ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 15, No. 2, 2020, *pp*.113-123

A Pattern Recognition and Performance Index Evaluation Model of Football Team based on Principal Component Analysis and PageRank Algorithm

Linjie Wu[1], Yujie Zheng[2] and Yunfei Fan[2]
[1]Department of Mathematics, Jinan University, Guangzhou, 510632, China
[2]Jinan University–University of Birmingham Joint Institute, Guangzhou, 510632, China
(Received May 10, 2020, accepted June 13, 2020)

Abstract. With the increasing knowledge integration and task complexity, individual ability demands a highly cohesive interdisciplinary team to amplify. To study the elements of successful team cooperation and explore valuable team strategies, this paper present a network pattern recognition model based on PageRank algorithm and principal component analysis method. Further, a team cooperation performance model based on group dynamic theory is built to capture the individual contribution and teamwork characteristic as a supplementary evaluation. By applying the model into football competition, we found our model has 73.68% accuracy, proving its outstanding adaptability. Based on the model, we can get the information of the inner network structure of a team, know the most significant contributors pertinent with team success, and make further justification plans and suggestions to achieve teamwork improvement.

Keywords: teamwork, network, cooperation performance, football

1. Introduction

In this era of more fierce competition and more miscellaneous challenges with high complexity, success is no longer only in favor of all-around individuals, but also more in favor of teams that are experts in division of work and cooperation. A successful team is always good at stimulating members' potentials and balancing their skills to work together as a whole to solve problems that are not attainable to an individual.

In competitive team sports, the significance of team strategies is self-evident. Football, particularly, requires 11 players and 3 substitutes, who are different in physical quality, technical expertise and carry out their own roles in the field to have extremely good cooperation between each other under the restraint of complicated rule and limited time, which is a rather demanding team work.

Due to high complexity of football and the changing situation of the game, finding a personalized but applicable way of football team cooperation is not only attractive but also challenging enough. Therefore, there is a need to build a specific controllable but adaptive model of team cooperation. This includes the analysis of the team's receiving and passing network, the establishment of group performance evaluation index system, we should also explore the value of the model in other team cooperation besides the controllable setting of such competitive team sports as football.

The Numbers Game [1] by Chris Anderson and David Sally studied how to use data analysis to understand the football game, arrange tactics and run club. At the end of the book, the authors predict that geometry "spatial, vector, triangular, and dynamic network" will be the focus of top-notch analysis. However, due to the complexity of football, the network analysis of football still needs to be developed for a long time.

E. Arriaza-Ardiles[2] analyzed the interactions between players in a football team from the point of view of graph theory and complex networks using data of a La Liga team. Clustering coefficient and centrality metrics (closeness and betweenness) are used to characterising the contribution of the players to the team. However, the paper only discusses the problem of passing without considering other contributions and lacks the analysis of game dynamics.

J. M. Buldú, J. Busquets, I. Echegoyen& F. Seirul.lo [3] analyzed the performance of Barcelona in the 2009-2010 season by using data on passing and possession. They focused on the temporal nature of football passing networks and calculated the evolution of all network properties along a match, instead of ignoring their average.

They found that the clustering coefficient, shortest path length, largest eigenvalue of the adjacency matrix, the algebraic connectivity and centrality distribution of Barcelona is better than the other teams. However, the variables selected for analysis in the paper may not be suitable for all teams. Similarly, the paper lacks sufficient research on the performance of teams over time. Further research is needed on which variables are more important for a particular style of team.

2. Models

2.1. Basic Model Description

In order to investigate the information of teamwork, we firstly use the pre-processed data of the a football team Huskies to establish a passing network and **Network Pattern Recognition Model** to learn its network features and then constructed a performance indicator system and Team **Cooperation Performance Model** to figure out key aspects for a team to success. After that, based on the result of models, we give several advice to the coach of the Huskies. Finally, we generalize the model out of the controlled setting of a team spot and into a normal kind of teamwork based on the **Group Dynamic theory**.

2.2. Data Pre-processing

2.2.1. Data Overview

The data is from problem D of The 2020 ICM(The Interdisciplinary Contest in Modeling) and can be downloaded on Internet. The Huskies have provided data with detailing information of last season, including all 38 games that they played against their 19 opponents. Overall, the raw data covers 23,429 passes between 366 players (30 Huskies players, and 336 players from opposing teams) and 59,271 game events (including duel, foul, free kick, goalkeeper leaving line, interruption, offside, others on the ball, pass, save attempt, shot and substitution).

2.2.2. Data Analysis and Pre-processing

Although most of the data is complete and reasonable, incomplete and abnormal data still often exists in a large amount of raw data, which may severely affect the efficiency of modeling and the accuracy of conclusions. It is therefore very important to pre-process the data. We listed the abnormal data and our corresponding pre-process methods and analysis as follows:

- Interference Data

After each shot, the coordinates of the destination of the ball are always (100, 100) or (0, 0). However, the goal situated at the middle near the edge of the field so that any destination of a successful shot (According to final score, it does exist) is impossible to be (100, 100) or (0, 0). Therefore these coordinates can be considered as interference data, which makes it difficult to identify which shot get scores and the deviation of each unsuccessful shot, which is a hinderance of modeling.

- Blank Data:

Blank missing data will affect the analysis process of the model, so it needs to be processed. We observe that the blank values in the data file generally appear in the four columns “EventOrigin_x”, “EventOrigin_y”, “EventDestination_x”, and “EventDestination_y”. Then we classified them into two kinds to tackle：

a) For the blank value due to data missing, we supplement it with the coordinate change before and after it. (Interpolation method)
b) For blank values due to "Substitution" such as column I, J, K, L in row 752 of “fullevents.csv”, , we will not fill it.

- Outliers

If one value in a set of data is more than twice the standard deviation of the average value, we call it the outlier. Statistically, we can use **block diagrams** to identify outliers. For the outliers, we use the average of the two adjacent observations to correct them.

- Wrong classification

Some data are classified wrongly in “fullevents.csv” and “passingevents.csv”, which leads to wrong selection of data if following the original classification standard. For example, in “fullevents.csv”, passes begin with Huskies_M8 were mistakenly classified as “Opponent7” in the column “team ID”（Obviously, it should be “Huskies”） in the 29th game of last season. We corrected these wrong classification.

2.3. Passing Network

Passing is not only a basic skill in football match, but also an vital way for players to interact with each other. High quality passing is not only a guarantee of effective attack but also has the function of direct attack and adjusting defense and attack. In addition, effective tactical cooperation is inseparable from high-quality passing. Therefore, passes worth investigation. Establishing the pass network and exploring its network identification pattern network attribute can help us analyze the team cooperation information effectively and deepen the understanding of the team cooperation in competitive games.

2.3.1. Establish Passing Network

According to graph theory, passing networks of the Huskies can be constructed from the ball exchange between players, where network nodes (or vertices) are players of the Huskies and links (or edges) account for passes between any two players of the team. Based on the husky's passing data obtained from the file "passingevents.csv", we draw the passing network map for the first game of last season of the Huskies by Gephi as shown in Fig. 1.

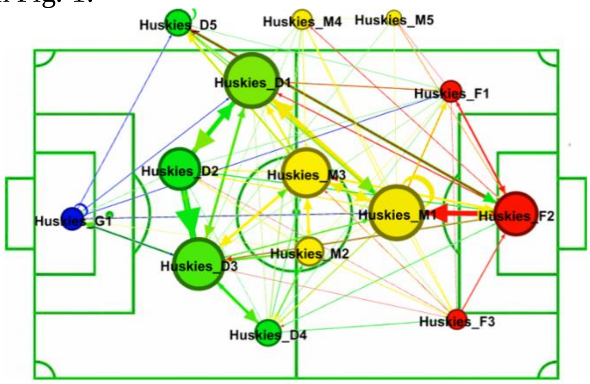


Fig. 1: Passing network of the first game of Huskies

In Fig. 1, there are 14 vertices and 111 edges in total, indicating that 14 players (including substitutes) had 111 times of passing in this game.

According to the game tradition of FIFA football games [4], the **colors** of defenders, midfield players, forwards and goalkeepers are respectively green, yellow, red and blue. (Note that the players drawn off the pitch are substitutes for this game.)

The **size** of each player's point is arranged according to the player's weighted outdegree (i.e., the number of successful passes of the player). The larger point represents that the player has more successful passes in this game

The **thickness** of the edge is determined by the number of successful passes between the two players as weight. It is worth mentioning that every edges in the picture are directed sides, a thicker side indicates that the player at the starting point gave the player at the end more successful passes.

2.3.2. Network Pattern Recognition(Formation Identification)

In football, each team has their own formation, which regulates the number of defenders, midfield players and forwards. For example, common formations are like 433(4 defenders, 3 midfield players and 3 forwards) , 442, 532 and so on. As a kind of tactical strategy, players will stand according to their position in the formation. Here, team formation can be considered as one kind of important network pattern.

Assume that in the process of competition, the team members will basically keep good formation. For instance, the forward player is generally in front of the midfield player. Then theoretically, we can identify the formation of the team according to their position. But in the actual process of the game, the difficulty of distinguishing the formation will be greatly increased when the players run. Therefore, we think it is necessary to establish a model according to the existing data to obtain the network mode of delivery network, which is the following **Network Pattern Recognition model**. Combine the model with the circumstance of Huskies, we have the following steps:

- Step 1: PageRank Algorithm:

  By PageRank algorithm, we can calculate the PageRank value (PR value for short) of each node based on the constructed passing network model. The PR value of player i (PR_i) is used to measure the importance of the node. By calculating the PR value of each node, we can measure the importance of each player in the current network, and then combine it with his approximate position on the field to attain his position in the formation.

Calculation:

(initial step) Given that the initial PR value of all nodes, $PR_i(0), i = 1,2, \dots, N$ satisfy:

$$\sum_{i=1}^{N} PR_i(0) = 1 \tag{1}$$

Correct PR value by the rule : Divide the PR value of each node in step k-1 equally to the node it points to so that each node's new PR value is corrected to the sum of the PR values it receives, i.e

$$PR_i(k) = \sum_{j=1}^{N} a_{ji} \frac{PR_i(k-1)}{k_j^{out}}, i = 1,2, \dots, N. \tag{2}$$

In the formula, k_j^{out} is out-degree of node j.

- Step 2: Unitization

Unitize the player's average position coordinates and the coordinate after unitization is denoted as $\overline{Co_i}$.

- Step 3: Fuzzy comprehensive evaluation method (FCE)

Use FCE to identify the weights of position coordinates and PR value, the outcomes are 0.7 and 0.3 respectively. Finally, we calculated the PR value score Sc_i of each team member by

$$Sc_i = 0.7 * \overline{Co_i} + 0.3 * PR_i \tag{3}$$

- Step 4: Identify Formation Identification Rules

After performing the previous steps on the data of the first game of the Huskies, we obtain the table 1.

Table 1: Different Players' Sc(i) in Match1

MatchID	TeamID	OriginPlayerID	AVEEventOrigin_x	PRvalue	Sc(i)
1	Huskies	Huskies_D1	34.32692308	0.147	2.549885
1	Huskies	Huskies_D2	34.36111111	0.101	2.506278
1	Huskies	Huskies_D3	38.83333333	0.145	2.863333
1	Huskies	Huskies_D4	52.30769231	0.061	3.722538
1	Huskies	Huskies_M1	52.07692308	0.157	3.802385
1	Huskies	Huskies_M2	50.1	0.053	3.56
1	Huskies	Huskies_M3	46.6	0.125	3.387
1	Huskies	Huskies_F3	64.875	0.034	4.57525
1	Huskies	Huskies_F2	59.71052632	0.128	4.307737
1	Huskies	Huskies_F1	74.36363636	0.046	5.251455

Table 2: Relationship between Score and Position

Role	Position	PR value	Score
Forward	Closer to the goal than other players	Relative high	Greater than 4
Midfield	Middle of the field	Medium	Between 3 and 4
Guard	Back position	Relative low	Less than 3

According to the known data provided by the Huskies, D, M, F, etc. are used to indicate the players' position in the formation. As seen from the table, the average x-axis coordinate of the forward (F) is the largest, that of central defender (M) is between F and D, and the defender (D) is generally the smallest.

In addition, we observe that the value of D4 is relatively large, so we calculate the average y value of its catch point, which is equal to 9.615. This indicates that D4 situated at rightward position on the court. Combining with the fact that left and right wing guards of a strong team will generally move forward to participate in some attack (that is, the average x value will larger), it is concluded that D4 is more likely to be a right back.

Integrating the known player position of the first game (corresponding team formation is 433) and the data in the table 3, we derive the role identification rules as shown in table 3, which can tell us the team formation from the data if unknow the D, F, M.

Combing the roles with their corresponding number, we obtain the formation of the football team.

- Step 5: Model verification

We can verify the model with the data of the second game:

Table 3: Different Players' Sc(i) in Match2

MatchID	TeamID	OriginPlayerID	AVEEventOrigin_x	PRvalue	Sc(i)
2	Huskies	Huskies_D1	33.69565	0.136364	2.495059
2	Huskies	Huskies_D2	31.5	0.066434	2.271434
2	Huskies	Huskies_D3	30	0.132867	2.232867
2	Huskies	Huskies_D4	34.6667	0.083916	2.510585
2	Huskies	Huskies_M1	44.2069	0.153846	3.248329
2	Huskies	Huskies_M2	43.5468	0.031469	3.079745
2	Huskies	Huskies_M3	43.94118	0.101399	3.177281
2	Huskies	Huskies_M4	47.88889	0.055944	3.408166
2	Huskies	Huskies_F1	73.33333	0.101399	5.234732
2	Huskies	Huskies_F2	54.94737	0.136364	3.982068

It can be seen from the calculation results in the table that the values of Sc_i of defenders (D) are all below 3, that of midfield(M) are between 3 and 4 and frontiers(F) are mostly greater than 4. Although for frontiers, one of them is less than 4, it's still very close to 4 and there is a certain gap with the value of M. We can tell that the formation is 442 in the second game. Therefore, through calculation and verification, we conclude that our model has high robustness.

What's more, using the known model to calculate the remaining games, the results are also almost the same as the known positioning.

2.3.3. Dyadic and Triadic Configuration

The dyadic configuration of the network reveals the relationship of pairs of players and the triadic configurations tells information of relationship among three players. As a measure of the dyadic and triadic configuration, we can calculate the **degree centrality** of nodes.

Among N nodes of a network, the degree centrality of each node is:

$$C_D^i = \frac{k_i}{N-1} \tag{4}$$

In the formula, k_i refers to the degree of node i , where $i \in [1, N]$ and $i \in \mathbb{Z}$.

According to the research of Hongren Yan et al in the paper "unsupervised feature selection based on the degree centrality of nodes"[5], if there are two nodes with degree centrality $C_D^i > \theta$, it is dynamic configurations. Similarly, if there are three nodes with $C_D^i > \theta$, it is triadic configurations. According to the article, we obtained the threshold value θ = 0.8.

Taking the first game as an example, we calculate the degree centrality of each node. The results show that there are two nodes whose C_D^i values are greater than θ, which are Huskies_M1 and Huskies_F2 respectively. Therefore, they are dyadic configurations, showing that these two players have close relationship.

2.4. Performance Indicator System and Strategy

Good teamwork is the key factor for a team to succeed. The success of a team is far more than the sum of the abilities of individual players. In fact, it is based on many other factors apart from individual capacities, among which the most significant is the overall team performance and cohesion. In order to capture best team work structure and configuration, we proceed from the determination of team performance indicators to evaluate a team.

In recent years, research has found that the social network of a team has an important impact on the performance of a team, which makes the research on the social network of a team gradually become a research hotspot. In the research on "The relationship between team social network and performance" by Yang Shangjian and Sun Youping, [6] team performance, personal performance, team relationship, team identity and helping team-mates are chosen as performance indicators. However, these indicators are abstract, difficult to collect data, and not conducive to analysis. Therefore, after collecting data from many ways, we have established a set of concrete performance indicator system of team cooperation according to the following indicators, which can reflect the aspects that Yang Shangjian and Sun Youping mentioned . After that we also considered other team level processes, so as to establish Team cooperation performance model.

2.4.1. Performance Indicators

We picked out several representative and concrete factors to measure the degree of success of teamwork of the Huskies:

i. **Offense:** In football, attack is mainly embodied by the probability of shot. Therefore, offensive ability directly determines the final scores of the game so that it has nonnegligible and irreplaceable places in teamwork performance evaluation. Thus we picked the ratio of the shooting amount of all team member of Huskies in the match field to the passing amount of the two teams in the whole match to represent the attacking aggressiveness, denoted as $\boldsymbol{\alpha}$.

ii. **Defense:** Corresponding to the offense, the purpose of defense is to prevent the other side's goal. Therefore, defensive ability ultimately determines the outcome of the game and that's why it is necessary to be included in the evaluation indicators. Practically, it can be reflected in various competitions of players, such as air duel, attack duel, defending duel and so on. Here we denote the number of air duel, attacking duel and defending duel respectively as $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ *and* $\boldsymbol{\epsilon}$.

iii. **Skills:** Skills like interruption, shot, free kick and so on constitute the main content of football game and are basic factors of team capability. Here we picked out player's passing success rate, denoted β and the ratio of player's passing amount on the match field to the passing amount of two teams in the whole match, denoted as $\boldsymbol{\gamma}$.

iv. **Player contribution**: This is an indicator that reflect the contribution of different players to the whole team. We will use **Principal Component Analysis** to analyze it from the perspective of shot, successful pass, successful catch, duel and total pass. From this, we can see how evenly different players in the team contribute and whether or not exist circumstance that some players are lazy and some are overfatigue.

v. **Team Style and tactical skills**: refer to the strategies that guide or determine the overall situation, the specific deployment of operations, and the strategy of conquering the opponents. It's importance to the success of a football team is self-evident. According to the players' coordinates of every event provided by the Huskies team, we can determine the location area of events of the Huskies team. Then we can roughly simulate the style and play of the team.

2.4.2. Team Cooperation Performance Model

The general idea of building the team cooperation performance model can be divided into two part. The first part focuses on the pertinent values of different event (from point i to iv) in the game, obtaining a algebraic indicator G. The second part pays attention to the team style (point v) by obtaining a geometric intuitionistic map of football passing hot spot.

<u>Part 1: analyze and calculate G</u>

The general idea of building the team cooperation performance model can be divided into two part. The first part calculate $\boldsymbol{G_1}$ to summarize the features of defense, offensive and skills of the football team while the second part use SPSS software to do **principal component analysis**, calculating $\boldsymbol{G_2}$, which convert the player contribution into an indicator with a few representative factors. Finally, we add them up to get the final teamwork performance indicator G.

- **Calculating $\boldsymbol{G_1}$**

Now we will pack the first six factors up since they have explicit value and can be directly integrate together. We combine the abstract harmonic mean of these six indicators, and sum up these behavior of a team into a single value to get the intermediate quantity $G1$ [7] of Team cooperation performance model as follow:

$$G_1 = \frac{6}{\frac{1}{\alpha^2} + \frac{2}{\beta} + \frac{1}{\gamma} + \frac{1}{\delta} + \frac{3}{\varepsilon} + \frac{1}{\epsilon}} \tag{5}$$

- **Calculating $\boldsymbol{G_2}$**

As an important aspect of team cooperation performance indicators, player contribution allocation can be analyzed in detail by using algorithms. We once considered **TOPSIS method** as a comprehensive evaluation method to analyze the contribution of players. But this method has a great influence on the player ranking with high quality and low quality indexes so that this method cannot accurately explain the level of player performance.

According to the characteristics of indicator performance and team level process, we use **principal component analysis** to analyze. Principal component analysis [8] is a kind of mathematical statistical method which transforms the mathematical statistics of multiple indicators into a few representative indicators through the idea of "dimension reduction", and solves the optimization problem of high-dimensional system with the idea of dimension reduction.

The main steps of the principle component analysis are concluded as shown in the following flow chart 1

Flow chart 1

Firstly, we load the data of all players of Huskies, which includes 5 indicators of performance of 30 players in this football season into the matrix X. Denote the number of players as p=30.
Then we do dimensionless process X. The formula comes as follows:

$$\overline{X_P} = \frac{\sum_{d=1}^{n} X_{dp}}{n} \qquad p = 1,2,\ldots. \tag{6}$$

Then we calculate the standard deviation of our sample by:

$$S_p = \sqrt{\frac{\sum_{d=1}^{n} X_{dp}}{n}} \qquad p = 1,2,\ldots. \tag{7}$$

Calculate its correlation coefficient matrix R by:

$$R = \frac{X'^{T} X'}{n-1} \tag{8}$$

According to the sample correlation coefficient matrix R, we can calculate the characteristic equation of the correlation coefficient matrix, and the corresponding eigenvalues $\lambda 1 \geq \lambda 2 \geq \ldots \geq \lambda m$, $m = 30$. According to the variance eigenvalue, the variance contribution rate is calculated as:

$$a_i = \frac{\lambda_i}{m} \tag{9}$$

Based on the team performance indicators established as above and the existing data, we will determine the following five evaluation indicators accordingly: shot, successful pass, successful receive, due, all pass. Then use SPSS software to analyze and calculate it. The specific calculation process is as follows:

Table 4

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	%of Variance	Cumulative %	Total	%of Variance	Cumulative %
1	3.813	76.265	76.265	3.813	76.265	76.265
2	1.017	20.334	96.599	1.017	20.334	96.599
3	.147	2.934	99.534			
4	.020	.404	99.937			
5	.003	.063	100.000			
Extraction Method: Principal Component Analysis						

Table 5

Component Score Coefficient Matrix		
	Component	
	1	2
shoot	.151	.780
successful pass	.245	-.336
successful catch	.257	-.151
duel	.229	.365
All pass	.247	-.326

Therefore, the expression of the factor can be obtained based on the component score matrix.

$$\mathrm{F}_1 = 0.151 * \mathrm{shoot} + 0.245 * \mathrm{successful\ pass} + 0.257 * \mathrm{successful\ catch} + 0.229 * \mathrm{duel} + 0.247 * \mathrm{all\ pass} \tag{10}$$

$$F_2 = 0.780 * \mathrm{shoot} - 0.336 * \mathrm{successful\ pass} - 0.151 * \mathrm{successful\ catch} + 0.365 * \mathrm{duel} - 0.326 * \mathrm{all\ pass} \tag{11}$$

Take the variance contribution rate of each factor as the coefficient, and multiply each factor by the coefficient to get the final team performance index evaluation system:

$$F = 0.3813 * F_1 + 0.1017 * F_2 \tag{12}$$

Use the above teamwork performance evaluation formula, we get the scores of the Husky team's players. (Only some of the rankings are listed below, see the appendix for the total rankings)

Table 6

Player	F_1	rank1	F_2	rank2	F	rank3
Huskies_M1	1116.33	1	-787.92	30	345.526	1
Huskies_F2	843.058	2	-587.21	29	261.739	2
Huskies_D1	787.29	3	-521.5	27	247.157	3

Based on the score F_i of each player and the average score $\bar{F}$, we calculate the variance G_2 by:

$$G_2 = \frac{1}{\frac{\sum(F_i - \bar{F})^2}{N}} = \frac{N}{\sum(F_i - \bar{F})^2} \tag{13}$$

- **Combine G_1, G_2**

Finally we can get the final indicator G by substitute the value of G_1, G_2 into formula of the following final formula:

$$G = G_1 + G_2 \tag{14}$$

Part 2: Football passing hot spot map

Based on the players' coordinates of every event during the whole season, we can determine the location area of events of the Huskies team.

According to the known coordinates, we can set up a Cartesian two-dimensional coordinate system with the football field as the coordinate system, the right corner of the husky team as the origin, the length of the football field as the x-axis and the width as the y-axis.

Then divide the length and width of the football field into 10 parts so that the whole football field is divided into 100 even areas (10 * 10).

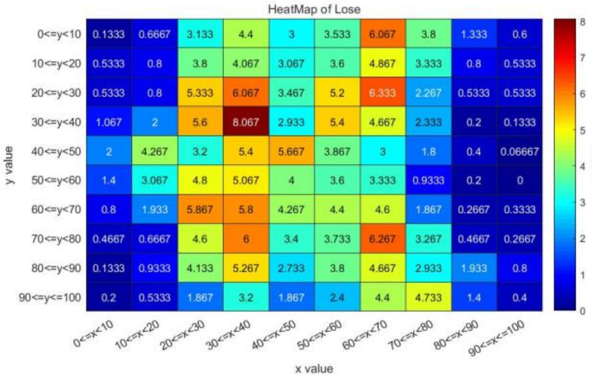
Now we count the total number of passes in each area of the game in which the husky team wins, draws and loses according to the data given of the Huskies. Then we can integrate these numbers into a matrix and we get three matrices $U_{win}, U_{tie}, U_{lose}$.

Among them, $U_{win} = [uw_{ij}]$, where uw_{ij} is the number of passes between $10 * (i-1) < x < 10 * i$ and $10 * (j-1) < y < 10 * j$.

Then we deal with three matrix matrices with nondimensional treatment. For example, let's set

$$vw_{ij} = \frac{uw_{ij}}{n} \tag{15}$$

where n is the numbers of matches of win.
Then we get $V_{win} = [vw_{ij}]$.Similarly, we can calculate $V_{win}, V_{tie}, V_{lose}$.
Via Matlab, we can draw the hot spot areas of football passing in win, lose and draw situations into heat chart form, as follows



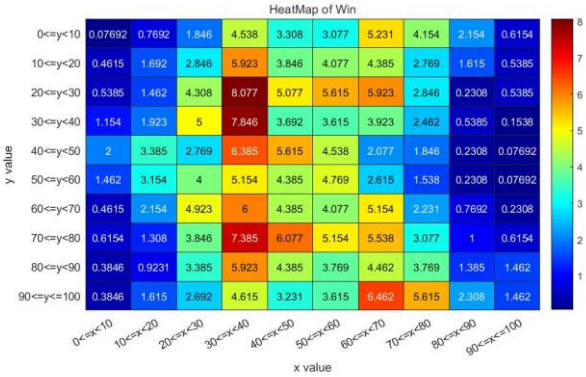


Fig. 2 Fig. 3: Match weight (total number of passes) for each of the 100 zones of the football field. Red color shows a higher density of game play, while blue shows a lower density.

From the passing hot spot chart, we can roughly see the main areas of husky team's ball control and holding cooperation on the field. It's easy to see that the team's heat map changes significantly with the result of the game.

We can see that compared with the middle way cooperation and promotion, the Huskies prefers to attack from both wings.

Comparing the winning and losing hot spots, it can be seen that when the Huskies loses, players pass more backcourt backfoot (x < 30 area), while when the Huskies wins, it often launches more attacks on the right side. When losing, the right side is often not very active.

However, in any case, we must keep in mind that the football passing hot spot map only talks about passes. That is why this number will not necessarily reflect the effectiveness or transcendence in game-play. A player can present a low participation, according to this criterion, but nevertheless he could be a key player in defensive functions, determinant for creating play space, or extremely efficient for scoring. [2]

3. Outcome and Evaluation

3.1. Outcome Analysis

In part 1, by **principal component analysis**, we can conclude [9] two main factors, shot and successful pass, from the Table 5 since these two factors whose eigenvalues are not less than 0.5, and the cumulative contribution of the two factors is 96.599% of the total contribution, which is extremely high. And the other three factors, successful receive, due and all pass have little effect on the team's performance.
So we can infer from the above conclusion that:

- In order to reach team success, the shooting training of players should be solid, and it is very important to improve the overall hit rate of the team.
- In order to reach team success, pass training is also significant. A successful team should pay attention to cultivate pass awareness and pass accuracy of each teammate. Only when the whole team has a high successful pass rate can the team reduce the probability of being snatched by the opponent, improve ball possession and finally better assisted scoring.

In part 2, we can conclude that

- The Huskies is relatively weak in arranging attacks when having large number of backcourt backfoot.
- In order to reach team success, they should pay more attention to catch the opportunity in the right side.

3.2. Experimental Results

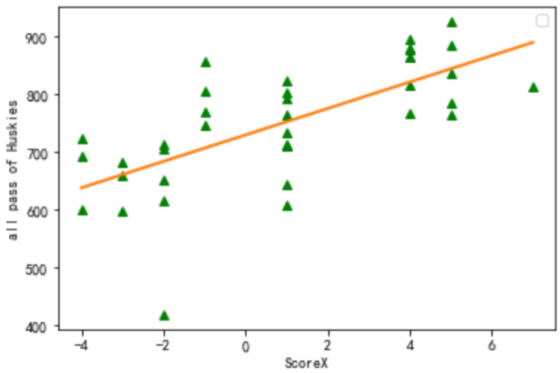
Passing ability: In the analysis of the previous team cooperation model, we concluded that "successful pass" and "all pass out" are the main factors for the performance on the field. Next, the correlation between the number of team passes and their victory in the game will be analyzed[5]. Because there are fewer goals in a game, there will be some errors in the direct analysis. Therefore, for the score of each game, we refer to the football game's points method, perform data processing, and then perform correlation analysis.[7]

Table 7 Match Points Rules

Win	Tie	Loss
3	1	0

Score calculation formula: $\text{Score}_X = \text{Score}_{\text{integral}} + \text{Score}_{\text{Goal difference}}$

In the formula, Score_X refers to Score of X, $\text{Score}_{\text{integral}}$ refers to Score of integral, $\text{Score}_{\text{Goal difference}}$ refers to Score of goal difference.



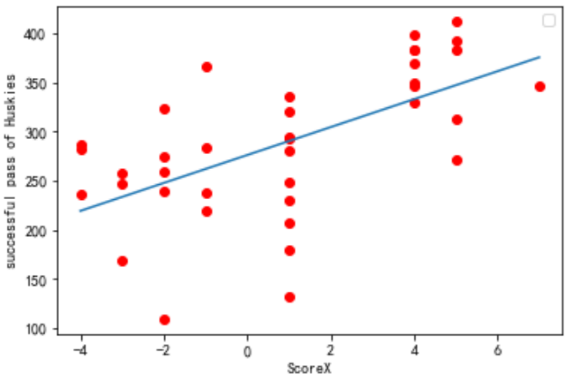


Fig. 4: (left) Figure-Correlation between scoring and Huskies' pass

Fig. 5: (right) Figure-Correlation between scoring and successful pass of the Huskies

As can be seen from the figure, the score is roughly proportional to the number of passes and successful passes of the Husky team. This shows that the team's passing activity is related to its success in the game. Because teams with higher passing activities tend to score more, score more opportunities and score higher. Therefore, the coach can formulate a corresponding training plan according to the passing characteristics of each person in the passing network.

3.3. Sensitivity Analysis

In the previous analysis, we established the Team coordinate performance model. In this section, we will analyze the sensitivity and reliability of the model. To do this, we use this model to simulate football matches by performing the following experiments.

After the modeling analysis of the previous model the Team coordinate performance model, we got the model formula：

$$G = \frac{N}{\sum(F_i - \bar{F})^2} + \frac{6}{\frac{1}{\alpha^2} + \frac{2}{\beta} + \frac{1}{\gamma} + \frac{1}{\delta} + \frac{3}{\varepsilon} + \frac{1}{\epsilon}} \tag{16}$$

We use this function to calculate the G value of each game to get the following figure

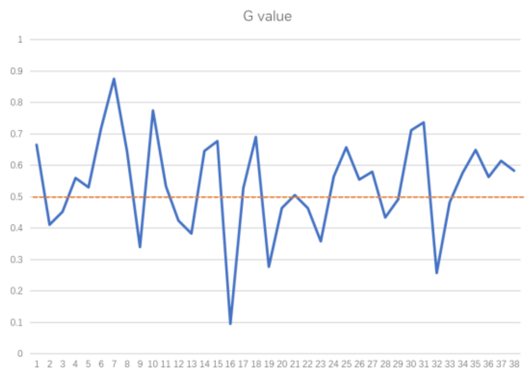


Fig. 6: G-values in 38 games

In the figure, if the score of a certain game is high, it means that the team played well. If the score is not high, it means that the team performs bad. We add a straight line at 0.5 in the picture to distinguish the performance of each match and use this to predict the final result of the match. That is, if G value ≥ 0.5, the team wins. We then compared our predictions to the original match results. Because tie is a relatively difficult case to predict, we classify it as loss.

The final comparison results show (see appendix) that we have 28 games with correct predictions and an accuracy rate of 73.68%. Therefore, our model can be considered to have high enough sensitivity.

4. Conclusion

With the increase of more complex and interdisciplinary challenges, the society has become more interconnected. It somehow indicates that team work will play a more and more significant role in the future. Compared with hardworking individuals, it seems that hard work favors teams that are good at division of labor and cooperation, stimulate members' knowledge and skills to work together and solve problems. He who see through the mystery of teamwork can get closer to success.

By our constructed model Network Pattern Recognition Model and Team Cooperation Performance Model, we can get the information of the inner network structure of a team, know the most important contributors pertinent with team success, and make justification plans and suggestions based on the outcome of the model.

Although the model is constructed from the data of a certain kind of team work, football, the model is still highly adaptable and accurate according to the sensitivity analysis. It has many noneligible and irreplaceable strengths and advantages so that it’s still an effective tool when analyzing team work.

References

[1] Chris Anderson , David Sally, The Numbers Game, Penguin,2014

[2] Arriaza-Ardiles, E., Human Movement Science (2017). Applying graphs and complex networks to football metric interpretation

[3] Buldú, J.M., Busquets J. and Echegoyen, I. (et al. (2019). Defining a historic football team: Using Network Science to analyze Guardiola’s F.C. Barcelona. Sci Rep, 9, 13602.

[4] https://www.ea.com/en-gb/games/fifa/fifa-20

[5] Yan hongren, ma guoshuai, qian yuhua. Unsupervised feature selection based on node degree centrality. Data collection and processing,2019,34(02):312-321

[6] Yang shangjian, sun youping. Research on the relationship between team social network and performance. Journal of tianjin university of physical education,2015,30(02):99-103.

[7] 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.

[8] Man shuai, longhua, xiong xin, li yimin, liu linxuan. Evaluation of NBA players' comprehensive ability based on principal component analysis. Software guide,2018,17(06):185-189+197

[9] http://dx.doi.org/10.1016/j.humov.2017.08.022

[10] Yin tianxiang. Factors influencing NBA team performance based on principal component analysis. Stationery and technology,2014(02):171.

[11] Zhang tao, yu jiong, liao bin, guo binglei, bian Chen, wang yuefei, liu yan. Graphx-based pass network construction and analysis. Computer research and development,2016,53(12):2729-2752.

[12] Hong zaikuan. Attaching importance to training football players' observation ability on the field. Sports science and technology,1980(06):32-34.

[13] Liao xiaoyan, Chen qi. Using the team member model to build an effective team. Market modernization,2006(29):270.

[14] Lu bo. A study on the relationship between position role and tactical rhythm of football players. Soochow university,2016.

[15] Kurt Lewin, Principles Of Topological Psychology, Munshi Press,2008

[16] Donelson R. FORSYTH, Group dynamics seven editions ,Wadsworth; International ed edition ,2019.