

# 基于集体评分行为的在线用户声誉识别

陈虎杰<sup>1</sup>, 宁梦霞<sup>1</sup>, 唐高文<sup>2</sup>

<sup>1</sup>上海理工大学管理学院, 上海

<sup>2</sup>安徽新华学院外国语学院, 安徽 合肥

收稿日期: 2024年5月14日; 录用日期: 2024年7月25日; 发布日期: 2024年8月1日

## 摘要

在线评级系统中, 用户声誉是基于历史评级行为而形成的信任度, 是区分正常用户与恶意用户的关键指标。未明确定义声誉可能导致系统易受恶意攻击。因此, 清晰界定用户声誉对于有效排除恶意用户和准确评估目标质量极为重要。本文从复杂网络的视角全面回顾了用户声誉评估方法, 特别着重于不同计算模型的探讨, 包括迭代模型、基于群组的模型和基于特殊分布的模型。此外, 本文对现有文献进行了系统分析, 说明了各方法的优势和局限性, 并对未来研究方向提出展望。

## 关键词

评级系统, 用户声誉, 复杂网络, 异常用户

# Identifying Online User Reputation in Terms of Collective Rating Behaviors

Hujie Chen<sup>1</sup>, Mengxia Ning<sup>1</sup>, Gaowen Tang<sup>2</sup>

<sup>1</sup>Business School, University of Shanghai for Science and Technology, Shanghai

<sup>2</sup>School of Foreign Languages, Anhui Xinhua University, Hefei Anhui

Received: May 14<sup>th</sup>, 2024; accepted: Jul. 25<sup>th</sup>, 2024; published: Aug. 1<sup>st</sup>, 2024

## Abstract

In online rating systems, user reputation is defined as a degree of trust formed based on historical rating behaviors and serves as a critical indicator to differentiate between genuine and malicious users. A lack of precise definition for reputation can render the system vulnerable to attacks by malicious entities. Thus, clearly defining user reputation is crucial for effectively excluding malicious users and accurately assessing the quality of the targets. This article comprehensively reviews methods for assessing user reputation from the perspective of complex networks, with a particu-

lar focus on different computational models, including iterative models, group-based models, and models based on special distributions. Finally, this paper provides a systematic analysis of the existing literature, illustrates the strengths and limitations of each method, and provides an outlook on future research directions.

## Keywords

Rating Systems, User Reputation, Complex Networks, Spammers

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在线评分系统涵盖众多用户及对象，用户可以在包含电影、书籍和网站的广泛数据库中选择并进行互动。例如，用户给电影的评分会直接决定人们对该电影质量的感知[1]。然而，一些用户可能会提供高质量的评分，而其他用户则可能进行恶意评分。因此，明确定义用户声誉，以区分正常与恶意的评分，成为一个亟待解决的问题。本文旨在通过系统回顾和分析现有文献，明确各种声誉评估方法的优劣和适用场景。

为准确通过用户评级识别对象质量，我们定义用户声誉基于历史评级行为在系统中建立的信任度。由于用户声誉缺乏明确定义，评分系统易受误导和操纵[2]。恶意用户可能会误导评分系统，对低质量对象进行高分评价[3] [4]，从而误导其他用户并影响系统的整体评价[5]。显然，将用户声誉作为一个指标，有助于有效区分正常评分和恶意评分，减少误导。此外，考虑用户声誉的角色还有助于筛选异常用户，提高评价的准确性。

用户评分行为是一个多维度的现象，受物品的内在质量、个人的评分倾向、物品的流行程度以及用户对物品进行评分的频次等多种因素的综合影响。Fu 等[6]的研究表明，随着时间的推移，“高度活跃”的用户在评分时趋于更加严格，倾向于给出较低的评分。这可能意味着，随着用户在评分系统中参与度的提高，其评价标准可能会变得更加严格和精细。Dai 等[7]则强调了个人偏好在用户评分形成中的关键作用，他们指出，由于个人偏见和心理因素的影响，即使是相同的物品也可能获得不同的评分。Liu 等[8]的研究进一步揭示了用户评分行为与物品流行度之间的显著相关性。研究发现，“低度”用户更倾向于评价那些已经广受欢迎的物品，而“高度”用户则更可能选择评价那些尚未广为人知的物品。此外，Liu 和 Jia [9]的实证研究确认了用户声誉与评价数量之间的正相关关系，即用户的评价数量越多，其在系统中的声誉也越高。依据用户的个人偏好和对象的受欢迎程度等评分模式，我们将用户声誉的计算方法分为三类：迭代模型[10]-[23]、基于群组的模型[6] [7] [24]-[26]和基于特殊分布的模型[8] [9] [27]-[30]。本文全面回顾了这三种模型，旨在为用户声誉研究领域提供连贯的理解和新视角，同时整合当前面临的挑战和复杂的方法论问题。

## 2. 迭代模型

迭代模型基于这样的假设：用户声誉和对象质量是相互影响、互为因果的。核心理念在于，对象具有固有质量，用户的评分可以通过其与对象质量的关系(如距离或相关系数)来评估用户声誉，进而利用用户声誉和评分来判定对象的质量。这一迭代过程持续进行，直至用户声誉和对象质量达到稳定状态。

Laureti 等[10]提出了一种迭代重构(IR)方法,用于从用户评分中精确提取对象的固有质量。该方法涉及  $N$  个用户对  $M$  个对象的评估,每个对象有其固有质量  $q_i$ 。用户给出的评分  $x_{ij}$  被用来重新评估对象质量,通过加权平均法估计对象的固有质量。迭代开始时,设置每位用户的初始权重为用户总数的倒数。接着,用户权重  $\omega_i$  通过公式  $\omega_i = V_i^{-\beta}$  计算,其中  $V_i$  是用户评分的样本方差,  $\beta$  是调整权重的参数。在迭代过程中,权重、样本方差和加权平均值不断更新,以改进对对象质量和用户判断的评估。迭代一直进行,直到成本函数  $E(\{q_i\}) = \sum_i [\sum_j (x_{ij} - q_i) V_i^{-\beta}]^2$  达到最小值为止。该方法在处理大规模数据集时表现出色,但其收敛速度较慢,并且在处理异质性较高的评分数据时可能存在不稳定性。

Kerchove 和 Van Dooren [11]通过引入信任矩阵  $T$  改进了迭代重构(IR)方法,以增强其鲁棒性。在初始阶段,每个用户对  $T$  中所有对象的信任值均设为 1。此方法通过信任矩阵计算每个用户  $i$  与对象  $j$  之间的权重  $W_{ij} = \frac{T_{ij}}{\sum_{k,k \rightarrow j} T_{kj}}$ 。接着,利用用户  $i$  对对象  $j$  的评级  $E_{ij}$  和上述权重,计算对象的质量  $r_j = \sum_{i,i \rightarrow j} W_{ij} E_{ij}$ 。此外,该方法还计算置信散度  $d_i = \frac{1}{m_i} \sum_j (E_{ij} - r_j)^2$ ,其中  $m_i$  是用户  $i$  评定的对象数量。置信散度用于更新信任矩阵,设定  $T_{ij} = c_j - d_i$ ,其中  $c_j$  是调整参数。同时,用户的可信度  $t_i$  由公式  $t_i = d_{\max} - d_i$  确定,  $d_{\max}$  是所有用户中的最大置信散度。这一计算过程持续重复,直到权重矩阵、声誉向量和信任矩阵稳定。

Zhou 等[12]提出了基于相关性的排名(CR)方法,该方法根据用户评级与相应对象质量之间的相关性来确定用户声誉。每个用户的初始声誉设定为其评级数与对象集总数的比例。对象质量通过用户评级的加权平均值估计,其中权重为用户声誉。用户声誉的高低取决于其评级与加权平均评级的一致性。相关系数通过以下公式计算:  $Corr_i = \frac{1}{k_{ui}} \sum_{\alpha \in O_i} \left( \frac{r_{i\alpha} - \bar{r}_i}{\sigma_{r_i}} \right) \left( \frac{Q_{\alpha} - \bar{Q}_{oi}}{\sigma_{Q_{\alpha}}} \right)$ ,其中  $Corr_i$  大于或等于 0 时,用户声誉等于相关系数;否则,声誉为 0。该方法反复计算相关性,直到连续迭代之间对象质量估计的差异低于特定阈值,从而达到稳定状态。

在 CR 方法的基础上,考虑到在线系统中恶意用户的普遍存在,Liao 等[13]开发了一种新的迭代排名方法,名为声誉再分配迭代法(IARR)。该方法的核心是一个非线性声誉再分配公式,具体为:

$$R_i = TR_i^{\theta} \frac{\sum_j TR_j}{\sum_j TR_j^{\theta}}, \text{ 其中, } \theta \text{ 是一个可调参数。当 } \theta=0 \text{ 和 } \theta=1 \text{ 时,该方法分别简化为基于平均值的方法和 CR。}$$

得到的  $R_i$  将作为用户  $i$  的声誉,用于评估对象的质量。该方法通过提高评分一致性高的用户的影响力,同时抑制一致性低的用户的影响力,有效地减少了迭代过程中的噪音,提高了系统的整体鲁棒性。

考虑到仅对少数对象评分的用户不应具有过高的声誉,以及仅被少数用户评分的对象也不应被认为是高质量的,Liao 等[13]提出了 IARR2 方法,在该方法中引入了两个惩罚因子来计算用户声誉和对象质量。该方法在声誉计算中考虑了用户评分对象的数量,以防止评分对象较少的用户获得不合理的高声誉。

用户声誉的修改后的计算公式为:  $TR_i = \frac{\lg(k_i)}{\max\{\lg(k_j)\}} \cdot \frac{1}{k_i} \sum_{\alpha \in O_i} \left( \frac{r_{i\alpha} - \bar{r}_i}{\sigma_{r_i}} \right) \left( \frac{Q_{\alpha} - \bar{Q}_{oi}}{\sigma_{Q_{\alpha}}} \right)$ 。在计算对象质量时,将用户声誉与最大声誉值比较并进行归一化,以减少少数高声誉用户对质量评估的不当影响。对象质量的修改后计算公式为:  $Q_{\alpha} = \max_{i \in U_{\alpha}} \{R_i\} \frac{\sum_{i \in U_{\alpha}} R_i r_{i\alpha}}{\sum_{i \in U_{\alpha}} R_i}$ 。这些调整有助于确保评分系统更加公正和准确。

针对传统方法忽略用户活动模式的问题,Liu 等人[14]提出了一种改进的对象质量和用户声誉排名迭代法(IRUA)。该方法认为,用户对少数对象的高评分不一定代表对象的高质量。因此,对象质量的评估

不仅基于加权平均评分，还考虑了评分用户的活跃度，计算公式为： $Q_\alpha = \max_{i \in U_\alpha} \frac{k_i}{|O|} \cdot \frac{\sum_{i \in U_\alpha} R_i r_{i\alpha}}{\sum_{i \in U_\alpha} R_i}$ ，其中， $|O|$ 是对象总数， $\max_{i \in U_\alpha} \frac{k_i}{|O|}$ 是基于活跃度的惩罚因子。接着，通过计算用户评分向量与对象质量向量之间的皮尔逊相关系数  $C_i = \frac{1}{k_i} \sum_{\alpha \in O_i} \left( \frac{r_{i\alpha} - \bar{r}_i}{\sigma_{r_i}} \right) \left( \frac{Q_\alpha - \bar{Q}_i}{\sigma_{Q_i}} \right)$ ，进而评估用户声誉。用户的声誉计算考虑了相关系数和用户度。如果  $C_i \geq 0$ ，则  $R_i = \frac{k_i^\theta}{k_{\max}^\theta} C_i$ ；否则  $R_i = 0$ 。迭代过程一直持续直到质量估计的变化低于设定阈值。

Wu 等[15]发现用户的评分偏差也会影响对象质量的评估，因此提出了迭代平衡法(IB)以减少这种影响。起初，每个对象的质量设置为收到的所有评分的平均值。通过更新质量值和用户评分来调整用户声誉。用户的声誉指数(IBR)计算公式为： $IBR_i(t) = \sqrt{\frac{\sum_{\alpha \in O_i} [Q_\alpha(t-1) - r_{i\alpha}]^2}{k_i}}$ ，每个对象的质量则根据用户声誉和评分更新： $Q_\alpha(t) = \frac{1}{k_\alpha} \sum_{i \in U_\alpha} [r_{i\alpha} + IBR_i(t) \text{sgn}(Q_\alpha(t-1) - r_{i\alpha})]$ 。迭代过程不断重复，直到对象质量的变化低于预设阈值。 $Q_\alpha(t)$ 和 $IBR_i(t)$ 的最终稳定值用于量化对象的内在质量和用户的声誉，以此来减少评分偏差的影响，提高系统的准确性和可靠性。

Liang 等[16]提出了一种稳健排名方法(SR)，基于评分波动较小的用户更可靠，应赋予更高声誉的前提。初始阶段，所有用户声誉设为 1，产品  $Q_\alpha$  的质量由其加权平均评分决定。该方法计算每个用户  $u_i$  对对象质量评分的均方偏差  $f_i$  来反映评分的稳定性，并使用下式计算评分波动指数： $s_i = \sqrt{\frac{1}{k_i} \sum_{\alpha \in O_i} (A_{i\alpha} - Q_\alpha - x_i)^2}$  其中  $x_i$  表示用户评分与对象估计质量之间的平均偏差。声誉  $R_i$  的计算公式为  $R_i = \frac{1}{f_i \times s_i}$ 。为防止用户评分较少导致声誉不成比例增长，声誉调整公式为： $R_i = \frac{\lg(k_i)}{\max\{\lg(k_i)\}} R_i$ 。迭代过程持续进行，直到对象质量的变化低于预定阈值。

Liao 等[17]在基于累积时间的排名(ATR)方法中引入时间依赖性，考虑网络的整个历史，以计算用户声誉和对象质量。第  $t$  年的用户  $i$  的权重  $W_i(t)$  是通过将评分数  $k_i(t)$  除以该年的总评分  $N(t)$  来计算的。对象  $\alpha(t)$  的权重  $W_\alpha(t)$  的计算方法类似。这些权重用于初始化用户  $i$  的声誉  $R_i$  和对象  $\alpha(t)$  的质量  $Q_\alpha$ ，其中： $R_i = \frac{\sum_{\alpha \in O_i(t), t \in Y} r_{i\alpha} \cdot W_i(t)}{\sum_{t \in Y} k_i(t)}$ ， $Q_\alpha = \frac{\sum_{i \in U_\alpha(t), t \in Y} r_{i\alpha} \cdot R_i(t) \cdot \tilde{W}_\alpha(t)}{\sum_{i \in U_\alpha(t), t \in Y} R_i(t) \cdot \tilde{W}_\alpha(t)}$ 。该方法采用迭代过程来更新用户的声誉  $R_i(t)$  和对象的质量  $Q_\alpha(t)$ ：

$$R_i(t) = \frac{1}{\text{accu}_i(t) \cdot \max\{\log k_i(t)\}} \sum_{\alpha \in O_i(t)} \left( \frac{W_i(t) \cdot (r_{i\alpha} - \bar{r}_i)}{\sigma_{r_i}} \right) \left( \frac{\tilde{W}_\alpha(t) \cdot (Q_\alpha(t-1) - \bar{Q}_\alpha(t-1))}{\sqrt{\frac{\tilde{\sigma}_\alpha}{k_i(t)}}} \right),$$

$$Q_\alpha(t) = \frac{\sum_{i \in U_\alpha(t)} \tilde{W}_\alpha(t) \cdot r_{i\alpha}}{\sum_{t \in Y} \tilde{W}_\alpha(t)} \cdot \text{accu}_\alpha(t) \cdot \tilde{k}_\alpha(t),$$

其中  $\text{accu}_i(t)$  和  $\text{accu}_\alpha(t)$  分别代表用户  $i$  和对象  $\alpha$  从不同时间段累计的用户  $i$  和对象  $\alpha$  的声誉值和质量值。该过程反复运行，直到  $|Q_\alpha(t) - \bar{Q}_\alpha(t)|$  低于指定阈值。

Sun 等[18]观察到, 可靠用户的评分偏差通常较小, 评分分布具有明显峰值, 而恶意用户的评分则无固定模式, 常带有偏差。基于此, 他们提出了一种迭代优化排名(IOR)方法, 利用预期分歧  $E$  和评分模式  $\zeta$  来评估用户评分差异和模式。方法首先使用用户评分的加权平均值初始化对象质量, 然后计算每个用户的评分差异向量  $E_i = \sum_{\beta} w_{\beta,i} * c_{\beta,i} (\beta=1,2,3,4,5)$ , 其中  $w_{\beta,i}$  和  $c_{\beta,i}$  分别为每个评分组的权重和平均值。接着, 创建用户评分频率矩阵, 定义评分模式为  $\zeta_i = \frac{r_{i,\max} - r_{i,\min}}{r_{i,\max} + r_{i,\min}}$ 。用户信誉和对象质量通过  $R_i = \frac{1}{E_i} + \zeta_i$  进行更新, 迭代直到声誉差异低于特定阈值。

Huang 等[19]针对电商系统中恶意评分者的问题, 提出了基于迭代偏差的用户声誉排名(IDR)方法。该方法关注用户评分与对象质量之间的差异  $\mu_i = \frac{\sum |r_{ia} - Q_a|}{k_i}$ 、用户的评分频率  $\eta_i = \frac{u_{is}}{k_i}$  和评分模式  $\delta_i = \frac{\eta_{is\max}}{\eta_{is\min}}$ 。初始阶段, 对象质量通过用户评分的加权平均值初始化, 迭代过程中用户的评分频次和模式结合起来就能揭示用户评分的一致性和差异性。用户声誉  $R_i = \delta_i - \mu_i$  则综合了平均评分差异和评分模式指标, 其值越小即用户的声誉越低, 越不可靠。迭代持续至声誉变化低于设定阈值。

Liu 和 Zhao [20]开发了一种基于特征向量分析的迭代声誉和质量排名方法, 称为 EigenRank。与许多缺乏收敛性证明的现有迭代法不同, EigenRank 的数值收敛性在理论上是有保证的。该方法通过  $r'_{ia} = \frac{2(r_{ia} - \min(r_i))}{\max(r_i) - \min(r_i)} - 1$ , 将用户的评分归一化为偏好度  $r'_{ia}$ , 并将评分映射到  $[-1, 1]$  区间。使用  $R_i = \frac{k_i}{|O|}$  初始化用户的声誉, 其中  $|O|$  是对象总数。对象质量的计算公式为  $Q = c_1 K^m B^T R$ , 其中  $c_1$  为缩放参数,  $K$  和  $B$  分别对应对象度和归一化评分。每次迭代都会使用  $Q = c_1 K^m B^T R$  和  $R = c_2 D^{m_2} B Q$  更新声誉向量  $R$  和对象质量向量  $Q$ , 其中  $D$  为用户度矩阵,  $c_2$  是另一个缩放参数。迭代过程一直持续到声誉向量  $R(n)$  趋于稳定, 即连续声誉向量之间的欧氏距离低于阈值。

Li 等[21]基于 CR 算法提出了一种改进的声誉评价方法(CRCN), 该方法结合了用户 - 对象二分网络结构。该方法将经典的声誉评价方法与用户的聚类系数相结合, 引入惩罚 - 奖励机制来更新用户的声誉。初始阶段, 用户声誉和对象质量的计算方式与 CR 方法一致。首先利用用户的聚类系数  $cc(e_i)$  调整临时声誉值  $TR'_i$ , 计算公式为  $TR'_i = \left( \frac{cc(e_i)}{\max cc(e_j)} \right)^{\frac{1}{2}} TR_i$ 。其中,  $cc(e_i, e_j) = \frac{|N(e_i) \cap N(e_j)|}{|N(e_i) \cup N(e_j)|}$ ,  $N(e_i)$  是用户  $i$  评定的所有对象的集合。其次, 引入非线性惩罚 - 奖励函数来计算最终声誉  $R_i$ , 公式为:

$$R_i = \begin{cases} 0 & \text{if } TR'_i = 0 \\ \left[ 1 + \left( \frac{1}{TR'_i} - 1 \right)^\beta \right]^{-1} & \text{if } 0 < TR'_i < 1 \\ 1 & \text{if } TR'_i = 1 \end{cases}$$

该方法迭代地更新用户信誉, 直到对象质量的变化低于预定阈值。

Li 等[22]提出了一种通过对评级系统中的对象进行分类来评估用户声誉的新方法(CRE), 强调每个用户在不同对象类别中的专业知识。用户声誉不再是单一的标量值, 而是用对象特征空间中的向量来表示。该方法首先计算出特征间的相似度  $\phi_{kl}$ , 计算公式为  $\phi_{kl} = \frac{A_{kl}}{A_{kk}}$ , 其中  $A_{kl}$  是特征  $k$  和  $l$  之间的边缘数,  $A_{kk}$  是特征  $k$  中的内部边缘数。每个用户  $i$  都会被分配一个声誉向量  $\vec{R}_i$ , 其中每个元素  $R_{ik}$  代表在特征  $k$  中的声誉, 初始阶段设置为用户在特征  $k$  中的度数  $k_{ik}$  与该特征中的对象数量  $|O_k|$  的比值, 即  $R_{ik} = \frac{k_{ik}}{|O_k|}$ 。用户  $i$

在特征  $k$  中的信誉  $R_{ik}$  是通过计算用户的评分向量和相应的对象质量向量之间的相关性得出的，表示为：

$$R_{ik} = \frac{1}{k_{ik}} \sum_{\alpha \in O_{ik}} \left( \frac{r_{i\alpha} - \bar{r}_{ik}}{\sigma_{r_i}} \right) \left( \frac{Q_{\alpha} - \bar{Q}_{ik}}{\sigma_{Q_i}} \right)$$
。对象质量  $Q_{\alpha}$  通过用户评分的加权平均值计算，权重为用户声誉  $R_{ik}$  与相似度  $\phi_{ki}$  的乘积。当对象质量估计值的变化低于一个较小的阈值时，迭代过程就会停止。

Zhu 等[23]提出了一种基于 Z 统计的新型声誉迭代方法(ZS)，重点关注经典方法经常忽略的棘手对象。每个用户的声誉初始设为其等级与对象总数之比。对象质量的用户评分向量的偏差度  $Z_i$  通过 Z 统计法计算，具体公式为：
$$Z_i = \sum_{\alpha \in O_i} \frac{|Q_{\alpha} - r_{i\alpha}|}{k_i \sqrt{\sum_{i \in U_{\alpha}} (Q_{\alpha} - r_{i\alpha})^2 / k_{\alpha}}}$$
。接着通过偏差度更新用户声誉  $R'_i = \theta^{-Z_i}$ ，其中  $\theta$  是

可调参数。重复该过程，直到所有对象的质量变化低于预定阈值。

迭代模型通过不断更新用户声誉和对象质量来提高评估准确性，能够动态调整以适应用户行为变化，并结合多维数据提供全面的声誉评估。然而，迭代模型计算复杂度高，处理大规模数据时耗时较长，收敛速度可能较慢，并且对初始参数设置敏感，可能需要多次调试才能获得最佳效果。

### 3. 基于群组的模型

基于群组的模型通过评分分类来量化用户声誉，主要依赖评分本身而非评分与对象质量之间的关系。此模型的核心策略是将给出相似评分的用户归为一个群体，利用该群体的规模来衡量相应对象的可信度，并据此计算用户的声誉。这种方法强调了群体共识对评价对象可信度的影响。

Gao 等[24]提出了一种基于组的排名(GR)方法，该方法考虑到用户背景的差异性，识别对象常收到有效但主观的评分。该方法通过评分行为将用户进行分组，并将评分矩阵转化为评分奖励矩阵。具体操作是将每个对象的评分组大小通过除以该对象的评级总数来进行计算和归一化，得到评级奖励矩阵  $\Lambda^*$ 。利用  $\Lambda^*$  的归一化过程，将原始评级矩阵  $A$  映射到奖励矩阵  $A'$ ，其中用户信誉  $R_i$  依据公式  $R_i = \frac{\mu(A'_i)}{\sigma(A'_i)}$  计算。

引入迭代机制后，Gao 等[25]提出了一种基于迭代组的排名(IGR)方法。IGR 通过将用户声誉纳入计算评分组的大小并通过迭代处理动态更新用户声誉来扩展 GR 方法。该方法考虑评分对象矩阵  $B(i)$  和用户声誉  $R_i$  (初始设置为 1)两者来计算用户评级组的加权大小  $\Lambda_s^{\alpha}$ ，公式为  $\Lambda_s^{\alpha} = \sum_{i=1}^m R_i B(s\alpha_i)$ ，其中  $m$  表示用户总数。对于已对对象  $\alpha$  进行评分的用户， $B(s\alpha_i)=1$ ；否则， $B(s\alpha_i)=0$ 。迭代地更新用户声誉  $R$  和奖励矩阵  $A'$ ，直到用户声誉的变化下降到预设阈值以下，进一步提高了 GR 算法的鲁棒性。

Fu 等[6]指出，评估用户声誉时应考虑分组行为与个人评级特征。他们观察到，与异常用户相比，正常用户的评级往往更集中且偏差较小，通常受个人偏好影响。基于此，Fu 等人提出了一种基于迭代组和差异排名(IGDR)的方法，这一方法将声誉重新分配过程集成进原有的 IGR 方法。新方法中  $R_i$  的计算公式为：
$$R_i = \sqrt{\mu(A'_i)} + \frac{1}{5\sqrt{\sigma(A'_i)} + \sigma(\omega_i)}$$
，其中  $\sigma(\omega_i)$  表示用户  $i$  的评级的标准偏差。该方法通过整合用户评级的集中度和变异性，更精确地反映了用户的声誉。

Dai 等[7]提出了一种将用户评级映射到统一标准的方法，以实现准确的声誉排名，并开发了一种基于用户偏好的在线用户声誉计算方法(PGR)。该方法首先对用户评级进行归一化，使用以下公式：

$$r'_{i\alpha} = \begin{cases} \frac{r_{i\alpha} - \mu_i}{r_i^{\max} - r_i^{\min}} & \text{if } r_i^{\max} \neq r_i^{\min} \\ 0 & \text{if } r_i^{\max} = r_i^{\min} \end{cases}$$
。对于新的评级矩阵  $D = \{r'_{i\alpha}\}$ ，设置类别区间  $\tau = \frac{r'_{\max} - r'_{\min}}{l}$ 。该矩阵通过

将映射的评分转换为特定范围内的在线评分构建，从  $[r'_{\min}, r'_{\min} + \tau)$  开始，以  $\tau$  为增量继续到  $[r'_{\min} + (l-1)\tau, r'_{\max}]$ 。用户根据评分分组，计算每个组的大小  $G_{s\alpha}$ 。然后，构造评级 - 奖励矩阵  $H$ ：

$H_{s\alpha} = \frac{G_{s\alpha}}{\rho_\alpha}$ , 其中  $\rho_\alpha$  表示对象  $\alpha$  的度。通过结合  $H$  和  $D$  构建奖励矩阵  $Z$ , 其中用户  $i$  对评分  $r_{i\alpha}$  的奖励是  $Z_{i\alpha} = H_{s\alpha}$ 。最终, 用户信誉  $R_i$  通过奖励向量  $Z_i$  的平均值  $\mu(Z_i)$  和标准差  $\sigma(Z_i)$  计算得出:  $R_i = \frac{\mu(Z_i)}{\sigma(Z_i)}$ 。

Zhang 等[26]提出了一种名为基于群组行为排序聚合的异常用户检测方法(GSRA), 以解决传统方法在检测由异常用户引起结构变化方面的不足。GSRA 利用 GMDD、GMMDD、GSOR 和 GGRV 指标来量化异常用户组评分中的失真。该方法首先从二分网络中选择邻居节点组作为初始候选。组异常度量  $S(c_{p_i})$  定义为:  $S(c_{p_i}) = 1 - \sqrt{\frac{1}{4}(F_{GMDD}(p_i)^2 + F_{GMRDD}(p_i)^2 + F_{GSOR}(p_i)^2 + F_{GGRV}(p_i)^2)}$ 。具有高的群体行为累积值的对象被识别为可能的异常用户。该方法通过合并前  $K$  个对象的邻居节点, 并利用 Jaccard 相似度和评分行为的哈希特征向量来分离异常用户群组。异常用户的得分通过排序聚合来优化, 每个异常用户  $s_i$  被分配一个原始异常值  $V_{s_i}^{\text{origin}}$  和优化后的声誉值  $V_{s_i}^{\text{agg}}$ , 后者由下式确定:  $V_{s_i}^{\text{agg}} = \alpha V_{s_i}^{\text{origin}} + W_{G_j} |R_{s_i} - \mu(G_j)|$ ,  $s_i \in G_j$ ,  $W_{G_j} = 1 - \frac{1}{e^{N_{G_j}/\sigma(G_j)}}$ , 其中,  $W_{G_j}$  是异常用户群组的影响因素,  $\alpha$  和  $\beta$  平衡了原始异常用户的度和行为特征之间的权重关系。

基于群组的模型通过将用户分组并计算组内一致性, 简化了计算过程, 能较好地处理用户评分行为的多样性, 并在评分数据更新时快速响应。然而, 该模型的准确性和鲁棒性高度依赖于分组策略, 缺乏对个体用户细粒度行为特征的分析, 可能产生偏见, 导致某些用户声誉评估不准确。

#### 4. 基于特殊分布的模型

基于特殊分布的模型通过应用统计分布理论来评估用户声誉, 假设用户的评级遵循特定的分布, 例如常见的正态分布或贝塔分布。在这种模型中, 用户声誉取决于其评分与预定分布的一致性程度, 其中偏差较小表示声誉较高。这种方法侧重于识别那些其评分行为与典型或预期分布紧密匹配的用户, 从而确保声誉系统的准确性和可靠性。

Lee 等[27]提出了基于偏差的排名(DR), 假设用户评分通常遵循正态分布, 这一分布受多个独立因素的影响。DR 方法通过评估用户评分行为与正态分布的偏差来评估声誉。该方法计算  $Z$  分数  $Z_{is} = \frac{\sqrt{|O_{is}|}(q_{is} - s)}{\sigma_i}$ , 以表示用户  $\mu_i$  对所有对象在评分等级  $s$  上给出的平均质量评级  $q_{is}$  与预期评分  $s$  的偏差, 通过预期分数  $s$  来度量用户声誉。然后, 将累积声誉  $r_i$  计算为所有评分等级的绝对  $Z$  得分之和的负值, 最终用户声誉  $r_i$  推导计算公式表示为:  $r_i = -\frac{\sigma_i \sum_{s=1}^L |Z_{is}|}{\sqrt{|O_i|}} = -\frac{\sum_{s=1}^L \sqrt{|O_{is}|} |q_{is} - s|}{\sqrt{|O_i|}}$ 。

Zhou 等[28]提出了一种基于用户评级模式和评级偏差(RPRD)的声誉排名方法, 该方法通过在 DR 的基础上引入熵和概率差来进一步衡量用户的评级模式和评级偏差。首先使用 DR 方法计算出用户声誉指数  $r_i$  后, 然后引入熵  $g_i$  来通过公式  $g_i = (-1)^k \sum_{s=1}^L p(t_{is}) \log_e p(t_{is})$  衡量评分模式的随机性。其中, 参数  $k$  的值为 1 或 -1, 用于区分随机用户和恶意用户,  $p(t_{is})$  表示用户  $i$  给出评分  $s$  的概率。概率差  $h_i = \frac{1}{2} \sum_{s=1}^{L-1} \sqrt[3]{p(t_{i(s+1)}) - p(t_{is})}$  反映了评分波动。最终的用户声誉得分  $R_i$  则为  $R_i = r_i + g_i + h_i$ 。此方法不仅考察了评分的偏差, 还考虑了评分的随机性和波动性, 提供了一种更全面的用户声誉评估方式。

王和刘[29]指出, 正常用户的评分通常表现为峰值分布, 这与恶意用户的分布不同, 并据此开发了一种基于用户评级模式(BPR)的排名方法。该方法包括统计每个用户的评分并计算评分模式变量

$P_i = \frac{r_{i,\max} - r_{i,\min}}{r_{i,\sup} + r_{i,\min}}$ ，其中符合峰值分布的用户具有较高的  $P_i$  值。评分模式变量有助于估算用户声誉和对象

质量。其中符合峰值分布的用户将具有较高的值。评分模式变量有助于估算用户声誉和对象质量。同时，计算评分偏差向量——用户评分与对象质量之间的绝对差异，以及平均偏差  $E_i$ 。最终的声誉值  $R_i$  则根据平均偏差和评分模式变量计算得出，公式为  $R_i = \frac{1}{E_i} + P_i$ 。

Sun 和 Chen [30]发现用户评分既符合正态分布也符合峰值分布，从而提出了基于区间划分的排名(IDR)方法。该方法使用 Z 分数公式对评级矩阵进行归一化，将每个评级  $r'_{ia}$  与标准正态分布对齐。指标函数  $\delta_{ia}$  用于评估评分的有效性，对于区间[-1, 1]内的评分设置为 1，否则设置为 0。评级准确性  $\eta_i$  通过公式  $\eta_i = \frac{s_i}{s_i + f_i}$  计算，其中  $s_i$  和  $f_i$  代表合理和不合理评级的数量。距离函数  $bd$  用于测量偏离合理区间的程度，

定义为： $bd = \begin{cases} -1 & \text{if } r'_{ia} \leq -1 \\ 1 & \text{if } r'_{ia} \geq 1 \\ r'_{ia} & \text{if } -1 < r'_{ia} < 1 \end{cases}$ 。累积距离  $d_i$  反映了评分偏差的总和，通过  $d_i = \frac{\sum_{\alpha \in O_i} (r'_{ia} - bd)^2 + \delta}{f_i + 1}$  计

算。用户声誉  $R_i$  的计算公式表示为  $R_i = \sqrt{\frac{1}{d_i}} + \eta_i + v_i$ ，其中  $v_i$  指示用户评级模式的特征值，计算为

$v_i = \frac{r_{ir}^{\max} - r_{ir}^{\min}}{r_{ir}^{\max} + r_{ir}^{\min}}$ 。此方法通过整合评分的正态和峰值特性，综合评估用户声誉，进一步提高了用户声誉的度量精度。

Liu 等[8]提出了一种基于贝叶斯分析和贝塔概率分布的无参数在线用户声誉排名方法(RBPD)。该方法将用户声誉定义为用户评分与多数派一致的概率。为了标准化不同的评分标准，应用归一化公式：

$r'_{iy} = \begin{cases} 2(r_{iy} - r_i^{\min}) / (r_i^{\max} - r_i^{\min}) - 1 & r_i^{\max} \neq r_i^{\min} \\ 0 & r_i^{\max} = r_i^{\min} \end{cases}$ 。评分的公正性通过归一化评分与多数评分的一致性来确定。

用户声誉  $R_i$  通过贝塔分布计算，公式为  $R_i = \frac{\alpha_s + s}{\alpha_s + s + \alpha_f + f}$ ，其中  $\alpha_s$ ， $\alpha_f$  分别代表初始阶段用户的公正

和不公正评分的数量， $s$ ， $f$  为观察到的计数。然后，对象质量  $Q_\gamma$  的评估使用  $Q_\gamma = \max_{i \in U_\gamma} \{R_i\} \frac{\sum_{i \in U_\gamma} R_i r_{iy}}{\sum_{i \in U_\gamma} R_i}$ ，

同时考虑评级和用户声誉。

Liu 和 Jia [9]提出了一种基于贝塔概率分布的迭代声誉排名方法(IBeta)，将迭代声誉分配过程融入到 RBPD 中。在该方法中，基于每个用户的评级  $r'_{iy}$  和声誉  $R_i$ ，正面意见  $PO_\gamma$  和负面意见  $NE_\gamma$  的计算重新定义了公平评分概率  $p_{iy}$  的公式，其中每个用户  $i$  最初被分配了相等的声誉  $R_i = 1$ ： $PO_\gamma = \sum_{r'_{iy} \geq 0, i \in U_\gamma} R_i$ ，

$NE_\gamma = \sum_{r'_{iy} < 0, i \in U_\gamma} R_i$ 。评分  $r_{iy}$  被认为是公平评分的概率  $p_{iy}$  表示为： $p_{iy} = \begin{cases} \frac{PO_\gamma}{PO_\gamma + NE_\gamma} & \text{if } r'_{iy} \geq 0 \\ \frac{NE_\gamma}{PO_\gamma + NE_\gamma} & \text{if } r'_{iy} < 0 \end{cases}$ ，该方法

迭代计算用户声誉和对象质量，直到声誉向量之间的差值下降到指定阈值以下。

基于特殊分布的模型依靠统计分布理论提供坚实的数学基础，能够识别和过滤异常评分，适应不同类型的评分数据。然而，这些模型依赖于评分数据是否符合假设的分布类型，当实际数据分布与假设不符时，评估结果可能失真。此外，某些统计分布模型计算复杂，尤其在高维数据或大规模数据集上，计算效率较低，不适用于非正态分布的数据。

## 5. 展望

声誉系统在识别和抑制恶意行为方面的有效性已经得到证明。目前, 声誉度量方法可分为迭代模型、基于群组的模型和基于特殊分布的模型。然而, 现有方法仍有进一步发展的空间:

(1) 在动态在线环境下, 声誉排名方法的适应性和鲁棒性方面尚存在理论上的差距。Laureti 等[10]的方法在处理大规模数据集时表现出色, 但在异质性较高的评分数据时可能不稳定。未来研究应探讨如何改进该方法以提高其收敛速度和稳定性。

(2) 探索如何利用多维数据去构建声誉排名算法, 声誉还可以从社交媒体行为和文本内容分析等多重因素中去进行度量。Zhou 等[28]的研究表明, 整合评分模式和评分偏差可以提高声誉评估的准确性, 但其计算复杂度较高。未来研究可以尝试降低计算复杂度以提高实用性。

(3) 如何有效地利用声誉排名方法适应动态变化的评分标准, 将网络社区结构与多维数据分析结合, 可能有助于提高这些方法的准确性和适用性。此外, 建立一个全面的数据库, 涵盖环境因素、社交网络、用户行为和评分动态, 对提高声誉排名方法的效率和准确性至关重要。Wang 和 Liu [29]的研究指出, 考虑评分分布特性能够更准确地反映用户声誉, 但其在大规模数据集上的应用效果仍需验证。

## 参考文献

- [1] Pan, W., Xiang, E. and Yang, Q. (2021) Transfer Learning in Collaborative Filtering with Uncertain Ratings. *Proceedings of the AAAI Conference on Artificial Intelligence*, **26**, 662-668. <https://doi.org/10.1609/aaai.v26i1.8197>
- [2] Wang, G., Xie, S., Liu, B. and Yu, P.S. (2011) Review Graph Based Online Store Review Spammer Detection. 2011 *IEEE 11th International Conference on Data Mining*, Vancouver, 11-14 December 2011, 1242-1247. <https://doi.org/10.1109/icdm.2011.124>
- [3] Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J. and Gonçalves, M. (2009) Detecting Spammers and Content Promoters in Online Video Social Networks. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, 19-23 July 2009, 620-627. <https://doi.org/10.1145/1571941.1572047>
- [4] Zeng, A. and Cimini, G. (2012) Removing Spurious Interactions in Complex Networks. *Physical Review E*, **85**, Article 036101. <https://doi.org/10.1103/physreve.85.036101>
- [5] Zhang, Q., Zeng, A. and Shang, M. (2013) Extracting the Information Backbone in Online System. *PLOS ONE*, **8**, e62624. <https://doi.org/10.1371/journal.pone.0062624>
- [6] Fu, Q., Ren, J. and Sun, H. (2021) Iterative Group-Based and Difference Ranking Method for Online Rating Systems with Spamming Attacks. *International Journal of Modern Physics C*, **32**, Article 2150059. <https://doi.org/10.1142/s0129183121500595>
- [7] Dai, L., Guo, Q., Liu, X., Liu, J. and Zhang, Y. (2018) Identifying Online User Reputation in Terms of User Preference. *Physica A: Statistical Mechanics and Its Applications*, **494**, 403-409. <https://doi.org/10.1016/j.physa.2017.12.020>
- [8] Liu, X., Liu, J., Yang, K., Guo, Q. and Han, J. (2017) Identifying Online User Reputation of User-Object Bipartite Networks. *Physica A: Statistical Mechanics and Its Applications*, **467**, 508-516. <https://doi.org/10.1016/j.physa.2016.10.031>
- [9] Liu, X. and Jia, S. (2019) An Iterative Reputation Ranking Method via the Beta Probability Distribution. *IEEE Access*, **7**, 540-547. <https://doi.org/10.1109/access.2018.2885551>
- [10] Laureti, P., Moret, L., Zhang, Y.-C. and Yu, Y.-K. (2006) Information Filtering via Iterative Refinement. *Europhysics Letters*, **75**, 1006-1012. <https://doi.org/10.1209/epl/i2006-10204-8>
- [11] de Kerchove, C. and Van Dooren, P. (2007) Iterative Filtering for a Dynamical Reputation System. arXiv: 0711.3964. <https://doi.org/10.48550/arXiv.0711.3964>
- [12] Zhou, Y., Lei, T. and Zhou, T. (2011) A Robust Ranking Algorithm to Spamming. *Europhysics Letters*, **94**, Article 48002. <https://doi.org/10.1209/0295-5075/94/48002>
- [13] Liao, H., Zeng, A., Xiao, R., Ren, Z., Chen, D. and Zhang, Y. (2014) Ranking Reputation and Quality in Online Rating Systems. *PLOS ONE*, **9**, e97146. <https://doi.org/10.1371/journal.pone.0097146>
- [14] Liu, X., Guo, Q., Hou, L., Cheng, C. and Liu, J. (2015) Ranking Online Quality and Reputation via the User Activity. *Physica A: Statistical Mechanics and Its Applications*, **436**, 629-636. <https://doi.org/10.1016/j.physa.2015.05.043>

- [15] Wu, L., Ren, Z., Ren, X., Zhang, J. and Lyu, L. (2018) Eliminating the Effect of Rating Bias on Reputation Systems. *Complexity*, **2018**, Article 4325016. <https://doi.org/10.1155/2018/4325016>
- [16] Liang, B., Jiang, S., Dou, L., Wang, L. and Wang, X. (2018) An Iterative Method for Evaluating Product Quality and User Reputation. 2018 37th Chinese Control Conference (CCC), Wuhan, 25-27 July 2018, 9714-9719. <https://doi.org/10.23919/chicc.2018.8483030>
- [17] Liao, H., Liu, Q., Huang, Z., Lu, K., Yeung, C.H. and Zhang, Y. (2022) Accumulative Time Based Ranking Method to Reputation Evaluation in Information Networks. *Journal of Computer Science and Technology*, **37**, 960-974. <https://doi.org/10.1007/s11390-021-0471-4>
- [18] Sun, H., Liang, K., Liao, H. and Chen, D. (2021) Evaluating User Reputation of Online Rating Systems by Rating Statistical Patterns. *Knowledge-Based Systems*, **219**, Article 106895. <https://doi.org/10.1016/j.knosys.2021.106895>
- [19] Huang, J., Sun, H., Chen, X., Liu, X. and Cao, J. (2021) An Iterative Deviation-Based Ranking Method to Evaluate User Reputation in Online Rating Systems. 2021 4th International Conference on Data Science and Information Technology, Shanghai, 23-25 July 2021, 15-21. <https://doi.org/10.1145/3478905.3478909>
- [20] Liu, X. and Zhao, C. (2022) A Converging Reputation Ranking Iteration Method via the Eigenvector. *PLOS ONE*, **17**, e0274567. <https://doi.org/10.1371/journal.pone.0274567>
- [21] Li, M., Han, C., Jiang, Y. and Di, Z. (2022) Improving the Performance of Reputation Evaluation by Combining a Network Structure with Nonlinear Recovery. *Frontiers in Physics*, **10**, Article 839462. <https://doi.org/10.3389/fphy.2022.839462>
- [22] Li, M., Jiang, Y. and Di, Z. (2022) Characterizing the Reputation of Evaluators Using Vectors in the Object Feature Space. *Expert Systems with Applications*, **201**, Article 117136. <https://doi.org/10.1016/j.eswa.2022.117136>
- [23] Zhu, H., Xiao, Y., Wang, Z. and Wu, J. (2022) A Robust Reputation Iterative Algorithm Based on Z-Statistics in a Rating System with Thorny Objects. *Journal of the Operational Research Society*, **74**, 1600-1612. <https://doi.org/10.1080/01605682.2022.2101952>
- [24] Gao, J., Dong, Y., Shang, M., Cai, S. and Zhou, T. (2015) Group-Based Ranking Method for Online Rating Systems with Spamming Attacks. *Europhysics Letters*, **110**, Article 28003. <https://doi.org/10.1209/0295-5075/110/28003>
- [25] Gao, J. and Zhou, T. (2017) Evaluating User Reputation in Online Rating Systems via an Iterative Group-Based Ranking Method. *Physica A: Statistical Mechanics and Its Applications*, **473**, 546-560. <https://doi.org/10.1016/j.physa.2017.01.055>
- [26] Zhang, Z., Zhou, M., Wan, J., Lu, K., Chen, G. and Liao, H. (2023) Spammer Detection via Ranking Aggregation of Group Behavior. *Expert Systems with Applications*, **216**, Article 119454. <https://doi.org/10.1016/j.eswa.2022.119454>
- [27] Lee, D., Lee, M.J. and Kim, B.J. (2018) Deviation-Based Spam-Filtering Method via Stochastic Approach. *Europhysics Letters*, **121**, Article 68004. <https://doi.org/10.1209/0295-5075/121/68004>
- [28] Zhou, J., Liu, Y. and Sun, H. (2022) A Reputation Ranking Method Based on Rating Patterns and Rating Deviation. 2022 5th International Conference on Data Science and Information Technology (DSIT), Shanghai, 22-24 July 2022, 1-6. <https://doi.org/10.1109/dsit55514.2022.9943923>
- [29] 王洁, 刘建国. 基于行为模式的用户声誉度量方法研究[J]. 上海理工大学学报, 2023, 45(1): 8-16.
- [30] Sun, H. and Chen, D. (2024) A Robust Ranking Method for Online Rating Systems with Spammers by Interval Division. *Expert Systems with Applications*, **235**, Article 121236. <https://doi.org/10.1016/j.eswa.2023.121236>