arXiv:2303.04963v1 [stat.AP] 9 Mar 2023

利用球员个人顺序统计量预测 NBA 精英阵容 Susan E. Martonosi[1], Martin Gonzalez[1],Nicolas Oshiro[2]1：哈维穆德学院 2：旧金山大学

摘要：NBA球队经理和老板会努力引进表现出色的球员。在这些决策中，一个重要的考量因素是新球员与队友搭档时的表现。我们的目标是找出精英五人阵容，并将其定义为每分钟正负值（PMM）为正的阵容。利用球员个人顺序统计量，我们的模型即使在阵容中的五名球员从未共同上场的情况下，也能识别精英阵容，从而为球员引进决策、薪资谈判和实时教练决策提供参考。我们将七种分类工具整合为一个一致同意分类器（全有或全无分类器，简称 ANC）：只有当七个分类器均预测某阵容为精英阵容时，该阵容才会被预测为精英阵容。通过这种方式，我们实现了较高的阳性预测值（即精确率），也就是被分类为精英阵容的阵容确实具有正PMM的可能性。我们使用2017-18赛季的球员个人数据和阵容数据训练并测试模型，并用该模型预测来自全部30支NBA球队2018-19常规赛名单的阵容表现。尽管 ANC 较为保守，会漏掉一些表现出色的阵容，但其精确率很高，并且推荐的位置配置均衡的阵容。**关键词：**篮球；分差；阵容；分类。

1 引言

职业篮球的策略不断演变，促使 NBA球队经理和教练持续寻找具备相关技能、能够赢得年度冠军的新球员。尽管球员个人表现是球队成功的重要因素，但我们更关注促成高水平球队以及构成这些球队的球员群体的总体因素。因此，我们利用球员顺序统计量，建立了一个用于预测高水平五人阵容的模型，其中包括此前从未共同出场过的阵容。该模型可以帮助考虑交易和自由球员的总经理评估潜在引进球员与现有球队成员搭配时可能产生的表现。教练也可以在比赛中使用该模型作为辅助工具，以判断坐在替补席上的哪名球员作为替补球员上场后可能产生积极影响。

文献中的大量研究主要关注预测篮球比赛、锦标赛¨或赛季结果的方法。Ozmen 量化了若干比赛统计数据每单位变化所带来的获胜概率增量变化 [23]。Loeffelholz 等使用神经网络预测 NBA比赛，其效果优于专家预测 [17]。Lin 利用球队历史数据，并在比赛期间进行实时更新，以预测比赛结果 [16]。Shen 等和 Hua 提出了预测 NCAA“疯狂三月”大学锦标赛结果的方法 [28, 13]„ 2015 年本刊特刊中的数篇文章也采用了类似方法 [11]。Gumm 等和 Zimmerman 等利用历史数据上的机器学习方法，预测 NCAA 和 NBA季后赛的结果 [12, 35]。Vaz de Melo 等利用网络分析预测 NBA比赛结果 [30], , 而 Cheng 等则运用最大熵原理预测 NBA季后赛结果 [7]。Ruiz 和 Cruz 提出了一个模型，通过将泊松因子分解方法与借鉴自足球的模型相结合，预测参加 NCAA 疯狂三月锦标赛的球队的获胜概率；该足球模型用于量化球队的进攻与防守系数 [27]。

已有大量研究探讨个别球员在决定球队成功方面所起的作用。多篇论文采用复杂社会网络领域的中心性指标及其他方法，识别关键球员并评估球队表现 [8, 25, 31]。Deshpande 和 Jensen 对个别球员为球队整体获胜概率所作的贡献进行量化 [9]。然而，也有研究指出，以真实正负值衡量的个人对球队成功的贡献，无法与其他球员的影响区分开来 [10]。虽然 Box Plus Minus（BPM）等统计指标试图衡量个别球员相对于队友对球队的贡献，但 BPM 并不能说明五人阵容协同作战时的表现。

我们的论文建立在一系列研究的基础上，这些研究探讨了 NBA 五人阵容的效能。例如，Maymin等人通过将各支 NBA 球队的五名球员阵容的效能与“部分之和” [18]进行比较，计算其协同效应。Robertson 使用图模型捕捉一次控球过程中发生的一系列事件，并根据场上球员估计比赛过程各类事件发生的概率 [26]。Kalman 和 Bosch 使用基于模型的聚类重新定义传统的篮球位置，并识别这些新位置之间的互动如何

形成净效率值最高的阵容 [14]。Pelechrinis 提出了 LinNet，该方法利用有向网络的动态特性，捕捉阵容对阵特定对手阵容时的表现 [24]。此外，Sisneros 和 Van Moer 使用分差（正负值）衡量个别球员对球队成功的贡献，并预测给定球队的胜率 [29]。Oh等人使用概率网络模型模拟比赛过程；不过，他们的研究有一项局限，即无法预测由球队之外的球员组成的阵容的表现 [22]。

我们的研究通过开发一种模型，为这一总体讨论作出了贡献。该模型能够预测未见过的阵容的表现，衡量指标是其每分钟得分差的预测符号。我们了解到的大多数研究都着眼于指导比赛中的换人等教练决策，而我们的研究更进一步，可用于自由球员市场决策、交易谈判以及从 NBA G 联盟召回球员。

我们的模型会识别出每分钟贡献正分差（也称正负值）的概率较高的阵容。正PMM 表明该阵容通常会对球队的相对得分作出正向贡献，而负PMM 则表明该阵容通常会拉低球队的相对得分。我们的目标是预测精英阵容，我们将其定义为很可能贡献正PMM 的阵容。我们将七种分类工具结合为一个一致同意分类器，称为全有或全无分类器 （ANC）。只有当所有七个子分类器都预测某阵容为精英阵容时，ANC 才会将其判定为精英阵容。通过这种方式，我们实现了较高的精确率，即被判定为精英阵容的阵容确实具有较高 PMM 的可能性。每种工具都以拟议阵容中五名球员各自的球员顺序统计量为输入，以此反映阵容的进攻和防守能力，并将阵容分类为精英或非精英。所用的子分类器包括决策树、随机森林、提升法、支持向量机、k近邻、逻辑回归和线性判别分析。

我们使用从 BigDataBall.com 购买的 2017-18赛季 NBA 常规赛球员个人数据，并结合来自 NBA.com 的拼抢数据，训练和测试模型。之所以纳入拼抢数据，是因为它们量化了球员对阵容的防守贡献。训练模型时，我们随机抽取了 712 个上场时间至少为 25 分钟的阵容。随后，我们用随机抽取的 176 个上场时间至少为 25 分钟的阵容测试模型。在测试集上，我们的精确率达到 86.7%，这表明 ANC 判定为精英阵容的阵容极有可能为球队带来正PMM。之后，我们使用完整的 2017-18赛季 NBA 常规赛数据训练模型，并用训练好的模型预测 2018-19常规赛期间全部 30 支 NBA球队的精英阵容。结果发现，即使模型用于预测下一个赛季的情况，分类器仍能达到较高的 76.9% 精确率（精英阵容的占比为 62.1%）。

本文结构如下。第 2 节概述我们在分类框架中采用的方法。第 3 节介绍用于训练和测试模型的数据集，并概述清理和合并数据所采取的步骤。第 4 节展示将模型应用于测试数据集的结果，以及对 2018-19常规赛的预测结果，并将预测结果与赛季的实际结果进行比较。第 5 节提出未来工作的建议并给出结论。

2 方法

本节将介绍我们的方法所要解决的分类问题、为解决该问题而构建的全有或全无分类器 (ANC)，以及我们用于调整超参数的交叉验证方法，以获得稳定且较高的正向预测值，即精确率。

2.1 对未知阵容的每分钟正负值进行分类

五人阵容的表现通常用其每分钟正负值（PMM）来概括。该指标衡量阵容每次共同上场时累计取得的分差（得分减去对手得分），再除以阵容共同上场的总时间。PMM为正的阵容有助于球队提高单位时间内的净得分，而PMM为负的阵容单位时间内失分多于得分。现有用于预测未观察到的阵容表现的方法，例如调整正负值，试图在控制场上其他球员的影响后，估算某位球员对得分的个人贡献

表 1：阵容 1 [1] 的个人球员统计数据示例。

球员	杰伦 布朗	凯里 欧文	马库斯 莫里斯	特里 罗齐尔	杰森 塔图姆
每分钟投篮命中数	0.17	0.28	0.18	0.15	0.16
每分钟防守盖帽数	0.01	0.01	0.01	0.01	0.02

表2：阵容 2 [2] 的个人球员统计数据示例。

球员	詹姆斯 哈登	克里斯 Paul	Eric 戈登	Nene 希拉里奥	克林特 卡佩拉
每分钟投篮命中数	0.26	0.20	0.19	0.18	0.22
每分钟防守盖帽数	0.02	0.01	0.01	0.02	0.07

在场上 [34]。然而，这些方法仅通过拟合观测到的正负值与球员是否在场的指标来进行；它们并未纳入有关球员个人表现的其他统计数据。

我们采用的方法是利用个人球员统计数据来预测未知阵容的 PMM。如需了解篮球统计数据的入门知识，请参阅 Kubatko 等 [15]。设 L 为所有五人阵容的集合， P 为球员集合。对于每个阵容 $l \in L$，设 $p(l) \in P$ 为阵容 l中五名球员组成的集合。对于每名球员 $p \in P$，设 s_{p1}、.....、 s_{pn} 为按每分钟统计的各项数据，例如每分钟投篮命中数或每分钟防守篮板数。因此，我们希望将阵容 l的每分钟正负值记为 $PMM(l)$，并根据每名球员 $p \in p(l)$的 s_{p1}、.....、 s_{pn} 值来预测：

$$PMM(l) = f(s_{pi}|p \in p(l), i \in 1, \ldots, n).$$

先前的研究表明，直接预测 $PMM(l)$颇具挑战性 [6]。因此，在本研究中，我们着重预测 $PMM(l)$的正负，将阵容 l 分类为精英阵容，当且仅当 $PMM(l) > 0$。

我们假设 PMM 与球员统计数据之间存在分布关系。对于每套阵容和每项统计数据，我们将该阵容五名球员的统计值从小到大排序。然后，我们在模型中将这些顺序统计量用作预测变量。实际上，这是使用该阵容五名球员各项统计数据分布的 $sign$（PMM）20^{th}、40^{th}、60^{th}、80^{th}和 100^{th} 分位数来预测 PMM。举例来说，一名球员可能对投篮命中率贡献很大，而另一名球员则可能在盖帽方面贡献突出，而所有这些因素都会影响该阵容的整体 PMM。

举一个简单的例子，考虑两项球员统计数据：每分钟投篮命中数（FGM）和每分钟防守盖帽数（DBM）。以阵容 1（波士顿凯尔特人队的杰伦·布朗、凯里·欧文、马库斯·莫里斯、特里·罗齐尔和杰森·塔图姆）和阵容2（休斯敦火箭队的詹姆斯·哈登、克里斯·保罗、埃里克·戈登、内内·希拉里奥和克林特·卡佩拉）为例，考察2017-18赛季常规赛期间的情况。表 1 和 2 [1, 2]列出了每位球员这两项统计数据的数值。

对于每个阵容，我们将五个 FGM 数值按从小到大的顺序排列，并将五个 DBM 数值也按从小到大的顺序排列，从而得到十个预测变量 $FGM_{(1)}, FGM_{(2)}, \ldots, FGM_{(5)}, DBM_{(1)}, DBM_{(2)}, \ldots, DBM_{(5)}$，其中下标（$i$）表示五名球员中第 i^{th} 小的数值（称为 i^{th} 顺序统计量）。对于给出的两个阵容，这样便得到表3所示数据集中的两行。因此，对于所收集的每项球员表现统计指标，我们的预测变量集合就是阵容中五名球员的顺序统计量。通过同时使用进攻和防守指标的顺序统计量，我们可以捕捉阵容的整体比赛风格。

2.2 分类器的选择

为了使我们的分类器成为人员决策的有效决策支持工具，我们希望确信被我们的分类器标记为精英的阵容确实表现出色。大多数分类框架都力求最大化分类器的总体准确率，即模型给出的分类与该数据点真实类别相符的概率。然而，在识别精英篮球阵容时，我们更关注较高的阳性预测值，也称为精确率。这是指在分类器预测某个阵容为精英阵容的条件下，该阵容实际上属于精英阵容（具有正PMM）的条件概率。

Table 3: Example player order statistics for the two lineups given in Tables 1 and 2. These order statistics serve as ten predictors in the classification framework.

Lineup	$FGM_{(1)}$	$FGM_{(2)}$	$FGM_{(3)}$	$FGM_{(4)}$	$FGM_{(5)}$	$DBM_{(1)}$	$DBM_{(2)}$	$DBM_{(3)}$	$DBM_{(4)}$	$DBM_{(5)}$
1	0.15	0.16	0.17	0.18	0.28	0.01	0.01	0.01	0.01	0.02
2	0.18	0.19	0.20	0.22	0.26	0.01	0.01	0.02	0.02	0.07

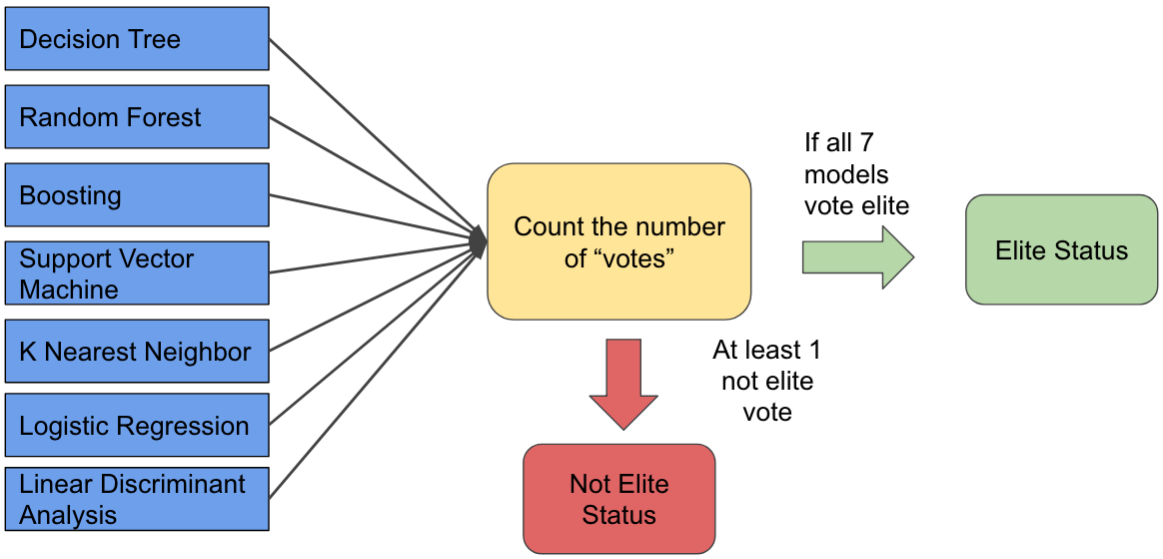


图1：全有或全无分类器用于对阵容进行分类的决策过程。

为提高分类器的精确率，我们将七种常用分类器（下文称为子分类器）组合成一个全有或全无分类器（ANC）：当且仅当全部七个子分类器均预测某个阵容为精英阵容时，该阵容才会被分类为精英阵容。共识分类器已成功应用于其他领域，包括计算生物学 [5]。这七个子分类器分别是决策树、随机森林、提升法、支持向量机、 k近邻、逻辑回归和线性判别分析。这些算法分别根据给定阵容中各球员的球员顺序统计量，对该阵容的分类结果投票。图1展示了ANC的示意图。

2.3 参数调优

这些子分类方法各自都有一组控制算法的参数。我们不单独调优每个子分类器，而是将这些参数作为一个集成进行调优，以使整个ANC分类器稳定地达到较高的精确率。我们在阵容训练集上使用10折交叉验证。（第3节将更详细地介绍这些数据。）下面简要介绍与各子分类器及整个ANC分类器相关的参数。随后，我们将提供概述集成调优流程的伪代码。

2.3.1 决策树参数

决策树会迭代地确定预测变量的阈值，并据此划分阵容是精英还是非精英。通常的做法是先构建一棵完整树，再根据成本复杂度权衡对其进行剪枝。R 的 rpart 软件包中用于构建和剪枝决策树的成本复杂度参数 cp，表示每次后续分裂所需达到的分支间变异性降低百分比。我们使用 $cp = -1$ 构建初始完整树，然后调整用于剪枝的 cp 值。此外，我们使用 weights参数，按照每个阵容的总上场分钟数为观测值赋权，从而降低总上场时间较少的阵容对树拟合的影响。为优先实现较高的阳性预测值，我们调整 parms 中损失矩阵选项的假阳性项，对将非精英阵容误判为精英阵容施加更重的惩罚。所有其他参数均采用 R 中 rpart函数给出的默认值。

2.3.2 随机森林参数

在随机森林分类器中，会基于训练数据的自助抽样样本构建许多决策树，并从全部预测变量中随机选取一个子集，供每次分裂决策考虑。每棵树产生的分类结果会作为投票，用于对单个观测值进行随机森林分类。二分类器中控制投票的参数是截断向量（$c,\ 1-c$）。如果观察到有 p 比例的树将某个阵容分类为精英，另有 $1-p$ 比例的树将该阵容分类为非精英，那么随机森林分类结果将是（$p/c,\ (1-p)/(1-c)$）中取值最大的类别。换言之，c 表示这样一个树的比例：当达到该比例时，随机森林分类器对两个组别将不置可否。由于我们力求最大化 ANC 的精确率，因此使用 $c \geq 0.5$ 的取值来调整随机森林。我们还调整控制森林中树的数量的参数 ntree。所有其他参数均采用 R 的 randomForest 包中 randomForest 函数给出的默认值，包括参数 mtry；该参数控制每个子集中包含的预测变量数量，其默认值为预测变量数量的平方根。

2.3.3 提升算法参数

提升算法会迭代构建分类树，使后续分裂由针对先前分裂的残差（错误分类）拟合的决策树决定。我们在分类器中调整了三个参数：要构建的树的数量（在 R 的 adabag 包 boosting 函数中称为 mfinal）、每棵树的深度（maxdepth），以及用于剪枝每棵树的代价复杂度值（cp）。此外，我们按照每个阵容的总上场分钟数对观测值加权，以降低上场时间较短的阵容所占的权重。其他所有参数均采用 R 的 adabag 包 boosting 函数给出的默认值。

2.3.4 支持向量机参数

我们的支持向量机（SVM）子分类器使用径向基核，并调整两个参数：代价和伽马。代价表示与错误分类相关的惩罚；伽马控制单个支持向量对一定距离外各点分类的影响衰减速率。较小的伽马会降低衰减速率，意味着该支持向量能够影响更远处点的分类；较大的伽马则产生相反效果。其他可调参数均采用 R 的 e1071 包中 svm 函数使用的默认值。

2.3.5 K近邻参数

k近邻子分类器根据预测变量空间中距离该阵容最近的 k 个阵容中最常见的标签，为阵容赋予标签。因此，我们使用 class 包中的 knn 函数来调整 k 的取值。

2.3.6 逻辑回归参数

逻辑回归将阵容属于精英阵容的对数优势 LO 预测为预测变量的线性函数。我们使用 R 包 stats 中的 glm 函数拟合模型。然后，通过计算精英分类的估计概率（$1+e^{-LO}$）$^{-1}$ 来确定阵容的分类；如果阵容属于精英阵容的估计概率超过指定阈值，则将其判定为精英阵容。自然的默认阈值为 50%；然而，鉴于我们希望最大化精确率，我们将阈值调整为大于或等于 50% 的取值，实质上要求有更强的证据才将阵容分类为精英阵容。glm 中的参数 thresh 指的是 1 减去所需的概率阈值。

2.3.7 线性判别分析参数

R 的 MASS 程序包中 lda 函数所实现的线性判别分析不依赖调优参数。不过，它允许我们根据每个阵容上场的总分钟数对观测值加权。

2.3.8 ANC 参数

整体 ANC 分类器的可调参数是：ANC 将阵容分类为精英阵容所需的、将其识别为精英阵容的子分类器数量。我们将其称为 numVotes。虽然我们假设要求七个分类器全部达成一致会获得最高精确率（正如“全有或全无分类器”这个名称所暗示的那样），但我们会通过调整将阵容分类为精英阵容所需的投票数来验证这一假设。

2.3.9 集成调优流程

我们不打算逐一调优每个子分类器，而是希望从所有子分类器的参数组合中选出最佳组合，使 ANC 在 10折交叉验证中获得稳定的高精确率。我们针对全部七个子分类器的参数组合以及 numVotes 参数执行网格搜索。对于每种参数组合，我们计算每一折上的 ANC 分类结果，并计算该折所达到的精确率。然后，我们对每种参数组合在各折上的精确率求平均值，从而得到衡量 ANC 在每种调优参数组合下预期精确率的良好指标。不过，除了最大化平均精确率之外，我们还希望分类器表现稳健。也就是说，在精确率相对较高的参数组合中，我们会优先选择精确率在各 k 折之间的标准差较低的组合（即稳健分类器），并且任一折上的精确率都不会太低的组合（即可靠分类器）。随后，我们可以考察所有参数组合的这些指标，并选择能带来足够高的平均精确率、较高的最低精确率和较低标准差的参数值。算法1概述了该流程的伪代码。

算法1ANC 参数调优的伪代码。

```
将数据划分为训练集和测试集
将训练集划分为 10 折，用于交叉验证
对每一折 k 执行
    对每个数值变量，减去均值并除以标准差，以此标准化折内训练数据
    在 k−1 折上计算得到的数值范围内。使用折内训练数据的均值和标准差对留出折进行缩放和平移
    以防止数据泄漏。
    对于每个子分类器 s 执行
        对于每种调优参数组合 c 执行
            用不在 k中的 9 折拟合一个模型。
            预测 k^th 折中观测值的类别。
            将这些预测结果存储为 preds [k,s,c]。
        结束循环
    结束循环
    对于 调优参数的每种组合 c （包括投票数）执行
        统计满足 preds [k,s,c] 为精英的子分类器数量 s。
        如果该数量至少达到以下数值，则该 k^th 折中观测值的 ANC 分类为精英
        投票数      ，否则为非精英。
        计算观测值的 ANC 分类与已知分类之间的混淆矩阵
        在 k^th 折上，并存储精确率 [k,c]。
    结束循环
结束循环
对于 每种调优参数组合（包括投票数）执行
    计算 10 折中的精确率平均值、最小值和标准差。
结束循环
选择能够稳定获得较高精确率的参数组合（平均值较高、最小值较高，
标准差较低）。
```

由于我们同时调整七种分类器和 ANC 的所有可能参数组合，因此随着每个参数所测试的不同取值数量增加，搜索空间会呈指数级增长。因此，我们采用粗略的

表4：网格搜索中使用的参数值。

子分类器	参数	测试值
决策树	cp（成本复杂度）	（-1, 0.01, 0.05）
	loss（误分类惩罚）	−1 表示完整树 (1, 1.5, 2)
随机森林	c（截止值）	(0.5, 0.7)
	ntree（树的数量）	(100,500)
提升法	mfinal（树的数量）	(100,500)
	maxdepth（每棵树的深度）	（1，2，3）
	cp（成本复杂度）	（0.01，0.05）
支持向量机	代价（误分类惩罚）	(0.1, 1, 10)
	伽马（影响衰减）	(0.01, 0.1, 1)
K近邻	k（邻居数量）	(3, 5,7)
逻辑回归	thresh（1-概率阈值）	(0.05, 0.25, 0.5)
全有或全无分类器（ANC）	投票数（所需一致意见数）	（1、2、3、4、5、6、7）

网格，为每个参数选取少量不同取值；这些取值根据未在此处报告的初步测试结果确定。所测试的具体取值见表4。

模型框架搭建完成后，接下来介绍所用数据。

3 数据与实现

本研究的新颖之处在于，利用组成阵容的球员经排序的个人统计数据来预测阵容表现。为此，需要合并来自两个来源的数据，清理数据以确保阵容与球员正确匹配，根据上场分钟数筛选数据以减少噪声，并将数据划分为训练集和测试集。本文将在此介绍这一流程。

3.1 数据来源

我们希望对 2018-19赛季 NBA 常规赛进行预测，并将预测结果与实际结果进行比较；这是最近一个未受新冠疫情影响的赛季。我们使用多个来源的数据，以 2017-18 NBA 常规赛的球员和阵容统计数据训练和测试模型。我们从 BigDataBall.com 购买了逐回合数据，其中记录了该赛季每场比赛每个回合的场上十名球员、发生的回合以及比赛时间。借助这一数据来源，我们能够重建每个阵容的每分钟得分差（PMM）。为求简便并提高准确率，我们没有使用这些逐回合数据计算个人球员统计数据，而是直接从 NBA.com 获取球员的基础统计数据和拼抢数据。

用于构建我们的140个顺序统计量预测变量的28项个人球员统计数据列于附录A。我们着重关注能够直接衡量比赛表现的统计数据，而非正负值（BPM）这类统计数据，因为后者本身是多项统计数据的汇总，因此较难解读。

3.2 数据清理与筛选

由于我们要合并来自两个来源的数据——NBA.com 和 BigDataBall.com——因此必须先清理两个数据集，统一球员姓名的写法，例如标点、昵称或后缀的使用。随后，我们将个人球员统计数据与阵容进行匹配，并按从小到大的顺序排列阵容中球员的各项统计数据；由此得到顺序统计量向量，并将其用作预测变量。

匹配球员和阵容后，应剔除仅出现在两个数据集之一中的球员（以及他们所在的阵容）。只有一名球员因此被剔除：泰·劳森，他的

季后赛数据出现在 theNBA.com 的数据中，但 BigDataBall.com 的上场时间段数据中没有常规赛信息。因此，包含劳森的阵容不会用于训练或测试。

接下来，我们剔除赛季上场时间少于50分钟的个别球员（以及他们所属的阵容）。这样做是为了确保按每分钟调整后的个人球员统计数据估计值具有较低的方差。在原始数据中出现的540名球员里，有483名达到这一上场时间阈值。同样，我们也剔除由这483名球员组成、共同上场时间少于25分钟的阵容。在2017-18赛季中出现的14,000多个不同NBA阵容里，若要求每名球员的上场时间至少为50分钟，则有888个阵容的阵容上场时间至少为25分钟。值得注意的是，一些先前研究建议为阵容设定更高的上场时间阈值，以提高模型准确率。然而，这样做会大幅缩减我们的数据集。例如，若要求阵容上场时间达到50分钟，则只有374个阵容可用于训练和测试。鉴于作为预测变量的顺序统计量数量庞大，这样做可能导致模型过拟合。因此，为了减轻上场时间较短的阵容所带来的影响，我们在线性判别分析、回归树和提升分类器中使用阵容上场时间作为权重。

3.3 数据划分

在这888个阵容中，我们按80%-20%的比例进行划分，得到包含712个阵容的训练集和包含176个阵容的测试集。在两个数据集中，若某个阵容的 PMM 严格大于零，则标记为精英；否则标记为非精英。训练集中有375个阵容标记为精英，337个标记为非精英；测试集中有95个阵容标记为精英，81个标记为非精英。

4 结果

现在，我们将介绍 ANC 参数调优的结果，以及将训练好的 ANC 应用于测试数据的结果。在确认训练好的 ANC 具有令人满意的预测能力后，我们会解读调优后的子分类器，以了解哪些预测变量对于预测阵容质量最为重要。随后，我们使用 ANC 对 2018-19赛季NBA 球队名单中的阵容组合进行预测。

4.1 参数调优结果

参数调优的常见做法是，选择在十个交叉验证折上的平均指标表现最佳的参数组合。在我们的案例中，这意味着选择平均精确率最高的参数组合。然而，在对参数调优结果进行初步分析时，我们发现，平均精确率最高的参数组合往往在十折之间的精确率波动较大。这意味着分类器偶尔会在某些数据集上表现不佳，而且这种情况难以预测。图2(a)展示了调优过程中使用的每种参数组合在十折上的平均精确率与最坏情况（最小值）精确率。我们排除了至少有一折的精确率计算结果为 NA 的参数组合。进一步调查发现，这些组合仅在 $\text{numVotes} = 7$和逻辑回归参数 $\text{thresh} = 0.05$ 时出现。在这种情况下，ANC 的判定非常保守，以至于某些折中没有任何阵容被分类为精英，因此在计算精确率时得到 NA。此外，那些能够计算出精确率数值的折中，被分类为精英的阵容数量非常少，导致精确率的标准差很高。我们更倾向于选择一种参数组合，使其平均精确率足够高，且十折中的最坏情况精确率也较高。具体而言，我们选择位于平均精确率和最坏情况精确率有效前沿上的参数组合，如图2(a)中的红色所示。此外，尽管我们更关注精确率而非总体准确率，图2(b)仍展示了所有参数组合在十折上的平均精确率与平均准确率，并以红色标出平均精确率和平均准确率的有效前沿。

图2(a)中的蓝色点对应的参数组合，其平均精确率为79.8%，最坏情况精确率为50%。图2(b)在平均精确率与平均准确率的图中以蓝色标出了这一点；可以看到，这些参数组合也位于精确率和准确率的有效前沿上，平均准确率达到53.2%。我们从中选择优化后的参数组合来训练最终模型，其中

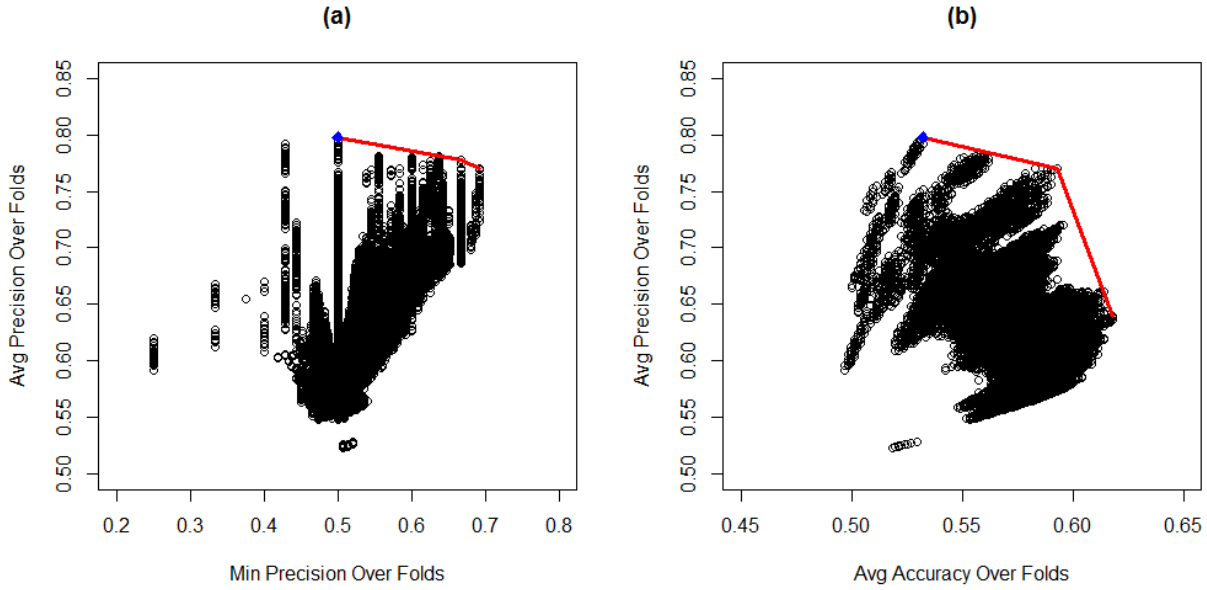


图 2：参数调优结果。对于每种参数组合，我们绘制：(a) ANC 在十折交叉验证中达到的平均精确率与该分类器在十折中的最低精确率；以及 (b) 平均精确率与平均准确率。在两张图中，有效前沿以红色表示，而我们通过优化后的参数组合所达到的最低精确率与平均精确率组合则以蓝色菱形标出。

该集合如表5所示。k（K近邻）、cp（决策树）、loss（决策树）、ntree（随机森林）、c（随机森林）、thresh（逻辑回归）和numVotes（ANC）的值均由此唯一确定。然而，参数调优对mfinal（提升法）、maxdepth（提升法）、cp（提升法）、cost（SVM）或伽马（SVM）的取值并不敏感。对于这些参数，我们选择在图2所示的有效前沿中取得次优结果的取值，对应的平均精确率为77.0%，最低精确率为66.7%，平均准确率为59.3%。

我们注意到，分类器的性能似乎对七个子分类器中有多少个投票赞成将某个阵容预测为精英最为敏感。当要求少于七个子分类器中的六个达成一致时，平均精确率便开始显著下降；相比六个，七个更为理想。此外，ANC 的性能似乎还对决策树的代价复杂度参数、逻辑回归的概率阈值以及随机森林的截断参数的选择较为敏感。ANC 的精确率对其余参数的选择似乎较为稳健。

4.2 测试结果

我们使用完整的标准化训练集拟合 ANC。为避免数据泄漏，我们使用训练数据的均值和标准差对测试数据进行平移和缩放，然后将训练好的 ANC 应用于测试数据。测试数据包含 176 个阵容，其中 95 个被标记为精英。混淆矩阵见表6。在预测为精英的 15 个阵容中，有 13 个的真实标签为精英，表明其 PMM 严格为正。对应的精确率为 86.7%。模型的总体准确率为 52.3%；鉴于我们的重点是预测精英阵容，这一结果可以接受。

此外，我们可以比较已知的 PMM：预测为精英的阵容与非精英阵容之间的 PMM。图3对预测为精英或非精英的阵容每分钟平均分差进行了箱线图比较。预测为精英的阵容的平均 PMM 为 $+0.30$，在统计上显著高于预测为非精英的阵容的平均 PMM -0.00049（$p = 0.00013$）。因此，虽然 ANC 并不直接预测 PMM，但它确实能很好地

表5：最终 ANC 中使用的调优参数值。

子分类器	参数	选定值
决策树	cp（成本复杂度）	0.05
	loss（误分类惩罚）	1
随机森林	c（截止值）	0.7
	ntree（树的数量）	500
提升法	mfinal（树的数量）	500
	maxdepth（每棵树的深度）	3
	cp（成本复杂度）	0.01
支持向量机	成本（误分类惩罚）	1
	伽马（影响衰减）	1
K近邻	k（邻居数量）	7
逻辑回归	阈值（1–概率阈值）	0.25
全或无分类器（ANC）投票数（所需一致意见数）		7

表6：调优后的 ANC 应用于测试数据集所得的混淆矩阵。在预测为精英的 15 个阵容中，13 个的真实标签为精英，对应的精确率为 86.7%。

	预测类别 精英	非精英
True Class		
精英	13	82
非精英	2	79

将球队划分为精英组和非精英组，预测前者具有正PMM，后者具有负PMM。值得注意的是，这些分布存在重叠，而且ANC会漏掉一些非常出色的阵容（例如多伦多猛龙的Delon Wright、德马尔·德罗赞、弗雷德·范弗利特、雅各布·珀尔特尔和帕斯卡尔·西亚卡姆组成的阵容，该阵容上场时间为40.0分钟，PMM为1.00）。尽管如此，精英阵容的PMM分布总体上高于非精英阵容的PMM分布，而被ANC误判为精英阵容的两个负PMM阵容，其PMM也只是中度偏低。这表明ANC对精英阵容具有良好的判断能力。

人们可能还会想知道，ANC 是否需要完整的五人次序统计量集合才能实现较高的精确率。附录B介绍了一个仅使用一阶统计量的更简单的模型（即阵容中每位球员统计量的最小值）。与 ANC 86.7% 的测试精确率相比，更简单的模型仅达到了 75% 的测试精确率。

4.3 子分类器的解释

在使用表5中给出的优化参数值训练 ANC 后，我们现在尝试解释每个子分类器给出的分类结果，以了解哪些变量对于确定阵容表现最为重要。第k近邻分类器和支持向量机并不特别容易解释，因此我们主要关注其他五个子分类器。

在这五个可解释的子分类器中，对于精英分类而言，最为一致且重要的预测变量无疑是阵容中五名球员的最小每分钟正负值，即 $\text{PMM}_{(1)}$。这并不令人意外，因为如果阵容中表现最差球员的正负值很高，那么五名球员的正负值都必然很高，这表明在这些球员参与的任何阵容中，整体表现水平都很高。

在决策树中，经过剪枝后，$\text{PMM}_{(1)}$ 是树中唯一保留下来的分支。精英分类需要标准化的 $\text{PMM}_{(1)} > 0.32$。

表7列出了通过后退逐步逻辑回归确定的24个标准化预测变量的估计系数。无论从系数大小还是统计显著性来看，模型中最重要的三个预测变量是 $\text{PMM}_{(1)}$（五名球员中每分钟正负值最低的球员）、$\text{FG3A}_{(5)}$（三分球出手率最高的球员）

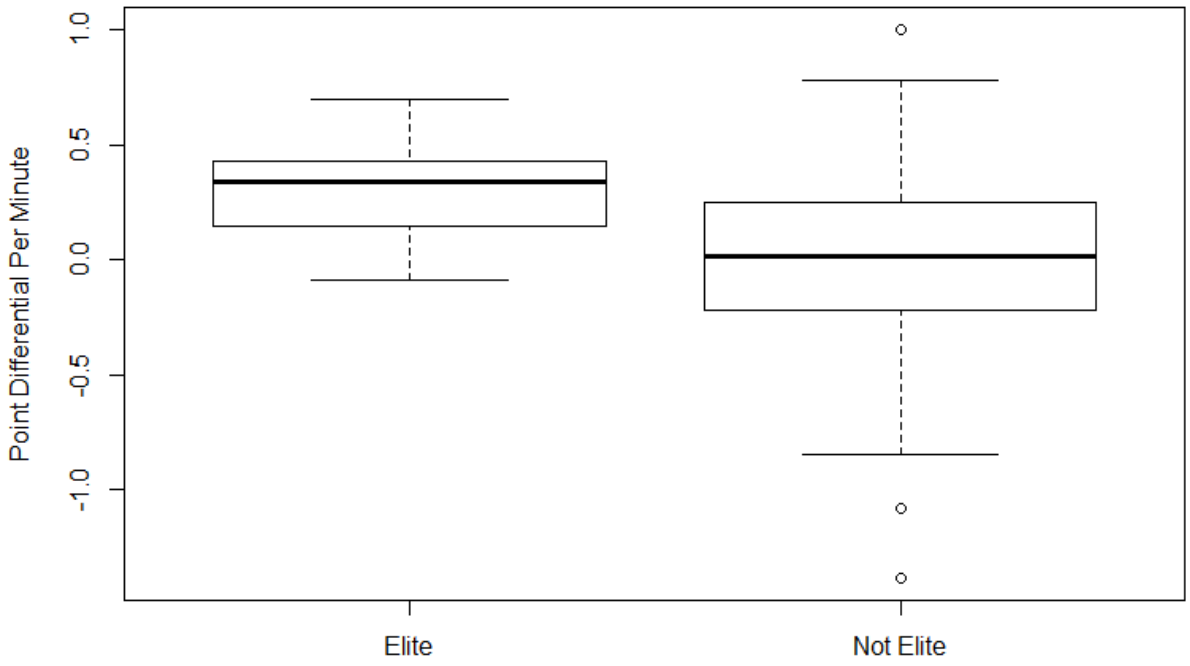


图3：按 ANC预测标签划分的阵容每分钟得分差（PMM）箱线图。

表7：逻辑回归模型的估计系数。显著性代码：***：0.001，**：0.01，*：0.05，.：0.1

预测变量	估计值	p-值
（截距）	0.12970	0.08068 .
每分钟投篮命中数(3)	0.17638	0.08454 。
投篮出手数(4)	-0.16022	0.10885
三分球命中数(5)	0.43471	0.04551 *
三分球出手数(5)	-0.60677	0.00666 **
三分球命中率(2)	0.22172	0.01398 *
三分球命中率(3)	-0.24260	0.01370 *
罚球出手次数(4)	0.21057	0.02255 *
罚球命中率(1)	0.19298	0.02183 *
罚球命中率(3)	-0.22849	0.03709 *
罚球命中率(4)	0.27984	0.02016 *
罚球命中率(5)	0.20627	0.04374 *
防守篮板(2)	0.17627	0.03486 *
失误(4)	-0.15526	0.06852 。
盖帽(5)	0.13514	0.10880
个人犯规(5)	0.20912	0.01139 *
造成犯规(1)	-0.19257	0.01774 *
PMM(1)	0.85462	¡ 2e-16 ***
受干扰投篮(4)	-0.19241	0.04944 *
受干扰投篮2PT(1)	0.19000	0.04680 *
受干扰投篮3分(5)	0.20103	0.03562 *
掩护助攻(2)	-0.17644	0.06542 .
掩护助攻(3)	0.30631	0.00984 **
掩护助攻(5)	0.23956	0.01057 *
挡拆卡位(3)	-0.26324	0.02719 *

五名球员每分钟的投篮次数，以及掩护助攻(3)，即促成进球的掩护每分钟发生率的中位数。此外，我们注意到，模型会选取多种进攻和防守数据（包括拼抢数据）来预测阵容表现。

对于线性判别分析、提升法和随机森林，可以通过图形评估预测变量的重要性，如图4所示。图4(a)展示了线性判别分析（LDA）模型中每个标准化预测变量的系数，并将其与正态分布的分位数进行比较。红色预测变量的系数绝对值尤其大，表明它们在将球队分类为精英或非精英时影响显著。除了PMM(1)之外，与受干扰投篮相关的数据在LDA模型中也具有较高的重要性。图4(b)展示了提升法的预测变量重要性，以基尼不纯度指数的降低量衡量，并将其与正态分布的分位数进行比较。红色点的重要性很高，对应的预测变量是PMM(1)、FTPCT(2)（第二低的罚球命中率）和FG3PCT(3)（三分球命中率的中位数）。图4(c)展示了随机森林子分类器的预测变量重要性，以基尼不纯度指数的降低量衡量，并将其与正态分布的分位数进行比较。重要性特别高的预测变量包括五名球员各自的正负值，以及第二低的三分球命中率（FG3PCT(2)）。我们得出结论，PMM(1) 始终是预测阵容表现最重要的预测变量，其次是三分球命中率和其他指标，但重要性相对较低。

4.4 2018-19赛季预测结果

鉴于分类器能够在2017-18赛季的数据中以较高的精确率识别出正PMM阵容，我们现在使用ANC为参加2018-19赛季NBA常规赛的三十支球队识别有潜力的阵容。我们使用2017-18赛季标准化后的训练数据和测试数据合并集，重新拟合ANC。我们使用2017-18赛季的数据均值和标准差对

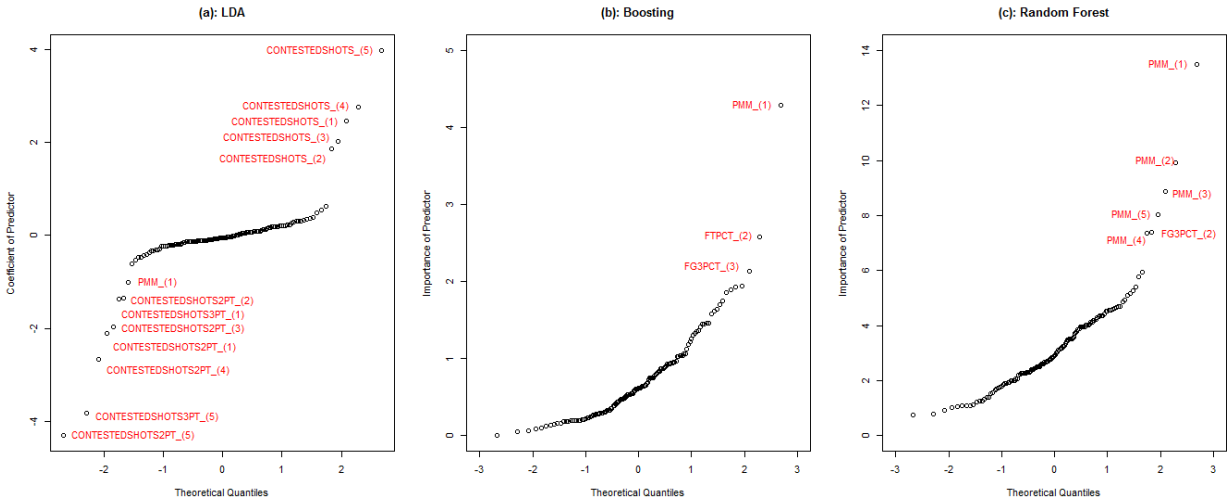


图4：(a) 线性判别分析、(b)提升法和(c)随机森林子分类器的预测变量重要性。以红色列出的预测变量对相应子分类器判定精英阵容具有重要影响。

2018-19常规赛数据进行平移和缩放，以避免数据泄漏；随后使用拟合后的ANC预测2018-19赛季的阵容表现。然后，我们可以将预测结果与该赛季的实际表现进行比较。

2018-19常规赛数据的收集方式如下。对于给定的球队名单，我们生成所有可能的五人阵容组合及其各球员针对 ANC 所用指标的顺序统计量（见附录A）。我们剔除任何球员在2017-18赛季常规赛上场时间不足50分钟的阵容。在30支 NBA 球队的2018-19常规赛名单所列的530名球员中，有392人在2017-18赛季常规赛期间拥有足够的上场时间，因此纳入分析。随后，我们可以预测这些可能的阵容中哪些会取得正PMM。这一过程有两点值得关注。首先，对于洛杉矶湖人队这类重建球队，例如2018-19赛季洛杉矶湖人队，我们可以预测新交易来的球员与现有球员的哪些组合可能会配合默契。其次，我们可以通过衡量各队板凳阵容中精英阵容的数量，比较 NBA 球队的相对实力。

表8列出了每支至少有一个精英阵容预测结果的球队所对应的精英阵容预测数量。预测精英阵容数量最多的六支球队分别是金州勇士队、休斯敦火箭队、印第安纳步行者队、波士顿凯尔特人队、犹他爵士队和夏洛特黄蜂队。其余球队的精英阵容预测数量均不超过四个，而三十支球队中有十四支没有任何精英阵容预测结果。由于优先考虑精确率，ANC 是一种对精英阵容采取保守预测的预测变量。

4.4.1 将2018-19赛季阵容预测结果与实际表现进行比较

现在，我们将 ANC 做出的预测与2018-19常规赛期间使用的阵容的实际表现进行比较。

我们首先考察混淆矩阵。对于任何给定的球队名单，大多数阵容（无论预测为精英阵容与否）可能从未被使用，而2018-19常规赛中使用的阵容偶尔会包含上场时间不足以纳入ANC预测的球员。因此，为了构建混淆矩阵，我们将分析范围限制在2018-19常规赛数据中出现且由ANC进行过预测的阵容。表9左侧是针对所有获得ANC预测的2018-19常规赛阵容计算出的混淆矩阵；表9右侧则是仅针对获得ANC预测且上场时间至少为25分钟的2018-19常规赛阵容计算出的混淆矩阵。限制范围的目的是聚焦于PMM估计较为可靠的阵容。在非限制情形下，我们看到，ANC预测为精英阵容的阵容中，有67.4%在2018-19常规赛期间取得了正PMM，

表8：ANC预测的2018-19赛季NBA常规赛精英阵容数量（针对精英阵容预测结果非零的16支球队）。

Team	预测的精英阵容数量
金州勇士队	127
休斯敦火箭队	62
印第安纳步行者队	30
波士顿凯尔特人队	25
犹他爵士队	18
夏洛特黄蜂队	14
俄克拉荷马城雷霆队	4
丹佛掘金队	4
波特兰开拓者队	3
新奥尔良鹈鹕队	3
明尼苏达森林狼	3
圣安东尼奥马刺	2
迈阿密热火	2
华盛顿奇才	1
多伦多猛龙	1
费城76人	1

表9：将经过调优的 ANC 应用于 2018-19常规赛数据所得的混淆矩阵。左侧的非限制情形包含所有具有 ANC预测结果的 2018-19常规赛阵容。在预测为精英的 46 个阵容中，有 31 个的真实标签为精英，对应的精确率为 67.4%。右侧的受限情形仅包含在 2018-19常规赛期间上场时间至少为 25 分钟且具有 ANC预测结果的阵容。在预测为精英的 26 个阵容中，有 20 个的真实标签为精英，对应的精确率为 76.9%。

		无限制			受限		
		预测类别			预测类别		
		精英	非精英		精英	非精英
True Class	精英	31	397	精英	20	201
	非精英	15	294	非精英	6	129

而精英阵容的总体占比为58.1%。当我们将分析范围限制在上场时间至少为25分钟的阵容时，精确率上升至76.9%，而精英阵容的占比为62.1%。我们得出结论：使用2017-18赛季的球员个人顺序统计量，通过ANC预测2018-19赛季的阵容表现，能够以较高精确率（尽管预测略显保守）预测出取得正PMM的阵容。

受限情形下，ANC预测为精英的26套阵容列于表10（2018-19赛季PMM为正）和表11（2018-19赛季PMM为非正），并附有其2018-19赛季的上场时间。在ANC预测为精英且最终在2018-19赛季确为精英的20套阵容中，进一步调查发现，其中有6套阵容包含在2017-18赛季与2018-19赛季之间新加入的球员，如表10最后一列所示。

我们还可以比较ANC预测为精英和预测为非精英的阵容在2018-19赛季实际取得的PMM。图5以箱线图展示了有ANC预测的阵容在2018-19常规赛期间观测到的PMM。图5(a)展示的是无限制情形，图5(b)则仅包括在2018-19常规赛中上场时间至少为25分钟的阵容。两种情形下都可以看到，ANC预测为精英的阵容，其PMM的方差低于预测为非精英的阵容。此外，当我们将考察范围限制在上场时间至少为25分钟的阵容时，由于这些阵容的PMM估计更为精确，可以看到预测为精英的阵容，其PMM分布几乎完全落在正值范围内。对均值进行单侧检验后发现，预测为精英的阵容的平均PMM很可能高于预测为非精英的阵容的平均PMM（$p = 0.090$）。

接下来，我们将探讨 ANC预测与球队整体板凳实力及表现之间的关系。

表10：ANC分类器预测为精英且具有正PMM的2018-19赛季阵容（仅限上场时间至少25分钟的阵容，与表9相同）。同时标出了包含新加入球员的阵容，这些球员在2017-18赛季未曾为该球队效力。

Team	阵容	上场分钟数	PMM	Note
波士顿凯尔特人队	A. Horford, K. Irving, M. Smart, J. Brown , J. Tatum	56	0.3	
夏洛特黄蜂队	M. Williams , N. Batum , K. Walker , J. Lamb、C. Zeller	593	0.16	
	M. Williams、N. Batum、K. Walker、J. Lamb、W. Hernangomez	34	0.03	威利·埃尔南戈麦斯曾效力于纽约尼克斯在2017-18赛季。
金州勇士队 riors	K. Durant、S. Curry、K. Thompson、D. Green、K. Looney	313	0.39	
	K. Durant、S. Curry、D. Cousins、K. Thompson、D. Green	268	0.29	DeMarcus Cousins 曾效力于新奥尔良鹈鹕队，效力赛季为2017-18赛季。
	A. Iguodala、K. Durant、S. Curry、K. Thompson、D. Green	178	0.69	
	A. Iguodala、K. Durant、S. Curry、K. Thompson、K. Looney	141	0.17	
	A. Iguodala、K. Durant、S. Curry、K. Thompson、J. Bell	36	0.73	
	A. Iguodala、S. Curry、D. Cousins、K. Thompson、D. Green	29	0.77	DeMarcus Cousins 曾效力于新奥尔良2017-18赛季的鹈鹕队。
	A. Iguodala、K. Durant、S. Curry、D. Green、K. Looney	25	0.8	
休斯敦火箭队	C. Paul , P. Tucker , E. 戈登 , J. 哈登、C. Capela	420	0.15	
	C. Paul、P. Tucker、J. Harden、A. Rivers, C. Capela	30	0.07	Austin Rivers 曾在洛杉矶快船队效力，时间是2017-18赛季。
印第安纳波利斯步ers	T. Young、D. Collison、B. Bogdanovic, V. Oladipo、M. Turner	555	0.1	
	T. Young、D. Collison、B. Bogdanovic, V. Oladipo、D. Sabonis	133	0.13	
	T. Young、C. Joseph、B. Bogdanovic、V. 奥拉迪波，M. Turner	29	0.03	
	D. Collison, B. Bogdanovic, V. Oladipo, M. Turner , D. Sabonis	26	0.39	
明尼苏达森林狼队-队	T. Gibson、R. Covington、A. Wiggins, T. Jones , K. Towns	77	0.32	Robert Covington 效力于费城76人队的2017-18赛季。
俄克拉荷马城雷霆	R. Westbrook、P. George、S. Adams、A. Abrines , J. Grant	88	0.21	
犹他爵士队	D. Favors、R. Gobert、J. Ingles、R. O'Neale , D. Mitchell	107	0.18	
	K. Korver、T. Sefolosha、R. Rubio、R. Gobert , D. Mitchell	26	0.7	凯尔·科沃尔效力于克利夫兰骑士在2017-18赛季。

表11：ANC分类器预测为精英、但PMM非正的2018-19赛季阵容（仅包括上场时间至少为25分钟的阵容，与表9相同）。此外，还标出了包含新引进球员的阵容，这些球员在2017-18赛季未曾效力于该球队。

Team	阵容	上场分钟数	PMM	Note
波士顿凯尔特人队	A. Horford、K. Irving、A. Baynes、J. 布朗 , J. Tatum	25	0	
金州勇士队 riors	K. Durant、S. Curry、J. Jerebko, K. 汤普森、D. Green	45	-0.22	Jonas Jerebko 于 2017 年效力于犹他爵士队-18.
	K. Durant、S. Curry、K. Thompson、D. Green , J. Bell	26	-0.38	
休斯敦火箭队	G. Green , P. Tucker , E. Gordon , J. 哈登，C. Capela	66	-0.32	
	C. Anthony、C. Paul、P. Tucker、E. Gordon，C. Capela	45	-0.11	卡梅罗·安东尼曾效力于俄克拉何马雷霆队在2017-18赛季。
印第安纳 步行者 队	T. Young、T. Evans、D. Collison、B. Bogdanovic、D. Sabonis	28	-0.43	泰瑞克·埃文斯曾效力于孟菲斯灰熊队, zlies 位于 2017-18赛季。

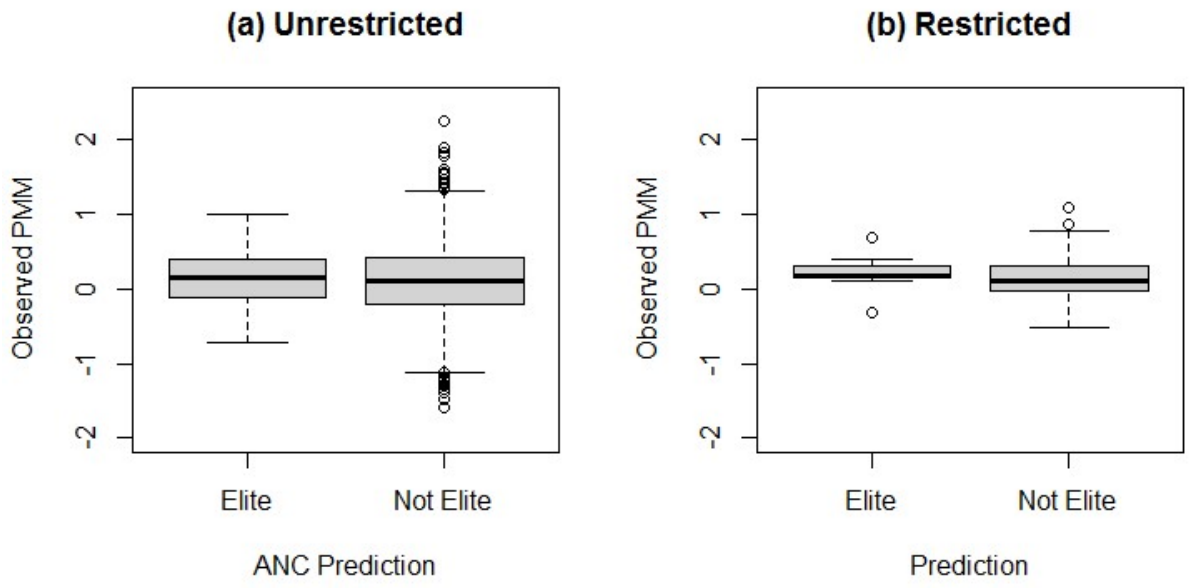


图5：按 ANC 预测标签划分的 2018-19赛季阵容每分钟得分差（PMM）箱线图。(a) 无限制情形。(b) 2018-19赛季阵容仅限上场时间至少为 25 分钟的阵容。

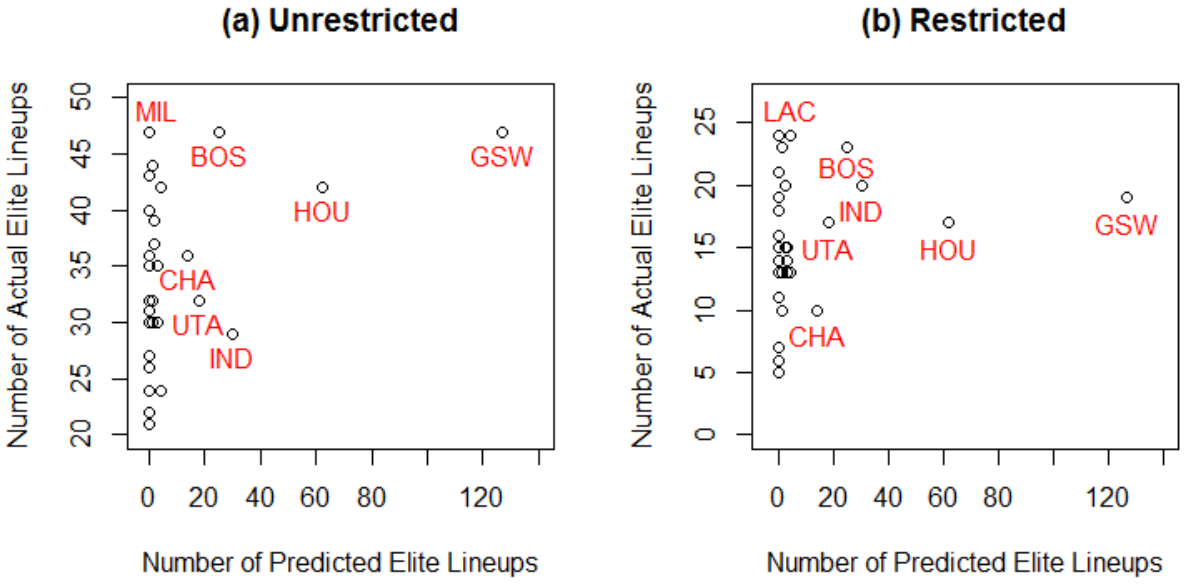


图6：2018-19常规赛期间各球队取得正PMM的阵容数量，与ANC预测各球队拥有的精英阵容数量。（a）非限制情形。（b）将2018-19赛季的阵容限定为上场时间至少达到25分钟的阵容。

图6绘出了30支NBA球队各自在2018-19常规赛中使用的、取得正PMM的阵容数量，以及ANC预测该球队会拥有的精英阵容数量。图6(a)展示的是非限制情形，图6(b)则将 y轴上的计数限定为2018-19赛季上场时间至少达到25分钟的阵容。首先值得注意的是，在预测拥有较多精英阵容的六支球队（金州勇士队、HOU、IND、BOS、UTA和CHA）中，波士顿和金州勇士队，以及在较小程度上的休斯敦，在非限制情形下也有相对较多的阵容于2018-19赛季取得正PMM。若将关注范围限定在上场时间至少达到25分钟的阵容（这些阵容的PMM估计更为精确），我们从图6(b)可以看到，波士顿，以及在较小程度上的印第安纳和金州勇士队，拥有实力雄厚的精英阵容替补阵容。

值得注意的是，ANC 会遗漏一些最终拥有许多正PMM阵容的球队。例如，从图6(a)中可以看到，ANC 预测密尔沃基雄鹿队的阵容没有一套会成为精英阵容，但该队有47套阵容取得了正PMM，其中19套的上场时间至少达到25分钟；这一表现与金州勇士队不相上下。对于密尔沃基雄鹿队而言，随着主教练Mike Budenholzer的聘任，2018-19赛季成为一个前所未有的赛季。球队取得了数十年来最佳的常规赛战绩，也是NBA整体常规赛战绩最佳的球队，赢得中央赛区冠军并闯入东部决赛 [32]。该赛季结束时，布登霍尔泽教练获选执教2019年NBA全明星赛东部队，并获评NBA年度最佳教练，还得到同行授予的美国国家篮球协会年度最佳教练奖 [33]。我们推测，尽管ANC能够很好地捕捉阵容的整体实力，但无法将教练组和执教风格的变化考虑在内。

同样，图6(b)显示，在2018-19赛季中，将上场时间限制为25分钟时，正PMM阵容数量最多的球队是洛杉矶快船队，而ANC也预测该队没有精英阵容。在这种情况下，该队在2018-19赛季使用的66套阵容中，有59套包含的球员在2017-18赛季个人上场时间不足，因此未能纳入ANC的考量。在2018-19赛季上场时间至少达到25分钟的36套阵容中，有34套包含的球员在2017-18赛季个人上场时间不足，因此未能纳入ANC预测。新秀球员数量较多的球队，以及从

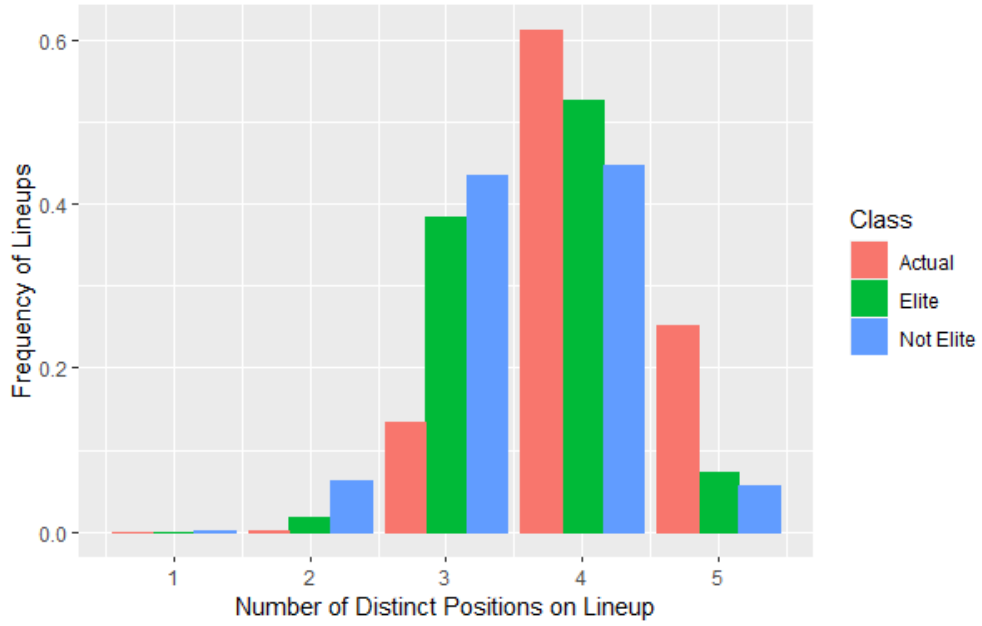


图7：实际阵容中不同位置数量的分布，与 ANC 预测为精英或非精英的阵容的对比。

G联盟，并上场　　前一赛季上场时间不足的球员，将无法从ANC获得许多预测结果　　.

4.4.2 ANC表现的其他方面

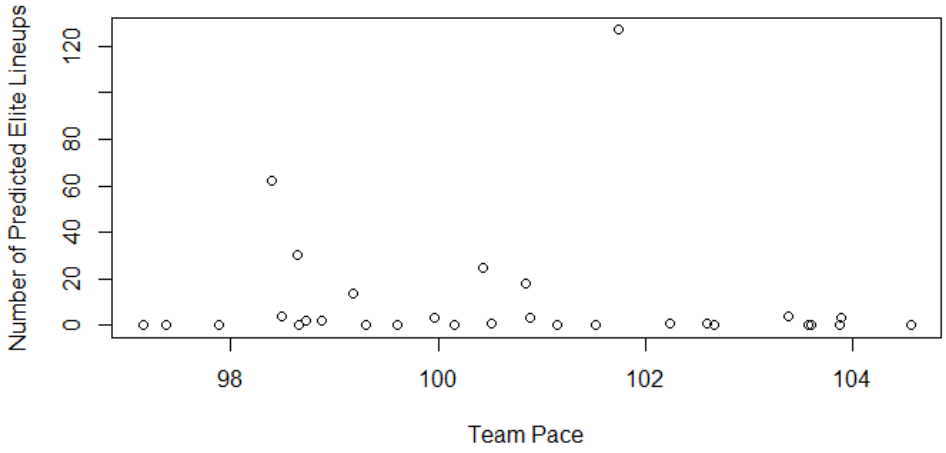
现在，我们将讨论 ANC 表现与球员位置和球队节奏的关系。

位置平衡。ANC 在预测阵容表现时只考虑球员顺序统计量，而忽略位置等其他球员信息。因此，一个自然的问题是：ANC 预测为精英的阵容是否具备位置平衡。利用球员位置数据 [3, 4], ，我们将位置定义为中锋、大前锋、控球后卫（专职控球后卫，或兼任控球后卫/得分后卫的球员）、得分后卫（专职得分后卫、兼任控球后卫/得分后卫的球员，或兼任小前锋/得分后卫的球员），以及小前锋（专职小前锋，或兼任小前锋/得分后卫的球员）。我们统计给定阵容中不同位置的数量。包含不同位置较多的阵容，比包含不同位置较少的阵容更加平衡。图7比较了实际阵容与 ANC 预测为精英或非精英的阵容中不同位置的分布。我们发现，与预测为非精英的阵容相比，ANC 预测为精英的阵容中，包含四个或五个不同位置的阵容更多。2018-19赛季实际获得上场时间的阵容整体上略微比 ANC 预测为精英的阵容更加平衡，但实际阵容中仍有 13.4% 使用了三个或更少的不同位置，其中许多阵容的 PMM 为正。这体现了 NBA 普遍逐渐摆脱僵化位置化打法的趋势 [19]。我们由此得出结论：ANC 预测为精英的阵容具备足够的位置平衡。

此外，ANC所使用的顺序统计量能够捕捉足够的位置角色信息，因此ANC不太可能推荐由五名相同位置球员组成的阵容。为说明这一点，我们构造了虚拟的2018-19赛季球员名单，其中所有球员都担任相同位置；分析仅限于纯位置，不包括控球后卫/得分后卫等混合位置。作为球员水平的替代指标，我们着重考察了2017-18赛季上场时间至少达到2100分钟的球员（这大致相当于各位置上场时间的最高四分位）。对于每个位置，我们从所有可能的五人阵容中均匀随机抽取200套阵容，并使用ANC预测该阵容是否会有正PMM。如表12所示，没有任何一套阵容被预测为精英阵容。因此，ANC使用的顺序统计量似乎

表12：ANC 对 2018-19赛季球员阵容的预测结果；这些阵容仅由单一位置的球员组成，且球员在 2017-18赛季的上场时间至少为 2100 分钟。

位置	球员	抽样阵容	预测为精英
得分后卫	22	200	0
大前锋	14	200	0
中锋	11	200	0
控球后卫	15	200	0
小前锋	19	200	0



Fi图8：ANC 根据球队节奏预测各球队拥有的精英阵容数量。

能够捕捉球场上不同位置的贡献。

球队节奏。ANC 依赖每分钟球员统计数据，而非每回合统计数据。由于众所周知各球队的节奏有所不同，一个自然的问题是：与慢节奏球队相比，ANC 是否会为快节奏球队预测出更多精英阵容？图8展示了根据 NBA.com [21] 报告的球队节奏，各球队被预测为精英阵容的数量。我们没有观察到球队节奏与 ANC 将阵容标记为精英阵容的倾向之间存在任何趋势。将球队节奏纳入 ANC，以预测由不同球队球员组合而成的阵容表现，将留待未来研究。

4.4.3 金州勇士队和洛杉矶湖人队案例研究

为进一步阐明ANC对2018-19常规赛的预测结果，我们以金州勇士队（此前两年的NBA总冠军）和洛杉矶湖人队为案例进行研究。洛杉矶湖人队在2018年休赛期引进勒布朗·詹姆斯，当时这一举动使许多人期待这支曾经强大的球队能够重返巅峰。

我们的ANC模型预测结果汇总于表13。洛杉矶湖人队在该赛季的球队名单中有15名球员具备足够的2017-18赛季个人上场时间，因此被纳入ANC预测，形成3003种可能的五人阵容。其中，没有任何一套阵容被预测为精英。金州勇士队在2018-19赛季的球队名单中同样有15名球员具备足够的2017-18赛季个人上场时间，形成3003种可能的五人阵容。与洛杉矶湖人相比，其中127套阵容被预测为精英。

为了理解为何洛杉矶湖人没有任何被预测为精英阵容的阵容，我们考察了表14中用于训练模型的2017-18赛季洛杉矶湖人和勇士队名单中各球员的PMM。我们看到

表13：ANC 为洛杉矶湖人队和金州勇士队 2018-19赛季名单预测的示例精英阵容。T

eam	Number of Pos- 可能阵容	Number of Pre- 预测的精英阵 ups	Sample Elite Li容
洛杉矶湖人队	3003	0	• 无
金州勇士队	3003	127	• D. Cousins、S. Curry、K. Durant、D. Green、A. Iguodala • S. Curry、K. Durant、J. Jerebko、K. Looney、K. Thompson • K. Durant、D. Green、A. Iguodala、K. Looney、K. Thompson • S. Curry、J. Bell、K. Durant、K. Looney、K. Thompson • D. Cousins, A. Iguodala, K. Looney、S. Livingston、K. Thompson

洛杉矶湖人名单中有几名球员的PMM为负。鉴于子分类器赋予PMM(1)的重要性，这有助于解释为何ANC没有预测洛杉矶湖人拥有任何精英阵容。预测不到精英阵容也与球队该赛季最终表现不佳相符。

作为两届卫冕冠军，金州勇士队进入2018-19赛季时，其全部阵容组合中被归类为精英阵容的比例很高，这并不令人意外。表13列出了预测为精英阵容的127套勇士队阵容中的一部分；这些阵容展现了多种选择，当“汉普顿五人组”（斯蒂芬·库里、克莱·汤普森、安德烈·伊戈达拉、凯文·杜兰特、德雷蒙德·格林）未能同时在场时，教练组可以依靠它们来优化球员的使用。此外，金州勇士队拥有大量精英阵容，也表明球队阵容深度出色，并与其当时的战绩相符。

现在，我们将基于2017-18赛季表现得出的ANC预测结果，与洛杉矶湖人队和金州勇士队在2018-19常规赛期间的实际表现进行比较。

表15分别给出了洛杉矶湖人和勇士队的混淆矩阵，显示在2018-19常规赛期间上场时间至少为25分钟的阵容中，预测为精英或非精英的阵容数量与其实际表现（正PMM与非正PMM）的对比 [20]。我们看到，在根据2017-18赛季数据被ANC预测为精英的金州勇士队9套阵容中，有7套最终在2018-19赛季取得了正PMM，对应的精确率为77.8%。（由于ANC没有将任何湖人阵容预测为精英，因此无法计算其精确率。）

表14：洛杉矶湖人队和金州勇士队2018-19赛季名单中球员在2017-18赛季常规赛期间的个人每分钟正负值（PMM）。表中未列出湖人队球员 Isaac Bonga、Jemerrio Jones、Scott Machado、Sviatoslav Mykhailiuk、Moritz Wagner 和 Johnathan Williams，以及勇士队球员 Jacob Evans 和 Marcus Derrickson，因为他们在2017-18赛季 NBA 常规赛的数据不足；也未列出湖人队球员 Ivica Zubac，他在赛季中期被交易至洛杉矶快船队，且 NBA.com 将其列在该队的球队名单中。

洛杉矶湖人队		金州勇士队	
球员	个人 PMM	球员	个人 PMM
Lonzo Ball	-0.01	Jordan Bell	0.20
Michael Beasley	-0.08	Andrew Bogut	-0.13
Reggie Bullock	0.01	Quinn Cook	-0.05
肯塔维厄斯·考德威尔-波普	-0.02	DeMarcus Cousins	0.05
亚历克斯·卡鲁索	-0.01	斯蒂芬·库里	0.30
泰森·钱德勒	-0.21	凯文·杜兰特	0.15
乔什·哈特	-0.05	德雷蒙德·格林	0.14
安德烈·英格拉姆	0.31	安德烈·伊戈达拉	0.17
布兰登·英格拉姆	-0.06	Jonas Jerebko	0.05
勒布朗·詹姆斯	0.03	达米安·琼斯	0.00
凯尔·库兹马	-0.05	达米恩·李	-0.06
迈克·穆斯卡拉	-0.07	肖恩·利文斯顿	0.10
贾维尔·麦基	0.06	凯文·卢尼	0.13
拉简·隆多	0.01	Alfonzo McKinnie	-0.22
兰斯·史蒂芬森	-0.06	克莱·汤普森	0.16

表15：基于2017-18赛季数据得出的ANC预测结果，与2018-19赛季实际情况相比，分别针对洛杉矶湖人队和金州勇士队的混淆矩阵；其中，阵容在2018-19赛季常规赛期间的上场时间至少为25分钟，以及阵容中的球员在2017-18赛季常规赛期间的上场时间至少为50分钟。

洛杉矶湖人队				金州勇士队			
		预测类别				预测类别	
		精英	非精英			精英	非精英
True Class	精英	0	12	True Class	精英	7	12
	非精英	0	9		非精英	2	4

附录C中的表19和表20分别列出了2018-19常规赛期间上场时间至少为25分钟的洛杉矶湖人和勇士队所有五人阵容的实现PMM，以及ANC预测结果。“–”表示未给出ANC预测的阵容。我们证实，虽然ANC始终将低PMM阵容预测为非精英，但它是一个保守分类器，偶尔会漏掉下一赛季实际取得高PMM的阵容。例如，两支球队中表现最佳的阵容——金州勇士队由McKinnie、Green、Looney、Livingston和Curry组成的阵容——被ANC预测为非精英。这一非精英预测可能是由于Alfonzo McKinnie在2017-18赛季的个体PMM为负；而实际PMM相对较高，则可能是因为该阵容在2018-19赛季的上场时间仅有28分钟，导致结果不精确。

5 结论

我们开发了一种全有或全无分类器，可根据球员个人顺序统计量预测五人篮球阵容的表现。通过调整分类器以实现高精确率，ANC 预测为精英阵容的阵容，其上场时间每分钟平均正负值高于预测为非精英的球队；这一结论在基于留出的测试数据集进行的同赛季预测和下一赛季预测中均成立。我们还发现，拥有大量预测为精英阵容的球队，往往也拥有大量取得高

PMM 的阵容，这表明 ANC 可以作为板凳深度的指标。ANC 纳入球员个人进攻与防守顺序统计量，既能捕捉位置化打法，也能推荐在位置平衡方面与实际比赛中使用的阵容相符的阵容。

虽然 ANC 分类为精英的阵容，其 PMM 在统计上确实高于被分类为非精英的阵容，但 ANC 有时会漏掉表现极佳的阵容。未来研究的一种可能方向是调整将阵容标记为精英所使用的 PMM 阈值。目前我们的阈值为零，但我们或许会希望识别并描述 PMM 高出零一定幅度的阵容；分类器在识别更极端的情况时可能效果更好。此外，有证据表明，对于教练组和战术风格发生变化的球队，以及拥有大量新手球员的球队，ANC 会失去预测能力。未来研究可以考察，能否利用其他联盟（如 G联盟或 NCAA）前一年的数据，为 NBA 历史数据不足的球员做出预测。

尽管如此，由于 ANC 优先考虑精确率，ANC 预测为精英的阵容往往确实会实现正PMM。我们得出结论，这种分类器可以纳入决策支持工具，用于辅助球员引进和换人。

致谢 本材料所述研究得到了美国国家科学基金会的资助（资助编号 DMS-1757952）。本材料中表达的任何意见、发现、结论或建议均代表作者的观点，并不一定反映美国国家科学基金会的意见。作者还感谢哈维穆德学院提供的资金支持。作者感谢 Isys Johnson、Lucius Bynum 和 Robert Gonzalez 对本研究早期阶段及代码库的贡献，其中部分代码经过改编后用于本文。最后，作者感谢匿名审稿人和编辑，他们的反馈大大改进了本文的分析。

参考文献

[1] Basketball-Reference.com（2021a）。2017–18赛季波士顿凯尔特人队的球队名单和统计数据。访问网址：https://www. basketball-reference.com/teams/BOS/2018.html，访问日期：2021年6月7日。[2] Basketball-Reference.com（2021b）。2017–18赛季休斯敦火箭队的球队名单和统计数据。访问网址：https://www. basketball-reference.com/teams/HOU/2018.html，访问日期：2021年6月7日。[3] Basketball-Reference.com（2022a）。2017–18赛季NBA球员统计数据：场均数据。访问网址：https://www. basketball-reference.com/leagues/NBA_2018_per_game.html，访问日期：2022年4月7日。[4] Basketball-Reference.com（2022b）。2018–19赛季NBA球员统计数据：场均数据。访问网址：https://www. basketball-reference.com/leagues/NBA_2019_per_game.html，访问日期：2022年5月10日。[5] Bendl, J., J. Stourac, O. Salanda, A. Pavelka, E. Wieben, J. Zendulka, J. Brezovsky, 和 J. Damborsky（2014）。PredictSNP：用于预测疾病相关突变的稳健且准确的共识分类器。《*PLoS Comput Biol*》*10*(1)。[6] Bynum, L. E. J.（2018）。子集行为建模：职业篮球数据的处方分析。本科毕业论文（克莱蒙特：哈维穆德学院）。[7] Cheng, G., Z. Zhang, M. Kyebambe, 和 K. Nasser（2016）。基于最大熵原理预测NBA季后赛结果。《*Entropy*》*18*，450。[8] Clemente, F., F. Martins, D. Kalamaras, 和 R. Mendes（2015）。篮球中的网络分析：利用中心性指标考察突出球员。《体育教育与运动杂志》*15*，212–217。[9] Deshpande, S. K. 和 S. T. Jensen（2016）。估算NBA球员对球队获胜机会的影响。《体育定量分析期刊》*12*(2)，51–72。[10] Ghimire, S., J. A. Ehrlich, 和 S. D. Sanders（2020）。衡量互补型团队环境中的个体员工产出：正则化调整正负值能否分离出NBA球员的个人贡献？《*PloS one*》*15*(8)，e0237920。

[11] Glickman, M. 和 J. Sonas（2015）。NCAA男子篮球预测方法专题导论。《体育定量分析期刊》*11*，1–3。[12] Gumm, J., G. Hu, 和 A. Barrett（2015）。一种预测疯狂三月获胜者的机器学习策略。载于第*16*届*IEEE/ACIS*软件工程、人工智能、网络与并行/分布式计算国际会议（*SNPD*）论文集，第1–6页。IEEE。[13] Hua, S.（2015）。比较NCAA疯狂三月的若干建模方法。博士学位论文（北达科他州立大学）。[14] Kalman, S. 和 J. Bosch（2020）。基于聚类球员倾向的NBA阵容分析：篮球位置研究的新方法，以及软阵容聚合的阵容效率建模。42 analytics（2020）。载于第*14*届*MIT*斯隆体育分析大会论文集，美国马萨诸塞州波士顿。[15] Kubatko, J., D. Oliver, K. Pelton, 和 D. T. Rosebaum（2007）。分析篮球统计数据的起点。《体育定量分析期刊》*3*(3)。[16] Lin, R.（2017）。Mason：实时NBA比赛结果预测。博士学位论文（亚利桑那州立大学）。[17] Loeffelholz, B., E. Bednar, 和 K. Bauer（2009）。利用神经网络预测NBA比赛结果。《体育定量分析期刊》*5*(1)，1–15。[18] Maymin, A., P. Maymin, 和 E. Shen（2013）。NBA化学反应：篮球中的正向和负向协同效应。《国际体育计算机科学期刊》*12*(2)，4–23。[19] McMahon, I.（2018）。无位置阵容如何（以及为何）席卷NBA季后赛。《卫报》。访问网址：https://www.theguardian.com/sport/blog/2018/may/01/how-and-why-position-less-lineups-have-taken-over-the-nba-playoffs，访问日期：2022年5月11日。[20] NBA.com（2019a）。NBA高级统计数据：统计数据主页/阵容/传统数据。https://www.nba.com/stats/lineups/traditional/?Season=2018-19&SeasonType=Regular%20Season&sort=MIN&dir=1&PerMode=Totals 访问日期：2021年5月20日。[21] NBA.com（2019b）。NBA高级统计数据：统计数据主页/球队/高级数据。https://www.nba.com/stats/ teams/advanced/?sort=W&dir=-1&Season=2018-19&SeasonType=Regular%20Season 访问日期：2022年5月9日。[22] Oh, M., S. Keshri, 和 G. Iyengar（2015）。用于篮球比赛模拟的图模型。载于*2015*年*MIT*斯隆体育分析大会论文集，第2728卷。.[23] Ozmen, M. U.（2016）。比赛统计数据对不同级别篮球比赛获胜概率的边际贡献：来自欧洲篮球联赛的证据。《国际运动科学与教练期刊》*11*，98–107。[24] Pelechrinis, K.（2019）。LinNet：通过网络嵌入进行概率阵容评估。载于 U. Brefeld, E. Curry, E. Daly, B. MacNamee, A. Marascu, F. Pinelli, M. Berlingerio, 和 N. Hurley（编），《数据库中的机器学习与知识发现》，第20–36页。施普林格国际出版。[25] Ribeiro, J., P. Silva, R. Duarte, K. Davids, 和 J. Garganta（2017）。透过社会网络理论视角分析团队运动表现：对研究与实践的启示。《运动医学》*47*，1–8。[26] Robertson, M.（2017）。NBA时空数据分析。硕士学位论文（杜克大学）。[27] Ruiz, F. J. 和 F. Perez-Cruz（2015）。一种预测大学篮球比赛结果的生成模型。《体育定量分析期刊》*11*(1)，39–52。

[28] Shen, G., D. Gao, Q. Wen, and R. Magel（2016）。使用三种不同方法预测疯狂三月的结果。《体育研究杂志》*3*，10–17。[29] Sisneros, R. and M. Van Moer（2013）。扩展正负值，以便对 NBA 技术统计数据进行可视化和统计分析。第 *1* 届 *IEEE* 体育数据可视化研讨会。[30] Vaz de Melo, P., V. Almeida, A. Loureiro, and C. Faloutsos（2012）。预测 NBA 及其他团队运动的结果：网络效应的实际应用。*ACM*《知识发现与数据汇刊》*6*，13。[31] W¨asche, H., G. Dickson, A. Woll, and U. Brandes（2017）。体育研究中的社会网络分析：一种新兴范式。《欧洲体育与社会杂志》*14*，1–28。[32] 维基百科（2021年）。2018–19赛季密尔沃基雄鹿队。访问网址：https://en.wikipedia.org/wiki/ 2018%E2%80%9319_Milwaukee_Bucks_season，访问日期：2022年5月4日。[33]维基百科（2022年）。迈克·布登霍尔泽。访问网址：https://en.wikipedia.org/wiki/Mike_Budenholzer，访问日期：2022年5月4日。[34] Winston, W. L.（2012）。《数学竞技：赌徒、经理和体育爱好者如何在棒球、篮球和橄榄球中运用数学》。普林斯顿大学出版社。[35] Zimmermann, A., S. Moorthy, and Z. Shi（2013）。使用机器学习技术预测美国大学篮球比赛结果——一些结果与经验教训。载于*MLSA@PKDD/ECML*。

A 个别球员统计数据作为 ANC 中的预测变量

表 16：在 ANC 中用作预测变量的个人球员统计数据。每分钟投中球数 FGM 每分钟投篮次数 FGA 投篮命中率 FGPCT 每分钟投中三分球数 FG3M 每分钟三分球出手次数 FG3A 三分球命中率 FG3PCT 每分钟罚球命中数 FTM 罚球出手次数 FTA 罚球命中率 FTPCT 每分钟进攻篮板 OREB 防守篮板 DREB 每分钟助攻 AST 每分钟失误数 TOV 每分钟抢断数 STL 盖帽 BLK 每分钟投篮次数 BLKA 每分钟个人犯规数 PF 每分钟得分 PTS 每分钟造成犯规 PFD 每分钟正负值 PMM 每分钟干扰投篮次数 CONTESTEDSHOTS 每分钟干扰两分球次数 CONTESTEDSHOTS2PT 每分钟干扰三分球次数 CONTESTEDSHOTS3PT 每分钟造成进攻犯规 CHARGESDRAWN 每分钟干扰传球 DEFLECTIONS 每分钟抢回松散球 LOOSEBALLSRECOVERED 每分钟助攻掩护 SCREENASSISTS 每分钟卡位次数 BOXOUTS

表17：基于一阶统计量的简单模型所使用的调优参数值。

子分类器	参数	选定值
决策树	cp（成本复杂度）	-1
	loss（误分类惩罚）	1
随机森林	c（截止值）	0.7
	ntree（树的数量）	100
提升法	mfinal（树的数量）	500
	maxdepth（每棵树的深度）	3
	cp（成本复杂度）	0.01
支持向量机	成本（误分类惩罚）	0.1
	伽马（影响衰减）	0.01
K近邻	k（邻居数量）	5
逻辑回归	阈值（1–概率阈值）	0.25
全或无分类器（ANC）	投票数（所需一致意见数）	7

表18：将基于一阶统计量的简单模型应用于测试数据集所得的混淆矩阵。在预测为精英的12个阵容中，有9个的真实标签为精英，对应的精确率为75.0%。

	预测类别 精英	非精英
True Class		
精英	9	86
非精英	3	78

B ANC与较简单模型的比较

人们或许还会想知道，ANC 是否需要完整的五人次序统计量集合才能达到较高的精确率。本节中，我们分析一个更简单的模型，该模型仅使用 ANC 所采用的每项单个球员指标的一阶统计量（即阵容的最小值）。

我们按照第4.1节所述，使用十折交叉验证来调整简单模型的参数。训练数据中各折的平均精确率和最坏情况精确率构成有效前沿，其对应的参数组合见表17。该组合在训练数据上的平均精确率为 86.5%，最低精确率为 57.1%，平均准确率为 51.8%。当模型性能对某个参数值不敏感时，我们选择与 ANC 中所用参数值相同的值。

调整好参数后，我们按照前文所述，将一阶统计模型拟合到完整的标准化训练集上，并将训练好的模型应用于测试数据。混淆矩阵见表18。

在预测为精英的十二个阵容中，有九个的真实标签为精英，表明其具有严格正PMM。与 ANC 的测试精确率 86.7% 相比，更简单的模型仅达到 75% 的测试精确率。

C 洛杉矶湖人队和金州勇士队案例研究中的实际阵容表现 dy

表 19：2018-19赛季洛杉矶湖人队的实际阵容表现与 ANC 预测结果对比，涵盖上场时间至少为 25 分钟的所有阵容。‘–’表示未提供 ANC 预测结果的阵容。

洛杉矶湖人队

阵容	分钟上场	实际 PMM	ANC预测
R. Rondo, K. Caldwell-Pope, B. Ingram, I. Zubac[1],J. Hart	25	0.68	–
L. James、B. Ingram、I. Zubac、L. Ball、K. Kuzma	55	0.36	–
T. Chandler、L. James、K. Caldwell-Pope、L. Ball、K. Kuzma	39	0.36	非精英
L. James , J. McGee , L. Ball , K. Kuzma , J. Hart	133	0.31	非精英
L. James、R. Rondo、J. McGee、K. Caldwell-Pope、K. Kuzma	47	0.23	非精英
T. Chandler、L. Stephenson、K. Caldwell-Pope、B. Ingram、J. Hart	37	0.21	非精英
T. Chandler、B. Ingram、L. Ball、K. Kuzma、J. Hart	36	0.14	非精英
T. Chandler、L. James、B. Ingram、L. Ball、K. Kuzma	61	0.13	非精英
L. James、R. Rondo、J. McGee、K. Caldwell-Pope、B. Ingram	31	0.13	非精英
B. Ingram、I. Zubac、L. Ball、K. Kuzma、J. Hart	39	0.13	–
L. James、J. McGee、K. Caldwell-Pope、L. Ball、K. Kuzma	34	0.12	非精英
L. James、J. McGee、R. Bullock、B. Ingram、K. Kuzma	73	0.11	非精英
T. Chandler、K. Caldwell-Pope、B. Ingram、K. Kuzma、J. Hart	31	0.10	非精英
T. Chandler、K. Caldwell-Pope、B. Ingram、L. Ball、K. Kuzma	45	0.04	非精英
T. Chandler , L. James , L. Ball , K. Kuzma , J. Hart	66	0.02	非精英
L. James, R. Rondo, J. McGee, B. Ingram, K. Kuzma	43	0.00	非精英
L. James, J. McGee, B. Ingram, L. Ball, K. Kuzma	234	0.00	非精英
L. James、R. Rondo、R. Bullock、B. Ingram、K. Kuzma	62	-0.05	非精英
J. McGee, K. Caldwell-Pope, 迈克·穆斯卡拉, 亚历克斯·卡鲁索, J. Jones[2]	31	-0.06	–
L. James、K. Caldwell-Pope、L. Ball、K. Kuzma、J. Hart	25	-0.16	非精英
L. James , R. Rondo , J. McGee , R. Bullock , K. Kuzma	62	-0.21	非精英
R. Rondo、K. Caldwell-Pope、B. Ingram、I. Zubac、K. Kuzma	29	-0.24	–
L. James、L. Stephenson、L. Ball、K. Kuzma、J. Hart	31	-0.25	非精英
R. Rondo、M. Beasley、K. Caldwell-Pope、B. Ingram、I. Zubac	25	-0.28	–
J. McGee、K. Caldwell-Pope、B. Ingram、L. Ball、J. Hart	25	-0.32	非精英
L. James、R. Rondo、B. Ingram、I. Zubac、K. Kuzma	33	-0.43	非精英
J. McGee、B. Ingram、L. Ball、K. Kuzma、J. Hart	83	-0.47	非精英
R. Rondo、J. McGee、K. Caldwell-Pope、亚历克斯·卡鲁索、莫里茨·瓦格纳[3]	27	-1.31	–

[1]Ivica Zubac 被交易至洛杉矶快船队，因此未纳入洛杉矶湖人的 ANC预测。[2]Jemerrio Jones 没有 2017-18 NBA 常规赛的数据。[3]Moritz Wagner 没有 2017-18 NBA 常规赛的数据。[4]Jacob Evans 没有 2017-18 NBA 常规赛的数据。

表20：2018-19赛季金州勇士队所有上场时间至少为25分钟的阵容的实际阵容表现与ANC预测的比较。“–”表示未给出ANC预测的阵容。

金州勇士队

阵容	分钟场次	实际PMM	ANC预测
A. McKinnie、D. Green、K. Looney、S. Livingston、S. Curry,	28	1.01	非精英
A. Iguodala、D. Green、K. Durant、K. Looney、S. Curry	25	0.80	精英
A. Iguodala、D. Cousins、D. Green、K. Thompson、S. Curry	29	0.77	精英
A. Iguodala、J. Bell、K. Durant、K. Thompson、S. Curry	36	0.73	精英
A. Iguodala、D. Green、K. Durant、K. Thompson、S. Curry	178	0.69	精英
A. Iguodala、K. Durant、K. Looney、K. Thompson、Q. Cook	35	0.63	非精英
A. McKinnie、J. Jerebko、K. Durant、K. Looney、S. Curry	26	0.61	非精英
D. Green、K. Durant、K. Looney、K. Thompson、S. Curry	313	0.39	精英
D. Cousins、D. Green、K. Durant、K. Thompson、S. Curry	268	0.29	精英
A. Bogut、D. Green、K. Durant、K. Thompson、S. Curry	83	0.27	非精英
A. Iguodala、D. Cousins、D. Green、K. Thompson、S. Livingston	67	0.24	非精英
D. Jones、J. Jerebko、K. Durant、K. Thompson、Q. Cook	29	0.20	非精英
A. McKinnie、A. Iguodala、K. Durant、K. Looney、S. Curry	48	0.19	非精英
A. Iguodala、K. Durant、K. Looney、K. Thompson、S. Curry	141	0.17	精英
A. Iguodala、J. Jerebko、K. Durant、K. Looney、K. Thompson	47	0.13	非精英
A. McKinnie、A. Iguodala、J. Jerebko、K. Looney、S. Curry	27	0.11	非精英
D. Jones、D. Green、K. Durant、K. Thompson、S. Curry	142	0.11	非精英
A. Iguodala、D. Green、J. Jerebko、S. Livingston、S. Curry	54	0.07	非精英
A. Iguodala、D. Cousins、K. Thompson、Q. Cook、S. Livingston	39	0.05	非精英
A. Iguodala、D. Green、J. Jerebko、K. Thompson、S. Livingston	26	0.00	非精英
D. Lee、J. Jerebko、K. Looney、K. Thompson、S. Livingston	30	0.00	非精英
D. Green、J. Jerebko、K. Durant、K. Thompson、S. Curry	45	-0.22	精英
A. Iguodala、D. Jones、K. Durant、K. Thompson、Q. Cook	77	-0.33	非精英
D. Green、J. Bell、K. Durant、K. Thompson、S. Curry	26	-0.38	精英
A. McKinnie, D. Cousins, D. Green, K. Durant, S. Curry	32	-0.56	非精英
A. McKinnie、J. Evans[4], J. Jerebko、J. Bell、Q. Cook	37	-0.57	–