

社交网络用户影响力的模糊综合评价^①

张琛^{1,2}, 汤鲲^{1,2}, 彭艳兵³

¹(南京烽火软件科技有限公司, 南京 210019)

²(武汉邮电科学研究院, 武汉 430074)

³(烽火通信科技股份有限公司, 武汉 430073)

摘要: 社交网络平台信息传播迅速, 为了有效地进行舆情预警, 定量地评估用户在消息传播网络中重要性, 将模糊综合评价方法引入用户影响力建模问题中. 通过对用户在社交平台上的行为分析, 构造了包含用户活跃粉丝数以及平均转发数等五项指标在内的评价体系. 并针对传统的模糊综合评价算法在应用于计算评价指标权重方面的缺陷与不足, 提出改进模糊合成算子的方法构建用户影响力评估模型. 利用新浪微博社交平台上的真实数据, 结合对比实验和实际评估, 改进的模糊合成算子能根据需求调整权重对评价结果的影响, 同时该方法能较准确地反应社交网络中用户的实际影响力.

关键词: 社交网络; 用户影响力; 模糊综合评价; 模糊矩阵合成

引用格式: 张琛, 汤鲲, 彭艳兵. 社交网络用户影响力的模糊综合评价. 计算机系统应用, 2017, 26(12): 18-24. <http://www.c-s-a.org.cn/1003-3254/6135.html>

Fuzzy Comprehensive Evaluation of Social Network User's Influence

ZHANG Chen^{1,2}, TANG Kun^{1,2}, PENG Yan-Bing³

¹(Fiberhome Starrysky Co. Ltd., Nanjing 210019, China)

²(Wuhan Research Institute of Posts and Telecommunications, Wuhan 430074, China)

³(Fiberhome Telecommunication Technologies Co. Ltd., Wuhan 430073, China)

Abstract: Information spreads quickly on social networking platform. In order to effectively carry out public opinion early warning and quantitatively evaluate the importance of users in social network, the fuzzy comprehensive evaluation method is introduced into the user influence modeling problem. Based on the analysis of the behavioral analysis of the user's behavior on the social platform, the evaluation system including five indicators, such as user active number of fans and average forwarding number is constructed. A new fuzzy synthesis operator is proposed to construct the user influence evaluation model based on the shortcomings of the traditional fuzzy comprehensive evaluation algorithm in calculating the weight of the evaluation index. This operator can adjust the weight of the impact on the evaluation results according to the demand. Using the real data of Sina microblogging social platform, combined with comparative experiments and practical assessment, the method can more accurately reflect the actual impact of the user in the social network.

Key words: social network; user influence; fuzzy comprehensive evaluation; fuzzy matrix synthesis

近年来, 社交网络平台的兴起对人们的信息获取和生活方式产生了不可低估的影响, 快速增长的社交平台用户数量也导致信息总量的爆炸式增长. 如何使

信息被更多的用户所接收, 即用户影响力评估问题, 已经成为社交平台及其应用的研究热点之一. 在社交网络中, 消息传播的速度、覆盖的范围以及深度反

① 基金项目: 十二五国家科技支撑计划项目 (2015BAK20B05)

收稿时间: 2017-03-31; 修改时间: 2017-04-20; 采用时间: 2017-04-24

映了用户影响力的大小. 因此, 计算网络中节点的影响力并分析消息在网络中的传播机制, 可以有效地找出信息传播网中的关键节点, 从而能够进行舆情预警和舆情分析, 对保障网络安全有重要的意义^[1].

最早关于社交网络中影响力的研究是以粉丝数作为衡量的标准. 近年来, 国内外很多学者在这方面的研究又有了新的进展. 例如, 康书龙等^[2]重点考虑了用户在社交网站上的不同行为方式, 在 PageRank 算法的基础上设计了 BRR 算法, 并通过网络节点的 BRR 值来计算用户影响力值. Y Yamaguchi 等^[3]重点考虑了用户与用户之间的关系, 提出了 TURank 模型来评价用户的影响力. 这些研究的重点在于探究用户与用户之间的“好友关系”, 通过确定“好友关系”的权重值, 来反映不同用户的影响力大小. 但在实际应用中, 粉丝数和关注数这些“好友关系”并不能真实地反映用户的影响力. 相反, 诸如点赞、转发、评论等这样一些用户行为却能够集中反映微博的实际影响力, 对舆论产生实质性地影响.

贾冲冲等^[4]为了减少低影响力的粉丝在评价用户影响力过程中带来的负面作用, 提出了基于 H 指数的 HRank 模型来计算用户的影响力. 宫秀文等^[5]面对用户影响力最大化的问题, 提出一种基于改进 PageRank 算法的信息传播模型 PRP, 利用贪心算法近似求解该问题. 胡勇等^[6]在社交网络意见领袖的识别及分析方向提出了通过用户特征属性来刻画社交网络领袖的领导模型. 这些研究拓宽了研究用户影响力的思路, 从多维度、多角度综合考虑影响力的评估问题. 但以上针对用户影响力的研究多基于 PageRank 算法的改进或对网络结构进行分析, 能合理地评价用户影响力. 但是在评价体系的确立以及评价指标权重的处理上仍有改进的空间.

本文在借鉴已有成果的基础上, 提出了基于模糊综合评价的社交网络用户影响力评估模型. 一方面, 本文对社交网络用户的行为特征进行提取和定量分析, 划分了属性指标. 另一方面, 在面对实施传统的模糊综合评价算法中常见的两大难点, 本文分别改进了权重的确定方法和模糊矩阵的合成算子, 以此定量地衡量用户的影响力. 对于改进后的模糊综合评价算法, 本文介绍了原理及实现过程, 并设置对比实验验证该算法的有效性.

1 理论基础

目前针对定量计算用户影响力的问题, 研究较多

的方向是利用 PageRank 算法^[7-9]. 但该算法在计算页面的 PR 值的时候是均匀地分配到链出页面, 并没有考虑到用户行为所来带的权值的改变. 而模糊综合评价 (Fuzzy comprehensive evaluation, 简记 FCE) 算法能自由选择评价指标, 同时能较好地解决非确定性问题或难以量化的问题, 把定性评价转化为定量计算^[10]. 因此非常适用于本文所研究的衡量用户影响力大小的问题.

一般而言, 模糊综合评价算法描述如下: 设 $X = \{x_1, x_2, \dots, x_m\}$ 为评价因素集, $Y = \{y_1, y_2, \dots, y_n\}$ 为评价集. $R_i = \{r_{i1}, r_{i2}, \dots, r_{in}\}$ 为因素评价集 x_i 的单因素评价, 也就是 r_{ij} 表示因素 x_i 对评价 y_j 的模糊评价. $W = [w_1, w_2, \dots, w_m]$ 为权重矩阵. 其中 w_i 表示第 i 个因素在评价中的影响因子或重要程度. 如果多因素作用下的综合评价结果设为 $B = [b_1, b_2, \dots, b_n]$, 则公式 (1) 给出了传统的模糊综合评价计算模型.

$$B = [w_1, w_2, \dots, w_m] \bullet \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} \quad (1)$$

$$= [b_1, b_2, \dots, b_n]$$

根据模糊集的运算方法, 表 1 中给出了四种常用的模糊合成算子^[11], 公式 (1) 中的符号 \bullet 表示表 1 中四种常用算子中的任意一种.

表 1 四种常用模糊算子

算子	类型	特点
(\wedge, \vee)	主因素决定型	结果主要是由数值最大的因素所决定
(\odot, \vee)	主因素突出型	结果能在一定程度上反映非主要指标
(\odot, \oplus)	加权平均型	结果对所有因素依权重大小均衡兼顾
(\wedge, \oplus)	取小上界和型	各因素均能参与, 减小主因素作用

在实际应用该模型的过程中, 关键问题是确定权重矩阵 W 和单因素评价矩阵 R . 下面将在第 2 节和第 3 节分别针对这两难点介绍改进模糊综合评价算法的实际应用.

2 指标选取及权重确定

在应用模糊综合评价对用户影响力作评估时, 首要问题是解决哪些因素会对综合评价结果产生影响, 即指标的选取问题. 其次, 由公式 (1) 知, 还需确定指标的权重矩阵 W .

2.1 用户影响力指标的选取

常见的社交网络平台诸如 twitter, facebook, 微博,

人人网等. 社交网络的影响力通过用户与用户之间的互动得以体现. 近年来, 微博作为一种便捷的社交网络平台, 日活跃用户过亿, 已逐渐成为人们进行信息交流的重要媒介, 影响着人们的生活和社交方式. 微博信息传播快速, 范围广, 极易形成网络舆情.

在新浪微博平台, 结合实际情况可知, 用户关注与被关注这样的用户关系; 用户发微博数的数量、评论数、转发等行为; 用户是否经过官方认证等个人身份这些因素均与用户影响力大小相关, 鉴于此, 本文从以下三个方面探究并分析.

首先, 从用户关系角度考虑. 用户的粉丝数能在某种程度上反映用户在社交网络中受重视的程度. 由于目前有许多人为了扩大自己影响力购买“僵尸粉”或“水军”, 因此粉丝数的多少并不能直接代表用户的影响力. 因此在对用户影响力进行衡量时需要剔除这一部分粉丝.

定义 1. 用户活跃度和用户活跃粉丝数. 假设用户在统计时间为 T 的时间段内, 进行原创微博数量为 U_0 , 转发微博数量为 U_r , 提及他人的微博数量为 U_a . 现定义用户活跃度指标为

$$A_i = \frac{SUM(U_0 + U_r + U_a)}{T} \quad (2)$$

当用户活跃度 A_i 低于一定的阈值的时候, 可判定该微博用户为不活跃用户, 在进行粉丝数统计时候不予考虑. 因此, 剔除掉不活跃用户后, 将用户活跃粉丝数记为 A .

其次, 从用户行为的角度考虑. 微博中用户的行为主要包括: 转发、评论、提及. 其他行为例如, 主持话题、参与微博活动、私信等等均是在这三种主要行为上的延伸. 另一方面, “提及”这一行为分主动和被动两种情况. 除去由转发这种情况引起的提及外, 不难看出, 无论是主动提及他人或者是被他人提及, 均是两个人或少数人之间的互动关系. 从中也只能反映出少数人之间有联系. 因此, 在分析用户影响力这一问题上, 仅考虑转发和评论两种用户行为. 下面给出相关定义.

定义 2. 平均转发数和平均评论数. 由微博中的“微博热搜”功能可知, 在热搜排行榜中的话题或人物都是能在短时间内引起网友大量的转发和评论, 从而成为关注焦点, 具有巨大的影响力. 针对这两种用户行为, 现定义平均转发数 \bar{R} 和平均评论数 \bar{C} 两个指标分别为

$$\bar{R} = \frac{\sum_{i=1}^n x_{ir}}{T} \quad (3)$$

$$\bar{C} = \frac{\sum_{i=1}^n x_{ic}}{T} \quad (4)$$

其中, T 为统计时间段, n 为该时段内用户发布微博总数 (包括原创和转发), x_{ir} 和 x_{ic} 表示第 i 个用户对该条微博是否转发和评论情况. 若转发, 则 $x_{ir} = 1$, 否则为 0, x_{ic} 同理. 其中第 i 个用户属于活跃用户集合 A .

最后, 从用户个人身份角度考虑. 对个人而言, 可以认证职业、作品或者获奖成就等. 并且对发微博数、粉丝数、好友质量等均有要求. 由此可以看出, 认证用户较普通用户而言, 身份更加真实可信. 因此, 在传播消息的过程中, 会更加得到关注, 影响范围更广.

现将新浪微博 2017 年 2 月每天热搜排行榜前 20 名的用户根据是否认证这一指标进行对比分析. 结果如图 1 所示.

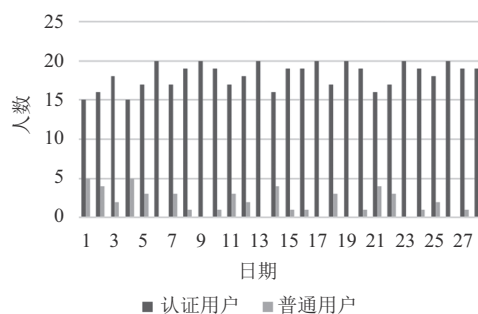


图 1 2017 年 2 月热搜前 20 名用户认证数对比图

可以看到, 在这一个月的热搜排行榜中, 微博认证用户的占比达到 95% 以上. 由此可见, 是否是认证用户在影响力评估过程中有较大影响.

综上所述, 结合对微博用户关系、用户行为和个人身份这三方面的分析, 最终选择用户活跃粉丝数 A 、是否认证 V 、发布微博数 N 、平均转发数 \bar{R} 和平均评论数 \bar{C} 这五个指标.

2.2 指标权重的确定

常用的确定权重的方法分为主观赋权法和客观赋权法^[12]. 其中主观赋权法包括德尔菲法 (专家法)、相邻指标比较法、层次分析法等. 相应地, 客观赋权法包括模糊定权法、熵权法、相关系数法等. 为了全面客观地衡量用户影响力大小, 本文在对评价指标权重进行计算的过程中, 选取了主观赋权法中的层次分析法^[13]和客观赋权法中的熵权法^[14]进行对比.

2.2.1 层次分析法

层次分析法的特点是利用 1-9 之间整数及其倒数作为标度来构造两两比较的判断矩阵. 比例的标度及含义见表 2.

表 2 比例标度

比例标度值	含义
1	同等重要
3	略微重要
5	重要
7	很重要
9	绝对重要
2,4,6,8	两相邻判断的中间值

通过判断矩阵可以求出各指标的权重分配. 目前, 在层次分析法的应用中, 一般都用近似的方法来求得最大特征值及对应特征向量的近似解. 如公式 (5) 和 (6) 所示.

$$w_i = \sqrt[n]{\prod_{j=1}^n a_{ij}} \quad (5)$$

标准化后可得评价指标的权重为:

$$w_i = \frac{\bar{w}_i}{\sum_{j=1}^n \bar{w}_j} \quad (6)$$

2.2.2 熵权法

根据信息论中对熵的定义, 熵权法充分利用了熵的特性, 即利用熵值来判断指标的离散程度. 熵值越大表示指标的离散程度越大. 从而对综合评价的结果影响也越大.

不同于层次分析法, 熵权法在计算权重时是按列进行概率比重化处理, 如公式 (7) 所示.

$$p_{ij} = \frac{z_{ij}}{\sum_{i=1}^n z_{ij}} \quad (7)$$

计算指标熵值:

$$E_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (8)$$

评价指标的权重为:

$$w_j = \frac{1 - E_j}{\sum_{j=1}^m (1 - E_j)} \quad (9)$$

3 改进模糊综合评价算法

现实情况下, 人们对事物的综合评价常常有着不同的方式. 根据实际需求, 人们有时会需要知道最突出

的指标. 有时则要求全面地考虑各个指标. 这些情况都可以通过不同的运算来实现.

考虑到在对用户影响力进行评估时, 指标平均转发数和平均评论数的重要性略高于其他指标. 在实际情况中, 可能出现权重并不能完全反映指标的重要程度的情况. 也就是说, 倘若某个评价指标远大于其他评价指标时, 可以认为该指标非常具有区分度. 但可能会由于权重的原因, 在应用某些算子的时候这个指标的影响力会被削弱, 从而导致最终的评价结果与实际情况不符.

为解决上述问题, 现对模糊合成算子进行改进, 其形式如下:

$$y = [1 + \lambda_1 w_1 x_1 + \lambda_2 w_2 x_2 + \dots + \lambda_i w_i x_i]^{\frac{1}{\lambda}} = \left[1 + \sum_{i=1}^n \lambda_i w_i x_i \right]^{\frac{1}{\lambda}} \quad (10)$$

其中, λ_i 表示所选指标对评价结果具有的影响力大小. 为便于计算, 将 λ_i 设定为 $\lambda_i \geq 1$. 若该指标对评价结果影响力不突出, 则取 $\lambda_i = 1$. $w = [w_1, w_2, \dots, w_n]$ 为权重集. 其中 $w_i \geq 0$ 且 $\sum_{i=1}^n w_i = 1$. 记 $X = [x_1, x_2, \dots, x_n]$ 且 $x_i > 1$.

(1) 单调性论证

现假设存在 $w'_i = w_i$, 若 $x'_i > x_i$, 那么

$$y(\lambda_i; w'_i; x'_i) > y(\lambda_i; w_i; x_i) \quad (11)$$

证明: 考虑到 y 的展开式, 必然存在子项 $(\lambda_i w_i x_i)^{\frac{1}{\lambda}}$. 由于 $\lambda_i \geq 1, 0 \leq w_i < 1$, 且 $x'_i > x_i$, 那么 $f(x) = w^x$ 是单调递减函数, 则 λ 越大, $w^{\frac{1}{\lambda}}$ 也越大. 因此由单调递增性可知, 必然存在 (11) 式. 即该函数是单调递增的. 也就是说, 若一个对象的评价指标值大于另一个评价指标值, 那么最终的评价结果也应该如此, 这也是符合实际情况的.

(2) 极限值论证

由于 $\lim_{\lambda \rightarrow \infty} \lambda^{\frac{1}{\lambda}} = 1, \lim_{\lambda \rightarrow \infty} (w_i x_i)^{\frac{1}{\lambda}} = 1$, 因此当 $\lambda \rightarrow \infty$ 时, 存在如下等式:

$$\lim_{x_i \rightarrow x'_i} y(\lambda_i; w_i; x_i) = \lim_{x_i \rightarrow x'_i} y(\lambda_i; w'_i; x'_i) \quad (12)$$

因此评价结果有上界, 这种函数会随着自变量的增大而增大, 但该函数值是平稳增加, 不存在跳跃点.

(3) 各个指标评价因素相同时论证

当 $x_1 = x_2 = \dots = x_i = C$ 时, 存在

$$y(\lambda_i; w; x_i) \geq c^{\frac{1}{\lambda}} \cdot \left[n(\lambda_1 x_1 \cdot \dots \cdot \lambda_n x_n)^{\frac{1}{n\lambda}} \right] \quad (13)$$

证明:

$$\begin{aligned}
 y(\lambda_i; w_i; x_i) &= [1 + \lambda_1 w_1 x_1 + \lambda_2 w_2 x_2 + \dots + \lambda_i w_i x_i]^{\frac{1}{\lambda}} \\
 &= [1 + C(\lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_i w_i)]^{\frac{1}{\lambda}} \\
 &\geq C^{\frac{1}{\lambda}} \cdot (\lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_i w_i)^{\frac{1}{\lambda}} \\
 &\geq c^{\frac{1}{\lambda}} \cdot \left[n(\lambda_1 x_1 \cdot \dots \cdot \lambda_n x_n)^{\frac{1}{n\lambda}} \right]
 \end{aligned}
 \tag{14}$$

令 $H = \left[n(\lambda_1 x_1 \cdot \dots \cdot \lambda_n x_n)^{\frac{1}{n\lambda}} \right]$, 则

$$\begin{aligned}
 H &\leq \left[\frac{1}{n} (n\lambda_1 x_1 + n\lambda_2 x_2 + \dots + n\lambda_n x_n) \right]^{\frac{1}{\lambda}} \\
 &= (\lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_i w_i)^{\frac{1}{\lambda}}
 \end{aligned}
 \tag{15}$$

当且仅当 $n\lambda_1 w_1 = n\lambda_2 w_2 = \dots = n\lambda_n w_n$ 时取等. 当 $n\lambda_1 w_1, n\lambda_2 w_2, \dots, n\lambda_n w_n$ 不恒等时,

$$H < (\lambda_1 w_1 + \lambda_2 w_2 + \dots + \lambda_i w_i)^{\frac{1}{\lambda}}
 \tag{16}$$

因此, 加入 λ 后, 可以根据需求调整权重与评价结果之间的关系. 减小 λ 意味着重视权重的影响. 反之, 增大 λ 表示评价结果受评价指标的影响更大.

综上所述, 通过对改进后的算法进行推导论证, 证明理论上能够实现根据需求调整权重与评价结果之间得而关系. 从而使评价结果更客观.

4 实验方法和结果

示例. 选取微博用户“迪士尼电影”与 2017 年 1 月 31 日发布的一条关于电影《美女与野兽》预告片的原创微博. 截至爬取这条微博相关内容的时, 这条微博共获得了 26532 人转发, 4120 条评论以及 8025 个点赞.

下面将根据本文提出的模糊综合评价模型对参与这条微博传播的用户进行影响力评估. 爬取到该条微博的部分相关内容如表 3 所示.

表 3 爬取微博部分内容示例

ID	Like	Repost	Comm	Pub_Time
1885831407	223	6718	30	2017-1-31 11:32
5037658133	312	6165	52	2017-1-31 22:47
3170785359	217	319	17	2017-1-31 15:26
.....

4.1 评价指标权重的确定

社交网络中用户影响力特征选取的五个维度分别是: 活跃粉丝数、发微博数、是否认证、平均评论数和平均转发数. 设下标 $i = 1, 2, 3, 4, 5$ 分别对应这五个维度. 则构造对比矩阵 R 如下:

$$\begin{bmatrix}
 1 & 1 & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\
 1 & 1 & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\
 3 & 3 & 1 & \frac{1}{2} & \frac{1}{3} \\
 4 & 4 & 2 & 1 & \frac{1}{2} \\
 5 & 5 & 3 & 2 & 1
 \end{bmatrix}
 \tag{17}$$

根据层次分析法确定的发微博数、活跃粉丝数、是否认证、平均评论数和平均转发数这五个指标的权重为 $[0.069, 0.069, 0.170, 0.269, 0.422]^T$, 而根据熵权法^[15]确定的这五个指标权重为 $[0.185, 0.185, 0.193, 0.202, 0.234]^T$. 如表 4 所示. 由此可以看出, 不论是由层次分析法或者熵权法, 得到的对用户影响力影响最大的指标均是平均被转发数.

表 4 权重比较表

方法	用户活跃 粉丝数	发微博 数	是否 认证	平均 评论数	平均 转发数
层次 分析法	0.069	0.069	0.17	0.269	0.422
熵权法	0.185	0.185	0.193	0.202	0.234

为避免出现非一致性的情况, 下面需进行一致性检验. 公式如下:

$$CR = \frac{CI}{RI}
 \tag{18}$$

其中, $CI = (\lambda \max - n) / (n - 1)$.

由公式 (18) 可求得一致性比率 $CR = 0.0697$. 由于 $CR < 0.1$ 故符合一致性检验, 即权重分配合理. 可进行下一步实验.

4.2 对比实验及结果

为了验证本文提出的改进模糊综合评价算法是否有效, 分别取加权平均型算子和主因素决定型算子建立模型进行对比实验.

记主因素决定型算子建立的模型为 M_1 , 加权平均型算子建立的模型为 M_2 , 改进后的模糊合成算子模型记为 M_3 . 其中 M_3 包含不同取值 λ_i 时的不同结果. 现将根据层次分析法进行权重确定的实验记为 A 组, 同理将熵权法进行权重确定的实验记为 B 组. 结果见表 5 和表 6.

表 5 A 组实验结果

模型	发微博数	用户活跃粉丝数	是否认证	平均评论数	平均转发数
M1	0.181	0.263	0.192	0.237	0.428
M2	0.178	0.182	0.129	0.221	0.289
M3-1	0.125	0.134	0.165	0.226	0.363
M3-2	0.142	0.155	0.172	0.31	0.322
M3-3	0.198	0.124	0.166	0.249	0.304

表6 B组实验结果

模型	发微博数	用户活跃粉丝数	是否认证	平均评论数	平均转发数
M1	0.185	0.202	0.19	0.23	0.334
M2	0.184	0.175	0.125	0.225	0.29
M3-1	0.169	0.172	0.183	0.231	0.285
M3-2	0.159	0.168	0.201	0.242	0.307
M3-3	0.193	0.178	0.244	0.278	0.301

将A组和B组的实验结果绘成折线图如图2和图3所示。

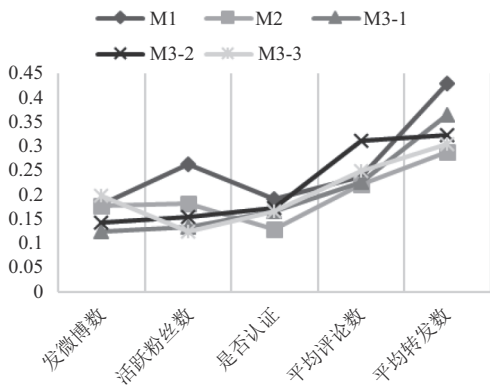


图2 A组实验结果

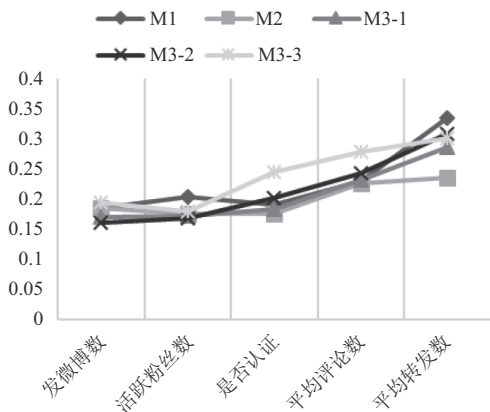


图3 B组实验结果

根据图2和图3结果可以看到,在确定权重方面,由层次分析法确定的权重对五个指标的区分度都高于由熵权法确定的权重.另一方面, M_2 模型由于采用了加权平均算子,因此对五个指标进行了平均,权重中平均转发数这一指标被削弱,而发微博数这一指标被加强.另外, M_1 模型采用主因素决定型算子,因此平均转发数这一指标明显高于其他指标,且其他四个指标之间的差异则变得不明显,不利于综合评判用户的影响力。

相比之下,改进后的模糊合成算子表现较好.由于

可以调整权重对评价结果的影响,因此根据 λ_i 的变化, M_{3-1} 、 M_{3-2} 和 M_{3-3} 的表现也体现出了差异.经过多次实验,该算法既能够有效区分各个指标,又综合考虑到多个指标对评价结果的影响,同时也不会过度重视权重占比最大的指标。

因此,采用改进模糊综合评价模型对涉及这条微博的用户进行影响力评估,推荐出前5名影响力最高的用户,他们的排名和综合得分如表7所示。

表7 影响力评分值前5名用户

用户昵称	影响力评分值
田喜碧Hebe	0.473
井柏然	0.425
迪士尼电影	0.294
艾玛沃特森资讯	0.183
小井先生工作室	0.179

另外,用户影响力与粉丝数并不一定成正比关系.现将影响力评分值前7名用户的粉丝数显示在表8中.由此可见,定义活跃用户粉丝数这一指标是有必要性的。

表8 影响力评分值前7名用户的粉丝数

用户昵称	粉丝数
田喜碧Hebe	966万
井柏然	2184万
迪士尼电影	462万
艾玛沃特森资讯	44万
小井先生工作室	14万
Taurus	392
Rajor	16万

5 结语

本文以新浪微博为例,通过分析社交网络中的用户的关系、用户的行为以及个人身份,选取发微博数、活跃粉丝数、是否认证、平均评论数和平均转发数作为判断用户影响力的指标.并对比了使用层次分析法和熵权法确定评价指标的权重过程中的差异.目前在社交网络中应用模糊综合评价法的难点在于如何合理地在将用户的影响力转为定量评价.考虑到传统的模糊合成算子并不能充分利用指标的实际意义,因此本文对模糊综合评价法进行了改进,提出了一种新的模糊矩阵合成方法评估用户影响力.经实验证明评价结果较为合理。

同时本文仍存在一些没有解决的问题.一方面,本文提出的改进算法耗时较长,并且在确定可调参数 λ_i 时,需要尝试多次实验才能取得较满意的结果.因此,

在提高算法的效率方面还有待加强. 另一方面, 本文仅针对人为选择的某一条微博中涉及的用户进行影响力评估, 涉及的范围较窄, 如何在更大的网路络中定量计算用户的影响力还需进一步的研究.

参考文献

- 1 丁兆云, 贾焰, 周斌, 等. 社交网络影响力研究综述. 计算机科学, 2014, 41(1): 48–53.
- 2 康书龙. 基于用户行为及关系的社交网络节点影响力评价——以微博研究为例[硕士学位论文]. 北京: 北京邮电大学, 2011.
- 3 Yamaguchi Y, Takahashi T, Amagasa T, *et al.* TURank: Twitter user ranking based on user-tweet graph analysis. Proc. of the 11th International Conference on Web Information Systems Engineering. Hong Kong, China. 2010. 240–253.
- 4 贾冲冲, 王名扬, 车鑫. 基于 HRank 的微博用户影响力评价. 计算机应用, 2015, 35(4): 1017–1020. [doi: [10.11772/j.issn.1001-9081.2015.04.1017](https://doi.org/10.11772/j.issn.1001-9081.2015.04.1017)]
- 5 宫秀文, 张佩云. 基于 PageRank 的社交网络影响最大化传播模型与算法研究. 计算机科学, 2013, 40(S1): 136–140.
- 6 胡勇, 张翀斌, 王楨学, 等. 网络舆论形成过程中意见领袖形成模型研究. 四川大学学报(自然科学版), 2008, 45(2): 347–351.
- 7 Riquelme F, González-Cantergiani P. Measuring user influence on twitter. Information Processing and Management: An International Journal, 2016, 52(5): 949–975. [doi: [10.1016/j.ipm.2016.04.003](https://doi.org/10.1016/j.ipm.2016.04.003)]
- 8 Zhang Y, Mo JQ, He TT. User influence analysis on micro blog. Proc. of the 2nd International Conference on Cloud Computing and Intelligent Systems. Hangzhou, China. 2012. 1474–1478.
- 9 Kwak H, Lee C, Park H, *et al.* What is Twitter, a social network or a news media? Proc. of the 19th International Conference on World Wide Web. Raleigh, North Carolina, USA. 2010. 591–600.
- 10 张跃, 邹寿平, 宿芬. 模糊数学方法及其应用. 北京: 煤炭工业出版社, 1992: 15–28.
- 11 李洪兴, 汪培庄. 模糊数学. 北京: 国防工业出版社, 1994: 67–72.
- 12 陈衍泰, 陈国宏, 李美娟. 综合评价方法分类及研究进展. 管理科学学报, 2004, 7(2): 69–79.
- 13 郭金玉, 张忠彬, 孙庆云. 层次分析法的研究与应用. 中国安全科学学报, 2008, 18(5): 148–153.
- 14 章穗, 张梅, 迟国泰. 基于熵权法的科学技术评价模型及其实证研究. 管理学报, 2010, 7(1): 34–42.
- 15 李玉琳, 高志刚, 韩延玲. 模糊综合评价中权值确定和合成算子选择. 计算机工程与应用, 2006, 42(23): 38–42. [doi: [10.3321/j.issn:1002-8331.2006.23.012](https://doi.org/10.3321/j.issn:1002-8331.2006.23.012)]