

基于机器学习和运筹学的排球排兵布阵研究

(浙江大学控制科学与工程学院, 杭州, 中国, 310000)

摘要: 鉴于目前国内大多数中低水平排球队缺乏高水平教练, 队员位置判断和全队阵容规划能力有所不足, 本团队以浙江大学竺可桢学院男排和控制学院男排为例, 以运动员的各项能力指标、身体素质指标为输入数据, 结合机器学习分类算法、能力评估模型建立和运筹学规划完成队员位置判断、能力评估以及全队规划。在机器学习判断位置方面, 本团队使用多种算法, 其中决策树算法、K 最近邻算法、支持向量机算法和一致性预测算法分别为每一位运动员提供其最适位置的一个或多个标签, 同时主成分分析法的使用可以使不同位置运动员的样本在低维特征空间中的分布显现, 以数据驱动方式提升判位客观性; 在能力评估方面, 本团队结合专家评审团队和层次分析法, 对每一位成员给出进攻、防守、前排、组织、综合能力评价, 有利于提升其训练的针对性; 在运筹规划方面, 本团队结合 0/1 规划和对策论, 由简到繁, 给出不同条件、不同目标下的最优阵容, 通过与实际比赛阵容相比较, 印证规划的可靠性。整个项目以数据驱动和数据挖掘为起点, 结合实例创新地结合应用一致性预测和运筹学于运动队规划, 有利于提升运动队的管理和能力优化。

关键词: 机器学习, 一致性预测, 决策树, 层次分析法, 对策论, 0/1 规划

引言

随着排球运动的发展, 运动员的技术以及战术都日趋完善, 竞技能力日趋提高, 从而使得运动员之间的组合以及相互之间的配合显得更加重要。在考虑阵容的时候, 除了需要考虑运动员的个人能力外, 更重要的是需要考虑其对于整个团队的贡献。而球队的阵容组合往往决定了球队的主流打法, 体现了一个球队的实力和竞技水平, 同时也在比赛的时候展现出一种独特的风格。

在比赛中, 有时能够明显感受到有的队伍进攻很犀利, 但弱于防守, 而有的队则恰好相反, 强于防守而欠于进攻。而在一个排球队中, 有的队员擅长防守, 而有的队员又长于进攻, 从而使得在阵容排布时需要在进攻和防守之间进行取舍。实际情况中, 队员的能力不只有进攻和防守两个能力指标, 同时具有二传组织能力、身体综合素质等指标。除以上外, 对于团队竞技, 有一个指标成为队伍的核心优势非常重要, 就是队员之间的默契程度。这些指标如何衡量, 在这些指标下如何判断队员适合什么位置, 在决策的时候如何根据目标进行取舍, 这些问题都是阵容排布时需要考量的问题。而如今科学理论已经为此类问题提供了两个强有力的工具——机器学习和运筹学, 机器学习基于大数据分析对确定球员的最佳位置具有很大的帮助, 而运筹学则对阵容排布具有重要的指导意义。¹

如今在诸多中低水平排球队伍中因缺乏高水平教练, 欠缺合理安排阵容的能力。通过

对文献的查阅，学界针对此类问题的研究较少，本篇文章旨在运用运筹学和机器学习的知识，以运动员的各项数据指标为原始数据，结合专家评委的意见建立能力评估模型，设计了根据队员指标数据获取队员适应位置的方案，同时给出了优化队伍阵容排布的方案。本文中提到的方法具有普遍适用性，因而可以推广至许多其他球类运动。

1. 方法

本项目的整体流程如下流程图图所示：首先获取运动队中每一位运动员的各项参数指标，主要包括身体素质指标、技术能力指标等等，随后输入数据进入算法平台，经历三种算法处理：机器学习分类算法、评价模型确定算法和运筹学规划算法，目标是获取某一运动员的最适位置、能力评价图谱和整支团队的阵容规划结果。

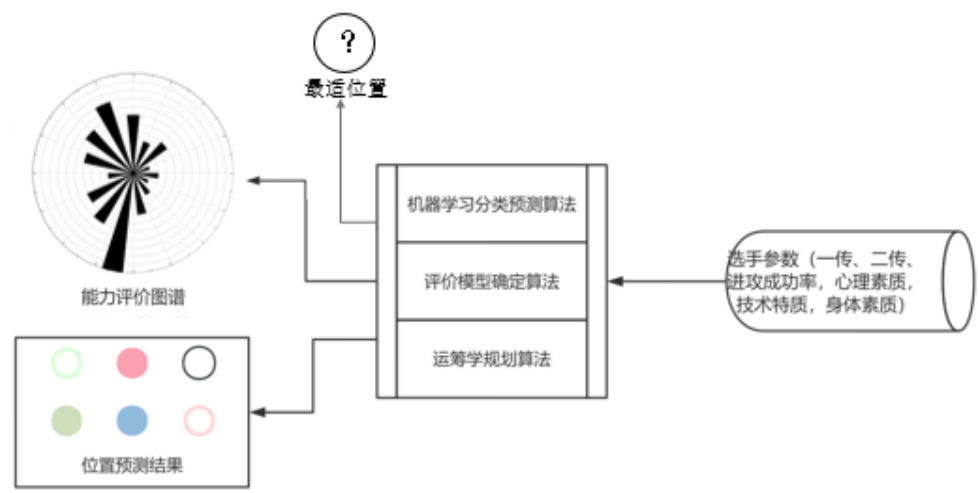


图 1-1 基于机器学习和运筹学的排球排兵布阵流程图

在数据方面，我们设计了 15 个客观指标、9 个主观指标、5 个基本指标，各项数据均来自于 2017-2018 学年的排球队联合训练以及学年内四次校级排球比赛中的数据记录。各项指标设计如下图所示：其中客观指标由场下队员严格记录，主观指标由以浙江大学公共体育与艺术部的高水平女排教练王先驰老师和诸多浙江大学校男排、女排成员、竺可桢学院男排、控制学院男排历任队长组成的专业评审团队给出打分，来平衡单一客观指标带来的评价不全面性。

表 1-1 不同类别各项参数指标

指标类别	指标数目	各项指标
------	------	------

客观指标	15	一传成功率，防守成功率，发球成功率，吊球成功率，扣球成功率，二传成功率，身高，体重，BMI，肺活量，50m 成绩，1000m 成绩，立定跳远成绩，坐位体前屈成绩，引体向上成绩
主观指标	9	一传质量评分，防守质量评分，发球质量评分，吊球质量评分，二传质量评分，场上意识评分，心理素质评分，有效拦网评分，氛围带动能力
基本指标	5	姓名，样本编号，年级，训练集/测试集，标签/参考标签 (主攻/副攻/二传与接应/自由人)

1.1 机器学习模型

要搭配出最佳的队伍，首先需要知道球队中的每名队员最适合的位置。这需要根据排球球队球员们在训练和比赛中的表现统计出各方面的评分和成功率，进行每个球员适合的位置的预测。从本质上来说这是一个分类问题。本次使用的机器学习算法主要有支持向量机 (SVM)、K-最邻近算法 (KNN)、决策树 (DT) 和在一致性框架下的 K-最邻近算法 (CP-KNN)。最后使用主成分分析法 (PCA) 将训练集的特征降至 3 维进行作图，从而更好的观察训练集的分布，来分析机器学习算法的结果。

1.1.1 支持向量机 (SVM)

支持向量机² (Support Vector Machine) 算法通过选取适当的核函数和参数 C ，将特征向量映射到高维空间，并通过在高维空间中选取距离两类样本“间隔”最大的划分超平面将不同种类的样本分开，并以此划分超平面作为后续进行标签预测的标准。

1.1.2 K-最邻近算法 (KNN)

K-最邻近³ (k-NearestNeighbor) 算法通过计算得出测试点附近最近的 K 个训练集中的点，统计这 K 个点中最多的类别来作为测试点的预测标签。为保证决策的成功进行， K 一般取奇数。计算点的距离可以采用距离为欧几里得距离、明可夫斯基距离、曼哈顿距离等。

1.1.3 决策树 (DT)

决策树⁴ (Decision Tree) 算法是基于信息增益、增益率或基尼指数等指标, 通过在特征中选取最优划分, 随着划分的不断进行, 不断使得决策树的分支节点包含的样本趋近于同一类别, 从而基于训练集产生一颗能够更好地处理未见新样本的决策树。

1.1.4 一致性框架 (CP)

上述的三种方法只能给出测试点的一种预测结果, 而在排球规划问题中, 存在多面手问题, 即一个人适合多个位置, 所以需要使用一种方法能给出多预测结果。

一致性预测框架⁵⁶⁷⁸⁹ (Conference Predict) 与 KNN 结合后, 通过如下计算可以得到测试点为任意一个标签时的奇异值度量。

$$\alpha_i = \frac{\sum_{s=1}^k d(x_{is}, y_{is})}{\sum_{s=1}^k d(x_{js}, y_{js})} \quad (1)$$

$d(x_{is}, y_{is})$ 表示两点之间的欧式距离

当奇异值越小时, (x_i, y_i) 与具有同类标签的点越近, 与不同类标签的点越远。有了如上的奇异值度量算法, 一致性预测框架可以将每一个样本的各个可能标签代入计算, 并且获取某一特定“特征-标签组合”与其他已知样本之间一致程度的衡量:

$$p^y = \frac{|\{i=1, \dots, n | \alpha_i^y > \alpha_n^y\}|}{n} \quad (2)$$

其中 p^y 即代表了当新的待预测样本 x_n 的标签被预测为 y 时候, 该预测与样本空间中的已知信息的一致程度大小。 p^y 越大, 则如果将该特征 x_n 预测为标签 y 就将获得与已知样本更一致的结果, 即更加可靠。

于是设定一个显著性水平参数, $\epsilon \in (0, 1)$, 当 p^y 大于 ϵ 时就认为这个预测标签显著, 可以作为预测标签输出, 则 CP-KNN 算法对每个测试点的输出是一组多预测标签:

$$\Gamma^\epsilon(z_1, z_2, \dots, z_{n-2}, z_{n-1}, z_n) = \{y | p^y > \epsilon\} \quad (3)$$

1.1.5 主成分分析 (PCA)

主成分分析法 (Principal Component Analysis, PCA) 从实质上来说是一个变换坐标轴的过程, 新坐标系的选择是由数据本身决定的。第一个坐标轴选择的是原始数据中方差最大的方向; 第二个坐标轴选择的是与第一个坐标轴正交且具有最大方差的方向; 第三个坐标轴选择的是与第一个坐标轴和第二个坐标轴都正交且具有最大方差的方向; 依次类推,

重复次数为原始数据的维数。

经过 PCA 处理后，特征空间的坐标系发生了变化，数据的表达发生了变化，但是其空间相对位置和距离是没有变化的。新坐标系的坐标轴顺序是按其包含的方差占总方差的比重进行排序的，大部分方差都包含在前几个坐标轴上，这几个坐标便是原始数据中的主成分。对于剩下的坐标轴，包含的方差很少，可以直接忽略，从而完成了降维的过程。

1.2 层次分析法

层次分析法(AHP)是一种多准则多因素决策方法。该方法主要解决最优方案的评估抉择问题，比较适合将人对难以量化的批判标准转化为客观量化的指标¹⁰。本项目中，层次分析法主要用来得到每个主评价层次下的各个二级评价指标的权重，结合各个指标的评价分数，得到进攻能力、防守能力、前排反应、综合素质和组织能力等 5 个主层次的各自的评价分数。这些评价分数会被第 3 部分运筹规划模型所用。

值得注意的是，一般的层次分析法，有如下四个步骤：(i) 建立递阶层次结构模型；(ii) 构造出各层次中的所有判断矩阵；(iii) 层次单排序及一致性检验；(iv) 层次总排序及一致性检验。但是本项目中没有进行五个一级指标的重要性判断，即省略了第(iii)步的内容。这样做的原因是，本项目使用层次分析法的目的是给数学规划提供数据，一级指标的分数会在更加复杂和详细的数学模型中，综合分析得出针对所有队员和所有位置的结果，而不是像在传统的层次分析法中那样，针对一个特定的位置，通过一级指标的重要度加权分数，得到这个位置最适合的选手。

得到评价矩阵的方法有专家调查法，优序图法等¹¹。本项目中使用专家调查法，由一名排队队长和三名普通队员组成评价队，分别独立地对层级结构指标的重要度按 1~9 分进行打分，得到重要度分数的中位数。之后重新进行第二次调查，每个人对上一一次的打分结果给出认同度，统计出认同度低的几项。第三轮调查中，针对认同度低的几项的，分别再次给出自己的重要度和理由，综合讨论后得出最后的分数。对矩阵进行一致性检验后，均通过。一致性检验时，通过如下所示的公式求出一致性指标 CI，CI 越接近于 0，一致性越好。

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (4)$$

最后，对每个一级指标下的共 5 个评价矩阵，分别得出特征值，最大的特征值所对应的特征向量即为每个指标的权重。同时在 matlab 中读取各指标的分数，结合权重加权平均得到每个队员每个一级指标下的分数。

1.3 运筹规划模型

1.3.1 Level 1 “群魔乱舞”——全部混合规划

根据前述工作得到的个人能力数值，挑选出能力较强的 30 位球员，设为 $x_i (i = 1 \dots 30)$ 。并记录下他们的五个能力指标的向量，分别为：进攻能力得分（简称 JG）、防守能力得分（简称 FS）、前排反应得分（简称 QP）、综合素质得分（简称 ZH）、组织能力得分（简称 ZZ）。在当前环节中，不考虑球员的位置属性来选择六个球员，选择的依据为某一能力之和最大化且其他能力值大于平均水平。这样，我们可以得到五种排兵布阵的方法，这些方法分别达到某种能力的最大化。具体模型如下：

$$\begin{aligned} \max z = & \sum_{i=1}^{30} JG_i * x_i \\ \text{s. t. } & \begin{cases} \sum_{i=1}^{30} FS_j * x_i \geq fs_{ave} * 6 \\ \sum_{i=1}^{30} QP_i * x_i \geq qp_{ave} * 6 \\ \sum_{i=1}^{30} ZH_i * x_i \geq zh_{ave} * 6 \\ \sum_{i=1}^{30} ZZ_i * x_i \geq zz_{ave} * 6 \\ 0 \leq x_i \leq 1, \in integer \quad j = 1 \dots 30 \\ \sum_{i=1}^{30} x_i = 6 \end{cases} \end{aligned} \quad (5)$$

其中目标的进攻能力得分 JG 可以换为其他四个能力指标，相应地，约束条件中要做适当的更改。

1.3.2 Level 2 “各司其职”——单标签规划以及与 AHP 的对比

在前一环节的基础上，给每位球员添加位置标签，分别为主攻 $x1_i$ 、副攻 $x2_i$ 和二传 $x3_i$ 。每个位置的总人数为 2 人。根据排球实战经验，对于二传标签的球员，其进攻能力置为零，对于主攻及副攻的球员，其组织能力置为零，其余属性不作变化。同样地，我们依次选择某一能力之和最大化来作为规划的目标，并且前述的约束条件依然在这个环节成立。具体模型如下：

$$\max z = \sum_i JG1_i * x1_i + \sum_i JG2_i * x2_i$$

$$s. t. \left\{ \begin{array}{l} \sum_i FS1_i * x1_i + \sum_i FS2_i * x2_i \\ + \sum_i FS3_i * x3_i \geq fs_{ave} * 6 \\ \sum_i QP1_i * x1_i + \sum_i QP2_i * x2_i \\ + \sum_i QP3_i * x3_i \geq qp_{ave} * 6 \\ \sum_i ZH1_i * x1_i + \sum_i ZH2_i * x2_i \\ + \sum_i ZH3_i * x3_i \geq zh_{ave} * 6 \\ \sum_i ZZ3_i * x3_i \geq zz_{ave} * 2 \\ 0 \leq x1_i \leq 1, \in integer \\ 0 \leq x2_i \leq 1, \in integer \\ 0 \leq x3_i \leq 1, \in integer \\ \sum x1_i + \sum x2_i + \sum x3_i = 6 \\ \sum_i x1_i = 2 \\ \sum_i x2_i = 2 \\ \sum_i x3_i = 2 \end{array} \right. \quad (6)$$

同理，其中目标的进攻能力得分 JG 可以换为其他四个能力指标，相应地，约束条件中要做适当的更改。

1.3.3 Level 3 “全面发展”——多面手问题

考虑到实际情况中，一个球员可能会有多个其擅长的位置。所以根据预测结果，将每个球员赋予两个标签。这样，我们得到变量标签一主攻 $x1_i$ 、副攻 $x2_i$ 和二传 $x3_i$ 与标签二的主攻 $y1_i$ 、副攻 $y2_i$ 和二传 $y3_i$ 。在球员可以胜任多个位置的情况下，对阵容进行规划。值得注意的是，每个球员最多以自己的一个标签位置出现在阵容中而不可以一人分饰两角。其余约束条件仍然与上一环节相同。具体模型如下：

$$max z = \sum_i JG11_i * x1_i + \sum_i JG12_i * x2_i + \sum_i JG21_i * y1_i + \sum_i JG22_i * y2_i$$

$$s. t. \left\{ \begin{array}{l} FS1 * X + FS2 * Y \geq fs_{ave} * 6 \\ QP1 * X + QP2 * Y \geq qp_{ave} * 6 \\ ZH1 * X + ZH2 * Y \geq zh_{ave} * 6 \\ \sum_i ZZ13_i * x3_i \\ + \sum_i ZZ23_i * y3_i \geq zz_{ave} * 2 \\ 0 \leq x1_i + y1_i \leq 1, \in integer \\ 0 \leq x2_i + y2_i \leq 1, \in integer \\ 0 \leq x3_i + y3_i \leq 1, \in integer \\ \sum x1_i + \sum x2_i + \sum x3_i + \sum y1_i \\ + \sum y2_i + \sum y3_i = 6 \\ \sum x1_i + \sum y1_i = 2 \\ \sum x1_i + \sum y1_i = 2 \\ \sum x1_i + \sum y1_i = 2 \end{array} \right. \quad (7)$$

同理，其中目标的进攻能力得分 JG 可以换为其他四个能力指标，相应地，约束条件中要做适当的更改。

1.3.4 Level 4 “固若金汤”——自由人规划

在这个环节中，先从之前的 30 名球员中挑选两名并赋予其新的位置标签——自由人 (x_4)。挑选的标准是

$$\begin{cases} fs \geq M \\ \max z = \frac{fs}{jg} \end{cases} \quad (8)$$

其中，M 是某个阈值。在防守能力大于该阈值的基础上，取 z 值最大的两位球员作为自由人的人选。

在添加了自由人标签后，规划的条件与约束也有改变。进攻能力只包括主攻与副攻，副攻不参与防守，前排拦截时自由人不参与，组织能力计算时只包括二传与自由人。同时阵容总人数提升至 7 人，即在 level 2 的基础上在阵容中添加一名自由人。具体模型如下：

$$\begin{aligned} \max z = & \sum_i JG1_i * x1_i + \sum_i JG2_i * x2_i \\ \text{s.t.} \left\{ \begin{array}{l} \sum_i FS1_i * x1_i + \sum_i FS3_i * x3_i \\ + \sum_i FS4_i * x4_i \geq fs_{ave} * 5 \\ \sum_i QP1_i * x1_i + \sum_i QP2_i * x2_i \\ + \sum_i QP3_i * x3_i \geq qp_{ave} * 6 \\ \sum_i ZH1_i * x1_i + \sum_i ZH2_i * x2_i \\ + \sum_i ZH3_i * x3_i + \sum_i ZH4_i * x4_i \geq zh_{ave} * 7 \\ \sum_i ZZ3_i * x3_i + \sum_i ZZ4_i * x4_i \geq zz_{ave} * 3 \\ 0 \leq x1_i \leq 1, \in integer \\ 0 \leq x2_i \leq 1, \in integer \\ 0 \leq x3_i \leq 1, \in integer \\ 0 \leq x4_i \leq 1, \in integer \\ \sum x1_i + \sum x2_i + \sum x3_i + \sum x4_i = 6 \\ \sum_i x1_i = 2 \\ \sum_i x2_i = 2 \\ \sum_i x3_i = 2 \\ \sum x4_i = 1 \end{array} \right. \end{aligned} \quad (9)$$

同理，其中目标的进攻能力得分 JG 可以换为其他四个能力指标，相应地，约束条件中要做适当的更改。

1.3.5 Level 5 “打到你服”——接应二传以及主攻、副攻、二传位置实际考虑

在本环节，从之前标签为二传及主攻的球员里选出两名接应二传 ($x5_i$)，选择的标准如下：

$$\begin{cases} zz \geq M \\ \max z = \frac{qp+jg+zh}{zz+zh} \end{cases} \quad (10)$$

在添加了接应二传的标签后，规划的条件与约束做出相应更改。计算进攻能力时，只

考虑主副攻及接应二传，计算防守能力时，只考虑主攻、自由人、二传及接应二传，前排拦截时不考虑自由人，组织时，只有二传、接应二传和自由人。并在计算进攻能力值时，将主攻的计入量放大；计算前排能力值时，将副攻的计入量放大；计算组织能力值时，将二传的计入量放大。具体方法表现为乘以一个放大系数 (α, β, γ) 。而在位置的人数限制上，派出主攻、副攻各两名，二传，接应二传和自由人各一名。具体模型如下：

$$\begin{aligned}
 \max z = & \sum_i JG1_i * \alpha * x1_i + \sum_i JG2_i * x2_i + \sum_i JG5_i * x5_i \\
 \text{s.t. } & \left\{ \begin{aligned}
 & \sum_i FS1_i * x1_i + \sum_i FS3_i * x3_i \\
 & + \sum_i FS4_i * x4_i + \sum_i FS5_i * x5_i \geq fs_{ave} * 5 \\
 & \sum_i QP1_i * x1_i + \sum_i QP2_i * \beta * x2_i \\
 & + \sum_i QP3_i * x3_i + \sum_i QP5_i * x5_i \geq qp_{ave} * 6 \\
 & \sum_i ZH1_i * x1_i + \sum_i ZH2_i * x2_i \\
 & + \sum_i ZH3_i * x3_i + \sum_i ZH4_i * x4_i \\
 & + \sum_i ZH5_i * x5_i \geq zh_{ave} * 7 \\
 & \sum_i ZZ3_i * \gamma * x3_i + \sum_i ZZ4_i * x4_i \\
 & + \sum_i ZZ5_i * x5_i \geq zz_{ave} * 3 \\
 & 0 \leq x1_i \leq 1, \in integer \\
 & 0 \leq x2_i \leq 1, \in integer \\
 & 0 \leq x3_i \leq 1, \in integer \\
 & 0 \leq x4_i \leq 1, \in integer \\
 & 0 \leq x5_i \leq 1, \in integer \\
 & \sum x1_i + \sum x2_i + \sum x3_i \\
 & + \sum x4_i + \sum x5_i = 6 \\
 & \sum x1_i = 2 \\
 & \sum x2_i = 2 \\
 & \sum x3_i = 1 \\
 & \sum x4_i = 1 \\
 & \sum x5_i = 1
 \end{aligned} \right. \quad (11)
 \end{aligned}$$

同理，其中目标的进攻能力得分 JG 可以换为其他四个能力指标，相应地，约束条件中要做适当的更改。

1.3.6 Level 6 “精挑细选” —— “12 人大名单” 问题

本环节中，选出 12 名球员组成参加比赛阵容的大名单，具体做法是以上一环节中以防守能力作为目标约束的求解结果选出的球员作为其中必选的七个球员。再将规划的目标换为进攻能力，这样又得到七名球员，从两次规划的结果中，筛选掉重复出现的球员，得到最后大名单的十二名球员。

1.3.7 Level 7 “随机应变” —— 对策应用

Level 7 采用 Level 5 规划的结果，有特定的自由人和接应二传，按照不同的目标(最大化进攻能力、最大化防守能力、最大化前排能力、最大化综合素质、最大化组织能力)得到

了五套方案，针对不同的对手我们可以从这 5 套方案中任选一套，那么现在我们的策略集可以表示为

$$S_w = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\} \quad (12)$$

假设对手也有 5 套方案，且我们已知他们不同方案的综合实力。对于我方的综合实力，根据经验可以按照下列计算公式得出：

$$ours_Ability = \sum_{i=1}^7 (20 \times JG_i + 30 \times FS_i + 10 \times QP_i + 20 \times ZH_i + 20 \times ZZ_i) \quad (13)$$

即被选中的 7 个人在进攻等 5 项能力上的得分乘以一个系数相加和即可得到我方的总和实力。在使用 MATLAB 编程进行模拟过程中，假设对方选手的各项能力都是在我们队员相应的能力值上加上一个随机数得到的，这个随机数的取值范围为 $[-0.035, 0.035]$ ，且其在该区间上服从高斯分布，MATLAB 的编程语言如下：

$$Enemy_Ability = ours_Ability + (0.07 * (rand(5,5) - 0.5)) \quad (14)$$

而在决策过程中，定义的决策函数如下：

$$H_{ij} = outs_Ability_i - enemy_Ability_j \quad (i, j = 1, 2, 3, 4, 5) \quad (15)$$

即我方采取第 i 套方案和敌方采取第 j 套方案时，我们的综合实力减去对方的综合实力，这样便可得到我方的赢得矩阵 $H_{5 \times 5}$ 。根据矩阵对策的假设——每个局中人对对方的实力都了如指掌且都是理性的，那么根据赢得矩阵 H 我们可以利用如下公式

$$\max_i \min_j H_{ij} = \min_j \max_i H_{ij} = H_{i^*j^*} \quad (16)$$

验证 H 是否存在平衡解、对于局中人是否存在最优纯策略¹²。

2 各模型结果

2.1 机器学习结果

本模型的数据集分为两部分，训练集来自于篮球队已有明确主攻、副攻、二传分工的老队员的测试数据，共计 26 人；训练集来自于刚加入排球队的新队员的测试数据，他们还未有明确分工，共计 12 人。并由排球队老师和队长就新队员的情况给出一个预估分工作为测试集的标准输出。

数据标签分为主攻（1）、副攻（2）、二传（3）和自由人（4）四种，观察数据集后发现，由于自由人需求较少，人员也较少。训练集中仅有一位自由人，测试集中也只有两位

自由人。考虑到机器学习需要较多的数据来进行训练，所以去除所有自由人数据。最终共计训练集 25 组数据，测试集 10 组数据。自由人的选择由规划算法给出。

在进行机器学习前，对所有数据进行了归一化处理，让所有特征的取值均位于 0-1 之间，保证每个特征的权重一样。

2.1.1 常规算法结果

常规算法是指 SVM、KNN (K=3) 和 DT 三种算法，测试方法分为两种：对训练集使用留一法进行测试；对测试集进行预测，然后和标准输出对比获得正确率。

SVM 算法只能计算二分类问题，将问题划分为三个二分类，正确率取三个二分类问题的平均值。结果如下表所示：

表 2.1-1 常规算法预测结果

算法	训练集留一法测试正确率	测试集正确率
SVM	78.67%	50%
KNN(K=3)	64%	10%
DT	76%	50%

从结果中可以看出，SVM 和 DT 算法较 KNN 算法取得了较好的结果。从原理上来分析，说明不同标签的数据在一些特定的特征上有着较为明显的区别，这种区别易于被 SVM 算法和 DT 算法发现并用来作为区别指标。

还可以发现，三种算法对训练集的测试结果较好，但是对于测试集的测试结果较差。这说明可能是测试集的标准输出不准确，从实际上考虑，标准输出是教练和队长给出的，对于新队员他们也没有较多的了解，确实存在标准输出不准确的可能性。

为得到测试集的可信度更高的正确标签，选用 DT 算法对测试集输出的预测标签作为测试集的正确标签，进行后续的运筹优化分析处理。

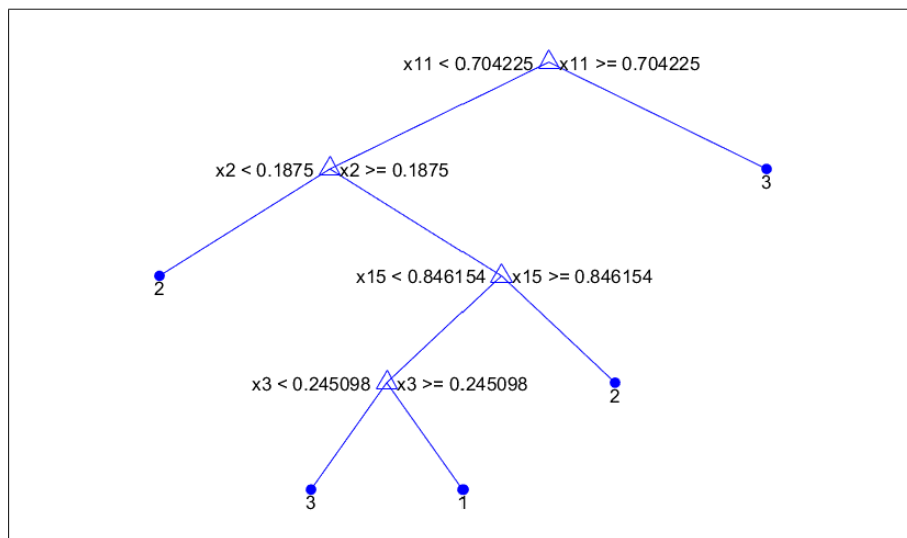


图 2-1 DT 算法最终得到的决策树

2.1.2 一致性预测框架结果

使用一致性预测框架结合 KNN 的 CP-KNN (K=3) 对数据进行预测，方法同上：对训练集使用留一法进行测试；对测试集进行预测，然后和标准输出对比获得正确率。

由于一致性预测需要解决多面手问题，所以需要输出多标签预测结果。从实际情况来看，多面手一般也是承受两个角色，所以综合考虑选择显著性水平 $\epsilon = 0.2$ ，使得算法对大部分数据都能输出两个预测标签同时也能保有较高的显著性水平。（显著性水平过高会导致算法只输出单预测，甚至输出空预测；显著性水平过低导致系统全部输出三标签预测，算法结果意义降低）

对于多预测结果的正确率计算，采用方法为：单预测结果等于正确标签算正确，多预测结果包含正确标签就算预测结果正确。算法结果如下所示：

表 2.1-2 CP-KNN (K=3) 预测结果

数据集	数据组数	单预测标签组数	双标签预测组数	三标签预测组数	正确率
训练集	25	10	10	5	84%
测试集	10	7	3	0	50%

从结果中可以看出，引入一致性预测框架后，算法可以输出多个预测标签，使得算法的准确率得到了上升。由于测试集标准输出不准确的原因导致测试集的准确率还是不高。

接下来随机展示几组数据在一致性预测框架下的输出，越靠前的预测结果说显著性水平越高：

表 2.1-3 CP-KNN (K=3) 对部分数据的输出

数据编号	正确标签	预测结果		
		预测结果 1	预测结果 2	预测结果 3
训练集 1	1	1	2	/
训练集 12	2	2	1	/
训练集 19	3	3	1	2
测试集 5	1	1	/	/
测试集 7	3	3	1	/

2.1.3 主成分分析法降维

从上述结果中可以看出，本次的数据明在一些特定的特征上有着较为明显的区别，所以适用于一些对个别特征敏感的机器学习算法。使用主成分分析法对数据进行降维，从而更容易看出数据的特征。

使用训练集进行降维分析：原始数据 16 维，使用 PCA 方法将其降至 3 维观察其空间分布，可以得到以下分布图像：

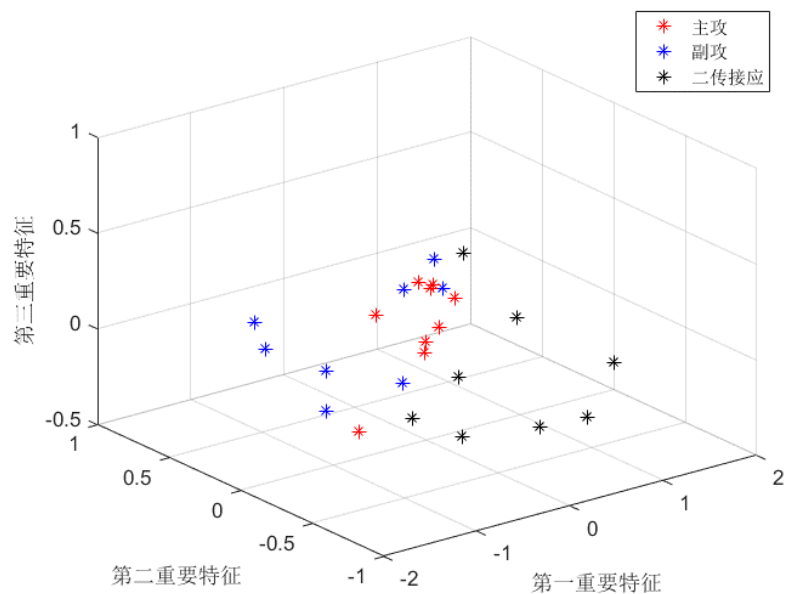


图 2-2 使用 PCA 降至 3 维后训练集数据分布

从图中可以看出二传接应在空间中的分布和主攻副攻有着较明显的区别，这是因为二传接应更要求的是二传能力、场上意识等；而主攻和副攻在主要功能上比较相似，在一些特征上有重合，所以较难区分。为了更好地区分主攻副攻，可能需要更高维数的特征值。

同时，还做出了维度重要性累计曲线，如下所示：

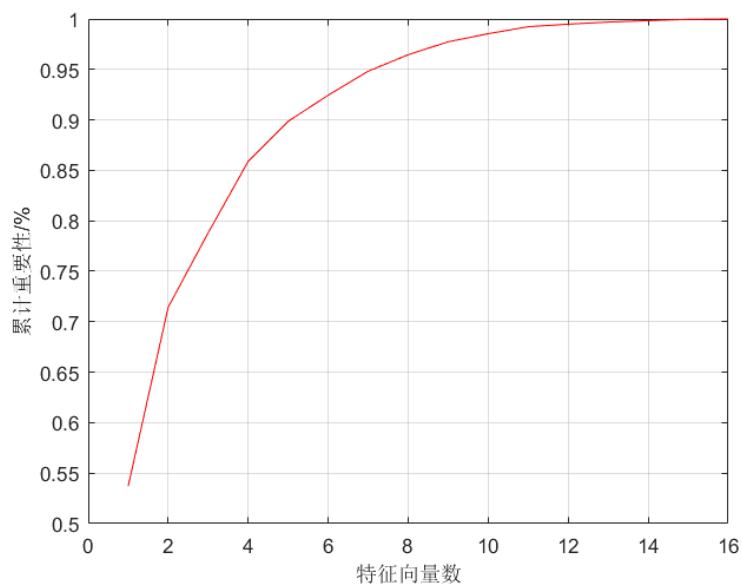


图 2-3 特征重要性累计趋势图

从图中可以看出，只使用三维数据只能体现原始数据 80%作用的方差，这会使得不同标记数据之间的区分度有着较大的下降，从而造成主攻和副攻在 3 维时难以区分的问题。

从特征重要性累计趋势图中还可以得出，为了达到较好的区分度，不过多的损失原始数据的方差，降至 7 维时应该还能较好的区分原始数据。

2.2 层次分析法结果

结合专家意见，可以得到如下的评价矩阵：

表 2.2-1 AHP 模型所使用的进攻能力评价指标

进攻能力	发球成功率	发球质量	扣球成功率	扣球质量评分	吊球成功率	吊球质量评分
发球成功率	1.00	1.00	0.33	0.33	3.00	3.00
发球质量	1.00	1.00	0.33	0.33	3.00	3.00
扣球成功率	3.00	3.00	1.00	1.00	5.00	5.00
扣球质量评分	3.00	3.00	1.00	1.00	5.00	5.00
吊球成功率	0.33	0.33	0.20	0.20	1.00	1.00
吊球质量评分	0.33	0.33	0.20	0.20	1.00	1.00

表 2.2-2 AHP 模型所使用的防守能力评价指标

防守能力	一传成功率	一传质量评分	防守成功率	防守质量评分
一传成功率	1.00	1.00	3.00	3.00
一传质量评分	1.00	1.00	3.00	3.00
防守成功率	0.33	0.33	1.00	1.00
防守质量评分	0.33	0.33	1.00	1.00

表 2.2-3 AHP 模型所使用的前排评价指标

前排反应	身高	拦网	场上意识评分
身高	1.00	0.67	2.00
拦网	1.50	1.00	3.00
场上意识评分	0.50	0.33	1.00

表 2.2-3 AHP 模型所使用的综合素质评价指标




综合素质	场上意识评分	心理素质评分	氛围带动能力评分
场上意识评分	1.00	1.00	1.50
心理素质评分	1.00	1.00	1.50
氛围带动能力评分	0.67	0.67	1.00

表 2.2-4 AHP 模型所使用的组织能力评价指标

组织能力	氛围带动能力评分	二传成功率	二传质量评分
氛围带动能力评分	1.00	0.33	0.33
二传成功率	3.00	1.00	1.00
二传质量评分	3.00	1.00	1.00

读取每个队员各个指标下的分数，和层次分析法的权重结合，得到每个队员每个主层次下的分数。对于每个队员，以分数的值为扇形半径，以分数占总分数的比为扇形角度，生成类似饼图的统计图。摘取 NO.1 NO.2 NO.3 的队员的结果如下表：

表 2.2-2 AHP 模型所使用的评价指标

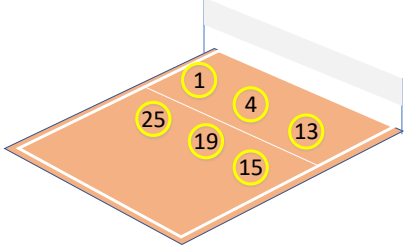
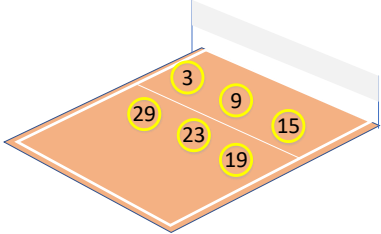
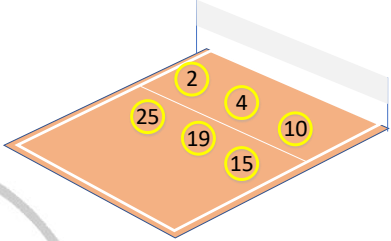
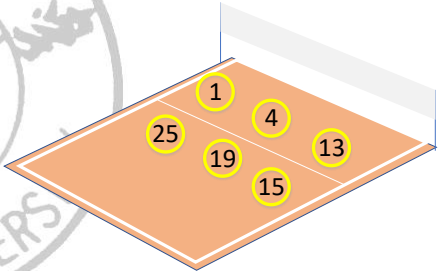
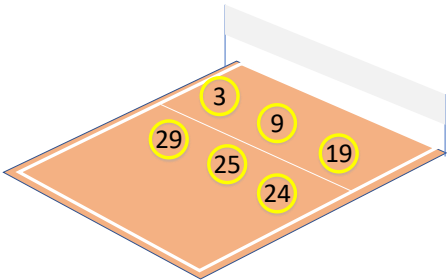
1	进攻能力得分	0.813	
	防守能力得分	0.758	
	前排反应得分	0.761	
	综合素质得分	0.636	
	组织能力得分	0.688	
2	进攻能力得分	0.666	
	防守能力得分	0.660	
	前排反应得分	0.860	
	综合素质得分	0.524	
	组织能力得分	0.622	
3	进攻能力得分	0.597	
	防守能力得分	0.961	
	前排反应得分	0.737	
	综合素质得分	0.711	
	组织能力得分	0.956	

2.3 运筹规划结果

2.3.1 Level 1 “群魔乱舞” ——全部混合规划结果

表 2.3-1 Level 1 全部混合规划模型规划结果

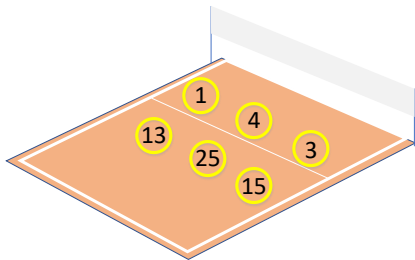
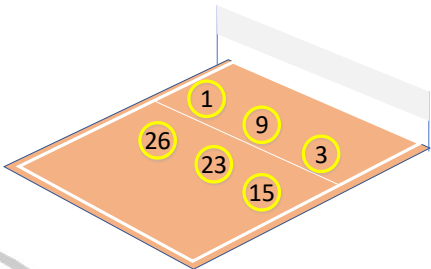
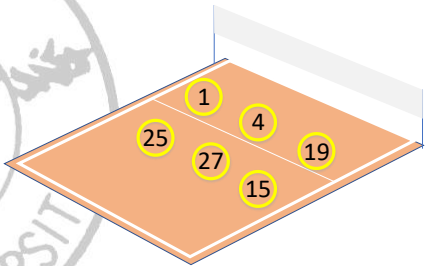
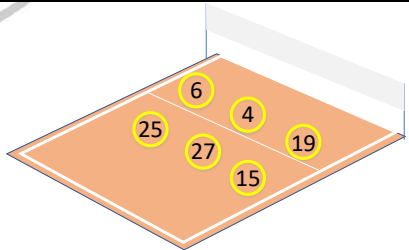
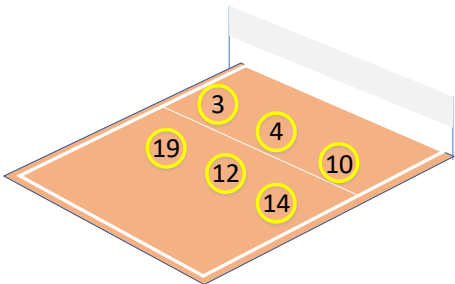
Target	Results	Plot Show
--------	---------	-----------

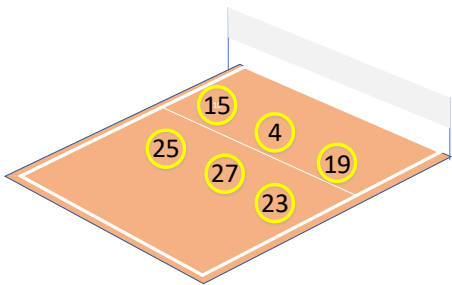
Max 进攻能力	NO. 1 & 4 & 13 & 15 & 19 & 25 进攻总能力为 4.9989	
Max 防守能力	NO. 3 & 9 & 15 & 19 & 23 & 29 防守总能力为 5.3806	
Max 前排能力	NO. 2 & 4 & 10 & 15 & 19 & 25 前排总能力为 5.2031	
Max 综合素质	NO. 1 & 4 & 13 & 15 & 19 & 25 综合总素质为 4.9467	
Max 组织能力	NO. 3 & 9 & 19 & 24 & 25 & 29 组织总能力为 5.5486	

2.3.2 Level 2 “各司其职” ——单标签规划以及与 AHP 的对比结果

基于文章 1.3 部分中的模型，我们可以获得如下表的规划结果。与此同时，在评估模型确定阶段我们也根据各位技术评审的意见为每一位队员给出了各个位置上的得分，如主攻能力评分、二传能力评分，我们也将该 AHP 方法给出的每个位置上的两个最优队员选出来作为 AHP 结果，与使用机器学习定标签和规划后的结果进行对比，具体结果如下：

表 2.3-2 Level 2 单标签规划以及与 AHP 模型规划结果

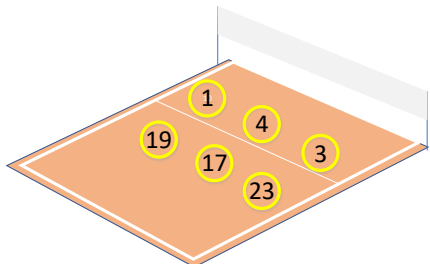
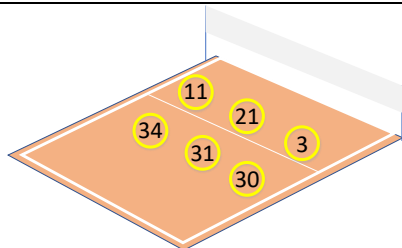
Target	Results	Plot Show
Max 进攻能力	<p>NO. 1 & 3 & 4 & 13 & 15 & 25</p> <p>进攻总能力为 3.37123</p>	
Max 防守能力	<p>NO. 1 & 3 & 9 & 15 & 23 & 26</p> <p>防守总能力为 5.1607</p>	
Max 前排能力	<p>NO. 1 & 4 & 15 & 19 & 25 & 27</p> <p>前排总能力为 4.96</p>	
Max 综合素质	<p>NO. 1 & 4 & 13 & 15 & 19 & 25</p> <p>综合总素质为 4.9467</p>	
Max 组织能力	<p>NO. 3 & 4 & 10 & 12 & 14 & 19</p> <p>组织总能力为 1.9492</p>	

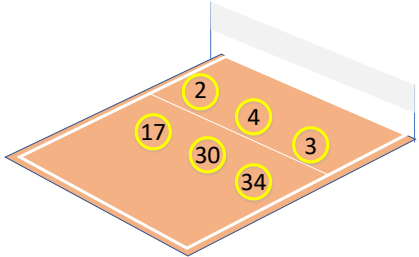
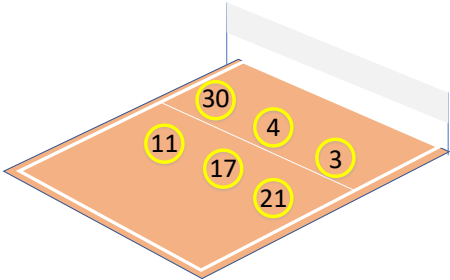
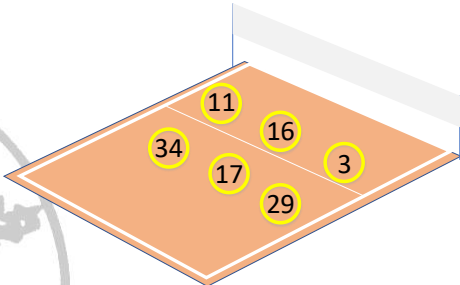
AHP	No. 4 & 15 & 19 & 25 & 27 & 21	
	进攻能力: 3.4051	
	防守能力: 4.3587	
	前排反应: 4.88	
	综合素质: 4.9333	
	组织能力: 1.1174	

由此可见，AHP 结果这种根据每个位置上“优中选优”的方法与我们进行机器学习定标签规划结果有一定的相似性，可以印证机器学习定标签规划的有效性。但是，由于 AHP 结果不考虑个人能力在团队中的贡献，即选择攻手时不很考虑其防守能力对于全队的影响，同时 AHP 方法中的评估权重也由人主观给出，因此该方法的评判与机器学习思路相比存在一定不足。

2.3.3 Level 3 “全面发展”——多面手问题规划结果

表 2-3.3 Level 3 多面手模型规划结果

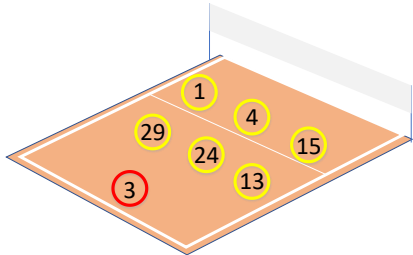
Target	Results	Plot Show
Max 进攻能力	NO. 1 & 3 & 4 & 17 & 19 & 21 进攻总能力为 3.4353	
Max 防守能力	NO. 3 & 11 & 21 & 30 & 31 & 34 防守总能力为 5.2941	

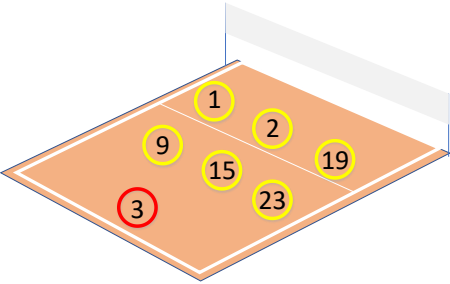
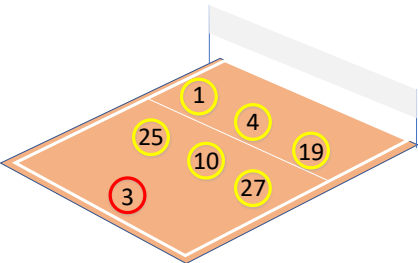
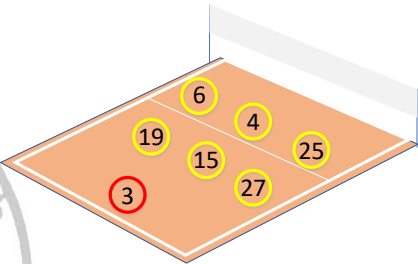
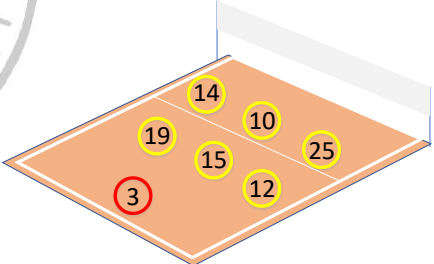
Max 前排能力	<p>NO. 2 & 3 & 4 & 17 & 30 & 34</p> <p>前排总能力为 4.96</p>	
Max 综合素质	<p>NO. 3 & 4 & 11 & 17 & 21 & 30</p> <p>综合总素质为 4.8889</p>	
Max 组织能力	<p>NO. 3 & 11 & 16 & 17 & 29 & 34</p> <p>组织总能力为 1.8483</p>	

2.3.4 Level 4 “固若金汤”——自由人规划结果

选出自由人两位：NO. 3 & 17

表 2-3.4 Level 4 加入自由人的模型的规划结果

Target	Results	Plot Show
Max 进攻能力	<p>NO. 1 & 3 & 4 & 13 & 15 & 24 & 29</p> <p>进攻总能力为 3.37128</p>	

Max 防守能力	NO. 1 & 2 & 3 & 9 & 15 & 19 & 23 防守总能力为 4.50189	
Max 前排能力	NO. 1 & 3 & 4 & 10 & 19 & 25 & 27 前排总能力为 4.97499	
Max 综合素质	NO. 3 & 4 & 6 & 15 & 19 & 25 & 27 综合总素质为 5.65777	
Max 组织能力	NO. 3 & 10 & 12 & 14 & 15 & 19 & 2 组织总能力为 2.8657	

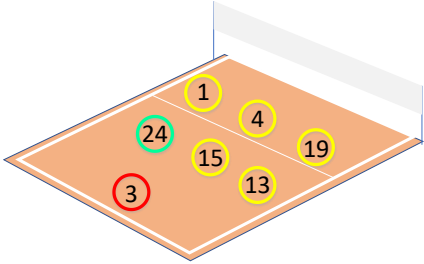
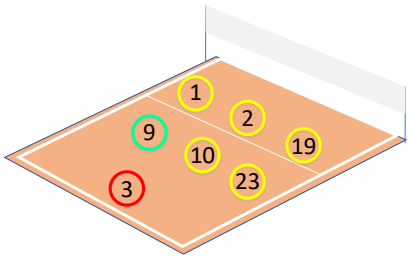
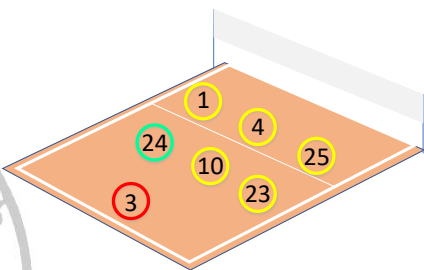
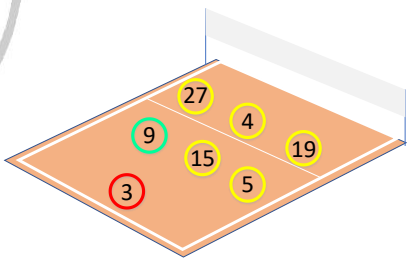
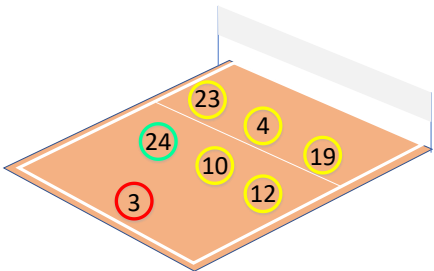
2.3.5 Level 5 “打到你服”——接应二传以及主攻、副攻、二传位置实际考虑规划结果

选出接应二传两位：NO. 9 & 24

选出自由人两位：NO. 3 & 17

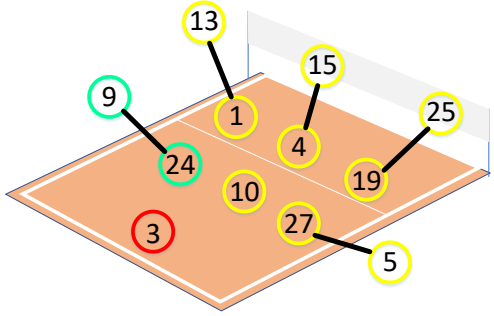
表 2-3.5 基于接应二传以及主攻、副攻、二传位置实际考虑的模型规划结果

Target	Results	Plot Show
--------	---------	-----------

Max 进攻能力	<p>NO. 1 & 3 & 4 & 13 & 15 & 19 & 24</p> <p>进攻总能力为 4.53448</p> 
Max 防守能力	<p>NO. 1 & 2 & 3 & 9 & 10 & 19 & 23</p> <p>防守总能力为 4.50189</p> 
Max 前排能力	<p>NO. 1 & 3 & 4 & 10 & 23 & 24 & 25</p> <p>前排总能力为 5.29028</p> 
Max 综合素质	<p>NO. 3 & 4 & 5 & 9 & 15 & 19 & 27</p> <p>综合总素质为 5.51111</p> 
Max 组织能力	<p>NO. 3 & 4 & 10 & 12 & 19 & 23 & 24</p> <p>组织总能力为 3.09512</p> 

2.3.6 Level 6 “精挑细选” —— “12 人大名单” 问题规划结果

表 2-3.6 Level 6 “12 人大名单” 模型规划结果

<p>主攻: No. 1 & 13 & 27 & 5</p> <p>副攻: No. 4 & 15 & 10</p> <p>二传: No. 19 & 25</p> <p>自由人: No. 3</p> <p>接应: No. 9 & 24</p>	
--	--

2.3.7 Level 7 “随机应变”——对策应用规划结果

$$H = \begin{bmatrix} 1.5751 & 37.3661 & 39.1868 & -2.9185 & 40.0783 \\ -34.7283 & 1.0627 & 2.8834 & -39.2218 & 3.7750 \\ -39.3884 & -2.5973 & -0.7766 & -42.8819 & 0.1149 \\ 5.9069 & 41.6979 & 43.5186 & 1.4133 & 44.4102 \\ -39.8774 & -4.0864 & -2.2657 & -44.3710 & -1.3741 \end{bmatrix} \quad (17)$$

在使用 MATLAB 模拟过程中, 我们得到的一个矩阵 H 如上所示。

根据矩阵对策的假设——每个局中人对对方的实力都了如指掌且都是理性的, 那么根据赢得矩阵 H 我们可以利用如下公式

$$\max_i \min_j H_{ij} = \min_j \max_i H_{ij} = H_{i^*j^*} \quad (18)$$

验证 H 是否存在平衡解、对于局中人是否存在最优纯策略。使用 matlab 编程可以得到

$$\max_i \min_j H_{ij} = 1.4133$$

$$\min_j \max_i H_{ij} = 1.4133$$

两者相等, 且在 $i=4, j=4$ 时取到 1.4133, 则对于局中人我方和对手存在最优纯策略, 我方的平衡策略是第 4 套方案(最大化综合素质得到的规划结果), 对手的平衡策略也是第 4 套方案。

3. 讨论

3.1 规划结果可靠性讨论

实践是检验真理的唯一标准。提供数据的浙江大学竺可桢学院男排在浙江大学排球比赛中, 真实的阵容如下所示:

利用运用机器学习和运筹学对阵容进行排布的结果有相同之处也有不同之处，如 Level 5 中规划得到的结果与上述阵容比较如下：

		1	3	4	13	15	2	9
19	24	1	3	4	13	15		

图 3.1-1 Level 5 最大化进攻能力规划结果与真实阵容比较

		1	2	3	9	4	13	15
10	19	23	1	2	3	9		

图 3.1-2 Level 5 最大化防守能力规划结果与真实阵容比较

		1	3	4	2	9	13	15
10	23	24	25	1	3	4		

图 3.1-3 Level 5 最大化前排能力规划结果与真实阵容比较

		3	4	9	15	1	2	13
5	19	27	3	4	9	15		

图 3.1-4 Level 5 最大化综合素质规划结果与真实阵容比较

		3	4	1	2	9	13	15
10	12	19	23	24	3	4		

图 3.1-5 Level 5 最大化组织能力规划结果与真实阵容比较

规划得到的结果与真实的阵容排布既有相同之处也有不同之处，其中最大化进攻能力与真实的阵容重合度最高，因此可以看出真实的阵容排布是偏向于增强团队进攻能力，从而可以初步断定该队伍在比赛中进攻取向较为明显。对于规划的结果是否真正的满足现实需要是应该通过实践来检验的，但规划结果对于阵容排布的参考价值依然是重大的。

3.2 模型优势与创新点

本团队鉴于目前国内大多中低水平排球队缺乏高水平教练的现实，并且学界关于排球

队优化策略和成员位置确定方面研究较少，创新地结合机器学习、数学建模和运筹学思想，提出了数据驱动的方式进行成员适合位置判定、综合能力分析以及队伍实力优化的模型和方法，并且以浙江大学竺可桢学院男子排球队和控制学院男子排球队为例佐证，证明模型的有效性；其次，本团队首次组合将各类机器学习分类算法、一致性预测算法和运筹学规划方法应用于运动队规划，并且能够在未来进一步扩展应用对象，完善应用场景，推广到更多球队和体育运动中；此外，本项目从设计到实现均为浙江大学本科生团队自主完成，属于兴趣驱动的探索研究，也将为学生课外学术研究起到一定的推动促进作用。

3.3 模型不足与展望

本项目目前还有诸多不足之处，首先，目前本项目仅以浙江大学竺可桢学院男排和控制学院男排为主要数据来源作为项目应用的尝试，样本容量较小，还不足以搭建样本较多、可靠性较高的训练集，未来本项目将会从两只院排球队扩展到更多排球队，收集更多样本数据完善机器学习、评估模型和运筹规划结果；其次，目前的各类指标的评判由评审团队制定，指标种类未来将有进一步丰富的空间，主观指标的评判客观性可以进一步提升。

4. 结论

本团队结合机器学习算法和运筹学规划算法，以浙江大学竺可桢学院男排和控制学院男排成员各项指标为数据，完成排球队的成员最适位置确定、能力评估图谱以及最优阵容排布。首先，决策树、KNN 和 SVM 和一致性预测四种算法给出根据团队内高年级运动员数据及位置给出每一位新队员最适位置的标签，其中在训练集中进行留一法交叉验证，支持向量机算法可以达到 78.67% 准确率，同时一致性预测算法可以给出运动员适合的多标签，分类准确率可以达到 84%，与此同时，PCA 算法的运用可以使不同位置，即不同标签的运动员的样本在特征空间中的分布得以可视化体现；随后，本团队结合专业评审团队意见，使用层次分析法完成运动员的能力评估，给出每位选手的进攻、防守、前排、综合和组织能力评估“雷达图”；最后，本团队结合 0/1 规划和对策论，设计并求解了由简到繁的七层规划问题，并且考虑了“多面手”问题、自由防守队员、接应二传、十二人大名单等复杂而实际的考量，给出不同条件下、不同目标下的最优阵容，通过与实际比赛阵容相比较，印证规划的可靠性。整个项目以数据和兴趣驱动，以数据挖掘为起点，结合实例创新地结合应用一致性预测和运筹学于运动队规划，有利于提升运动队的管理和能力优化。未来，本团队将结合更大量的数据来完善预测模型，同时根据实践结果和比赛效果来调整能力评估模型和运筹规划模型中的诸多参数，达到系统的优化目的。

参考文献

- [1] 倪婷婷.浅谈排球比赛阵容配备和进攻战术的运用[J].科技资讯,2010(07):222.
- [2] Vapnik V N. The nature of statistical learning theory[M]// The nature of statistical learning theory /. Springer, 1995:988-999.
- [3] T. Cover, P. Hart. Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 1967.13(1):21-27.
- [4] Breiman L I, Friedman J H, Olshen R A, et al. Classification and Regression Trees (CART)[J]. Encyclopedia of Ecology, 1984, 40(3):582-588.
- [5] V. Vapnik. The nature of statistical learning theory, in: Conference on Artificial Intelligence. 1995, pp:988-999.
- [6] A. Gammerman, V. Vovk, Hedging predictions in machine learning the second computer journal lecture, Computer Journal 10 (2) (2007) 151-163.
- [7] Nouredinov, D. Devetyarov, V. Vovk, B. Burford, S. Camuzeaux, A. Gentry-Maharaj, A. Tiss, C. Smith, Z. Luo, A. Chervonenkis, Multi-probabilistic prediction in early medical diagnoses, Annals of Mathematics and Artificial Intelligence 74 (1-2) (2015) 1-20.
- [8] V. Vovk, Conditional validity of inductive conformal predictors, Machine Learning 92 (2-3) (2013) 349-376.
- [9] Z. Wang, X. Sun, J. Miao, Y. Wang, Z. Luo, G. Li, Conformal prediction based on k-nearest neighbors for discrimination of ginsengs by a home-made electronic nose, Sensors 17 (8) (2017) 1869.
- [10] 王柯. 基于层次分析法的亚龙商贸物流园区选址研究[D]. 昆明理工大学, 2017.
- [11] 王孝宁, 何苗, 何钦成. 层次分析法判断矩阵的构成方法及比较[J]. 中国卫生统计, 2002, 19(2):111-113.
- [12] 韩中庚. 实用运筹学——模型、方法与计算[M]. 北京: 清华大学出版社, 2007

团队成员与分工



团队合影

团队展示照片



竺可桢学院男排成员合影



控制学院男排成员合影

