预测模型——思路

在前两问中，我们基于一些场上的指标，提出了momentum score模型，对于任意一场比赛的每个回合都可以计算出此回合结束后双方的momentum。如果我们想要知道一场比赛中的“转折点“——双方的形势发生逆转的回合，场上的实际情况如何，攻防双方都采取了什么策略从而导致转折，那么显然寻找两条momentum曲线的相交点是基于第一问的正常思路。但是这同时也带来了两个问题——

1. 第一问的模型是基于一些相对简单和符合直观认知的假设，在第三问中，教练员想要知晓比赛的转折点，这就要求我们的预测模型势必要包括更多与接发球类型相关的特征（'p1\_ace', 'p2\_ace'，'p1\_double\_fault', 'p2\_double\_fault'， serve\_width\_B', 'serve\_width\_BC', 'serve\_width\_BW', 'serve\_width\_C', 'serve\_width\_W', 'serve\_depth\_CTL', 'serve\_depth\_NCTL', 'return\_depth\_D', 'return\_depth\_ND'），这里就涉及到机器学习方法学习特征权重拟合momentum曲线。
2. 根据我们在第一问的模型假设，momentum score与最近的若干回合的表现相关，我们认为“momentum“是一个运动员兴奋度的体现，在一场高水平网球比赛中双方互有胜负，momentum的变化应当是频繁的。这导致了两条momentum 曲线交点过多。我们需要采取一些判断方法，排除掉那些”琐碎重复“的胶着点，把重点放在进入或离开胶着状态的那些回合。

预测模型——方法

关于第一个问题，我们使用随机森林回归器对数据进行训练，这些数据经过第一问的momentum score模型计算过，并将momentum score当作监督学习过程中关键的标签。使用R² score 作为评估指标，并观察训练后特征的重要性，为避免模型过拟合，增强预测模型泛化能力，需要挑选主要的特征再次继续训练。

关于第二个问题，我们先求得在前后变化为不小于一定阈值的所有交点，然后再挑出一定邻域内没有其他交点的交点，即那些前若干回合或后若干回合不再有momentum交叉的回合。找到这些转折点，并结合场上的具体情况具体分析预测的合理性。

预测模型——结果

下图为初次训练后特征重要性的排名：

图表

描述已自动生成

再次训练，得到实际曲线和拟合曲线

图表

描述已自动生成

寻找转折点，分析场上情形

以比赛编号为1302的这场比赛为例，对战双方分别是Alexander Zverev（P1） 和 Matteo Berrettini（P2）

图表

描述已自动生成

1. Intersection 1: (x=26, y=0.53) - Change in Momentum 1: 0.65, Change in Momentum 2: -0.71

大比分——0：0

小比分——1：2

当前——40: 15（1发球）

拍数：1

跑动距离：5.6m-6.9m

发球速度：133mph

结果：1一发保发成功，小比分2：2

模型预测：此球后，2优1劣->胶着

1. Intersection 2: (x=44, y=0.64) - Change in Momentum 1: -0.47, Change in Momentum 2: 0.64

大比分——0：0

小比分——3：4

当前——15：30（1发球）

拍数：2

跑动距离：7.2m-8.8m

发球速度：112mph

结果：2二发得分，1非受迫性失误，15：40

模型预测：此球后，2优势进一步扩大

1. Intersection 3: (x=64, y=0.72) - Change in Momentum 1: 0.36, Change in Momentum 2: -0.44

大比分——0：1

小比分——1：1

当前——40：0（1发球）

拍数：1

跑动距离：0.6m-0.9m

发球速度：126mph

结果：1一发ace保发成功，小比分2：1

模型预测：此球后，1一转颓势，重新将比赛打入胶着状态

1. Intersection 4: (x=112, y=0.58) - Change in Momentum 1: -0.81, Change in Momentum 2: 0.82

大比分——0：1

小比分——6：6

当前——15：40（1发球）

拍数：1

跑动距离：0.8m-0.9m

发球速度：134mph

结果：1一发ace得分，30：40

模型预测：此球后，2再次优势

1. Intersection 5: (x=135, y=0.73) - Change in Momentum 1: 2.16, Change in Momentum 2: -3.30

大比分——0：2

小比分——1：1

当前——30：40（2发球）

拍数：5

跑动距离：26.5m-16.1m

发球速度：127mph

结果：2一发保发成功，小比分1：2

模型预测：此球后，2重新将比赛拉入胶着

尽管对于个别点的momentum增减存在歧义，但是结合mementum曲线的趋势和场上实际事件来看，我们认为预测模型比较好的找到了影响比赛momentum的那些关键转折点。找到这些转折点后，我们只需要将这些转折点单独拿出，分析接发球的类型、对拍数、跑动距离等实际比赛策略偏好。

When Player1 plays server:

Coefficient of features:

Feature Coefficient

0 p1\_sets -0.660716

1 p2\_sets 1.079027

2 p1\_games 0.523895

3 p2\_games -0.461889

4 p1\_double\_fault 0.360961

5 p1\_unf\_err -0.310627

6 p2\_unf\_err -0.924556

7 rally\_count 0.404886

8 speed\_mph -0.641765

9 serve\_width\_B -0.342782

10 serve\_width\_BC -0.237314

11 serve\_width\_BW -0.029538

12 serve\_width\_C 0.361789

13 serve\_width\_W 0.111184

14 serve\_depth\_CTL -0.393256

15 serve\_depth\_NCTL 0.391367

16 distance\_diff -0.314234

17 p2\_distance\_run -0.570710

18 p1\_distance\_run -0.515582

Accuracy of model: 0.75

When Player1 plays servee:

Coefficient of features:

Feature Coefficient

0 p1\_sets 0.124868

1 p2\_sets -0.149697

2 p1\_games 0.519961

3 p2\_games -0.155793

4 p2\_double\_fault 0.643224

5 p1\_unf\_err -0.136075

6 p2\_unf\_err -0.087853

7 rally\_count -0.305280

8 speed\_mph 0.205257

9 return\_depth\_D 0.259512

10 return\_depth\_ND 0.372136

11 distance\_diff -0.093827

12 p2\_distance\_run 0.399639

13 p1\_distance\_run 0.333224

Accuracy of model: 0.782608695652174

预测模型——建议