|  |  |  |
| --- | --- | --- |
| **Problem Chosen** D | **2020 MCM/ICM Summary Sheet** | **Team Control Number** 2008495 |

Use this template to begin typing the first page (summary page) of your electronic report. This template uses a 12-point Times New Roman font. Submit your paper as an Adobe PDF electronic file (e.g. 1111111.pdf), typed in English, with a readable font of at least 12-point type.

Do not include the name of your school, advisor, or team members on this or any page.

Papers must be within the page limit specified in the problem statement.

Be sure to change the control number and problem choice above.  
You may delete these instructions as you begin to type your report here.   
  
**Follow us @COMAPMath on Twitter or COMAPCHINAOFFICIAL on Weibo for the most up to date contest information.**

**Summary**

**Have a wonderful soccer game**

本文利用图论，概率论和微积分的方法，建立，，，和数据优化模型，为足球教练和球员的训练提供策略。

首先，根据图论知识，在球员之间的传球建立一个网络，每个球员都是一个节点，每一次传球便将球员连接起来。球员之间传球次数用数组记录下来，在每一场球赛中，数组大小为14\*14，在一个赛季中，数组大小为30\*30。

第二，

第三，

最后，

**Contents**

**1.Introduction 1**

1.1 Background 2

1.2 Our work 3

**2.Problem restatement 4**

**3.Terminology 1**

3.1 Terms 2

3.2 Symbols 3

**4.Preparation of the Models 4**

4.1 Assumptions 5

4.2 Notations 6

**5.The Models 1**

5.1 Model 1 2

第3级 3

5.2 Model 2 2

第3级 3

5.3 Model 3 2

第3级 3

5.4 Model 4 2

第3级 3

**6.Strengths and Weaknesses 4**

6.1 Strengths 5

6.2 Weaknesses 6

**References**

**1.Introduction**

1.1 Background

足球运动有着悠久的历史，从推广以来就受到了全世界的喜爱，球迷遍布全世界。足球可以说是当今世界普及度最高的运动项目之一，足球这看似简单的一种运动，背后蕴藏着个人能力及团队协作的奥秘。随着时代的发展和科技的进步，足球运动员们和教练在技术上不断提升，给观众们呈现出一场场精彩纷呈的比赛。众所周知，一场精彩的足球赛离不开球员和球队贡献的，通过研究团队中每个人的行动，平衡队内力量，协调团队关系，合理安排上场时间和每个人所在的位置，可以在最大程度上得分。

1.2 Our work

我们的工作是从传统模型开始，在问题一中，我们首先建立每一场中十四位球员（其中包括三位后补球员）的传球次数统计图，再建立整个赛季中

我们的任务时为足球教练提供战略决策，在千变万化的球场上找出一些较为普遍适用在作战规律

在建立模型时，我们使用了Python，vscode，OneNote（开始胡编乱造，要删掉）

**2.Problem restatement（缩短）**

To respond to the Huskie coach’s requests, your team from ICM should use the provided data to address the following:

Create a network for the ball passing between players, where each player is a node and each pass constitutes a link between players. Use your passing network to identify network patterns, such as dyadic and triadic configurations and team formations. Also consider other structural indicators and network properties across the games. You should explore multiple scales such as, but not limited to, micro (pairwise) to macro (all players) when looking at interactions, and time such as short (minute-to-minute) to long (entire game or entire season).

Identify performance indicators that reflect successful teamwork (in addition to points or wins) such as diversity in the types of plays, coordination among players or distribution of contributions. You also may consider other team level processes, such as adaptability, flexibility, tempo, or flow. It may be important to clarify whether strategies are

universally effective or dependent on opponents’ counter-strategies. Use the performance indicators and team level processes that you have identified to create a model that captures structural, configurational, and dynamical aspects of teamwork.

Use the insights gained from your teamwork model to inform the coach about what kinds of structural strategies have been effective for the Huskies. Advise the coach on what changes the network analysis indicates that they should make next season to improve team success.

Your analysis of the Huskies has allowed you to consider group dynamics in a controlled setting of a team sport. Understanding the complex set of factors that make some groups perform better than others is critical for how societies develop and innovate. As our societies increasingly solve problems involving teams, can you generalize your findings to say something about how to design more effective teams? What other aspects of teamwork would need to be captured to develop generalized models of team performance?

**3.Terminology**

**4.Preparation of the Models**

**5.The Models**

5.1 model 1

影响整个游戏的结构指标和网络属性有多个indicators，比如：微观上两两球员之间的行为，宏观上多个球员之间的行为以及时间尺度。

两人传球评价指数，用于评价两人配合程度。在一场球赛中，宏观来看，球员相对于球场可以视为一个个节点，球场可以视为一个网络，每一次传球可以视为节点之间的连线。在每一场球赛中，共有十四位球员有机会上场，其中包括三名后补球员。我们调取passingevents中的数据，“F”表示前锋，“D”表示防守，“M”表示中场，“G”表示守门员，统计出每场球赛中每两位球员间传球的次数并绘制出38张14×14的矩阵和一张代表整个赛季的30×30的矩阵

其次我们考察多人传球网络体系。在三人传球的评价体系中，我们引入边权的概念，将三个节点用三条边连接起来，图的每条边上带的一个数值，其代表的含义可以是长度等等，这个值就是边权。

边的权值我们考虑由以下四部分组成：距离权，干扰权，类型权。其中，类型权是传球的类型，包括“头传球”、“简单传球”、“发射”、“高传球”、“手传球”、“聪明传球”、“交叉传球”之一。

通过生活经验和数据挖掘所发现的规律可以得出以下公式表格：

1. 传球类型权重表：



1. ，其中，x为球员到对方球门的横坐标，与受到的防守压力成负相关
2. 那么单次的价值评价：

传球评价，其中，为传球人坐标带入得到的值，为接球人坐标带入得到的值。

那么，由我们可以得出~~单次传球~~的价值评价图：



整个赛季单次评价图

时间尺度也是影响球赛结构指标的重要因素之一，球赛刚开始时，球员才稍加热身，身体还未舒展开而导致传球概率密度较小，5—10分钟过后，球员们已经活动开了，传球效率逐渐提高并大致趋于稳定，赛程过半时，球员们体力消耗较大，传球概率密度开始降低，即传球概率的增速减缓（虽然一场球赛中成功传球次数仍在增加，但传球失败概率开始增加），此后传球概率密度大致呈现下降趋势。纵观整个赛季中38场球赛球员们的传球概率密度，整体上看可以与单场球赛所展示出的概率密度变化趋势相同，大体体现为先增加，在趋于稳定一段时间后开始降低。我们以时间为横坐标，成功传球密度（公式化）为纵坐标作图：



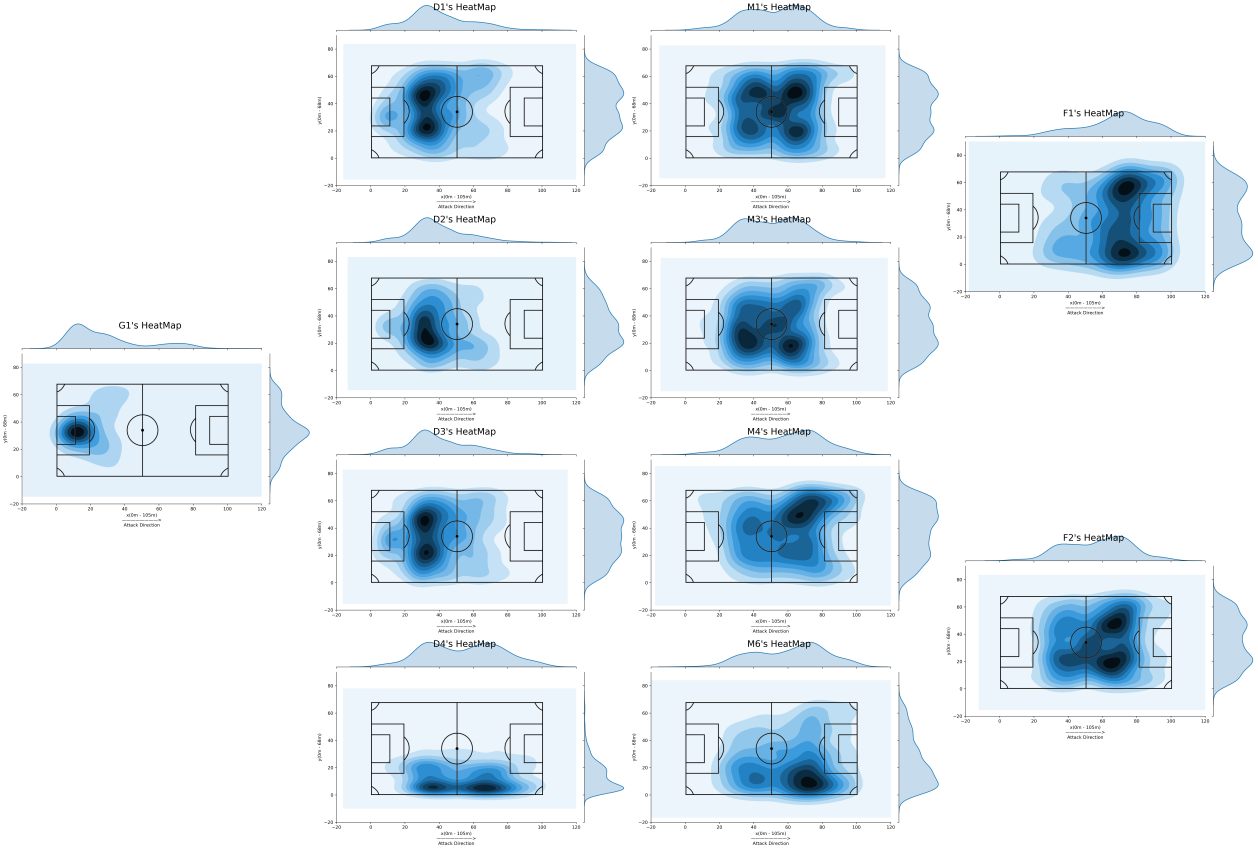
第1场比赛传球密度 整个赛季传球密度

5.2 model 2

成功团队合作有许多指标，通过fullevents的数据分析和实际经验，我们主要考虑以下指标：球员位置、球队阵型、球员协调配合和动态类型多样性。

为了考虑球员位置分布，我们采出每个球员在整个赛季中的位置坐标，用Python做出球员运动位置的热点图

颜色越深则表示出现在此处的概率较大，越浅表示出现的概率越小。（define heatmap怎么反映球员位置）主力11人的位置热点图如下：



球员位置分布热点图

在一场球赛中，球队的阵型对团队协作起到重要作用，我们考虑在一场球赛中球员阵型，我们采取每一场比赛中每一位球员的运动坐标，采用坐标对时间积分（写个积分公式）的方法，找出每场球赛中，每一位球员平均坐标。将在数据中可以获取（球员出现在Origin/Destination）的时间点作为新的横坐标，X或Y坐标作为新的纵坐标，得出函数。我们近似认为在任意两个有记录的时间点，球员在X或Y方向上匀速移动，这样就将离散型的数据集转换为了连续性的数据集（每个）。因此平均坐标，以X坐标为例，Y坐标同理：

将这十一位球员的位置标在图中绘制出每场球赛的阵型图，部分阵型图如下：（结论）



Match 1 and Match 11球员阵型图

球员间的协调配合是影响成功团队合作的关键因素，在球员阵型的基础上，考虑球员之间的联系，宏观上体现为球员间传球的次数多少，我们用线将球员之间连起来，用线的深浅表示传球次数的多少，颜色越深，传球次数越多，反之，颜色越浅，传球次数越少。（公式化定义）部分球员传球图如下：



第一场球员传球次数图

动态类型的多样性我们考虑其影响因素有比赛类型多样性、教练多样性、对手水平、主客场。比赛类型我们可以简单分为传球，进攻，失败和进球四个方面。进攻主要包括了射门控球率，传球，传球成功率，盘带过人，触球，角球，越位，失去控球权，防守，铲球（尝试铲球），抢球，空中对抗胜利。拦截主要包括了解围，封堵，扑救，犯规。

（这一段要改）在我们建立动态类型的模型时，我们以作为每场比赛评价标签，通过已经通过标准化等方法处理的数据来量化比赛的特征：

特征装箱 降为思想 PCA(Principal Component Analysis) 用One-Hot编码 --> 哑变量 维度增加 主客场，（讲了啥我又忘了，自己写吧TT 我只是个关键词记录者）描述一下得分（ratings）是什么？即包含什么。

通过Pearson相关系数的计算来估计出变量间两两特征相关性。

令矩阵，得：



动态类型的多样性评价图

我们怎样取出。。。数据集，以主客场为横坐标



主客场与得分的关系图

从图中可以看出，得分情况越好，净胜球数越高，得分情况与净胜球成正相关。主场得分情况整体上比客场得分情况要好也就是整体上看主场比客场净胜球多。

我们怎样取出。。。数据集，以Goal为横坐标，分别以Attack，Defence，Fail，Passes为纵坐标，黑线代表平均值，画出在不同Coach下的关系图。通过。。。方法得到的关系图如下：





不同教练指导下失误与进球之间关系图

从boxes square我们可以看出，在Coach 1，Coach 2，Coach 3，的指导下，在Coach 3指导下，球队整体的净胜球较多，其次是Coach 2和Coach 1。我们还可以看出，在同一个教练的指导下，净胜球越多，进攻强度越大且进攻强度的分布越集中，防守强度越大，传球率越高且分布越集中，失误率的绝对值越小。

我们怎样取出。。。数据集，通过。。。方法得到进攻、传球与净胜球之间关系图如下：



进攻、传球与净胜球之间关系图

从图中我们可以看出，在不同净胜球数下，进攻和传球大体上为正相关

随着净胜球数的增加，净胜球数越多，攻击和传球系数越高，且分布越集中，攻击和传球系数的方差较小（要解释攻击，防守，传球系数如何得到）。我们可以得出结论：在一场球赛乃至整个赛季，净胜球越多说明比赛的传球率越高，进攻强度越强。

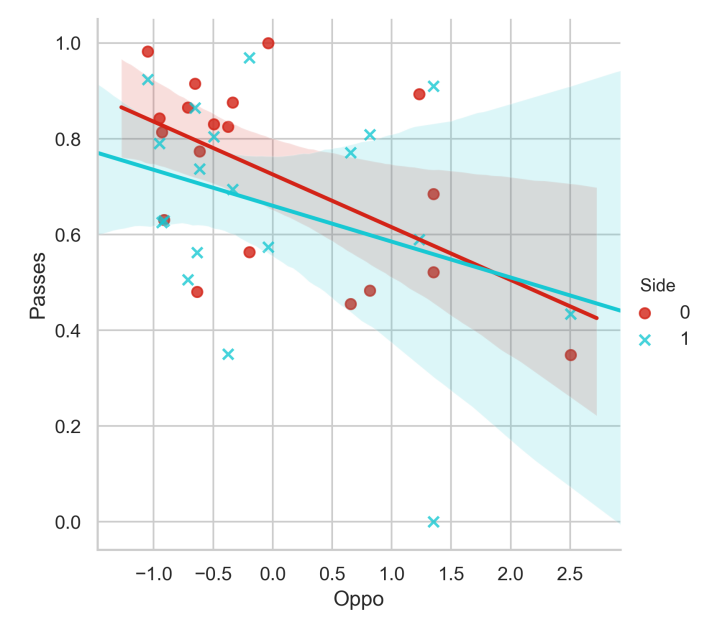
我们怎样取出。数据集，通过。方法得到防守、失败与净胜球之间关系图如下：



防守、失败与净胜球之间关系图

净胜球数小于零时，失误系数范围主要为-1.0 — -0.2，防守系数范围主要为0—0.5；净胜球数等于零时，失误系数范围主要为-0.9—0，防守系数范围主要为0—1.0；净胜球数大于零时，失误系数范围主要为-0.6— -0.2，防守系数范围主要为0—0.8。随着净胜球数的增加，净胜球数越多，防守系数绝对值越高，失误系数绝对值越小，且分布越集中，防守和失误（用失误代替失败会不会更好）系数的方差较小（要解释攻击，防守，传球系数如何得到）。观察：图2左1的点分布在下方，因此防守不好会导致输球；右1左半边没有点，因此期望赢球失误不能多。

我们怎样取出。。数据集，通过。。。方法得到进攻和防守之间关系图如下：

进攻和防守之间关系图 对手水平与传球关系图

从左图中我们可以看出数据重心分布在左下角，认为整个赛季上（进攻表现）显著优于（防守表现）。从图中我们可以看出不论是在主场还是客场，，但主场更可能有较小提升；结论是对手水平越高，我方传球率越低。

我们怎样取出。。。数据集，通过。。。方法得到对手水平与传球关系图如上图所示。

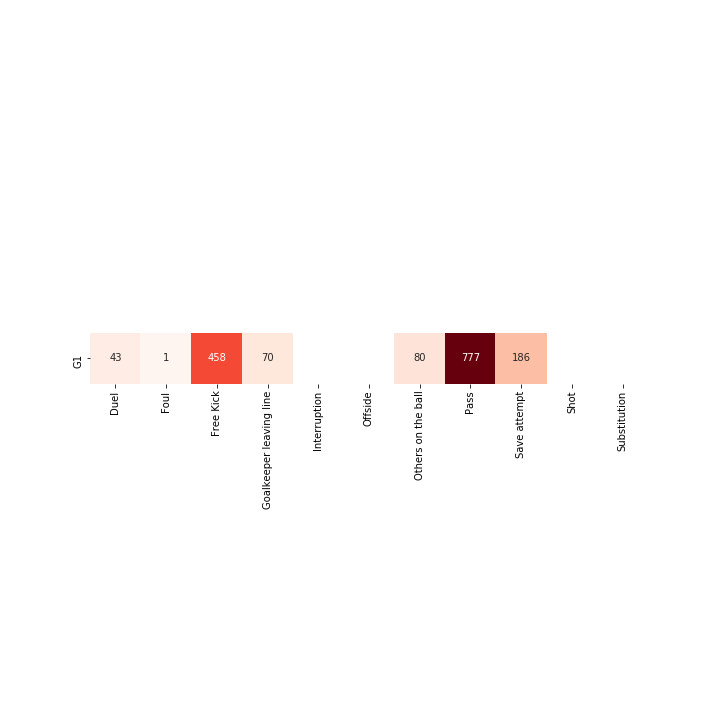
（对手不是非常强时，主场的红点分布较高）

交叉验证（我觉得可以不要）（这玩意不是在这的）

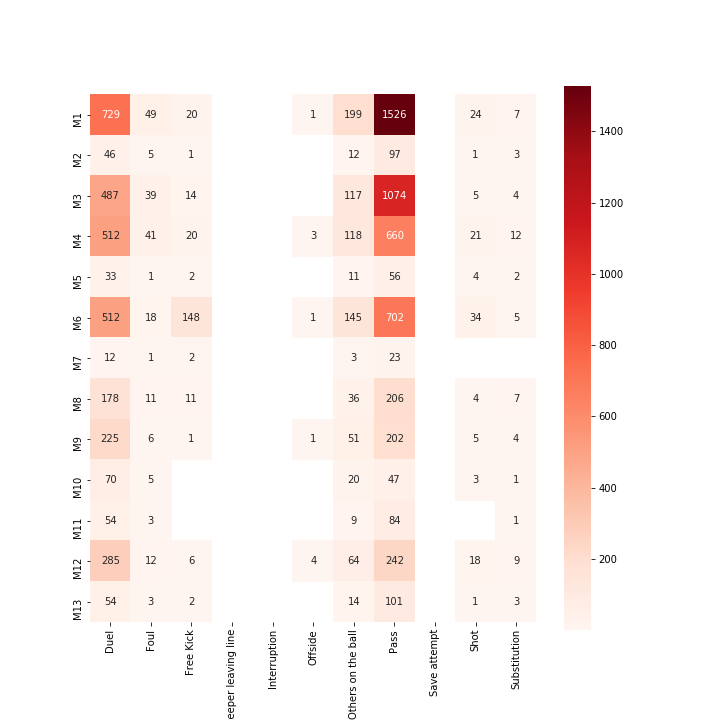
5.3 model 3

一次成功的团队合作离不开结构策略。作为一名成功的教练应该具备较好的统筹规划，协调合作，人员安排能力。我们认为，具体的结构策略应该最主要体现在以下两个个方面：球员位置安排和球队阵型。此外，还应考虑球员间默契度，主客场影响，教练安排。

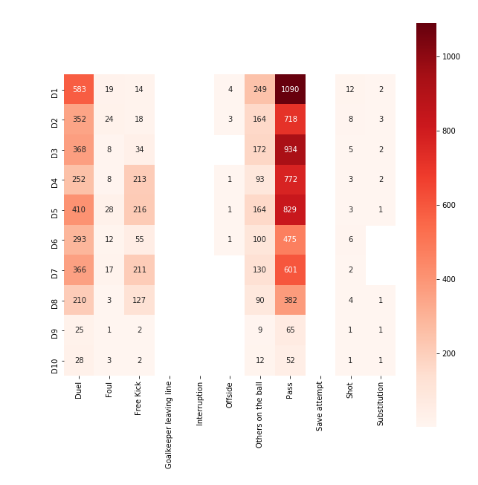
在考虑球员位置安排时，我们首先需要选出球队中每个位置上表现较好的球员，我们考虑守门员，前锋，中场，后卫四个位置不同球员的贡献值。我们在fullevents表格中采集哈士奇球队三十位球员的EventType，并以其为横轴，球员编号为纵轴，统计出每个球员在整个赛季中每个EventType共执行了多少次，用颜色的深浅来表示次数的多少，以下分别为守门员，前锋，中场和后卫的EventType统计图：



守门员EventTypes统计图

前锋EventTypes统计图 中场EventTypes统计图



后卫EventTypes统计图

由以上四个图我们可以看出G的主要EventType是pass，F的主要EventType是pass和Duel，D的主要EventType是pass,Duel和Freekick,M的主要EventType是pass,Duel,Others on the ball和Freekick。F中贡献最大的是F2，其次是F1,F6,F5,F4。M中贡献最大的是M1，其次是M3,M4,M6。D中贡献最大的是D1，其次是D3,D5,D4,D2,D7,D6,D8。

最后我们分析球队的29位球员（除守门员）分别在G,F,M位置上的表现情况，权衡球员各方面情况（需要具体写吗），为球员在不同位置上的表现打分。颜色越红表示越适合这个位置，反之越蓝则表示越不适合。

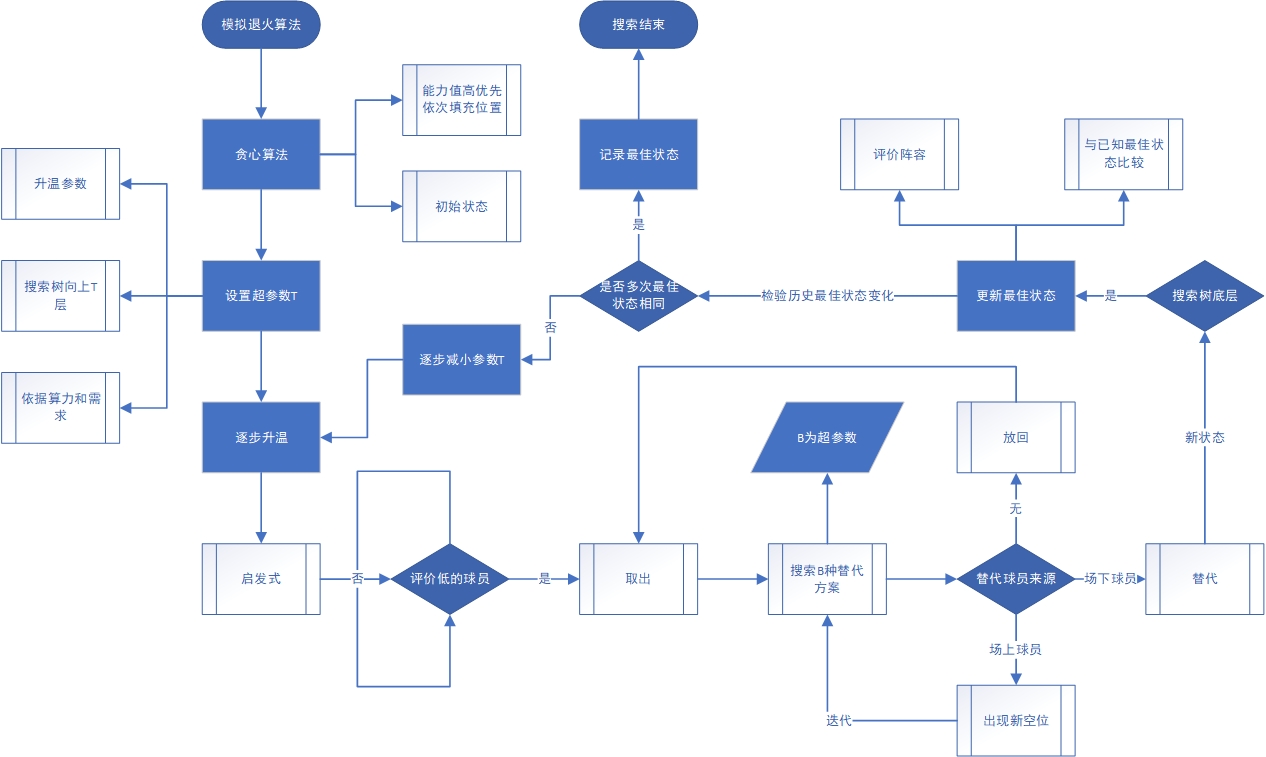


不同球员在不同位置评分表

我们分析整个赛季三十八场比赛中主力阵容/首发阵容line-up，可以得到球队的阵型主要有三种，442,433,4321。在考虑球队阵容时，我们采用模拟退火算法。（过渡）

该模型的目标是要找到一个最优的有序组合，使场上11人在各自位置的能力之和最大。在搜索树极为庞大、算力资源有限的情况下，我们选择模拟退火算法。模拟退火算法的思想来源于对固体退火降温过程的模拟：将固体加温至充分高，再让其徐徐冷却。模拟退火算法在迭代的过程中不断能够接受使目标函数向好的方向前进的解；其主要优点之一就是能以一定的概率接受目标函数值不太好的状态，这使得算法即便落入局部最优的陷阱中，理论上经过足够长的时间也可跳出局部最优。令随算法进程递减的控制参数担当固体退火过程中温度的角色，则对于的每一取值，算法采用接受准则，持续进行迭代过程而达到该温度下的平衡点。把场上11个位置有序排列，用11位的30进制格雷码表示当前状态；例如格雷码0A1GRD739KI表示11个位置依次有第0,10,1,16,26,13,7,3,9,11,18号球员。模拟退火算法的具体步骤如下：

1. 给定冷却进度表参数及迭代初始解.以及,其中冷却进度表参数包括：控制参数的初值，衰减函数，终值以及链长度；
2. 参数时，按照如下过程作次试探搜索：
   1. 根据当前解的性质，产生一个随机偏移量，从而得到一个当前解邻域的新的试探点；
   2. 产生一个在区间上均匀分布的随机数,计算出在给定当前迭代点和温度下与接受准则相对应的转移概率：
   3. 试探搜索小于次，返回步骤1，否则进入步骤3；
3. 根据给定的温度衰减函数产生新的温度控制参数，及链长度，转入步骤2，进入下一温度点的平衡点寻优。



在实际试探搜索中，我们很可能现入局部最优，需要进行判定以退出。假设搜索进入局部最优点，那么由模拟退火算法的求解过程可知，当前解的优化程度小于当前最优解的优化程度的时候,新解被接受的概率为,而当温度足够低的时候，较差解被接受的概率趋近于。依据最近的次搜索中都没有优化程度更高的解出现这一特征，可以根据具体问题确定阐值而后判定搜索己经进入局部最优，且由于温度过低，若要尽快跳出局部最优就需要进行升温。

考虑完主要策略后，我们考虑以下四个次要影响因素：球员间默契度，主客场影响和教练安排。

首先，选择默契度高的小分队有利于提高传球和进球的效率。默契度高的小组往往配合能力较强，有助于比赛的成功。通过我们的调查发现，传球效率较高的球员往往适应能力较强，与其他球员的配合度也较好。所以应尽量安排贡献值较高的球员上场。

1. 主客场因素也是必须要考虑的，有的球员适应性较强，在主场和客场都能较好的发挥出原有的水平，我们称之为适宜性强的选手。而有些适应性较差的球员只在主场发挥出原有的水平，甚至有时会超常发挥，环境对他的表现有较大的影响，在客场时，环境的变化有较大可能会使他不能发挥出原有水平。那么在主客场时，应该安排不同的球员上场。

最后，教练安排上，整个赛季中，Coach 1,Coach 2,Coach 3,分别指导了9,5,24场比赛，通过我们在第二题中的数据分析也可得出，Coach 3的水平较高。

纵观整个模型，为了在下个赛季中提高球队成功率，我们团队给出的建议是，球队聘用Coach 3作为球队主教练，采用442的line-up，将F1,F6,F2,M3,M1,M6,D3,D1,D2,D5作为主力球员并将他们的位置按下图安排。



球员阵型图

5.4 model 4