Team Dynamic Structure and Strategy Evaluation: Mathematic Model based on Decision Trees

Zefeng QIAN[1], Xietian WANG[1], Yiru ZHANG[1], Feng YE[2]

(1.School of Telecommunication and Engineering, Xidian University, Xi' an, Shaanxi, China 710126)

(2.School of Mathematics and Statistics, Xidian University, Xi' an, Shaanxi, China 710126)

Summary

In this paper, we developed some models to reflect the dynamic configuration of an anonymous soccer team (Huskies team) and evaluate team performance indicators.

First, a ball passing network is built by analyzing the exchanges between players. We counted the passing events and positions of players and looked at their passing preferences and positions. Each soccer player is treated as a node and the number of passes is the link between any two players. In this way, we can establish weighted and unidirectional football delivery networks.

Then, using the ball passing network team dynamic model analysis, including the team defense and attack chain configuration, binary and ternary configuration to build our models.They will be more flexible, being the core of the offensive or defensive team. They will be more flexible, being the core of the offensive or defensive team. In addition, we also analyzed the performance of Huskies from individuals to teams, and established five indicators to evaluate individuals, forming five-dimensional ability diagrams. Analyze the favorable factors for the team from the team position and side.

Next, a Comprehensive Evaluation Index model is designed. By using PCA analysis method to reduce the dimension of five individual indexes, the individual comprehensive evaluation indexes are obtained. For the configuration teams at different positions (Forward, Midfield and Defense), the contribution coefficients of the players to the position were calculated respectively, and the comprehensive evaluation data were obtained. Based on the individual comprehensive score and the contribution distribution of teams in different locations, we can get the comprehensive evaluation index of the team.

In order to verify the accuracy of the model, we established three types of decision trees for prediction, and the accuracy is close to 80%.Since our strategy is based on a sample of data from the season, the winning strategy for the Huskies is a universal strategy,

Finally, based on the above research, we proposed the changes of huskies in the next season from the aspects of team configuration, individual contribution index, decision tree model and passing chain position. Combined with the actual situation of society, the model is generalized. Besides, 5W1H analysis method is used to analyze the interaction and cooperation process within the team, so as to build an efficient team.

Keyword: Soccer, Team Strategy, Mathematic Model, Decision Tree

Author Description:

Zefeng QIAN, male, born in 1999, undergraduate, research area: artificial intelligent, data analysis. E-mail: zfqian@stu.xidian.edu.cn

Xietian WANG, male, born in 1999, undergraduate, research area: data analysis, digital signal processing. E-mail: xtwang_53@stu.xidian.edu.cn

Corresponding Author:

Feng YE, Associate Professor, Faculty of Applied Mathematics.

2000, Graduated from Mathematics Department of Northwest University with a master of Science degree;

2012, obtained a doctorate in Applied Mathematics from the school of mathematics and statistics, Xidian University;

2000, taught in Mathematics Department of Xidian University;

From May 2015 to May 2016, the national government sent an exchange visit to the University of Texas (Dallas) for one year;

Mainly engaged in optimization theory, algorithm and application, intelligent information processing, machine learning, support vector machine and other aspects of research.

Contents

1. Introduction

1.1 Background

Team is a term and concept we are all familiar with. There is no doubt that team needs the cooperation of individuals. In competitive team sports, the success of a team involves many factors, including team structure, the role of the players, whether the team has diverse skills, the ability to effectively coordinate the team, how the team balances individual and collective skills, and so on. In a rapidly developing society, how to improve the overall effectiveness of the organization and the best strategy for team building become the key challenges we face.

The current huskies coach asked to explore how the complex interactions between players on the field affect their success. Our goal is not only to study the interactions that lead directly to scoring, but also to explore team dynamics throughout the game and throughout the season to help identify specific strategies that can improve teamwork next season. We can explore the best strategies for teams by quantifying and formalizing the structural and dynamic characteristics of success (and failure).

1.2 Restatement of the problem

To explore the dynamics of the team throughout the game and the season, and identify specific strategies that can improve teamwork for the next season, while the following tasks are to be accomplished :

- Build a passing network with each player as the node and each pass between players as the chain. Use the delivery network to identify network patterns. At the same time, observe and analysis from micro to macro, a game to the whole season of multi-angle interaction.

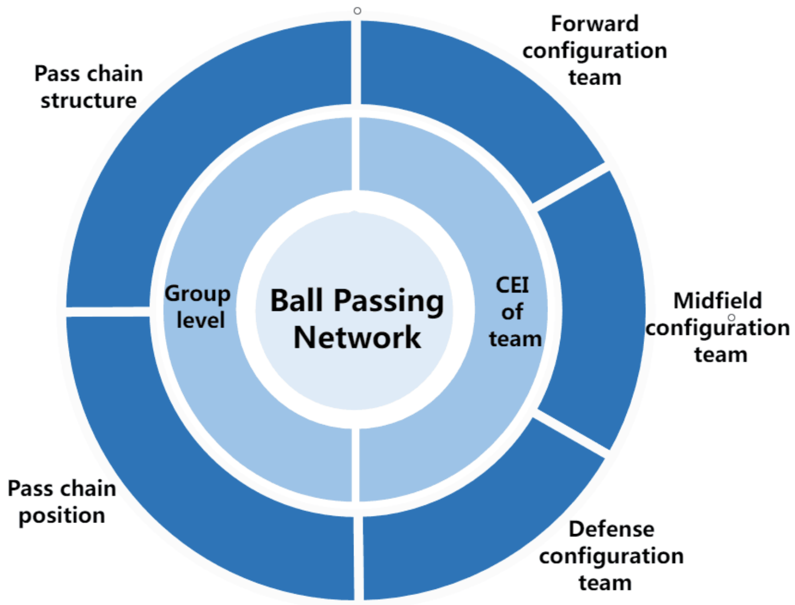
- Determine the scoring indicators of team cooperation, reflecting team performance and team performance, and build a model to capture the structure, configuration and dynamic aspects of team work.

- Based on the above network analysis, determine the structure strategy for next season and suggest changes that the coach should make to improve the team's success.

- Explore how to design more effective teams in conjunction with the wide range of issues related to teams in society. In addition, in order to develop general model of team performance, we also need to know what aspects of the team cooperation.

1.3 Views of our work

In this paper, we do the following work. The transmission network and comprehensive



evaluation indexes are established, as shown in the figure below:

Figure 1: Mind mapping

2. Preparations of the models

2.1 Assumptions

- All Huskies players' ability indicators are reflected in the player statistics for the season, ignoring the uncertainty of injuries or poor performance. Since our model is a generalized model for the Huskies team, we need comprehensive data for all players, and chance is not necessary, so this assumption is reasonable.

- We assume that the Huskies team members perform their own duties on the court, which is not purely a case of players switching roles. In practice, the probability of professional players changing roles on the field is very small, so our hypothesis is reasonable.

- It is assumed that the influence factors of Huskies team in each match are only related to the players' comprehensive ability and the way of cooperation within the team, while ignoring the players' psychological factors and other unstable factors. Such instability is out of control in real life, so the assumption is reasonable.

2.2 Notation Description

Table 1: Notations

Notation	Description
R	correlation coefficient matrix
λ	the eigenvalue
F	individual comprehensive evaluation index
C	judgment matrix
D	paired comparison matrix
α	contribution coefficient
Z	team comprehensive evaluation index
m	principal component number

3. Ball Passing Network

In the past decade, network science has become one of the most active fields in applied physics and mathematics. A football team can be thought of as a complex network in which nodes interact to overcome an opponent network. The passing network is constructed by analyzing the ball exchanges between players, treating each soccer player as a node and the number of passes as a link between any two players. In this way, we can build weighted and unidirectional football delivery networks.

3.1 The Passing preference of Players

To identify the ball passing events, we have to find the exact pattern of how often a player pass ball to his fellows, and how often he gets the ball passed by his fellows. We count the passing ball event of players, and build a network to evaluate the weight of one player in ball passing chain (Table 2), and find that each player have their own ball passing preference.

Table 2: The passing ball weight of player in the whole season

	D1	D2	D3	D4	D5	D6	D7	D8	F1	F2	F4	F5	F6	G1	M1	M12	M3	M4	M6	M8	M9
D1	0.0	0.3	0.6	0.4	0.3	0.1	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.6	0.5	0.1	0.3	0.1	0.1	0.0	0.0
D2	0.3	0.0	0.3	0.2	0.3	0.3	0.1	0.0	0.1	0.3	0.0	0.0	0.0	0.3	0.3	0.0	0.2	0.1	0.1	0.0	0.0
D3	0.5	0.2	0.0	0.3	0.1	0.0	0.1	0.2	0.1	0.3	0.0	0.1	0.1	0.7	0.3	0.0	0.3	0.2	0.1	0.1	0.0

D4	0.3	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.4	0.1	0.2	0.1	0.1	0.3	0.1	0.3	0.2	0.2	0.0	0.0
D5	0.2	0.2	0.1	0.0	0.0	0.1	0.0	0.0	0.2	0.5	0.1	0.0	0.0	0.1	0.4	0.1	0.4	0.2	0.4	0.0	0.1
D6	0.1	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.2	0.0	0.0	0.0	0.2	0.2	0.0	0.2	0.1	0.0	0.0	0.0
D7	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.2	0.2	0.1	0.0	0.1	0.2	0.3	0.0	0.1	0.2	0.1	0.2	0.0
D8	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.3	0.1	0.1	0.1	0.0	0.2	0.0	0.0	0.0
F1	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.1	0.1	0.1	0.0
F2	0.3	0.2	0.3	0.4	0.5	0.2	0.2	0.2	0.2	0.0	0.1	0.1	0.1	0.0	0.6	0.1	0.3	0.3	0.3	0.1	0.1
F4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.0
F5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.0
F6	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.0	0.0
G1	0.4	0.2	0.3	0.1	0.1	0.1	0.1	0.0	0.4	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.0
M1	0.5	0.3	0.3	0.4	0.4	0.3	0.4	0.3	0.2	1.0	0.2	0.1	0.2	0.1	0.0	0.1	0.8	0.4	0.4	0.1	0.1
M12	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
M3	0.4	0.3	0.3	0.5	0.4	0.2	0.2	0.1	0.1	0.3	0.1	0.1	0.1	0.1	0.9	0.2	0.0	0.1	0.2	0.1	0.1
M4	0.2	0.1	0.1	0.2	0.1	0.1	0.2	0.1	0.1	0.3	0.1	0.0	0.1	0.0	0.4	0.0	0.1	0.0	0.2	0.1	0.0
M6	0.1	0.1	0.1	0.3	0.4	0.0	0.1	0.0	0.2	0.3	0.1	0.1	0.1	0.0	0.3	0.1	0.2	0.1	0.0	0.1	0.1
M8	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
M9	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0

Among them, column is the destiny of passing ball events, row is the originate of passing ball events. For example, we find (1, 2) = 0.3, (1, 3) = 0.6, which means that D1 seldom pass ball to D2 as well as D1 sometimes pass ball to D3. From the table, we identify the ball passing rate for every player. We also find that some of them would love to pass ball to others, while others would always get ball passed to them. In this way, we have the ability to identify the possible position of each player, and how they pass the ball.

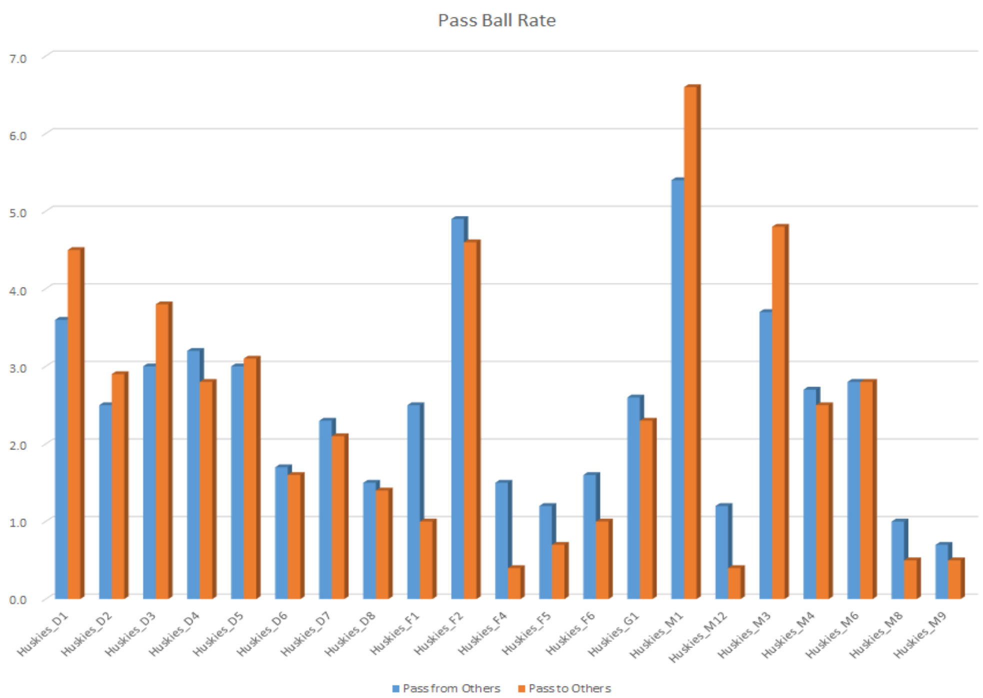


Figure 2: College of each player's passing frequency

We summarize the 'Passing to 'and 'Passing from 'events data of each player. The number of Y-axis is meaningless, while the relative height shows the difference of each player. Also, the total height of one player means how often this player appear on the ground.

3.2 The Position of Players

Different player has their own position in the game. In the soccer team, we already know the position related to player's duty, such as the forwards will appear closer to the goal of opposite team. With the data analysis from ball passing events, we can estimate the position of every player.

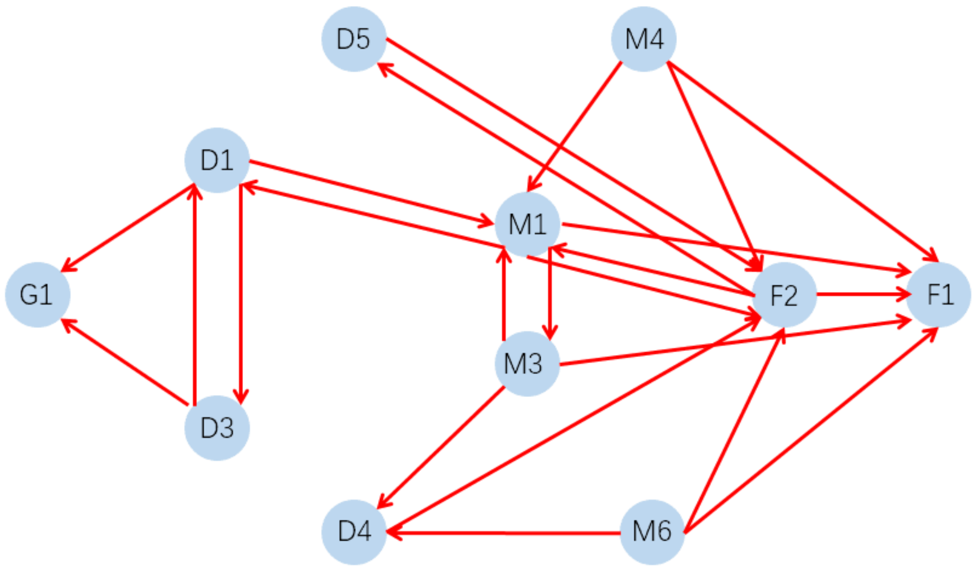


Figure 3: Passing position and trajectory

This figure creates a clear view of the original position of the most often appeared players at the game start time. The arrows show the most often ball passing tracks in games.

Through the estimate of most often appear players and their ball passing event, we can assume the Huskies team is using 4-4-2 formation in the season, which can give them a great strength in defense. However, this formation also lacks in attacking, and it makes Huskies cannot get goal very easy.

To make the advantage of 4-4-2 formation maximize, we find chain defense may be used by Huskies. M4-M1-M3-M6 makes the first line of defense, D5-D1-D3-D4 is the second line of defense. This technique is proved to be effective in our module. It helps to weaken many times of attacking.

3.3 The establishment of passing network

Players will have lots of interactions, there is lot of coordination such as passing, attacking and defense. In order to make those interactions achieve best outcome, we find every player have to control the relevant position to others.In this way, we analyzed the player position data for the whole season, as well as the passing events. [1]

The passing network is then constructed, with players represented by circular nodes placed in the average position of the field, while the radius of the nodes is proportional to the number of passes, ensuring their importance in the network structure. (The average position of each player is shown in the Appendix1) The width of the links is proportional to their weight, which weighs the number of passes between players. The diagram below:

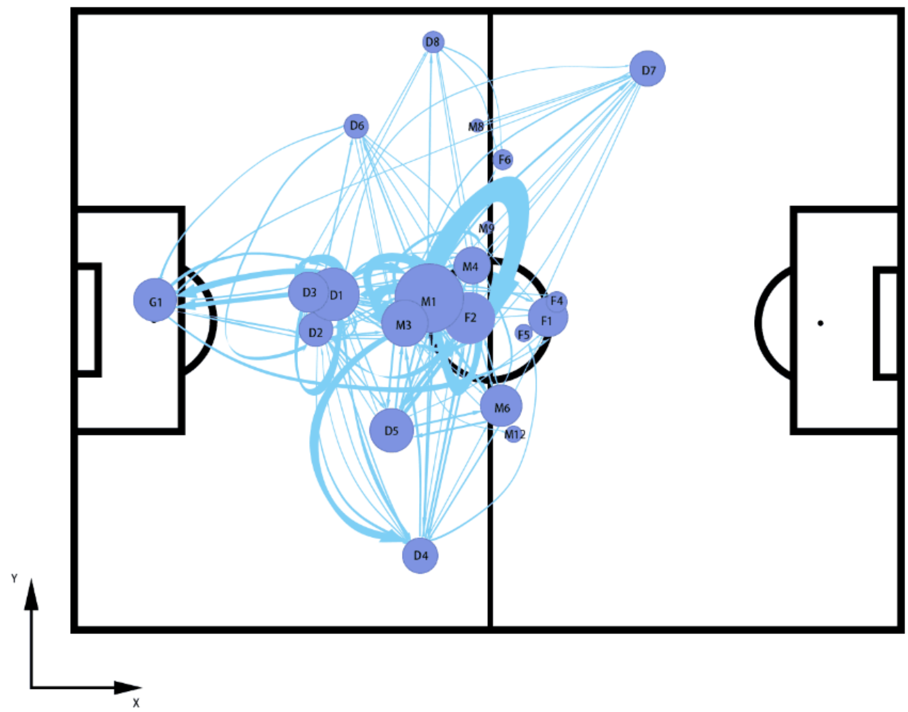


Figure 4: Schematic diagram of player passing network

If two players communicate with each other frequently, and these interactions are important to the game, then nodes and links become stronger symbols.We find that, if we want to make those interactions have higher successful rate, the players involved in need to keep them closer to each other, and make sure less opponent around the moving track of the ball.

Also, when decide player's position, it's a need to take player's duty into account. The forwards (F1, F2 etc.) have to go further, and the defender (D1, D2 etc.) need to be closer to their own goal. In this way, we could get the estimation of player best position and how they make the attack and defense.

4. Dynamic model analysis

4.1 Attacking Chain

Through rebuild the track of ball in every attacking events, we summarize those condition into two set, attacking from the backfield, and attacking from the midfield. Those attacking event is thought to start at the classified position.

4.1.1 Attacking from the Backfield

Those attacking are most generated by the failure of opponent team attacking events. There are primarily two kind of backfield attacking.

The first condition is made by the failure of shot. When the opponent team have a failed shot, Huskies_G1 will have the chance of goal kick, and pass the ball to attacking team, to say, he can directly pass the ball to Huskies_F1. This is a very common event appear in the game. After the passing, Huskies_F1 will have the opportunity to take the ball for a direct shot. Also, Huskies_F1 can pass ball to others, but this have a great chance of fail, and not a good choice for attacking.

Another condition is switch defense into attacking. When there is a chance of attacking, two option is presented to defender: pass ball to the midfielder or to the forwards.If pass to the midfielder, then the midfielder will pass the ball to forwards. For example, Huskies_D1 pass ball to Huskies_M1, after several duel, Huskies_M1 can pass the ball to Huskies_F2. If there is a shot after, one attacking finished. Also there have several times the defender pass ball directly to forwards. We find that Huskies_D5 sometimes offer Huskies_F2 an opportunity of shot.

4.1.2 Attacking from the Midfield

It's the most common condition in a typical attacking event.In a pretty large possibility, Huskies_M1 will pass ball to Huskies_F2. We put a great effort in research this pair of players. There is a great chance that this pair makes up a small attacking group. Beside this, Huskies_M4 and Huskies_F2 also form a chain of attacking.

If the opponent team is strong, Huskies must face fierce duel in the match. When this happens, midfielder to midfielder passing ball will be very common, and can possibly continue for a long time. Huskies will use this tactic to avoid opponent team get the ball. After the duel, midfielder can pass the ball to forwarder to continue attacking, or, they have to begin to defense.

4.2 Defense Chain

The defense can be started from the backfield or the mid field.There has little chance of playing tricks when in the defense at the backfield. The attacking is fierce and little room to withdraw. Defender can pass ball to another defender in such event, they can also pass ball to midfielder, if it is possible, the midfielder can switch defense to attacking.

Because Huskies using 4-4-2 formation, it's easy for them to resist attacking by using left and right guard to defense. If the defense start at the midfield, Huskies can have a better choice to defense. The ball passing between Huskies_M1 and Huskies_M3 can reduce the probability of opponent team continue attacking. Also, this tactic can easily change defense into attack.

4.3 Dyadic and Triadic Configurations

We divided the players who worked well with each other into teams of two and three, giving them different roles in different competitions. These teams will be the heart of the offense or defense.

By comparing the frequency of duels won with the chance to score, huskies were more likely to win with small teams. First, calculate the number and frequency of passes between two players. Then, assuming some random group, use iteration to clear the less relevant players and add more relevant players. After a few iterations, we can form a team. In each group, each member in the group has more connections with each other than anyone outside the group.

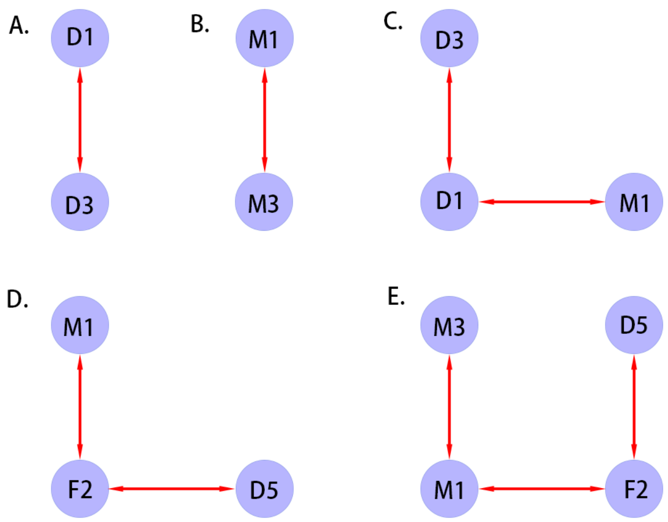


Figure 5: Binary and ternary teams

After analysis of the whole team, we can find those small groups. A, B is the dyadic team, while A usually have the function of defense, and B serves as attacking. C, D is the triadic team. C often appears in defense, D is used in attacking. E is a quaternary team, which primarily serve as attacking, also can used to defense.

The defensive team is mainly composed of Huskies_D1 and Huskies_D3 pairs (configuration A), which have different functions when defending. If the ball is in the backcourt, the pair can defend with Huskies_G1. The Huskies_M1 defensive formation (configuration C) can be formed between the two of them when it is a center and back court.

The main attack consists of Huskies_M1 and Huskies_F2. However, the offense is not always ideal. Players must face the defense, so other formations may be a better choice (configuration D, E). In addition, Huskies_M1 and Huskies_M3 formations are common between two separate stacks or between attack and defense (configuration B).

The huskies chose a 4-4-2 formation, and forming a defensive chain was a major advantage. In our defense studies, we can see that huskies have two defense chains. The first one is m4-m1-m3-m6. This chain is used to limit the opponent's advance and form a defensive midfield. Huskies_M1 and Huskies_M3 face each attack and kick the ball back. The back line is D5- d1-d3-d4. When your opponents reach this line, it's best to ask them not to take a step forward. If they can't resist the attack, they are likely to lose a point

4.4 Evaluation of Each Player

Through the game data annotation, as well as the actual football game statistics, we can get the following five indicators: **Clearance Rate**, **Duel Rate**, **Foul Rate**, **Pass Rate**, **Shot Rate**. From the game data set given in the question, the frequency of each player in different indicators is obtained. Take the Foul Rate for example:

$$\text{Foul Rate\_id} = \text{id player's Foul count/all Huskies players' Foul count} \quad (1)$$

Then, the index data are normalized:

$$\text{Foul Rate\_id} = [\text{Foul rate\_id} - \min(\text{Foul Rate\_id})]/[\max(\text{Foul Rate\_id}) - \min(\text{Foul Rate\_id}) \quad (2)$$

The other four indicators are calculated similarly. So you get a score sheet for each player in five categories. (see Appendix 2) By selecting the 10 players with the highest frequency of playing (except the goalkeeper), we can draw the five-dimensional ability diagram of all players:

Figure 6: Five-dimensional ability diagram

4.5 Performance of Team

4.5.1 The effect of position

By comparing 19 opponents' matches, huskies played each opponent twice. If there are different results for the same opponent, we observe and compare the two games. With these comparisons, we can see that there are big differences, especially in player positions.

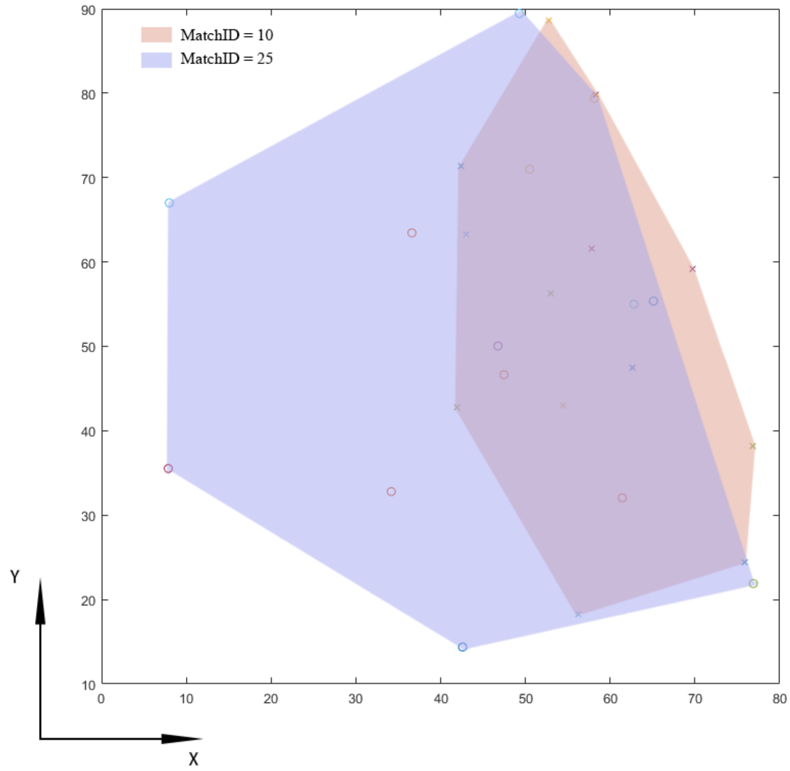


Figure 7: Position influence diagram

Player average position draw from MatchID 10 and MatchID 25. Huskies vs. Opponent10. The average position of players in the first match is marked with ‘x’, in the second one is marked ‘o’. The colored shape illustrate the area covered by the team members.

Huskies won MatchID 10 and lost MatchID 25. From the picture we can easily get the huskies team position by defensive position is more advantageous.

4.5.2 The effect of sides

We find that when Huskies have matches as ‘home’ team, they can play much better than play as ‘away’ team. in fact, Huskies only win 3 times when ‘away’, compare to 10 wins when ‘home’. We advice that Huskies need to find out strategies to deal with this issue, mental health or travel agency, etc.

5. CEI Model Design

5.1 The establishment of individual Comprehensive Evaluation Index

In order to obtain the comprehensive ability indicators of all husky players, based on the analysis of the five indicators of each player in the previous section, we will use the principal component analysis method to try to replace the original indicators with a new set of independent comprehensive indicators. It is to reduce the dimension of all players. The steps are as follows: [2]

Step1: Calculate the correlation coefficient matrix R of 5 indicators.

$$R=\begin{bmatrix} 1.000 & 0.383 & 0.214 & 0.603 & -0.072 \\ 0.383 & 1.000 & 0.879 & 0.682 & 0.743 \\ 0.214 & 0.879 & 1.000 & 0.649 & 0.680 \\ 0.603 & 0.682 & 0.649 & 1.000 & 0.295 \\ -0.072 & 0.743 & 0.680 & 0.295 & 1.000 \end{bmatrix} \quad (3)$$

Step2: Determine the number of principal components. Before determine the number of principal components, we need to first give a control value value value, value =15%.

Ingredient	Contribution rate(%)	Cumulative contribution rate(%)
1	62.955	62.955
2	25.174	88.130
3	6.645	94.130
4	3.771	98.546
5	1.454	100.000

Table 3: Principal component contribution rate

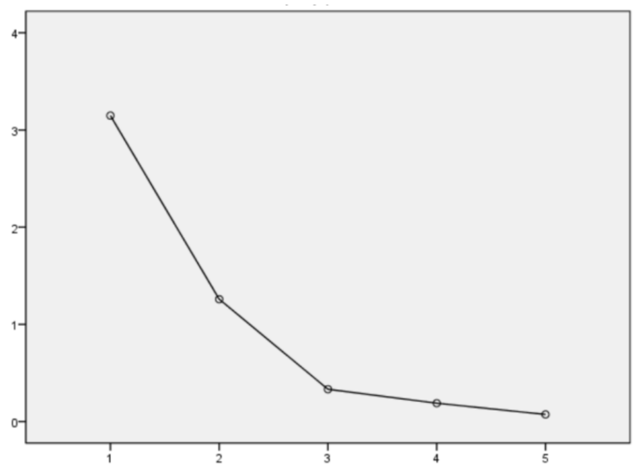


Figure 8: scree plot

The figure above is the figure of the contribution rate of the low gravel on the cliff, which is the hope that the graph is very steep at the beginning, like a cliff, and the remaining value is very small, like the gravel at the bottom of the cliff. The lithotriptic graph is an intuitive representation of the contribution rate of principal components. When one component is taken, its cumulative contribution rate is 62.955%. When two components were taken, the cumulative contribution rate reached 88.130%, meeting the setting conditions. Therefore, it can be determined that the number of retained principal components m is 2.

Step3: Find the eigenvalues of the correlation coefficient matrix R and sort $\lambda 1 \geq \lambda 2 \geq \lambda 3 \geq \lambda 4 \geq \lambda 5$(see Table 4), then the i principal component is expressed as a linear combination Fi of normalized indicators X_k.

Table 4: The eigenvalues of the correlation coefficient matrix

λ_1	λ_2	λ_3	λ_4	λ_5
3.148	1.259	0.332	0.189	0.073

Table 5: Component score coefficient matrix

Ingredient	Clearance	Duel	Foul	Pass	Shot
1	0.148	0.306	0.290	0.258	0.228
2	0.649	-0.085	-0.164	0.334	-0.476

$$F_1 = 0.148\, x_1 + 0.306\, x_2 + 0.290\, x_3 + 0.258\, x_4 + 0.228\, x_5 \qquad (4)$$

$$F_2 = 0.649\, x_1 - 0.085\, x_2 - 0.164\, x_3 + 0.334\, x_4 - 0.476\, x_5 \qquad (5)$$

Step4: Calculate the comprehensive score. First, the score of each principal component was calculated, and then the contribution rate of variance of each principal component was taken as the weight to obtain the comprehensive evaluation index of each player (see Appendix 3).

$$F = 62.955 / 88.130\, F_1 + 25.174 / 88.130\, F_2 = 0.7143F_1 + 0.2856F_2 \qquad (6)$$

5.2 The establishment of team Comprehensive Evaluation Index

First of all, we set up corresponding analysis methods for the configuration teams of different positions[3] (forward, midfield and defense) :

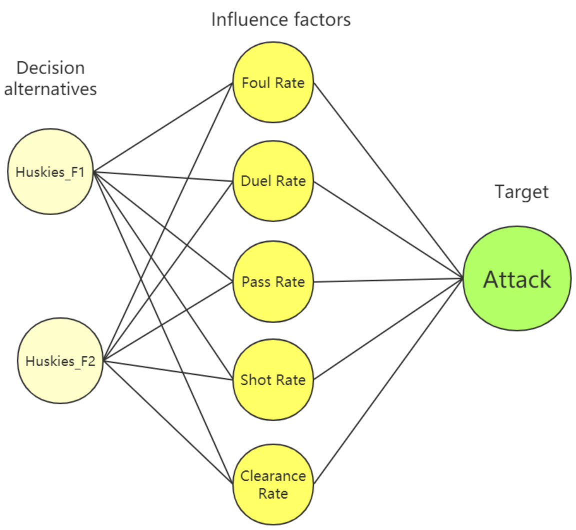


Figure 9: Offensive team decision analysis

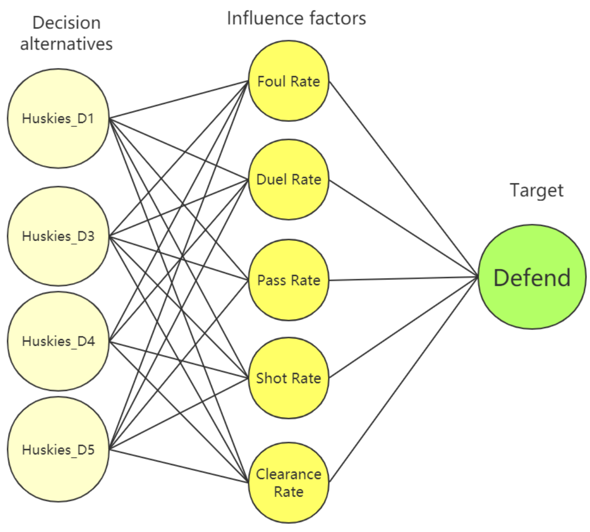


Figure 10: Defensive team decision analysis

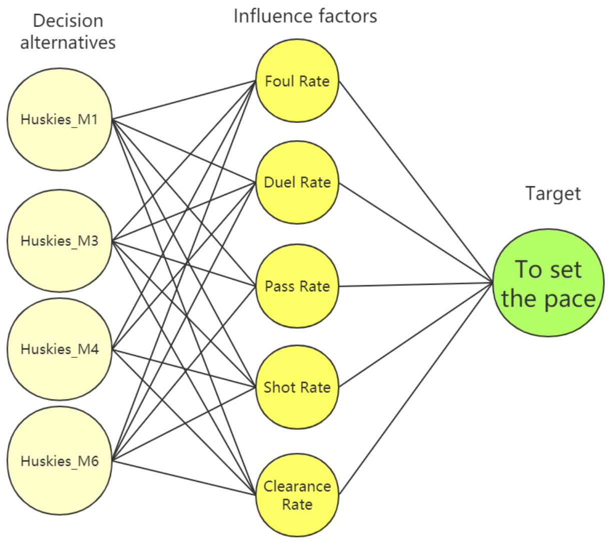


Figure 11: Control the rhythm

Because different positions have different needs for players' abilities, we can obtain players' contributions to the team through principal component analysis. Taking forward players as an example, combined with the actual situation, we believe that the importance of the five indicators for forward position is

Shot Rate = Duel Rate > Foul Rate = Pass Rate > Clearance Rate

According to the nine-scale method, the judgment matrix is established:

$$C=\begin{bmatrix} 1 & 1/3 & 1 & 1/3 & 3 \\ 3 & 1 & 3 & 1 & 5 \\ 1 & 1/3 & 1 & 1/3 & 3 \\ 3 & 1 & 3 & 1 & 5 \\ 1/3 & 1/5 & 1/3 & 1/5 & 1 \end{bmatrix} \quad (7)$$

Then, the weight of different indicators for the position of the forward can be calculated by Matlab.

Table 6: The weight of different indicators to forward position

Clearance	Duel	Foul	Pass	Shot
0.0551	0.3435	0.1290	0.1290	0.3435

In addition, by analyzing the two most common forward players in the lineup (Huskies_F1, Huskies_F2), we can establish a pair comparison matrix for each indicator data:

$$D[Clearance]=\begin{bmatrix} 1 & 0.52 \\ 1.923 & 1 \end{bmatrix} (8) \qquad D[Duel]=\begin{bmatrix} 1 & 1.85 \\ 0.54 & 1 \end{bmatrix} (9)$$

$$D[Foul]=\begin{bmatrix} 1 & 1.13 \\ 0.88 & 1 \end{bmatrix} \quad (10) \qquad D[Pass]=\begin{bmatrix} 1 & 0.31 \\ 3.23 & 1 \end{bmatrix} \quad (11) \qquad D[Duel]=\begin{bmatrix} 1 & 1.087 \\ 0.92 & 1 \end{bmatrix} \quad (12)$$

From this, we can obtain the contribution of two forward players to the five indicators of forward, as shown in the table below:

Table 7: The contribution of two forward players to the five indicators

	Clearance_per	Duel_per	Foul_per	Pass_per	Shot_per
Huskies_F1	0.3421	0.6492	0.5312	0.2365	0.5208
Huskies_F2	0.6579	0.3508	0.4688	0.7635	0.4792

Based on the contribution of two players to the five indicators and the weight of each indicator to the striker position, the total contribution coefficient of the two players to the striker position can be obtained:

α= Shot Rate_weight*Shot_per+Duel Rate_weight*Duel_per+Foul Rate_weight*Foul Rate_weight+Pass Rate_weight*Pass_per+Clearance Rate_weight*Clearance_per (13)

The contribution coefficient of Huskies_F1 α_1 is 0.52 and that of Huskies_F1 α_2 is 0.48. Similarly, we can calculate the contribution coefficients of different players at different positions.

Among the four midfields with the highest attendance rate:
Huskies_M1: 0.2147 Huskies_M3: 0.3579 Huskies_M4: 0.1816 Huskies_M6: 0.2450

Among the four defenses with the highest attendance rate:
Huskies_D1: 0.23490 Huskies_D3: 0.2490 Huskies_D4: 0.1403 Huskies_D5: 0.2618

To sum up, combined with individual comprehensive evaluation index F, the comprehensive evaluation index of forward configuration team is:

$$Z_1 = \alpha_1 * F + \alpha_2 * F(\text{another forward's individual score}) \quad (14)$$

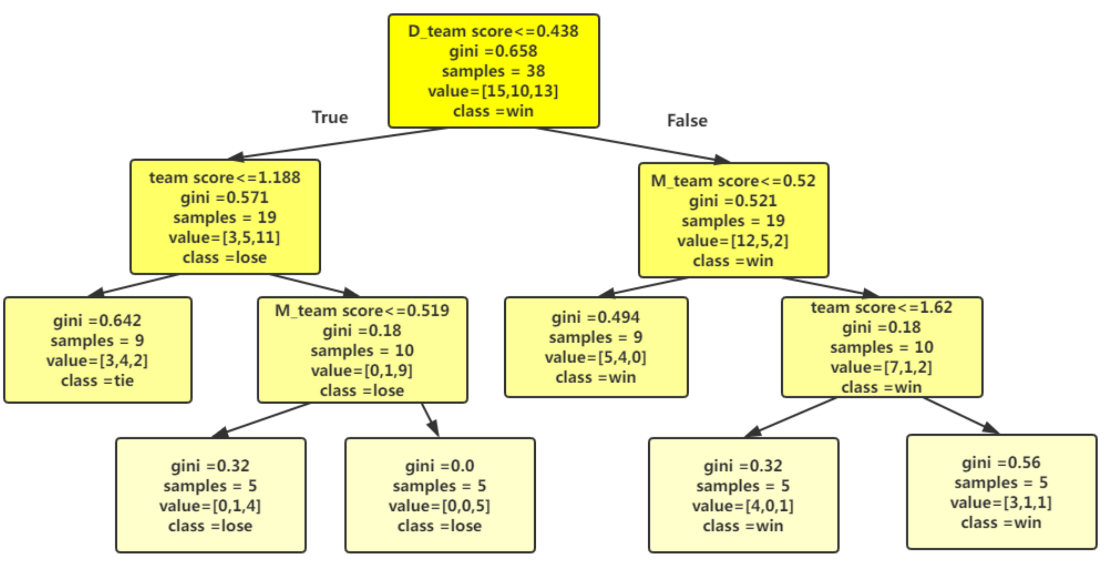
In the same way, we can obtain the comprehensive index Z_2 of the midfield configuration team and the comprehensive index Z_3 of the defense configuration team. Since every attack and defense of the team is linked, it cannot be expressed by three independent indicators. Therefore, the comprehensive evaluation index Z of the team is the sum of the index of offense, the index of the contribution of center guard ball to offense, the index of defender, and the index of the contribution of center guard ball to defense.

$$Z = Z_1 + 655/3863 * Z_1 * Z_2 + Z_3 + 1668/3863 * Z_2 * Z_3 \quad (15)$$

5.3 Model prediction

Decision tree learning essentially generalizes a set of classification rules from existing data sets. Decision tree learning is a conditional probability model estimated by training data sets. There are an infinite number of conditional probability models based on feature space partition, and the conditional probability model we choose should not only fit the training data well, but also have a good prediction for the unknown data.Therefore, we introduce the classification model of decision tree.[4]

Since the above data are combined with the data of appearances in the whole season, in order to verify the accuracy of our model, we calculated the Huskies' individual comprehensive evaluation scores for forwards, midfields and defenses in each game, and configured the team's index data and the team's comprehensive evaluation scores. Based on these data,



we established the following decision tree model a:

Figure 12: Decision tree model a

Using the decision tree model to predict, we found that the prediction effect of the model is close to 70%. Analyzing the model, we found that the Huskies' failure rate decreased to 2/19 when the average combined value of the Huskies' defense teams was greater than 0.438 Huskies teams have a winning ratio of 0.7 if the average combined score of the Huskies' midfields is greater than 0.52.

To explore the ability of team level flexibility in different games to influence the outcome of the game. We counted the number of passes from 1 to 25 in the passing chain and analyzed the number of passes in different matches together with the game results. We established decision tree classification model b.

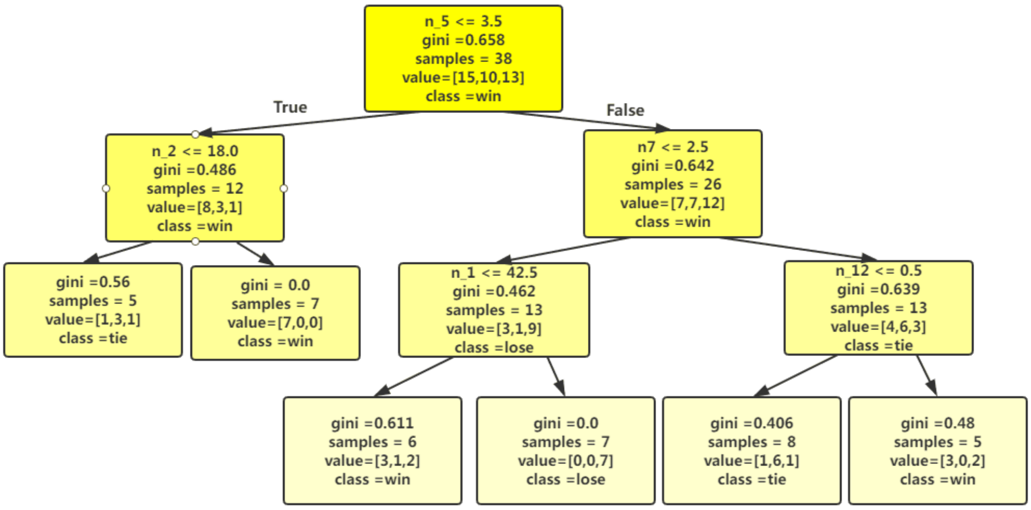
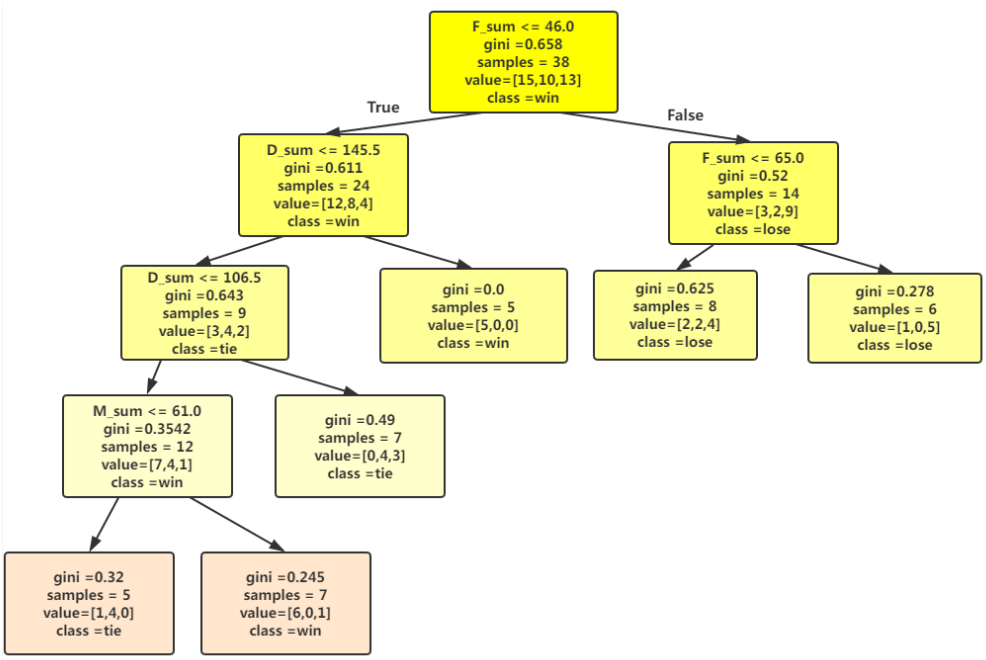


Figure 13: Decision tree model b

Using the decision tree classification model, the prediction accuracy of the game results reached 73%. We can see that the probability of winning is better when there is a longer passing chain.

In order to explore the position of passing chain, we once again counted the total passing



times of players at different positions in the whole season, and established decision tree model c:

Figure 14: Decision tree model c

The prediction accuracy of the decision tree model was 76%. We found that players at different positions had different average passes per game, while Huskies had a 7/12 chance of winning when the forward player passed less than 65 passes and the defense passed less than 145.5 passes and the midfield passed less than 61.

To sum up, through the prediction of the three decision tree models established above, we analyzed the ability indicators of team members, the number of passing chains and the position of passing chains in the team cooperation, which can predict the Huskies' winning and losing situation in the season accurately

6. Suggestions to Huskies

Based on the above analysis of the team network structure and the establishment of the comprehensive evaluation index of the team, in order to improve the Huskies' winning probability next season, we will make the following Suggestions to the coaches:

- From the team configuration analysis, the players who cooperate with each other more can be divided into two or three small teams, so that they have different responsibilities in different competitions. These teams will be the heart of the offense or defense.

- From the analysis of individual contribution index, the coach needs to make reasonable changes to the players so that each position matches the player with the highest score in the comprehensive index. If necessary, the original position of the players can be changed according to the difference of the five indicators, so as to achieve the optimal comprehensive evaluation index matching the team and the individual.

- From the team configuration decision tree model analysis, Huskies should use a 4-4-2 formation that is more capable of holding the ball in defense and midfield, Huskies team's winning probability will increase to 0.7 if there are enough strong full-back configuration team and centre-back configuration team. [5]

- From the perspective of the passing chain decision tree model, Huskies should reduce the number of invalid passing paths in the passing chain to minimize the number of passing chains per offense. For the Huskies, the longer the passing chain, the lower the probability of a successful goal. The Huskies are better suited to Strong Side Overloads and quick returns. Corner kicks and free kicks are also very effective for Huskies as they minimize the number of passes in the attack.

- From the ball chain position analysis, Huskies to win as much as possible, should pass the ball between the lower guard configuration team number, the best solution is tightening defender defense, expand the centre-half chain of passing, can make the ball faster is passed to the front of the striker's position through the defensive counter-attack to compress the opponent's formation, rather than through the defender passed between tactics to ensure the stability of the ball.

- From the side of the team analysis, the huskies team in the home win probability is greater. Husky team needs to adjust their mentality in time and seek help from a psychologist when necessary.

7. Development team performance promotion

In real society, the types of teams are diversified, so the indicators of team members will change according to the demands of team tasks. However, what remains unchanged is that in the real world, there are three common conditions for teams:

（1）Scattered team members

（2）Team members and a single team core

（3）Multi-team core and team members under multi-team core

Therefore, we can establish a decision tree model to determine what is the most beneficial for the overall team to limit the number of core members. In addition, the process of interaction and cooperation between different groups within the team is also crucial. Due to the existence of team tasks, cooperation relationship will be formed between groups, which will lead to cooperation coefficient and friction coefficient between groups. By establishing the classification model of decision tree, we can identify and judge to what extent the cooperation coefficient and friction coefficient between groups are limited, which is more beneficial to the whole team

Combined with the actual situation of the society, we also need to use the **5W1H** analysis method[6] to analyze the interaction and cooperation process within the team, so as to figure out how to build an efficient team.

1. Who

In-depth analysis of team members, clear the advantages and disadvantages of each team member, the coordination of team members and the distribution of contributions.

2. Where

Every team has its own strengths and weaknesses. The success of the team needs to face risks and opportunities in the society. We need to analyze the current environment of the team to clarify how the team plays its strengths, avoids threats and meets challenges.

3. What

Focus on the tasks of the team and form a common purpose and belief. The generation of common purpose can provide guidance and motivation for the members, forming a strong cohesion.

4. When

Taking the right action at the right time is the key to team success. When to ease conflicts and avoid risks. When to meet challenges, seize opportunities.

5. Why

In order to run the team efficiently, it is necessary for team members to make clear the motivation for joining the team and why they choose to join the team, so as to enhance their sense of responsibility and mission.

6. How

How to operate within the team involves the allocation of team roles. Each member needs to have a clear position responsibility, and cooperate with each other, and good communication.

8. Strengths and Weaknesses

8.1 Strengths

- PCA can effectively reduce data dimensions and simplify data calculation.

- Through the establishment of comprehensive evaluation index data model, it can accurately describe the role of individuals in the team and their contributions to various aspects of the team. Under the same team, the abilities and contributions of different team members are highly comparable with data

- The team comprehensive evaluation indexes established by us have been verified by the prediction of the decision tree model and have high accuracy. And with the increase of sample data. The accuracy of our predictions can be improved.

- The team comprehensive index data model for Huskies teams can be extended to the general team model, which has a high practical significance.

- The results of the cooperation behaviors within the team established by us are consistent with those of the team in real life. Through our model, we can more accurately grasp how to control these behaviors within the team, so that the overall team output tends to be optimal.

8.2 Weaknesses

- The results of our model are the results of the Huskies team analysis for the whole season. The significance of this model is that the Huskies team optimization based on this model is limited to the current season data and cannot take into account the changes in players' individual abilities in the future.

- Our decision tree classification model is more dependent on the decision classification of large sample data, and the limitation of team season data will affect the prediction results to some extent

- Because our model is based on the analysis results of the general situation. Therefore, it is of general significance to the team, but we cannot give specific solutions to the opponents' temporary tactical changes.

9. References

[1] Cintia, P., Rinzivillo, S., and Pappalardo, L. (2015). “A network-based approach to evaluate the performance of football teams,” in Machine Learning and Data Mining for Sports Analytics Workshop (Porto).

[2] Frey D F, Pimentel R A. Principal component analysis and factor analysis[J]. 1978.

[3] Saaty T L. Decision making with the analytic hierarchy process[J]. International journal of services sciences, 2008, 1(1): 83-98.

[4] Ibomoiye Domor Mienye,Yanxia Sun,Zenghui Wang. Prediction performance of improved decision tree-based algorithms: a review[J]. Procedia Manufacturing,2019,35

[5] https://en.wikipedia.org/wiki/Association_football_tactics_and_skills

[6] Tianxiao Jiang.A brief discussion on 5W1H analysis method in scientific and technological novelty checking[J].Information Research, 2011, (5): 96-97.

Appendix

Appendix 1

Table 8: The average position of each player

ID	n	x	y
Huskies_F1	1476	65.7737127371273	48.8035230352303
Huskies_F2	1935	53.6739018087855	49.425322997416
Huskies_F3	173	63.9653179190751	53.3179190751445
Huskies_F4	808	68.444306930693	49.5222772277227
Huskies_F5	679	63.9823269513991	44.8247422680412
Huskies_F6	757	60.7424042272126	69.9933949801849
Huskies_M1	2555	46.6211350293542	53.5878669275929
Huskies_M2	165	53.1818181818181	47
Huskies_M3	1740	44.6827586206896	48.4206896551724
Huskies_M4	1387	55.1795241528478	55.9524152847873
Huskies_M5	109	60.2477064220183	42.9633027522935
Huskies_M6	1565	58.9284345047923	36.2140575079872
Huskies_M7	41	40.5609756097561	57.170731707317
Huskies_M8	517	57.4816247582205	74.3520309477756
Huskies_M9	495	59.1030303030303	59.630303030303
Huskies_M10	146	53.4657534246575	44.9246575342465
Huskies_M11	151	50.1589403973509	58.3973509933774
Huskies_M12	640	62.5375	30.228125
Huskies_M13	178	51.1853932584269	52.3483146067415
Huskies_D1	1973	33.8849467815509	52.9878357830714
Huskies_D2	1290	32.8441860465116	46.5093023255813
Huskies_D3	1523	32.8187787261982	52.4865397242285
Huskies_D4	1344	47.7008928571428	14.1748511904761
Huskies_D5	1652	43.090799031477	32.77602905569
Huskies_D6	942	39.3333333333333	75.0976645435244
Huskies_D7	1327	45.4513941220798	84.1431801055011
Huskies_D8	817	50.6328029375765	86.9130966952264
Huskies_D9	104	42.1730769230769	29.2211538461538
Huskies_D10	99	34.5656565656565	28.2424242424242
Huskies_G1	1615	26.5919504643962	49.240866873065

Among them, n represents the number of passes made by players, x and y represent the average position coordinates of players.

Appendix 2

Table 9: Individual index score

	Clearance	Duel	Foul	Pass	Shot
Huskies_F1	0.1019	1	0.8542	0.2255	1
Huskies_F2	0.1944	0.5412	0.7500	0.7299	0.9211
Huskies_F3	0	0.0588	0	0.0379	0.2368
Huskies_F4	0.0370	0.4780	0.6042	0.1257	0.9737
Huskies_F5	0.0926	0.3661	0.5000	0.1357	0.4737
Huskies_F6	0.0926	0.3311	0.0833	0.2023	0.4737
Huskies_M1	0.1667	0.8102	1	1	0.6316
Huskies_M2	0.0093	0.0384	0.0833	0.0492	0.0263
Huskies_M3	0.2778	0.5367	0.7917	0.6993	0.1316
Huskies_M4	0.1111	0.5650	0.8333	0.4238	0.5526
Huskies_M5	0	0.0237	0	0.0220	0.1053
Huskies_M6	0.2222	0.5650	0.3542	0.4518	0.8947
Huskies_M7	0	0	0	0	0
Huskies_M8	0	0.1876	0.2083	0.1218	0.1053
Huskies_M9	0.0093	0.2407	0.1042	0.1191	0.1316
Huskies_M10	0.0093	0.0655	0.0833	0.0160	0.0789
Huskies_M11	0.0093	0.0475	0.04167	0.0406	0
Huskies_M12	0.0926	0.3085	0.2292	0.1457	0.4737
Huskies_M13	0.0278	0.0475	0.0417	0.0519	0.0263
Huskies_D1	1	0.6452	0.3750	0.7099	0.3158
Huskies_D2	0.9537	0.3842	0.4792	0.4624	0.2105
Huskies_D3	0.9444	0.4023	0.1458	0.6061	0.1316
Huskies_D4	0.3241	0.2712	0.1458	0.4983	0.0789
Huskies_D5	0.6667	0.4497	0.5625	0.5363	0.0789
Huskies_D6	0.4815	0.3175	0.2292	0.3007	0.1579
Huskies_D7	0.5278	0.4000	0.3333	0.3846	0.0526
Huskies_D8	0.1111	0.2237	0.0417	0.2389	0.1053
Huskies_D9	0.0463	0.0147	0	0.0279	0.0263
Huskies_D10	0.0741	0.0181	0.0417	0.0193	0.0263
Huskies_G1	0.1574	0.0350	0	0.5017	0

Appendix 3

Table 10: The combined score of each player

	F1	F2	F
Huskies_F1	3.06170274754161	1.02022339025789	2.47851947121248
Huskies_F2	2.75596095949437	-0.5127938130563	1.82221565028093
Huskies_F3	-1.71719798516306	0.420068562231622	-1.10667230419417
Huskies_F4	1.28929854829254	0.611649833770462	1.09571285340449
Huskies_F5	0.311279408087971	0.257681483975792	0.29596501754531
Huskies_F6	-0.347914853290596	0.595650571726929	-0.0783842930880938
Huskies_M1	3.80345175637014	-0.916730852027469	2.4550972266872
Huskies_M2	-1.86010442588132	0.158292765619695	-1.28353130025586
Huskies_M3	1.83611898959643	-0.953081271726318	1.03936938608478
Huskies_M4	1.88919706624545	-0.232694903059335	1.28306066402249
Huskies_M5	-1.98449259846496	0.320612144320359	-1.32602086621093
Huskies_M6	1.66436199248196	0.419246562146051	1.30867644920034
Huskies_M7	-2.20481301363716	0.262864402976259	-1.49989996290902
Huskies_M8	-1.12335907376152	0.189141138264474	-0.748432911170108
Huskies_M9	-1.14945286459256	0.417839848207432	-0.701745878011986
Huskies_M10	-1.79257936621578	0.283074694270796	-1.19965199034291
Huskies_M11	-1.95889352080872	0.224514685226475	-1.3351842758783
Huskies_M12	-0.246473379955079	0.470905539949346	-0.0415535103727405
Huskies_M13	-1.89030604663507	0.211049116207658	-1.29003588550505
Huskies_D1	2.24669510693678	-0.528344209340644	1.4539847726679
Huskies_D2	1.28208891889825	-0.735561308864936	0.70573725269374
Huskies_D3	0.895938471979731	-0.565124563857418	0.478579116906385
Huskies_D4	-0.168262887699912	-0.353853793131997	-0.221273831792538
Huskies_D5	1.26680957215865	-0.773446722033706	0.684000787413767
Huskies_D6	-0.0275550250368048	-0.0971612499980416	-0.0474373790017317
Huskies_D7	0.368307733507277	-0.289493774982386	0.180404207504969
Huskies_D9	-2.05313082618049	0.235032135808256	-1.39949748490868
Huskies_D8	-1.03228497372905	0.214395794691761	-0.676161146963083
Huskies_D10	-1.9666090060488	0.191259578429811	-1.35019494384074
Huskies_G1	-1.1478198517694	-0.545212438235651	-0.975671385045214

Appendix 4

```
import pandas as pd
import matplotlib.pyplot as plt
import os
os.chdir(r'C:\Users\12082\Desktop\2020_Problem_D_DATA')
df = pd.read_excel('score_result.xlsx')

print(df.shape)
print(df.dtypes)

x = df[['D_sum','M_sum', 'F_sum','sum']]
y = df['result']
print(x.shape)
print(y.shape)

from sklearn.tree import DecisionTreeClassifier

clf_tree = DecisionTreeClassifier(criterion='gini',
                                  splitter='best',
                                  max_depth=3,
                                  min_samples_split=2,
                                  min_samples_leaf=5,
                                  max_features=None,
                                  max_leaf_nodes=8,
                                  min_impurity_decrease=0.0,
                                  min_impurity_split=None,
                                  class_weight=None)

clf_tree.fit(x,y)
print(clf_tree)

y_predict = clf_tree.predict(x)

import graphviz
# https://graphviz.gitlab.io/_pages/Download/Download_windows.html
import os
os.environ["PATH"] += os.pathsep + 'C:/Program Files (x86)/Graphviz2.38/bin/'
from IPython.display import Image
import pydotplus
features = ['D_sum','M_sum', 'F_sum','sum']
classes=['1','0','-1']
#
from sklearn import tree
tree_graph_data = tree.export_graphviz(clf_tree,
                        feature_names=features,
                         class_names=classes,
                        filled=True,
                        rounded=True)
```

```
tree_graph = pydotplus.graph_from_dot_data(tree_graph_data)
Image(tree_graph.create_png())

from sklearn.metrics import accuracy_score

print(f'accuracy score is: {accuracy_score(y,y_predict)}’)
```

Appendix 5

```
#
import pandas as pd
import matplotlib.pyplot as plt
import os
os.chdir(r'C:\Users\12082\Desktop\2020_Problem_D_DATA')
df = pd.read_csv('pass_count.csv')
#

#
print(df.shape)
print(df.dtypes)
#
x = df[['n1','n2', 'n3', 'n4','n5',
      'n6','n7', 'n8', 'n9','n10',
      'n11','n12', 'n13', 'n14','n15']]
y = df['result']
print(x.shape)
print(y.shape)
```

Appendix 6

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import os
os.chdir(r'C:\Users\12082\Desktop\2020_Problem_D_DATA')
import warnings
warnings.filterwarnings('ignore')

fullevents = pd.read_csv('fullevents.csv')
passingevents = pd.read_csv('passingevents.csv')

passingevents_H = passingevents[passingevents['TeamID'] == 'Huskies']
fullevents_H = fullevents[fullevents['TeamID'] == 'Huskies']
Huskies_id = list(set(fullevents[fullevents['TeamID'] == 'Huskies']['OriginPlayerID']))
Huskies_M = ['Huskies_M1','Huskies_M2','Huskies_M3','Huskies_M4','Huskies_M5',
          'Huskies_M6','Huskies_M7','Huskies_M8','Huskies_M9','Huskies_M10',
```

```
           'Huskies_M11','Huskies_M12','Huskies_M13','Huskies_M14','Huskies_M15']
Huskies_F = ['Huskies_F1','Huskies_F2','Huskies_F3','Huskies_F4','Huskies_F5',
           'Huskies_F6','Huskies_F7','Huskies_F8','Huskies_F9','Huskies_F10']
Huskies_D = ['Huskies_D1','Huskies_D2','Huskies_D3','Huskies_D4','Huskies_D5',
           'Huskies_D6','Huskies_D7','Huskies_D8','Huskies_D9','Huskies_D10']

#
def match_passing(df):
    lst=[]
    for i in range(1,39):
        dic={}
        df_temp = df[df['MatchID'] == i]
        dic['M_F'] = len(df_temp[df_temp['OriginPlayerID'].items() in Husk-
ies_M][df_temp['OriginPlayerID'] in Huskies_F])
        dic['M_D'] = len(df_temp[df_temp['OriginPlayerID'] in Husk-
ies_M][df_temp['OriginPlayerID'] in Huskies_D])
        lst.append(dic)
    pass_df = pd.DataFrame(lst)
    return pass_df
pass_df = match_passing(passingevents_H)
outputpath=r'C:\Users\12082\Desktop\2020_Problem_D_DATA\match_pasing.xlsx'
match_passing_df.to_excel(outputpath,index=True,header=True)

def match_passing(df):
    lst=[]
    for i in range(1,39):
        dic={}
        df_temp = df[df['MatchID'] == i]
        for per_name in Huskies_id:
            dic[per_name] = len(df_temp[df_temp['OriginPlayerID'] == per_name])
        lst.append(dic)
    match_passing_df = pd.DataFrame(lst)
    return match_passing_df
match_passing_df = match_passing(passingevents_H)
outputpath=r'C:\Users\12082\Desktop\2020_Problem_D_DATA\match_pasing.xlsx'
match_passing_df.to_excel(outputpath,index=True,header=True)

def match_pos(df,n):
   list_x=[]
   list_y=[]
   list_n =[]
   df = df[df['MatchID'] == n]
   for i in Huskies_id:
      list_n.append(len(df[df['OriginPlayerID'] == i]))
      list_x.append(df[df['OriginPlayerID'] == i]['EventO-
rigin_x'].sum()/len(df[df['OriginPlayerID'] == i]))
      list_y.append(df[df['OriginPlayerID'] == i]['EventO-
rigin_y'].sum()/len(df[df['OriginPlayerID'] == i]))
   df_position = pd.DataFrame({'ID':Huskies_id,
```

```
                    'n':list_n,
                    'x':list_x,
                    'y':list_y})
  return df_position
for i in range(1,39):
  df_temp = match_pos(fullevents,i)
  outputpath=r'C:\Users\12082\Desktop\2020_Problem_D_DATA\pos_'+str(i)+'.xlsx'
  df_temp.to_excel(outputpath,index=True,header=True)

df_temp=pd.DataFrame()
list_temp = []
for i in Huskies_id:
  df_temp1 = passingevents_H.loc[passingevents_H['OriginPlayerID']==i]
  dic_temp = {}
  for j in Huskies_id:
     df_temp2 = df_temp1.loc[df_temp1['DestinationPlayerID']==j]
     dic_temp[j]=len(df_temp2)
  list_temp.append(dic_temp)
df_temp = pd.DataFrame(list_temp,index=Huskies_id)

outputpath=r'C:\Users\12082\Desktop\2020_Problem_D_DATA\passing_f.xlsx'
df_temp.to_excel(outputpath,sep='\t',index=True,header=True)

print('finished!')
```