

基于 Apriori 算法的产品众筹项目超额融资率影响因素研究

王 为, 李淑敏, 李 俊, 夏恩君

(北京理工大学 管理与经济学院, 北京 100081)

摘要:利用京东众筹平台的 1 031 个产品众筹项目,通过 Apriori 关联规则算法得到众筹项目的各属性与超额融资率之间的关联规则。研究发现:目标筹集金额与超额融资率之间存在紧密的关联关系;超额融资率与各档金额的设定没有必然的关联关系;在支持人数偏低、点赞人数偏低的情形下,电子科技类产品不易获得较高的超额融资率,这与目标筹集金额的设定没有必然的关联关系;各档支持人数与超额融资率之间存在紧密的关联关系。研究结论对于提高产品众筹超额融资率,进而提升众筹项目成功率具有重要意义。

关键词:产品众筹;超额融资率;关联规则;Apriori 算法

中图分类号:F203.9 **文献标志码:**A **文章编号:**1671-1807(2022)04-0009-10

2011—2016 年中国众筹行业经历了启蒙、起步、成长、量变、质变和规范的过程,且融资规模保持持续增长,互联网巨头在众筹领域也完成初步布局。但近年来,中国正常运营的众筹平台数量逐年下降,优质项目缺失、筹资人在众筹平台难以筹集到所需资金、众筹平台风险较大等问题成为众筹平台倒闭的主要原因。此外,中国产品众筹平台,如京东众筹、淘宝众筹等产品众筹项目在众筹期间大部分都可完成目标筹集金额,超额融资率能够有效地定义众筹项目的进度与实际筹款效果。通过对产品众筹超额融资率影响因素的研究,可以有效探究影响众筹项目成功的要素,帮助中小企业利用合法途径高效筹集投入生产和扩大生产所需要的资金,帮助众筹平台改善缺乏优质众筹项目的现状,并实现筹资人、投资人和平台三赢的效果。鉴于此,本文从数据挖掘的角度出发,基于京东众筹平台的 1 031 个产品众筹项目,通过 Apriori 算法计算关联规则,探讨众筹项目的各属性与超额融资率之间的关联关系。

1 文献回顾与理论分析

1.1 文献回顾

众筹的概念来源于众包,Lambert 和 Schwi-

bacher^[1]对众筹的定义为:在一个开放的互联网平台上,通过货币捐赠,或者通过换取未来产品、服务或股权的形式来为特定项目或企业提供金融支持的融资手段。可以看出,众筹包括两个关键要素,互联网平台和大众直接融资,对众筹项目的支持可以有偿交换,也可以是无偿捐赠。目前,不同学者对于众筹现状的研究也逐渐集中在细分领域。夏恩君等^[2-3]对国外众筹研究进行综述和展望,在对基于知识图谱的众筹现状和对未来研究方向进行探讨后指出现有研究主要集中于参与动机、影响因素和面临的风险等。魏可鑫^[4]以众筹理论、大众参与理论和网络环境下的用户动机研究为基础,构建起投资人参与产品众筹的行为模型,研究发现预期收益、众筹平台信任、促进因素以及参与意愿对参与行为的影响呈正相关关系。刘畅^[5]通过运用精细加工可能性模型对影响众筹产品众筹项目成功率的因素进行研究,发现中心线索项目信息模糊程度、风险关键词提示、项目动态更新次数、边缘线索发起人社会资本、项目持续时间对项目融资成功率有显著影响。Ahlers 等^[6]从项目投资角度人角度出发,认为项目投资信息传递、沟通、交流与反馈等会

收稿日期:2021-12-10

基金项目:国家自然科学基金面上项目(71572012,7217020877)。

作者简介:王为(1992—),女,山东潍坊人,北京理工大学管理与经济学院,博士研究生,研究方向为技术创新管理;李淑敏(1994—),女,山西运城人,北京理工大学管理与经济学院,硕士研究生,研究方向为技术创新管理;李俊(1995—),男,湖北十堰人,北京理工大学管理与经济学院,博士研究生,研究方向为技术经济;通信作者夏恩君(1962—),男,吉林松原人,北京理工大学管理与经济学院,教授,博士研究生导师,研究方向为技术创新管理、技术经济评价、区域经济发展、产业经济等。

影响项目成功。Mollick^[7]认为与筹资人的距离也会成为投资人影响项目成功的因素。胡金焱和韩坤^[8]探究了信息不对称下产品众筹市场逆向选择和道德风险问题,发现融资完成率高的产品,质量水平更低且逾期率更高。

根据众筹项目成功的方式可以将众筹分为全额众筹和非全额众筹两类。全额众筹是指当众筹项目在众筹期间内达到筹资人设定的目标筹集金额时,该众筹项目被认定为成功,否则已筹集资金需退还给投资人,即众筹项目失败;非全额众筹是指无论目标筹集金额是否达到,投资人所投资金会悉数归为筹资人,投资人获得产品或服务相应回报。目前关于众筹成功的影响因素研究多是将筹款率达到 100% 的众筹项目作为研究目标,一般将项目筹款进度或成功率作为解释变量,以是否达到 100% 来定义融资成功或失败,来探究项目类型、规模、筹资额度、社会公众对项目的关注度等因素对融资能否成功的影响。近年来,部分学者也会选择用众筹项目超额融资率作为被解释变量,探究影响众筹项目超额完成的因素。陈波等^[9]以淘宝众筹的 7 091 个成功众筹项目为样本,并对众筹项目超额完成的影响因素进行了分析,得出项目筹资的过程中,超额成功率与支持的人数呈现倒“U”形的关系、支持喜欢人数对超额融资率有正面影响等结论。此外,方兴和王博^[10]利用分位数回归研究了影响众筹发起人超额筹资的因素,结果发现众筹的基本属性、投资门槛、发起人的努力程度以及社交网络都对众筹项目的超额筹资存在显著影响。

然而,当前关于众筹的相关研究很少涉及众筹项目成功融资之后的情况,探究影响项目超额融资的因素对于提升产品众筹项目成功度以及市场欢迎程度具有实践应用价值。本文基于 Apriori 算法研究产品众筹项目超额融资率的影响因素,一方面填充了众筹超额融资率研究领域的部分理论空白,另一方采用数据挖掘的方法,拓展了研究众筹领域的方法和思路。

1.2 理论分析

在产品众筹中,筹资人会设置不同的回报档和公益支持,筹集金额一部分为各档支持人数与该档设置金额成绩之和,另一部分为公益筹集金额。因此,产品众筹的超额融资率可以表示为

$$T = 100\% \times \left(\sum_i^n U_i V_i + W - D \right) / D \quad (1)$$

式中: T 为超额融资率; U_i 为第 i 档的设定金额; V_i

为第 i 档的支持人数; U_i 与 V_i 的乘积为该档的实际筹集金额; W 为公益筹集金额; D 为每个项目的目标筹集金额。由超额融资率的表达式可知,在实际筹资活动中,每档设定金额、每档投资人数,以及公益筹集金额,这 3 个因素影响最终筹集金额,也正向影响超额融资率。每个众筹项目的各档支持金额在发起之后不能更改,且过高的支持金额会影响支持人数,每档设定金额对超额融资率的影响体现为对支持人数的调节作用。另外,目标筹集金额越高超额融资率越低。

此外,从筹资人角度来看,项目筹资人历史发起项目数作为项目筹资人的背书,能够为该众筹项目提供一定的信息保障。一般认为,项目筹资人发起的历史项目越多(该平台项目都会成功),获得的点赞数量越多,越容易赢得投资人的信任,投资人更有可能产生投资行为。同时,项目筹资人也可以支持其他项目,这在一定程度上会增加投资者对该筹资人的信任。

从投资人角度来看,项目关注人数、项目赞人数、项目进展(筹资人关于项目的后续声明),以及话题互动数、项目评论数等都会成为影响投资决策和投资人数的因素。王月升等^[11]认为,项目的关注数体现了项目潜在投资者的数量,当浏览项目的人选择关注项目时,代表对项目感兴趣且有可能产生投资行为。根据搭便车理论,投资者会选择得到多数人赞许的项目,跟随受欢迎的发起人和项目,付出更少的成本获得更大的收益。支持者与潜在支持者可对项目发表评论表达对项目的热情或者不满,这能在很大程度上影响其他潜在投资者的投资决策。Huang 和 Chen^[12]通过对用户在网上进行产品选择的研究认为项目的评价数量会对下一个项目的支持人数与前期支持人数之间的正向关系起到调节作用。Gerber 等^[13]认为投资者在追求物质回报的同时,会通过众筹平台与投资人进行互动在众筹过程中获取到参与感,而社交互动会影响众筹的成功率。

从项目信息来看,项目声明是在众筹页面筹资人发表的关于该众筹项目的声明,当筹资人发表声明时,会补充众筹的产品说明、提供优惠信息等,这在一定程度上可以向投资人传递更多的信息,从而促进投资人的投资行为,且声明通过影响支持人数进而影响筹款率。众筹项目的行业属性也会影响投资人的选择。比如陈波等^[9]发现科技类项目具有明显的融资劣势,设计类、漫画类项目具有较为明显的融

资优势。Schulz 等^[14]基于双边市场和金融中介理论可以将众筹中介分为以盈利为目的、快乐主义和利他主义 3 类,投资人会发生基于不同目的的投资行为,如喜欢科技类投资人可能倾向于投资科技众筹项目,公益项目会获得更多公益支持等。项目的类别与项目筹款率之间存在关系,但不同平台定位以及产品众筹类别差异,无法推断具体相关关系。

总体来看,项目筹资人的信用信息,众筹项目自身的公开信息,投资人对该项目的支持、评论等是影响产品众筹超额融资率的潜在因素。

2 研究方法

2.1 关联规则

关联规则(association rule)可用来挖掘数据的隐含特征,最早是由 Agrawal 等^[15]提出,它反映一个事物与其他事物之间的相互依存性和关联性。作为数据挖掘领域的重要方法,关联规则运用广泛,最常见和最经典的应用场景是购物篮分析,它可以用于分析顾客的购物习惯,并为制订营销策略和商品促销提供有价值的参考。关联规则可以发现大量数据中项集之间有趣的关联或相关关系,如何从众多的变量中快捷地选出关联性最强的两组或更多组变量是关联规则算法的核心问题^[16]。

关联分析首先要找出频繁项集,即经常一起发生的事物的集合,然后再从频繁项集中发现强关联规则。评估频繁项集的标准有 3 个,分别为支持度、置信度和提升度,它们的定义如下^[17]:

1)支持度(support)。指关联事件在项集 X 中出现的次数占总事物集 S 的比重,类似于频率,即 $\text{support}(X) = \text{count}(X)/S$ 。若支持度 $\text{support}(X)$ 大于或等于给定的最小支持度 $\text{support}(\min)$,则 X 称为频繁项集。

2)置信度(confidence)。一个事件 X 出现后,另一个事件 Y 出现的概率,相当于事件的条件概率。对于一条关联规则 $X \rightarrow Y$,置信度 $\text{confidence}(X \rightarrow Y) = \text{support}(XY) / \text{support}(X)$ 。

3)提升度(Lift)。指在表示含有事件 X 的条件下,同时含有事件 Y 的概率,与 Y 总体发生的概率之比。公式表示为 $\text{Lift}(X \rightarrow Y) = \text{support}(Y|X) / \text{support}(Y)$ 。若 $\text{Lift}(X \rightarrow Y) > 1$,则规则“ $X \rightarrow Y$ ”被认为是有效的强关联规则;若 $\text{Lift}(X \rightarrow Y) < 1$,则规则“ $X \rightarrow Y$ ”被认为是无效的强关联规则;当 $\text{Lift}(X \rightarrow Y) = 1$,则 X 与 Y 相互独立。

2.2 Apriori 算法

作为挖掘频繁项集的经典数据挖掘算法,Apri-

ori 算法是通过使用频繁项集的先验性质来压缩搜索空间。依据频繁项集的所有非空子集也一定是频繁的这一特征,使用一种逐层搜索的迭代方法,由 k 项集探索 $(k+1)$ 项集。每次通过支持度和置信度阈值进行筛选,直到不能再找到频繁 k 项集为止。每找出一个 L_k 就需要一次数据库的完整扫描。

Apriori 算法的步骤核心可以总结为扫描事务并根据支持度获得频繁项集^[18],依次产生频繁项集 L_1, L_2, L_3 等,最终获得频繁 k 项集,具体操作过如下:

1)算法扫描所有的事务,获得每个项,并且生成 C_1 。然后对每个项进行计数。随后根据最小支持度从 C_1 中删除不满足的项,至此获得频繁 1 项集 L_1 。

2)根据剪枝策略, L_1 的自身连接生成的集合可以产生候选 2 项集的集合 C_2 ,再扫描所有事务,对 C_2 中每个项进行计数。同理,根据最小支持度从 C_2 中删除不满足要求的项,获得频繁 2 项集 L_2 。

3)根据剪枝策略,同第 2)步,由 L_2 自身连接生成的集合可以产生候选 3 项集的集合 C_3 ,并且扫描所有的事务,同时对 C_3 的每个项进行计数。同理,根据最小支持度从 C_3 中删除不满足项,获得频繁 3 项集 L_3 。

以此类推,通过剪枝策略, $L_{(k-1)}$ 的自身连接生成的集合可以产生候选 k 项集 C_k ,再扫描所有事务,并且对 C_k 中的每个项进行计数。根据提前设定好的最小支持度再从 C_k 中删除不满足的项,以此来获得频繁 k 项集。Apriori 算法的具体实现过程如图 1 所示。

3 研究设计

3.1 属性构造

属性是指一个数据字段,也表示数据对象的一个特征,通常可以和特征、变量、维度等词互换使用,例如描述顾客对象的属性可以包含顾客 ID、姓名、地址等,给定属性的观测值称为观测,如果用来描述一个对象的一组属性就被称为属性向量,或者特征向量^[19]。良好的数据属性构造能够提高数据挖掘的精度,对业务问题有用的信息才能更好地解释业务问题。在关于项目影响的融资影响因素实证研究中,学者通过不同的研究和建模方法,得出对众筹成功率影响的各种因素。钟超^[20]的研究结果表明,支持者人数、评论数量、“喜欢”的数量、社交网络分享数量对项目成功率呈现正向影响。

基于文献回顾与理论分析,本文属性的构造从

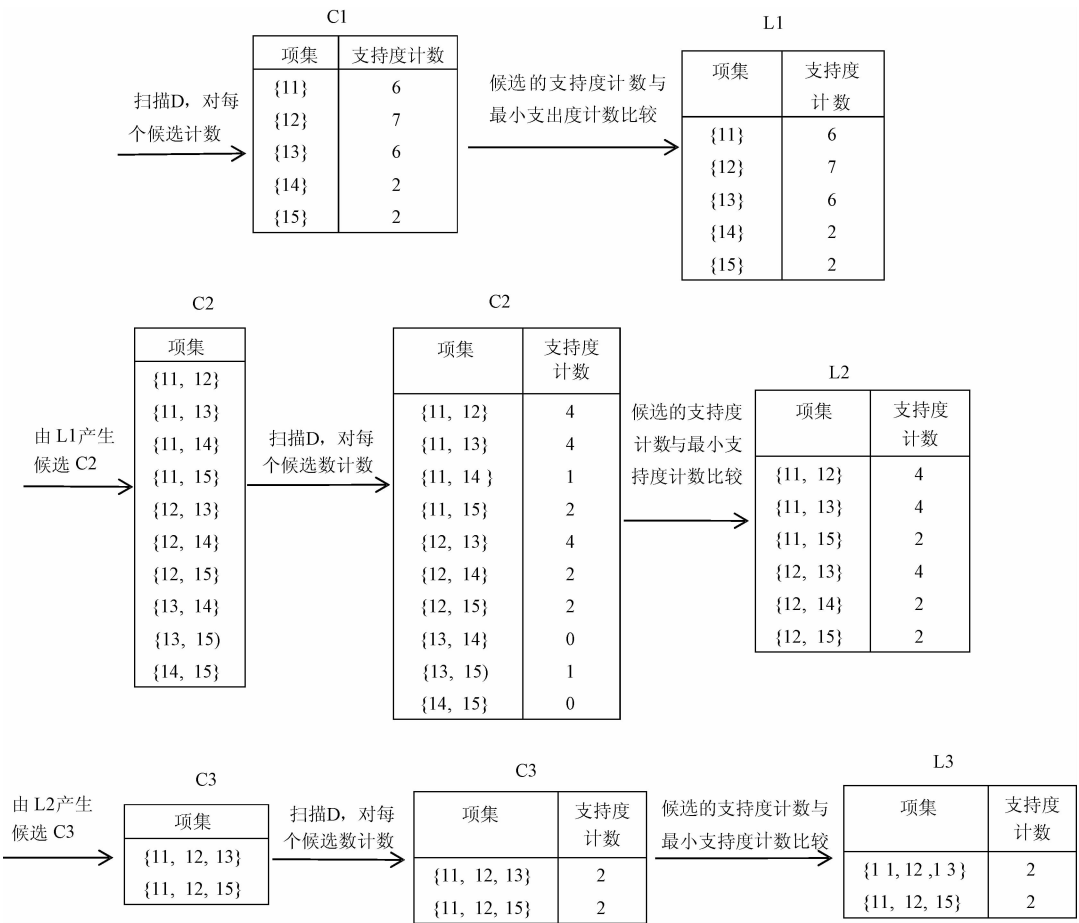


图 1 Apriori 算法过程示意图

项目筹资人、众筹项目、投资人 3 个维度进行。京东众筹平台一般有 4 档产品回报设置，每档支持人数又产生 4 个属性，再加上公益支持人数。一共包括 17 个属性，见表 1。

表 1 属性构造

维度	说明	属性	编号	含义
项目筹资人	项目筹资人信用信息	历史发起项目数	A	项目筹资人“发起”字段值
		是否投资其他项目	B	项目筹资人“支持”字段值
众筹项目	众筹项目自身的公开信息	所属行业	C	筹款项目的类别，共划分 15 个行业类别，根据标签字段划分
		目标筹集金额	D	该众筹项目的预期筹款金额
		项目声明数	E	包括不局限于项目进度说明、发货时间说明等，在网页表现为“项目进展”字段
		第一档设定金额	F	筹资人设置的第一档产品的金额，投资人选择支持该档则需投资该档的金额
		第二档设定金额	G	同上，筹资人设置的第二档产品的金额
		第三档设定金额	H	同上，筹资人设置的第三档产品的金额
		第四档设定金额	I	同上，筹资人设置的第四档产品的金额
投资人	投资人对该项目的支持、评论等	关注人数	J	对该项目进行关注的用户数，对应网页“关注”字段
		点赞人数	K	对该项目表示喜欢的用户数，对应网页“赞”字段
		话题互动数	L	话题互动数为用户提问以及筹资人回应的具体数值，表现为网页的“话题”字段
		第一档支持人数	M	筹资人设置的第一档投资金额下的支持用户数
		第二档支持人数	N	筹资人设置的第二档投资金额下的支持用户数
		第三档支持人数	O	筹资人设置的第三档投资金额下的支持用户数
		第四档支持人数	P	筹资人设置的第四档投资金额下的支持用户数
		公益支持人数	Q	公益支持即不需要实物回报的用户数

3.2 数据来源

由于本文探究的是产品众筹项目属性与超额融资率的关系,因此研究对象须是在规定的期限内达到目标筹集金额的产品众筹项目。本文选择京东众筹平台为数据采取平台,选择上线并已结束了的产品众筹项目,通过网络爬虫技术(八爪鱼采集器)收集项目基本信息,数据采集时间为 2018 年 12 月 21 日。最终爬取到 1 600 个产品众筹项目的基本信息,提取每个众筹项目的筹集金额、超标金额、当前进度、总支持人数、截止日期、目标筹集金额、剩余时间、关注人数、点赞人数、项目声明、话题互动数、公益支持人数、所属行业、筹资人、筹资人历史发起项目数、筹资人支持人数、第一档设定金额、第一档支持人数、第一档回报产品数、第一档回报产品剩余人数、回报时间、第二档设定金额、第二档回报人数、第二档回报产品数、第二档回报产品剩余数、第二档回报周期、第三档设定金额、第三档支持人数、第三档汇报产品数等 36 个字段的数值。

3.3 描述性统计

在爬取到的源数据中,存在较多字段缺失数据,因此需要对源数据进行进一步的清洗和处理,按以下两个步骤进行:①对于多字段缺失数据采用删除法,并对构建属性不必要字段直接删除。②对文本类字段进行处理,并将所属行业划分为一级行业和二级行业两个层级,关联行业标签字段。

在处理完缺失数据后,一共得到 1 031 条样本数据,各属性的描述性统计见表 2。其中,超额融资率平均值为 356%,说明项目平均筹款度远超过 100%,但中位数为 60%,表明过半数项目的超额融资率超过了 50%,中位数远小于平均值,表明超过半数项目的超额融资率都远小于平均超额融资率,超额融资率的最大值可达 25 008%,最小值为 0%。关注人数和点赞人数的数据特征与超额融资率较为相似,平均值远大于中位数,数据标准差较大。另外,选取样本的总目标筹集金额约为 1 亿元。

表 2 各属性的描述性统计

属性	平均值	中位数	众数	标准差	最小值	最大值	峰度	偏度
超额融资率	356	60	1	1 517	0	25 008	152	11
发起项目数	6	1	1	13	1	100	29	5
是否投资其他项目	0	0	0	0	0	1	0	1
目标筹集金额	102 097	50 000	50 000	111 669	1 000	1 000 000	29	4
项目声明数	7	4	0	7	0	44	2	2
第一档设定金额	769	199	99	4 060	5	70 000	217	14
第二档设定金额	779	299	299	2 197	1	45 996	209	12
第三档设定金额	1 910	528	399	5 425	1	99 999	128	9
第四档设定金额	6 194	1 188	10 000	16 781	15	259 800	79	7
关注人数	677	348	1 000	1 028	0	10 000	35	5
点赞人数	514	64	1 000	1 199	0	20 000	88	8
话题互动数	92	46	19	164	0	1 863	48	6
第一档支持人数	207	49	0	726	0	16 720	272	14
第二档支持人数	158	26	0	1 010	0	24 479	412	19
第三档支持人数	76	14	0	285	0	6 279	239	13
第四档支持人数	59	7	0	321	0	7 637	325	16
公益支持人数	35	6	0	114	0	1 606	83	8

3.4 数据转换

由于不同属性之间数据的差异较大,往往不在同一个数量级,因此需要对数据进行规范化处理。数据规范化是需要按照一定的比例将原数据缩放至一个区间,本文采用零-均值规范化法^[9]。零-均值规范化法一方面是其可以在未知最值的情况下进行数据规范化,弥补了当未知最值的情况下无法使用最大-最小规范化的缺陷,另一方面根据均值和标准差进行规范化的数据消除了强影响点对数据

分析的影响。

然后根据数据分布特征将离散区间划分为 2~15 个不等的数量,然后利用 Python 对连续型数据源进行离散化处理。最后对所有属性的数据区间进行编码,每个属性的区间编码对应一个标准化区间和原始数据区间。如表 3 所示,目标属性超额融资率共有 15 个区间编码,其中 T1 的标准化区间为 [-0.23, -0.21],原始数据区间 [0, 45],即对应超额融资率为 0%~45%的项目^①。

注:①鉴于篇幅,其他属性的离散化分类表未罗列,感兴趣的读者可向作者联系索取。

表 3 目标属性离散化及分类

属性	区间编码	标准化区间	原数据区间	范围元素数量/个
超额融资率	T1	$[-0.23,-0.21]$	$[0,45]$	453
	T2	$[-0.2,-0.15]$	$[46,136]$	243
	T3	$[-0.14,-0.04]$	$[137,299]$	129
	T4	$[-0.03,0.1]$	$[304,512]$	87
	T5	$[0.11,0.28]$	$[526,780]$	28
	T6	$[0.3,0.54]$	$[807,1171]$	44
	T7	$[0.65,0.91]$	$[1337,1737]$	8
	T8	$[1.01,1.15]$	$[1891,2098]$	15
	T9	$[1.34,1.58]$	$[2395,2744]$	5
	T10	$[1.91,2.34]$	$[3245,3908]$	4
	T11	$[3,3.47]$	$[4906,5611]$	8
	T12	$[6.3,6.38]$	$[9902,10021]$	3
	T13	$[9.63,9.63]$	$[14955]$	1
	T14	$[13.59,14.16]$	$[20961,21826]$	2
	T15	$[16.26,16.26]$	$[25008]$	1

3.5 建模流程

采用 Apriori 关联规则算法对转换后的数据进行分析,需要进行重复迭代。建模流程如图 2 所示,输入部分包括包含转换后的数据和建立众筹项目筹款率关联规则模型的建模参数,输出部分则是通过处理过程从输入的众筹属性数据转化而成的隐含的关于各属性与目标属性之间的未知关联规则。建模参数最终设定为,支持度 support=0.4,置信度 confidence=0.8,通过搜索与重复迭代,发现与挖掘目标一致的有趣规则。

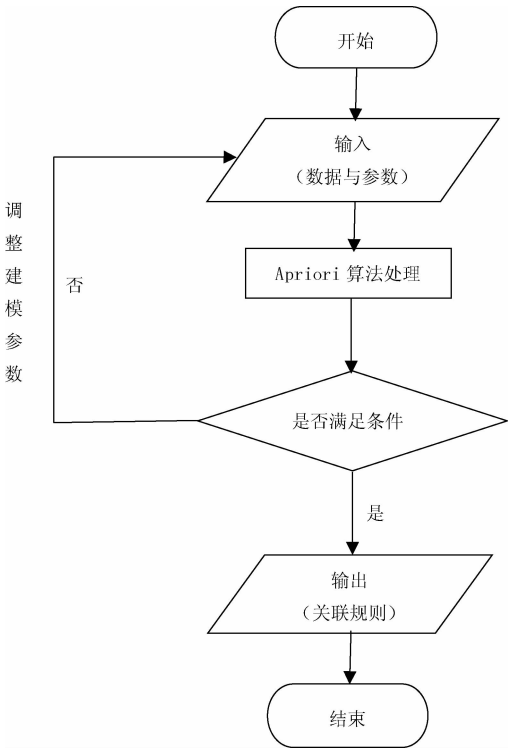


图 2 建模流程

4 研究结果与分析

在得到的众多关联规则中,需要挖掘出有效且有用的关联规则。在支持度 support=0.4,置信度 confidence=0.8 的条件下,共得到 99 条关联规则,这些关联规则中无趣的关联规则,因此需要筛选有趣的关联规则进行分析。为了方便分析结论,分别从三元、四元和七元关联规则来结果进行分析。

4.1 三元关联规则结果分析

在支持度 support=0.4,置信度 confidence=0.8 的条件下,共得到三元关联规则 8 条,见表 4。整体而言,提升度都大于 1,表明每条关联规则的属性之间都具有相关性。具体来看:

1)对于编号 1 关联规则而言,项目筹资人未投资其他项目、目标筹集金额在 10 万、话题互动数在 33 条以下时,超额融资率偏低。

2)对于编号 2 关联规则而言,项目声明数在 0~2 条、以 10 万为目标筹集金额且话题互动在 33 条以下时,众筹项目的超额融资率偏低。

3)对于编号 3 关联规则而言,话题互动在 33 条以下、以 10 万为目标筹集金额且项目关注人数在 0~207 时,众筹项目的超额融资率偏低。

4)对于编号 4 关联规则而言,第 4 档支持人数在 0~12、以 10 万为目标筹集金额且项目关注人数在 0~207 时,众筹项目的超额融资率偏低。这也符合基本的常识认知,因为一般众筹项目第三档、第四档的投资额度较大,在低档资金相差不大的情形下,若高投资额度的支持人数相比于其他众筹项目较少,那该项目实际筹集资金较少,在目标筹集金额相同且不变时,众筹项目的超额融资率就偏低。

5)对于编号 5 关联规则而言,公益支持人数在 0~8、以 10 万为目标筹集金额且项目关注人数在 0~207 时,众筹项目的超额融资率偏低。由于公益支持不限定投资资金,属于项目的纯收益(无须回报商品或服务),因此适当提高公益支持人数对于提升超额融资率也有着重要的作用。

其余关联规则与上文分析思路相似,综合上述结论,由三元关联规则分析可知:

1)目标筹集金额设定为 10 万的众筹项目,具备项目关注人数偏低、项目声明偏低、话题互动数偏低、筹资人未投资过其他项目时,该项目的实际超额融资率即偏低。在这种情形下,筹资人可以适当地改变筹资目标筹集金额或在不改变目标筹集金额的情况下,通过项目声明(补充众筹产品/服务信息)、与用户互动(使用户获得更多信息)、投资其他

项目作为背书等行为提高超额融资率。

2)目标筹集金额设定为 10 万的众筹项目,在第三档、第四档或者公益支持人数较少的情形下一定会获得较低的超额融资率,这就意味着在不改变目标筹集金额的情形下,投资人需要通过提高或改善第三档或第四档产品回报或者体现产品公益性等方式提高众筹项目的超额融资率。

4.2 四元关联规则结果分析

在支持度 $\text{support}=0.4$,置信度 $\text{confidence}=0.8$ 的条件下,得到四元关联规则 27 条,见表 5。可以发现,在三元关联规则的基础上,挖掘出的四元关联规则增加了项目点赞人数、筹资人历史发起项目数等属性,并且挖掘出行业等属性特征,对比三元属性可知:

1)将编号为 9 的关联规则与编号为 4 和 7 的关联规则对比分析可知,四元关联规则在三元关联规则的基础上进行组合,即目标筹集金额在 10 万、项目关注人数在 0~207、话题互动数在 33 条以下且第四档支持人数较少的众筹项目会获得较低的超额融资率;编号 9~12 的四元关联规则是将三元关联规则中的部分属性进行组合,与三元关联规则结论类似。

2)编号 13~19 的关联规则在三元关联规则的基础上,挖掘出其他档(如第一档、第二档)支持人数属性与低超额融资率之间的关系,以编号 13 的关联规则为例,在目标筹集金额为 10 万、话题互动数在 33 条以下、项目关注人数在 0~207 且第二档支持人数在 0~42 时,众筹项目的超额融资率偏低。综合三元和四元关联规则可知,在目标筹集金额设置偏高时,每档支持人数较低(1、2、3、4)、话题互动、关注、声明等偏低的情况下,不易获得高的超额融资率。

3)由 20 和 21 条关联规则可知,筹资人历史发起项目数也会成为影响低超额融资率的因素,筹资人历史发起项目数偏低(1~4 个)的众筹项目,在未投资其他项目、目标筹集金额较高且话题互动数较低的情况下,超额融资率偏低。

4)同 3),由 22~26 条关联规则可知,项目点赞人数是影响低超额融资率的因素。

5)由第 27 条关联规则可知,第二档支持人数为在 0~42、筹资人历史发起项目数偏低(1~4 个)、第三档支持人数偏低(1~11 个),且目标筹集金额在 20 万得众筹项目不易获得较高的超额融资率。同理第 28 条也得出相似的规则和结论,即个档支持人数较少、历史发起项目少且目标筹集金额偏高的众筹项目不易获得超额融资率。且伴随着目标筹集金额的

上升,筹资人历史发起项目数这个指标得到关注。

6)第 29~34 条关联规则与三元关联规则类似,得出在目标筹集金额在 5 万~7 万的众筹项目,具备项目关注人数偏低、项目声明偏低、话题互动数偏低、投资人未投资过其他项目、项目点赞人数偏低、各档支持人数较低这些属性中一个或多个特性时,项目超额融资率偏低。

7)第 35 条关联规则可知,在第二档和第四档支持人数偏低、点赞人数偏低时,电子科技类产品不易获得较高的超额融资率,这与目标筹集金额的无必然关系。

综上,四元关联规则可以得到如下结论:

1)目标筹集金额在中档及以上的(5 万~7 万、10 万)众筹项目,具备项目关注人数偏低、项目声明偏低、话题互动数偏低、筹资人未投资过其他项目、项目点赞人数偏低、各档支持人数较低这些属性中一个或多个特性时,项目超额融资率偏低。

2)伴随着目标筹集金额的升高,挖掘出筹资人历史发起项目数这一关键属性,关注人数、项目声明、点赞人数等属性重要性减弱,即目标筹集金额高的众筹项目更应该关注这一指标,可在上市新的大型高成本产品或服务之前通过小的低成本项目来提升历史发起项目数这一属性值,为进一步筹集高超额融资率的产品做准备。

3)在支持人数偏低、点赞人数偏低的情形下,电子科技类产品不易获得较高的超额融资率,这与目标筹集金额的设定没有必然的关联关系。

4.3 七元关联规则结果分析

五元、六元关联规则的分析与上文类似,挖掘出项数最多的关联规则为 7 项,即七元关联规则。在支持度 $\text{support}=0.4$,置信度 $\text{confidence}=0.8$ 的条件下,得到七元关联规则 4 条,见表 6。这 4 条七元关联规则,每条关联规则都包含目标筹集金额为 5 万~7 万,项目关注人数为 0~207,第二档支持人数为 0~42,项目点赞人数为 0~207 这 4 个属性,但是不包含每档金额,说明每档产品的设定金额与低融资率之间没有必然的关联关系,但是每档的支持人数与融资率之间存在关联关系。

由以上关联规则分析可得,在关注人数、点赞人数、话题互动数偏低且未投资过其他项目的情况下,目标筹集金额在中档及以上的(主要在 5 万~7 万),各档(主要为第二档和第四档)支持人数和公益支持人数较少,获得的超额融资率就会偏低,而各档金额的设定与超额融资率之间不存在关联关系。

表 4 三元关联规则

编号	关联规则	支持度	置信度	提升度
1	{是否投资其他项目=否,目标筹集金额=[100 000],话题互动数=[0, 33]}=>{超额融资率=[0, 45]}	0.05	0.82	1.87
2	{项目声明数=[0,2],目标筹集金额=[100 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.04	0.83	1.88
3	{话题互动数=[0,33],关注人数=[0, 207],目标筹集金额=[100 000]}=>{超额融资率=[0, 45]}	0.05	0.88	2.00
4	{第四档支持人数=[0,12],关注人数=[0, 207],目标筹集金额=[100 000]}=>{超额融资率=[0, 45]}	0.06	0.82	1.86
5	{公益支持人数=[0,8],目标筹集金额=[100 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.05	0.82	1.87
6	{第三档支持人数=[0,11],目标筹集金额=[100 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.04	0.82	1.87
7	{第四档支持人数=[0,12],目标筹集金额=[100 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.05	0.83	1.89
8	{公益支持人数=[0,8],目标筹集金额=[100 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.06	0.83	1.88

表 5 四元关联规则

编号	关联规则	支持度	置信度	提升度
9	{关注人数=[0, 207],话题互动数=[0,33],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.04	0.90	2.04
10	{话题互动数=[0, 33],公益支持人数=[0,8],目标筹集金额=[100 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.04	0.93	2.13
11	{关注人数=[0, 207],公益支持人数=[0,8],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.88	2.01
12	{话题互动数=[0,33],公益支持人数=[0,8],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.86	1.97
13	{话题互动数=[0,33],第二档支持人数=[0,42],目标筹集金额=[100 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.04	0.86	1.95
14	{关注人数=[0, 207],第二档支持人数=[0,42],目标筹集金额=[100 000],第一档支持人数=[0,36]}=>{超额融资率=[0, 45]}	0.04	0.81	1.84
15	{关注人数=[0, 207],第一档支持人数=[0,36],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.04	0.83	1.89
16	{关注人数=[0, 207],第二档支持人数=[0,42],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.82	1.87
17	{关注人数=[0, 207],第二档支持人数=[0,42],公益支持人数=[0,8],目标筹集金额=[100 000]}=>{超额融资率=[0, 45]}	0.05	0.84	1.91
18	{话题互动数=[0,33],第二档支持人数=[0,42],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.04	0.82	1.87
19	{话题互动数=[0,33],第二档支持人数=[0,42],公益支持人数=[0,8],目标筹集金额=[100 000]}=>{超额融资率=[0, 45]}	0.05	0.82	1.86
20	{是否投资其他项目=否,话题互动数=[0,33],目标筹集金额=[100 000],筹资人历史发起项目数=[1,4]}=>{超额融资率=[0, 45]}	0.05	0.82	1.88
21	{话题互动数=[0,33],公益支持人数=[0,8],目标筹集金额=[100 000],筹资人历史发起项目数=[1,4]}=>{超额融资率=[0, 45]}	0.04	0.82	1.87
22	{话题互动数=[0,33],点赞人数=[0,207],目标筹集金额=[100 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.05	0.87	1.98
23	{关注人数=[0, 207],点赞人数=[0,207],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.81	1.84
24	{公益支持人数=[0,8],点赞人数=[0,207],目标筹集金额=[100 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.05	0.82	1.86
25	{话题互动数=[0,33],点赞人数=[0,207],目标筹集金额=[100 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.04	0.86	1.96
26	{公益支持人数=[0,8],点赞人数=[0,207],目标筹集金额=[100 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.04	0.85	1.93
27	{第二档支持人数=[0,42],筹资人历史发起项目数=[1,4],第三档支持人数=[0,11],目标筹集金额=[200 000]}=>{超额融资率=[0, 45]}	0.04	0.81	1.85
28	{第四档支持人数=[0,12],第二档支持人数=[0,42],筹资人历史发起项目数=[1,4],目标筹集金额=[200 000]}=>{超额融资率=[0, 45]}	0.04	0.82	1.87

续表 5

编号	关联规则	支持度	置信度	提升度
29	{话题互动数=[0,33],公益支持人数=[0,8],目标筹集金额=[50 000,70 000],关注人数=[0, 207]}=>{超额融资率=[0, 45]}	0.05	0.85	1.93
30	{关注人数=[0, 207],第一档支持人数=[0,36],目标筹集金额=[50 000,70 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.81	1.84
31	{关注人数=[0, 207],公益支持人数=[0,8],目标筹集金额=[50 000,70 000],第一档支持人数=[0,36]}=>{超额融资率=[0, 45]}	0.04	0.85	1.93
32	{关注人数=[0, 207],公益支持人数=[0,8],目标筹集金额=[50 000,70 000],第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.05	0.80	1.83
33	{公益支持人数=[0,8],点赞人数=[0,207],目标筹集金额=[50 000,70 000],话题互动数=[0,33]}=>{超额融资率=[0, 45]}	0.05	0.82	1.87
34	{话题互动数=[0,33],公益支持人数=[0,8],目标筹集金额=[50 000,70 000],第一档支持人数=[0,36]}=>{超额融资率=[0, 45]}	0.05	0.84	1.91
35	{第二档支持人数=[0,42],点赞人数=[0,207],行业=电子产品类,第四档支持人数=[0,12]}=>{超额融资率=[0, 45]}	0.04	0.81	1.85

表 6 七元关联规则

编号	关联规则	支持度	置信度	提升度
36	{目标筹集金额=[50 000,70 000],关注人数=[0, 207],第二档支持人数=[0,42],点赞人数=[0,207],是否投资其他项目=否,话题互动数=[0,33],公益支持人数=[0,8]}=>{超额融资率=[0, 45]}	0.04	0.89	2.03
37	{目标筹集金额=[50 000,70 000],关注人数=[0, 207],第二档支持人数=[0,42],点赞人数=[0,207],是否投资其他项目=否,第四档支持人数=[0,12],第一档支持人数=[0,36]}=>{超额融资率=[0, 45]}	0.04	0.81	1.85
38	{目标筹集金额=[50 000,70 000],关注人数=[0, 207],第二档支持人数=[0,42],点赞人数=[0,207],是否投资其他项目=否,第四档支持人数=[0,12],公益支持人数=[0,8]}=>{超额融资率=[0, 45]}	0.04	0.83	1.88
39	{目标筹集金额=[50 000,70 000],关注人数=[0, 207],第二档支持人数=[0,42],点赞人数=[0,207],第一档支持人数=[0,36],第四档支持人数=[0,12],话题互动数=[0,33],}=>{超额融资率=[0, 45]}	0.04	0.82	1.87

5 结论与建议

本文基于 Apriori 关联规则算法,研究影响产品众筹项目超额融资率的因素,通过三元、四元和七元关联规则的实证结果分析,可以得到以下结论:

- 1)目标筹集金额与超额融资率之间存在着紧密的关联关系,目标筹集金额在 5 万~7 万、10 万和 20 万的众筹项目会获得较低的超额融资率,且随着目标筹集金额的提升,对于筹资人历史发起项目数这一指标的关注逐渐提升。
- 2)各档金额的设定与超额融资率没有必然的关联关系,公益支持人数与超额融资率之间存在着一定程度的关联。
- 3)在支持人数偏低、点赞人数偏低的情形下,电子科技类产品不易获得较高的超额融资率,这与目标筹集金额的设定没有必然的关联关系。
- 4)各档支持人数与超额融资率之间存在着密切的关联关系,伴随着目标筹集金额的提高,若想获得较高的超额融资率,需要提升第二、第三、第四档等高投资档的支持人数。

为了提高众筹项目超额融资率,使投资人更多

地了解关于众筹项目的必要信息,项目筹资人应该根据众筹项目的目标筹集金额、关注人数、历史发起项目数等属性规划和运营众筹项目,本文从筹资人角度提出以下 4 点对策建议:

- 1)一般众筹项目的目标筹集金额不宜过高。低融资率目标属性总与中档(5 万~10 万)、高档(20 万及以上)目标筹集金额相关联,高目标筹集金额的众筹项目虽然可以获得项目成功,但是不容易获得较高的超额融资率。
- 2)中档目标筹集金额众筹项目应关注指标。对于目标筹集金额在 10 万以下的众筹项目而言,在目标筹集金额不变的前提下,应该重点关注项目关注人数、项目点赞人数、众筹项目发布声明数量、投资人筹资人活动数量、是否投资过其他项目等指标,通过运营、增加互动等形式增加用户对产品或服务的了解,从而提高支持人数,达到提升超额融资率的目的。
- 3)高档目标筹集金额众筹项目应加强对筹资人历史众筹项目的重视。对于目标筹集金额超过 10 万的众筹项目而言,关注、点赞、评论等属性的影

响作用减小,筹资人以往发起的众筹项目数成为影响众筹项目超额融资率的关键指标。筹资人可以通过团队形式,增加历史众筹项目数量,继而为高目标筹集金额项目背书,这样更容易获得投资人信任。同时,筹资人应增加高档金额的投资人数,妥善设置高档对应的回报产品或服务,从而吸引更多用户选择高档金额投资。

4)众筹项目应关注公益支持人数。公益支持人数与超额融资率之间存在着一定程度的关联,因此在实际运营过程中,众筹筹资人应该适当提升公益支持档的回报,例如发放成本较低的感谢信、署名款纪念品等,可以有效地吸引投资人进行投资,小额但大量的投资金额也能够显著提高实际筹集金额,从而提升众筹项目的超额融资率。

参考文献

- [1] LAMBERT T, SCHWIENBACHER A. An empirical analysis of crowdfunding[EB/OL]. (2010-05-06). <http://ssrn.com/abstract=1578175>.
- [2] 夏恩君,王素娟,王俊鹏. 基于知识图谱的众筹研究现状及发展趋势分析[J]. 科研管理, 2017(6): 1-8.
- [3] 夏恩君,李森,赵轩维. 国外众筹研究综述与展望[J]. 技术经济, 2015, 34(10): 10-16, 125.
- [4] 魏可鑫. 大众参与众筹行为影响因素的实证研究[D]. 杭州: 浙江财经大学, 2013.
- [5] 刘畅. 产品众筹成功率影响因素: 兼论创新外部性问题[D]. 广州: 暨南大学, 2016.
- [6] AHLERS G K C, CUMMING D J, GUENTHER C, et al. Signaling in equity crowdfunding[EB/OL]. [2012-10-14]. <http://ssrn.com/abstract=2161587>.
- [7] MOLLICK E. The dynamics of crowdfunding: An exploratory study[J]. Journal of Business Venturing, 2014, 29(1): 1-16.

- [8] 胡金焱, 韩坤. 网络众筹融资表现与风险: 逆向选择还是道德风险? 基于投资人评价的文本挖掘[J]. 山东大学学报(哲学社会科学版), 2021(4): 152-162.
- [9] 陈波, 鲁斯玮. 众筹超额成功的影响因素: 基于淘宝众筹的实证分析[J]. 经营与管理, 2018(1): 78-82.
- [10] 方兴, 王博. 什么为众筹发起人带来了超额筹资? 基于羊群效应和粉丝经济视角的研究[J]. 经济学(季刊), 2019, 18(3): 1061-1080.
- [11] 王月升, 杜朝运, 陈雪君. 奖励型众筹项目融资影响因素的实证分析[J]. 黑龙江工业学院学报(综合版), 2018(2): 86-91.
- [12] HUANG J, CHEN Y. Herding in online product choice[J]. Psychology & Marketing, 2010, 23(5): 413-428.
- [13] GERBER E, HUI J, KUO P Y. Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms[C]//Computer Supported Cooperative Work, 2012.
- [14] SCHULZ M, HAAS P, SCHULTHEISS K, et al. How idea creativity and hedonic value influence project success in crowdfunding[C]//Wirtschaftsinformatik, 2015: 948-962.
- [15] AGRAWAL R, IMIELINSKI T, SWAMI A. Mining association rules between sets of items in large databases[J]. ACM SIGMOD Record, 1993, 22(2): 207-216.
- [16] 陈治, 吴娟娟. 基于关联规则的医疗数据挖掘研究[J]. 统计与决策, 2020, 36(6): 174-177.
- [17] 邱均平, 崔腾腾, 陈仕吉. 基于聚类 and 关联规则的 Altmetric TOP 榜文献特征分析[J]. 现代情报, 2021, 41(9): 12-21, 63.
- [18] AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules in large databases[C]//Proceedings of the 20th International Conference on Very Large Data Bases. IEEE, 1994.
- [19] 韩家炜, KAMBER M. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2012.
- [20] 钟超. 中国众筹平台的羊群行为研究: 基于众筹网的实证分析[D]. 南京: 南京大学, 2015.

Research on the Influencing Factors of the Excess Financing Rate of Product Crowdfunding Projects Based on Apriori Algorithm

WANG Wei, LI Shumin, LI Jun, XIA Enjun

(School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Based on 1 031 product crowdfunding projects on JD crowdfunding platform, the association rules between the attributes of crowdfunding projects and the excess financing rate are obtained through the Apriori association rule algorithm. Study results showed that: There is a close correlation between the target amount raised and the excess financing rate. There is no necessary correlation between the excess financing rate and the setting of each amount. When the number of supporters is low and the number of “likes” is low, electronic technology products are not easy to obtain a high excess financing rate, which is not necessarily related to the setting of the target fundraising amount. There is a close correlation between the number of supporters and the excess financing rate. The research conclusion is of great significance to improve the excess financing rate of product crowdfunding and the success of crowdfunding projects.

Keywords: product crowdfunding; excess financing rate; association rule; Apriori algorithm