

在线知识社区中用户付费行为的发展 轨迹:基于潜类别增长模型的研究

马 玲 丁 枫 李 嘉

【摘 要】在线知识社区内知识付费产品得以发展的关键因素在于用户付费行为的持续性。但是,用户的付费意愿和付费行为,在与知识产品以及社区交互的过程中会发生改变。因此,了解用户付费行为随时间的动态变化规律至关重要。本研究利用潜类别增长模型对社区内用户的付费行为进行纵向分析,刻画用户付费行为的发展轨迹,并根据发展轨迹的异质性识别社区内存在的用户亚群(谨慎型用户、领先型用户和学习型用户)。研究结果有助于管理者理解付费知识社区中不同亚群用户的角色、行为特征和演化特点,优化管理策略。

【关 键 词】在线知识社区;知识付费;用户亚群;潜类别增长模型

【作者简介】马玲,华东理工大学商学院教授,硕士生导师,博士;丁枫,华东理工大学商学院,硕士;李嘉(通讯作者),华东理工大学商学院教授,硕士生导师,博士(上海 200237)。

【原文出处】《管理评论》(京),2023.3.136~147

【基金项目】国家自然科学基金项目(72271093);教育部人文社会科学研究青年基金项目(18YJC630068)。

引言

互联网时代,越来越多的人参与到知识生产和传播过程中,以 Quora 和知乎为代表的在线知识社区快速增长。近年来,许多在线知识社区开始设置知识付费专区,并推出知识付费产品,如知乎 Live、分答等。知识付费产品通过激发用户的消费意愿,将知识变成产品或服务,实现其商业价值。知识付费机制的兴起,使在线问答平台形成了付费专区与免费专区并存的局面。然而,由于知识免费是根深蒂固的用户习惯,用户针对知识付费的意愿仍普遍较低^[1]。因此,理解在线知识社区中用户付费的规律,具有重要的现实意义。

与视频观看、课程学习等在线付费内容类似,在线知识社区内容付费得以发展的决定性因素在于用户在社区内存在持续付费的行为^[2]。这是因为,用户的付费意愿和付费行为在与知识产品以及社区交互的过程中会发生改变。例如,用户在付费后感知到的内容质量、学习回报和服务质量等因素,会改变后续的付费意愿。因此,理解社区内用户付费行为的发展变化情况对社区管理者至关重要。现有文献多为用户付费行为的静态分析(如初始付费意愿和行

为)^[1,3,4],将付费行为看作一成不变的规律,忽略了对用户付费行为发展轨迹的研究。同时,不同亚群用户在知识需求、尝新意愿和支付能力等内在特征上的差异,决定了他们的付费行为发展规律也存在明显差异^[2,5]。因此,对于用户付费行为的分析应该针对不同亚群分别探讨其发展轨迹特征,实现对用户付费行为的动态把握和付费用户的分群管理,从而实现更有效的社区运营管理。

本文以在线知识社区内的知识付费用户为研究对象,针对用户付费行为的动态性进行分析,刻画用户付费行为的发展轨迹,并根据发展轨迹的异质性定义社区内潜在用户亚群,进一步深入探索影响用户付费行为发展轨迹的因素。之前相关研究采用的多数方法不考虑时间的影响,或假设因变量和自变量之间的关系是静态的,这些方法无法捕获变量的动态变化轨迹^[6]。此外,现有的聚类分析等分群方法也主要聚焦于个体间的差异而不考虑个体的变化。潜类别增长模型(latent class growth model, LCGM)通过对纵向数据的分析刻画个体的发展轨迹,并通过识别个体的初始水平和发展速度来定义用户的类别^[6]。LCGM 已在社会学和心理学等领域广泛应用

于疾病发展问题和人口属性变化的研究。因此,本文运用 LCGM 拟合在线知识社区中用户付费行为的发展轨迹,根据发展轨迹的异质性定义用户类别,并分析不同类别用户付费行为的影响因素。

文献综述

1. 知识社区的用户付费行为

用户在知识社区中的付费行为,已经吸引了信息系统领域的大量研究。研究内容主要围绕用户付费行为的影响因素,如付费区知识供给者个人信息和实绩及用户反馈^[2,7,12,13]、知识接受者的个人特征及感知所得和满意度等^[1,3,4,8,15-17]、免费区与付费区知识贡献的相互影响^[5,9-11,14]等。这些研究使用了多样化的理论基础,包括社会资本理论^[7]、感知价值理论^[1]、现状偏差理论^[3,8]、推-拉-锚定模型^[9-11]、计划行为理论^[4]、信号理论^[2,12,13]、社会交互视角^[14]、精细加工可能性模型使用与满足理论^[15]、信任理论^[16]、价值感知差异性理论^[5]、信息捕食理论^[17]等。从研究方法上看,一些研究使用的是二手数据(从知识社区抓取的客观数据)分析方法^[2,5,7,12,16,17],一些研究使用的是问卷调查的方法^[1,3,4,8,15,17]。此外,还有少数研究使用了关键事件技术、熵权法^[9,10]以及机器学习^[14]等方法。

现有关于知识社区用户付费的研究把付费行为看成静态的。大部分研究关注的都是用户的初始付费意愿^[1,3,4]或一段时间内固定的付费规律^[7,8,12]。有的研究虽然关注的是用户的持续行为(如持续参与意愿^[18]、继续付费意向^[15]),但是这些研究本质上只在一个时间点利用问卷进行了测量。因此,这些研究都没有真正关注用户付费行为随时间变化的发展轨迹。

2. 知识社区的用户亚群

信息系统领域关于知识社区的研究,很早就意识到了用户亚群的重要性。这是因为不同亚群的用户,在知识社区中会表现出不同的行为。知识社区的相关研究,也对用户亚群进行了多种维度的分类(见表 1)。这些研究用到的分类标准包括用户的专业水平^[5]、性别^[19]、领导水平^[20]、协作模式^[21]、角色^[22]、人

格特质^[23]等,其研究场景包括知识付费^[5]、知识共享^[19,20]、用户协作^[21]和持续参与^[22,23]等。这些研究表明,用户亚群是知识社区研究场景下的一个重要变量,合理划分用户亚群对深入理解知识社区的用户行为具有重要意义。

然而,现有研究在划分知识社区中的用户亚群时,基本上是采用理论驱动的方法,即亚群的分类标准是根据研究场景和理论分析得到的。如何让数据来说话,根据数据分析的结果来对用户亚群进行分类,现有研究未曾考虑过。因此,本文将利用潜类别增长模型,基于实际数据并根据发展轨迹的异质性来定义在线知识社区中的用户类别。

研究假设

用户与知识产品和社区交互的过程,会改变用户对付费知识的感知价值,从而体现为用户在不同阶段的不同付费行为。用户对付费知识的感知价值受到多方面因素的影响:第一类因素是用户特质类因素,如用户的免费观念、用户对选择新事物的谨慎程度、用户对新鲜感保持的程度、用户的知识水平和学习意愿等^[3];第二类因素是知识产品本身的因素,如用户感知到的内容质量、学习回报和在互动过程中感知的服务质量等^[3,12];第三类因素是社区的环境因素,如整个知识社区对知识付费的态度(社会规则)、社区中免费知识的供给等^[8,9,17]。这三类因素都会影响用户对付费知识的感知价值。用户最终的付费行为是这三类因素综合影响的结果。由于用户对知识产品感知价值的影响因素很多,这些因素随时间的发展变化趋势必然是不平衡的。用户对知识产品的感知价值和付费行为,将随时间呈现非线性的变化趋势。因此,本研究假设:

H1:在线知识社区中,用户的知识付费行为随时间呈非线性的变化趋势。

不同的用户特质对应了用户知识付费行为的不同初始水平。首先,大多数用户都对在线内容和服务持有免费心理,并且这种免费心理对用户的知识付费

表 1 知识社区关于用户亚群的相关研究

分类标准	分类结果	场景	来源
专业水平	专业用户、普通用户	知识付费	Zhang 等 ^[5]
性别	男、女	知识共享	Chai 等 ^[19]
领导水平	领导型用户、非领导型用户	知识共享	Jappesen 等 ^[20]
协作模式	领导者、领域专家、词条贡献者、词条维护者、边缘用户	开放知识社区用户协作	陈耀华和杨现民 ^[21]
角色	专业用户、健康达人、普通用户	知识社区持续参与	张薇薇和蒋雪 ^[22]
人格特质	大五人格特质	在线问答社区用户持续使用	杨敏 ^[23]

行为有显著负向影响^[24]。免费心理弱的用户对应的付费行为起始点较高,更容易做出初始付费行为。其次,不同用户对新事物的谨慎程度是不同的。拥有较强创新性特质(innovativeness)的用户,更容易做出初始付费行为^[25]。此外,知识付费用户的其他特征也存在很大的差异,例如参与动机、教育背景、感知价值^[24]等,这些差异也会影响用户的初始付费行为。

同时,用户知识付费行为的发展速度会有所差异。首先,用户特质类因素会导致用户知识付费行为的发展速度有所差异。例如,用户的心理惯性就会影响用户付费行为的发展速度^[24]。心理惯性较弱的用户,其付费发展速度会更加迅速。其次,用户在付费知识产品的使用过程中,可能选择了不同的知识产品和服务,导致用户感知到完全不同的价值(如内容质量、学习回报和服务质量等)。而且即使是同样的知识内容和知识服务,据价值感知差异性理论^[5],不同的用户感知到的价值也是不同的。由于用户对付费知识的感知价值会影响付费行为,有不同心理惯性和体验效果的用户必然有不同的知识付费发展速度。因此,本研究假设:

H2:在线知识社区中,用户知识付费行为的初始水平(a)和发展速度(b)存在差异。

在知识付费的场景下,不同用户在初始付费意愿、感知内容质量、感知学习回报、感知服务质量上有所差异,因而在付费行为的发展轨迹上会有所不同。前人关于知识社区的研究,已经表明用户亚群是一个重要的变量^[15,19,20],同一亚群内的用户有类似的行为规律。在知识付费的场景下,也会有一些用户在初始付费意愿、感知内容质量、感知学习回报、感知服务质量等因素上表现出相似性,从而展现出类似的付费行为发展轨迹。因此,本研究假设:

H3:在线知识社区中,存在付费用户亚群。

价格是用户获取知识而付出的经济成本。前人的研究表明,价格对知识付费行为有重要影响^[16],因为只有当用户对知识产品的感知价值大于感知价格时,他们才会做出付费行为。因此,一般而言,价格与用户的付费行为呈负相关^[26]。但是,感知价格是一个相对概念而不是绝对概念。首先,不同的用户有不同的支付能力。支付能力更强的用户,感知价格会更低。其次,用户会根据已有的消费经验形成平均适应水平,并以此作为估计感知价格的基础。Zhang 等^[5]的研究表明,知识付费用户会将历史价格

与当前价格进行比较。历史价格越高的用户平均适应水平越高,其感知到的当前付费知识产品的价格也就越低。但是,用户知识付费的历史经验不同,导致他们的适应水平存在异质性,所以不同类别的用户对价格有着不同的敏感程度。用户购买了不同价格的知识产品后,其不同的感知价格水平(例如“超值”或“血亏”)会影响用户付费行为的发展趋势。因此,本研究提出假设:

H4a:知识付费产品的价格会影响用户付费行为的发展趋势。

H4b:知识付费产品的价格会影响用户亚群的划分。

知识付费产品的时长是指阅读或收听完该知识所需花费的时间。不同的用户对知识时长的偏好是不一样的。有的用户为节约时间成本,更愿意购买时长更短的知识付费产品^[27]。但是,另一些用户更喜欢购买时长更长的知识付费产品,因为更长的时长意味着蕴含了更多的知识,从而体现出更好的效费比^[28]。用户购买了不同时长的知识产品后,其感知到的使用效果(如知识精炼程度、知识效费比)会影响用户付费行为的发展趋势。因此,本研究提出假设:

H5a:知识付费产品的时长会影响用户付费行为的发展趋势。

H5b:知识付费产品的时长会影响用户亚群的划分。

评论效价是指用户对一个知识付费产品的评分。在线知识社区中,知识付费用户可以对知识付费产品进行评分,反映对付费知识质量的认可程度。前人的文献表明,评论效价是用户判断知识质量的一个重要依据,会对用户的付费行为产生显著影响^[2]。但是,不同的用户对评论效价的感受是不一样的。首先,当用户购买了不同评论效价的产品后,会将主观感受到的质量同评论效价进行对比,并计算其中差距作为期望不证实(expectation disconfirmation)的程度。但是,不同用户感知到质量会有所不同,因而期望不证实的程度也会有所不同。其次,与高知识水平的用户相比,评论效价对低知识水平的用户的影响更大,因为高水平的用户会更多地依据自身的经验和专业能力来判断知识产品的质量^[5]。更加依赖评论效价的低知识水平用户,会更加准确地记忆该产品的评论效价,并更愿意将其与实际感受质量进行对比。因此,即使两个用户有相同的实际感受质量,也会因为知识水平的差异而有不同的期望不证实程度。这种期望不证实的程度会进一步

影响用户的满意度和付费行为的发展趋势^[29]。因此,本研究提出假设:

H6a:知识付费产品的评论效价会影响用户付费行为的发展趋势。

H6b:知识付费产品的评论效价会影响用户亚群的划分。

付费知识平台通常同时提供付费知识(付费区)和免费知识(免费区)。免费区知识的数量和质量,对用户后续付费行为的发展可能有不同的影响。一方面,知识提供者在免费区知识的数量越多,用户感知到付费区知识的相对价值就会下降。这是因为用户在免费区可能也能找到该知识提供者发布的类似知识产品,从而降低用户未来的付费意愿。另一方面,免费区知识的数量越多,用户越容易判断答主发送知识的质量。由于在线知识社区的虚拟性和知识的不确定性,知识提供者无法让用户直接观察或事先评估其提供的知识的质量,只能基于他们发送的信号来推断所提供的知识的质量。例如,医疗在线咨询平台上医生通过发布免费文章或提供免费在线咨询服,向患者展现专业水平和医德水平^[30]。同样,在线知识社区中,用户也会通过答主在免费区的回答,来判断答主的能力和热心程度^[9]。因此,用户通过知识提供者在免费问答区提供的关于自身能力、态度和知识产品质量的信号,来消除付费区知识产品质量的不确定性。而知识产品质量的不确定性是阻碍用户付费的一个重要因素^[31]。消除了知识产品质量的不确定性后,用户的付费意愿会有所上升。综上,知识提供者提供的免费知识的数量,对不同的用户(如有的用户在意付费知识的相对价值、有的用户在意如何更加准确地判断付费知识的质量)会产生不同的影响,从而进一步影响用户付费行为的发展趋势。因此,本研究提出假设:

H7a:知识供给者的免费供给会影响用户付费行为的发展趋势。

H7b:知识供给者的免费供给会影响用户亚群的划分。

研究方法

1. 研究设计

基于上述的探索和分析,本研究需要刻画知识付费用户付费行为的初始水平和发展速度,依据用户付费行为的发展轨迹对用户进行分类,并且检验知识付费产品的时长、价格、评论效价和知识供给者的免费供给这4个协变量对付费行为发展轨迹的影响。

针对假设 H1,本研究使用潜在增长曲线模型(latent growth curve models, LGCM)进行检验。针对假设 H2~H7,使用潜类别增长模型(LCGM)进行分析。LGCM 模型是 LCGM 模型的特殊情况, LGCM 模型假定截距和斜率的方差和协方差均等于 0,即所有个体具有完全相同的初始水平和发展速度,而 LCGM 模型先将个体进行分类,并且不同类别的个体具有不同的初始水平和发展速度。

使用 LCGM 模型的目的是获取用户付费行为的发展轨迹的差异。已有研究证明了当观察到的变量差异是由未观察到的异质性所导致时,使用 LCGM 模型是有效的^[32,33]。人口特征,例如性别、年龄等属性,属于可观察到的异质性,可以在模型中直接使用指标衡量。但是,当存在不可观察的异质性时,则无法通过简单的回归模型进行解释。利用潜类别分析对个体进行分类,形成潜类别,可以解决未观察到的异质性问题。每个潜类别都代表了一个不可观察的子类别,该子类别具有自己独有的发展轨迹^[34]。

2. 样本选择

本研究以知乎网站为背景,因为它是中国领先的在线知识社区。知乎不仅提供免费的知识共享服务,并于 2016 年 5 月 14 日推出了有偿的知识服务——知乎 Live,知乎逐渐成长为免费与付费服务并存的知识社区。截至 2018 年底,知乎的用户总量突破 2.2 亿,其问题数超过 3000 万,回答数超 1.3 亿。同时,知乎的知识付费模式也发展得较为成熟,知乎 Live 是知乎知识付费模式的核心业务。因此,本文以知乎 Live 为研究对象进行实证研究。

在知乎的免费专区中,用户可以发布文章或答案以共享知识,也可以通过提问或关注他人的问题以表达自身的兴趣和所需知识。而在知乎 Live 中,用户可以扮演两个不同的角色:主讲人(即知识提供者)和听众(知识消费者)。知乎 Live 鼓励主讲人通过文本、视频、音频等多种形式发起讲座,听众要收听这些 Live 需支付一定的入场费,并可以对 Live 进行评价。

本研究针对知乎中 881 个最新更新的热门 Live,抓取了每个 Live 最新的 20 条评论的用户 ID,最终获取 12556 名订阅过知乎 Live 的用户作为知识付费用户。知乎 Live 是知乎网站于 2016 年 5 月上线的知识付费产品,并且知乎于 2019 年 3 月推出了“盐选会员”机制,知乎的盐选会员可以免费订阅知乎 Live,此项机制会影响知乎 Live 的销量。本文利

用 Python 爬虫技术,从 2016 年 5 月至 2019 年 3 月,以一个季度为时间间隔,追踪收集付费用户的 Live 订阅行为及 Live 的详细信息,共收集了 11 期数据。剔除掉因违规行为而被注销个人资料的无效用户之后,共保留了 12502 名知识付费用户的数据,涉及 5927 条知乎 Live,知乎 Live 类型包括互联网、生活方式、职业、经济金融、教育等。

3. 研究模型

本文的研究模型如图 1 所示。模型使用截距和斜率来描述用户付费行为的发展轨迹。换句话说,

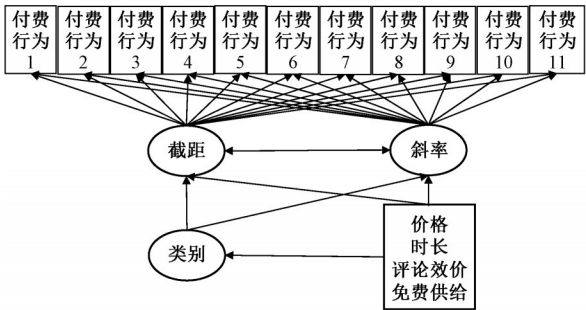


图 1 研究模型

表 2		描述性分析和相关性分析								
变量	均值	方差	最大值	P1	P2	P3	P4	P5		
P1	0.156	1.301	45	1.000						
P2	0.774	3.908	113	0.727**	1.000					
P3	1.221	5.290	144	0.670**	0.939**	1.000				
P4	1.528	5.956	185	0.648**	0.908**	0.983**	1.000			
P5	1.767	6.347	215	0.634**	0.883**	0.966**	0.988**	1.000		
P6	2.026	6.741	221	0.607**	0.852**	0.937**	0.963**	0.982**		
P7	2.380	6.972	227	0.590**	0.832**	0.917**	0.944**	0.965*		
P8	2.704	7.108	227	0.579**	0.816**	0.901**	0.928*	0.949**		
P9	2.950	7.182	227	0.574**	0.807**	0.890**	0.918**	0.938**		
P10	3.050	7.232	227	0.570**	0.802**	0.886**	0.913**	0.933**		
P11	3.296	7.632	227	0.540**	0.759**	0.838**	0.864*	0.884*		
价格	19.979	14.372	598	0.043**	0.052**	0.053**	0.049**	0.045*		
时长	58.260	19.743	218	-0.035**	-0.053**	-0.059**	-0.058*	-0.053**		
评论效价	4.498	0.280	5	-0.079**	-0.136**	-0.137*	-0.135*	-0.129**		
免费供给	7.323	5.901	53	-0.010**	-0.028**	-0.028*	-0.029*	-0.025*		
变量	P6	P7	P8	P9	P10	P11	价格	时长	评论效价	免费供给
P6	1.000									
P7	0.987**	1.000								
P8	0.972**	0.989**	1.000							
P9	0.962**	0.979**	0.993**	1.000						
P10	0.957**	0.974**	0.988**	0.997**	1.000					
P11	0.908**	0.925**	0.939**	0.949**	0.952**	1.000				
价格	0.040*	0.031**	0.023*	0.015*	0.013	0.001*	1.000			
时长	-0.046*	-0.040*	-0.034	-0.029*	-0.028*	-0.024	0.071*	1.000		
评论效价	-0.118*	-0.108	-0.096*	-0.088*	-0.084*	-0.073	0.018	0.337*	1.000	
免费供给	-0.020*	-0.018	-0.007*	-0.004*	-0.002	-0.003	0.058*	0.011*	0.139*	1.000

注:*p<0.05,**p<0.01,下同。

不同的截距和斜率对应了不同的发展轨迹,进而对应不同的用户类别。而用户类别可以进一步通过用户对价格、时长、评论效价、免费供给等因素的偏好来定义。本研究采用了 11 期的面板数据,因此模型中包含了 11 个付费行为变量(P1—P11,每个变量对应一期)。本研究使用 Mplus17.0 软件构建 LGCM 和 LCGM 模型并进行统计检验。

具体来说,付费行为用该期用户购买的付费知识产品的次数来测量。价格用该期用户购买付费知识产品的平均价格来测量。时长用该期用户购买付费知识产品的平均时长来衡量。免费供给用该期用户订阅的知乎 Live 主讲人在社区中免费贡献的答案数来测量。

4. 描述性分析和相关性分析

表 2 提供了各变量的描述性分析和相关性分析结果。由于数据间的量级差距较大,因此本研究对部分计数变量进行了取对数处理,以消除量级差距。

研究结果

1. 潜在增长曲线模型 LGCM 结果

为了验证假设 H1,本研究构建了 LGCM 模型。

具体来说,本研究进行了4种LGCM模型的拟合,分别是线性估计模型LGCM1、二次估计模型LGCM2、三次估计模型LGCM3以及自由估计模型LGCM Free,以确定描述用户付费行为发展轨迹的最佳模型,结果见表3。根据表3显示的拟合指标,LGCM2较为合适,即二次估计模型的拟合结果更好(LGCM2拥有更大的CFI和TFI指标,以及更小的SRMR和RMSEA指标),支持假设H1,即知识付费用户的付费行为为随时间呈非线性变化。

2. 潜类别增长模型 LCGM 结果

为了验证假设H2~H7,本研究构建了LCGM模型。表4显示了LCGM二次估计模型的均值和方差,截距、线性斜率因子和二次斜率因子的方差估计分别为0.076、0.050和0.001,且均在0.001的水平上显著,这表明用户付费行为的初始水平和发展速度方面存在显著的个体间差异,在此类情况下可以对个体进行分类分析。因此,支持假设H2。

本研究根据潜类别分析的一般步骤,对LCGM

表 3		LGCM 模型		
模型	CFI	TFI	SRMR	RMSEA
LGCM1	0.728	0.755	1.073	0.310
LGCM2	0.926	0.932	0.102	0.086
LGCM3	Analysis did not converge after repeated try			
LGCM Free	0.760	0.751	1.150	0.312

表 4		LCGM 的均值和方差			
参数	测量项	系数		标准差	
截距	均值	0.061***		0.003	
	方差	0.076***		0.001	
线性斜率	均值	0.117***		0.002	
	方差	0.050***		0.001	
二次斜率	均值	-0.003***		0.000	
	方差	0.001***		0.000	

注:***p<0.001,下同。

表 5		LCGM 模型拟合结果							
类别	自由度	LL	AIC	BIC	aBIC	entropy	LMR	BLRT	
1	14	-153468.066	306964.131	307068.202	307023.712	—	—	—	
2	18	-104702.970	209441.940	209575.746	209518.544	0.981	0.3332	<0.0001	
3	22	-83168.447	166380.894	166544.435	166474.521	0.969	0.0005	<0.0001	
4	26	-70294.114	140640.228	140833.503	140750.878	0.970	0.0913	<0.0001	
5	30	-60831.984	121723.968	121946.977	121851.640	0.971	0.1305	<0.0001	
6	34	-54227.741	108523.481	108776.225	108668.177	0.986	0.0237	<0.0001	
7	38	-46818.437	93712.874	93995.352	93874.592	0.981	<0.0001	<0.0001	
8	42	-42385.775	84855.550	85167.763	85034.291	0.981	0.1593	<0.0001	
9	46	-38772.306	77636.611	77978.559	77832.376	0.979	0.0042	<0.0001	
10	50	-35842.988	71785.977	72157.659	71998.764	0.981	0.0148	<0.0001	

的模型进行了拓展,依次建立了1~10个类别的LCGM模型,模型拟合结果如表5所示。

对数似然值(LL)、Akaike 信息标准(AIC)、贝叶斯信息标准(BIC)和样本大小调整后的贝叶斯信息标准(aBIC)是评价模型拟合程度的主要指标,这四个指标的值越小,表示模型的拟合程度越高。图2的碎石图显示,aBIC的值在类别3和类别4处存在明显的拐点,因此将用户分成3类或4类较为合适。

熵(entropy)是判断分类质量的另一个指标,用于评估个体是否可以且仅可以被划分到一个类别中^[35]。

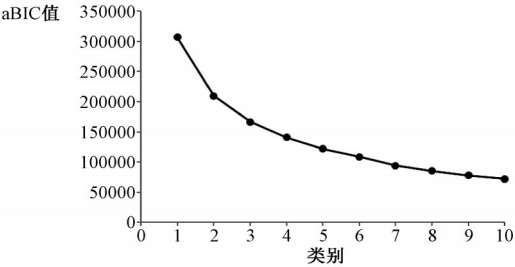


图2 aBIC值的碎石图

一般情况下,熵值越接近 1,代表分类情况越好,熵值高于 0.80 可以被视为有合适的分类。此外,似然比检验(LMR)和基于 Bootstrap 的似然比检验(BLRT)也可作为判断指标,一般而言,LMR 和 BLRT 显著则证明模型的拟合程度较高。LMR 和 BLRT 的值越高,表示将个体分为 K+1 类未必比 K 类好。因此,综合上述评价指标,将 12502 名知识付费用户分为 3 类较为合适。

表 6 提供了知识付费用户被划分到 3 个潜在类别的平均概率。结果显示,用户被正确分到各个类别的概率均超过 0.95,说明分类结果是可靠的。

表 6	用户划分到 3 个潜在类别的平均概率		
类别	Class1	Class2	Class3
Class1	0.989	0.011	0.000
Class2	0.015	0.981	0.004
Class3	0.000	0.010	0.990

表 7 中展示了 3 个潜在类别的参数估计结果,所有类别的截距和斜率的均值均显著,表明各个类别用户的付费行为的发展轨迹存在显著差异。因此,支持假设 H3。由表 7 可知,类别 1 用户占比最大,此类用户在社区中较晚付费,知识付费行为呈正向增长趋势。类别 2 用户占比最小,此类用户最早在社区中付费,属于社区中的领先用户,但后续呈现明显的下降趋势,知识付费行为呈倒 U 型发展轨迹。类别 3 用户的二次项均值为负且显著,此类用户知识付费行为的增长速度呈轻微下降趋势。3 个潜在类别用户的变化轨迹图如图 3 所示。

表 8 提供了知乎 Live 的价格、时长、评论效价以

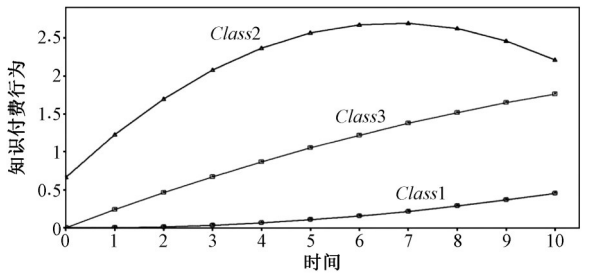


图 3 潜在类别变化轨迹

表 7 含有 3 个潜在类别的 LCGM 模型的参数估计结果

类别	用户数	用户占比	截距	线性斜率	二次斜率
Class1	8214	65.701%	0.005(0.001)***	-0.003(0.001)**	0.005(0.000)***
Class2	1252	10.014%	0.668(0.037)***	0.604(0.009)***	-0.045(0.001)***
Class3	3036	24.284%	0.007(0.005)*	0.243(0.010)***	-0.007(0.001)***

注:*p<0.1,下同。

表 8	协变量对付费行为发展趋势的影响		
变量	参数	β 系数	标准差
价格	截距	0.050***	0.004
	线性斜率	0.011***	0.003
	二次斜率	-0.003***	0.000
时长	截距	-0.015***	0.005
	线性斜率	-0.017***	0.004
	二次斜率	0.002***	0.000
评论效价	截距	-0.116***	0.009
	线性斜率	-0.064***	0.006
	二次斜率	0.009***	0.001
免费供给	截距	0.003***	0.001
	线性斜率	-0.013***	0.001
	二次斜率	0.001***	0.000

及主讲人的免费供给对知识付费行为发展趋势的影响。结果显示,知乎 Live 的价格对知识付费行为的二次斜率有显著的负向影响($\beta=-0.003, p<0.001$),而时长($\beta=0.002, p<0.001$)、评论效价($\beta=0.009, p<0.001$)以及主讲人的免费供给($\beta=0.001, p<0.001$)对知识付费行为的二次斜率均有显著的正向影响。因此,假设 H4a、H5a、H6a 以及 H7a 成立。

为验证协变量对用户付费行为发展轨迹的影响,本研究以 3 个潜在类别的分类结果作为因变量,以订阅的知乎 Live 的价格、时长、评论效价以及主讲人的免费供给作为协变量进行了 Logit 回归,回归结果如下页表 9 所示,模型将类别 3 用户作为参照组。

类别 1 的价格估计系数不显著。类别 2 的价格估计系数正向显著($\beta=1.020, p<0.001$),说明价格每增加 1 个单位,用户被划分为类别 2 的概率是类别 3 的 2.773 倍。类别 1 和类别 2 的时长估计系数均为负且显著,表明时长每增加 1 个单位,用户更有可能属于类别 3。同理,评论效价每增加 1 个单位,用户被划分为类别 1 的概率是类别 3 的 3.201 倍,被划分为类别 2 的概率是类别 3 的 0.295 倍。主讲人的免费供给每增加 1 个单位,用户被划分为类别 1 的概率是类别 3 的 0.873 倍,被划分为类别 2 的概率是类别 3 的 1.349 倍。因此,假设 H4b、H5b、H6b 以及 H7b 成立。

表 9 协变量对潜在类别的 Logit 回归结果

变量	Class1 VS. Class3			Class2 VS. Class3		
	β 系数	标准差	发生比	β 系数	标准差	发生比
价格	0.049	0.050	1.050	1.020***	0.072	2.773
时长	-0.421***	0.064	0.657	-0.220**	0.072	0.803
评论效价	1.163***	0.130	3.201	-1.221***	0.110	0.295
免费供给	-0.136***	0.020	0.873	0.300***	0.024	1.349

稳健性检验

作为一个数据驱动的研究,其结论可能会受到产品异质性的影响。因此,本文考察了在不同内容板块的知乎 Live 上,知识付费行为发展变化轨迹和用户亚群分类的结论是否会发生变化。本文挑选了一些占比较大的主要内容类型(如互联网、生活方式、职业、经济金融、教育等),并在这些具体的类型上做了进一步的分析。

为检验内容类型因素对本研究结果的影响,本研究加入知乎 Live 类型变量。表 10~表 14 是加入了内容类别变量后协变量对付费行为发展趋势的影响。结果显示,在区分了产品内容的不同类型后,价格、时长、评论效价和免费供给等协变量的方向和显著性都没有变化。由于用户付费行为的发展趋势是用截距、线性斜率和二次斜率刻画的,这说明在区分了不同的内容板块后,其发展趋势的规律并没有显著的不同。因此,本文知识付费行为发展变化轨迹和用户亚群分类在不同的内容类型上是稳健的。

结果讨论

根据以上研究,可以对 3 个类别的用户进行命名,并提出分类管理建议。

表 10 加入知乎 Live 类型(互联网)变量后协变量对付费行为发展趋势的影响

变量	参数	β 系数	标准差
价格	截距	0.049***	0.004
	线性斜率	0.013***	0.003
	二次斜率	-0.004***	0.000
时长	截距	-0.013***	0.004
	线性斜率	-0.016***	0.004
	二次斜率	0.002***	0.000
评论效价	截距	-0.119***	0.009
	线性斜率	-0.065***	0.006
	二次斜率	0.009***	0.000
免费供给	截距	0.004***	0.001
	线性斜率	-0.013***	0.001
	二次斜率	0.001***	0.000
类型(互联网)	截距	-0.055***	0.007
	线性斜率	0.036***	0.006
	二次斜率	-0.001*	0.001

表 11 加入知乎 Live 类型(生活方式)变量后协变量对付费行为发展趋势影响的稳健性检验

变量	参数	β 系数	标准差
价格	截距	0.051***	0.004
	线性斜率	0.011***	0.003
	二次斜率	-0.003***	0.000
时长	截距	-0.015***	0.005
	线性斜率	-0.016***	0.004
	二次斜率	0.002***	0.000
评论效价	截距	-0.115***	0.009
	线性斜率	-0.065***	0.006
	二次斜率	0.009***	0.001
免费供给	截距	0.003***	0.001
	线性斜率	-0.013***	0.001
	二次斜率	0.001***	0.000
类型(生活方式)	截距	0.009*	0.005
	线性斜率	-0.019***	0.004
	二次斜率	0.001**	0.000

表 12 加入知乎 Live 类型(职业)变量后协变量对付费行为发展趋势影响的稳健性检验

变量	参数	β 系数	标准差
价格	截距	0.050***	0.004
	线性斜率	0.011***	0.003
	二次斜率	-0.003***	0.000
时长	截距	-0.016***	0.005
	线性斜率	-0.017***	0.004
	二次斜率	0.002***	0.000
评论效价	截距	-0.118***	0.009
	线性斜率	-0.063***	0.006
	二次斜率	0.009***	0.001
免费供给	截距	0.004***	0.001
	线性斜率	-0.013***	0.001
	二次斜率	0.001***	0.000
类型(职业)	截距	-0.031***	0.006
	线性斜率	0.015**	0.006
	二次斜率	-0.001	0.000

(1)类别 1:谨慎型用户。类别 1 用户较晚在社区中付费,但其知识付费行为一直呈增长趋势,且偏好评论效价较高的知识付费产品,因而属于“谨慎型”用户。此类用户在社区中的占比最多。社区管理者应向此类用户推荐高评分的产品,以培养用户忠诚度。

表 13 加入知乎Live类型(金融经济)变量后协变量
对付费行为发展趋势影响的稳健性检验

变量	参数	β 系数	标准差
价格	截距	0.054***	0.004
	线性斜率	0.012***	0.003
	二次斜率	-0.004***	0.000
时长	截距	-0.013**	0.005
	线性斜率	-0.016***	0.004
	二次斜率	0.002***	0.000
评论效价	截距	-0.114***	0.009
	线性斜率	-0.064***	0.006
	二次斜率	0.009***	0.001
免费供给	截距	0.003**	0.001
	线性斜率	-0.013***	0.001
	二次斜率	0.001***	0.000
类型(金融经济)	截距	-0.036***	0.009
	线性斜率	0.015	0.011
	二次斜率	0.000	0.001

表 14 加入知乎Live类型(教育)变量后协变量
对付费行为发展趋势影响的稳健性检验

变量	参数	β 系数	标准差
价格	截距	0.058***	0.005
	线性斜率	0.033***	0.010
	二次斜率	-0.005***	0.000
时长	截距	-0.018**	0.006
	线性斜率	-0.014**	0.004
	二次斜率	0.001***	0.000
评论效价	截距	-0.120***	0.010
	线性斜率	-0.064***	0.006
	二次斜率	0.008***	0.001
免费供给	截距	0.006**	0.001
	线性斜率	-0.005 ⁺	0.003
	二次斜率	0.001 ⁺	0.000
类型(教育)	截距	0.057***	0.007
	线性斜率	-0.008	0.009
	二次斜率	0.001	0.001

(2)类别 2: 领先型用户。类别 2 用户最早在社区中付费,所购买的知识付费产品也较多,因而属于“领先型”用户。这类用户人数最少,但是在整个知识付费的扩散行为中,起到带头示范作用。其他类型的用户(如谨慎型用户和学习型用户),都要看到领先型用户的付费行为后,才会有自己的支付行为。因此,管理者应重点关注此类用户,着重发挥他们的领先价值。同时,这类用户支付能力较强,偏好价格较高的知识付费产品(认为价格是质量的信号),关注知识供给者的能力,但其知识付费行为呈倒 U 型的发展趋势。此类用户喜欢尝鲜但热度不持久,容易

被功能相似的竞品吸引。本文发现,类别 2 用户付费行为拐点出现的时间,正是知乎推出了“知乎大学”产品的时间。这说明类别 2 的用户很可能是被“知乎大学”这种新产品所吸引,从而出现热度下降。因此,社区管理者应注意跟踪类似竞品并保持对竞品的比较优势,以延长此类用户付费行为发展轨迹的拐点出现时间。

(3)类别 3: 学习型用户。与类别 1 和类别 2 用户相比,类别 3 用户支付能力有限,对价格较为敏感,更倾向于购买时长较长的知识付费产品(从而获得更好的知识效费比,即单位花费买到更多的知识)。时长较长代表其蕴含的知识量也较大,因此这类用户属于“学习型”用户。此类用户虽然支付能力较弱,但是人数较多,而且其购买行为有较稳定的增长。社区管理者可向其推荐性价比更高的知识产品,从而促使其在较长一段时间内保持对知识付费的积极参与。

结论与展望

1. 主要结论

本文利用 LCGM 模型刻画了用户知识付费行为的发展轨迹,根据发展轨迹的异质性确定了潜在的用户亚群,并根据分类结果进一步深入探讨了影响用户付费行为发展轨迹的因素。具体来说,本研究得到了以下结论:

首先,在线知识社区内知识付费用户的付费行为的发展轨迹具有异质性,并且根据付费行为发展轨迹的差异可以将知识付费用户划分为 3 类潜在的用户亚群,即谨慎型用户、领先型用户和学习型用户。

其次,知识付费产品的价格、时长、评论效价以及知识供给者的免费供给会影响知识付费行为的发展轨迹,同时会影响用户的类别转化。因此,网站可以通过用户在上述因素上的表现,判断其所属的用户亚群,从而预测其后续的发展轨迹和全生命周期内的客户价值。

再次,领先型用户属于社区管理者需重点关注的用户。不仅是因为这类用户有较强的支付能力,是平台利润的主要来源(很多付费平台中,超过 80% 的利润都来自不到 20% 的头部客户),而且其具有重要的示范带头作用。如果没有领先型用户的付费,其他用户就会继续观望,从而无法带动整个知识付费的繁荣。即使领先型用户比较喜爱追求新鲜事物,容易被类似竞品所吸引并出现倒 U 型的发展趋势,但是处在下降阶段的领先型用户对社区的历史

使命已经基本完成。通过对领先型用户的观察学习,谨慎型用户和学习型用户逐渐培养起了知识付费的习惯,从而在领先型用户逐渐退却后,成长为知识付费社区的利润主体。同时,知识付费社区的管理者,应紧盯类似竞品并保持对竞品的比较优势,以延长此类用户付费行为发展轨迹的拐点出现时间。

最后,谨慎型用户和学习型用户虽然较为谨慎(不愿意为新鲜事物承担风险),支付能力也不如领先型用户强,但是他们人数众多(谨慎型用户占了社区总人数的65%,学习型用户占了社区总人数的24%),而且不太容易被其他竞品所吸引,因而能够在较长一段时间内维持参与行为和付费行为的持续增长。付费知识社区的发展,同时需要积极的用户参与和持续的盈利能力。但是,用户参与和持续盈利的目标,可能必须通过不同的用户亚群来实现。换句话说,有一些亚群的用户(如谨慎型用户和学习型用户)负责持续参与,另一些亚群的用户(如领先型用户)负责平台盈利。

2. 研究贡献

本文具有以下几个研究贡献:

首先,本研究将LCGM模型应用于在线知识社区的环境中,关注了用户知识付费行为随时间推移的动态变化。学者们已在社会学和心理学等领域广泛使用LCGM模型来研究观测变量间的纵向关系,但潜类别增长模型在信息系统领域的应用还不充分。在线知识社区的场景下,还没有看到使用了LCGM模型的相关研究。因此,本研究打破了用户知识付费行为静态分析的框架束缚,不再将付费行为看作一成不变的规律,为用户知识付费行为的研究提供了新思路。

其次,本研究通过一种数据驱动的方法,识别出了付费知识社区中存在的3类关键用户。虽然前人关于知识社区的研究中很早就意识到了用户亚群的重要性,但是这些研究都是采用理论驱动的方法进行用户亚群的划分。在大数据时代,充分利用数据的优势,让数据来说话,能够更好地捕捉到新的管理理论,比纯粹通过理论演绎的方式更容易得到有实践意义的结论。

最后,本文不仅刻画了不同类别用户的发展轨迹,而且识别出了不同类别用户在社区中的作用。例如,领先型用户不仅是付费知识社区中利润的主要来源,而且是整个知识付费生态的示范者。正是

通过观察领先型用户的示范行为,其他类型的用户才会发展出自己的知识付费习惯。同时,谨慎型用户和学习型用户虽然为平台利润贡献不大,但是对应了大量的知识交易。他们有持续增长的潜力,是社区繁荣的中坚力量。

3. 局限与展望

本研究仍然存在一些局限性,并据此提出了部分未来的研究方向。首先,本研究根据付费行为的发展轨迹确定了付费知识社区内3个不同的潜在用户类别,但是没有深入探究这些类别形成的原因。未来的研究可以尝试结合定性分析方法来深入探索形成潜在类别的原因。其次,本研究是在知乎网站中进行的。知乎属于社交性较强的在线知识社区,但是在网络环境中也存在着专业性较强的在线知识社区(例如经管之家),本文的研究结果可能并不适用于这类社区。未来的研究可以考虑加入对其他类型在线知识社区的探索。第三,受限于数据,本文使用了评论价效(用户评分)来综合衡量内容质量和服务质量。内容质量和服务质量作为影响用户持续付费行为的关键要素,在本文中并没有区分开来。未来在其他的数据集上,可以进一步区分内容质量和服务质量的不同影响。最后,本文使用数据驱动的方法展开研究,与传统的基于演绎的理论构建方法相比,捕捉到的管理理论较为细碎。因此,后期要进一步加强数据结果的理论解读,从中发现崭新的管理理论。

参考文献:

- [1]方爱华,陆蒙蒙,刘坤峰.虚拟社区用户知识付费意愿实证研究[J].图书情报工作,2018,62(6):105-115.
- [2]蔡舜,石海荣,傅馨,等.知识付费产品销量影响因素研究:以知乎Live为例[J].管理工程学报,2019,33(3):71-83.
- [3]卢恒,张向先,张莉曼,等.理性与偏差视角下在线问答社区用户知识付费意愿影响因素构型研究[J].图书情报工作,2020,64(19):89-98.
- [4]李钢,卢艳强,滕树元.用户在线知识付费行为研究——基于计划行为理论[J].图书馆学研究,2018,(10):49-60.
- [5]Zhang J., Zhang J., Zhang M. From Free to Paid: Customer Expertise and Customer Satisfaction on Knowledge Payment Platforms[J]. Decision Support Systems, 2019, 127: 113140.
- [6]Zheng Z., Pavlou P. A., Gu B. Latent Growth Modeling for Information Systems: Theoretical Extensions and Practical Applications[J]. Information Systems Research, 2014, 25(3): 547-568.
- [7]赵杨,袁析妮,李露琪,等.基于社会资本理论的问答平

台用户知识付费行为影响因素研究[J]. 图书情报知识, 2018, (4): 15-23.

[8]卢恒, 张向先, 张莉曼. 语音问答社区用户知识付费意愿影响因素研究——基于现状偏差的视角[J]. 情报科学, 2019, 37(6): 119-125.

[9]刘周颖, 赵宇翔, 朱庆华. 付费知识问答平台中提问者转移行为的影响因素分析[J]. 图书情报工作, 2020, 64(12): 75-86.

[10]Liu Z., Zhao Y. C., Chen S., et al. Exploring Askers' Switching from Free to Paid Social Q&A Services: A Perspective on the Push-Pull-Mooring Framework[J]. Information Processing & Management, 2021, 58(1): 102396.

[11]赵宇翔, 刘周颖, 朱庆华. 从免费到付费: 认知锁定对在线问答平台中提问者转移行为的影响研究[J]. 情报学报, 2020, 39(5): 534-546.

[12]Cai S., Luo Q., Fu X., et al. What Drives the Sales of Paid Knowledge Products? A Two-Phase Approach[J]. Information & Management, 2020, 57(5): 103264.

[13]Zhang M., Zhang Y., Zhao L., et al. What Drives Online Course Sales? Signaling Effects of User-Generated Information in the Paid Knowledge Market[J]. Journal of Business Research, 2020, 118(9): 389-397.

[14]邓胜利, 蒋雨婷. 用户交互特征对知识付费行为预测的贡献度研究[J]. 图书情报工作, 2020, 64(8): 93-102.

[15]魏武, 谢兴政. 线上知识付费用户继续付费意向影响因素研究[J]. 数据分析与知识发现, 2020, 4(8): 119-129.

[16]Zhao Y., Zhao Y., Yuan X., et al. How Knowledge Contributor Characteristics and Reputation Affect User Payment Decision in Paid Q&A? An Empirical Analysis from the Perspective of Trust Theory[J]. Electronic Commerce Research and Applications, 2018, 31(9): 1-11.

[17]Shi X., Zheng X., Yang F. Exploring Payment Behavior for Live Courses in Social Q&A Communities: An Information Foraging Perspective[J]. Information Processing & Management, 2020, 57(4): 102241.

[18]张杨琰, 彭子健, 刘齐凯. 问答平台用户付费围观持续参与意愿的影响因素[J]. 图书馆论坛, 2018, 38(6): 86-94.

[19]Chai S., Das S., Rao H. R. Factors Affecting Bloggers' Knowledge Sharing: An Investigation across Gender[J]. Journal of Management Information Systems, 2011, 28(3): 309-342.

[20]Jeppesen L. B., Laursen K. The Role of Lead Users in Knowledge Sharing[J]. Research Policy, 2009, 38(10): 1582-1589.

[21]陈耀华, 杨现民. 开放知识社区用户分类研究——以

中文维基百科为例[J]. 现代教育技术, 2016, 26(6): 47-53.

[22]张薇薇, 蒋雪. 在线健康社区用户持续参与动机的演变机理研究[J]. 管理学报, 2020, 17(8): 1245-1253.

[23]杨敏. 人格特质对在线问答社区用户持续使用意愿的影响研究[D]. 西南大学硕士学位论文, 2015.

[24]胡珍苗, 程岩, 崔华玉. 在线内容用户服务升级意愿研究: 基于增值体验的心理惯性视角[J]. 管理评论, 2016, 28(11): 116-128.

[25]Midgley D. F., Dowling G. R. Innovativeness: The Concept and Its Measurement[J]. Journal of Consumer Research, 1978, 4(4): 229-242.

[26]Gaur V., Fisher M. L. In-Store Experiments to Determine the Impact of Price on Sales[J]. Production and Operations Management, 2005, 14(4): 377-387.

[27]杨东红, 贺红梅, 徐畅. 移动音频有声阅读平台用户知识付费行为研究[J]. 情报科学, 2020, 38(7): 105-111.

[28]赵庆亮, 王培勇, 陈吉. 付费问答社区用户围观行为研究[J]. 数字图书馆论坛, 2019, (11): 12-20.

[29]Ho Y.-C., Wu J., Tan Y. Disconfirmation Effect on Online Rating Behavior: A Structural Model[J]. Information Systems Research, 2017, 28(3): 626-642.

[30]Li J., Tang J., Jiang L., et al. Economic Success of Physicians in the Online Consultation Market: A Signaling Theory Perspective[J]. International Journal of Electronic Commerce, 2019, 23(2): 244-271.

[31]Wu J., Wu Y., Sun J., et al. User Reviews and Uncertainty Assessment: A Two Stage Model of Consumers' Willingness-to-Pay in Online Markets[J]. Decision Support Systems, 2013, 55(1): 175-185.

[32]Morin A. J., Morizot J., Boudrias J.-S., et al. A Multifoci Person-Centered Perspective on Workplace Affective Commitment: A Latent Profile/Factor Mixture Analysis[J]. Organizational Research Methods, 2011, 14(1): 58-90.

[33]Wang M., Chan D. Mixture Latent Markov Modeling: Identifying and Predicting Unobserved Heterogeneity in Longitudinal Qualitative Status Change[J]. Organizational Research Methods, 2011, 14(3): 411-431.

[34]Muthén B., Muthén L. K. Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes[J]. Alcoholism: Clinical and Experimental Research, 2000, 24(6): 882-891.

[35]王孟成, 毕向阳, 叶浩生. 增长混合模型: 分析不同类别个体发展趋势[J]. 社会学研究, 2014, 29(4): 220-241.