

新榜指数NRI 算法说明

Powerby newrank.cn

版本号 3.4

2018年11月08日



目录

什么是新榜指数	3
算法简述	3
算法详解	4
1、选定指标	4
2、指标标准化	4
3、指标加权	5
算法所使用数据的统计口径	6
FAQ	7
1、 为何要设定常数?	7
2、 1000 分是最高分吗?	7
3、 常数设定中的点赞数的最高值为何没用 10 万, 而是 1 万?	7
4、 算法会经常变么, 如发布上限从 8 篇/天变成了 10 篇/天?	7
5、 新榜指数可以跨期比较么?	8
拓展: 两微一端的独立及跨平台应用	8
1、 微博、腾讯微社区、新闻客户端等独立应用	8
2、 双微及两微一端的应用	8

什么是新榜指数

“新榜指数” (New Rank Index, NRI): 由新媒体排行榜 (newrank.cn) 基于海量数据、用户深度反馈及专家建议而推出, 用于衡量新媒体 (主要是两微一端) 的传播能力, 此指数反映该新媒体主体的热度和发展趋势。

以下算法主要针对微信进行阐述。

算法简述

新榜指数是由原始数据参照基数通过计算公式推导出来的标量数值, 用以衡量原始数据在其所属维度的相对表现。例如, xx 点赞数为: 379, 我们会将这一数据和点赞常数进行比较, 以确定其点赞数的相对位置, 进而转换为其点赞指数, 其余指标以此类推。

该算法相较于常规加权有四大优势: 1、不仅显示绝对表现, 还能反映所处位置; 2、标准化后, 不同维度的指标之间可以相互比较; 3、指标之间可以相互运算, 从而得到综合考察各维度的新榜指数; 4、形成的指数为单调增函数, 不仅可以用于用户间的比较, 也可以用于用户自身的跨期比较。

算法详解

1、选定指标

在微信的计算中，为兼顾其数量与质量，我们使用总阅读数、最高阅读数、平均阅读数、头条阅读数、总点赞数五个指标来对账号进行评估，其中：

- 整体指标（总阅读数， R ）：指统计周期内所有发布内容的阅读数总和；
- 优异指标（最高阅读， R_m ）：是指统计周期内所有发布内容中的单篇最高阅读数；
- 质量指标（平均阅读数， R_a ）：统计周期内所有发布内容的阅读数平均值；
- 主动预判指标（头条阅读， R_h ）：统计周期内位置为头条的内容的阅读总数；
- 互动指标（总点赞数， Z ）：统计周期内所有发布内容的点赞数总和。

2、指标标准化

基于不同维度指标的差异性，我们首先将所得数据标准化，即通过对各维度原始数据与该维度常数的比值计算，得到该样本各项指标的干分化数值。

在新榜指数中，以天为最小计算周期，将常规（1天推送一次，每次最多8篇）的公众号为对象在这一时间段内总阅读数、最高阅读数、平均阅读数、头条阅读数和点赞数共五个指标的常数分别记为 \bar{R} , \bar{R}_m , \bar{R}_a , \bar{R}_h , \bar{Z} , \bar{R} 。

结合新榜历史数据样本库及微信平台的发布规则——1个常规的微信公众号每天最多发布8篇，每次单篇阅读数最高为10万，故而我们记录80万为“总阅读数”这一指标理论上的1000分，对于某些特别账号——例如每天可以推送多次的人民日报、央视新闻，则理论上，在该

指标上可以获得比 1000 分更高的得分。

故而

$$\bar{R} = 8 * 100000 = 800000, \bar{R}_m = 100000, \bar{R}_a = 100000, \bar{R}_h = 100000, \bar{Z} = 8 * 10000.$$

在不同的时间周期，即日、周、月、年 (n=1,7,30,365) 下，以上常数具体数值为：

$$n * \bar{R}, \bar{R}_m, \bar{R}_a, n * \bar{R}_h, n * \bar{Z}$$

在此基础上，各指标按以下公式进行标准化：

$$R' = \frac{\ln(R+1)}{\ln(n*R+1)} \times 1000, R'_{max} = \frac{\ln(R_m+1)}{\ln(\bar{R}_m+1)} \times 1000, R'_a = \frac{\ln(R_a+1)}{\ln(\bar{R}_a+1)} \times 1000, R'_h = \frac{\ln(R_h+1)}{\ln(n*\bar{R}_h+1)} \times 1000,$$

$$Z' = \frac{\ln(Z+1)}{\ln(n*\bar{Z}+1)} \times 1000$$

注：此处"+1"是为了避免出现类似“0 阅读”时而产生无效值

3、指标加权

由于此五项数据影响微信传播能力的程度不同，我们赋予不同的权重，分别记为

$\omega_R, \omega_{R_m}, \omega_{R_a}, \omega_{R_h}, \omega_Z$ 。最终，新榜指数的计算公式表述为：

$$NRI = \omega_R R' + \omega_{R_m} R'_m + \omega_{R_a} R'_a + \omega_{R_h} R'_h + \omega_Z Z'$$

目前，我们使用的系数为 $\omega_R = 0.75, \omega_{R_m} = 0.05, \omega_{R_a} = 0.1, \omega_{R_h} = 0.05, \omega_Z = 0.05$ 。

新榜始终相信，不同类别之间的简单比较是没有意义的，所以，在发布需要跨类别比较的月

榜单“中国微信 500 强”时，我们会赋予类别系数，彰显渠道价值。具体系数表如下：

类别	系数	类别	系数
时事	1	文化	0.99
民生	0.98	百科	0.98
财富	0.99	健康	0.98
科技	0.98	时尚	0.98
创业	0.98	美食	0.98
汽车	0.98	乐活	0.98
楼市	0.98	旅行	0.98
职场	0.98	幽默	0.937
教育	0.98	情感	0.98
学术	1.02	体娱	0.98
政务	1	美体	0.98
企业	0.98	文摘	0.92

算法所使用数据的统计口径

日榜的统计周期是所示日期 0 时至 24 时间，统计截止时间是次日中午 12 时，以确保子夜发布者亦可得到至少 12 小时的传播。周榜的统计周期是所示日期范围（周一 0 时至周日 24 时），统计截止时间是下一个周一的中午 12 时。周榜并非日榜数据的简单加总，而是重新更新并统计。月榜、年榜亦然。

需要特别说明的是：根据腾讯官方政策，微信影响力榜单所列微信阅读数仅指前台显示数据，且超过 10 万者一律显示为 10 万+，与微信管理后台所显示数据可能不一致。

FAQ

1、为何要设定常数？

正如我们将“1 米”定义为平面电磁波（光）在“1/299,792,458 秒”的持续时间内在真空中传播行经的长度一样。我们基于理论值和实际观测值设定了各个指标的相对参考值，用来对各个指标进行归一化处理。

2、1000 分是最高分吗？

不是。但是 1000 分可以视为一个常规号（一天发布 1 次）的能够达到的极限——发满 8 篇，每篇阅读数 10 万，点赞 1 万。对于少数公众号，例如人民日报、央视新闻则存在理论上超越这一得分的可能；

3、常数设定中的点赞数的最高值为何没用 10 万，而是 1 万？

在实际的观测中（超过 1000 万个样本），新榜的技术人员观测到过超过 10 万的单篇点赞数，但是这是属于极其特别的个例，通过测试样本分布（单篇点赞超过 1 万的样本比例远小于万分之一），我们选定 1 万作为理论上限；

4、算法会经常变么，如发布上限从 8 篇/天变成了 10 篇/天？

新榜指数的核心价值是为所有公众号提供了一个参照系，只要这一个参照系是稳定的，

那么比较就是有意义的。一个稳定的体系更有利于长期的观察和度量；

类似微信策略的变化只会使得总阅读一项指标变得更容易得到 1000 分，甚至超越这一分数，但是并不对这一算法有着决定性的影响。

新榜技术人员会对数据环境进行长期跟进，如果确有极大的影响，会及时优化调整算法版本，并通过公开渠道统一说明；

5、新榜指数可以跨期比较么？

可以。新榜指数为单调增函数，只要时间维度一致（日、周、月、年等），公众号之间、自身与自身都是可以比较的。

拓展：两微一端的独立及跨平台应用

1、微博、腾讯微社区、新闻客户端等独立应用

通过标准化指标使得转发、评论、PV、UV、楼中楼回复等变得独立可比较的指标，每一个指标均反映该对象在全部样本中的独立表现，进行加权得到独立应用场景下的新榜指数；

2、双微及两微一端的应用

与独立应用类似，跨平台的新榜指数分别考察某一对象在单一平台上的相对表现，最后在使用平台系数进行加权从而得到综合后的新榜指数（如微博+微信）。