

2024美赛论文初稿

一、捕捉比赛流程

问题内容

开发一个能够捕捉比赛中随着得分发生的比赛流程的模型，并将其应用于一个或多个比赛。你的模型应该能够识别在比赛中的某个时刻哪位球员表现更好，以及他们表现得有多好。提供一个基于你的模型的可视化，来描述比赛流程。注意：在网球中，发球方赢得得分/局的概率要高得多。你可能希望以某种方式将这一因素纳入你的模型中。

可视化模型的确定

定义一场比赛中两个球员在某时刻的“势能”量化为：

$$Q(i, t)$$

其中*i*取值为1或2，*t*为连续的时间值，我们考虑对一场完整的比赛中对两位球员的势能各自独立的进行量化，并使用两者的差值来具体体现某个时刻哪位球员表现的更好，以及表现的多好：

$$\Delta Q(t) = Q(1, t) - Q(2, t)$$

绘制 ΔQ 关于*t*的图像，用来反映比赛过程中势能的转化，并以可视化的形式体现。

这里比赛的数据是离散化的出现，但我们最终给出连续化的函数图像。

势能量化的确定

考虑每个球员在每个得分点的势能所构成的序列为连续的时间序列，该时间序列的每一项都依赖于前一项和当前项有关因素的集合

$$Q(i, t) = Q(i, t - 1) + \sum_{k=1}^n f(\theta_k)$$

其中*f*是对当前项有关因素对势能量化贡献值的衡量函数

θ_k 为对当前项有关因素，此处考虑：

发球权

（此得分点之前）连续得分或失分

此得分点的得分或失分

此得分点发球方的失误次数

此得分点分值差距

此得分点是否为局点、盘点（此点结合发球权）（破发点）

此得分点是否为发球直接得分

此得分点球员体力 $H(i, t)$

其中体力依赖于：

前一时刻的体力、此得分点的跑动距离、此得分点的击球次数、此得分点的击球速度

此外，考虑特殊得分点，如抢七抢十。

考虑对所有因素的系数带入某类机器学习进行训练，得到有效的系数集合。注意要将离散化转为连续化（函数拟合）

二、判断势能有无

考虑使用蒙特卡洛方法进行随机模拟，观察是否与实际情况相吻合

考虑选手实力（以世界排名为标准）、发球权（发球放有更高的概率得分）

三、判断势能转换

考虑哪些因素影响势能的转换，并预测势能的转换

前面我们已经成功模拟了比赛流程中的势能量化，其中 $\Delta Q = 0$ 时即发生势能转换，此处注意，势能转换具有方向性。

此处不难认为，加权系数大的因素，对势能的波动更具有影响力。

前面我们是在已知该得分点的全部数据下，计算出该得分点时的球员势能，现在要求我们对此势能波动进行预测，此处考虑是仅对势能转换点进行预测，还是全部波动进行预测。

前者为二分类问题，可以采用随机森林、微分方程方法

后者难度较大，考虑预测后与真实值进行校对后进行下一阶段的预测

四、模型检验与泛化

建立检验标准来量化预测准确程度

进行迁移学习来检验泛化能力

考虑不同数据集应当更改、添加和删除的关键影响因素