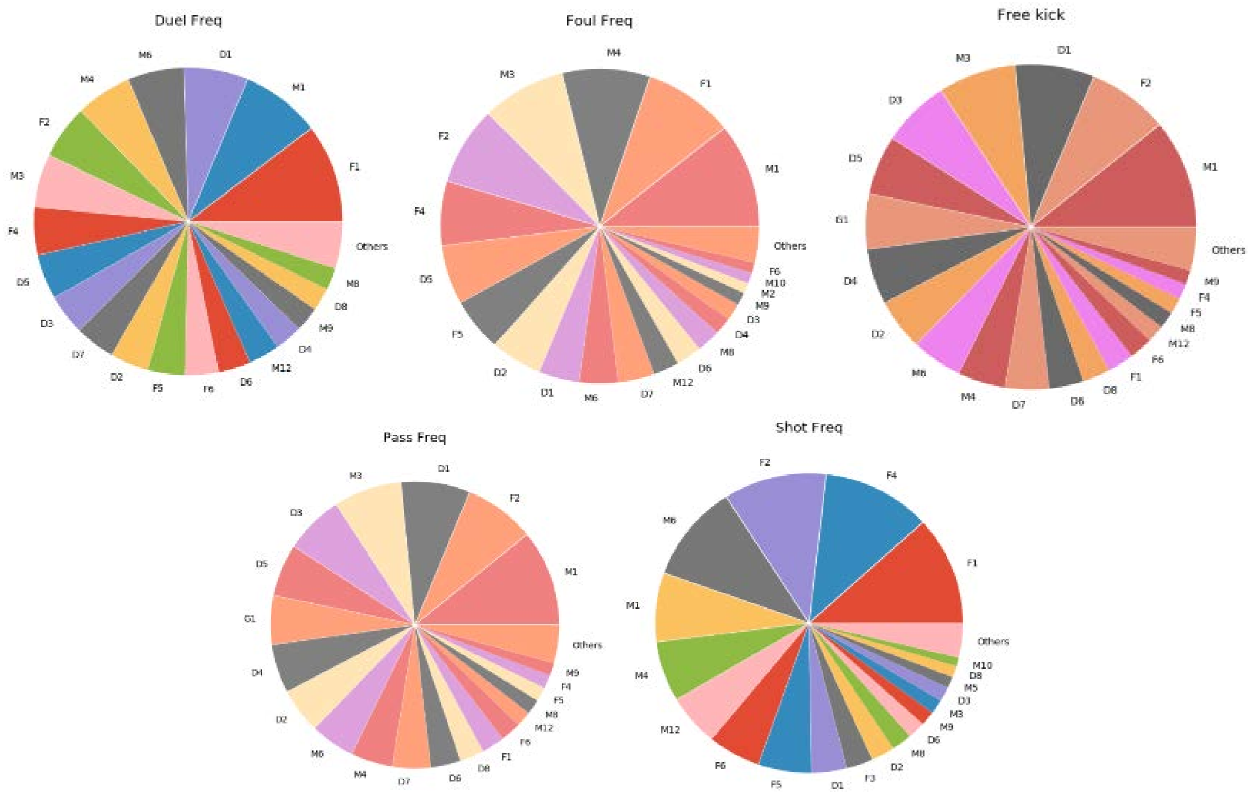


Fig.7 (a, b, c, d, e) Five competency indicators Player comparison chart

By analyzing these pictures, we can find some outstanding players. For example, M1 attended 33 matches, making the most fouls, free kick and pass. Maybe he is good at making tactical fouls and good at passing balls. F1 attended 32 matches, which is good at shot and foul, but not good at passing balls, so that he's average position is near opponent's door. F1 made many fouls, too. D1 is good at the head duel, who attended 30 matches at all.

We evaluated the abilities of the forwards, defenders, midfielders by counting data. We used ball controlling time, passing times and the success rate of passing, shot times, speed, dueling ability and times of the free-kick, times of breaking the opponent team's passing to represent the controlling ability, passing ability, shooting ability, speed, power and defense. In order to make data seem reasonable, we scaled the data to the range of 0-1. After that, we timed 0.45 to all the data and added 0.35 eventually. The ability of F1, D1 are shown in Fig [8, a, b].

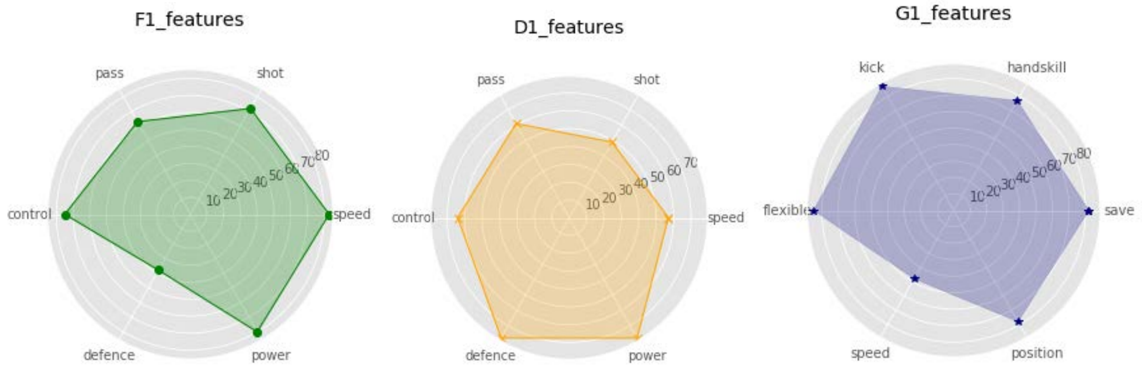


Fig. 8 (a, b, c) Personal Capability Radar Map

On the other, Goalkeeper is unique in the team. We counted the times of saves, the distance between their average positions and the goal, moving speed, the variance of their positions, the times of their Free kicks, times of hand pass, respectively. By doing a similar processing, we got the scores of the goalkeeper. We visualized the data in the Fig [8, c].

4.1.2 Evaluating Players with Indicators in Network Structure

We use Weighted Node Entropy to describe the players' distribution. The contribution of a node to the stability of the entire network depends on the Weighted Directed Node Entropy of the node. The node whose weighted directed node entropy is big has large distribution to the stability of the network. If this node hurt, the network will face a big problem. In our passing network, the nodes that have big number of Weighted Node Entropy has large distribution to the team. What's more, we can control the network more efficient by controlling this node. So, the fig [3] tells us that the most obvious players with this trait are M1, F2, M3, D1. So, in that sense, they are the most important players in the Huskies.

4.3 Use Some Artificial Intelligence Algorithms to Predict the Outcome of The Game

4.3.1. Use BP Neural Network && Indicators in Network Structure&&Actions to predict results

We selected 13 features: Closeness Centralization, Weighted Directed Network Entropy,

Clustering Coefficient, the number of freekicks, shot, duel and pass taken by our team, the rate of the successful pass, possession rate, offside rate, foul rate, successful duel rate, their number of shots. What we want to forecast is the result of the game. We have 38 matches' data and result, A match will generate two sets of data, namely the data on both sides of the match, so we have 76 sets of data. For the convenience of the learning process, we digitized the results. 3 represents 'win', 2 represents 'tie' and 1 represents 'loss'. We used 50 samples as training data and 26 samples as testing data. The Neural Network has 14 inputs and 1 output. We set the Hidden Layer number was 9.

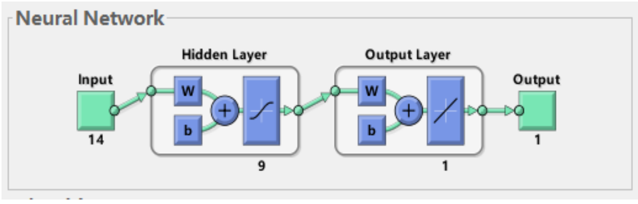
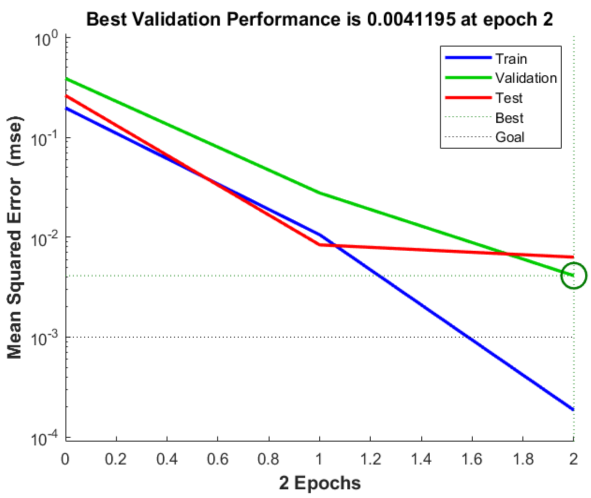
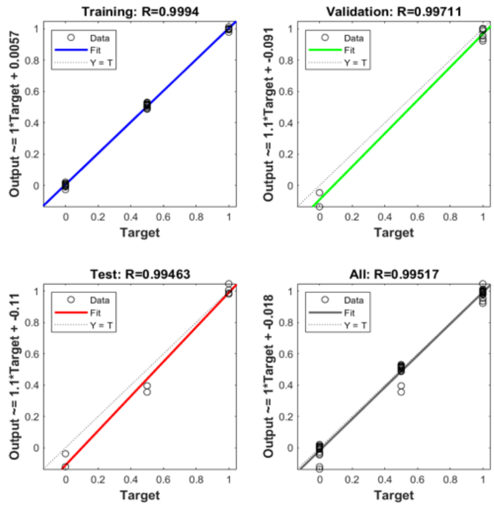


Fig.9 The Neural Network

The parameters that this machine learning model has run out was shown in the following diagrams.





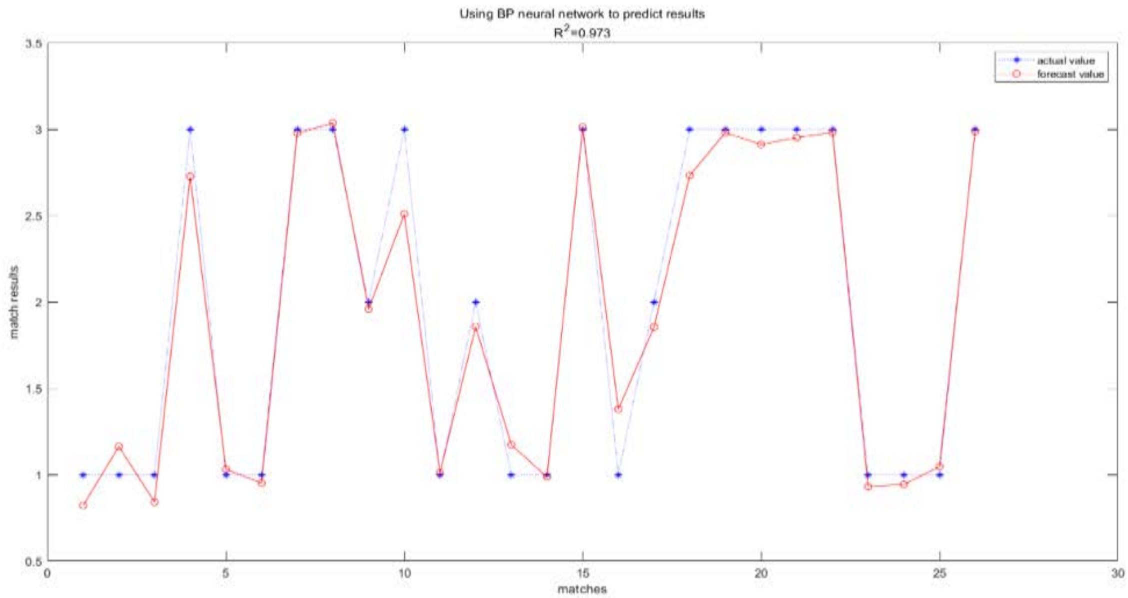


Fig. 10, 11 (a, b, c, d), 12 Parameters and Outcome

The results of the BP Neural Network are shown above. Since the results of the competition are discrete, So, we do the corresponding discrete operation of the predicted value. The predicted value near 1 is considered to be 1, It's the same for 2 and 3. We can see that all 21 sets of forecasts are perfectly matched to the real data.

5. Conclusion

1). Our model covers a lot of factors, including the team's network structure characteristics.

2). We use a scientific quantitative method rather than an inaccurate qualitative method to determine the mathematical expression of the evaluation system.

3). Using the model we've built to predict the outcome of the game can get a very high accuracy, indicating that the factors we choose are very representative, which shows that our model is very reasonable.

4). Requires a lot of data and can't give results quickly and easily

References

[1] Sun H J, Wang. S. R. FCM-based model selection algorithms for determining the number of clusters [J]. Pattern Recognition, 2004, 37(10): 2027-2037.

[2] Wang Feng. FCM-based model selection algorithms for [M].Taiyuan :Shanxi University, 2013.

[3] http://myportal.sxu.edu.cn/index.portal，2/8/2018.

[4] Journal of Wuhan Institute of Physical Education 2019,Vol.53. Competition Performance of Chinese Football Super League Tean under Background of Big Data

[5] Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D, & Malvaldi, M. (2015). The harsh rule of the goals: Data-driven performance indicators for football teams. 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 1-10, 7344823.

[6] WuJiang, HeChaocheng, GongZheng (2018) Research on the Decentralization Level and Performance of 2017 NBA Finals Teams Based on Weighted Directed Network Entropy

[7] Pappalardo, L., Cintia, P., Rossi, A. et al. A public data set of spatio-temporal match events in soccer competitions. Sci Data 6, 236 (2019).