新提出的评价指标： PBG (Position Biased Gain)，关于指标的名字，刘老师有没有更好的建议？

基本的思路是，考虑在用户浏览的长度上收益的累积，以高度（height）作为衰减的因素

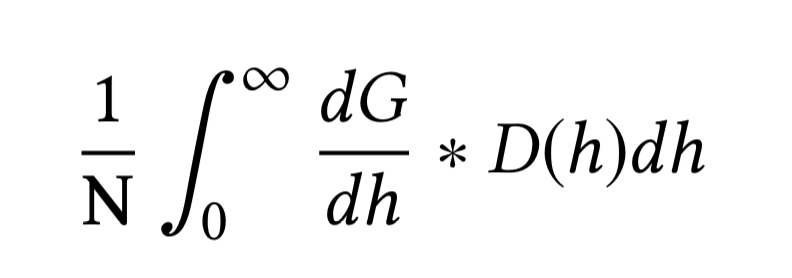
// Sakai建议用高度（height）,而不是长度（length），长度容易误解成文本的单词数等等

**主要面临的三个挑战：**

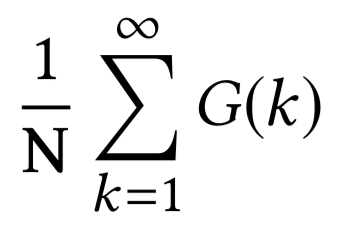
1. 在异质化的移动搜索结果中，不同的结果在屏幕上所占的大小可能差异很大，用户获取信息的成本发生了变化。这与传统的Rank-based 评价指标的基本假设是不一致的。我们需要考虑结果本身的差异。
2. 在异质化的搜索结果中，用户和搜索结果的交互方式发生了变化，有很多结果用户不需要点击，就可以获得收益，甚至有的搜索结果是没有外链的。传统的指标基于ten blue links并没有考虑这样的情况。
3. 在移动设备上，由于屏幕尺寸、交互方式的影响，用户阅读信息更加困难，因此在桌面设备上用户“看到结果即获得收益”的假设不再稳固。我们需要考虑Gain在SERP上的snippet、以及landing page上的收益，采用用户期望浏览长度作为衰减因素。

**指标设计：**

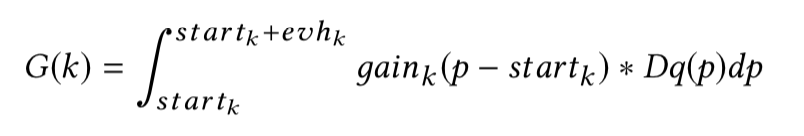
基本的评价指标形式：



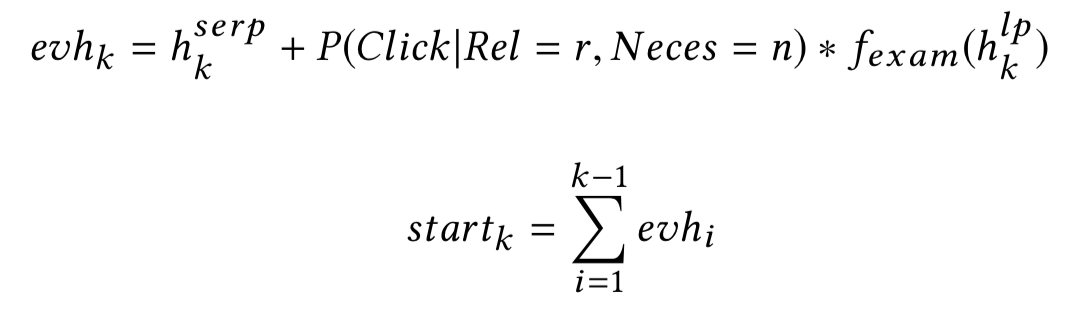
如果考虑在每一个结果上的收益，上面的式子等价于



其中表示在第k个结果上的收益，可以写成



evh（Expected Viewed Height）预期浏览长度的计算方式为：



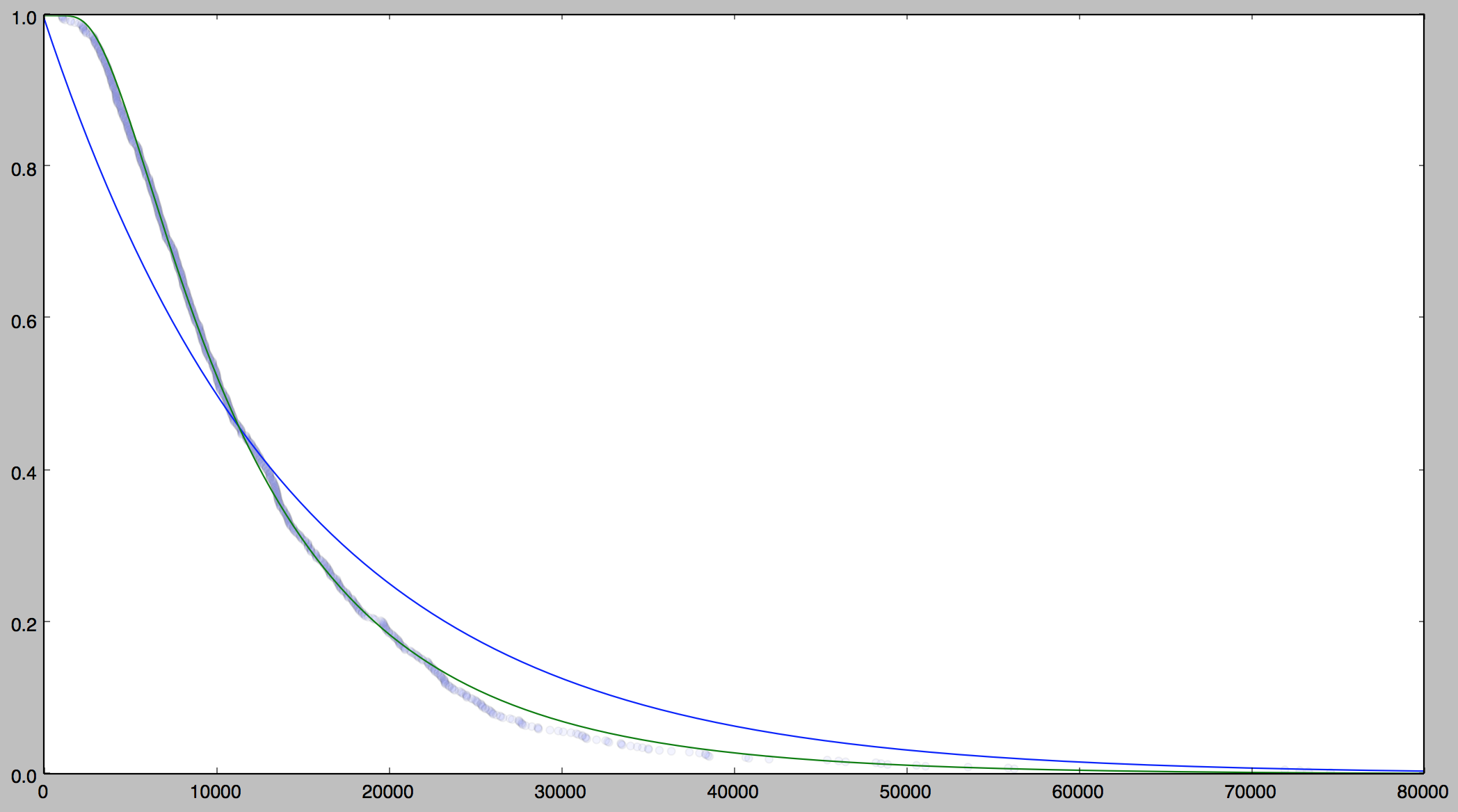
用serp上的高度加上P\*f\_exam(h^lp). 这背后的假设是用户一定会浏览serp上的snippet，并且以一定的概率点开landing page，这里的f\_exam是用户在landing page上的浏览模型，我们的指标中认为如果用户点开了，就会看到底（张帆的用户行为实验提供了数据支持，avg(用户最深浏览位置/页面长度)> 0.9）。

P(Click| Rel = r, Neces=n)是根据张帆的用户行为实验数据估计出来的：

|  |  |  |  |
| --- | --- | --- | --- |
| Possibility | N=1 （必须点击） | N=2（可能需要点击） | N=3（几乎不需要点击） |
| R=1 | 0.403 | 0.067 | 0.093 |
| R=2 | 0.438 | 0.313 | 0.040 |
| R=3 | 0.607 | 0.500 | 0.147 |
| R=4 | 0.884 | 0.757 | 0.647 |

表示结果的Gain在第k个结果上的分布，在之前计算的指标上，我把这个分布算成是在整个evh上的一个均匀分布；改进后的结果，根据参考文献上认为通常用户在snippet和landing page上花费的实际比为4:6，所以新的分布是40%的gain均匀分布在evh中的snippet段，60%的均匀分布在landing page段。

对于Decay function，我们采用Inverse\_Gaussian\_distribution，利用L1-norm 拟合，残差约为采用指数形式的1/10，可以看到新的decay function更好的fit了用户离开的概率。



基于目前的指标，在SBS上的一致性超过了传统的rank-based metrics（这里面也应该包含一些对传统指标的讨论）。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Metrics Name | #Agr | #Dis | RATE | Metrics Name | #Agr | #Dis | RATE |
| PBG\_HCS | 128 | 22 | 85.33% | NCUgu,P | 91 | 59 | 60.67% |
| PBG\_TBG | 120 | 30 | 80.00% | NPBG\_TBG | 84 | 66 | 56.00% |
| Q@0003 | 113 | 37 | 75.33% | ERP | 80 | 70 | 53.33% |
| Q@0005 | 113 | 37 | 75.33% | NPBG\_HCS | 78 | 72 | 52.00% |
| MSnDCG@0003 | 111 | 39 | 74.00% | Q-measure | 77 | 73 | 51.33% |
| nERR@0003 | 109 | 41 | 72.67% | AP@0005 | 73 | 77 | 48.67% |
| nERR@0005 | 108 | 42 | 72.00% | NCUrb,P | 72 | 78 | 48.00% |
| nDCG@0003 | 107 | 43 | 71.33% | AP@0003 | 71 | 79 | 47.33% |
| P-measure | 102 | 48 | 68.00% | AP | 71 | 79 | 47.33% |
| ERR | 102 | 48 | 68.00% | P@0005 | 70 | 80 | 46.67% |
| MSnDCG@0005 | 102 | 48 | 68.00% | P@0003 | 70 | 80 | 46.67% |
| P-plus | 101 | 49 | 67.33% | RR | 58 | 92 | 38.67% |
| NCUgu,BR | 100 | 50 | 66.67% | RBP | 56 | 94 | 37.33% |
| nDCG@0005 | 99 | 51 | 66.00% | Hit@0003 | 48 | 102 | 32.00% |
| O-measure | 98 | 52 | 65.33% | Hit@0005 | 45 | 105 | 30.00% |
| NCUrb,BR | 92 | 58 | 61.33% |  |  |  |  |

流程图把user model画出来；