DEEPLINK 技术报告 · 2026 年 11 月

小红书搜索情境下的声量占有率与群体认同：度量方法

Measuring Share of Voice and Group Affinity in Xiaohongshu Search

深通国际 数据科学组

深通（杭州）国际教育科技有限公司 · DeepLink International

摘要　中国学生选择德国语言学校时，主要通过小红书的搜索结果及其评论区获取信息。搜索结果因时间和用户而异，单次观察不可复现；笔记之间、评论之间的影响力也相差悬殊。本文给出一套可重复的度量方法：以多轮搜索的折半一致性判定样本是否稳定，以名次折扣与互动热度定义笔记权重，并在此基础上定义三项指标，即信息供给率、声量占有率与群体认同。三项指标均为有界的无量纲比例，对刷屏和个别爆款内容不敏感，可在不同搜索词和不同时间之间比较。

关键词　声量占有率；群体认同；折半一致性；名次折扣；社交媒体搜索

1　问题与记号

给定一个搜索词，对其重复搜索 K 轮，每轮记录前 60 个结果。记笔记 n 在第 r 轮的名次为 $p_{n,r}$，未出现时其名次折扣取 0。稳定出现且与行业相关的笔记构成样本集 N，其中由本方账号发布的子集记为 N°。笔记 n 下的评论集合记为 C_n。被讨论的对象（学校）记为 b。每一次对学校的提及带有一个情感极性标注 $\sigma \in [-1, 1]$，−1 为明确否定，1 为明确认可。

2　抽样稳定性

搜索排序带有随机成分，需要先判断观察到的排序是否稳定。借用测量学中折半信度的思路 [2, 3]：把 k 轮搜索随机分成等大的两半 A、B，各自按累计名次折扣排出前 20 篇 $T_{20}(A)$、$T_{20}(B)$，以两者的重合比例作为一次折半的一致性：

$$O(A,B) = \frac{|T_{20}(A) \cap T_{20}(B)|}{\min(\,|\,T_{20}(A)\,|\,,\,|\,T_{20}(B)\,|\,)} \tag{1}$$

对 $S = 60$ 次独立的随机折半取平均（随机种子固定，结果可复现）：

$$C_k = \frac{1}{S}\sum_{j=1}^{S} O(A_j, B_j) \tag{2}$$

自第 6 轮起每 2 轮计算一次 C_k。当最近 4 次取值的极差不超过容差 $\varepsilon = 0.04$ 时判定收敛并停止抽样，因此最少需要 12 轮：

$$\max_{i \in I_k} C_i - \min_{i \in I_k} C_i \le \varepsilon, \quad I_k = \{k-6, k-4, k-2, k\} \tag{3}$$

收敛值低于 0.60 的搜索词，其结果被标注为「分散」，相应指标仅供参考。未收敛的批次不计算任何指标。

3　笔记权重

名次越靠后，被看到的概率越低。采用信息检索评价中通用的对数名次折扣 [1]：

$$g(p) = \frac{1}{\log_2(p+1)} \tag{4}$$

笔记的曝光定义为各轮名次折扣的平均值：

$$E_n = \frac{1}{K}\sum_{r=1}^{K} g(p_{n,r}) \tag{5}$$

互动热度由点赞数 L_n、收藏数 F_n、评论数 M_n 加权得到，取该批次内最近一次观测值：

$$H_n = 10 L_n + 15 F_n + 20 M_n \tag{6}$$

笔记权重为曝光与热度平方根的乘积。对热度开平方，使权重随互动量次线性增长，避免个别爆款笔记支配全部结果：

$$w_n = E_n \sqrt{H_n} \tag{7}$$

4　评论可见度

评论的影响力取决于它被看到的程度。设评论 c 获得 ℓ_c 个赞，在所属楼层中是第 j_c 条回复（一级评论取 $j_c = 0$）。楼内前两条回复不折减，其后按次序衰减：

$$v_c = \frac{1 + \ell_c}{1 + 0.5 \cdot \max(0, j_c - 1)}, \quad s_c = \sqrt{v_c} \tag{8}$$

后续计算一律使用 s_c，理由与笔记权重的定义相同。

5　指标定义

5.1　信息供给率

信息供给率衡量搜索结果中有多大比重的内容由本方账号提供：

$$R = \frac{\sum_{n \in N^{\circ}} w_n}{\sum_{n \in N} w_n} \tag{9}$$

5.2　声量占有率

声量只统计被提及的多少，不区分正负。笔记正文提到学校 b 时，b 获得整篇笔记的权重；评论提到 b 时，按该评论的可见度占本篇评论总量的比例分得相应份额：

$$W_b = \sum_{n \in N} w_n \left(\mathbb{1}_{b \in n} + \frac{1}{D_n} \sum_{c \in C_n(b)} s_c \right) \tag{10}$$

其中 $C_n(b)$ 是笔记 n 下提到 b 的评论，D_n 是本篇评论可见度的归一化总量。为抑制刷屏，对在同一篇笔记下发表不少于 3 条评论、且占该篇评论数不低于 20% 的用户，其对同一学校的多次提及合并为一次，取其中最大的 s_c，在 D_n 中也只计这一个最大值。声量占有率为归一化后的份额：

$$V_b = \frac{W_b}{\sum_{\mathrm{b}'} W_{\mathrm{b}'}} \tag{11}$$

显然 $V_b \in \left[0, 1\right]$ 且 $\sum_b V_b = 1$。

5.3　群体认同

群体认同只使用评论，衡量评论区整体上认可谁。先在单篇笔记内计算。每位评论者 u 的发言权重取其在本篇中可见度最高的一条评论，并在评论者之间归一化：

$$\pi_{u,n} = \frac{\max\limits_{c \in C_n(u)} s_c}{\sum_{\mathrm{u}'} \max\limits_{c \in C_n(\mathrm{u}')} s_c} \tag{12}$$

同一评论者对同一学校的多次表态取平均，记为 $\bar{\sigma}_{u,b,n}$。学校 b 在本篇的认同净值为：

$$\nu_{b,n} = \sum_{u} \pi_{u,n} \bar{\sigma}_{u,b,n} \tag{13}$$

读者对一篇笔记的印象通常由评论区的主流意见决定。以指数 $\gamma = 3$ 对净值作锐化，使领先者获得更大的份额，同时保留符号：

$$a_{b,n}=\operatorname{sgn}(\nu_{b,n})\frac{|\nu_{b,n}|^{\gamma}}{\sum_{b'}|\nu_{b',n}|^{\gamma}} \tag{14}$$

另设两项修正。表态密度 δ_n 反映本篇评论区表态的明确程度；人数系数 $\varphi_{b,n}$ 对只有个别人提及的学校作折减，其中 $U_{b,n}$ 为本篇中提到 b 的不同评论者人数，阈值 $U_0=3$：

$$\delta_n=\min\Big(1,\sum_{b}|\nu_{b,n}|\Big),\quad \varphi_{b,n}=\min\Big(1,\frac{U_{b,n}}{U_0}\Big) \tag{15}$$

跨笔记按笔记权重汇总，再以全部学校贡献的绝对值之和归一化：

$$A_b=\sum_{n\in N} w_n a_{b,n}\delta_n\varphi_{b,n} \tag{16}$$

$$G_b=\frac{A_b}{\sum_{b'}|A_{b'}|} \tag{17}$$

因此 $G_b\in[-1,1]$，$\sum_b|G_b|=1$。$G_b<0$ 表示否定意见占优。

6　性质

有界与可比。三项指标均为比例，不依赖样本的绝对规模，可在不同搜索词、不同批次之间比较。

对极端值稳健。热度与可见度均取平方根；评论者按人计权而非按条计权；高频评论者的重复提及被合并。单篇爆款或单个用户的刷屏不能决定结果。

可复现。指标计算不含随机成分，折半检验使用固定随机种子。同一批数据重复计算结果相同。

口径对所有对象一致。合作院校与其他学校使用完全相同的公式和参数。

7　参数与样本

表 1　参数取值

符号	含义	取值
—	每轮记录的结果数	60
T_{20}	折半比较的名单长度	20
S	随机折半次数	60
ε	收敛容差（最近 4 次的极差）	0.04
—	热度系数（赞 / 收藏 / 评论）	10 / 15 / 20
—	楼内回复衰减系数	0.5
γ	认同锐化指数	3
U_0	人数系数阈值	3
—	高频评论者判定	≥ 3 条且 ≥ 20%

表 2　本站引用批次的样本概况（2026 年 11 月 8 日至 9 日）

搜索词	搜索次数	一致性	样本笔记	评论	信息供给率
德国语言班	31.5	77.4%	38	664	51.6%
德国语言学校	38.8	67.0%	35	533	39.1%
埃森语言班	17.5	79.3%	26	423	37.6%
德国语言班费用	15.8	74.3%	38	538	33.2%

8　局限

其一，数据只来自小红书一个平台。其二，每个批次是一个时间截面，不代表长期平均。其三，样本规模有限，每个搜索词约 30 至 40 篇笔记和数百条评论，排名靠后的对象提及很少，其名次不稳定。其四，合作院校的声量中包含本方账号发布的内容；我们另行计算只含第三方内容的份额作为对照。其五，对象名称按评论中的实际叫法记录，可能与学校的正式名称不同。

参考文献

1. Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4), 422–446.

2. Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271–295.

3. Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3(3), 296–322.

© 2026 深通（杭州）国际教育科技有限公司 www.deeplink-edu.com