[**数学之美：Xbox评分系统TrueSkill**](http://www.biaodianfu.com/xbox-trueskill.html)

Written by [标点符](http://www.biaodianfu.com/author/admin) on 2012年07月24日 in [程序设计](http://www.biaodianfu.com/category/web-development/programming)

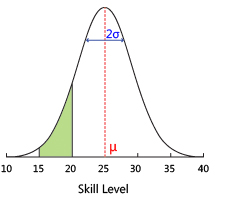
在电子竞技游戏中，特别是当有多名选手参加比赛的时候需要平衡队伍间的水平，让游戏比赛更加有意思。这样的一个参赛选手能力平衡系统通常包含以下三个模块：

* 一个包含跟踪所有玩家比赛结果，记录玩家能力的模块。
* 一个对比赛成员进行配对的模块。
* 一个公布比赛中各成员能力的模块。

事实上目前已经有的游戏评分系统是[Elo评分](http://www.biaodianfu.com/fackbook-facemash-algorithm.html)，但是[Elo评分](http://www.biaodianfu.com/fackbook-facemash-algorithm.html)仅只是两名选手参加的游戏。TrueSkill系统是基于[贝叶斯推断的评分系统](http://www.biaodianfu.com/imdb-rank.html)，由微软研究院开发以代替传统[Elo评分](http://www.biaodianfu.com/fackbook-facemash-algorithm.html)，并成功应用于Xbox Live自动匹配系统。TrueSkill评分系统是[Glicko](http://en.wikipedia.org/wiki/Glicko_rating_system)评分系统的衍伸，主要用于多人游戏中。TrueSkill评分系统考虑到了你水平的不确定性，综合考虑了玩家的胜率和可能的水平涨落。当玩家进行了更多的游戏后，即使你的胜率不变，系统也会因为对你的水平更加了解而改变对你的评分。

**怎样进行能力计算**

TrueSkill排名系统是针对玩家能力进行设计的，以克服现有排名系统的局限性，确保比赛双方的公平性，可以在联赛中作为排名系统使用。它为玩家排名使用的为 [贝叶斯定理](http://zh.wikipedia.org/wiki/%E8%B4%9D%E5%8F%B6%E6%96%AF%E5%AE%9A%E7%90%86)。 系统的特点是假设每一个玩家的能力不是固定的，其能力水平的表现为一个钟型曲线（[正态分布](http://zh.wikipedia.org/wiki/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83)或高斯分布）。



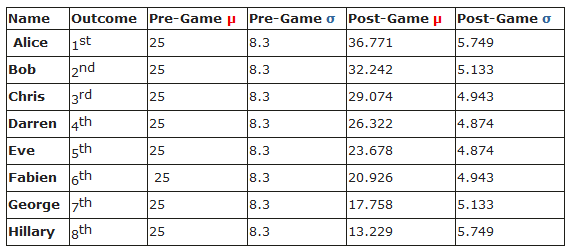
绿色区域15～20代表了Ranking System对的评分。可以看出系统的评分是比较保守的。**σ**越小则越靠近**μ** ，相应的玩家的能力水平就较高。总的来说玩家的水平受“平均得分”和“玩家稳定性”综合影响。

由于TrueSkill排名系统使用高斯信仰分布来描述一个玩家的能力，也就意味着玩家的能力始终落在4倍的**σ**内（概率为99.993666%）。从微软追踪的65万玩家数据显示，有99.99%都落在了3倍的**σ**内。 有趣的是，TrueSkill排名系统可以使用1作为最初的不确定性做所有的计算，将相乘**μ**和**σ**可以缩放到任何其他的范围。假设所有的计算都以初始值μ=3和σ=1，如果一个玩家有50级，几乎所有的**μ**的发生是在±3倍的初始**σ，σ**可得50/6 = 8.3。 两个玩家最大的区别在于**μ**值得大小。假设**σ**相当，那么**μ**高的玩家赢得机会就越大，这一原则也适用在TrueSkill排名系统。但并不表示**μ**高的就一定会赢。在单个的配对比赛中，玩家的个人表现与玩家的能力是相当的，游戏结果也是有个人表现决定的。因此，可以认为能力的一个玩家在TrueSkill的排名是在大量游戏中的平均表现。个人表现的变化原则是能力表现的一个参数。

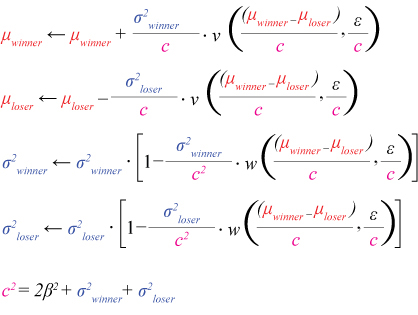
**怎样更新能力值**

TrueSkill排名系统只会根据比赛结果更新**μ**和**σ**，它假设的情况为一个玩家的表现围绕着他的能力水品进行变化，如果一个玩家玩一个基于点数的游戏，他战胜了所有的其他10个对手和他和战胜了另外一场比赛只有一个对手的积分是一样的，但是这样两场比赛确实反映了不同选手的能力情况。通常会使**σ**下降。在计算一场新的比赛结果之前，TrueSkill排名系统会计算比赛的排名与选手在比赛前的排名的变化情况。排名的变化最终影响了玩家技能的不确定性**σ**。这个参数可以被TrueSkill用来记录玩家的技能的变化。并且σ永远不可能为0。

下面这张表格来自微软研究院，此表格给出了8个新手在参与一个8人游戏后**μ**和**σ**的变化。



这里有个很有意思的现象：注意第四名Darren和第五名Eve，他们的**σ**是最小的，换句话说系统认为他们能力的可能起伏是最小的。这是因为通过这场游戏我们对他们了解得最多：他们赢了3/4个人，也输给了4/3个人。而对于第一名Alice，我们只知道她赢了7个人。 如果想知道更详细的定量分析可以先考虑最简单的两人游戏情况



在上述的方程式中，唯一未知的就是选手的表现。另外还有就是游戏的模式。系数**β2**代表的是所有玩家的平均方差。 **v**(.,.) 和**w**(.,.) 是两个函数，比较复杂。**ε**是个与游戏模式有关的参数。 简而言之，你赢了\mu 就增加，输了\mu 减小；但不论输赢，\sigma都是在减小，所以有可能出现输了涨分的情况。

**怎样进行选手匹配**

势均力敌的对手能带来最精彩的比赛，所以当自动匹配对手时，系统会尽可能的为你安排可能与水平最为接近的玩家。TrueSkill评分系统采用了一个值域为(0,1)的函数来描述两个人是否势均力敌：结果越接近0代表差距越大，越接近1代表水平越接近。 假设有两个玩家A和B，他们的参数为(**μA**,**σA**) 和 (**μB**,**σB**)，则函数对这两个玩家的返回值为 

c的值由如下公式给出 http://www.biaodianfu.com/wp-content/uploads/2012/07/c.png

如果两人有较大几率被匹配在一起，光是平均值接近还不行（e指数上那一项），还得方差也比较接近才行（d）。

**怎样创建能力排行榜**

TrueSkill假设玩家的水平可以用一个正态分布来表示，而正态分布可以用两个参数：平均值和方差来完全描述。设Rank值为R，代表玩家水平的正态分布的两个参数平均值和方差分别为**μ** and **σ**，则系统对玩家的评分即Rank值为 **R=μ-k\*σ** k值越大则系统的评分越保守。在Xbox Live上，系统为每个玩家赋予的初值是μ = 25 以及 σ = 25 / 3，k=3。所以玩家的起始Rank值为R=25-3\*25/3=0。