假设：考虑表现得分时不考虑对手表现的影响

没有考虑一发得分率和二发得分率

数据处理

异常值处理

数据转换

特征提取

计算各场比赛的优胜比例

连续得分

连续失分

问题一

（1）筛选建模指标

|  |  |  |
| --- | --- | --- |
|  |  |  |
| X1 | Point\_victor | 是否得分 0.，1 |
| X2发球 |  | ACE 0，1 |
| X3 发球 |  | 双误 -1，0 |
| X4 |  | 非受迫性失误-1，0 |
| X5 |  | 上网得分 0，1 |
| X6 发球 |  | 发球宽度 |
| X7 发球 |  | 发球深度 |
| X8 发球 |  | 速度 |
| X9 |  | 距离 |
| X10接球 |  | Winner 0，1 |
| X11接球 |  | 返回球深度 |
| X12 接球 |  | 是否破发点 -1，0，1 |

(2)构建表现得分模型

发球表现得分模型：

接球表现得分模型：

采用AHP-CRITIC综合方法计算权重

AHP步骤

发球判断矩阵（由于考虑到发球优势较明显，所以在设置发球判断矩阵时将X1相对于其他指标的重要性适当减弱了）：

接球判断矩阵：

判断矩阵设计时考虑到了发球优势：由于发球方胜率为0.65，

CRITIC步骤：

AHP-CRITIC组合权重

（3）可视化

Pl和p2的表现得分可视化

从图中可以看出，xxx的表现得分总体高于xxx，符合比赛结果，说明模型准确性较高

（4）与历史数据结果与趋势对比，说明模型的准确性（可小波分析表现得分趋势 后面有时间补） 并分析在p1发球的5球内，p1的发球得分是否高于p2，说明发球优势

发球和接球

考虑破发局和保发局

（5）预测模型

问题二：

模拟随机数据，检验说明非随机，接着进一步说明其与动量的关联 （待定）

1.计算指标

（1）计算动量

解释为何以局为单位来构造动量 合理性

1. \*\*局的重要性\*\*：在网球比赛中，每个局都是赢得比赛的基本组成部分。因此，球员在每个局中的表现对于整场比赛的结果至关重要。

2. \*\*局内的动量\*\*：在每个局内，球员的表现会受到前几个球的影响，比如发球、接发球、得分和失误等。这些事件的连续发生会构成局内的动量变化。

3. \*\*局间的转折点\*\*：当一个球员赢得一个局时，他们可能会在心理上获得优势，并在下个局中保持动量。相反，输掉一个局可能会影响球员的信心和表现。因此，局与局之间的胜负关系可以作为动量变化的重要指标。

4. \*\*动量的传递\*\*：在网球比赛中，局内的胜利通常会传递到下个局中，从而影响到整个比赛的走势。因此，局内的动量变化可以帮助我们理解比赛中的关键时刻和转折点。

通过以局为单位构造动量，我们可以更好地理解球员在比赛中的表现，并分析他们在不同局中的动态变化。这有助于教练和分析员更好地制定战术和策略，以应对比赛中的不同情况。

定义动量，说明为何选取该指标

动量 连续得分win\_points 连续失分 lose\_points 连续赢的局数win\_games 连续输的局数lose\_games 连续赢的盘数win\_sets,连续输的盘数lose\_sets, 发球优势serve

CRITIC法赋权

连续赢的局数:在本盘中到目前为止连续赢的最大局数

连续赢的盘数：在本次match中到目前为止连续赢的最大盘数

连续得分：在本局中最高的连续得分次数

题目中提到动量被定义为“力量或通过运动或一系列事件获得的力量”，并指出在比赛中球队或球员可能会感觉到他们拥有动量，但很难对这种现象进行量化和衡量。因此，在构建动量模型时，我们应该尽可能地接近这个定义，并尝试从比赛中可测量的因素中推导出动量的概念。

考虑到这一定义，我们的动量模型应该能够反映出球员在比赛中的“力量”或“势头”的变化，以及这种变化是如何由一系列事件或表现所驱动的。因此，我们选择的因素和指标应该能够捕捉到球员在比赛中的表现变化，并将其转化为动量的量化度量。

1. \*\*表现得分的斜率\*\*：考虑到斜率可以反映球员表现的变化速度，因此将其纳入动量的计算中是合理的。较大的正斜率可能表示球员表现的快速改善，而较大的负斜率可能表示球员表现的迅速恶化。

2. \*\*连续赢的局和盘\*\*：连续赢得局和盘是一个重要的动量指标，因为它们可以反映球员在比赛中建立的优势和动力。连续赢得局和盘的球员往往会在心理上更有信心，同时也会对对手产生压力。

3. \*\*发球优势\*\*：在网球比赛中，发球优势可以带来重要的竞技优势，因为发球方有更大的控制权和更高的获胜概率。因此，将发球优势纳入动量的考虑是合理的。

4. \*\*连续得失分\*\*：连续得失分可以反映比赛中的连续性，这对于捕捉到比赛中的动态变化和转折点是很重要的。连续得分可能表示球员的持续优势，而连续失分可能表示球员的快速下滑。

计算动量

CRITIC法赋权：CRITIC法是一种常用的赋权方法，它考虑了多个指标之间的相对重要性，并根据其对比值来赋予权重。这种方法能够有效地平衡各个指标之间的影响，从而更准确地反映出动量的变化。

划分种类 （为了根据不同的类别提供不同的建议）

（2）计算局势波动 =我当前的表现得分-对方的表现得分 和连续成功次数

为正表示走势偏向我，为负表示比赛偏向对方

将局势波动定义为你的表现得分与对手的表现得分之差是一个合理的方法，因为这可以捕捉到比赛中你与对手之间的竞争情况。当你的表现得分高于对手时，局势波动值会为正，反之为负，这可以反映出比赛中的优劣势和局势的变化。

为了说明这种定义的合理性，你可以考虑以下几点：

1. \*\*竞争性体现\*\*：局势波动的定义直接反映了比赛中你与对手之间的竞争情况。正值表示你处于优势地位，负值表示对手处于优势地位。

2. \*\*动态性\*\*：局势波动随着比赛的进行而动态变化，能够捕捉到比赛中局势的变化趋势。这有助于分析比赛中的关键时刻和转折点。

3. \*\*量化指标\*\*：局势波动的定义提供了一个量化的指标来衡量比赛中的局势变化。你可以通过比较局势波动的大小和方向来评估比赛中的优劣势。

4. \*\*可解释性\*\*：局势波动的定义直观且易于理解，能够直接反映出你与对手之间的得分差异。这样的定义有助于深入分析比赛中的各种情况和策略。

综上所述，将局势波动定义为你的表现得分与对手的表现得分之差是合理的，因为它能够全面地反映比赛中的竞争情况和局势变化，为比赛分析提供了有力支持。

待定：

用小波变换提取局势波动和连续成功的特征，分析趋势

求导数，如果导数趋于一致，说明二者有很大的相关性

画动量与结果的趋势图，

（3）画局势波动、连续成功与动量的折线图，判断是否随机 应为不随机 可视化分析

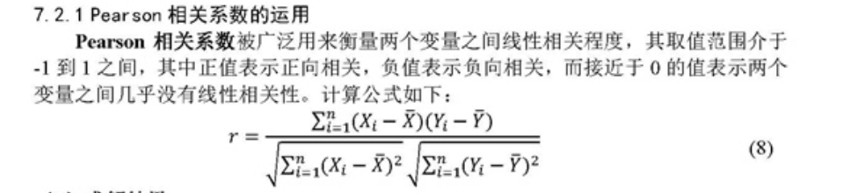
例如：动量高的时候，局势对我也是有利的，连续成功次数也是高的（仅供参考！！！请自己撰写）

2.建立回归模型判断关联关系

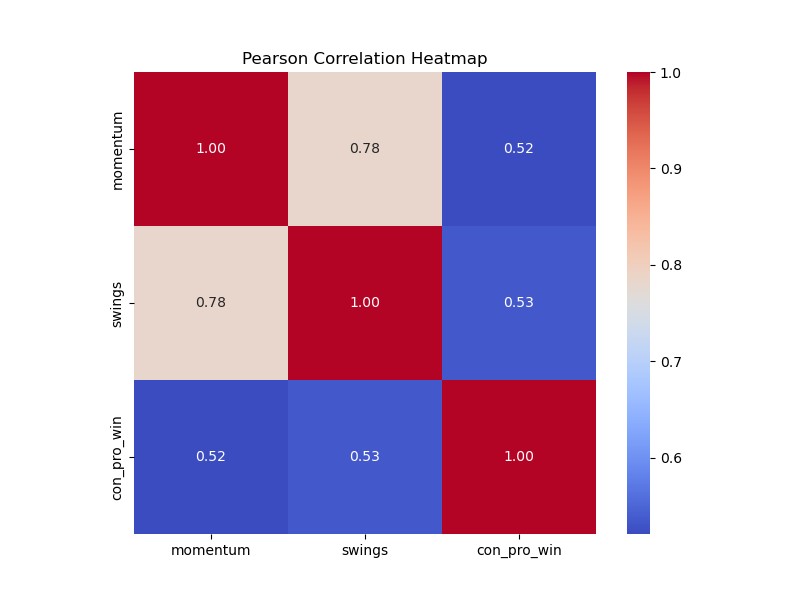
相关性分析 进行相关系数假设检验 进一步说明有关联

介绍pearson相关系数

参考：



（1 求解结果：



（2） Pearson相关系数检验

对局势波动、连续成功和动量进行pearson相关系数检验，检验步骤如下：

Step1：提出假设

原假设：pearson相关系数R≠0

备择假设：pearson相关系数R=0

设定置信水平为95%

Step2：计算P值

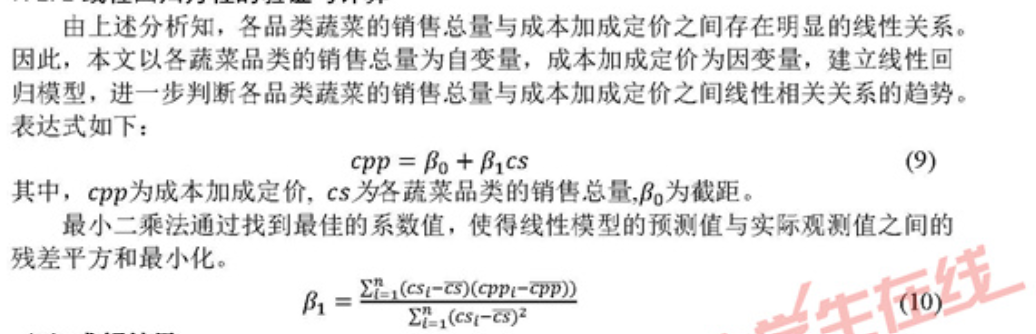
利用Python进行相关系数检验，结果如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | P值检验 | |  |
|  | Momentum | Swings | Con\_pro\_win |
| Momentum | 1 | 0.0 | 0.000594 |
| Swings | 0.0 | 1 | 0.000096 |
| Con\_pro\_win | 0.000594 | 0.000096 | 1 |

由表X可知，Swings和Con\_pro\_win的P值均小于置信水平005，拒绝原假设即Swings和Con\_pro\_win均与Momentum呈现显著的线性相关关系

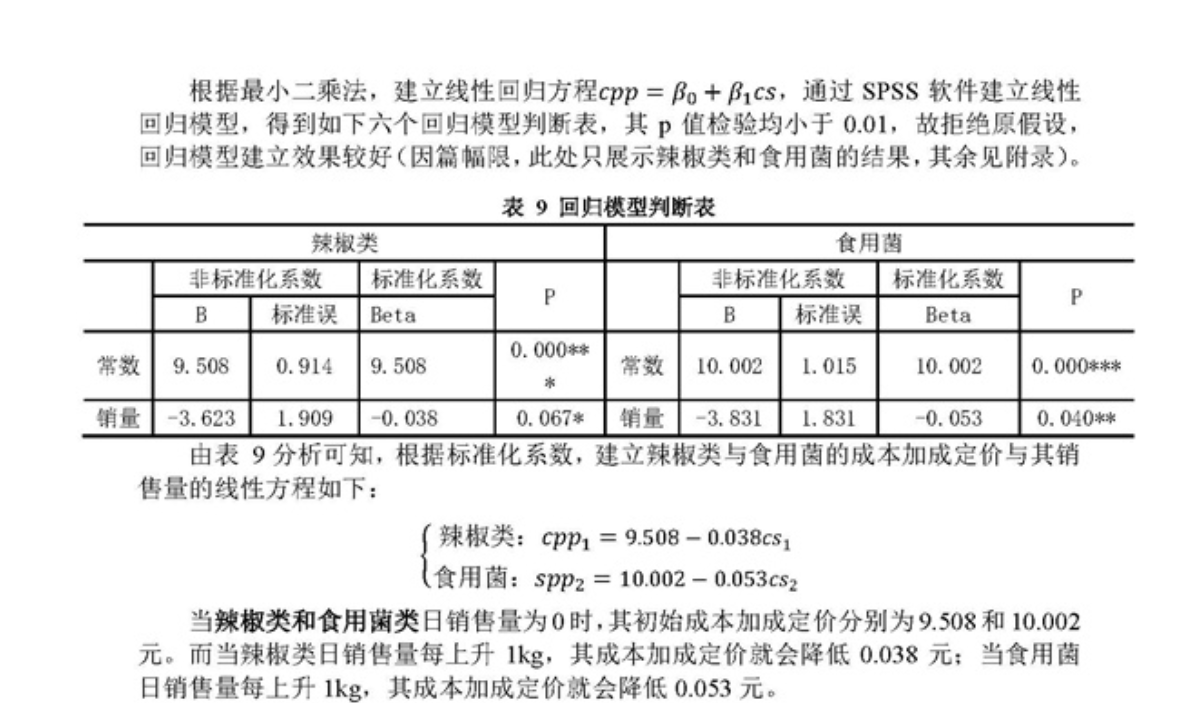
* 线性回归方程的验证与计算

由于Swings和Con\_pro\_win(考虑删掉，不是很显著)均与Momentum呈现显著的线性相关关系，因此可分别构建其线性回归方程

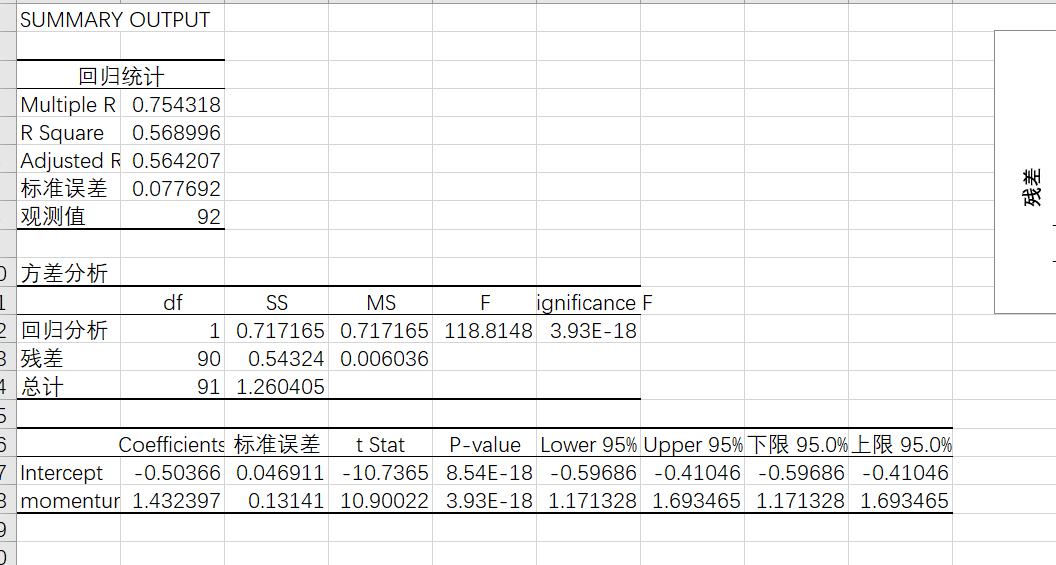


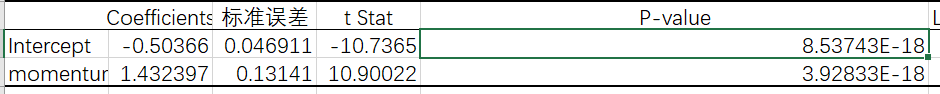
1. 求解结果

分析参考

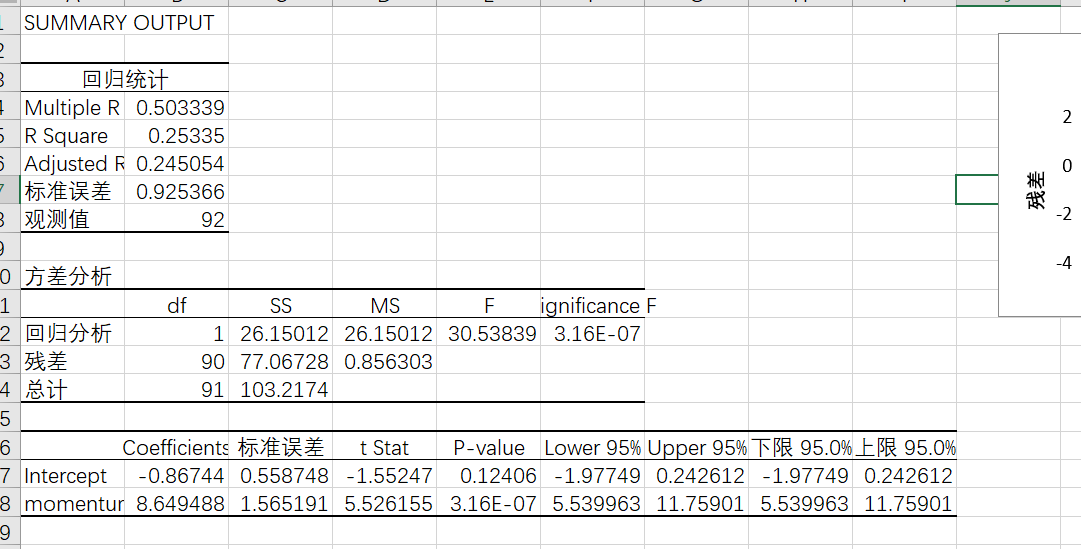


momentum和swings Y = 1.43X + -0.50





1. momentum和Con\_pro\_win



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Coefficients | 标准误差 | t Stat | P-value |
| Intercept | -0.86744 | 0.558748 | -1.55247 | 0.124060073 |
| momentum | 8.649488 | 1.565191 | 5.526155 | 3.15545E-07 |

问题三

（1）联系问题一

有问题2可知，局势波动是双方表现得分的差异，接下来我们将具体分析模型1中的每一个指标对局势波动的影响。

我们用问题一中的指标与问题二中的走势波动一起构建随机森林模型，根据指标对局势波动进行预测，并计算出变量的重要性，选出与走势最相关的因素

（2）问题二中已经证明动量与局势波动有关联，我们可以根据已有的动量差异，预测未来的动量差异，为一名选手在与不同选手的新比赛中提供建议，从而使局势对我方更加有利（自圆其说！）

* k-means聚类

首先，我们对动量的差异进行聚类，划分等级（为了根据不同的类别提供不同的建议）

对已有的7个指标差异值进行聚类分析，聚为不同类 比如4类，用1，2，3，4代替

【合理性】

对动量差异进行聚类并划分等级是一个合理的做法，因为它能够帮助我们理解和描述比赛中双方动量的差异程度，并据此制定相应的应对策略。以下是几点合理的理由：

1. \*\*提供直观的分析结果：\*\* 将动量差异分成几个等级，可以使结果更加直观和易于理解。这样的等级划分可以帮助教练和运动员更清晰地了解比赛中双方动量的差异情况。

2. \*\*辅助决策制定：\*\* 根据动量差异的等级，可以制定相应的应对策略。例如，当动量差异较大时，可能需要采取更积极的策略来应对；而当动量差异较小时，则可能需要更稳健的策略来保持优势或缩小差距。

3. \*\*定量化比赛表现：\*\* 动量差异的等级可以将比赛表现定量化，使得评估和比较不同比赛的结果更为方便。这有助于分析比赛结果，并从中获取更多的启示和经验。

4. \*\*指导训练和准备：\*\* 了解双方动量差异的等级可以为球员的训练和比赛准备提供指导。他们可以根据动量差异的等级调整训练计划和比赛策略，以更好地应对不同的比赛情况。

因此，对动量差异进行聚类并划分等级是一个合理且有益的分析方法，可以帮助球员和教练更好地理解比赛局势，并做出相应的决策。

对球员1来说，

双方动量差异==

动量差异受到，……的影响，因此本文基于影响动量差异的指标进行聚类

【采用指标进行聚类的合理性】

因为这些指标反映了比赛中双方表现的关键因素，可以帮助我们更全面地理解动量的形成和变化。以下是一些合理性的理由：

1. \*\*指标的重要性：\*\* ∆win\_points，∆lose\_points等指标直接反映了球员在比赛中的得分情况和失分情况。这些指标与比赛的胜负密切相关，也影响着比赛的动量变化。

2. \*\*全面性和代表性：\*\* 考虑到了多个影响动量的因素，包括得分、失分、赢得的局数、赢得的盘数等。通过综合考虑这些因素，可以更全面地描述比赛中双方的表现和动量差异。

3. \*\*可解释性和实用性：\*\* 这些指标具有很强的可解释性，能够清晰地反映出比赛中双方的得分情况和局势变化。基于这些指标进行聚类分析可以产生直观且易于理解的结果，为教练和球员提供有用的信息和指导。

4. \*\*数据可获得性：\*\* ∆win\_points，∆lose\_points等指标通常是从比赛记录中直接获取的，数据的获得相对容易。这意味着我们可以更方便地进行聚类分析，并且能够利用大量的实际比赛数据进行建模和分析。

因此，基于影响动量差异的指标进行聚类分析是合理的，可以帮助我们更好地理解和分析比赛中双方的动态变化，为球员和教练提供更有效的决策支持。

k-means聚类 原理+结果+分析

例如：可选一段时间出来分析（如一个sets）如果我比对方动量大，那我这局是否赢了，且是否连续得分高（如问题二中已分析，此处可一句话带过 ）

例如分为4类

则 1：动量比对方差很多

2：动量比对方差一点

1. 动量比对方好一点
2. 动量比对方好很多

* 然后用马尔可夫链（or 其他预测模型） 预测未来我与对手的动量差异等级

【使用马尔可夫预测的原因】：

马尔可夫链是一种基于状态转移概率的模型，可以用于预测系统在不同状态之间的转移情况。在这种情况下，我们将动量差异的不同等级视为状态，并利用历史数据来估计不同状态之间的转移概率。使用马尔可夫链进行动量预测的合理性体现在以下几个方面：

1. \*\*局势演变具有马尔可夫性质：\*\* 在网球比赛中，当前的局势演变通常受到前几个局势的影响，而且当前的局势转变往往只与前一个或几个局势的情况有关，符合马尔可夫性质。

2. \*\*历史数据反映了未来趋势：\*\* 马尔可夫链模型基于历史数据对未来的状态转移进行预测，而在网球比赛中，历史比赛数据包含了丰富的信息，可以用来估计不同动量状态之间的转移概率。

3. \*\*适用于稳定的系统：\*\* 如果系统的状态转移概率在短期内基本保持不变，那么马尔可夫链模型可以很好地预测未来的状态。在网球比赛中，虽然局势会因为球员的表现、策略等因素而发生变化，但在短期内通常不会发生剧烈的变化，因此马尔可夫链模型可以适用于这种情况。

综合来说，使用马尔可夫链进行动量预测在网球比赛中是合理的，因为它可以基于历史数据反映系统的状态转移情况，并利用这些信息来预测未来的局势演变趋势。

针对不同等级采取不同的战术，提供不同建议

例如：预测下一盘中好几局动量等级都处于1，教练则可以在这一盘结束后喊队员下来休息（具体需要了解规则）（这边只是举例，假设在盘结束后可以休息）

【建议参考】

当根据预测模型的结果发现动量将有较大波动时，我们可以制定如下建议策略：

1. \*\*调整战术\*\*：根据预测结果，分析动量的变化趋势，有针对性地调整战术。例如，在动量有利的时候，可以选择更进攻性的战术，积极主动地发起进攻；而在动量处于劣势时，则可以采取更保守的战术，稳健地守住局面。

2. \*\*加强心理素质\*\*：面对动量的波动，球员需要保持良好的心理素质，不受外界因素的影响。可以通过心理训练、专业心理咨询等方式，提高球员的心理抗压能力，保持冷静、沉着的心态。

3. \*\*调整发球策略\*\*：发球是影响比赛走势的重要因素之一，根据预测结果，可以调整发球策略。例如，在动量有利的情况下，可以选择更具攻击性的发球策略，以增加得分的机会；而在动量劣势时，则可以选择更为保守的发球策略，减少失误的可能性。

4. \*\*及时反应变化\*\*：根据模型预测的动量变化趋势，球员和教练团队需要及时反应，灵活调整战术和策略。保持对比赛的敏锐观察和分析，根据实际情况及时做出调整，以最大程度地利用动量的优势。

当预测结果显示动量波动较小时，我们可以建议球员继续保持稳定的表现，并保持当前的战术和策略不变。此时的重点在于稳定表现，保持比赛的掌控力，避免因过度调整而引起的不必要风险。同时，球员也应保持对局势的关注，随时准备应对可能的变化。

问题四

1. 在其他场比赛中测试开发模型

用问题三中已训练好的随机森林模型预测未来的局势波动并于实际数据比赛结果进行对比，说明预测效果

考虑

1. 未来因素 选手经验、水平 分析精度是否会提升
2. 普适性

女子

锦标赛

球场表面 参考：

敏感性分析

具体分析每一个指标变化对局势波动的影响，提出建议（待定）

随机森林 调参查的变化看MAPE（均方误差）

对于随机森林等模型，调整模型的超参数可能会对性能产生影响。通过调整树的数量、树的深度、每棵树的特征数量等参数，可以尝试改善模型的性能。