



数字访问
学者病理学

DASH. HARVARD. EDU



HARVARD LIBRARY
Office for Scholarly Communication

从任何得分中产生职业网球比赛的获胜概率

引用

Gollub, Jacob. 2017. Producing 赢得任何分数职业网球比赛的概率。学士论文，哈佛学院。

永久连接

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41024787>

使用条款

本文从哈佛大学的DASH存储库下载，并在适用于其他已发布材料的条款和条件下提供，详见
[http://](http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA)

nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

分享你的故事

哈佛社区已经公开了这篇文章。

请分享此访问对您的好处。提交一个 故事

[可访问性](#)

内容

12Hcod TefsFsal	
gMyPat 0	
ung	
2 dr	
22个疯狂	
24 ddngatdeie	
4 M	
25 WhPat和Eyalan	
皮格斯数学奥德顿	
1.10reviam	
80kRma	
33ATP端。	
34 Pait ensad Msit	
35 JmusSain Esimute	
36OppaealLess 西纳夫劳维	
37 Rads	

14464

41.1 Closs Vaic

福姆

al M kew Nosell.

3.1 BeE m

43.2 Bolndond Sve Picab

43.3 hpotnce*单

4.1 Modd

44.2 Byset

44.3 Bj Surce

45 Wadzag WaPotakay

1 就像波邦克

52F

5. 3美国大圣马切斯市

5. 3. 1是费德勒2017年的冠军

2015年532年

533 Noak Dphove vs. 2011年美国潘汽车大赛

27

-2

-

-3

-3

-41

44

1. 介绍

1.1 游戏内体育分析

“ESPN上的获胜概率图表/讨论简直是拿起一把剑，把它塞进棒球中剩下的乐趣。”

—KennyDucey (@KennyDucey) 2017年4月2日

“ESPN显示获胜的概率非常好。接下来，运行预期！”

—尼尔·温伯格 (@尼尔·温伯格44日) 2017年4月4日

“感谢上帝，我们有ESPN的获胜概率数据来告诉我们，领先的球队有很好的机会获胜。”

—克里斯蒂安·施耐德 (@施耐德·CM) 2017年4月17日

近年来，获胜概率在体育广播中越来越普遍。美国橄榄球联盟评论员布莱恩·伯克在他的网站上发布了过去几年NFL季后赛的获胜概率图表（Burke，2014）。今年早些时候，ESPN开始取得胜利在美国职业棒球大联盟的电视比赛中，处于得分框之上的概率（斯坦伯格，2017）。尽管球迷的反应不一，如上所示，这些发展代表了体育广播现代叙事的转变。

任何比分交流的获胜概率都表明一个球队或球员更容易获胜。虽然人们可以从任何选择的模型中产生这一点，但体育分析的人努力产生最明智的估计。随着最近在线博彩的激增，在比赛中获胜的概率决定了它自己的整个市场。贝特费尔的平台

在2011年法网公开赛决赛中匹配了超过4000万欧元（Huangetal.，2011）；其他高调的比赛通常会吸引相当的交易量。虽然大多数网球预测论文都承认赛前预测，但大约80%的在线投注发生在比赛进行的时候（Sipko和诺顿贝特，2015）。至少是，在比赛中获胜概率考虑到这个波动市场的所有参与者。

1.2网球预测的历史

统计预测模型应用于网球比赛已经有20多年了。克拉森和马格纳斯检验了网球比赛中的积分是独立的假设

卡利分布，或id（克拉森和Magnus，2001）。虽然这个假设是错误的，但他们

得出结论，偏差足够小，常数概率提供了一个合理的

近似在此假设下，他们构造了一个层次马尔可夫模型

与网球界的得分系统相结合。从这个模型中，他们提供了一个解析方程

来自任何分数的比赛获胜概率，给定每个玩家的获胜概率

发球得分（克拉森和马格努斯，2003年）。巴米特和克拉克然后提供了一种方法

从历史数据中估计每个球员的发球概率（巴尼特和克拉克，2005年）。

多年后，Bevc提出用beta分布来更新这些服务概率

用上述模型计算相应的获胜概率

(Bevc, 2015). 最近，科瓦尔奇克评估了11种不同的赛前预测的性能

2014年ATP巡回赛级别比赛的模型。虽然基于市场的庄家共识模型总体上表现最好，但Elo评级在依赖历史数据的方法中最为有效（Kovalchik，2016）。

在过去的几年里，杰夫·萨克曼发布了最大的公开图书馆

网球数据集通过github。这个集合包含了开放时代每一场ATP和WTA比赛的比赛摘要，近10万巡回赛水平和卫星的点点摘要

比赛，以及一个超过3000场比赛的众包比赛图表项目，志愿者记录整个比赛的类型和方向。虽然538和科瓦尔奇克使用杰夫·萨克曼的匹配数据来生成Elo评级，但上述的论文没有一篇用他的逐点数据集测试模型。

由于过去的论文跨越了几十年，数据集和模型评价并不一致。克拉森和马格纳斯最初的基于点的模型和贝夫克的贝塔实验都适用

从1991-94年到大约500场温布尔登比赛。巴奈特深入地探索了基于点的模型，但主要是将它们应用于安迪·罗迪克之间的一场马拉松比赛

以及2002年澳大利亚网球公开赛上的尤尼斯·艾纳维（Bamettetal.，2006）。除了Bevc之外，这些论文中没有一篇使用准确性或交叉熵等指标来评估游戏内的匹配预测。虽然贝夫克在500场温布尔登比赛中创造了准确的记录，但他只是

报告第三组预测的准确性，整个数据集的子集。换句话说，没有人采用所有的游戏内预测模型并进行大规模的测试。出于这个原因，我将在过去七年内的数千场比赛中测试所有相关的匹配预测方法。

本文结合Elo评级，丰富的数据和当前的技术，提供类似的调查，匹配预测方法表现最好。我们建立在过去的研究基础上，通过测试以前最先进的方法的变化，并将新的概念应用于这些数据集，从足球中使用的随机森林模型到橄榄球诱导的服务百分比。

1.3逐匹配数据集

这个项目使用了两种不同类型的数据集：一种是匹配的汇总统计数据，另一种是逐点信息。两者都可以在github上公开获得，由Jeff Sack-提供

曼(<https://github.com/JeffSackmann>)。我们使用“网球atp”下的匹配数据集来测试预匹配的预测方法。这个数据集涵盖了超过15万场巡回赛级别的比赛，dat-

回到1968年。相关功能包括：

- 比赛信息（日期、比赛日期、球员姓名）
- 比赛统计（每个球员的发球/回收比赛统计）

“网球点点”中的数据提供了关于单个比赛点的更详细的细节

连续每场比赛都包含一个字符串，上面列出了球员赢得积分和交换发球的顺序。虽然atpworldtour.com并没有公开列出逐点排列的摘要，

杰夫·萨克曼已经从一个博彩网站上提供了这些信息。与匹配数据集不同的是，这个数据集涵

盖了自2010年以来的大约12,000场巡回赛级的比赛。具有相关性的

功能包括：

- 比赛信息（日期、比赛日期、球员姓名）
- 逐点计算（详细说明得分顺序的字符串）

1.4实施

由于上述数据集作为本项目的基础，需要对数据进行彻底的重新格式化，以便将逐点字符串连接到匹配数据集中它们相应的观察结果。由于这两个数据集都包括玩家的姓名、年份和分数，我们联系起来

使用哈希方案跨数据集进行匹配。¹由于观察到的不一致性，我们将比赛和逐点数据集之间的球员名字进行了标准化，以最大化

具有逐点摘要的可用匹配项的数量。虽然部分点-逐点的总结来自卫星事件，类似于网球小联盟，但我们能够做到

以大约10600/12000的逐点串进行各自的巡回赛比赛。

然后，我们可以将匹配数据集的整体（Elo排序、年调整的服务统计数据等）生成的信息与相应的逐点字符串相结合。

要更深入地查看此过程或访问任何结果数据集，请访问<https://>

1g. 哈希（manchX）= “罗杰·费德勒托马斯·伯蒂奇2012年3-67-5
7-5”2eg.”斯坦·瓦林卡，斯坦尼斯拉斯·瓦林卡，费德里科·德尔博尼斯

github.com/igollub1/tennis_match_prediction .这其中每个方法的实现
本文还提供了项目3，以及关于如何生成所有相关统计数据的说明。

³本项目中涉及的所有方法都需要在两个数据集上进行特征构建。

2. 评分

网球的得分系统由三个层次组成：套装、比赛和积分。考虑一场网球比赛

在两个实体之间， p_i 和 p_j 。我们可以将任何分数表示为 $(s_i, s_j, g_i, g_j, \text{服务于 } x_i, x_j)$ ，其中 p_i 是

s_k, g_k, x_k 分别代表 p_k 在集合、游戏和得分中的得分。1岁的球员

每场比赛都替补发球，直到有人赢得两盘锁定比赛

（三局三胜制）或三组（五局三胜制）。2

大多数的网球模型利用体现层次的图形结构

在网球得分系统内。马杜尔斯卡将其称为一个层次化的马尔可夫模型

(Madurska, 2012)。巴奈特正式定义了网球分数的表示

等人，2002年)。 p_i 和 p_j 在发球时获胜，概率为 f_{ij}, f_{ji} ，每个比赛中

比分 $(s_i, s_j, g_i, g_j, x_i, x_j)$ 进展到它的两个邻居 $(s_i, s_j, g_i, g_j, x_i+1, x_j)$ 和 $(s_i, s_j,$

$g_i, g_i, g_i, x_i, x_j+1)$ 中的一个，其转移概率依赖于当前服务器。假设匹配中的所有点都是

ii. d. between服务器，然后我们可以使用下面的模型进行递归化

确定获胜概率。看看下面的记分牌，拉斐尔·纳达尔为胡安

马丁·德尔·波特罗7-63-66-62-1，在2011年温网第四轮。表示分数

从… 那 服务器的 观点 的 视图，我们 发现

- 1网球正式将比赛的前三分指分为15、30、40，为了简单起见，我们称它们为1、2、3。
- 2最好的模式通常用于男子大满贯和戴维斯杯赛事。



图2.1：纳达尔在第三盘比赛中送达德尔波特罗

$P_m(1, 1, 6, 6, 2, 1) = P(\text{Pi赢得比赛, 从这个比分开始发球})$

$P_m(1, 1, 6, 6, 2, 1) = f_{iy}P_m(1, 1, 6, 6, 3, 1) + (1 - f_{ij})P_m(1, 1, 6, 6, 2, 2)$

在下面的部分中，我们将为层次模型的每个级别指定边界值。

2.1 建模游戏

在一场比赛中， p_i 或 p_j 发球每一场比赛从 $(0, 0)$ 开始，要赢得一场比赛，玩家必须以至少2分的优势赢得4分或4分以上。因此，所有具有有效分数 (x_i, x_j) 的游戏，其中 $x_i + x_j > 6$ ， $k_{iy} x_{jly} 1$ 被简化为 $(3, 3)$ 、 $(3, 2)$ 或 $(2, 3)$ 。此外，在 $(3, 3)$ 处还可以直接计算出它的获胜概率。从 $(3, 3)$ 开始，

服务器赢得概率 (i, j) 2. 翻新者赢得了接下来的两场比赛

概率 $(1 - f_{ij})/2$ 。或者两个玩家将两分分开，分成 $(3, 3)$

概率为 $2f_{ij}(1 - f_{ij})$ 。将游戏的余数与一个几何序列联系起来，我们发现

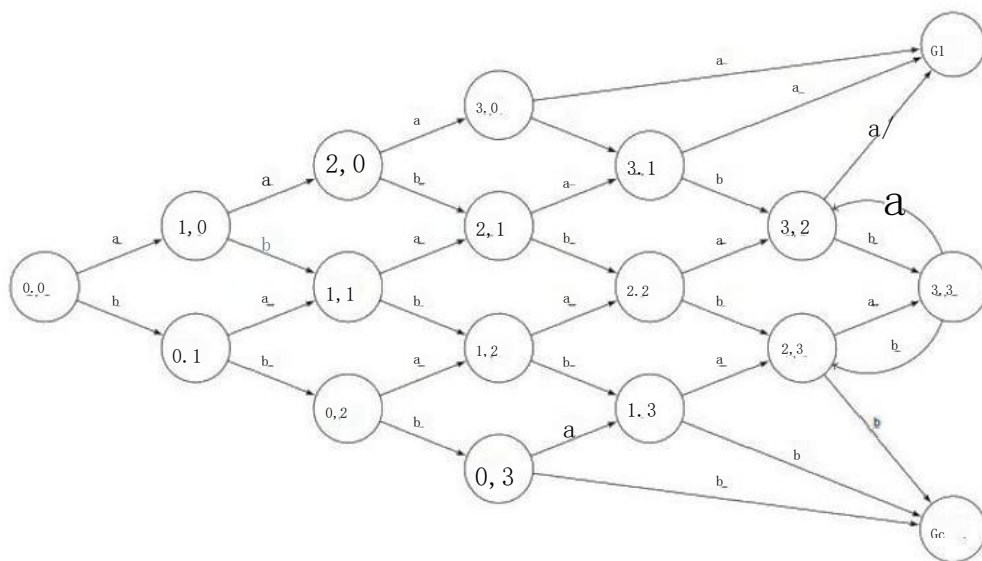
. 因为每个点都有6r

这两个玩家，这个图假设在一个游戏中的分数是独立的。在下面的关卡中，我们假设集合内的游戏和匹配内的集合都是独立的。

在一个游戏中可能出现的分数序列：

$a - p_i$ 将赢得以下一分

$b - p_j$ 赢得了以下一分



边界值:

$$Pg\left(x_i,x_j\right)=\frac{\left(f_{ij}\right)^2}{\left(sacrifice\right)^2\left(1-y_{ij}\right)^2},$$

根据上述说明，我们可以有效地计算出任何恐慌的胜利

Pg (xi, 对) 。

2.2 建模集

每一场比赛， p_i 或 p_j 交替发球。每一组都从(0,0)开始。要赢得一场比赛，a

玩家必须拥有至少2场以上的6场或更多的游戏。如果达到了设定的分数（6、6）。

进行一场决胜局来决定比赛的结果。

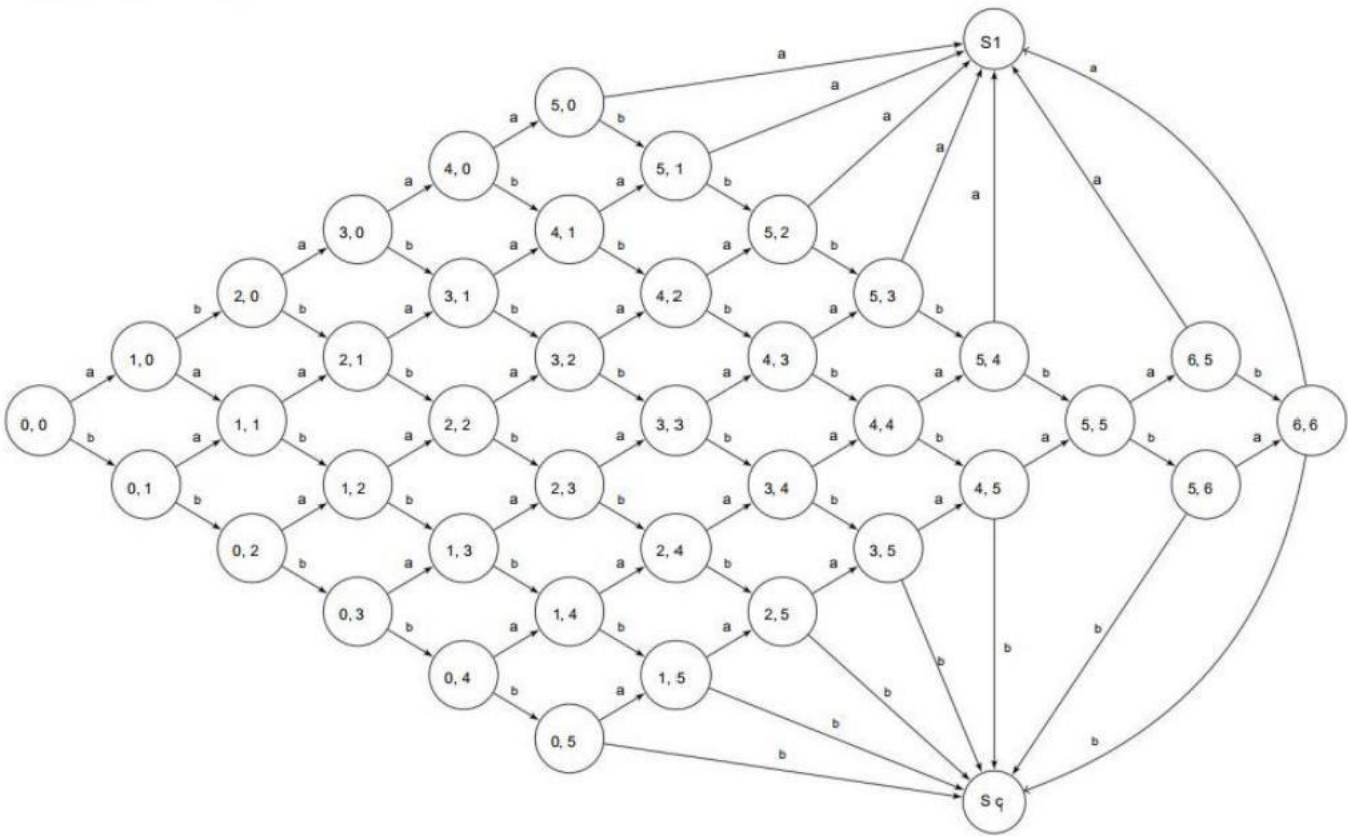
在一个游戏中可能出现的分数序列：

a-玩家1赢得了假的游戏

b-玩家2是跟随游戏

a-玩家1赢得了老虎决胜局的游戏

b-玩家2破坏破坏游戏



边界值：

$$P_s(g_1, g_2) = \begin{cases} 0, & \text{if } g_1 \leq 6, g_2 \leq 6 \\ P_{tb}(s_1, s_2), & \text{if } g_1 = g_2 = 6 \\ P_g(0, 0) \times P_s(g_2, g_1 + 1) + (1 - P_g(0, 0)) \times P_s(g_2 + 1, g_1), & \text{否则} \end{cases}$$

(2.2)

2.3模拟一个抢七游戏

在抢七比赛中，一名选手发球，每两个选手轮流发球
从那时起。游戏从（0,0）开始，玩家必须以7分的优势获胜
两个或更多的人来赢。

边界值：

$$P_t(p_1, p_2) = \begin{cases} 0, & \text{if } p_1 \leq 7, p_2 \leq 7 \\ P_{fj}(p_1 + 1, p_2) + (1 - P_{fj})P_t(p_1, p_2), & \text{if } (p_1 + p_2) \bmod 2 = 1 \\ (1 - P_{fj})P_t(p_1 + 1, p_2) + P_{fj}P_t(p_1, p_2 + 1), & \text{if } (p_1 + p_2) \bmod 2 = 0 \end{cases}$$

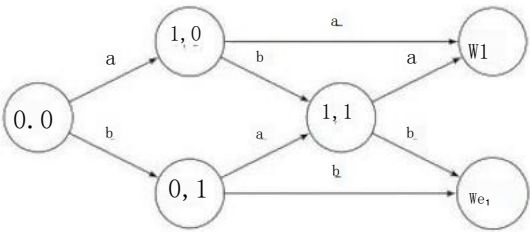
(2.3)

2.4建模一场三局三胜制的比赛

在三局三胜的比赛中，第一个赢两盘的选手拿下比赛。本项目
主要将模型应用于三局三胜的比赛。在最佳比赛中，图的边界值为s1j 3，s2j 3为3，因为玩家必须赢得三局
才能完成
比赛

a-玩家1赢得以下一盘

b级选手2将赢得以下一盘比赛



边界值:

ifslg 2

$$P_m(s_1, s_2) = \begin{cases} 0, & \text{if } s_1 \geq 24 \\ P_s(0, 0)P_m(s_1 + 1, s_2) + (1 - P_s(0, 0))P_m(s_1, s_2 + 1), & \text{otherwise} \end{cases}$$

2.5赢概率方程

结合上述等量，我们可以递归地计算出细化时的赢概率

自 (si、可、gi、gi、xi、xj) 为:

$$P_m(s_i, s_j, g_i, g_j, x_i, x_j) = P_m(s_i, s_j, g_i, g_j, x_i, x_j + 1) + (1 - P_s(s_i, s_j)) P_m(s_i, s_j, g_i, g_j, x_i, x_j)$$

我们会递归地调用自己

除非 (xi、x) 满足边界条件，否则Pm (si、s)、gi、gi、

关于这两种可能的跟踪分数。因为我们可以计算Pg (3, 3)，所以我们可以最终表示

每个重复概率的边界值，这允许我们计算概率-

赢得目前的比赛的幸福。通过遍历集合和匹配相似的模型

通过递归的方式，我们最终可以确定圆周率赢得比赛的可能性。

3. 赛前比赛预测

3.1 概述

在比赛开始之前，比赛内预测模型不能从比赛本身获取信息。然后，在 p_i 和 p_j 开始匹配之前，最明智的匹配前预测 $y_{ij}(t)$ 应该作为匹配内预测的基础。因此，我们首先探索匹配前模型。

今年早些时候，科瓦尔奇克发布了一项对11种不同的赛前预测模型的调查，并排评估它们的不准确性、日志损失、校准和鉴别。FiveThir-

基于elo的轮胎灯模型和庄家共识模型（BCM）表现最好。基于elo的预测包含了 p_i 和 p_j 的整个比赛历史，而BCM模型包含了在各种博彩市场中编码的信息。在本节中，我们将探讨以下方法和对她的调查中最成功的模型的变体：

- Elo评级
- ATP等级
- 基于点的模型
- 詹姆斯-斯坦估计器
- 竞争对手-调整后的服务/回报统计数据

3.2 Elo评级

Elo最初是作为一个为国际象棋玩家开发的头对头的评级系统开发的（Elo，1978）。

最近，5338的Elo变体在媒体中获得了突出的地位（Bialik et al.，2016）。对于时间为 p_i 和 p_j 之间的 t 匹配，Elo评级为 $E_i(t)$ 和 $E_j(t)$ ， p_i 为

预测将获胜的可能性为：

$$\hat{y}_{ij}(t) = \frac{1}{1 + 10^{\frac{E_j(t) - E_i(t)}{400}}}$$

p_i 的评级为以下匹配 $t+1$ ，然后相应地更新：

$$E_i(t+1) = E_i(t) + K(t)(y_{ij}(t) - \hat{y}_{ij}(t))$$

$W_i(t)$ 是 p_i 是否赢得了给定匹配的指标，而 $K_i(t)$ 是 p_i 在 t 时刻的学习率。根据538的分析师，当允许 $K_i(t)$ 随着时间的推移缓慢衰减时，Elo评级表现最佳（Bialik et al.，2016）。用 $m_i(t)$ 代表 p_i 的职业生涯比赛，我们更新了我们的学习率：

$$K_i(t) = \frac{250}{(5 + m_i(t))^{.4}}$$

当我们没有关于玩家的信息，并且随着 $m_i(t)$ 的积累时，这个变体会最快地更新玩家的Elo。将此Elo评级方法应用于我们的

数据集，我们初始化每个玩家的Elo评级在 $E_i(0)=1500$ 和匹配历史 $m_i(0)=0$ 。

然后，我们按时间顺序遍历从1968-2017年的所有巡回赛级别的比赛，为每场比赛存储 $E_i(t)$ 、 $E_j(t)$ ，并相应地更新每个玩家的Elo。¹

3.3 ATP等级

当克拉森将ATP等级纳入他的预测模型时（克拉森和马格努斯，2003），科瓦尔奇克和538一致认为Elo的表现优于基于排名的方法（科瓦尔奇克，

2016）。根据2010年至今的ATP匹配数据，我们发现：

¹“网球”公开赛时代始于1968年，当时职业选手被允许进入大满贯赛事。在此之前，只有业余选手参加这些项目

方法一致	acy ATP
等级66.5%	andard elo
69.0 %	

在每个评级系统中，我们预测排名较高的玩家将是获胜者。准确度代表憎恨被正确预测的匹配项的比例。因为ATP不提供球迷a相应的赢概率方程和他们的排名系统，我们不能计算-减少日志损失或校准方面的方法。不过，考虑到它优越的准确性近年来，本文中的模型使用Elo评级来代表玩家的能力代替官方的旅游排名。

3.4基于点模型

层次马尔可夫模型提供了一个解析解，以赢得概率 “yij(t)之间 玩家pi和pj，给定发球概率fij.fji。这取决于马尔可夫假设 到任何状态的转移概率只依赖于当前状态。在网球比赛中 意味着所有的点都是独立的，赢得一个给定的点的概率只取决于 在当前服务器上。虽然这是违反直觉的，但克拉森和马格纳斯证明了这一点 偏离i.i.d. 匹配范围内的假设足够小，足以合理地证明 基于点的模型（Klaassen和Magnus，2001）。后来，巴奈特和克拉克展示了一个 方法估计fij，fji，给定球员的历史发球/回报平均数(巴奈特和克拉克， 2005)。

fij=ft+(fi y fav)y (gi y gav)

fi=ft+(fy fav)g (gi y gav)

每个球员的发球率是他们自己的发球能力和对手的函数 返回能力。ft表示比赛的平均发球率， 而fi，fj和gi，gj代表pi和pj在发球和回传上的得分比例， 分别。喜欢，gav是发球和回复率。因为所有的点 由服务器或返回，喜爱=lygav。通过结合能力相对的差异

对于平均值，这个公式假设它们的影响是相加的。

根据Bamett和Clarke的公式，我们使用前一年的比赛服务统计数据来计算给定比赛和年份的 f_t ，其中 (w, y) 表示在年份 w 的比赛中进行的所有比赛的集合。

$$f_t(w, y) = \frac{k_y(w, y-1) \text{ \# of points won on serve in match } k}{k_y(w, y) \text{ \# of points played in match } k}$$

在他们的论文中，巴米特和克拉克只将这种方法应用于一场比赛：罗迪克vs。

2003年澳大利亚网球公开赛。此外，他们计算服务和服务的能力

百分比受到atpworldtour.com提供的总统计数据限制。也就是说，他们只能使用今年迄今为止的服务和返回的统计数据来计算 f_i 、 g_i 、 f_j 、 g_j 。由于统计数据没有列出相应的样本量，他们必须假设每一次给予3次匹配持续165分，这增加了估计玩家能力的另一层不确定性。

用今年迄今的统计数据实现这种方法很麻烦，因为 f_i 、 g_i 删除

随着 p_i 率全年累积匹配，不确定性。由于可用性

届时，9月份的预测将比1月份更可靠。然而，有了我们的巡回比赛水平的比赛数据集，我们可以在任何时间点为每个球员保存一个为期一年的发球/回报统计数据。式中 (p_i, y, m) 表示 p_i 在年 y 、月 m 中的匹配项的集合，我们得到以下统计数据：2

$$f_i(y, m) = \frac{\sum_{k=1}^{12} k_y(y-1, m+t) \text{ \# of points won on serve by } p_i \text{ in match } k}{\sum_{k=1}^{12} k_y(y-1, m+t) \text{ \# of points played on serve by } p_i \text{ in match } k}$$

$$g_i(y, m) = \frac{\sum_{k=1}^{12} k_y(y-1, m+t) \text{ \# of points won on return by } p_i \text{ in match } k}{\sum_{k=1}^{12} k_y(y-1, m+t) \text{ \# of points played on return by } p_i \text{ in match } k}$$

为了保持此格式的一致，我们还计算了 f_{av} 、 g_{av} ，其中 (y, m) 表示 y 年、 m 个月的巡回赛级别比赛：

$$f_{av}(y, m) = \frac{\sum_{t=1}^{12} k_y(y-1, m+t) \text{ \# of points won on serve in match } k}{\sum_{t=1}^{12} k_y(y-1, m+t) \text{ \# of points played in match } k} = 1 - g_{av}(y, m)$$

现在， f_i 、 g_i 的变化不再依赖于一年中的时间。因为他们赢得了自己的分数

每场比赛都记录发球数，我们也知道球员的发球数/回球得分

演奏下面，我们使结合演员■

2013年，凯文·安德森（RSA）和费尔南多·沃达斯科（ESP）之间的第三轮比赛
澳大利亚公开赛。

球员的名字	s分赢得了s points 3292		r分赢得了r分ts 1726	gi
凯文·安德森	484225723981	是	49621560411	.3478
弗达斯科		.6799.6461		.3795

从2012年澳大利亚网球公开赛的统计数据，ft=.6153。从2010-2017年的比赛水平数据来看，
fav=0.6468；gav=lyfav=3532使用上述02/12-01月13日的服务/返回统计数据，
我们可以计算：

$$f=f+(fiy fav)y (giygav)=.6153+(.6799y.6468)g (3795y.3532)=.6221$$
$$fji=ft+(fy fav)y (giygav)=.6153+(.6461y.6468)y (.3478y.3532)=.6199$$

有了以上的发球率，凯文·安德森就有希望赢得五胜最强的比赛
与概率Mp (0,0,0,0,0,0) =.5139

3.5詹姆斯-斯坦估计师

几十年前，埃弗隆和莫里斯提出了一种方法来估计样本均值组 (Efron
和莫里斯，1977年)。詹姆斯-斯坦估计器将样本均值缩小到总体均值，
收缩与估计器的方差成正比。不管y的值如何，这个方法
产生的期望结果优于极大似然估计。

为了估计那些不经常参加巡回比赛的球员的发球/发球参数，
fi，gi必须根据有限的样本量来计算。因此，这是基于匹配概率的
偏离这些估计值可能会受到噪音的影响。詹姆斯-斯坦的估计人提供了一个更多的理由
能够估计对比赛历史有限的球员的发球和回归能力。收缩
服务百分比，我们计算在匹配数据集中所有记录的fi统计数据的方差

Dm³

$$\hat{t}^2 = \frac{fiyDm (be \hat{y} fav)^2}{|Dm| \hat{y} 1}$$

³每个fi都是根据前12个月的玩家数据计算出来的。

然后，每个估计器 f_i 都基于 n 个服务点。每个估计量 f_i 代表 f/n_i 。如果在发球时获胜，我们可以计算估计器 f_i 的方差和相应的归一化系数

$$s_i^2 = \frac{f_i(1 - f_i)}{n_i}$$
$$\text{With } w_i = \frac{s_i^2}{s_i^2 + \bar{y}_i^2}$$

最后，詹姆斯-斯坦估计器采用这样的形式：

$$JS(f_i) = f_i + B_i(f_i)$$

我们对 g_i 重复同样的过程，以获得返回统计量的詹姆斯-斯坦估计量。向看看收缩如何使我们的模型对小样本量健壮，请考虑下面的例子。2015年，当丹尼尔·伊拉希（COL）和伊沃·卡洛维奇（CRO）面对波哥大世界杯时，伊拉希已经做到了

过去一年只参加了一场巡回赛级别的比赛。从之前的一个单边的胜利，他的一年-长发球率， $f_i = 51/64 = .7969$ ，与一年的时间相比异常高。旅游水平平均 $off_{fav} = .6423$ 。

玩家名称 p	恩特斯赢得了积分			r 点数赢得了 r points-22	g_i	elo评级
丹尼尔伊拉希	51	为64.7	969	67	.3284	1516.92
卡罗维奇	3516	4654.75	55	1409	4903	2874

$$f_{ij} = f_i + (f_i - f_{fav}) w_i (g_i - g_{fav}) = .6676 + (.7969 - .6423) w_i (.2874 - .3577) = .8925$$

$$f_{ji} = f_j + (f_j - f_{fav}) w_j (g_j - g_{fav}) = .6676 + (.7555 - .6423) w_j (.3284 - .3577) = .8101$$

根据克拉森和马格努斯结合球员结果的方法，我们计算了伊拉希的结果。服务比例为89.3%。这是非常高的，日食了卡洛维奇的81.01%。serve预测这很奇怪，考虑到卡洛维奇是最有效的服务器之一。游戏的历史。从服务统计数据，我们的层次马尔可夫模型计算Elahi的获胜概率为 $M_p(0, 0, 0, 0, 0, 0) = .8095$ 。这个预测似乎有不合理的信心。埃拉的胜利，尽管他只收集了一场比赛的球员统计数据。卡洛维奇氏症360分的Elo优势计算出伊拉希的获胜概率为

_____00

$$\chi(t)=(1+10$$

$$1876.981516.92$$

$$g^1_{1}=-.1459$$

这导致我们在使用有限的历史数据时进一步质疑这种方法的有效性。因此，我们转向詹姆斯-斯坦估计器来缩小伊拉希的服务和返回概率到总体平均喜爱， gav 。

$$JS(f_i) = f_i + B_i(f_{av} - f_i) = .7969 + .7117(.6423 - .7969) = .6869$$

$$JS(g_i) = g_i + B_i(g_{av} - g_i) = .3284 + .7624(.3577 - .3284) = .3507$$

$$JS(j) = f_j + B_j(f_{av} - f_j) = .7555 + .0328(.6423 - .7555) = .7518$$

$$JS(g_i) = g_i + B_j(g_{av} - g_i) = .2874 + .0420(.3577 - .2874) = .2904$$

$$JS(f_{ij}) = f_{ij} + (JS(f_i) - f_{av}) + (JS(g_i) - g_{av}) = .6676 + (.6869 - .6423) + (.2904 - .3577) = .7795$$

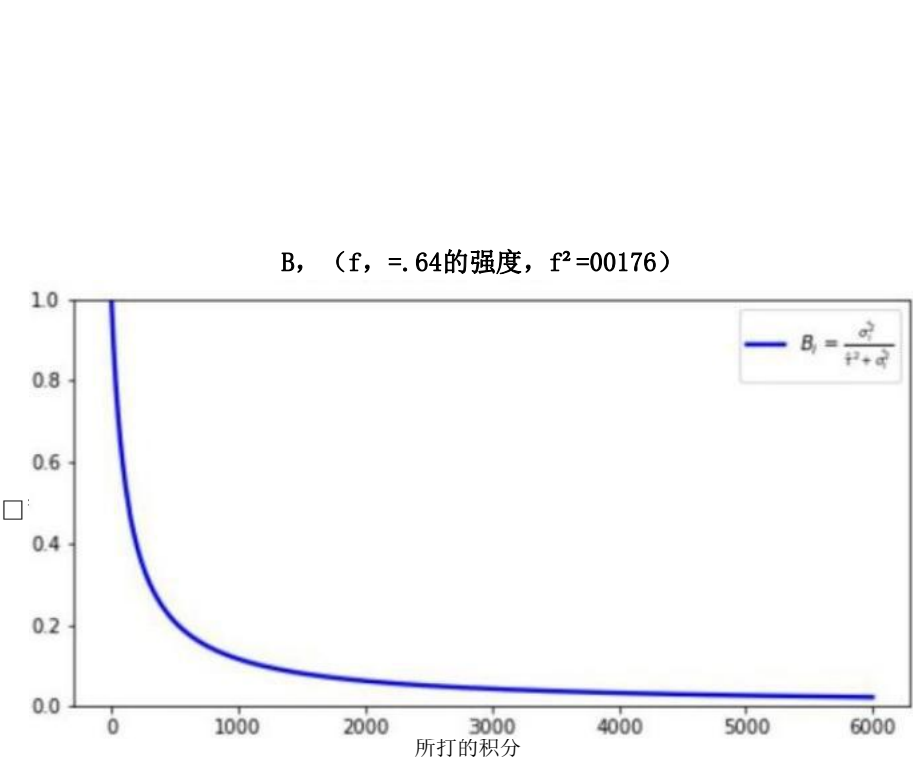
$$JS(j) = f_{ij} + (JS(f_i) - f_{av}) + (JS(g_i) - g_{av}) = .6676 + (.7518 - .6423) + (.3507 - .3577) = .7841$$

上图，我们可以看到，詹姆斯-斯坦的估计者缩小了伊拉希的统计数据，远比卡洛维奇的更多，因为卡洛维奇在过去的一年里参加了许多巡回赛级别的比赛。给定 $JS(f_i)$ ， $JS(f)$ ，我们

计算 $M_p(0, 0, 0, 0, 0, 0) = .4806$ 。通过缩小服务/餐厅的统计数据，我们的模型降低了Elahi膨胀的服务百分比，并在小样本量下变得更加稳健。

由于有限数据的基于点的预测威胁模型性能，特别是交叉熵，詹姆斯-斯坦估计器允许一种更安全的方法来预测匹配结果。稍后，我们将使用詹姆斯-斯坦估计器不仅规范化一年的发球/回报统计，而且还规范化表面特定和对手调整的百分比。对于一个在过去12个月里平均为 $.64$ 的球员来说，这张图表说明了如何标准化

在我们的数据集中，系数 B_i 随 n 而变化：



3. 6对手-调整后的服务/回报统计数据

虽然巴梅特和克拉克的方程式确实考虑了对手的发球和回传能力，但它不追踪球员历史中普通对手的能力。这是很重要的

球员的发球/回传率可能会因面对较弱的对手或邪恶而膨胀

反的在本节中，我们提出一个对巴内特和克拉克方程的变化来代替

f_{av} , g_{av} 与支持者调整平均19放弃, l_{gf} 尤尔为 π 。然后方程

成为

$$\begin{aligned} f_j &= f_t + (f_y (1_y \text{放弃} o \text{身}) 9 (g_{iy} (1 \text{以上}) - \\ f_{ji} &= f_t + (= (1 \text{以上以上})) g (= (1 \text{以上})) \end{aligned}$$

a_v 代表了他在过去12个月里所面对的对手的平均回归能力。为了计算这一点，我们加权每个对手的

返回能力 g_j

他们各自的比赛。

放弃 = $\frac{t^2}{t} \text{ kg } (yy1, m+t)$ (p的返回数 比赛k中的得分” (比赛k前选手的返回能力)

$\frac{t m l}{(yyt, m+t) \text{ kg}}$ pf在比赛k中的回报得分

为了阐明跟踪对手调整后的统计数据的效果，我们考虑了2014年美国网球公开赛

米哈伊尔尤日尼（罗斯）和基尔乔斯（奥斯）的回合比赛。

球员的名字	s分赢得了s points 1828		r分赢得了r分ts 1445	gi	elo评级		
米哈伊尔尤日尼	2960	是6176	2947	.388517	49.77		
尼克凯吉奥斯	900	1370	.6569	424	1323	.3205	1645.03

从标准方法中，使用 $\text{ft}=.6583$ ，我们计算出 $\text{fij}=.6446$ ， $\text{fji}=.6159$ 。然后，转到

计算调整后的统计数据，我们考虑在服务器/回报中获得的加速分数，

鉴于尤日尼和基尔高斯过去的对手。正如我们从 fi 中观察到的，放弃，放弃，基尔高斯

在过去的12个月里，他面对着比尤兹尼稍强一些的对手。

球员的名字	交给 在下面	米哈伊尔	s expected s ad	l的jfi	729.74.0332	r预期r adj	-
尤兹尼。4156	尼克 基尔	gios.4313	6812			939.60.0697	
			779.19	0882	.7060	388.92	0265

sadj^l 、 radji 和调整服务概率（ ij ）、 $\text{adj}(\text{i})$ 计算如下：

$$\text{adii} - \text{#Pain} \text{os} \text{#9} \text{ (1y放弃)} = \text{fig} \text{ (19放弃)} -$$
$$\text{point} \text{如此赢} \text{sn} \text{在} \dots \text{上} \text{r} \text{fe} \text{t} \text{unm} \text{-g} \text{ (1yfi} \text{#up-oD} = \text{gi} \text{9} \text{ (19fi-超过)}$$
$$(\text{ij}) = \text{ft} + (+ \text{ (放弃)}) \text{ y (演出(1))} = \text{ft} + \dots = .6288$$
$$\text{adj}(\text{fi}) = \text{ft} + (19 \text{美元}) \text{ y (giy (1yfi以上))} = \text{ft} + \text{sadjjy radji} = .6406$$

使用常规发球概率，尤兹尼有望赢得 $\text{Pm} \text{ (0,0 , 0,0,0,0)} = .6410$ 。

调整发球率，我们在基尔高斯的比赛中表现较强的对手

历史和尤兹尼的胜利冠军 $\text{(0,0,0,0,0,0)} = .4411$ 。同时进行调整

发球/回统计并不总是导致获胜概率的显著变化，他们是一致的

考虑到球员的对手产生发球概率的相对能力。

3.7结果

以下结果是在我们的逐点比赛中使用ATP最佳匹配获得的

数据集，不包括戴维斯杯。4我们观察了基于elo的预测变体的性能

和基于点的模型。因为所有的基本实现都提供了显式的公式

另外，我们直接评估了他们在2014年巡回赛级别的数据（2.409场比赛）上的表现。

4. 戴维斯杯和戴 是由国家组织的参赛队伍之间为期一年的比赛。由于科瓦尔奇克被排除在她的评估之外（科
维斯杯比赛的球员 瓦尔奇克，2016），而且许多比赛涉及相对未知
，我们排除在外 所有匹配项。

在逻辑回归的情况下，该模型是根据2011–2013年的比赛水平数据进行训练的

（7828个匹配项）。我们根据精度来衡量每个模型的性能，如第2节所述

3. 3. 我们还考虑了对数损失，一个模型预测的过度可能性的度量。日志

损失很重要，因为它会惩罚那些错误地分配过度自信或分配不足的模型

自信赢得概率。由于我们希望优化匹配预测模型的对数损失，

考虑各种赛前预测的可能性是很重要的。以下条款

表示基于点和Elo模型的变化：

基于km点的分层马尔可夫模型结合服务/服务

来自巴米特/克拉克的百分比

James-Stein-版本的KM模型，与服务返回百分比或-

与詹姆斯-斯坦的估计

模型的表面版本，其中所有的评级和百分比都是分层的

表面（硬质、粘土、草）

538-具体表示538的“衰减k因子”方法

计算Elo评级

表3. 1：结果表

方法变化	精度日志丢失	
公里	64. 8	649
KM詹姆斯斯坦	65. 4	. 616
KM表面	63. 3	. 707
KM曲面詹姆斯-斯坦	63. 6	639
KM调整	67. 8	632
KM调整詹姆斯-斯坦	67. 9	617
纳穆赫	69. 1	. 586
Houmuh surface	68. 4	. 591
Houmuh 538	69. 2	. 587
Houmueh表面538	69. 4	592
日志现在538，表面现在538	69. 5	577

3.7.1 讨论

正如预期的那样，James-Stein归一化显著地改善了每个基于点的模型的对数损失。而基于表面的Elo与常规的Elo竞争，限制了克拉森和Mag-

基于nus点的方法对特定数据的表面显然损害了性能。这可能是由于有限的表面特定数据。由于许多玩家在不同的表面上的表现也有所不同，因此从特定于表面的数据中计算服务百分比提供了表达玩家的表现如何因表面而变化的潜力。然而，即使使用詹姆斯-斯坦归一化，样本量的损失似乎超过了从特定表面的数据中获得的好处。因为我们没有钻机

大胆地探索加权玩家的特定表面统计数据的可能性，但仍然有可能使用一个基于点的模型，在计算服务/回报百分比时有效地考虑表面。最后，使用对手调整的发球回报统计进一步提高了表现。通过将每个对手的能力充分融入到玩家的比赛历史中，该模型接近于与预测相似的Elo评级，以及两个共享 $R^2 = .76$ 之间的获胜概率预测。

通过将Elo和表面Elo插入到一个逻辑回归模型中，我们实现了一个对数损失。577. 在匹配训练数据后，该模型学习了每个变量的系数，这些变量将56%的权重分配给56%的常规Elo评级，44%的权重分配给表面Elo评级。这两种评级的接近平均权重表明，表面特定能力对预测有多重要

比赛除了可以直接从博彩市场获取信息的模型外，没有

其他模型已经记录了更好的日志损失。科瓦尔奇克报告了使用538的Elo方法的70%的准确率（科瓦尔奇克，2016），使用卫星事件计算Elo评级(ATP挑战者

以及未来)，除了巡回赛级别的活动5。虽然这可以解释了精度的小幅提高，但她的方法实现了0.59的日志损失，这并不优于我们的实现。为了简单起见，我们只计算了巡回赛级别比赛的所有Elo评级和统计数据。6

5可以观察到<https://github.com/skoval/deuce>.

6每年大约有3000场巡回赛级别的比赛。

最后，我们基于Elo的逻辑回归比单一Elo评级的增量改进，这意味着一个有效的评级系统应该考虑整体能力和表面特异性能力

Sipko最近探索了机器学习方法，测量逻辑

回归，共同对手模型，和一个人工神经网络(Sipko和克诺顿带，

2015)。虽然他声称通过人工神经网络在博彩市场获得了4.3%的投资回报率，但他的机器捆绑模型并没有超过日志损失。在预测2013-2014年ATP匹配时，有61个7。尽管目前在整个行业和

学术界(Lewis-Kraus, 2016)，没有人发表的发现表明他们在预测网球比赛方面的优势。由于缺乏基于市场的模型，最有效的方法似乎是坚持埃洛的“无情的贝叶斯设计”，它经常被吹捧(Bialik et al., 2016)。经过过渡，我们将考虑这些发现如何可以提供一个有效的匹配预测模型。

⁷这些模型在2004-2010年的ATP比赛中进行了训练，并在2011-2012年的ATP比赛中进行了验证。

4. 游戏内匹配预测

以下方法将在我们有逐点数据的巡回赛级别的比赛中进行测试。比赛跨度为2010–2017年，占这段时间内所有巡回赛比赛的近一半。萨克曼数据集中的逐点记录采用以下字符串的形式：

米哈伊尔Youzhny和。EvgenyDonskoy（2013年澳大利亚网球公开赛，第二轮）

P=“SSRSS; RRRR; SRSS; SRRSRSS; SRSSRS; RSRSS; SRSRSS; RSRSRSS; SSS

S. SSSRRRSS; RSSSS; SSRSS; SSSRS; SSSS; RRRSSSRSSRRSRSS; SRSRSS; SSS

RS; RSRSSRSS; SSSS; SRSS; RSRSSRRSS; R/SR/SS/RR/RS/SR. RRRR; ...”

S表示服务器赢一分，R表示返回者赢一分。个人游戏是

以“; ”分隔，由“ ”设置。，以及与“ ”之间的服务变化。通过迭代通过

这个字符串，可以从n个总点的匹配中构造n个数据点 $\{P_0, P_1, \dots, P_{n-1}\}$ ，

P_i 表示比赛的子集，之后我就开始得分了。

$P_0 =$ “

$P_1 =$ “S”

$p_2 =$ “ss”

$p_3 =$ “ssr”

使用 $M = \{M_1, M_2, \dots, M_k\}$ 表示我们的逐点数据集中完整的匹配字符串，

我们枚举的数据集的大小变为米|。总共超过120万

2010–2017年ATP比赛中的个人积分。而许多不匹配的预测

模型采用层次马尔可夫模型结构，我们将首先测试几个机器-

以学习方法作为基线。现在，我们将探讨以下方法：

- 逻辑回归
- 随机森林
- 层次马尔可夫模型
- 测试实验
- Elo诱导服务百分比

4.1 逻辑回归

考虑一个逻辑回归模型，其中每个观测值 $x_i = [x_{i1}, x_{i2}, \dots, x_{ik}]$ 都有 k 个特征

缝线。当 P 表示 p_i 的获胜概率时，

$$\log \frac{P}{1-P} = \gamma_0 + \gamma_1(x_{i1}) + \gamma_2(x_{i2}) + \dots + \gamma_k(x_{ik})$$

从任何分数（ s_i 、 s_j 、 g_i 、 g_j 、 x_i 、 x_j ）中，我们都可以简单地将这些值输入到我们的模型中

x_i 的特点。逻辑回归的结构使得它很容易考虑附加的特征

对于每个玩家，如Elo差异，表面Elo差异，或打破优势。之前

在模型中添加更多的特性，我们考虑两个基线：一个使用分数差异的模型

（ $s_i s_j$ ， $g_i g_j$ ， $x_i x_j$ ）和另一个训练了Elo差异和领导启发式 L_i 的模型。

这个启发式估计了 p_i 在集合、游戏和积分方面的总领先优势：

$$L_{ij} = (s_i - s_j) + \frac{1}{2} (g_i - g_j) + (x_i - x_j)$$

这些系数保持了集合、游戏和点之间的顺序，因为一个人不能领先六个游戏

没有赢一盘或四分，没有赢一场比赛。而它假设这些都是相对的

排序，启发式是从任何比分近似多少pi领先pj。在
该模型，我们在预测比赛获胜者时考虑以下特征：

表4. 1： Logistic回归特征

V ariable	描述
铅裕度	铅启发式Lij
eloDiff	现在（第0页）-现在（第1页）
s eloDiff	s Now（p0）-s Now（p1）
setDif	SetsWon（p0）-SetsWon（p1）
gameDif	inSetGamesWon（p0）-inSetGamesWon（p1）
pointDiff	inGamePointsWon（p0）-inGamePointsWon（p1）
breakAdv	服务（p0）-服务游戏（p1）+I（目前服务）
brkPointAdv	1（保持断点）-I（面向 破碎 观点
百分比 的 积分赢 在...上 p0's 服务 所以 遥远地	
sv点件 1%的 积分赢 在...上 p1's 服务 所以 遥远地	

接下来，我们测试以下特性的组合：

- 1)setDiff+游戏差异+点Diff（差异）
- 2)领先缘+系数+系数（Elo）
- 3)所有功能（全部）

4. 1. 1交叉验证

在我们的三局三胜制数据集中，每一场比赛平均约有160分。在调谐超
机器学习模型的参数，我们实现k =5实现组验证k。这个
将训练集随机分割成五个大小大致相等的子集，同时总是进行分组
在同一比赛中的得分。通过将来自同一匹配项的点分组在一起，
这种方法可以防止训练、验证和测试集之间的重叠。将每个子集视为
作为一个验证集，我们在测试性能之前对剩余数据进行训练
相应的验证集。然后我们平均所有k个验证集的性能
选择模型表现最好的超参数。当我们在验证和测试集上评估模型性能时，结果反映了从未见过的
匹配模型的性能
在...之前

4.2随机森林

伯克的获胜概率模型是体育界最著名的（Burke，2014）。

他们根据历史数据计算一个球队在比赛中任何时候的获胜概率

通过组合的装箱输入和平滑它们的结果概率。内特尔顿住所名称

和锁建立在这个方法之上，通过使用随机森林的方法(锁和Netleton，2014)。

一个随机森林由一个随机生成的分类树的集合组成。每一个

树形成决策函数的特征子集与分裂产生最大的分裂

犯罪能力。因为内特尔顿和洛克都有足球特有的功能，比如向下，

得分，和码的目标，我们这样做类似的网球特点，并训练我们的验证

放置

表4.2：随机森林特征

变量	描述
表面	硬，粘土，草
放置	第一、第二、第三
e1oDiff	现在（第0页）-现在（第1页）
setDiff	SetsWon（p0）-SetsWon（p1）
gameDiff	inSetGamesWon（p0）-inSetGamesWon（p1）pointDiff
inGamePointsWon	（p0）-inGamePointsWon（p1）
brkPointAdv	ServeGamesWon（p0）-ServeGamesWon（p1）+ （curently
serving)breakAdv	我（保持断点）-I（面对断点）

4.3层次马尔可夫模型

由于服务百分比已经从历史数据中计算出了服务百分比，我们的层次马尔可夫模型

能够产生在比赛中获胜的概率估计。使用解析方程

利用球员的发球能力 f_{ij} 、 f_{ji} ，我们从每个比分中计算出 $P_m(s_i, s_j, g_i, g_j, x_i, x_j)$

$(s_i, s_j, g_i, g_j, x_i, x_j)$ 在一场比赛中。为了评估这个模型的性能，我们在每个模型上都重复这个模型

在我们的数据集中匹配，与 f_{ij} 、 f_{ji} 计算与一个人的方法选择。

4.3.1 测试实验

上述方法在计算获胜概率时只考虑了当前得分和预先计算的发球百分比。然而，在许多情况下，相关的信息可能是

同事来自Pk。请考虑以下匹配中的子字符串，

P="SSSS; RSSRRSS; SSSS; SRRSRSSSS; SSSS; RRRSSSRSSS; ,

上面的序列显示了目前三场比赛的比分。然而，pi有

赢得了12/12的服务积分，而pj已经赢得了18/30的服务积分。如果双方继续发球速度相似，圆周率更有可能打破发球并赢得比赛。由于最初的预测是fij，fji是基于历史上的发球百分比，所以在比赛中的发球百分比可以帮助我们更好地确定每个球员在某一天的发球能力是有意义的。要做

这样，我们就可以在比赛的t时刻更新fij，fji，以考虑到每个球员的发球表现迄今

Bevc探索了这种方法，通过建模每个玩家的服务百分比fij之前

可以在匹配中间更新以产生后验估计。通过二项式共轭，我们可以更新一个球员赢得赢得积分时的优势
审判

BpiorityBeta (a、b)

(+型，b+标准)；

如果这是我们的先验，那么a，b必须满足E(b先验)=。那么a是一个超参数，它决定了我们的先验的强度。为了找到一个合适的优先强度，我们将测试各种强度

通过在我们的训练集上的交叉验证得到的值。无论a如何，比赛的影响都会对我们的后验发球估计总是随着得分的增加而增加。在比赛的任何时候，我们都可以获得任何一名球员的后发球百分比

$$f_{ij} = E(\text{bposterior}) = a \frac{a + swon}{a + b + stotal}$$

4.3.2 elo诱导的服务概率

早些时候，克拉森和马格努斯提出了一种推断服务概率的方法

赛前胜利预测 y_{ij} 。通过施加一个约束条件 f_{jit} ，我们可以创建一个一对一的约束条件 $S: S(+, t) y(f, f)$ ，它生成服务概率 $\tilde{f}_{j, j}$ for

两个玩家都认为 $P_m(0, 0, 0, 0, 0, 0) = y_{ij}$ 。由于这篇论文发表于2002年，克拉森说

和Magnus反转了他们的匹配概率方程，以产生基于ATP等级的预测的服务概率。然而，由于Elo

优于ATP排名，我们将此方法应用于Elo

预测

由于分支的复杂性（见第2.5节），我们的层次马尔可夫模型的匹配问题-

能力方程的逆没有解析解，即使我们指定了五+小二=。

因此，我们转向以下近似算法来生成服务百分比

这符合我们的Elo预测中的获胜概率：

算法1 elo诱导的 服务概率

```

    程序服务（问题，总和）
        currentProb
        y.5 diff y和/4
        而|
        当前问题-问题|>做：如果当前问题
            <问题然后s0+=差异其他s0-
                差异差异=
                diff/
                2
        currentProk
        y匹配Prob（s0，和s0）返回s0，和s0

```

为了生成给定匹配的elo诱导服务概率，我们运行上述算法

使用 $PROB=y_{ij}$ ， $SUM=f_{ij}+f_{ji}$ ，并设置到所需的精度水平。²在每一步，我们呼叫

（）计算从比赛开始开始，如果 p_i 和 p_j 有发球的获胜概率

概率分别为 $f_{ji}=s_0$ ， $f_{ji}=和ys_0$ 。然后我们比较当前的问题和问题和

1 f_{ji} 和 f_{ji} 按照3.4中规定的方法与James-Stein标准化计算，以防止极端的结果。

2 对于本项目而言，设置 $\epsilon=0.001$ 已足够准确。

差异增加 s_0 ，每次迭代增加一半。这个过程一直持续到服务结束

概率 s_0 ，和 s_0 在PROB范围内产生一个获胜概率，取 $0(\log^{-})$ 呼叫

matchProb.

这种逆算法是有用的，原因有几个原因。给定任何有效的匹配前预测 y_{ij} ，我们可以根据我们的层

次结构产生与 y_{ij} 一致的服务概率

马尔可夫模型。通过设置约束 $f_j+f_{ji}=t$ ，我们也确保了我们的和

球员的发球概率与历史数据一致。而克拉森和马格努斯则这样认为

$=f_{ij}$ 在很大程度上独立于 y_{ij} ，在预测特定得分匹配时更重要（Barnett et al., 2006）。使用层

次马尔可夫模型的方程，巴内特特别计算了达到给定 f_{ij} 的任何集合分数的概率。当

比较匹配的概率，可以观察到随着 t 的增加，更接近的集合分数和

抢七局的比赛变得更有可能是了。4也就是说， t 编码了关于比赛比分的可能轨迹和每场比赛获胜的相

对重要性的信息。对于较高的 t ，

服务游戏赢得的次数更高，服务休息相对更重要，而较低的 t 则相反。现在，给定 y_{ij} 和 t ，我们可以

为任何两个球员产生投球诱导的发球概率。

4. 3. 3 $f_{ij}+f_{ji}$ 的重要性

为了说明 $t =f_{ij} +f_{ji}$ 的重要性，我们考虑了以下两个匹配项：

匹配的	乌尔法斯伊普1费里奇	阿诺洛佩兹约翰一世	sner ATmshangnai 2014	困难的478	1. 55	t
m2 An	德雷亚斯塞皮胡安	摩纳哥ATP基茨布	uhel 2014 clay.4761.14			

在这两个比赛中， y_{ij} 近似相同，而 t 差异显著。如何可视化

服务中断可以有不同的重要性，考虑与以下点的级数对应的赢概率图，P。在每种情况下， p_i 和 p_j

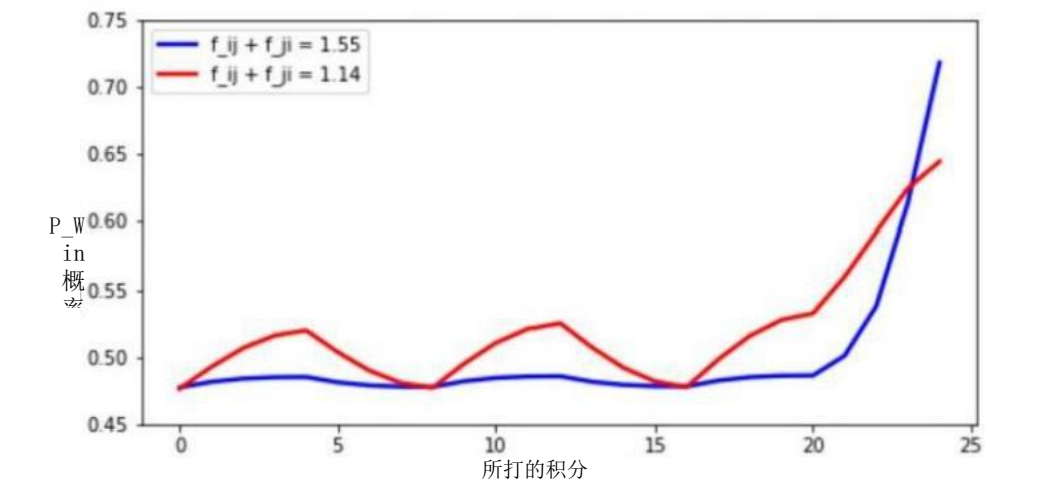
的服务概率为

由方法4. 3. 2计算。

- 3虽然这只是一个估计，但 $t=f_{ij}+j_i$ 是我们从这个项目中涵盖的方法中获得整体服务能力的最佳预测器。
- 。 47-6、6-7、7-5、5-7等。

P=“SSSS; SSSS; SSSS; SSSS; SSSS; RRRR”

赢概率图



在前五场比赛中，两名球员都赢得了发球局的每一分。当 $t=1.55$ 时，服务保持是预期的，和赢的概率很难移动。在第六局破发发球后，

他的获胜概率突然上升，这表明了一个巨大的优势。当 $t=1.14$ 时，我们的模型假设玩家赢得服务游戏的概率要低得多。因此，每次球员保持发球，他的胜概率显著增加。当圆周率在第六局休息时，他的胜率会上升，但没有前一场那么显著。当服务游戏以较低的概率获胜时，服务中断的优势就会变得更弱。

t 对比赛轨迹的影响表明，当 t 准确反映球员的整体发球能力时，任何比赛中比分的预测都是最有效的。虽然Elo评级或投注赔率产生最有效的赛前预测（Kovalchik，2016），但他们的隐含赔率没有提供给定比赛中 t 的信息。因为 t 对于

来自特定得分线的匹配内预测，方法4.3允许我们将任何有效的匹配前预测 y_{ij} 与 t 结合起来，以产生新的服务概率。 . 2

4. 4结果

以下模型在2011-2013年逐点数据（4584场比赛）上进行训练，并在2014年逐点数据（1855场比赛）上进行测试。为了保持一致性，我们只使用了ATP巡回赛三场三胜制的比赛，并排除了戴维斯杯的比赛。从交叉验证

训练集，我们确定设置超参数 $\alpha = 300$ 在beta实验中产生最佳性能。然后，我们在KMelo诱导和KMlogit-中添加了一个beta变异

诱导的，我们的两个表现最好的基于点的模型。由于观察结果对应于每个匹配中的单个点，较长的匹配对我们的数据集贡献更大，并有更大的流感-

对评估指标的影响。由于解决这种巨大的影响可能会使结果的可解释性复杂化，我们在逐点的基础上提出模型评估。

指定特征集的LR-Logistic回归模型

在4.2中指定特征集的RF-随机森林模型

等效km模型，服务概率设置为旅游级平均（假设同等球员能力）

利用基于elo的获胜概率生成的服务概率的elo诱导-KM模型

基于Elo和表面Elo的逻辑回归生成的km模型

表4.3：结果表

方法变化	精度日志丢失校准		
LR差值	71.4	540	993
LR Elo	75.2	505	1.006
LR所有	76.2	488	993
射频	70.1	562	902
KM等效	71.5	538	1.040
KM詹姆斯斯坦	74.9	501	1.002
KM调整詹姆斯-斯坦	75.9	498	.992
Elo诱导KM	76.2	486	1.001
elo诱导KM (a=300)	76.3	481	1.003
KM日志项目	76.5	481	996
KM logit (a=300)	76.5	477	999

在四种模型中，随机森林的预测能力最低。而这可能是因为难以构建一个有效的特征空间，接近所有的特征输入到每台机器排队模型对获胜概率有加性影响。积极的分数差异，打破优势和Elo的差异都将提高玩家的预期获胜概率。唯一的所考虑的非附加特征是“sef和表面”。而随机森林模型则是能够考虑这两种特征中的各种情况，我们得出结论，它倾向于将类似的例子分组，而不是估计变量之间的线性关系，最后解释了它的表现。

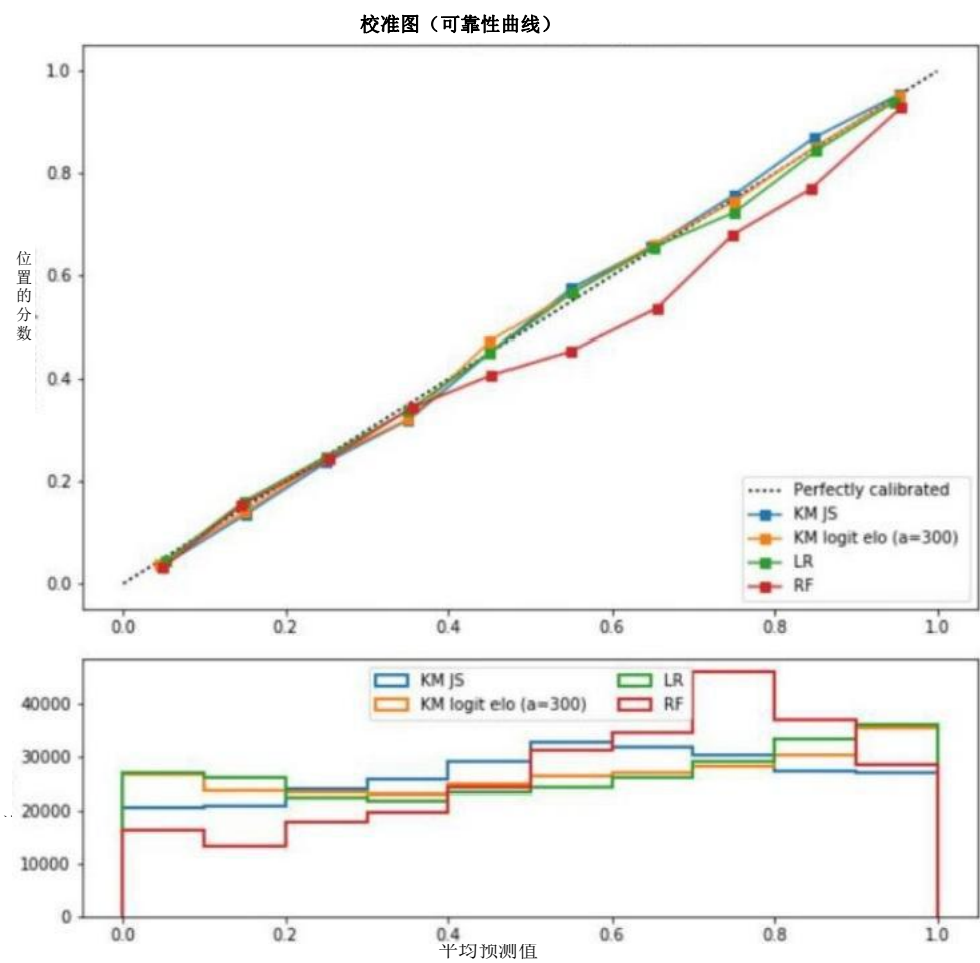
另一方面，逻辑回归是为了学习fea之间的加性关系方法及其概率输出。它的性能与完整的功能集是有竞争的我们的积分模型，尽管他们熟练地掌握了网球的得分系统。5. 大概是，与基于点的模型相比，该模型在其自身的训练能力方面具有优势数据除了我们对beta实验的超参数搜索外，基于点的模型没有提供类似的方法来优化关于训练集的性能。

5考虑一下，例如，逻辑回归是如何不知道抢七游戏何时发生的。

在基于点的模型中，KMlogit-elo ($\alpha=300$) 获得了最好的性能，尽管运行beta实验的改进很少。因为Elo的评级产生了更多的效果

可靠的赛前获胜预测比估计，从历史发球/回报数据的估计，使用从Elo评级诱导的服务概率，然后从我们的logit Elo模型，显著提高了我们的模型性能。在接下来的几节中，我们将检查模型跨数据集子集的校准和性能。

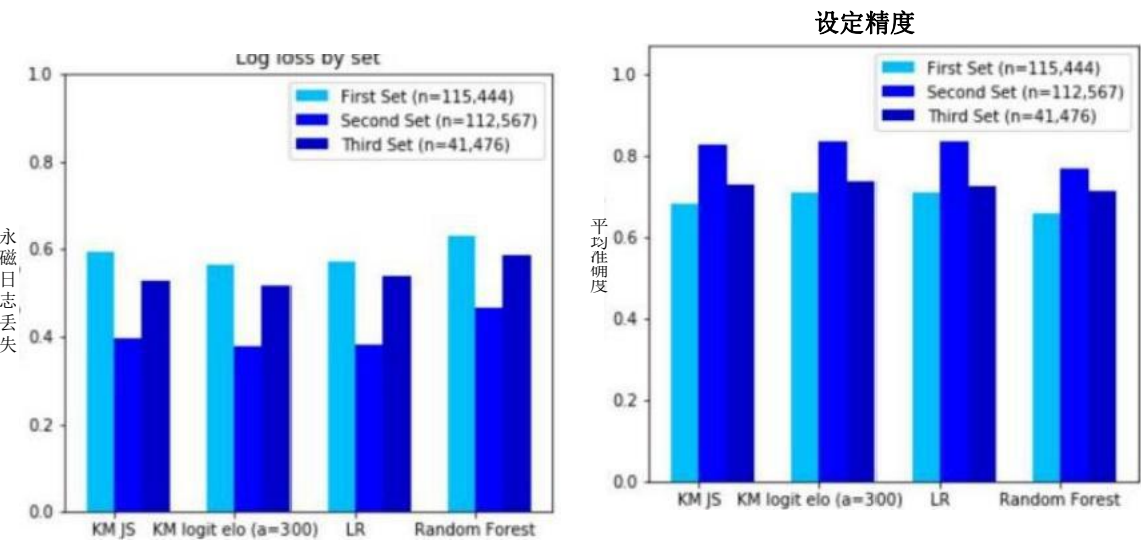
4. 4. 1模型校准



校准表示观察到的成功率和平均预测的成功率之间的比率。从一个完美校准的预测 p 获胜概率的模型来看，我们总是可以期望获胜以概率 p 发生。而随机森林模型的预测不足获胜后，我们的其他模型非常像一个完美校准的模型。从我们的桌子上，

KM logit elo和KM JS近乎完美地校准了1.001和0.999，而Logistic
回归reported. 993. As在测量校准中存在不确定度，我们不能合理地期望产生更多校准良好的
模型。给定获胜概率预测在我们的测试集中匹配的任何点的 y_{ij} ，我们期望 p_i 以概率 y_{ij} 获胜。

4. 4. 2由集

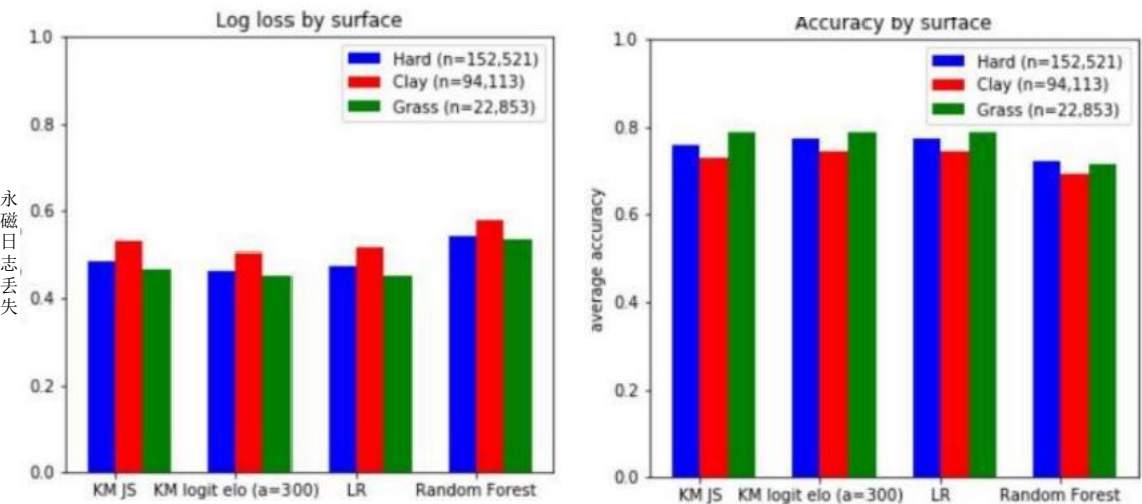


由于我们早期关于比赛的信息较少，所以总是很难预测
在第一集设置中进行匹配。然而，在这两个指标中，模型在第二组中是最有效的。到比赛的这个
时候，第一盘的获胜者往往是获胜的热门人选
在比赛中，除非他们的对手有明显更高的Elo评级或发球率。由于在我们的测试集中有64.3%
的比赛是直落集的胜利，许多比赛在第二集比赛中为我们的模型提供了相对容易的预测6
一旦比赛进入第三盘，不确定性就会增加，因为比分趋于1-1。虽然第一组和第三组都是平的
，但在第三组预测结果更容易，因为它的结果直接对应于匹配结果。当

6场三局三胜制的两盘比赛都是直落两盘。

在第一盘比赛中预测，赢得第一盘比赛的球员仍然可以继续输掉比赛，这增加了不确定性。

4. 4. 3表面



从表面上看，模型预测草地场地匹配效果最有效，红土地匹配效果最低。从历史上的平均英尺可以观察到，草地球场往往比红土球场产生更高的发球率。大概，这是

因为球在草地球场上反弹的速度比红土球场更快。对观察到的差异的一种解释是，在红土球场上，由于 f_{ij} ， f_{ji} 更低。随着服务中断的可能性的增加，我们的模型就更难预测

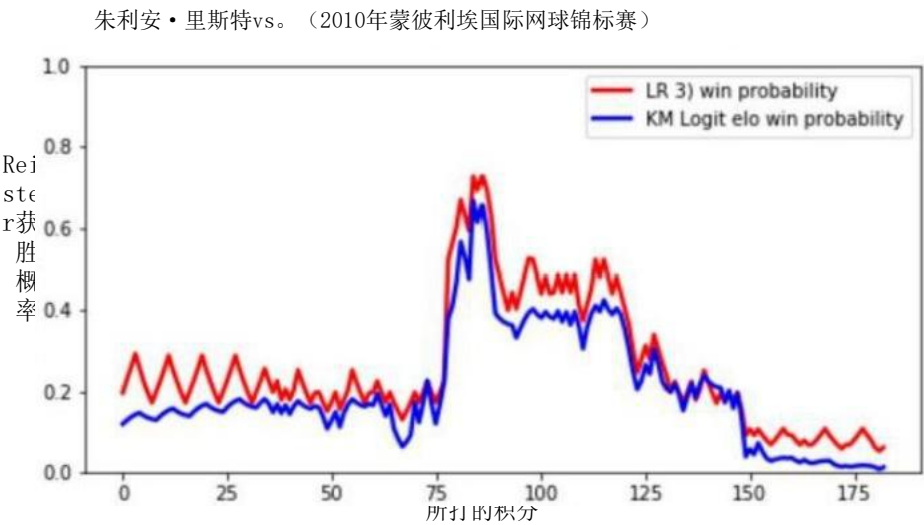
比赛是因为来自当前比分的优势更有可能被逆转。大多数比赛都在硬地球场进行，最终证明了比赛在草地和泥土之间的可预测性。对比赛波动性的进一步了解可能会正式揭示出来

在不同的表面上得到的结果也不同。⁷

⁷我们可以从逐点序列或模型的获胜概率输出中计算出波动率的度量。

4.5可视化获胜概率

由于最有效的Logistic回归和KM模型在预测性能上相差不远，我们现在用几场巡回赛级的比赛来可视化它们的预测。



在比赛的前30分中，我们的模型之间的对比让人想起了

4.3.3. 因为我们的逻辑回归模型已经学习到了“点差异”的特定系数

特征，赖斯特的获胜概率在他和加斯凯特交换时明显地上下上升

游戏服务。我们的=日志模型似乎将fji设置到了一个高值，

期望频繁的服务保持并在这个时间框架内产生相对静态的获胜概率。而我们可以通过约束来传达

服务中断的相对重要性

因此，没有直接的方法来用逻辑控制概率波动

回归。

在比赛的最后30分中，考虑一下赖斯特的获胜概率是如何慢慢下降的

根据KM模型，而根据Logistic模型，它则保持近似不变

回归。我们的逻辑回归特征集的一个缺点是它们不能

区分得分差异相等的情况。由于加斯奎特在第三组结束时保持了打破优势，这解释了逻辑回归的行为，因为

分数差异保持相似。

一般来说，考虑一个在（1、1、5、4、3、0）服役的玩家，以及一个在（1、1、1、0、3、0）服役的玩家。在第一种情况下，发球的球员如果赢得接下来三分中的任何一分，就赢得比赛。从

第二种情况是，发球者在最后一盘一开始就拥有破发优势，而回传者有更多的机会回来。假设每个玩家在*f_i* =*f_j*=.64服务，

我们基于点的赢概率方程表明这两者之间有实质性的差异

情节

$$P_m(1,1,5,4,3,0)=.994$$

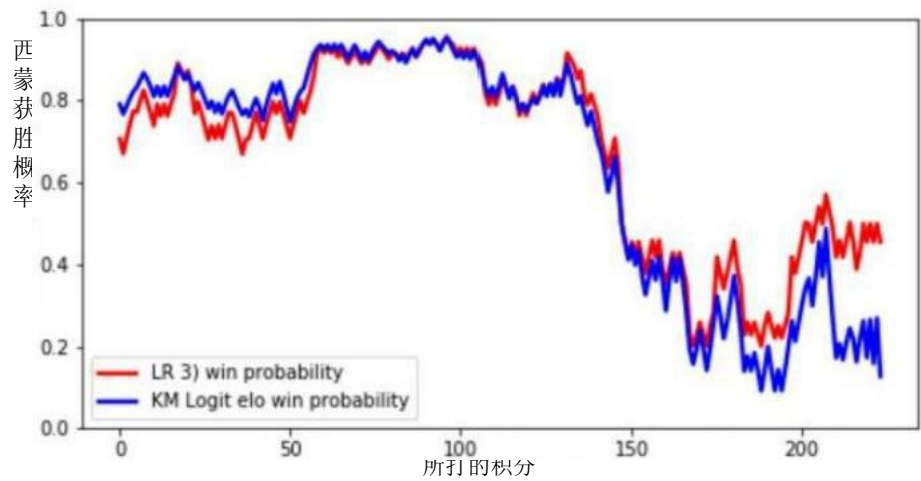
$$P_m(1,1,1,0,3,0)=.800$$

虽然第一种情况显然是有利的，但Logistic回归将观察到等价的差分（集差异、博弈差异和点差异），并计算出近似相同的概率

在这两种情况下。⁸

在一个类似的例子中，Logistic回归也可能没有注意到一个排名更高的玩家将输掉一场剂量比赛。

Gilles西蒙和。（2014年蒙特卡洛网球锦标赛）



西蒙在比赛的大部分时间里都在领先，直到加巴什维利重新回到第三盘。在

$$\frac{8Aesfting\ cheiate\ PN)=Paye.aiya.,x9x)=1t}{}$$

和 $P(1, 0, 5, 4, 3, 0) \Rightarrow P(1, 0, 1, 0, 3, 0)$ 的对称性。

第三盘结束时，加巴什维利以5-4的比分局，但模型未能检测到

西蒙就要输了。这是因为，设置为1-1，比赛差接近

甚至，西蒙的埃洛优势也超过了他目前的比分劣势。结果，当他输掉比赛时，他的获胜概率徘徊在50%之间，而不是跌向0。当我们

考虑了几个基于分数的交互术语来防止出现这样的场景，并将它们添加到

我们的特征集并没有显著地改变结果。不过，当比较模型的性能时

在特定的情况下，需要注意的是，KM模型是专门为网球预测而设计的，而我们的Logistic回归模型只是输入了似乎有用的分数相关特征。尽管有明显的缺陷，观察到的性能表明有进一步探索的潜力

机器学习方法。

5. 结论

第一次，我们记录了在匹配预测方法的性能

成千上万的比赛。虽然之前的工作只参考了个人比赛或1992–1995年温布尔登比赛数据集（克拉森和马格努斯，2003），我们已经应用了数千场比赛。从这个项目中得到的结果和杰夫·萨克曼的逐点分析

数据集应该作为该领域未来工作的适当基准。

虽然许多论文引用了克拉森和马格努斯的点模型，但没有一个引用

明确地提出了一种方法来纠正球员发球/训练历史中小样本量的不确定性。应用一种归一化方法，如詹姆斯-斯坦估计器

3.5，显著提高了基于点的模型的性能。通过考虑到整体上的因素

玩家的对手在过去12个月的能力，计算 f_{ij} ， f_{j-}

3.6中的调整方法提供了额外的预测能力。不过，这些变化都不能像Elo评级那样有效地预测预测。使用4.3.2中的近似算法，我们可以从任何匹配前预测 y_{ij} 和历史服务参数 $t = f_{ij} + j_i$ 中得到服务概率。该方法的结合，使用来自基于elo的逻辑回归的 y_{ij} ，和我们的KM模型最终产生了最可靠的匹配预测模型。

尽管如此，使用基于分数的特征集的逻辑回归的表现仍然值得注意。机器学习模型提供了从训练集中学习的潜力。当服务参数-

f_{ji} 可以从机器学习模型的预测中得到，

我们的传统

基于点的层次马尔可夫模型没有提供可比的方法来“训练”从a本身

资料组未来的工作是基于点的模型，提供了优化的某些方面

尊重一个训练集可能对球迷和赌徒都很感兴趣。然而，我们暂时概述了几种方法，可以在网球比赛的任何时候产生良好校准的获胜概率预测。

5.1对博彩市场的应用程序

本项目所探索的方法可以从多种方式应用于博彩市场。首先，我们可以采取任何赛前预测 jij ，从隐含的市场赔率，与历史服务

参数 $t=f_{ij}+f_{ji}$ ，并产生相应的保存概率 f_{ij} ， f_{ji} 从方法

4.3.2. 正如科瓦尔奇克发现，庄家共识模型的表现优于Elo评级

（Kovalchik, 2016），我们可能会通过使用pre-进一步改进匹配预测模型

匹配的胜算。接下来，我们可能会根据历史上的赛内赔率来测试我们的任何一种赛内预测技术。在线博彩公司必发公司提供历史市场数据

它有时间戳，但缺乏一致的匹配分数。在映射时间戳时

在我们的逐点数据集中，相应记分点的市场概率被证明超出了范围，对匹配内投注数据的分析确实存在。Huang展示了来自投注赔率的分数推断（Huang et al., 2011），而市场暗示的赔率可以作为比较模型的基准。最后，我们也可以测试任何这些模型的投注策略

实时的，提供最新的Elo评级和 f_{ij} ， f_{ji} 。

5.2未来步骤

除了beta实验外，本文中所有基于点的模型都在假设每个服务器播放的点都是i. i. d. 的假设下运行而克拉森和马格努斯得出的结论是

与ii. d. 的偏差足够小，这样的模型可以提供合理的近似（克拉森和Magnus, 2001），事实是匹配中的点不是独立的

虽然球员训练自己忽略记分板，但没有一个球员真正能够在每场比赛中保持恒定的比赛水平。一旦一个玩家的关卡出现波动，

顺序点变得依赖，表明在整个比赛中形式下降或峰值。未来通过点、游戏和集之间的依赖关系来探索势头的工作，可能会为我们的基于点的模型增加另一个预测维度。首先，我们可以量化发球得一分会影响发球得下一分的概率。总的来说，杰夫·萨克曼的逐点数据集提供了超过10万场比赛来探索

动量的影响。¹

本文所探索的方法只对ATP匹配进行了测试。本文中进行的所有方法和分析也可以应用于萨克曼的逐点数据集的女子WTA匹配，并进行对比，以说明男子的匹配可预测性和女性的旅游

5.3可视化的大满贯比赛

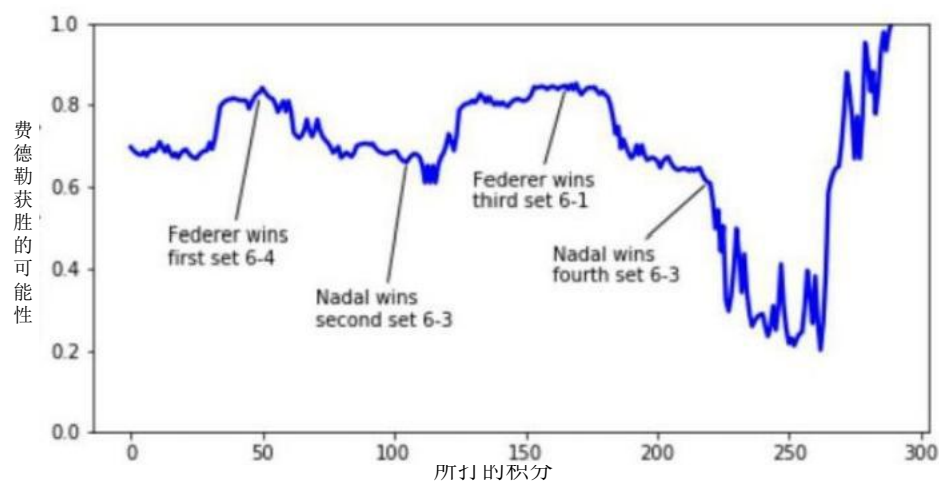
下面，我们将查看几个历史比赛的获胜概率图。概率由KM logit-elo ($\alpha=300$) 生成。

5.3.1的罗杰·费德勒vs.拉斐尔·纳达尔澳大利亚网球公开赛2017年决赛

今年早些时候，罗杰·费德勒在澳网决赛中以6-4-3-6战胜了拉斐尔·纳达尔6-13-66-3。

虽然我们检查了ATP巡回赛级别的比赛，但每个巡回赛也有相应的WTA比赛和卫星级别的赛事。

罗杰费德勒d. 拉斐尔6月43日6日6月13日6日3



5. 3. 2 • 斯坦尼斯拉斯 • 瓦林卡vs。诺瓦克 • 德约科维奇2015年法网公开赛决赛

德约科维奇面临着多年的压力，完成了多年来的职业生涯大满贯，2冠军终于出现在他的手中，他击败了纳达尔，然后9次法网公开赛

冠军，正在前往2015年决赛的途中。然而，瓦林卡在这一天有其他的计划，扭转比赛，以坚定的决心击败了德约科维奇。

斯坦尼斯拉斯瓦林卡d. 诺瓦克 • 德约科维奇4-66-46-36-4

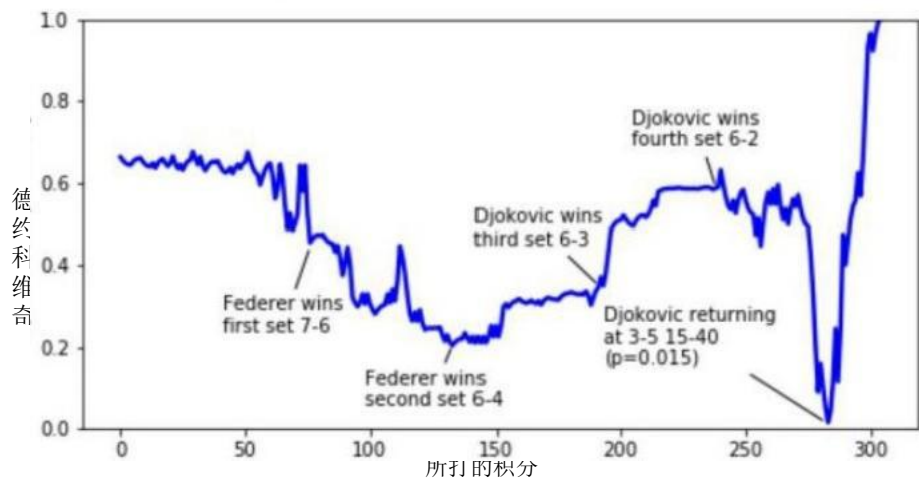


这需要赢得一次大满贯：澳大利亚公开赛，法国公开赛，温布尔登，美国公开赛。

5. 3. 3诺瓦克德约科维奇vs。罗杰·费德勒，2011年美国网球公开赛半决赛

在网球史上最著名的复出之一中，德约科维奇从两盘的落后中恢复过来，在最后一盘费德勒的发球局以3-515-40的比分落败。³

诺瓦克德约科维奇d。罗杰·费德勒6-74-66-36-27-5



³如果考虑到德约科维奇在第五盘以3-515-40的比分打出全胜或全无的决定，我们可以计算出他获胜的胜率甚至低于 $p=0.015$ ，就像费德勒那样
臆测<https://www.theguardian.com/sport/2011/sep/11/us-open-2011-federer-djokovic>). 由于杰夫·萨克曼的比赛图表项目和最近对鹰眼跟踪数据的数据分析，基于投篮选择在积分内计算比赛获胜概率是现实的。

参考书目

Bamett, T. and Clarke, S. R. (2005). 结合网球选手的统计数据来预测网球比赛的成绩
比赛IMA管理数学杂志, 16(2): 113-120。

Bamett, T. J., Clarke, S. R. 等人 (2002)。使用微软的excel来模拟一场网球比赛。在第六
关于体育领域的数学和计算机的会议, 第63-68页。澳大利亚昆士兰州:
邦德大学

Bamett, T. J. et al. (2006). 具有特定参考的层次游戏中的数学建模
网球。博士论文, 史文本理工大学。

Bevc, M. (2015). 根据逐点的数据来预测网球比赛的结果。

比克, 莫里斯, B. 和Boice, J. (2016). 我们是如何预测2016年的u. s. open. [http://的
fivethirtyeight.com/features/how-were-forecasting-the-2016-us-open/](http://fivethirtyeight.com/features/how-were-forecasting-the-2016-us-open/). 已访问的内容:
2017-10-30.

伯克, 学士 (2014)。赢得概率和wpa. [http://www.advancedfootballanalytics.com/
index.php/home/stats/stats-explained/win-probability-and-wpa](http://www.advancedfootballanalytics.com/index.php/home/stats/stats-explained/win-probability-and-wpa). Accessed:2017-
10-30.

埃弗龙, B. 和Moris, 中国国立台湾大学 (1977)。统计学中的斯坦悖论。弗里曼。

Elo, A. E. (1978). 棋子玩家的等级, 过去和现在。Arco酒吧。

黄, X, 诺顿带, W., 和布拉德利, J. (2011)。从内部推断网球比赛的进展

玩是很困难的。伦敦帝国理工学院，南肯辛顿校区，
伦敦，SW72AZ。

Klaassen, F. J. and Magnus, J. R. (2003). 预测一场网球比赛的获胜者。欧洲运筹学杂志，148(2)
： 257-267。

Klaassen, F. J. G. M. and Magnus, J. R. (2001). 网球点是独立的和同分布的？来自动态二元面板数
据模型的证据。美国人的朱玛尔
统计学协会，96(454)： 500-509。

Kovalchik, S. A. (2016). 寻找网球胜利的预测。定量杂志
体育领域的分析，12(3)： 127-138。

刘易斯-克劳斯，G. (2016)。伟大的爱伊的觉醒。《纽约时报》杂志。

锁上, D. and Nettleton, D. (2014)。之前使用随机森林来估计获胜概率
每一场nfl的比赛。《体育定量分析的判断》，10(2)： 197-205。

Madurska, A. M. (2012). 一种预测专业人员研究结果的逐组分析方法
单打网球比赛。伦敦帝国理工学院计算机技术系。棱纹平布

Sipko, M. and 钮扣带, W. (2015)。机器连接的专业预测
网球比赛。

斯坦伯格，医学博士(2017年)。为什么espn会使用那些让一些棒球迷发疯的比赛内获胜概率数
据。[https://www.washingtonpost.com/news/dc-sports-bog/wp/2017/04/23/
为什么要espn-uses-those-in-game-win-probability-stats-that-drive-some-baseball-fans-nuts/
?utm_term=.8163323a5ee2](https://www.washingtonpost.com/news/dc-sports-bog/wp/2017/04/23/为什么espn-uses-those-in-game-win-probability-stats-that-drive-some-baseball-fans-nuts/?utm_term=.8163323a5ee2). Accessed: 2017-10-30.