

Analysis of User Influence of Social Investment Platform—Taking Snowball Network as an Example

Shaohua Lou¹, Man Zhou², Qixing Qu¹

¹School of Information Technology & Management, UIBE, Beijing

²Department of Electronic and Electrical Engineering, University College London, London

Email: loushaohua1121@163.com

Received: Oct. 29th, 2019; accepted: Nov. 14th, 2019; published: Nov. 21st, 2019

Abstract

With the popularity of Internet technology, mobile social media has gradually entered people's lives, but at the same time all kinds of complex information noise has increased the cost of screening effective information. At present, the representative noise reduction method is to evaluate the influence of users, but few studies on influence models in the vertical field at home and abroad. Therefore, this paper explores how to establish a user impact assessment model by analyzing Snowball Network, a mature investment socialization platform, so as to enrich the academic research in this field and also provide a low-cost, high-efficiency information screening environment for investors in social platforms. Combining data characteristics, this paper selects the PCA method to determine the weights of 10 indicators such as the number of forwarding, fans, portfolio concerns. The coefficients of each index in the linear combination of principal components are calculated by the principal component and variance contribution rate, and then the weighted average and normalized calculations are performed to obtain the final influence model. In order to verify its validity, the AHP and entropy method were used to calculate the influence again. The experimental results show that the user impact assessment method proposed in this paper can be effectively applied to social platform of investment field and accurately present the influence of different users.

Keywords

User Influence, PCA Principal Component Analysis, Social Investment Platform

社交投资平台用户影响力分析——以雪球网为例

娄韶华¹, 周曼², 屈启兴¹

¹对外经济贸易大学信息学院, 北京

²伦敦大学学院电子与电气工程学院, 伦敦
Email: loushaohua1121@163.com

收稿日期: 2019年10月29日; 录用日期: 2019年11月14日; 发布日期: 2019年11月21日

摘要

随着互联网技术的普及, 移动社交媒体逐渐步入人们的生活, 但各类复杂的信息噪声也增加了人们筛选有效信息的成本。目前具有代表性的降噪方法是评估用户的影响力, 但目前国内外在垂直领域方面研究的影响力模型较少。因此, 本文选取了雪球网这一成熟的投资领域社会化平台, 探究在该领域如何建立用户影响力评估模型, 以丰富用户影响力相关的学术研究, 同时也为社交平台的投资者提供一个低成本、高效率的信息筛选环境。通过借鉴、分析不同平台的影响力模型, 结合数据特征, 本文选择了主成分分析法, 以确定转发数、粉丝数等10个指标的权重。通过各主成分及方差贡献率计算各指标在主成分线性组合中的系数, 再进行加权平均和归一化计算, 得到了最终的影响力模型。为了验证其有效性, 本文采取了层次分析法和熵值法, 再次计算影响力并进行对比。实验结果显示, 本文所提出的用户影响力评估方法, 能够有效地适用于投资领域的社交平台, 准确呈现不同用户的影响力。

关键词

用户影响力, PCA主成分分析, 社交投资平台

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

1.1. 社交平台与用户影响力

纵观社交平台发展的近二十年, 可以发现信息的流动性和复杂性在呈指数型上升, 信息的呈现方式日新月异, 但含金量却参差不齐。1999年诞生的腾讯QQ着眼于即时通讯领域, 而后在该平台上的QQ空间成为信息发布和交流最为密集的网络。2005~2010年, 新浪微博、人人网、开心网等社交网站上线, 以博文、帖子等方式呈现的网络信息开始爆发。2010年后, 现象级社交应用微信的诞生, 标志着移动端社交成为主流阶段, 各类的公众号推送令人眼花缭乱。垂直社交领域也同样面临信息过于冗杂、妨碍用户正常使用平台的问题。在最为讲求信息对称性的投资社交领域, 鱼龙混杂的投资信息将显著降低平台的有效性。而信息的有效性由发布该信息的用户决定, 通过判别和筛选更有影响力的用户, 可以帮助社交平台推荐有效信息, 过滤无效信息。

用户影响力评价在目前的各类社交平台上应用甚广, 体现形式多样。以国内一些热门的社交平台或带有社交功能的平台为例: 微博的实时影响力评价指标有昨日阅读微博数和昨日互动数(转发微博、评论微博、赞微博、回复评论、赞评论的数量); 知乎的影响力评价包括了赞同数、感谢数、收藏数; 抖音则包括了获赞数、关注数、粉丝数; 而36氪只包括了关注数。在垂直社交领域也比较类似: 专业IT技术社区CSDN以访客人数为指标, 对所有用户进行排名作为影响力评估方式; 职场社交平台LinkedIn以好

友数和关注者人数为评估指标；而创业服务社交平台 36 氪只包括了关注数。国外软件的影响力评估也主要以关注数、点赞数、转发数等指标反映。Twitter 使用正在关注人数和已关注人数衡量，Facebook 使用点赞数和超级粉丝拥有数衡量，Instagram 则是纳入了帖子数、粉丝数、关注数等指标。通过评估用户的影响力，平台可以清晰地发掘社交网络中的意见领袖，准确的把握舆论导向，提取实时热点，并激励用户发布优质内容；用户可以便捷地发现高质量用户发布的内容，减少信息筛选的时间，提升浏览效率。社交网络的用户影响力对信息的传播也有显著影响[1]，高影响力的用户的产出内容往往更易传播。

1.2. 研究意义

纵览各平台的影响力评估方式，独立的单指标影响力评价体系目前是主流。但这种评估方法不具备综合评估的能力，难以结合各个方面合理的进行影响力评估，故而学术界一直致力于研究社交平台的用户影响力评估新方法。目前来看，大部分研究模型都针对微博、Twitter 等大型热门社交平台，例如 TwitterRank, WBRank, DomainRank 等。此类模型适应主题繁多、集群复杂的社交平台。但相比之下，针对垂直领域社交平台的模型研究几乎是一片空白。近年来新兴的专业领域社交平台异军突起，成为了分割市场的有力力量，在投资领域最有代表性的雪球网就是一个很好的例子。从 2011 年的 300 万美元红杉资本的 A 轮到 2018 年蚂蚁金服 1.2 亿元的 D 轮，这家融合社交和交易的投资平台成功获得了广大投资者的青睐，有效的 UGC (User Generate Content) 机制让越来越多的投资者在该平台分享、讨论投资信息。这便要求该平台拥有良好的用户影响力评价体系，以降低用户的信息筛选成本，减少多余的信息噪音，鼓励创造优质内容，营造一个良好向上的平台环境。本文通过分析雪球网 8937 条用户数据，为以投资领域为代表的垂直领域社交平台提供了一种评价用户影响力的方法和模型，帮助垂直社交平台营造一个低成本、高效率的信息筛选环境，同时也丰富了用户影响力相关的学术研究，为该领域的探索前行提供新鲜力量。

2. 文献综述

2.1. 影响力有关的研究

影响力传播模型最早可追溯到 1927 年由生物学家 Kermack 和 Mckendrick 提出的传染病模型，该模型最初用来对人群中疾病的传播过程进行建模[2]。针对社交网络影响力的研究则始于 1984 年，被誉为“影响力教父”的 Robert Cialdini 发表了《Influence: The Psychology of Persuasion》一书，提出了决定影响力的六大原则：互惠性，承诺和一致性，社会证明，权威性，偏好，稀缺性[3]。进入互联网时代后，在社交网络节点影响力的评估方法主要可以分为两类：基于指标权重的影响力分析和基于关系网络的影响力分析。

根据《辞海》，影响力是指一方发生一种动作而引起他方发生变化或行动的力量。在社交网络的研究中，影响力一般指在线社交网络中的用户影响网络中的其他用户改变自己的行为和思想的能力。在现有的研究中，影响力一般被分为两种：个人影响力和群体影响力。个人影响力以社交网络上某一个节点为中心，探讨其对周围节点的影响。而群体影响力需集合多个相关节点，探讨该集群的行为活动对集群周边的节点和其他集群的影响力。本文以雪球网上的单个节点为主作分析，故在分析中以个人影响力为基准。

2.2. 基于指标权重的影响力分析

基于指标权重是指以网络中单个节点为中心，对其静态的属性特征赋予权重，进而计算其影响力。如度中心性，认为一个节点的邻居节点越多，影响力越大，在网络中就越重要。在有权网络中，节点的

度可以看作强度，即边的权重之和。度中心性刻画了节点的直接影响力[4]，在粉丝数这一指标下，一个粉丝就是该节点一个邻居节点，该节点的度即为总粉丝数。当点赞数、评论数、转发数等同时纳入指标时，一个节点就有了多重身份(指标)，每一指标的权重乘以该指标下的度决定了节点的强度，即影响力。

指标的确定由影响力形成的过程决定。在这方面，McGuire (1989)提出的 12 个步骤得到了广泛的认可，并在后续逐步简化归纳为“接触 - 认知 - 说服 - 二次传播”四个环节。许多指标可以对应这四个环节，比如关注数(接触)、点赞数和评论数(认知)、收藏数(说服)、转发数(二次传播)等。

明确各指标后，则需确定权重。常用的方法分主观和客观两类。主观分为 Delphi 专家调查法，AHP 层次分析法、二项系数法、环比评分法等；客观分为 PCA 主成分分析法、熵值法、均方差法等。刘海涛(2017)利用层次分析法构建了新的综合排序方法[5]。魏杰明(2019)即基于主成分分析算法，系统研究了社交网络中用户行为和贴文特征，将各组成因素进行相关性研究，得到了社交网络节点影响力函数表达式[6]。李晓(2016)采用熵值法和综合指数模型测算出了十七地市公安微博的受众影响力综合指数，提出了提升其受众影响力的对策建议[7]。

可以看出，基于指标权重的算法特点是直观、计算复杂度低，模型相对静态。在主题垂直、集群鲜明、指标明确的社交网络区域应用效果较好。

2.3. 基于关系网络的影响力分析

基于关系网络的影响力算法关注节点在整个网络中的重要性。经典的算法是 90 年代末期提出的 PageRank [8]和 HITS 算法(Hyperlink-Induced Topic Search) [9]。这两种算法原本应用在万维网中用来评估网页的流行性，而社交平台中的关注和粉丝关系与网页的链入与链出十分相似，因此它们也被应用在了社交网络中节点影响力的评估中。PageRank 算法模型，是 Google 在搜索引擎结果中对网站排名的核心算法，核心思想是通过计算页面链接的数量和质量，来估计网站的重要性，HITS 算法模型中，有两类节点：权威节点和枢纽节点。权威节点在网络中具有高权威性，枢纽节点具有很个指向边的节点。通过计算网络中每个节点的权威值(Authority)和枢纽值(Hub)来寻找高权威性的节点。即求值过程是在迭代中计算 Authority 和 Hub 值，直到收敛状态。郭博等(2018)即通过提出改进的 PageRank 算法和 HITS 算法,分别基于知乎用户社交网络、问答网络构建用户影响力挖掘模型[10]。黄贤英等(2019)也基于 PageRank 算法引入了用户博文的传播率来挖掘用户的潜在影响力[11]。

除了上述经典的算法，越来越多的创新算法被提出。陈志云(2013)借用学术论文评价中的 H-Index 来评价微博中用户的影响力。即一名微博用户的 h 指数是指他至多有 h 个粉丝数超过 h 的粉丝。王仁武等(2018)利用领域字典和话题识别模型将目标用户的主题范围进行限定，同时结合社交网络用户中的个人信息综合指标，基于用户关注关系建立链路网络，并充分纳入用户评论的情感评分，提出了针对专业影响力节点挖掘的 DomainRank 算法[12]。刘威等(2019)提出了一种面向微博话题的用户影响力分析算法——基于话题和传播能力的用户排序 TSRank 算法[13]。

可以看到，以上算法基于动态的用户关系网络提出，关注节点在整个网络中的动态强度，比基于权重指标的算法更加灵活，但在算力的需求远大于后者。适用于集群复杂、主题丰富的社交网络。

结合两种方式的优劣、现有的数据和雪球网平台垂直领域的特性，本文选取了第一种方式中的主成分分析法作分析。

3. 数据的收集与分析

3.1. 数据概况

本文选取了雪球网 8937 条清洗后的用户数据。识别码为用户编号(user ID)根据“接触 - 认知 - 说服

- 二次传播”四个环节，分别抽取了粉丝数、关注数、微博数、关注股票数、投资组合数、认证、投资组合关注数、评论数、转发数、收藏数十个指标作为变量，见表 1。除了认证指标为布尔值(已认证为 1，未认证为 0)，其他变量均为正整数。其中，关注数和粉丝数属于“接触”环节，最直接的体现了该节点的度。微博数、关注股票数、投资组合数代表了该用户在平台上展现的内容数量，属于“认知”环节。认证和投资组合关注数体现了平台和社会证明给该用户的背书，归为“说服”环节。最后，评论数、转发数、收藏数属于“二次传播”环节，因为评论、转发和收藏都起到了创造社会证明、扩大影响力的作用。

Table 1. Variable table
表 1. 变量表

传播环节	雪球用户信息	标志符
接触	粉丝数	FC
	关注数	AC
认知	微博数	TC
	关注股票数	SC
	投资组合数	GC
说服	认证	V
	投资组合关注数	GFC
二次传播	评论数	TRC
	转发数	RTC
	收藏数	TFC

3.2. 数据有效性分析

通过 SPSS 软件，我们可以轻松的得到各主成分的得分，但第一步需要检验数据是否具有的一定相关性，即是否通过 KMO 检验。标准规定 0.6 以上即可通过检验。

Table 2. KMO and Bartlett test
表 2. KMO 和巴特利特检验

KMO 取样适切性量数		0.628
巴特利特球形度检验	近似卡方	29950.172
	自由度	45
	显著性	0.000

检验结果得出 KMO 值为 0.628，表明数据可使用主成分分析法求权重。

从表 2 可知，前 4 个主成分对应的特征根 > 1，提取前 4 个主成分的累计方差贡献率达到 66.109%，超过 60%。因此前 4 个主成分基本反映了指标的信息，可以代替原来的 10 个指标，见表 3。

从表 4 可知第一主成分与第二主成分对原来指标的载荷数。例如，第一主成分对认证指标的载荷数为 0.008。

4. 影响力计算

4.1. 确定指标权重

用主成分分析确定权重的方法是：指标权重等于以主成分的方差贡献率为权重，对该指标在各主成

分线性组合中的系数的加权平均的归一化。

Table 3. Total variance interpretation
表 3. 总方差解释

成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%
1	3.077	30.769	30.769	3.077	30.769	30.769
2	1.492	14.917	45.686	1.492	14.917	45.686
3	1.036	10.360	56.047	1.036	10.360	56.047
4	1.006	10.062	66.109	1.006	10.062	66.109
5	0.894	8.945	75.054			
6	0.827	8.266	83.319			
7	0.643	6.433	89.753			
8	0.592	5.924	95.677			
9	0.368	3.682	99.359			
10	0.064	0.641	100.000			

提取方法：主成分分析法。

Table 4. Component matrix
表 4. 成分矩阵

	成分			
	1	2	3	4
V	0.008	-0.060	-0.007	0.987
GF	0.046	0.067	0.886	0.027
TC	0.619	0.376	-0.100	0.040
FC	0.674	-0.040	0.070	0.130
AC	0.328	0.645	-0.198	0.030
SC	0.191	0.723	-0.094	-0.008
GC	0.222	0.396	0.431	-0.040
TFA	0.730	-0.381	0.016	-0.087
TRC	0.831	-0.126	-0.036	-0.011
RTC	0.907	-0.290	-0.023	-0.053

因此，要确定指标权重需要明确三点：指标在各主成分线性组合中的系数、主成分的方差贡献率以及指标权重的归一化。

(1) 指标在不同主成分线性组合中的系数，见表 5。

用表 4 中的载荷数除以表 2 中第 1 列对应的特征根的开方，即可求出指标在不同主成分线性组合中的系数。

例如，在第一主成分 F_1 的线性组合中，认证变量 $V(X_1)$ 的系数为：

$$X_1 = \frac{0.008}{\sqrt{3.007}} = 0.005$$

按此方法，基于表 2 和表 4 的数据，可在 Excel 中分别计算出各指标在两个主成分线性组合中的系数，见表 6：

Table 5. The number of loads of the indicator and the characteristic root of the principal component
表 5. 指标的载荷数与主成分的特征根

指标	主成分			
	F ₁	F ₂	F ₃	F ₄
V (X ₁)	0.008	-0.06	-0.007	0.987
GF (X ₂)	0.046	0.067	0.886	0.027
TC (X ₃)	0.619	0.376	-0.1	0.04
FC (X ₄)	0.674	-0.04	0.07	0.13
AC (X ₅)	0.328	0.645	-0.198	0.03
SC (X ₆)	0.191	0.723	-0.094	-0.008
GC (X ₇)	0.222	0.396	0.431	-0.04
TFA (X ₈)	0.73	-0.381	0.016	-0.087
TRC (X ₉)	0.831	-0.126	-0.036	-0.011
RTC (X ₁₀)	0.907	-0.29	-0.023	-0.053
主成分的特征根	3.077	1.492	1.036	1.006

Table 6. Coefficients of each indicator in the linear combination of principal components
表 6. 各指标在主成分线性组合中的系数

指标	主成分			
	F ₁	F ₂	F ₃	F ₄
V (X ₁)	0.005	-0.049	-0.007	0.984
GF (X ₂)	0.026	0.055	0.870	0.027
TC (X ₃)	0.353	0.308	-0.098	0.040
FC (X ₄)	0.384	-0.033	0.069	0.130
AC (X ₅)	0.187	0.528	-0.195	0.030
SC (X ₆)	0.109	0.592	-0.092	-0.008
GC (X ₇)	0.127	0.324	0.423	-0.040
TFA (X ₈)	0.416	-0.312	0.016	-0.087
TRC (X ₉)	0.474	-0.103	-0.035	-0.011
RTC (X ₁₀)	0.517	-0.237	-0.023	-0.053

由此可得各主成分关于十个指标的线性组合：

$$\begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{pmatrix} = \begin{pmatrix} 0.005 & 0.026 & 0.353 & 0.384 & 0.187 & 0.109 & 0.127 & 0.416 & 0.474 & 0.517 \\ -0.049 & 0.055 & 0.308 & -0.033 & 0.528 & 0.592 & 0.324 & -0.312 & -0.103 & -0.237 \\ -0.007 & 0.870 & -0.098 & 0.069 & -0.195 & -0.092 & 0.423 & 0.016 & -0.035 & -0.023 \\ 0.984 & 0.027 & 0.040 & 0.130 & 0.030 & -0.008 & -0.040 & -0.087 & -0.011 & -0.053 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{10} \end{pmatrix}$$

如主成分 F1 的表达式为:

$$F_1 = 0.005X_1 + 0.026X_2 + 0.353X_3 + 0.384X_4 + 0.187X_5 + 0.109X_6 + 0.127X_7 + 0.416X_8 + 0.474X_9 + 0.517X_{10}$$

根据先前的输出结果, 各主成分的方差贡献率如下表所示, 见表 7:

Table 7. Variance contribution rate of each principal component

表 7. 各主成分的方差贡献率

	主成分			
	F ₁	F ₂	F ₃	F ₄
主成分的方差% (方差贡献率)	30.769	14.917	10.36	10.062

根据方差贡献率的加权平均各系数, 见表 8, 可得:

Table 8. Weighted averaged coefficients based on principal component contribution rate

表 8. 根据主成分贡献率加权平均后的系数

指标	V (X ₁)	GF (X ₂)	TC (X ₃)	FC (X ₄)	AC (X ₅)
系数	0.140	0.165	0.224	0.202	0.180
指标	SC (X ₆)	GC (X ₇)	TFA (X ₈)	TRC (X ₉)	RTC (X ₁₀)
系数	0.169	0.192	0.113	0.190	0.176

由于权重相加总和应为 1, 因而对各系数进行归一化处理, 见表 9:

Table 9. Normalized indicator weights

表 9. 归一化后的指标权重

指标	V (X ₁)	GF (X ₂)	TC (X ₃)	FC (X ₄)	AC (X ₅)
系数	0.080	0.094	0.128	0.115	0.103
指标	SC (X ₆)	GC (X ₇)	TFA (X ₈)	TRC (X ₉)	RTC (X ₁₀)
系数	0.096	0.110	0.064	0.109	0.100

由此得到的综合得分模型为:

$$Y = 0.140X_1 + 0.165X_2 + 0.224X_3 + 0.202X_4 + 0.180X_5 + 0.169X_6 + 0.192X_7 + 0.113X_8 + 0.190X_9 + 0.176X_{10}$$

4.2. 计算影响力

将原始数据套入上述模型计算可得样本中所有用户的影响力结果, 以下是进行降序排序后得到的前 20 名影响力最大的用户, 见表 10。

5. 对比验证

为了验证本文所得模型的准确性, 本节根据数据的特征选取了确定权重的另外两种方法——层次分析法与熵权法, 推演出用户影响力排名结果, 与之前所得结果进行比较。

Table 10. Influence calculation results after principal component analysis to determine weights
表 10. 主成分分析确定权重后的影响力计算结果

排名	user ID	影响力 Y	排名	user ID	影响力 Y
1	5335375589	25388.11	11	1496137555	2552.01
2	7051804479	10434.655	12	1748578716	2531.721
3	2785562330	7767.093	13	3905629206	2408.931
4	1744347952	5998.15	14	7889012679	2364.009
5	5267456690	5137.6	15	9943910551	2283.469
6	1645252804	4414.73	16	1304843196	2195.683
7	4047266173	2930.559	17	5514474822	2102.344
8	4505284940	2806.272	18	3191943504	2038.44
9	7813497513	2675.405	19	9472407023	1972.823
10	3793075520	2664.708	20	6157001146	1963.021

5.1. 使用层次分析法计算影响力

层次分析法(Alytic hierarchy process, 简称 AHP)法是美国运筹学家 T.L.Saaty 等人在 20 世纪 70 年代中期提出了一种定性和定量相结合的, 系统性、层次化的多目标决策分析方法。AHP 法的核心是将决策者的经验判断定量化, 增强了决策依据的准确性, 在目标结构较为复杂且缺乏统计数据的情况下更为实用。应用 AHP 法确定评价指标的权重, 就是在建立有序递阶的指标体系的基础上, 通过比较同一层次各指标的相对重要性来综合计算指标的权重系数。具体步骤如下:

1) 构造判断矩阵

构造判断矩阵首先要确定判断的标度。本文的标度定义如下, 见表 11:

Table 11. Scale definition of analytic hierarchy process

表 11. 层次分析法的标度定义

标度	定义
1	两个变量重要性相同
3	两个变量中前者比后者稍微重要
5	两个变量中前者比后者明显重要
7	两个变量中前者比后者特别重要
9	两个变量中前者比后者极端重要
2, 4, 6, 8	重要性位于上述相邻数值的中间
1~9 的倒数	两个变量中后者比前者的重要程度

确定标度后, 本文构建了如下判断矩阵, 并邀请了 20 位专家对四个环节和各指标两两比较评分, 见表 12。

2) 判断矩阵一致性的检验

为了度量不同阶数判断矩阵是否具有满意的一致性, 需引入判断矩阵的平均随机一致性指标 RI 值。

当 $n = 9$ 时, $RI = 1.46$ 。一致性指标 $CI = \frac{\lambda_{\max} - n}{n - 1}$ 。当阶数大于 2, 判断矩阵的一致性比率 $CR = CI/RI < 0.10$ 时, 即认为判断矩阵具有满意的一致性, 否则需要调整判断矩阵, 以使之具有满意的一致性。本方法中所有专家给出的判断矩阵的 CR 值均通过了一致性检验。

Table 12. Judgment matrix of analytic hierarchy process
表 12. 层次分析法的判断矩阵

影响力	V	GF	TC	FC	AC	SC	GC	TFA	TRC	RTC
V	1									
GF	-	1								
TC	-	-	1							
FC	-	-	-	1						
AC	-	-	-	-	1					
SC	-	-	-	-	-	1				
GC	-	-	-	-	-	-	1			
TFA	-	-	-	-	-	-	-	1		
TRC	-	-	-	-	-	-	-	-	1	
RTC	-	-	-	-	-	-	-	-	-	1

3. 计算权重

在一致性检验后, 本方法对 20 名专家打分矩阵得出的结果进行了加权平均。得出了最终的权重系数, 见表 13。

Table 13. Indicator weights under the analytic hierarchy process
表 13. 层次分析法下的指标权重

传播环节	标志符	权重	指标	标志符	权重
接触	X_1	0.441	粉丝数	FC	0.275
			关注数	AC	0.166
			微博数	TC	0.186
认知	X_2	0.242	关注股票数	SC	0.021
			投资组合数	GC	0.035
说服	X_3	0.074	认证	V	0.043
			投资组合关注数	GFC	0.031
二次传播	X_4	0.243	评论数	TRC	0.108
			转发数	RTC	0.135

根据本方法求出的影响力排名, 见表 14, 为:

Table 14. Ranking of influence under analytic hierarchy
表 14. 层次分析法下的影响力排名

排名	user ID	影响力 Y	排名	user ID	影响力 Y
1	2785562330	9253.983	11	9472407023	4191.359
2	5335375589	8346.049	12	4065692543	4030.717
3	5267456690	7633.226	13	9455873430	3986.889
4	1744347952	6797.272	14	3303254887	3881.304
5	1645252804	6385.258	15	1748578716	3713.572
6	7051804479	5388.712	16	8438645214	3687.569
7	4047266173	5380.743	17	8745013228	3610.108
8	3905629206	4525.836	18	6157001146	3580.826
9	9943910551	4222.508	19	1455780333	3569.325
10	7813497513	4201.977	20	8157997076	3534.31

5.2. 验证分析

将本文中方法的结果与层次分析法结果进行比较发现前 7 名都是同一组用户，重合率达到 100%，其中用户“1744347952”、“4047266173”的排名完全相同。同时，在 PCA 排名中的前 20 名有 13 位仍然保持在 AHP 排名前 20 名的榜单中。可以看出本文中的排序方法与 AHP 算法能够很大程度上接近，有效性通过了验证，见表 15。

Table 15. Sorting comparison between AHP and PCA
表 15. AHP 与 PCA 两种方法的排序对比

AHP 排名	PCA 排名	user ID	AHP 排名	PCA 排名	user ID
1	3	2785562330	11	19	9472407023
2	1	5335375589	12	26	4065692543
3	5	5267456690	13	24	9455873430
4	4	1744347952	14	31	3303254887
5	6	1645252804	15	12	1748578716
6	2	7051804479	16	32	8438645214
7	7	4047266173	17	42	8745013228
8	13	3905629206	18	20	6157001146
9	15	9943910551	19	45	1455780333
10	9	7813497513	20	38	8157997076

6. 总结与启示

6.1. 主要结论

通过借鉴、分析不同平台的影响力模型，结合数据特征，本文选择了主成分分析法，以确定转发数、粉丝数、投资组合关注数等 10 个指标的权重。在 KMO 检验确定数据相关性适合分析后，选取了 4 个特征根大于 1 的主成分，其方差贡献率总和达到了 66.11%。通过各主成分及方差贡献率，计算各指标在主

成分线性组合中的系数，再进行加权平均和归一化计算，得到了最终的影响力模型：

$$Y = 0.140X_1 + 0.165X_2 + 0.224X_3 + 0.202X_4 + 0.180X_5 + 0.169X_6 + 0.192X_7 + 0.113X_8 + 0.190X_9 + 0.176X_{10}$$

为了验证其有效性，本文采取了层次分析法，再次计算了影响力并进行了对比。实验结果显示，本文所提出的用户影响力评估方法，能够有效地适用于投资领域的社交平台，准确呈现其中不同用户的影响力。

6.2. 创新点

本文为垂直型社交平台的用户影响力评价提供了一种方案。在实践中具有计算复杂度低、结果准确、易于调整的优势。通过分析雪球网 8937 条用户数据，本文为以投资领域为代表的垂直领域社交平台提供了一种评价用户影响力的方法和模型，帮助垂直社交平台营造一个低成本、高效率的信息筛选环境，同时也丰富了用户影响力相关的学术研究，为该领域的探索前行提供新鲜力量。

6.3. 不足之处

在主成分分析中，本文提取出的四个主成分的累计贡献率不是非常理想，仅有 66.11%，如果达到 80% 以上甚至更高的水平，即变量降维后的信息量保持在一个更高的水平，模型的结果可能会更加具有实践意义。

另外，在验证方法层次分析法中，由于专家打分具有一定的主观性，不同的专家组会给出不同的打分矩阵，在对比时可能存在相对的偏差。下一步的工作是探究能够综合主观和客观的建模方法。

参考文献

- [1] 林青, 李立煊, 杨腾飞. 社交网络用户影响力量化模型研究——以新浪微博为例[J]. 情报杂志, 2018, 37(8): 202-207.
- [2] 于尚尚. 多主题社交网络用户影响力研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2018.
- [3] Cialdini, R. (2009) *Influence: Science and Practice*. Pearson Education, Boston, MA.
- [4] 王梓. 社交网络中节点影响力评估算法研究[D]: [硕士学位论文]. 北京: 北京邮电大学, 2014.
- [5] 刘海涛, 樊重俊. 复杂网络视角下的社交网络用户影响力研究[J]. 科技和产业, 2017(11): 118-121.
- [6] 魏杰明, 何慧. 社交网络中用户行为及影响力评估算法研究[J]. 智能计算机与应用, 2019, 9(2): 162-167, 171.
- [7] 李晓. 政务微博受众影响力评估研究——以山东省十七地市公安微博为例[J].
- [8] Page, L., Brin, S., et al. (1998) The PageRank Citation Ranking. Bringing Order to the Web. Stanford InfoLab., 1-14.
- [9] Kleinberg, J.M. (1999) Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, **46**, 604-632. <https://doi.org/10.1145/324133.324140>
- [10] 郭博, 许昊迪, 雷水旺. 知乎平台用户影响力分析与关键意见领袖挖掘[J]. 图书情报工作, 2018, 62(20): 122-132.
- [11] 黄贤英, 阳安志, 刘小洋, 刘广峰. 一种新的微博用户影响力评估算法研究[J/OL]. 计算机工程: 1-7.
- [12] 王仁武, 周威, 张文慧. 融合用户情感评分的节点专业影响力分析[J]. 现代情报, 2018, 38(7): 54-61.
- [13] 刘威, 张明新, 安德智. 面向微博话题的用户影响力分析算法[J]. 计算机应用, 2019, 39(1): 213-219.