马龙—樊振东乒乓球实战数据分析

摘要：

我们将乒乓球实战的比赛过程设定为一连串状态的转换。两位选手每击出一拍，就表征一次状态转换。

我们设定在特定情况下，某些击球策略是有优势的。在每个状态上选取具有优势的击球策略，优势逐渐累积，最终可以得分。

我们认为可以为每个情况下的每个击球策略设定一个对于胜率的影响值。我们调整这个影响值，使得最后得分时的胜率最高。这时候，我们认为这个影响值是有代表性的。

最后，我们利用这个影响值，尝试反推并解释比赛中的若干现象。

关键词：极大似然法、线性规划。

**模型与实验方法：**

**模型假设：**

1. 将对手的击打技术与落点位置的组合**定义**为当前的"**情景**"。面对某一种“情景”时，己方也会采取某种击打技术将球击打到某个落点，将己方为应对“情景”而作出的击打技术和落点位置的组合**定义**为“**击回选择**”（例如，马龙发球的击打方式和落点位置是樊振东所要面对的“情景”，樊振东对该“情景”所作的“击回选择”是接下来马龙要面对的“情景”，马龙对这一“情景”所作的“击回选择”是樊振东下一次要面对的“情景”，依此类推）。

假设在面对每一种情景时，两位选手均会以固定的概率选择一种击回选择将球击回。概率较高的击回选择即为该选手所习惯于使用的技术。我们假设这种习惯以及概率不会改变。

1. 用取值范围为[0,1]的“对局评分”描述选手在应对某种情景时赢球的几率（注意，对局评分不是赢球概率）。对局评分越接近0，输球几率越大。反之，越接近1，赢球几率越大。
2. 假设在每种情景下，各选手作出的每一种击回选择均会对当前对局评分产生固定的影响。即：多次给定一种情景M，只要采用一定的击回选择N，就会使当前的对局评分加一个固定的值Q。如果N是有利的击回选择，则Q为正值。如果N是不利的击回选择，则Q为负值。
3. 假设各个选手的每次发球都能给自己带来 0 ～ 1 不等的对局评分。我们假设这是定值，不随时间变化或者选手心态而改变。

**数学描述：**

对于樊振东：

发球时，共有“顺旋转”、“逆旋转”等四种旋转方式，“正手短”、“正手长”等六种落点位置。因此共有24种组合方式，定义为“发球方式”。故用一个长度为24的列向量 Fstart描述樊振东采用每种发球方式时的固有胜率。

接球时，分两种情况：接对方的发球或接对方击回的球。

（i）接发球时，马龙有24种发球方式，根据统计，樊振东有34种击回选择。因此用大小为 24 \* 34 的矩阵 Fjie描述樊振东在马龙的各种发球方式下采用各种击回选择时对于己方胜率的影响。

（ii）接击回的球时，马龙有35种击回选择，即樊振东需要应对的情景有34种。对于每种情景，樊振东都有34种击回选择。因此用大小为 35 \* 34 的矩阵 Fback描述樊振东击回各种回球对于己方胜率的影响。

对于马龙，类似地：

有长度为24的列向量 Mstart，用于描述马龙每种发球的固有胜率。

有大小为 24 \* 34 的矩阵 Mjie，用于描述马龙接各种发球对于己方胜率的影响。

有大小为 34 \* 35 的矩阵 Mback，用于描述马龙击回各种回球对于己方胜率的影响。

在实际计算时，我们可以根据选手的倾向适度减小矩阵的尺寸，例如消除掉原始数据中没有出现的发球方式和击回选择等。

**模型建立：**

我们已经假设，每位选手在各个情景下的各种击回选择都会对胜率造成固定的影响。我们将这些影响设为未知量。

我们已知，在这些影响下，两位选手通过来回击球，最终会有一方得分，一方失分。

我们将问题转化为：求出这些未知量，使得最终得分时，己方的胜率最高；失分时，己方的胜率最低。

例如，当樊振东发出a类发球，马龙b类接回，樊振东与马龙分别以c，d，e击打，樊振东得分。则我们规定，最后樊振东的获胜几率P=Fstart[a]-Mjie[a][b]+Fback[b][c]-Mback[c][d]+ Fback[d][e]

则我们天然地有如下约束：

0≤Fstart[a]≤1

0≤Fstart[a]-Mjie[a][b]≤1

0≤Fstart[a]-Mjie[a][b]+Fback[b][c]≤1

0≤Fstart[a]-Mjie[a][b]+Fback[b][c]-Mback[c][d]≤1

0≤Fstart[a]-Mjie[a][b]+Fback[b][c]-Mback[c][d]+ Fback[d][e]≤1

我们需要在已知的比赛记录上，使得最后的平均获胜几率P最大，则可以求得一组参数。这组参数即是我们所认为的两位选手在各种场景下作出应对的优劣评价。

**实验方法**

我们可以在建立模型之后，通过对比球员评分较高的击球选择与球员倾向的击球选择，科学地分析球员的失误在哪里，以及劣势/优势是怎样逐渐累积的。

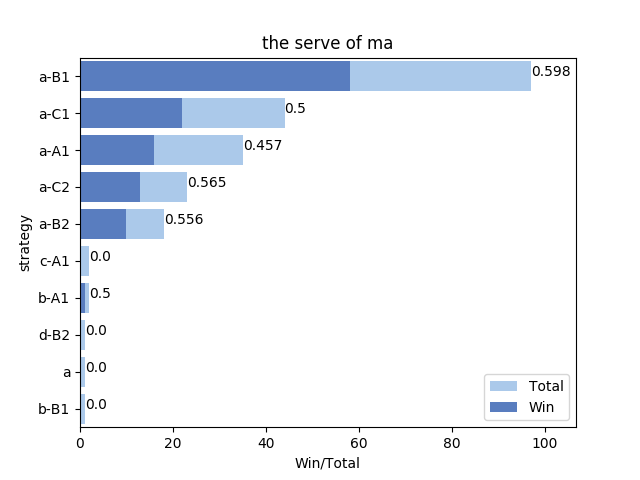
我们可以通过已经获得的击球选择的评价，去模拟一次比赛，以推算出较好的战术选择。

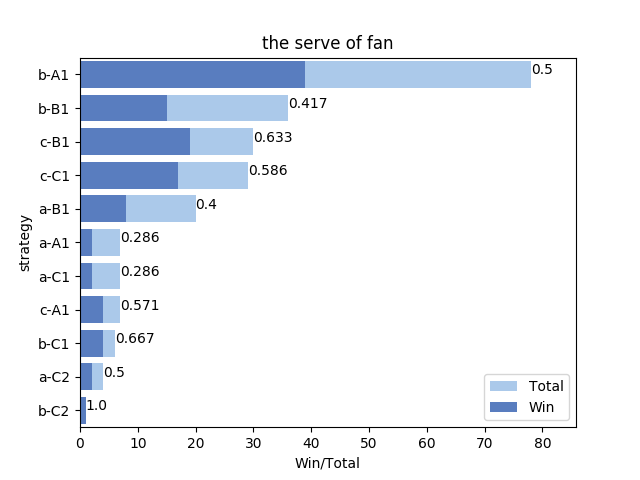
我们使用全部5场比赛建立约束条件，用除2017年全运会男单决赛外的数据建立优化目标，求解上述线性规划，再回推2017年全运会男单决赛的对局。对每一分球，在每次击球都可以计算分数，如果大于0.5，则樊振东有希望获胜，否则马龙有希望获胜。

可以计算出，在最后一次有效击球（除出界、下网外）时，模型的预测准确率为0.7849，在倒数第二次有效击球时，模型的预测准确率为0.7097。这意味着模型已经获取了对局的有效信息，我们可以用这一模型来辅助分析两名球员。

我们也不局限于此模型，还使用了图表分析这五场比赛。

统计两位球员采取不同发球方式时的得分情况，结果如下：





以上数据可大致反映两球员的**发球习惯**：马龙最习惯采用“顺侧旋”的方式发球，落点大部分在对手中间短的位置。而樊振东最常用“逆侧旋”方式发球，落点大部分在对手中间短的位置。

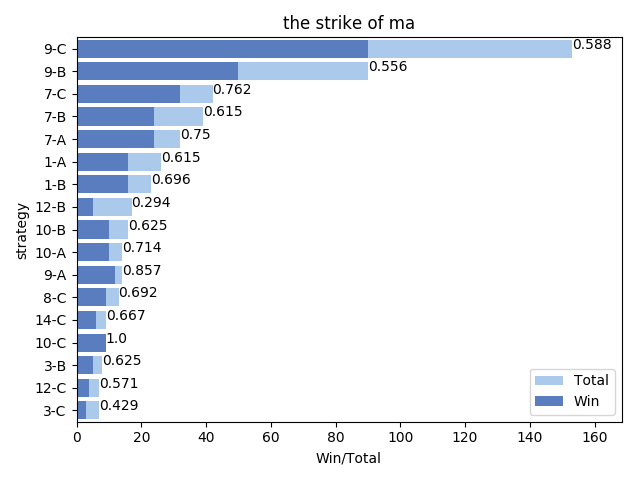
比较各种发球方式的**胜率**可知，马龙采用“顺侧旋”向对手“中间短”的位置发球胜率最高，达59.8%，这同时也是他最习惯采用的发球方式。而樊振东采用“转不转”方式向对手“中间短”的位置发球胜率最高（虽然他采用“逆侧旋”方式向对手“反手短”位置发球时胜率更高，但由于他采用这种方式发球的次数过少，结果的偶然性较大，因此不计这组数据），达63.3%，这是他比较习惯采用的发球方式，但不是最习惯的。

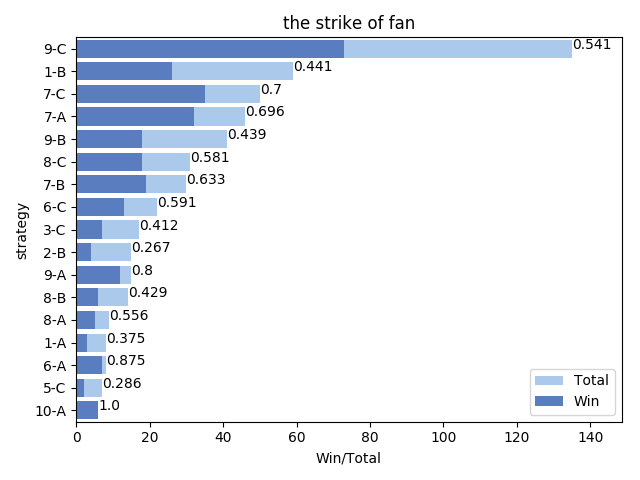
因此，从发球胜率的角度看，马龙的**最优发球策略**是采用“顺侧旋”向对手“中间短”的位置发球，而樊振东的最优发球策略是采用“转不转”方式向对手“中间短”的位置发球。

在接发球时，两名球员也有一定策略（略去少于5次的打法）：

应对马龙擅长的“顺侧旋中间短”，樊振东多采用反手拧拉落到反手，但效率其实不高，胜率要低于50%。我们可以注意到樊振东较少采用的正手摆短到正手对应“顺侧旋中间短”可能有一定效果，但要避免正手摆短到中间。我们还能注意到樊振东接发球失误较多，建议能针对马龙常用的发球策略进行练习。

应对樊振东的“转不转中间短”，马龙尝试了很多打法，但只有他最常用的正手摆短到中间达到5次，效果却不好。这意味着马龙需要在应对此种方法上多加练习，在有结果之前樊振东也能通过攻击这一弱点有效得分。此外，马龙的反手摆短到中间在接发球时效果比较好，也许他可以在这一技术上多家磨练，以适应更多的发球情况。

随后我们可以看一下两名运动员的击球特点：

两人的共同杀招是反手快撕到反手和正手拉到三个落点，胜率均大于50%。马龙的杀招比较丰富，还有反手快撕到正手和正手摆短到正手或中间。樊振东的杀招还有反手拉到反手。对比而言，马龙的技术更加全面，樊振东还有一些短板，比如常用的正手摆短效果不佳，在训练时应该多加练习。我们也能从整体上看出，两人都是进攻型选手，防守的招法不多。

我们的模型有Fback和Mback数组，他们的结果对每一击都给出了分数，可以用于指导选手应对对方的杀招。

对樊振东擅长的反手拉到正手，模型认为马龙采用正手拉到正手或反手会比较好（给分均接近1）。对付樊振东的反手快撕到反手，模型认为正手拉到三个落点或反手防御会比较好。对樊振东的正手拉，模型发现马龙用过4次的反手切到反手效果不佳(给分-0.2317)，而应采用对拉、正手快带或反手防御等手段。

对马龙擅长的反手快撕，模型认为樊振东应该采用正反手拉或者反手快撕（给分接近1）。对付正手摆短，模型给出了许多建议，包括正手摆短、反手摆短、正手劈长、正手挑打、反手拧拉、正手拉、反手拉，这可能是因为樊振东尝试过这么多应对策略，模型认为这些策略都有可能是好的方案，需要樊振东多实践。对马龙的正手拉，模型也发现樊振东使用过的正手防御效果不佳(-0.3894分)，建议使用对拉或者反手快撕的策略。

**总结**

我们用统计和线性规划两套模型，分析了马龙和樊振东两位乒乓球选手，时间有限，考虑难免有不周之处，请读者不吝赐教。

**附录**