

A0301

互联网背景下知识付费发展与付费意愿影
响因素探析——基于 LDA 与结构方程模型的
分析

目录

摘要	V
一、 引言	1
(一) 研究背景 (PEST)	1
(二) 研究目的与意义	4
(三) 文献综述	5
(四) 研究方法与技术路线	7
三、 知识付费的发展影响因素分析	8
(一) 发展现状	8
(二) 数据介绍	9
(三) 模型与回归过程	10
(四) 分析结论	12
四、 基于 LDA 模型的知识付费评论主题挖掘	12
(一) 数据的获取与处理	12
(二) 词频统计和词云图	14
(三) 评论文本的情感分析	15
(四) 文本评论的 LDA 主题分析	16
五、 知识付费市场的消费需求分析	19
(一) 模型介绍与建立	19
1. 结构方程模型介绍	19
2. 模型构建	20
(二) 数据处理与结论分析	21

1. 数据处理与检验	21
2. 模型结果与分析	26
六、 结论与建议	29
(一) 结论	29
(二) 建议	29
参考文献	31
附录一：调查问卷	33
致谢	38

表格和插图清单

图 1 调查技术路线图	8
图 2 部分文本数据展示	13
图 3 词云图	15
图 4 情感分直方图	16
图 5 LDA 模型结构图	17
图 6 文本分词散点图	18
图 7 模型与假设示意图	20
图 8 相关系数矩阵热力图	24
图 9 结构方程模型图	27
表 1 变量说明	10
表 2 回归结果	11
表 3 部分分句结果表	13
表 4 部分分词结果表	13
表 5 评论文本词频统计	14
表 6 网络课程主题词表	18
表 7 潜变量测量与问卷对应表	21
表 8 描述性统计分析表	21
表 9 测量变量的相关矩阵、均值和方差	23
表 10 Cronbach ' s alpha 系数与折半系数	24
表 11 表 KMO 和巴特利球形度检验	25

表 12 验证性因子分析，平均方差提取量和组合信度	25
表 13 结构方程结果表	26

摘要

近年来,随着互联网的普及与数据的市场化运用,在线知识付费行业完成了从零到一的跨越,创造了全新的商业风口,本文旨在探明知识付费行业发展的影响因素及消费者付费意愿的影响因素。

第一部分,探析知识付费行业的发展影响因素。本文在多个官方网站获取了2004年至2019年期间,中国与美国的在线教育行业的市场规模、消费者价格指数(CPI)、居民人均纯收入、劳动人口平均受教育年限、互联网普及率这五个变量的数据,将市场规模作为被解释变量,其余四个变量作为解释变量构造面板数据模型,通过豪斯曼检验、LSDV回归判定采用混合回归模型进行分析,并利用stata软件加以实现,得出CPI、居民受教育平均年限、互联网普及率这三个因素均对知识付费行业的发展具有显著的正向影响效应。

第二部分,探析用户对知识付费产品的评价与需求状况。本文以慕课在线教育平台的一门计算机语言课程为例,采用爬虫技术利用集搜客网页爬取了大量评论数据,并利用python软件进行词频统计、词云图等初步的内容分析,再利用LDA模型深度分析评论的主题,得出用户的真实需求在于有针对性的、灵活的知识产品与服务,且教师的教学方式是一个重要的需求关注点。

第三部分,探析用户付费意愿的影响因素。本文通过发放调查问卷获取消费者反馈数据,采用SPSS软件对这些数据进行初步分析,再利用Amos软件构造结构方程模型来分析包括专业性、有趣性、便捷性、监督性、主观规范在内的产品与服务的体验因素和需求因素对消费者付费行为意愿的影响,得出需求是决定用户是否选择付费的主要因素,专业性、便捷性以及主观规范起到促进作用,且用户对于趣味性的期望对其支付意愿影响较小,监督性的期望则会产生负向影响。

本文利用了多个数据处理软件、数据分析模型揭示了知识付费行业发展的主要影响因素，量化了数据生产力在其中所挥的作用，证实了数据可以为行业深度赋能。同时，通过研究消费者基于切身体验而给出的回馈信息，进一步明晰了市场中用户的真实体验，及其对专业性、便捷性等特性的切实诉求，这表明企业需有针对性地更新优化所提供产品与服务以谋求长足的发展。

关键词：知识付费 付费意愿 面板数据模型 LDA 主题分析 结构方程模型

一、 引言

(一) 研究背景 (PEST)

近年来,在线知识付费行业完成了从零到一的跨越,其发展与进步渗透在我们生活的许多方面。但无疑这一行业仍处于起步与探索的阶段。在对此行业进行分析之前,我们利用 PEST 模型,对其行业背景进行分析。

1. 政治 (politics)

2017 年,国家制定并印发了《文化部“十三五”时期文化产业发展规划》,此次《规划》的出台为我们绘制了经济新常态背景下对“十三五”时期文化产业发展蓝图,立足新阶段,明确新要求。《规划》中指出,在新时代发展国家的文化产业,增强文化产业的竞争力与软实力,提升行业发展的质量和速度,有着十分重要的意义。《规划》也强调了科技创新驱动力量的重要性。提出应该大力推动“文化+”与“互联网+”战略,促进高新科技与文化产业的交融与合作。由此看来在线知识付费产业由于起内容与其特殊的呈现方式,实际上很好地契合了未来我国文化产业宏观发展方向。同时《规划》提出完善现代文化市场体系,加强相关法制建设,出台了有关综合行政、互联网营业规范、产业促进等一系列法律条例,为文化产业发展提供制度保障以及一个更规范良好的发展空间。

另外,在 2020 年 11 月 11 日,人大常委通过了《关于修改<中华人民共和国著作权法>的决定》体现了我国版权制度的完善以及版权保护力度的进一步加大。新修改案完善了知识产权领域的惩罚性赔偿制度,同时扩大了产权保护范围,对以互联网为媒介、以内容产出为主的知识付费服务提供了保障,有利于该行业的健康发展。

2. 经济 (economy)

相关报告显示，从 2016 年到 2020 年的 4 年间，我国经济发展形势向好，居民人均可支配收入稳步上升。可支配收入从 23821 元上涨到了 32189 元，涨幅接近 40%。在经济上行、居民收入攀升的背景下，居民更加注重精神的修养，互联网、大数据等新兴产业也有了稳定的增长动力。

其次，知识付费实际上是内容付费的后续产物，只是在内容上更聚焦于“知识”。在我国，从 2003 年起点中文网的付费观看内容到后来的音乐视频网站会员付费，至今日多种知识付费服务，内容付费已经有了相对较长的发展历程，也形成了相对完备的市场体系与一系列规范。在内容付费发展中形成了许多不同的支付方式，如专栏订阅、支付会员以及知识讲座、一对一问答等一次性付费等，根据不同需求为消费者提供了多种选择，改变了消费者的支付习惯，实际上降低了付费门槛，为知识付费服务的发展打下了一定的基础，争取到了更大的潜在消费人群。

纵观整个市场，目前知识付费处于起步阶段，但由于进入成本并不高，现已涌入了大量不同类型不同规模的平台，如平台式，垂直式等。同时也存在着许多问题，其中最常见的即内容同质化严重或者内容质量不高，以及平台服务有所欠缺，一些类型的服务无法满足高质量高效率的学习要求等问题。所以，各个平台还需要不断做出改进，做出真正令人满意的知知识服务产品，促进这一行业不断向前发展。

3. 社会 (society)

随着经济和文化的发展，我国的国民素质不断提高，社会中对知识和产权的保护、尊重意识有显著提升。除了提供资源之外，知识付费还有信息整理与收集的功能。诚然，互联网降低了信息的传播成本，给人们带来海量信息，那么想要挑选出有价值，有意义，个性化的知识则需要付出巨大的成本。而理想的知知识付费产品可以提供信息筛选服务，并根据产品种类的不同选择性地提供

信息整理，帮助人们获得真正高质量，有体系的信息。

与信息爆炸这一现实相适应的即为现代人普遍的信息焦虑情绪。随着社会节奏的提高与社会竞争的日益激烈，在这种日新月异的发展速度下人们为自身有限的知识与技能而感到焦虑。这种心理促使那些已经脱离学校以及系统性学习的人群去接触，获取多种多样的、新的知识，成为了社会终身学习氛围形成的一个不容忽视的因素。

值得一提的是，知识付费也适应当前碎片化时间利用的潮流。知识付费的内容一般较为短小精炼，加之其以互联网为媒介，知识学习不受时空的限制，满足人们随时随地进行知识吸收的需求。

最后，由于去年疫情影响，线上教育又迎来了一次巨大的正向冲击，更多人认识到并开始关注在线上学习这一种学习方式。

4. 技术（technology）

21 世纪以来，新技术、新制度、新模式不断为行业发展赋能，大数据与互联网行业的发展迅速，随着相关行业的资源不断丰富，再结合日益完善的产业链，使得融合创新应用持续涌现。2019 年，我国的大数据行业的行业规模已达到 5397 亿元。预计到 2022 年，这一数字将突破万亿元。而依托于大数据技术的大数据平台建设为知识付费平台提供了收集信息交互信息的基础，而对信息数据的使用也为知识付费行业生产高质量个性化产品提供帮助。

线上支付技术的成熟与普及也成为了知识付费行业有力的促进因素。《中国互联网络发展状况统计报告》显示，我国的移动支付用户数量已达到 4.55 亿，移动支付覆盖率达到 64.7%，其中使用智能手机进行支付的用户数为 4.21 亿。电子支付技术已成为人们生活中不可或缺的一部分，为线上知识付费服务的交易提供了便利以及技术基础。

总的来说，知识付费行业正处于一个集“天时，地利，人和”的大环境中，

在此发展的关键时期，各方需把握好机遇，正确应对处理各种问题，为形成一个健康有活力的新产业做出努力。

（二）研究目的与意义

1. 研究目的

基于当下经济快速发展，居民生活水平日益提高，综合考虑目前我国居民对内容文化、多元知识的需求，本文以在线知识付费行业作为研究对象进行实证分析。我国知识付费行业的发展阶段主要分为三段：2011-2015 发展初期，企业探水知识付费；2016 爆发增长期，市场出现强劲企业；2017-至今成长期，市场中逐步形成行业巨头。随着行业的进一步成熟化，知识付费市场的格局逐步形成，当下该市场中的企业个体主要可以分为四类：综合平台（比如得到）、垂直平台（比如微课）、在线教育（比如腾讯课堂）、社交平台（比如微信的知识星球）。本文将以上四种类别的知识付费服务终端都纳入所研究的范畴。

本文旨在明晰在互联网蓬勃发展的大背景下，知识付费产业链近年来的发展态势及其主要影响因素，并进一步探明数据作为一种生产要素在其发展过程中所发挥的作用。在此基础上，通过挖掘舆情数据、消费数据等信息，以及调查消费者的相关消费态度、行为，进而深度分析行业用户所表现出的需求偏好，消费关注点，消费情感倾向，以期探究行业未来的发展趋势，为企业提供符合市场内在演化规律的可持续的发展道路。

2. 研究意义

当下，我国的经济发展处于新常态的发展模式，逐渐从单一的高速度转变到高质量发展，推进供给侧结构性改革。同时，创新驱动成为时代经济发展的主旋律，高素质人才成为其中关键一环。为适应经济环境的发展，人们普遍开始接受“终身学习”的生活理念，对知识、教育表现出巨大的需求。知识付费行业紧紧

抓住市场机遇，开始走进人们的视野。因而，对这一新兴行业的深度研究就显得尤为重要且迫在眉睫。

首先，探索知识付费产业链近年来的发展路径及发展的主要影响因素，有利于我们深层次了解知识付费产业的发展脉络、行业生存逻辑，及内在的本质发展规律。此外，量化数据在行业发展过程中所发挥的作用，可以进一步为如何促进数据为产业赋能指明方向。

其次，通过研究消费者基于切身体验而给出的回馈信息，可以明晰市场中消费者的切实需求、消费情感诉求，这有利于企业认识到行业消费者对行业的真实期望值，进而可以有争对性地更新优化所提供的服务模式、产品定位。基于此，市场中的企业可以与消费者保持一致的发展频次，有利于行业整体的稳健可持续健康发展。

此外，知识付费平台作为一种全新的现代化的先进知识传播路径，具有高度的教育属性，为居民提供了大量学习资源、学习平台、虚拟化学习环境。该行业的健康发展在一定程度上可以保障居民持续的学习可获得性，有利于强化社会的学习风气、学习意识，提高民众的科学文化素质、认知广度、眼界宽度，促进国家的文化建设、社会建设。

（三）文献综述

1. 知识付费的相关概念与发展情况概述

知识付费指的是用户为自己所需使用的知识付出资金或其他代价的过程。在这个过程中，知识获得者可以依托互联网，利用移动支付手段向知识传播与筛选者给予相应的报酬。其他的知识传播链条中间人则不能够通过搭便车的行为，利用流量或者广告等方式赚取收益。

知识付费中的“知识”比日常一般认知的知识有着更加丰富的含义。通过用户的付费购买，生产者创造的知识能够实现商业上的变现。简而言之，知识付费是指知识消费者为获得知识所支付的货币成本。因此这里的知识，可以理解为知识产品或知识服务，即人们将认知积累或认知盈余通过交易手段实现共享，付费是指为获取知识产品或服务而进行的支付行为。

2016 年被称为“知识付费元年”。在行业兴起和发展的几年里，我国也出现了许许多多关于知识付费的探讨与研究。不论是内容付费，还是知识付费，知识的共享化与盈利性在互联网时代，逐渐成为一种趋势和潮流。知识付费的概念在我国兴起的时间不长，行业发展也是处于起步和开端，因而存在一些问题。

莫诗怡指出，新媒体环境下，朋友圈等社交媒体会降低付费意愿、知识售卖方创造的拟态环境不稳定、市场混乱损害版权方利益。

据头豹研究院《2019 年中国互联网知识付费行业概览》，从 2017 年到 2018 年，中国互联网知识付费的市场规模从 3.2 亿元增长到 73.7 亿元，涨幅超过 22 倍。目前，随着互联网行业的发展，知识付费行业也得到了相当的助力，且用户付费意识提升，提升了行业的渗透率。不过也存在用户体验不佳、知识内容交付不足、版权问题等制约因素。

艾瑞咨询预测，随着用户知识付费习惯的逐步养成，高价值的付费用户将不断增长，知识付费产业规模将持续扩大。总之，尽管存在或多或少的问题，但从长远来看，知识付费仍然具有十分广阔的前景。因而对于知识付费行业发展的影响因素探究与消费需求调查也就具有十分重要的意义。

2. 知识付费意愿影响因素概述

目前，在知识付费意愿和需求等关于用户行为机理和规律方面的研究多是基于现有的理论模型，在此基础上进行发散和创新。

魏武，谢兴政通过建立结构方程模型、模糊集定性比较分析等方法，分析了用户持续使用知识付费的因素有认同感、人际互动和信息可信度等因素。用户的内在需求的满足与否，将直接决定其继续付费意愿。

杜智涛，徐敬宏研究提出，用户的付费意愿是付费行为的主要影响因素，感知行为控制具有较弱的抑制作用，在线知识付费用户群体呈现出性别、年龄、文化差异。

刘心如发现影响用户知识付费意愿的主要因素是：需求确认、感知价值、主观规范和知觉行为控制等因素。付费意愿、感知价格和决策水平将影响知识付费行为。

(四) 研究方法与技术路线

1. 研究方法

本文主要采用的调查研究方法有模型建立法、网络爬虫数据收集法、文献调查法等。

网络爬虫数据收集法。本文的文本挖掘分析需要大量评论文本数据，用以进行词云图绘制、情感分析、LDA 主题分析等，因此我们利用爬虫技术工具来获取数据，有效节省成本，提升效率。

文献调查法。围绕本文主题，需要针对知识付费行业、结构方程模型、文本挖掘分析等多方面的研究成果进行查阅，以此对知识付费行业的发展现状、背景意义、研究现状、理论基础做一个全面的了解。

模型建立法。依托计量经济学相关知识，利用 stata 软件分析数据，建立以面板数据为基础的计量模型分析知识付费行业发展的影响因素。运用文本挖掘技术，进行 LDA 主题分析，探究用户对知识付费服务的态度。采用结构方程模

型，对消费者付费意愿进行研究。

2. 技术路线

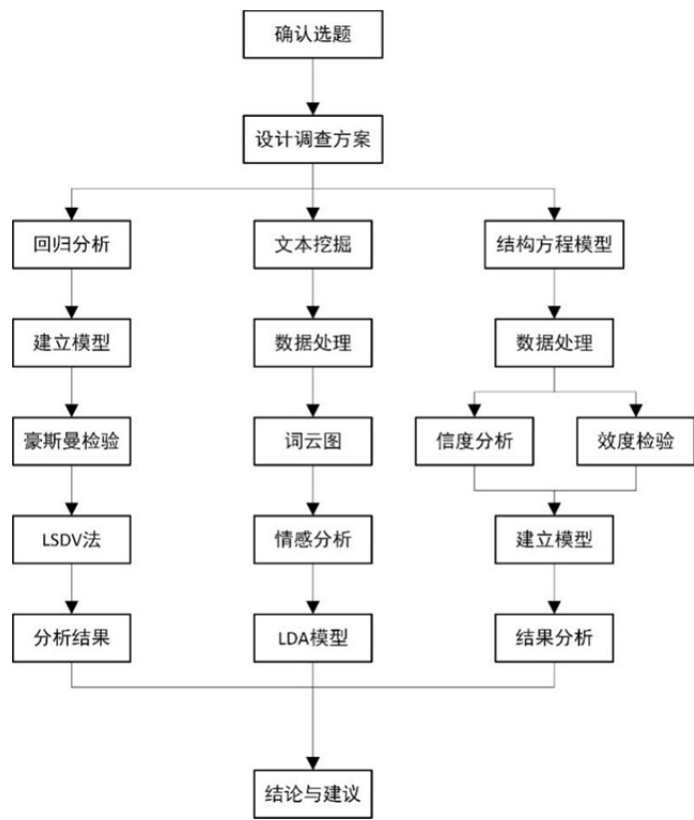


图 1 调查技术路线图

三、 知识付费的发展影响因素分析

(一) 发展现状

近几年来，随着互联网技术的进步、居民消费生态的优化升级，知识付费行业作为时代新秀日益发展壮大。其市场规模逐渐扩大、用户数量不断攀升自然离不开诸多的外部利好因素，如经济大环境的持续向好、消费市场的自然增长、居民平均受教育水平的逐步提高、数字科技的发展等等。

由于知识付费行业具有线上传播的自然属性，数据作为一种全新高能的生产力在其发展过程中发挥了重大作用。首先，相比较于提供纸质教学资源的线下教育，知识付费行业正是凭借提供海量多元的数据与信息服务打破传统的知识分享

模式。其次,企业端通过分析用户的行为习惯来刻画消费者肖像,利用机器学习、推荐算法,可以实现精准营销,提高潜在用户的转化率。再者,企业可以通过对用户的学习行为进行建模、跟踪用户的学习过程,从海量的日志数据中提取有价值的信息,实现精准的个性化定制化服务,提高用户留存率。

知识付费行业兴起于 2016 年,其更早期的形态为在线教育。后期随着付费概念与付费意识在消费者群体中的普及,加之消费者对教育的需求从单一的应试教育类的技能学习逐渐演化为包含软技能在内的多元需求,知识付费便应运而生。在线教育行业的早期表现在很大程度上可以反映知识付费行业的状况,二者并没有本质区别,仅是同一事物在不同社会时期的不同表现形式。

(二) 数据介绍

1. 数据来源

为进一步探明上述因素对知识付费行业的影响,收集了 2004 年至 2019 年期间,中国与美国知识付费行业的有关数据,并运用回归分析技术,构造面板回归模型得到量化结果。

将在线教育行业的市场规模作为被解释变量,消费者价格指数(CPI)、居民人均纯收入、劳动人口平均受教育年限、互联网普及率作为被解释变量。

其中,有关中国知识付费行业规模的数据来源于国泰君安、智研咨询等机构所整理,美国在线教育行业的市场规模数据来源于 Neil Patel Digital 统计整理的数据库。消费者价格指数(CPI)、互联网普及率数据,消费者价格指数(CPI)、居民人均纯收入、劳动人口平均受教育年限、互联网普及率数据,均来源于世界银行统计整理的数据库。中国的居民人均纯收入、劳动人口平均受教育年限数据分别来源于国家统计局统计整理的数据库、百度文库运用爬虫技术爬取的数据库。

2. 数据处理

对于变量缺失的个别数据，采用自回归拟合、线性拟合等多种方法进行插值补全。

相较于消费者价格指数（CPI）、劳动人口平均受教育年限、互联网普及率这三个变量的数量级，市场规模、居民人均纯收入这两个变量的数量级过大，因而通过取对数的方法，减小其变化范围。通过对数据做对数变换，可以在不改变数据的性质和相对关系的前提下，压缩变量的尺度，获得更平稳的数据，并在一定程度上消除异方差。

（三）模型与回归过程

1. 符号说明

表 1 变量说明

变量类型	变量含义	变量名称
被解释变量	在线教育市场规模（亿元）	SI_{it}
解释变量	CPI(100)（以 2010 年作为基年）	CPI_{it}
	居民人均纯收入（元）	IN_{it}
	劳动人口平均受教育年限（年）	ED_{it}
	互联网普及率（10%）	ln_{it}

2. 模型建立与求解

本文采用 2004~2019 年间中国、美国这两个国家的相关数据来构造面板数据模型。假设该模型中，解释变量不包含被解释变量的滞后值，且行业主要的影响因素都被引入。

面板数据模型的数学形式为：

$$\ln SI_{it} = \beta_0 + \beta_1 CPI_{it} + \beta_2 \ln IN_{it} + \beta_3 ED_{it} + \beta_4 ln_{it} + \varepsilon_{it}$$

Hausman 检验

处理面板数据时，首先利用 *Hausman* 检验来推断采用固定效应或是随机效应模型。

Hausman 统计量为：

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})^T \left[\widehat{Var}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \xrightarrow{d} \chi^2(K)$$

其中， K 为 $\hat{\beta}_{FE}$ 的维度，即 X_{it} 中的解释变量个数。

原假设 H_0 : μ_i 与 X_{it} 不相关(随机效应为正确模型)

备择假设 H_1 : μ_i 与 X_{it} 相关

导入数据，在 5% 的显著性水平下，用 stata 软件求解得到 P 值接近 0，因而可以拒绝原假设。此时，宜采用固定效应模型进行进一步的计算和分析。

LSDV 法

确定采用固定效应模型后，再根据 LSDV 法进行回归，通过观察各解释变量的显著性水平，来推动采用混合回归或是个体固定效应模型。

LSDV 模型的数学形式为：

$$\ln SI = \beta_0 + \mu_1 D_2 + \mu_2 D_2 + \beta_1 CPI_{it} + \beta_2 \ln IN_{it} + \beta_3 ED_{it} + \beta_4 IN_{it} + \mu_{it}$$

原假设 H_0 : $\mu_1 = \mu_2 = 0$

备择假设 H_1 : μ_1 、 μ_2 不全为 0

导入数据，通过 stata 得到存在个体虚拟变量在 5% 水平上显著，故不可拒绝原假设 H_0 。“全部个体虚拟变量值都为 0”，认为不存在个体固定效应模型，宜采用混合回归。

回归结果

表 2 回归结果

	混合回归			固定效应			随机效应		
	Coef	t	P	Coef	t	P	Coef	t	P
CPI	42.628	2.63	0.013	56.896	8.30	0.076	42.628	4.47	0.000
IN	-16.760	-4.65	0.000	-12.880	-12.97	0.049	-16.760	-6.48	0.000
ED	2.541	2.10	0.043	-2.751	-17.28	0.037	2.541	3.29	0.001
IN	1.784	2.88	0.043	1.166	12.25	0.052	1.784	3.11	0.002
CONS	93.485	6.99	0.000	89.715	22.23	0.029	93.485	6.24	0.000

（四）分析结论

Hausman 检验结果说明该面板数据模型应该采用固定效应模型，不宜采用随机效应模型。鉴于此，采用 *LSDV* 法进行回归的结果表明，混合回归优于个体固定效应，宜运用混合回归模型。这表明，近十年来，中美两国的知识付费行业的发展状态没有显著的差异。该模型分析结果与现实行业状态相吻合，得益于国家经济、科技等方面的均衡快速发展，及消费者个人对教育的重视，知识付费作为新兴行业在国内具有与发达国家同样利好的成长土壤。

模型结果表明，CPI 每增加 1 单位，知识付费市场规模增加 0.4268%；居民平均受教育年限每增加 1 年，知识付费市场规模增加 2.541%；互联网普及率每增加 10%，知识付费市场规模增加 1.784%。换言之，物价整体水平的上升会带动知识付费行业的名义市场规模增大。居民受教育程度提高后，自身学习能力也得到提升，由于工作、进一步学习的需要，对新知的需求量也随之增加，促进知识付费行业的发展。此外，互联网是实现数据生产力为行业赋能的技术前提与基础，而互联网普及率则进一步测度了数据在该行业的渗透率。根据实证结果，数据对在线知识付费行业的发展具有显著的正向作用。

四、 基于 LDA 模型的知识付费评论主题挖掘

（一）数据的获取与处理

1. 数据来源

数据来源于中国大学生 MOOC（慕课）在线教育平台。为调查用户对知识付费产品的评价与需求状况，选取该平台上参与人数较多、内容质量较高的一门计算机语言课程，找到网页源码地址，利用集搜客网页爬取和内容分析工具，爬取 2018 年 3 月到 2020 年 8 月的用户评论，共计 10008 条。将数据以 CSV 格式的文

件保存，如下图所示：

文本	时间
感觉还是可以的。现在我的地方没有更上，还不理解意思，我想先学一遍在看看那里不熟悉在来学习	发表于 2020-08-26
师傅领进门，修行在个人。任重道远	发表于 2020-08-26
讲的很好 不错!	发表于 2020-08-26
很适合碎片时间的学习	发表于 2020-08-26
嵩天老师的课通俗易懂，对新手学员较为友好，感谢嵩天老师团队为此付出的努力	发表于 2020-08-26
老师讲的非常好	发表于 2020-08-26
课程非常适合非专业的人学，课后题安排也很好	发表于 2020-08-26
学到了很多	发表于 2020-08-26
真的很好，老师讲课很细致	发表于 2020-08-26
课程设计很精心，老师讲的也很棒。	发表于 2020-08-26
挺好的，虽然作业有一些拓展题对我而言挑战性比较强	发表于 2020-08-26
非常棒的课程	发表于 2020-08-26
老师讲的很仔细	发表于 2020-08-26
6666666666	发表于 2020-08-26
感受很好，一点点学起来，只是感觉都学完一遍再看实例比较好（个人建议）	发表于 2020-08-26
非常好的一门课程，受益匪浅，感谢嵩天老师的教师团队!	发表于 2020-08-26
挺好 讲的很细致	发表于 2020-08-26
有一定的效率提升，但无法与课堂完全相同。	发表于 2020-08-26
以前学了一点，没有系统学习，有老师带，可以有一个循序渐进的过程，讲的很仔细，适合初学者	发表于 2020-08-26
很好的入门课程，条理清晰，目标明确，很有收获~	发表于 2020-08-26

图 2 部分文本数据展示

2. 文本分句与分词

为了方便词云图的生成和 LDA 模型的构建 ,我们对原始评论数据进行分句与分词处理。同时，对文本进行预备处理，通过去除特殊字符、去除过短评论、删除重复评论的处理，共删除评论 249 条，剩余评论 9759 条。

在本文的数据中，文本评论多为短句的集合。为了将文本分词，首先应该分句。通过识别短句符号以及根据句尾的标点符号（如“。！？；”）来判断一句话是否结束，如下表：

表 3 部分分句结果表

原句	分句
嵩天老师的课通俗易懂，对新手学员较为友好，感谢嵩天老师团队为此付出的努力	[‘ 嵩天老师的课通俗易懂 ’ ‘ 对新手学员较为友好 ’ ‘ 感谢嵩天老师团队为此付出的努力 ’]
课程非常适合非专业的人学，课后题安排也很好	[‘ 课程非常适合非专业的人学 ’ ‘ 课后题安排也很好 ’]
挺好的，虽然作业有一些拓展题对我而言挑战性比较强	[‘ 挺好的 ’ ‘ 虽然作业有一些拓展题对我而言挑战性比较强 ’]
感觉很好，老师很认真	[‘ 感觉很好 ’ ‘ 老师很认真 ’]

为了方便接下来的词频分析和情感分析，在分句的基础上，我们进一步做分词处理。分词部分结果如表 4：

表 4 部分分词结果表

原句	分词
课程非常适合非专业的人学 ,课后题安排也很好	[‘ 课程 ’ ‘ 非常 ’ ‘ 适合 ’ ‘ 非专业 ’ ‘ 课后题 ’ ‘ 安排 ’ ‘ 很好 ’]

感觉很好，老师很认真	[‘ 感觉 ’ ‘ 很好 ’ ‘ 老师 ’ ‘ 很 ’ ‘ 认真 ’]
------------	--

(二) 词频统计和词云图

经过文本预处理之后，文本数据有了较好的分析性质。接下来对评论数据进行词频统计。我们采用 Pos-tag 的方式对词语进行标记，选取其中词频大于 530 的词语，如下表：

表 5 评论文本词频统计

词语	频数	词语	频数
老师	3686	不错	774
课程	1966	内容	618
非常	1848	很棒	597
学习	1233	感觉	577
讲解	987	清晰	538

由表 3 可知，用户提到最多的三个词是“老师”“课程”“非常”。在此，由于数据来源于某在线课程的评论，因此“老师”“课程”出现的频率多，属于正常现象。“非常”表示程度很深，既可以用以表示非常好，也可以用以表示非常差。结合其他词频数据，我们可以推断，这里的“非常”基本上适用于称赞该课程的。这表示出用户在主观上是认同该课程的，展现了课程讲师的水平和平台服务的水准。

基于 python，利用分句和分词后的评论构建词云图，使评论数据更具有可视性，如下图：



图 3 词云图

通过词云图我们可以清晰地看到，“老师”“讲解”“基础”“清晰”“编程”等是出现频率较高的词汇，说明用户对课程内容、讲解方式等关注较多。

（三）评论文本的情感分析

情感分析的作用为分析目标文本所包含的情感色彩。无论是网购还是选择知识付费服务，用户都会参考他人对现有服务的评价，做出相应的消费选择。在此基础上，我们不难看出，文本评论所承载的情感倾向能够切实地影响到用户的使用意愿。因此，情感分析的重要性不言而喻。

文本情感分析过程利用 python 的 SnowNLP 包对每个词语的情感色彩进行评估。根据评论中情感色彩的积极因素占比来打分,其分值在[0,1]区间内。分值越大,说明该评论的正面感情色彩越多。通过参考其他文章和比较,将阈值定位在 0.5。将情感分数大于 0.5 的评论视为好评;小于 0.5 的评论则判为差评。

将每条评论的情感分整理成直方图，如下图所示：

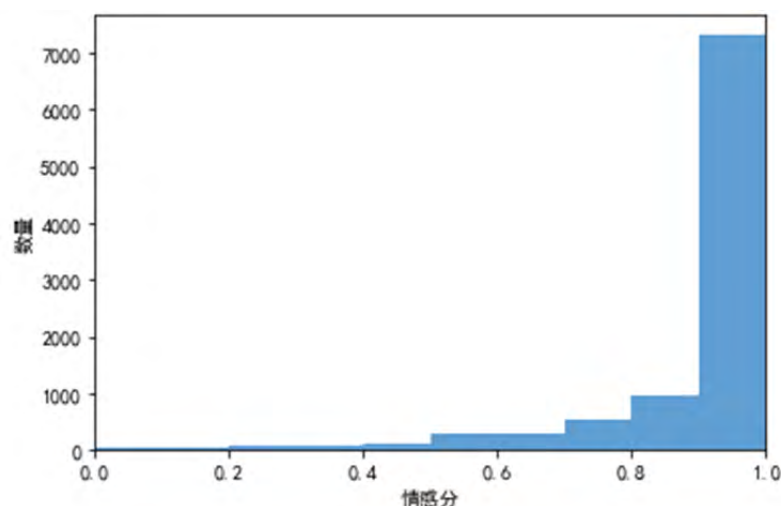


图 4 情感分直方图

从以上的情感分直方图中，我们不难看出，大部分用户评论都是积极的，其中，绝大多数的用户评论的情感分集中在 0.8-1 之间。说明用户对该课程的使用体验和评价均是良好的。

情感分析为我们观察和了解用户的情感倾向提供了直观而又有效的方式，但是情感分析不能为我们提供具体的主题内容和对应的反馈意见，因此我们采用 LDA 分析，挖掘评论文本中的主题词，做更深入的分析。

(四) 文本评论的 LDA 主题分析

1. LDA 模型介绍

LDA (Latent Dirichlet Allocation) 是一种文本主题模型。在模型的假设条件下，文档、主题和分词均服从多项分布。模型对每一篇文档均抽取一个主题，在根据主题的分布抽取分词，对每个文档、主题、分词均重复以上过程。LDA 采用词袋模型，词之间没有先后顺序的关系，仅考虑一个词汇是否出现。以此，利用每个文档所抽取的主题和分词数量，就可以用一个词频向量来表示每篇文档，从而达到将文本数据转化为数字数据的目的。一个 L 维向量 $(1, 0, 0, \dots, 0)$ 表示一个词汇。 M 篇评论分布着个主题，记为 $Z_i (i=1, 2, \dots, K)$ 。由个词构成的评论记为

$$d = (w_1, w_2, \dots, w_i)$$

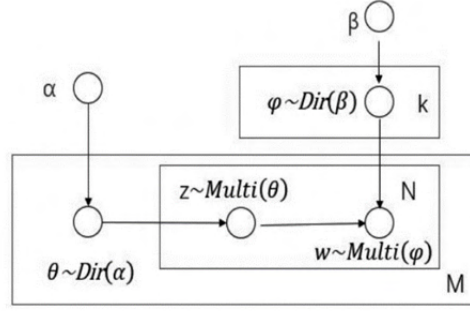


图 5 LDA 模型结构图

LDA 模型结构如图 5 所示。 α 和 β 是狄利克雷函数的先验参数； θ 是主题在文档中的多项分布的函数，其服从超参数为 α 的 *Dirichlet* 先验分布； ϕ 是词在主题中的多项分布的参数，其服从超参数 α 的 *Dirichlet* 先验分布。LDA 模型假设每条评论是所有的主题按照一定比例组合而成的，而组合比例服从于多项分布，记为：

$$Z | \theta = \text{Multinomial}(\theta) \quad (1)$$

每个主题词典是由所有的分词按照比例组合而成，组合比例也服从多项分布，记为：

$$W | Z, \phi = \text{Multinomial}(\phi) \quad (2)$$

在评论 d_j 条件下生成词 w_i 的概率表示为：

$$P(d_j | w_i) = \sum_{s=1}^K P(w_i | z = s) * P(z = s | d_j) \quad (3)$$

其中， $P(w_i | z = s)$ 表示词 w_i 属于第 s 个主题的概率； $P(z = s | d_j)$ 表示第 s 个主题在评论 d_j 中的概率。

2. 文本评论的 LDA 主题分析

在 LDA 主题分析之前，我们利用百度云 NLP 算法和 t-SNE 降维算法，对数据进行降维并输出可视化图像，用以大致判断分词之间的关系，输出散点图如下：

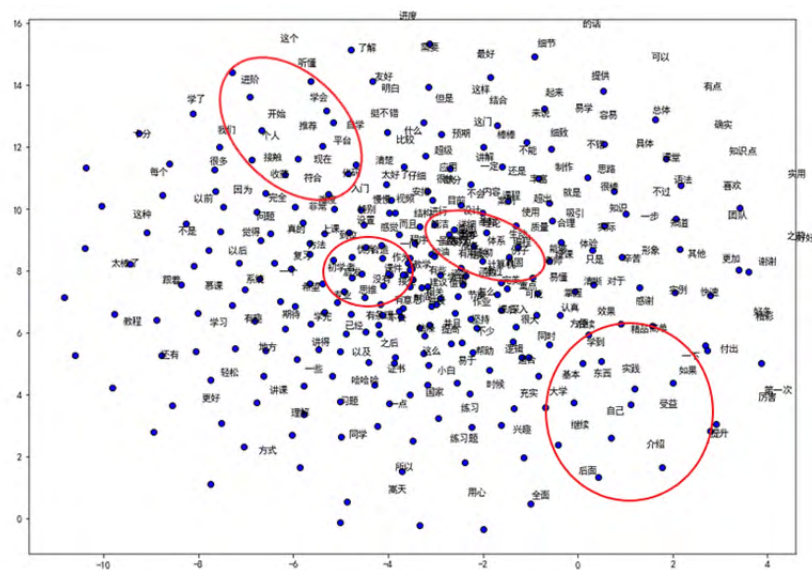


图 6 文本分词散点图

通过分词散点分布以及语义之间的联系，我们大致可以将文本主题词分为 4 类，如下表所示：

表 6 网络课程主题词表

主题一	主题二	主题三	主题四
通俗易懂	很棒	课程	老师
适合	收获	有用	讲得
初学者	很大	自习	细致
清晰	详细	实用	好
简单	学习	强烈推荐	喜欢
课程内容	学到	太棒了	Good

由上表可知，主题一主要是用户认为课程内容比较容易让人接受，作为计算机语言初级课程，能够让初学者听懂，老师水平较高，授课条理清晰，节奏掌握较好；主题二主要是用户学习课程之后的感受，大多认为从该课程中收获很大；主题三主要是对于课程内容的评价，大多认为该课程十分实用；主题四是对老师的评价，从分词中可以看出，用户非常接受老师的教学方式，也对老师有较高的认可。

通过 LDA 主题分析，我们可以初步得出关于知识付费相关课程和服务的部分结论。在课程设计上，应该针对不同水平的课程设置不同的难度，教师授课也

应该对节奏和可接受度多加把握；在内容选择上，应该注重内容的应用性和理解性，能够让用户在学习之后有良好的感受，并能够付诸实践；在师资水平上，我们可以看出教师的水平与用户的体验有密切的关系，因而需要重点关注。

五、 知识付费市场的消费需求分析

（一） 模型介绍与建立

1. 结构方程模型介绍

在有关心理、教育、社会等的研究中，有的概念如满意度、动机、雄心、智力等往往难以直接准确度量，这类变量我们一般称之为潜变量。因此我们通过一些外显指标，也即可以直接观测的指标来间接测度潜变量。采用结构方程模型，可以直观准确的为我们呈现出外显变量和潜变量之间的关系。

测量模型与结构模型为构成结构方程模型的两部分。如下方程所示，测量模型体现了我们所观测的变量和潜变量之间的关系：

$$\begin{aligned}X &= \Lambda_X \xi + \delta \\Y &= \Lambda_Y \eta + \varepsilon\end{aligned}$$

其中 ξ 与 η 分别为潜变量， X 与 Y 为不同的潜变量对应的观测变量， Λ_X 和 Λ_Y 分别为 X 与 Y 的因子载荷， δ 与 ε 是影响观测变量的残差。

结构模型分析的是潜变量之间的关系：

$$\eta = B\eta + \Gamma\xi + \zeta$$

ξ 为自变量潜变量， η 为因变量， Γ 为因变量对自变量的影响系数， ζ 为方程的误差项。

2. 模型构建

① 理论依据

杜志涛、徐敬宏概括整合了一系列理论模型，总结在线知识付费意愿主要受需求与体验两个维度的因素影响。需求分为内在需求与外在需求，而针对知识付费这一内容、形式皆较为特殊的消费，体验主要由所感知到的专业性、趣味性、便捷性与主观规范四个因素组成。

同时在朱鸿军与张萌对于劳动快感的分析中，加之当今人们所面对的外在内在压力，提出部分平台用户实际上是在压力驱动下为知识投入劳动与金钱来满足快感，其付出劳动或金钱成本的过程实则是在为知识赋值。所以我们认为用户在知识付费产品或服务的消费过程中所感知到的成就感与收获感将会对其支付意愿造成一定的影响。所以我们通过平台所提供的监督来间接地度量用户在知识学习过程中所期望能得到的劳动快感。

② 理论假设与模型构建

根据以上理论基础，我们构建如下模型，且假设以下 8 个项目均对用户的支付意愿有着促进作用。

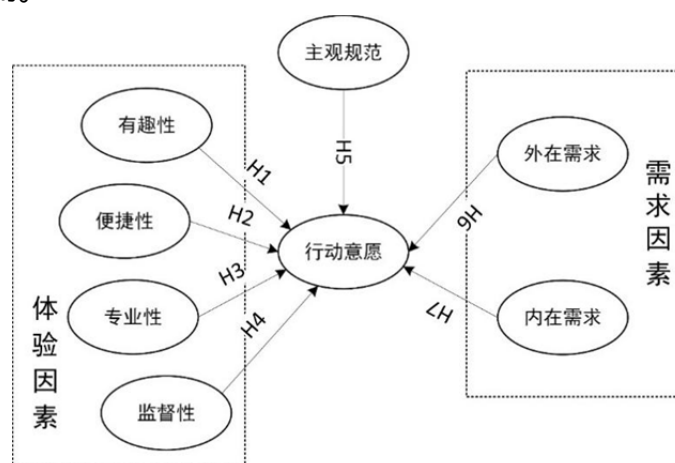


图 7 模型与假设示意图

为了得到潜变量的信息，需要通过各外显变量来测度。我们通过阅读相关文献进行总结加工，最后得到了以下 18 个测度变量，见下表。以下的潜变量均由

“1=完全不赞同”到“5=完全赞同”的量表测度。

表 7 潜变量测量与问卷对应表

项目	题号
行动意愿 (Willingness to act) WTA	T7, T8
外在需求 (External demand) ED	T9, T10, T11
内在需求 (Internal demand) ID	T12, T13
专业性 (Specialty) SP	T15, T16
有趣性 (Interesting) I	T17, T18
便捷性 (Convenience) C	T19, T20
主观规范 (Subjective norm) SN	T14, T21, T22
监督性 (Supervision) SU	T23, T24

(二) 数据处理与结论分析

1. 数据处理与检验

数据收集与描述性统计分析

本次调查利用问卷星平台发放问卷，共收集问卷 230 份，首先剔除填写时间少于一分钟的问卷，得到有效问卷 218 份。其中数据处理与分析采用 SPSS 软件，结构方程模型构建与分析采用 Amos 软件，同时利用 python 语言进行部分数据可视化展示。数据描述性统计分析指标如表 8。测量变量的相关矩阵、标准差以及均值如表 9。

表 8 描述性统计分析表

项目	类别	频率	百分比
性别	男	102	47.2
	女	115	52.8
年龄	18 岁以下	4	1.8
	18—25	105	48.2
	26—35	32	14.7
	36—45	33	15.1
	46 及以上	44	20.2
受教育程度	初中及以下	8	3.7
	高中/中专	43	19.7

	本科/大专	142	65.1
	硕士	21	9.6
	博士及以上	4	1.8
月可支配收入	1000 元以下	13	6.0
	1000—3000 元	99	45.4
	3001—5000 元	42	19.3
	5001—8000 元	36	16.5
	8000 元以上	28	12.8
消费金额	0—50 元	78	35.8
	50—150 元	72	33.0
	150—500 元	37	17.0
	500—2000 元	19	8.7
	2000 元以上	12	5.5
使用时长	0—5 分钟	46	21.1
	6—30 分钟	77	35.3
	31—60 分钟	54	24.8
	1—2 小时	29	13.3
	2 小时	12	5.5
	总计	218	100.0

表 9 测量变量的相关矩阵、均值和方差

	WTA1	WTA2	ED1	ED2	ED3	ID1	ID2	SP1	SP2	I1	I2	C1	C2	SN1	SN2	SN3	SU1	SU2
WTA1	1.000																	
WTA2	0.678	1.000																
ED1	0.498	0.546	1.000															
ED2	0.601	0.682	0.596	1.000														
ED3	0.627	0.636	0.656	0.688	1.000													
ID1	0.629	0.684	0.581	0.640	0.725	1.000												
ID2	0.569	0.654	0.531	0.633	0.686	0.664	1.000											
SP1	0.451	0.446	0.586	0.475	0.506	0.534	0.462	1.000										
SP2	0.517	0.489	0.681	0.506	0.542	0.537	0.535	0.675	1.000									
I1	0.507	0.559	0.591	0.521	0.543	0.556	0.500	0.620	0.662	1.000								
I2	0.450	0.470	0.577	0.507	0.578	0.557	0.552	0.504	0.585	0.629	1.000							
C1	0.520	0.461	0.625	0.528	0.517	0.568	0.455	0.656	0.656	0.634	0.538	1.000						
C2	0.546	0.510	0.558	0.506	0.553	0.582	0.492	0.515	0.559	0.561	0.573	0.642	1.000					
SN1	0.596	0.592	0.600	0.595	0.608	0.639	0.605	0.590	0.636	0.576	0.581	0.544	0.525	1.000				
SN2	0.589	0.532	0.686	0.539	0.582	0.609	0.567	0.649	0.714	0.666	0.615	0.665	0.614	0.626	1.000			
SN3	0.536	0.447	0.604	0.506	0.519	0.525	0.486	0.614	0.661	0.611	0.585	0.708	0.623	0.600	0.725	1.000		
SU1	0.526	0.474	0.644	0.538	0.595	0.643	0.530	0.669	0.692	0.624	0.589	0.744	0.655	0.586	0.740	0.650	1.000	
SU2	0.586	0.551	0.605	0.583	0.629	0.613	0.603	0.599	0.663	0.658	0.652	0.635	0.632	0.640	0.692	0.650	0.683	1.000
Mean	3.587	3.436	3.881	3.688	3.693	3.560	3.486	3.876	4.046	3.872	3.757	3.789	3.679	3.725	3.894	3.844	3.908	3.830
S.D.	1.185	1.206	1.088	1.079	1.137	1.127	1.161	1.111	1.115	1.108	1.099	1.128	1.089	1.143	1.100	1.066	1.069	1.145

注：WTA 为行动意愿；ED 为外在需求；ID 为内在需求；SP 为专业性；I 为有趣性；C 为便捷性；SN 为主观规范；SU 为监督性。

利用 python，将相关系数矩阵利用热力图进行可视化展示，其中颜色深浅表示数值大小，颜色越深数值越大，变量之间相关性越高。

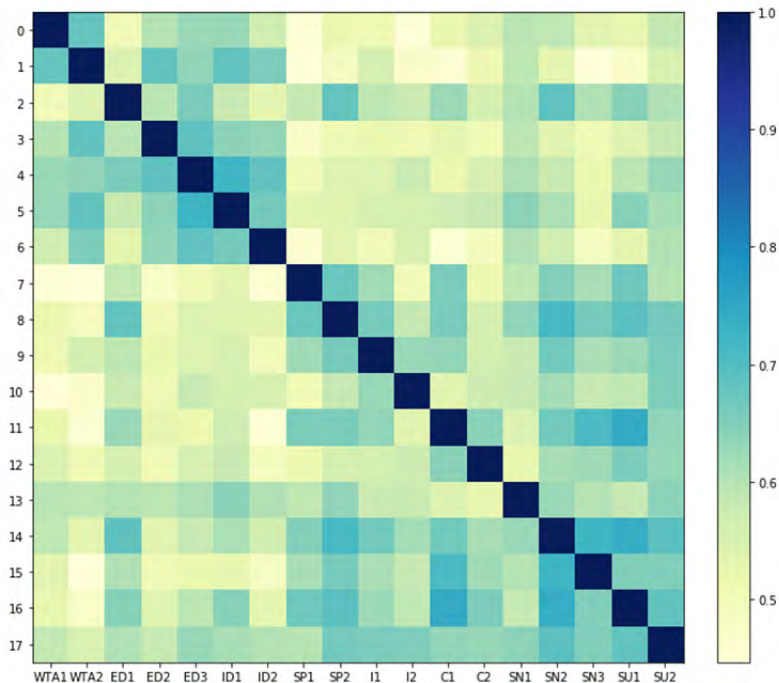


图 8 相关系数矩阵热力图

信度分析与效度检验

在进行结构方程模型分析之前，对数据的信效度进行检验和分析，检验数据的性质是否能够支持我们进行进一步的分析，保证分析结果的可靠性。

信度检验主要可以分为同质性信度与分半信度。同质性主要利用 α 技术，测验内部所有题目之间的一致性。一般而言，Cronbach's α 系数大于0.7，说明内部一致性较好。

分半信度主要利用斯皮尔曼-布朗系数对可靠性进行检验。其主要过程为，将量表奇偶分为两部分，分别利用每部分的Cronbach's α 系数，对两分量表的信度进行比较。

如下表，Cronbach's α 系数，斯皮尔曼-布朗系数均大于0.75，说明量表具有较好的信度。

表 10 Cronbach ' s alpha 系数与折半系数

	Cronbach ' s alpha	斯皮尔曼-布朗系数	项数
行动意愿 (WTA)	0.808	0.808	2
外在需求 (ED)	0.846	0.858	3
内在需求 (ID)	0.798	0.798	2
专业性 (SP)	0.806	0.806	2
有趣性 (I)	0.772	0.772	2
便捷性 (C)	0.782	0.782	2
主观规范 (SN)	0.847	0.864	3
监督性 (SU)	0.811	0.812	2
总计	0.957	0.919	19

在问卷数据具有良好信度的基础上,对其进行效度检验。如下表所示,问卷数据的 KMO 统计量值为 0.964,说明问卷数据适合做因子分析。

表 11 表 KMO 和巴特利球形度检验

KMO	0.964
近似卡方	3180.885
自由度	190
显著性	0.000

利用验证性因子分析,测量各因子标准化载荷 (*estimate*)、平均方差提取量 (*AVE*)和组合信度 (*CR*)如下表所示。其中,所有测量变量指标的因子载荷均大于 0.7,表示数据具有较好的聚合效度。此外,表中数据显示,组合信度 (*CR*)均大于 0.75,平均方差 (*AVE*)均大于 0.5,表明问卷数据具有良好的信效度。

表 12 验证性因子分析,平均方差提取量和组合信度

路径	estimate	AVE	CR
Q8 <--- 行动意愿	0.837	0.6529	0.7898
Q7 <--- 行动意愿	0.778		
Q9 <--- 内在需求	0.722	0.6472	0.8454
Q10 <--- 内在需求	0.808		
Q11 <--- 内在需求	0.876		
Q12 <--- 外在需求	0.837	0.6655	0.799

Q13	<---	外在需求	0.794		
Q16	<---	专业性	0.845	0.6754	0.8061
Q15	<---	专业性	0.798		
Q20	<---	便捷性	0.784	0.6419	0.7818
Q19	<---	便捷性	0.818		
Q21	<---	主观规范	0.871		
Q14	<---	主观规范	0.724	0.6561	0.8505
Q22	<---	主观规范	0.828		
Q24	<---	监督性	0.792	0.6852	0.8129
Q23	<---	监督性	0.862		
Q18	<---	有趣性	0.755	0.632	0.7741
Q17	<---	有趣性	0.833		

2. 模型结果与分析

在结构方程模型中，经过 Amos 软件拟合后的结果与各变量的路径系数如表 13 与图 9 所示。

表 13 结构方程结果表

因变量	自变量		路径系数
行为意愿	需求因素	外在需求	1.406
		内在需求	-0.531
	体验因素	专业性	0.246
		有趣性	0.031
		便捷性	0.148
		监督性	-0.442
		主观规范	0.19

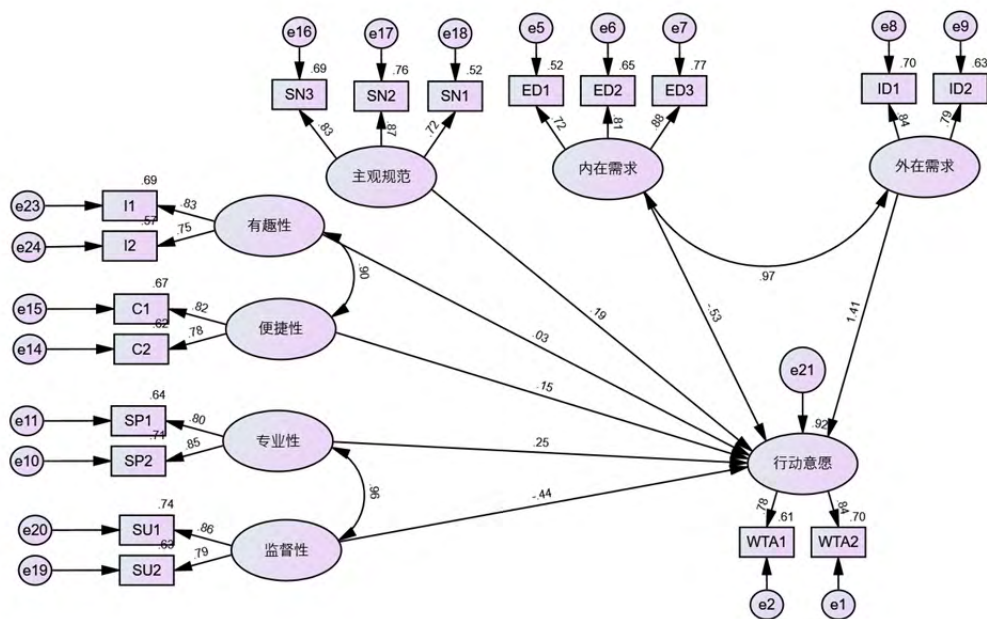


图 9 结构方程模型图

由图 2 与表 1 可得，除去内在需求与监督性对于用户的支付意愿有明显的负向影响，其余的因素均对知识付费意愿有促进作用，与一开始的预设有一定的偏差，同时各路径系数的不同反应了各因素的影响有明显差异，但均可以得到合理的解释，也较好地反映出当下用户对于知识付费服务各维度的需求、偏好情况。

需求是决定用户是否选择付费的主要因素：

需求因素中的各潜变量的路径系数均明显大于体验因素中的各变量，同时外在需求的系数远大于其他因素，成为决定用户付费意愿的最主要因素。说明用户的知识付费意愿更多地来自于对工作学习中实际的需求，工具性与功利性较强。而内在需求与支付意愿是负向相关的，证明大多数人更愿意通过知识付费学习一些偏实用性的知识而很少通过这种方式来达成情感目的。

专业性、便捷性以及主观规范对用户知识付费起到促进作用

用户在使用知识付费产品或服务中对专业性、便捷性的期望均对其付费意愿

产生显著的促进作用。这一结果与外在需求对于用户付费意愿的影响相适配。人们在知识付费过程中更多寻求的是一种实用性与功利性,所以付费知识的专业性与便捷性无疑是提高了付费知识学习过程的效率,一定程度上满足了用户的需求。

而主观规范的正向影响则揭示社会氛围、规范信念等对于付费知识生产者的重要影响。而知识付费作为通互联网平台进行实现的产业,其传播有着独特优势,在如今互联网与生活密切相关的时代,更容易在人群中形成一种风尚,促进人们的消费行为。

用户对于趣味性的期望对其支付意愿影响较小

由数据可知,有趣性对支付意愿虽有正向影响,但影响较小,系数绝对值为各路径系数最小。说明当人们考虑是否进行知识付费时,较少考虑有趣与否的问题。这一结果与人们更注重在知识消费中自身需求是否得到满足一致,在满足知识需求的过程中,用户可能更注重成效与效率。同时也可能是因为现有的知识付费行业中少有趣味与专业兼具的产品,亦或是受线下知识传递中整体较为严肃的氛围影响。

对于监督性的期望对用户的支付意愿产生显著的负向影响

由分析结果可得,在在线知识付费中,对于平台监督性的感知与期望并不如我们所预期那样带来更多的成就感与劳动快感,反而成了抑制支付意愿的因素。我们认为这也许是因为知识付费服务相比传统的知识教育更具灵活性,以用户更大的空间,营造一个较为轻松的学习环境。而平台的监督更可能会打破这种灵活与轻松感,相较学习效果提升的期望,更多给用户带来一种压力与压迫感。所以平台若是需要通过平台自身服务来保证用户的学习效果,则需要对方式的采取这一问题多加注意。

六、 结论与建议

(一) 结论

本研究基于 LDA 与结构方程模型对互联网背景下知识付费发展与付费意愿影响因素进行分析,得出以下结论:

1. 社会经济以及相关行业的发展助力在线知识付费行业发展

物价水平的提升带动了知识付费行业的规模增大,同时在线知识付费行业的发展也与居民受教育程度以及互联网行业的发展保持一致,体现了在线知识付费行业作为商品与其他商品的共通性以及其在内容以及媒介上的特殊性。无疑互联网以及大数据的发展,会对在线知识付费行业产生巨大的推力。

2. 在线知识付费服务需更具灵活性

通过 LDA 主题分析,知识付费内容的难易程度,知识的实用性以及输出方式等都对用户的感受有重要影响。加之知识付费使用人群具有一定的差异性,所以生产个性化的服务产品、避免同质化将是行业的发展重点。

3. 服务需保证工具价值并发展体验价值

在付费意愿的分析中,我们得到现有用户的付费意愿更多地来源于对知识技能的实际要求,其消费中的关注点更集中于在消费中的收益。所以我们认为在在线知识付费行业发展的过程中,一方面需要满足用户这方面的需求,保证内容的质量,学习过程的效率;另一方面改善用户在消费过程中的体验,以在消费心理层面对用户进行引导,创造更大的市场。

(二) 建议

由以上分析,我们认为在线知识付费行业是一个极具潜力的新兴产业,且正处于一个良好的发展环境中。行业生产者需要切实了解用户的心理与需求,提供

更具个性化、针对性的服务。同时政府需要加强市场规范，保证行业健康有序发展。除此之外，在线知识付费与互联网、大数据、教育等行业关系紧密并以此为基础，因此这些相关行业的成长也值得关注。

参考文献

- [1]杜智涛,徐敬宏.从需求到体验:用户在线知识付费行为的影响因素[J].新闻与传播研究,2018,25(10):18-39+126.
- [2]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(08):119-129.
- [3]莫诗怡.新媒体环境下知识付费发展的现状、问题与对策[J].传媒论坛,2019,2(21):105+108.
- [4]刘心如.用户知识付费行为的影响因素研究[D].哈尔滨工业大学,2019.
- [5]姜萌.知识付费 APP 营销路径探析——以“樊登读书”为例[J].现代视听,2019(08):55-56.
- [6]赵蔚华.知识付费对传统出版的冲击与救赎[J].编辑学刊,2018(04):16-20.
- [7]张厚栋,徐爱民.基于 LDA 模型的电商用户评价分析[J].浙江万里学院学报,2020,33(06):91-96.
- [8]王少平,杨继生,欧阳志刚.计量经济学(第二版)[M].北京:高等教育出版社,2020-12,293-305
- [9]陈强.高级计量经济学及 Stata 应用[M].b 北京:高等教育出版社,2014-4,272-302
- [10]沈波,蒋岩汐.知识付费产品研究综述[J].商业经济,2021(01):120-123.
- [11]孟迪云.论数字经济时代有声阅读的商业模式创新[J].中国经贸导刊(中),2020(03):126-127.
- [12]汪海心.基于互联网经济新形态的知识经济的 IP 化生存探讨[J].现代商贸工业,2021,42(06):4-5.
- [13]严炜炜,陈若瑜,张敏.基于元分析的在线知识付费意愿影响因素研究[J].情

报学报,2021,40(02):204-212.

[14]陈海砚.后疫情时代,知识付费行业的困局与机遇[J].中关村,2021(03):89.

[15]张慧.数据驱动下的出版业知识服务模式创新微探[J].新闻世界,2021(03):47-50.

[16]赵子忠,杨少彤.互联网内容付费的特征[J].青年记者,2021(04):13-15.

[17]朱鸿军,张萌.劳动快感的供给及其调适:再论知识付费平台的实践逻辑[J].现代出版,2021(02):52-58.

[18]胡妍妍.知识付费成为知识服务新潮流[N].人民日报,2021-04-02(020).

[19]范周.深度解读《文化部“十三五”时期文化产业发展规划》

<http://scim.cuc.edu.cn/2018/0904/c2672a79462/page.htm> 2021.5.28

[20]2020年全国居民人均可支配收入

http://www.gov.cn/guoqing/2021-04/09/content_5598662.htm 2021.5.28

附录一：调查问卷

关于线上知识付费的支付意愿调查

您好！感谢您参与本次问卷调查，该问卷用于调查人们对知识付费服务的支付意愿以及各影响因素，一共 24 题，大约用时 3 到 4 分钟，请根据您的实际情况和看法进行填写。

本次调查采用匿名形式，所有回收的数据也仅供学术研究参考，请放心填写，谢谢您的支持与配合。

1. 您的性别为 [单选题] *

☐男

☐女

2.您的年龄为

[单选题] *

☐18 岁以下

☐18—25

☐26—35

☐36—45

☐46 及以上

3.您的受教育程度 [单选题] *

☐初中及初中以下

☐高中/中专

- 本科/大专
- 硕士
- 博士及以上

4.您的月平均可支配收入为 [单选题] *

- 1000 元以下
- 1000—3000 元
- 3001—5000 元
- 5001—8000 元
- 8000 元以上

5.您对知识付费产品的年平均消费金额大约为（包括但不限于知乎会员、各平台付费课程、喜马拉雅付费讲座、各类付费问答等） [单选题] *

- 0—50 元
- 50—150 元
- 150—500 元
- 500—2000 元
- 2000 元以上

6.您平均一次使用知识付费产品或服务的时长为 [单选题] *

- 0—5 分钟
- 6—30 分钟

○31—60 分钟

○1—2 小时

○2 小时以上

7.条件允许情况下，我愿意经常购买在线付费知识服务或产品 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

8.相对于线下课程，我更愿意购买在线知识付费课程 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

9.我使用或将要使用在线知识付费产品是为了提升我的工作学习能力 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

10.使用在线知识付费产品能够让我较为系统、全面地掌握一门知识 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

11.使用在线知识付费产品能解决我在生活中遇到的困难，让我更深刻地认识，了解自己。 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

12.使用在线知识付费产品让我觉得生活特别充实 [单选题] *

完全不赞同	○1	○2	○3	○4	○5	完全赞同
-------	----	----	----	----	----	------

13.我认为在线付费学习在一定程度上可以缓解我的焦虑 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

14.我觉得上进、优秀的人都愿意为获取新知识而付费 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

15.我在网上学习和获取知识时，十分看重知识提供者或是服务平台的权威性与专业性

[单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

16.在线付费学习中，内容的质量和专业度是决定我是否付费的重要因素之一 [单选题]

*

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

17.我在网上学习和获取的知识都是自己感兴趣的内容 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

18.我更愿意参与那些看起来轻松有趣，不那么严肃的知识获取过程 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

19.在线学习和获取知识的操作是否便捷对于我是否进行支付起着很关键的作用 [单选

题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

20.支付的便捷性对于我是否购买在线付费知识内容起关键作用 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

21.其他用户的评价和评论会直接影响我对于在线知识付费产品的支付意愿 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

22.朋友或者有影响力的人的推荐会影响我对于在线知识付费产品的支付意愿 [单选题]

*

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

23.我觉得平台在付费过程中应该起到更积极主动的作用（如在线课程服务中多与用户沟通以形成反馈，或起到学习监督作用等） [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

24.我认为我需要在在线学习的过程中，有更多与其他用户以及知识生产者讨论，交流的机会 [单选题] *

完全不赞同 ☐1 ☐2 ☐3 ☐4 ☐5 完全赞同

25.在选择在线知识服务或产品时，您最关注的因素为 [排序题，请在中括号内依次填入

数字] *

至多选 3 项，至少选一项

☐ 学习内容、过程的专业性

☐ 学习内容、过程的有趣性

☐ 价格

☐ 平台服务的合理及时性

☐ 产品、服务使用的便捷性

致谢

该作品的完成诚然是小组成员共同努力的结果，但是也同样离不开其他人的帮助。

首先要感谢我们的指导老师 xxx，xxxxxx 大学 xxxx 学院教授。xxx 老师对我们作品的选题和调查方法上给出了指导性意见，他的指导和鼓励，是我们前进的动力。同时要感谢我们的计量经济学 xxx 老师，在计量模型的建立过程中给予我们重要的帮助和建议。

此外，还需要感谢在作品完成过程中给我们带来帮助的同学朋友们，他们给我们提供生活中的灵感和交流讨论的契机。也要感谢参考文献中的作者们，通过参考前人的文章，我们才能够走得更远。

最后还要感谢各位评审论文的老师，比赛的顺利完成，离不开他们的努力。由于团队成员水平和精力有限，作品中涉及的问题诚然还有许多等待解决，您的批判将成为我们下一阶段学习的动力。