

智能手环使用现状及产品发展分析报告

——基于电商平台消费者购买评价的文本挖掘与武汉市的抽样调查



中南的后裔

指导教师：张海波

成 员：万文娇

孔慧琳

魏羽璠

黄 菡

摘要

随着人工智能技术不断取得新的突破,一系列智能生活硬件逐步走进普通大众的日常生活中。智能手环凭借入门级的售价和产品方案的标准化,市场渗透率在智能可穿戴设备领域中一马当先,远远超过智能手表、智能眼镜等其他产品。智能手环作为新兴产品,其市场潜力还未完全释放,正处于承前启后的发展阶段。现阶段消费者对智能手环的评价如何?智能手环厂商今后应采取什么样的发展策略?针对第一个问题,我们通过抓取电商平台评论文本,使用大数据挖掘的方法,了解消费者对产品的好评度以及关注点。针对第二个问题,通过实地调查的方式,结合现有用户使用情况和非用户潜在价值,对智能手环未来发展策略提出有针对性的可取建议。

考虑到在当今的移动互联网时代,通过电商平台进行网络购物已成为人们生活中一种普遍的消费方式。利用文本挖掘的方法通过爬取天猫平台和京东平台上,小米手环的全部可抓取评论数据,共计 468100 条。经过数据清洗后,使用中文分词、构建语义网络与 LDA 主题模型等算法和模型,实现对文本评论数据的情感倾向性判断以及所隐藏信息的挖掘与分析,以有效、准确的从互联网电商评论数据中分析出消费者对小米手环的好评度、关注点以及抱怨点,对小米手环的功能、优缺点进行初步的评价,为线下问卷调查提供方向和思路。

在问卷调查实施中,我们进行了试调查和正式调查两轮调查。在正式调查中,我们以武汉市社区居民为调查对象,采用分层抽样和三阶段抽样相结合的概率抽样调查方式,将武汉市的分为中心城区和远城区两层,其中三阶段抽样是从武汉市所有行政区中利用 PPS 抽样法抽取行政区,再从行政区中抽取社区,最后利用系统抽样的方式抽取常住居民,保证最终单元入选样本的概率均相同。试调查问卷各项检验均通过,问卷结构合理。正式调查数据进行信度、效度分析,并且进行游程检验和独立性检验均通过,调查结果真实可信。

针对问卷调查的结果,我们运用描述性统计方法,分别对用户和非用户的基本信息进行分析。借助 Logit 二元选择模型,分析得到性别、年龄、学历、兴趣爱好、性格、运动频率是购买智能手环的显著影响因素。基于回归结果,利用 K-means 聚类的方法,定量分析潜在用户性别、年龄、学历、爱好、年收入、购买意愿,将潜在用户分为 5 类,以定性为基础结合定量方法分析了潜在客户价值。

通过文本挖掘以及问卷调查的结果,我们得到以下主要结论:第一,智能手环已具备一定的市场认知率,从认知到实际购买的转化率较低,功能实用性是集

中关注点。第二，消费者对智能手环总体的满意度差强人意，天猫和京东两个电商网站的消费者评价数据中 19.9%的消费者购买的智能手环给予了高度满意的评价，10.7%的用户做出了负面消极的评价；实地进行的问卷调查，我们进一步发现，约 61%的用户对购买的智能手环整体较为满意。第三，易用人群特征明确，通过实地调查后的建模分析，我们发现智能手环目前的消费者主要是有健身需求的男性，对健康检测有需求的中老年消费者市场有待开发。第四，接近 45%潜在用户有较大可能成为智能手环用户。针对调查结果我们提出了相应的营销与产品发展策略。

关键字：智能手环；不等概率三阶段抽样；文本挖掘；LDA 主题模型；logit 二元选择模型；聚类分析法

Abstract

With artificial intelligence technology keeping getting breakthroughs , a series of intelligent life hardware is stepping into the daily life of publics. Because of the Entry-level price of smart bracelets and the standardization of its program , the market penetration of smart wearable art equipment lead the way, far more than smart watches, smart glasses and other products. As a new product, smart bracelet is now in the exploratory stage of development, so how consumers evaluate smart bracelet ? what kind of future development strategies should smart bracelet manufacturers design? For the first question, we capture the comment text in the electronic business platform and use large data mining methods to comprehend the concerns and satisfaction of consumer. As for the second question, we combined field survey with the usage experience of existing users and characteristics of non-users to make some advices about the future development strategy of smart bracelet.

Online shopping has become a general method of consumption in the era of mobile Internet. We first use text mining method ,capturing data about comment of MI- bracelet in Jingdong Mall and Tmall, and we got a total of 468,100 pieces. After data cleaning, we used the way of Chinese word segmentation and built a semantic network with LDA topic model algorithms, achieving judging the emotional tendency of those text comment data and mining hidden information, so that we can analyze the Internet electricity suppliers Reviews data effectively and accurately ,besides finding the focus points and complain about MI-bracelet, which can help us make a preliminary assessment of advantages and disadvantages of this product .Then provide direction and ideas for the further study.

We conducted a pilot survey and a formal survey. In the formal investigation, we treat whole community residents in Wuhan as respondents, using stratified sampling and probability sampling mode from three-stage sampling, and we divided Wuhan into 2 different parts—— the central area and urban area .In the three stage sampling ,we first used PPS sampling method to sample administrative of Wuhan, then taken further communities from selected administrative regions. Finally, the system sampling method was used in the community blocks to sample permanent residents, ensuring that the probability of the final unit selected samples are same. items of questionnaire had passed several statistic test, which means the structure of the questionnaire was

reasonable. Formal survey data are also tested to be reliable, and valid , besides the questionnaire passed run test and independence test as well., so the survey data are authentic.

Based on the Results of the survey, we made corresponding descriptive statistical analysis which reviewed the users and those non-users basic information respectively. With analysis of LOGIT binary choice model ,we concluded that sex, age, education, interests, personality and exercise frequency are significant factors of whether to buy a smart bracelet. Then based on the regression results, we made quantitative analysis about gender, age, education, hobbies, income and purchasing intention of potential users by using clustering method which can divide potential users into five categories and identified potential users' value.

From the results of the text mining and the survey, we came to some conclusions: First, smart bracelets have had a certain market awareness, but the transformation rate from cognitive to the purchase is low, besides, practicality is a main attention. Secondly, the overall satisfaction of consumers is mediocre. 19.9% consumers of Tmall and Jingdong website have relatively high satisfaction , 10.7% of users made a negative evaluation; From the survey, we found that about 61% of the users satisfied with the products. Third, features of potential users is distinct. After analyzing, we found that main consumers of smart bracelets are men who would like to exercise, and the health monitor in the elderly market needs to be developed. Fourth, about 45% potential users have a greater chance to become a real users. And according to these results, we put forward the corresponding marketing and product development strategies.

Keywords: smart bracelet; three-stage PPS random sampling ; Chinese word segmentation; sentiment analysis; LDA topic model; Logit model; cluster analysis

目录

摘要.....	I
一、问题综述.....	- 1 -
(一) 研究背景.....	- 1 -
(二) 文献综述.....	- 3 -
二、研究思路与创新.....	- 4 -
(一) 研究思路.....	- 4 -
(二) 研究意义.....	- 6 -
(三) 特色与创新.....	- 7 -
三、基于电商平台评论数据的文本挖掘.....	- 9 -
(一) 评论数据采集.....	- 9 -
(二) 数据预处理.....	- 11 -
(三) 中文分词.....	- 13 -
(四) 基于情感词典的情感倾向性分析.....	- 12 -
(五) 基于 LDA 模型的主题分析.....	- 14 -
四、调查方案.....	- 17 -
(一) 调查目的.....	- 17 -
(二) 调查对象和调查单位.....	- 17 -
(三) 调查项目.....	- 17 -
(四) 调查方式与方法.....	- 18 -
(五) 抽样调查基本步骤.....	- 20 -
五、调查实施.....	- 23 -

(一) 调查组织工作.....	- 23 -
(二) 调查时间和工作期限.....	- 24 -
(三) 质量控制.....	- 24 -
(四) 抽样效果控制.....	- 25 -
六、试调查的数据初步分析.....	- 28 -
(一) 区分度分析.....	- 28 -
(二) 信度检验.....	- 28 -
(三) 效度检验.....	- 29 -
七、正式调查数据处理与检验.....	- 31 -
(一) 数据处理.....	- 31 -
(二) 数据分析方法.....	- 31 -
(三) 数据检验.....	- 32 -
八、武汉市智能手环市场及用户使用现状分析.....	- 35 -
(一) 智能手环市场现状.....	- 35 -
(二) 智能手环用户特征.....	- 37 -
(三) 智能手环使用现状.....	- 40 -
九、智能手环用户特征的模型分析.....	- 42 -
(一) 模型的选择.....	- 42 -
(二) 二元选择模型的建立.....	- 42 -
(三) 对智能手环用户群体特征的研究.....	- 43 -
(四) 模型结果分析.....	- 44 -
十、智能手环潜在用户挖掘.....	- 45 -

(一) 模型选择.....	- 45 -
(二) 模型运用与分析.....	- 45 -
十一、调查结论与产品发展策略.....	- 49 -
(一) 调查结论.....	- 49 -
(二) 产品发展策略.....	- 50 -
参考文献.....	- 53 -
附录一.....	- 55 -
附录二.....	- 56 -
附录三.....	- 59 -
附录四.....	- 66 -
致 谢.....	- 71 -

表格清单

表 1-1 各品牌智能手环的优缺点对照表.....	- 1 -
表 3-1 从天猫平台上抓取的数据结构示意图.....	- 10 -
表 3-2 从京东平台上抓取的数据结构示意图.....	- 10 -
表 3-3 LDA 差评主题分析结果.....	- 15 -
表 3-4 LDA 好评主题分析结果.....	- 15 -
表 4-1 分层抽样样本比重表.....	- 18 -
表 4-2 中心城区代码法抽样表.....	- 19 -
表 4-3 随机数表抽样.....	- 19 -
表 4-4 居民社区抽样框.....	- 21 -
表 5-1 调查人员记录表.....	- 23 -
表 5-2 抽样效果计算表.....	- 27 -
表 6-1 信度分析表.....	- 29 -
表 6-2 单项与总和相关效度分析.....	- 29 -
表 7-1 单项与总和相关效度分析.....	- 33 -
表 7-2 用户选择偏好与年龄阶段列联表.....	- 34 -
表 8-1 智能手环用户满意度评价中位数分布.....	- 41 -
表 9-1 logit 模型最终回归结果（筛除后）.....	- 43 -
表 10-1 潜在客户人物指标含义表.....	- 46 -
表 10-2 潜在客户类型聚类中心表.....	- 47 -

图片清单

图 2-1 研究思路图..... - 5 -

图 2-2 文本挖掘流程图..... - 6 -

图 3-1 天猫商城评论界面示意图..... - 9 -

图 3-2 京东商城评论界面示意图..... - 9 -

图 3-3 词云统计图..... - 14 -

图 3-4 语义网络图..... - 12 -

图 3-5 文本情感分析结果..... - 13 -

图 5-1 调查工作流程图..... - 24 -

图 8-1 智能手环市场认知率与使用率..... - 35 -

图 8-2 智能手环认知渠道..... - 35 -

图 8-3 智能手环市场认知率与使用率..... - 36 -

图 8-4 智能手环来源..... - 36 -

图 8-5 智能手环用户性别分布..... - 37 -

图 8-6 智能手环用户年龄分布..... - 37 -

图 8-7 智能手环用户学历分布..... - 38 -

图 8-8 智能手环用户职业分布..... - 38 -

图 8-9 智能手环用户年收入分布..... - 39 -

图 8-11 智能手环用户流失情况..... - 40 -

图 10-1 用户关注功能分布表..... - 46 -

图 10-2 特征分析图..... - 47 -

一、问题综述

（一）研究背景

1.智能可穿戴产品简介

随着人工智能技术的不断取得新的突破，一系列智能生活硬件逐步走进普通大众的日常生活中。智能硬件是通过软硬件结合的方式，对传统设备进行改造，使之具备连接的能力，实现互联网服务的加载，形成“云+端”的典型架构，具备大数据等附加价值。

中国信息通信研究院 2015 年 12 月发布的《可穿戴设备研究报告》显示，2015 年，中国智能可穿戴设备市场规模为 125.8 亿元，增速高达 471.8%。目前，中国可穿戴智能设备市场仍处于探索期。从 2016 年开始，部分市场领域的巨大潜力将开始释放，可穿戴市场将正式进入启动期，预计 2016 年可穿戴市场规模将达到 200 亿元。

2.智能手环品牌格局

据互联网消费调研中心(ZDC)发布的《中国智能穿戴设备市场研究报告》显示，截至 2015 年年底，智能手环市场占有率的分别为：小米 42%、三星 7%、Jawbone6%、NIKE6%、咕咚 6%、索尼 4%。各大科技公司纷纷推出智能手环的背景下，小米手环凭借成功的低价策略，呈现压倒式优势，几乎占据了一半的市场份额，而其他的品牌的市场份额差距不大。我们通过浏览产品官网、ZEALER 等科技网站总结出以下各品牌旗下智能手环的优点和缺点：

表 1-1 各品牌智能手环的优缺点对照表

品牌	小米	三星	Jawbone
主要优缺点	<ul style="list-style-type: none">•功能较多•低价	<ul style="list-style-type: none">•弯曲柔性屏幕•显示手机消息•价格较高	<ul style="list-style-type: none">•饮食追踪
品牌	咕咚	索尼	NIKE
主要优缺点	<ul style="list-style-type: none">•基于百度云开发•硬件配置较差	<ul style="list-style-type: none">•内置麦克风和扬声器•可控制摄像头•配对的 APP 功能丰富	<ul style="list-style-type: none">•主打运动•佩戴较轻便•续航能力差•功能单薄

3.智能手环发展机遇与面临的挑战

(1) 发展机遇

在 2015 年，国内出台了首个智能可穿戴设备行业标准，促进该行业的正规化，包括《可穿戴产品分类与标识》、《可穿戴产品数据规范》、《可穿戴产品应用服务框架要求》。这三项标准对于智能穿戴设备的发展具有正向激励作用。

在人们越来越关注健康的大环境下，能提供运动数据分析和监测的运动手环很有市场；在饱受慢性病侵袭的老年人中，相比于繁重的医疗设备轻巧的健康监测手环很有市场；对于对世界充满好奇又缺乏危险意识的小朋友，一款颜色鲜艳的定位手环很有市场；面对商场超市排长龙的结账队伍，快速高效的支付手环很有市场，但是这一切需要一个完整的市场体系来支撑。数据记录的真实准确、食物摄入的精准分析、睡眠质量的监测提示，GPS 定位的正确与否，快捷支付的账户安全等问题将成为制约智能手环产品未来的难题。

纵观目前智能手环市场的，作为一个较新兴的产品，智能手环在可穿戴领域的关注度非常高，并且销量也在逐步上涨，但由于该产品还处于起步时期且并不是生活必需品，其不可能像智能手机一样改变我们的生活习惯，不过随着整个行业的高速发展，智能手环终将会出现革命性产品来改变整个行业。

(2) 面临的挑战

第一，关键技术有待突破。国内针对智能可穿戴设备的定制化芯片和操作系统产品成熟度仍待提高，尤其是操作系统差距较大。同时，国内传感器主要满足中低端市场需求，缺乏面向可穿戴设备的技术含量和产品附加值更高的中高端产品。

第二，产品应用服务仍需强化。产品同质化竞争严重，多以实现运动健康管理功能为主，在实际应用中和竞争对手没有太大区别，只是监测的数据类型和活动项目不同而已。产品的用户体验需进一步提升，由于对数据监测精度不够，在一定程度上制约了应用的发展。

第三，安全与隐私保护有待加强。智能可穿戴设备对个人数据的收集能力较其他智能设备更为强大，如个人信息、生理状态、日常活动、周边环境等，个人隐私泄露的危险大大增加。

第四，智能穿戴设备独立性问题。大多数的智能手环不具备单独工作的能力，需要连接手机工作，智能手环在手机没电的时候几乎丧失所有功能。其次长时间通过蓝牙连接智能手机，令智能手机的待机时间大大缩短。

综上，智能手环行业现在发展正处于承前启后的发展摸索阶段，要在未来格局中占有一席之地就必须对行业现状和行业发展方向有清醒的认识。在手环自身的应用软件的支撑下，根据消费者的实际需求不断改善用户体验。

（二）文献综述

由于智能设备在我国是近年来才兴起的一个新兴概念，国内关于智能穿戴设备的研究理论还不是特别完善。刘思言^{3M6}指出可穿戴设备作为物联网和移动互联网融合的重要体现，器产品形式丰富多彩，主要分为四类产品：网络延伸类、独立应用类、医疗健康类和控制娱乐类。业界对于可穿戴智能设备即将引领终端市场的大趋势，持一致认同态度。

中国调研网^{3m6}基于多年来对客户需求的深入了解，全面系统地研究了智能穿戴行业市场现状及发展前景，预计未来可穿戴设备行业的投资亮点将集中在以下三方面：其一，健康、运动及医疗产品；其二，专注男性消费的相关产品；其三，专注于解决电池问题的相关企业。

在对智能穿戴设备的发展前景方面，姜梦静^{3N8}指出，智能穿戴设备真正将人体作为大数据时代的入口，是人的智能化眼神，在智能网络的辅助下，人们可以更好地感知外部与自身的信息，实现无缝交流，智能穿带设备是未来智能设备的代表。

而对于智能穿戴设备在健康领域的发展，赵亦俊、张涛^{3P8}指出，穿戴性、智能化、专业化是智能穿戴设备的特点。健康监测、智能管理、健康产业是智能穿戴设备在健康领域的主要方面。同时，智能穿戴设备在健康领域发展中也有一系列问题。首先是产品易用性，包括产品的续航能力以及物理连接。再者是产品的专业性及数据标准相关的问题。智能穿戴设备只有解决了这些方面的问题，才能在健康领域有更大的发展。

李红岩、段莹^{3p8}等人指出，智能手环作为一种简单方便的可穿戴设备，对于大样本人群睡眠状况的研究可能做出贡献。智能手环可以用于失眠、睡眠呼吸障碍、周期性肢动障碍、昼夜节律失调性睡眠障碍等睡眠障碍的评估。

^{3M6}刘思言. 可穿戴智能设备引领未来终端市场 诸多关键技术仍待突破[J]. 世界电信,2013,12:38-42.

^{3m6} 2016 年中国可穿戴设备现状调研及市场前景走势分析报告.

^{3N8}姜梦静. 关于智能穿戴设备发展前景的思考与建议[J]. 移动通信,2014,17:64-68

^{3P8}赵亦俊,张涛. 智能穿戴在健康领域发展的现况分析[J]. 中国卫生信息管理杂志,2015,04:354-358.

^{3p8}李红岩,段莹,卢烨,常远,佟雅婧,孙瑶,孙书臣. 智能手环的应用评价[J]. 世界睡眠医学杂志,2014,06:341-344.

二、研究思路与创新

（一）研究思路

为了解小米智能手环用户使用情况与产品未来的改善方向，本文制定的研究思路如下：

首先，考虑到在当今的移动互联网时代，通过电商平台进行网络购物已成为人们生活中一种普遍的消费方式。同时，有一定比率的消费者会在购买商品后留下自己对产品的主观评论，这些信息不仅包含了消费者对商品的情感倾向，还有消费者对商品最直接的使用评价。本文拟通过爬取天猫平台和京东平台上，小米手环的全部可抓取评论数据，使用中文分词、构建语义网络与 LDA 主题模型等算法和模型，实现对文本评论数据的情感倾向性判断以及所隐藏信息的挖掘与分析，以期有效、准确的从互联网电商评论数据中分析出消费者对小米手环的好评度、关注点以及抱怨点，从而对小米手环的功能、优缺点进行初步的评价。

其次，从目前的智能手环市场来看，在消费者购买及购买后的使用过程中，消费者对该品牌的产品和所提供服务的满意度决定了对品牌的忠诚度，影响其是否再次购买该品牌产品及推荐率。因此我们针对文本挖掘得到的结果，在此基础上设计了调查问卷，在武汉市范围内开展线下调查，从基本情况、用户使用现状、客户满意度等方面设计了度量表。另一方面设计了可以提取出使用群体和潜在客户相关特征的有关问题，以深入分析客户群体的购买动机和购买行为，便于提取有针对性的营销推广策略。

我们根据研究目的设计了相应的调查问卷，根据分析设计的指标设计调查问卷，从调查单位基本信息，智能手环发展现状，客户满意度，潜在客户特征设计了度量表。考虑数据搜集的科学性，我们采用了分层抽样与不等概率三阶段抽样相结合，使总体中的单位科学地区分开来，并且每个单位都有一个非零概率被抽中。

在搜集到数据之后，我们对数据进行编码和录入，对结果进行检验。我们对智能手环使用率及认知率现状，潜在客户特征，客户群体特征，客户群体满意度等进行了描述性分析。我们运用 logit 二元选择模型分析智能手环客户特征进行分析，再利用 K-means 聚类分析法进行对客户群体分级，发掘潜在在客户特征。

最后，本文将综合评论文本的挖掘结果和实地调查的调研结果，结合市场营销学理论，根据小米手环用户的使用反馈和满意度，针对产品目标客户群体和潜在客户群体，提出关于小米手环产品未来研发改进方向以及营销策略的可取建议。

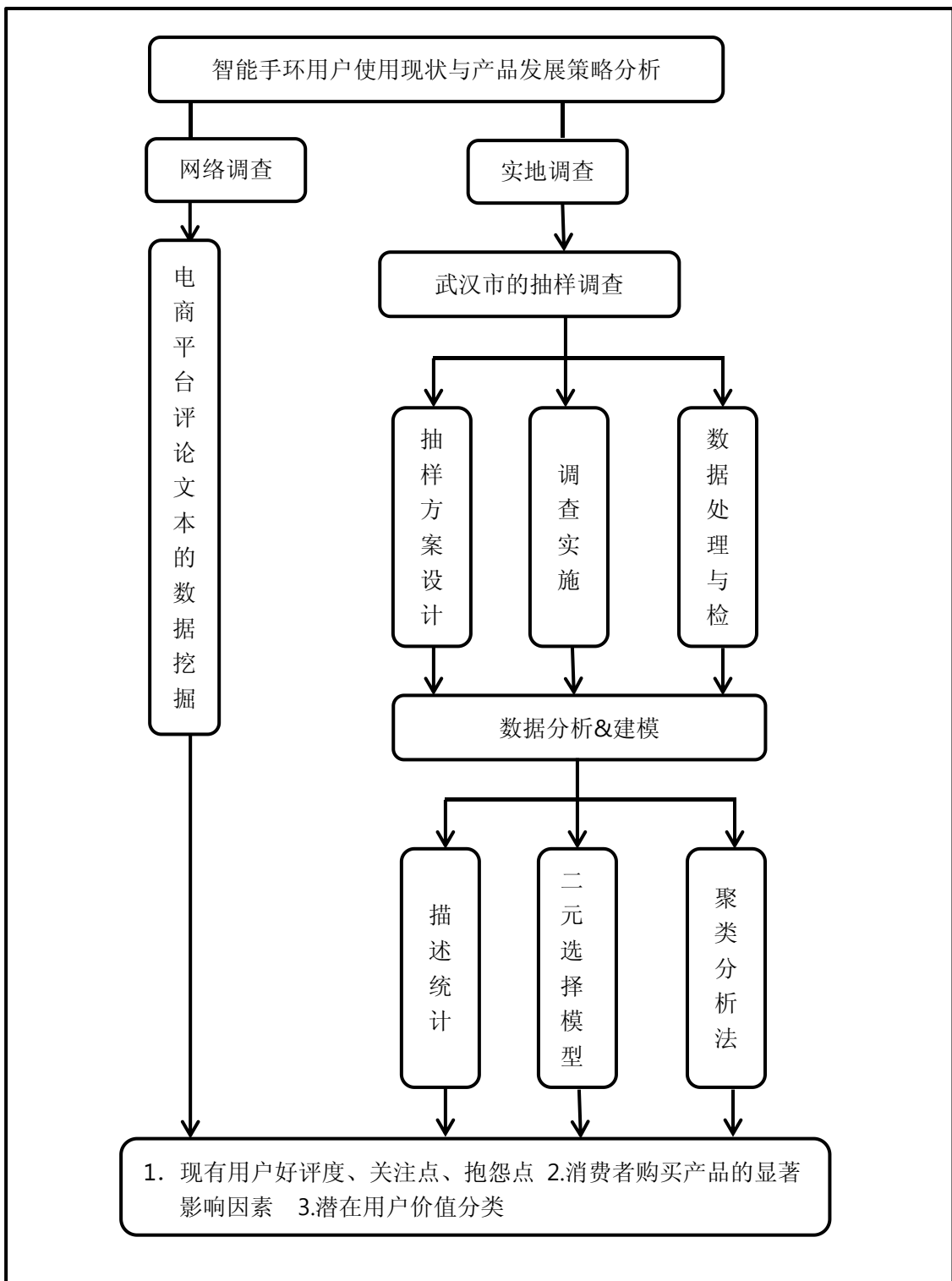


图 2-1 研究思路图

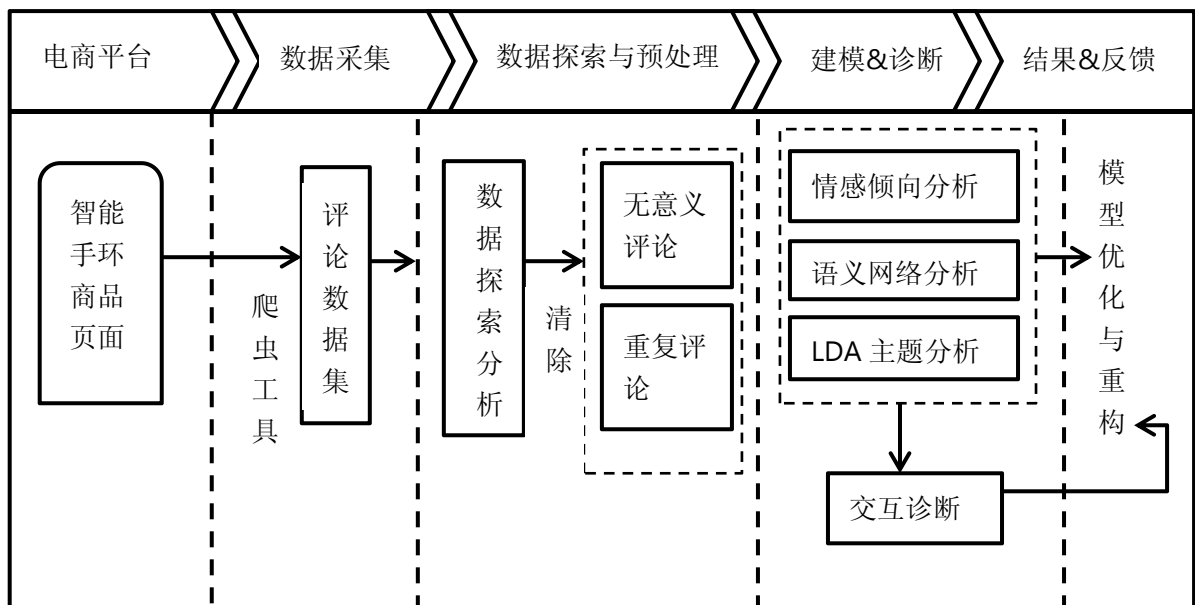


图 2-2 文本挖掘流程图

(二) 研究意义

1. 基于大数据分析智能手环用户使用情况

通过电商平台网络购物已成为人们生活中一种普遍的消费方式，有一定比例的消费者在购买商品后留下自己对产品的主观感受，这些非结构化与半结构化评价数据以爆炸性速度增长。从京东和淘宝两个网络电商平台抓取销售和评价数据，使用中文分词、构建语义网络与 LDA 主题模型等文本挖掘方法，可以挖掘出消费者对商品的好评度、关注点以及抱怨点，从而对智能手环的功能、优缺点进行初步的评价。大数据挖掘的方法不仅让我们在短时间内获得大量信息，同时让我们对消费者的使用情况有一个整体感知，更有针对性的设计智能手环用户填写的调查问卷，为后续的实地调查做铺垫。

2. 通过抽样调查了解智能手环认知度及使用现状

通过分层抽样与三阶段抽样相结合的概率抽样调查方式，了解智能手环使用率、认知率现状，对产品的满意度等产品使用情况。

3. 潜在客户分级，挖掘最有价值的潜在客户

运用 logit 二元选择模型分析影响消费者购买智能手环的显著影响因素，基于回归结果，利用 K-means 聚类分析法对潜在客户群体分级，发掘潜在客户特征。通过这些工作，可以为智能手环厂商设计有针对性的宣传和营销策略，帮助

提高产品的销售量。

（三）特色与创新

1.基于大数据的文本挖掘

通过抓取京东和淘宝两个网络电商平台所有可获取的小米手环评价数据，通过中文分词、构建语义网络与 LDA 主题模型等文本挖掘方法，从消费者购买评价数据中分析出消费者对小米手环的好评度、关注点以及抱怨点，从而对产品的功能、优缺点进行初步的评价。该方法在短时间内获得了实地调查无法完成的数据信息，可以提供实时的用户使用反馈信息。

2.实地电子问卷调研提高数据质量

传统调研中纸质问卷是最常见的调研形式。我们的调研过程中采用“云调研电子问卷”，以智能客户端为展现形式，在平板电脑上运行电子问卷页面，调研过程中只需要将平板电脑给被调查者填写，替代传统纸质问卷形式。这种调研方式带来以下优点：电子问卷内设跳问控制，防止被调查者出现漏答的情况；省去了数据录入环节，避免出现录入错误，缩短了项目周期和时间；省去纸质问卷的打印成本和保存成本。

3.调查对象新颖

我们选择了时下关注度较高的智能穿戴设备——智能手环作为研究的对象，具有时效性、科学性和创新性。作为近几年才出现的产品，智能手环这一产品还在发展更迭中，课题具有一定的研究价值。

4.调查分析形式多样

除了运用描述性统计分析，对客户群体和非客户群体进行定性分析外。我们还借助 Logit 二元选择模型分析影响消费者购买手环的显著影响因素，借助聚类分析法挖掘潜在客户，并对潜在客户进行分级，确定最有价值的潜在客户。

5.研究思路逻辑严密

本课题分为网络数据挖掘和实地问卷调查两部分，结合两部分的数据，获取智能手环用户对产品的使用体验和反馈，提取出用户对于该产品的关注点和抱怨点，为后续调查提供数据支撑并优化调查问卷的内容。对武汉市民进行抽样调查后，

了解目前智能手环认知率和使用情况，得到手环的用户特征，挖掘出最有价值的潜在客户。最后两部分的分析结果提出智能手环的未来发展策略。

三、基于电商平台评论数据的文本挖掘

通过电商平台网络购物已成为人们生活中一种普遍的消费方式，有一定比例的消费者会在购买商品后留下自己对产品的主观感受，这些评论不仅包含了消费者对商品的情感倾向，还有消费者对商品最直接的使用反馈。充分利用好这些信息，一方面可以帮助消费者了解产品的优缺点，优化购买决策；另一方面可以帮助企业了解消费者的心声，规划产品未来发展方向，定制营销策略，提升企业在行业内的竞争力。这些富有价值的信息往往杂乱的分布在海量的文本数据中，通过人工操作进行分析显然费时费力。运用自然语言处理技术和机器学习的方法处理非结构化或半结构化的文本数据成为大数据时代的一个关注热点。

基于上述基本情况和实际的背景，本节拟通过抓取电商平台上小米手环的评论数据，使用中文分词、构建语义网络与 LDA 主题模型^{3M8}等文本挖掘方法，实现对文本评论数据的情感倾向性判断以及所隐藏信息的挖掘与分析，以期有效、准确的从互联网电商评论数据中分析出消费者对小米手环的好评度、关注点以及抱怨点，从而对小米手环的功能、优缺点进行初步的评价。进而优化实地调查环节中的调查问卷，改进实地调查方案。

（一）评论数据采集

B2C 网站商品页面上用户评价的数量越多，说明商品的销量越高。天猫、京东的商品评论界面如下：

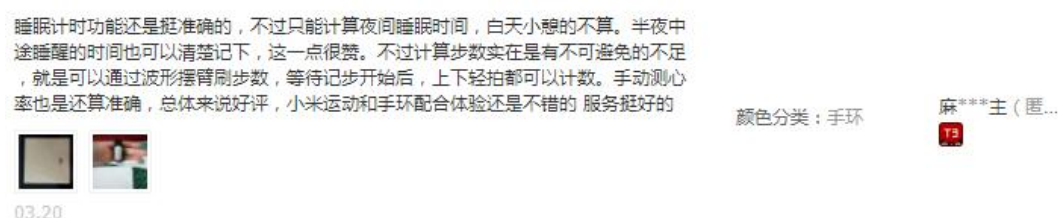


图 3-1 天猫商城评论界面示意图



图 3-2 京东商城评论界面示意图

^{3M8} 谢金星，基于 SVM 的中文微博情感分析的研究[D].2012

天猫商城用户评价的基本元素有：评论内容、用户名、商品名称和评论时间。京东商城界面上显示的内容相较天猫商城则更为丰富，增加了用户等级、用户所在省份、用户发表评论所使用的客户端等信息。

本文的数据采集工作是使用 R 完成的，编写好数据采集逻辑后运行程序，成功抓取到京东平台和天猫平台小米智能手环全部可抓取数据 441488 条和 26612 条。（注：天猫商城每个在售商品的评论只能显示前 99 页，也即每个商品最多只能抓取 1998 条评论数据）在数据抓取过程中，同时将非结构化的文本数据转化为可检索的结构化数据，用表格形式储存，结构示意图如下：

表 3-1 从天猫平台上抓取的数据结构示意图

评价内容	评价时间	商品种类	用户名
配合 iPhone 用了几天,总体可以。就是不知为什么手环记录的步数会比苹果健康少了将近一半。另外,想要用震动提醒的果粉们,如果你有越狱,强力推荐 Miband Alert。亲测很有效,现在设置了微信 QQ 和短信提醒,比较钟爱这个功能,所以好评。睡眠监测也挺准的,希望步数那里可以更准确。	03.01	颜色分类: 石墨黑	v***r (匿名)
用了 20 天才来客观评价: 1.电池很经用, 还没提示充电呢。2. 智能闹钟很方便, 不会吵醒别人。3.计步器不太准, 手动的时候会被记录为走路运动。总的来说很不错啦, 性价比非常高, 下次再买我会选择升级版的, 继续支持!	02.13	颜色分类: 石墨黑	秋***水 (匿名)
本来打算明天在取, 可是实在等不及还是坐车去取快递了, 收到之后就使用了, 敏感度不错, 戴起来很舒服, 黑色很精致, 我很喜欢, 朋友买过之后我看着好也买了, 夏天快到了, 到了锻炼身体的日子, 不错, 和苹果手机很相配, 赞一个	02.22	颜色分类: 石墨黑	一***y (匿名)

表 3-2 从京东平台上抓取的数据结构示意图

评价时间	商品类别	评价内容	用户名称	用户等级	用户省份
2016-03-02 20:57	新款光感心率手环	不错可以。就是太难弄上去了	维***堡	金牌会员	
2015-12-25 11:19	黑色原封(白色 LED 提示灯)	不止一次购买.爸妈退休天天运动比赛	j***q	银牌会员	北京
2016-02-20 22:34	黑色原封(白色 LED 提示灯)	不错	j***w	铜牌会员	

（二）数据预处理

在数据挖掘中，海量的原始数据中存在着大量不完整、不一致、有异常的数据，严重影响到数据挖掘建模的执行效率，甚至可能导致挖掘结果的偏差。对于电商平台的评论数据来说，由于存在商家雇佣水军采取作弊行为人为刷高销量的现象，进行原始数据的清洗尤为重要。在前人研究的基础上，有许多文本去重算法，大多都是先通过计算文本之间的相似度，再以此为基础进行去重，包括编辑距离、Simhash 算法去重^{3M8}等，但算法的缺陷以及漏洞较为严重，特别是容易将情感倾向相反但表达相近的文本视为重复，出现错删现象。因此本文采用相对简单的文本比较方法，按如下规则进行处理，尽量留存更多的语料。

本次预处理的主要工作是清除水军发表的重复数据和消费者随意发表的无意义数据，处理规则如下：

1.评论文本内容与用户名均相同的数据视为水军发表的，予以删除。例如：

2015-07-01 19:24	一次性购买 150 个，店家赚的最多的估计就是运费。东东还可以，送 叮客户，可以刻字。	***u
2015-07-01 19:24	一次性购买 150 个，店家赚的最多的估计就是运费。东东还可以，送 叮客户，可以刻字。	***u
2015-07-01 19:24	一次性购买 150 个，店家赚的最多的估计就是运费。东东还可以，送 叮客户，可以刻字。	***u

2.评论内容完全为英文字母、数字和标点符号的，视为随意发表的评论，予以删除。例如：

2015-07-28 21:52	0000000000	x***4
2015-11-12 21:11	。 。 。 。 。 。 。	q***i
2015-12-24 21:52	asdsadasdsadasdasd	jd_132486ijt

3.天猫平台为避免一些客户长时间不发表评论，系统会自动替客户做出评论，这些数据没有任何分析价值，大量重复出现，视为无意义的评论，予以删除。如：

2016-01-18 11:22	初次评价:	—***2（匿名）
2016-03-01 19:10	此用户没有填写评论！	车***2（匿名）

经过预处理后，共得到评论数据 22605 条。无效的数据达 445495 条，文本有效率仅为 4.83%。我们分别统计了天猫平台和京东平台的文本有效率，天猫和京东上小米手环的评论有效率分别为 57.4%和 1.6%。进一步分析后发现，造成

^{3M8} 张丽，在线评论的客户参与动机与评论有效性研究[D].2011

京东平台文本有效率如此低下的原因是其京东自营店，392795 中仅为 549 为有效评论，水军刷高销量的现象十分严重。

(三) 中文分词

不同于在英文中，每个单词之间都有空格来隔开。在中文里，只有字、句和段落能够通过明显的分界符进行快速的划界，对于“词”和“词组”来说，其边界模糊没有一个形式上的分界符。而词是最小的，能够独立运用的有意义的语言单位，文本分词的工作也就是将连续的字序列按照一定的规范重新组合成词序列。本文的分词工作是使用 R 软件的 **jiebaR** 包来完成的，基本实现原理及步骤如下：

1.加载字典，形成 Trie 树，也即字典树。JiebaR 中的 Trie 树是基于作者根据人民日报语料等资源训练得出的，包含两万多条词语。目的是利用字符串的公共前缀来减少查询时间，最大限度减少无谓字符串比较。

2.加载待分词的文本,基于前步生成的 **Trie** 对文本生成有向无环图 **DAG**,也即对每一条评论文本从第一个字符开始,生成所有可能成词情况的 **DAG**。

3.使用动态规划查找最大概率路径。对每条文本从右往左反向计算每个 DAG 下的词频，找出使得词频最大的分词组合。

4.对于未登录词，也即 `jiebaR` 中没有记录的词，采用了基于汉字成词能力的 **HMM**(隐马尔科夫)模型，将中文词汇按照 **B** (开始位置)、**E** (结束位置)、**M** (中间位置) 和 **S** (单独成词位置) 四个状态来标记。使用 **viterbi** 算法，找到一个使得概率最大的最佳 **BEMS** 序列。

将样本数据按照上述步骤，分词后，提取出高频词画出词云：



图 3-3 词云统计图

上图中，词频越高，词的字号就越大。观察上图我们发现，通过电商平台购买小米手环的消费者主要的关注点是小米手环的心率和睡眠监控功能。可以认为，真正吸引广大消费者的还是产品自身的功能。除此之外，从这幅图上我们还可以发现不少消极词汇，如：“失望”、“垃圾”、“不准”、“玩玩”等高频词。尽管它们的频率在整体上排名不算靠前，但频率最高的词语中多是“手机”、“功能”等中性的名词，对于消极词汇的出现更应格外关注。

由于机器不能直接识别语句的整体结构思想，在经过上一步的分词工作后，为了挖掘出各个词之间的联系和语义结构，得出单独的词语无法清晰表达的内容，基于分词后的文本，抽取其中的高频词和行特征词，得到一个共现矩阵，以此建立语义网络图。

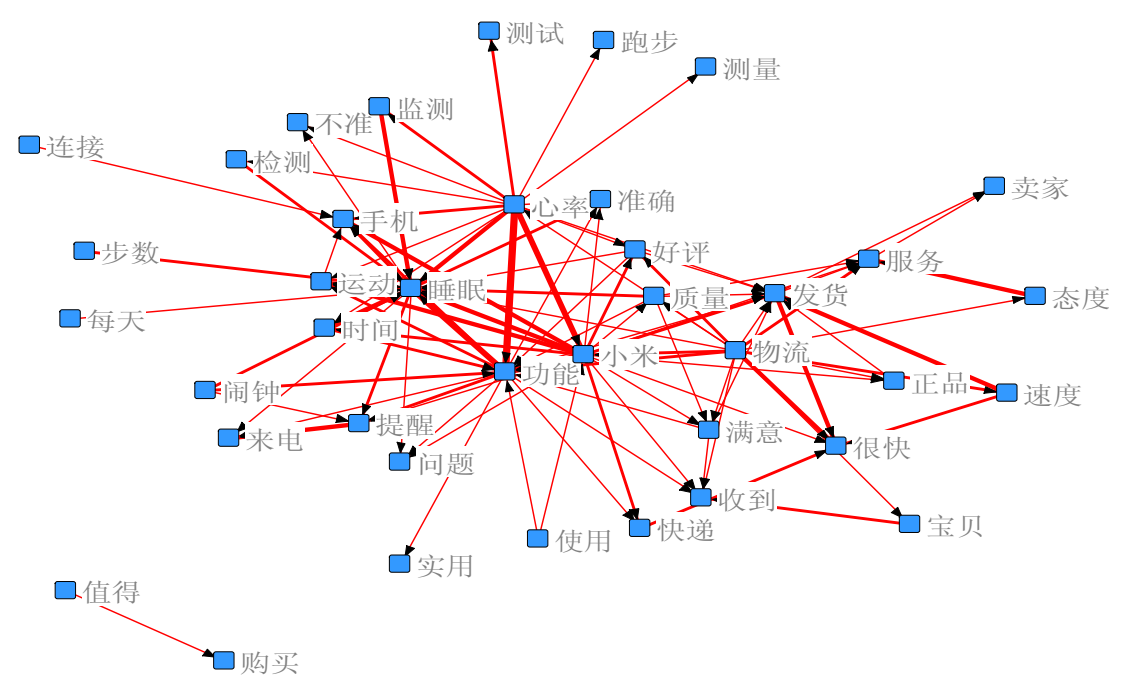


图 3-4 语义网络图

在上图的左边主要是跑步、睡眠、提醒等与手环功能密切相关的内容，右边主要是消费者对电商平台整个购买流程的评价、包括物流、发货、赠品、卖家、态度等内容。由于发货、物流等和手环自身设计无直接联系，这里我们主要关注上图左边及中间的词语。在抓取的评价文本中，消费者对手环的质量和功能较为满意，最为关注的功能是睡眠监测、心率测量以及运动记录步数、充当闹钟等提醒功能，有消费者认为手环测量数据准确、也有消费者认为数据不准，整体而言，大部分消费者认为小米手环是一个值得购买的产品。

（三）基于情感词典的情感倾向性分析

情感倾向分析的通用方法都是对大规模语料集进行统计分析、预先对有代表性的词语采用人工标注方法选为基准词，然后对候选词作语义相似度计算来获取新的情感词，从而扩展情感词典的覆盖面。本文基于已有资源，使用中国知网发

布的“情感分析用语集”，作为判断文本情感倾向、情感极性的参照，进行情感倾向性分析^{3M8}的方法如下述公式：

$$P(d_i) = \begin{cases} 1 & (sumPos(d_i) > SumNeg(d_i)) \\ -1 & (sumPos(d_i) < SumNeg(d_i)) \\ 0 & (sumPos(d_i) = SumNeg(d_i)) \end{cases} \quad (3.1)$$

上式中 $P(d_i)$ 表示文本 d_i 的情感极性值， $sumPos(d_i)$ 表示文本 d_i 中包含的正面情感词个数， $SumNeg(d_i)$ 表示文本 d_i 中包含的负面情感词个数。对经过预处理的文本评论进行判别分析后，结论如下：

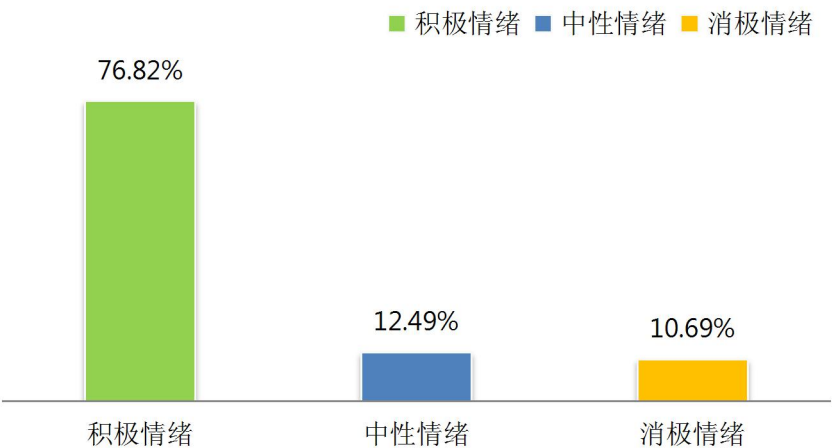


图 3-5 文本情感分析结果

天猫平台上商品是没有好评、差评的区分的，而京东平台上每个销售小米手环的页面其好评度都在93%以上。经过统计得到样本中正面情绪占比76.82%。产生这种差异的原因，一方面是店家雇佣水军刷高了好评度，另一方面是多数消费者给出的好评是在综合多个因素后给出的。

（五）基于 LDA 模型的主题分析

潜在狄利克雷分配（Latent Dirichlet Allocation ,LDA）是用在一系列文档中发现抽象主题的一种统计模型，认为每一篇文档的每一个词都是通过“一定的概率选择了某个主题，并从这个主题中以一定的概率选择了某个词语”。LDA 模型包含文档、主题、词三层结构，能够有效地对文本进行建模，挖掘数据集中的潜在主题，进而分析数据集的集中关注点及其相关特征词。

LDA 模型^{3M8}假定每篇评论由各个主题按一定比例随即混合而成，混合比例服从多项分布，记为：

^{3M8} 乐国安，情感分析技术及应用[D],2015
^{3M8} Ben Fry,张羽（译），数据可视化[M],电子工业出版社.2009

$$Z|\theta = \text{Multinomial}(\theta) \quad (3.2)$$

而每个主题由词汇表中的各个词语按一定比例混合而成，混合比例也服从多项分布，记为：

$$W|\theta, \phi = \text{Multinomial}(\theta) \quad (3.3)$$

在评论 d_j 条件下生成词 w_i 的概率表示为：

$$P(w_i|d_j) = \sum_{s=1}^k P((w_i|z = s) \times P((z = s|d_j) \quad (3.4)$$

式中， $P((z = s|d_j)$ 表示词 w_i 属于第 s 个主题的概率， $P((w_i|z = s)$ 表示第 s 个主题在评论 d_j 中的概率。LDA 模型对参数的近似估计使用马尔可夫链蒙特卡洛（MCMC）算法中的一个特例 Gibbs 抽样。本文处理时将狄利克雷函数的先验函数 α 和 β 设置为经验值： $\alpha = 50/K$, $\beta = 0.1$, 主题个数 K 采用统计语言模型中常用的评价标准困惑度来选取，即 $K = 50$ 。

为了提高主题分析在不同情感倾向下热门关注点反映情况的精确度，本章在语义网络情感分类结果的基础上，对不同情感倾向下的潜在主题分别进行挖掘分析，从而得到不同情感倾向下用户对小米手环不同方面的反映情况。

经过 LDA 主题分析后，评论文本被聚成 3 个主题，每个主题下生成 10 个最有可能出现的词语以及相应的概率。

表 3-3 LDA 差评主题分析结果

主题 1	主题 2	主题 3
一般	不错	速度
这样	还行	自己
不知	就是	有点
没用	玩玩	以为
没 想到	东西	发货 慢
不准	不过	售后
电池	小米	天
以前	感觉	不会
刻字	还是	费劲
效果	就那样	必须

表 3-4 LDA 好评主题分析结果

主题 1	主题 2	主题 3
点赞	测量	不错
达标	家人	的
腕带	喜欢	就 到货
绑定	满意	很好
计步 准确	看	东西
方便	手腕	非常
太 喜欢	据说	小米
公里	材质	物流
超值	不错	快递 快
以前	感觉	不会

根据小米手环差评的 3 个潜在主题的特征词提取，主题 1 中的高频特征词，即一般、这样、不知、没用、没想到、不准、电池、以前、刻字、效果，主要反映小米手环的功能不顾准确，没有达到部分消费者的期望值；主题 2 中的高频特征词，即热门关注点主要是不错、还行、就是、玩玩、东西、不过、小米、感觉、还是、就那样，主要反映出消费者认为小米手环就是随便玩玩，使用后感觉就那样，不够惊喜；主题 3 的高频特征词主要是速度、自己、有点、以为、发货慢、售后、天、不会、费劲、必须；主要反映出部分商家的发货慢、与店家的沟通不够顺畅等。

根据小米手环好评的 3 个潜在主题的特征词提取，主题 1 主要反映出大部分消费者认为小米手环性价比较高，使用较为方便、计步比较准确，主题 2 反映出小米手环的材质不错，对商品的整体比较满意，同时家人一词也在热门词中，这有可能是因为存在相当一部分比例的消费选择将手环作为礼物送给家人，主题 3 主要反映出消费者对店家使用的快递和物流较为满意。

四、调查方案

（一）调查目的

第一，搜集武汉市城镇常住居民对智能手环的使用情况，并搜集智能手环客户群体的基本信息，包括性别、年龄、职业、年收入等数据，分析使用智能手环的受众特征。

第二，搜集调查中使用智能手环的用户的购买动机及其相应的基本信息，研究智能手环各方面因素对用户选择的影响。

第三，搜集未使用智能手环的用户不使用智能手环的原因及其相应的基本信息，在此基础上利用数据挖掘技术对不同影响因子和潜在客户进行挖掘归类，以针对不同特征群体进行不同方式的推广。

第四，搜集智能手环用户对产品各方面的评价信息，包括智能手环需要改进的地方、产品外形、功能等关于用户体验各方面的评价，从而为智能手环这一产品的改善及未来发展方向提供一定的参考。

（二）调查对象和调查单位

武汉市是湖北省省会，中国中部地区的中心城市，也是全国重要的工业基地、科教基地和综合交通枢纽。在“十三五”时期，武汉市确定了战略性新兴产业发展战略，一直以来武汉对于新型产业和产品都是极具包容性和鼓励性的。

2015年武汉城市一卡通有限公司与上海金雅拓智能卡有限公司、武汉优讯通公司合作开发了既可以用于乘坐公交、地铁、轮渡，还能在自动贩卖机、电影院、大型超市等场所进行消费，同时可以进行武汉通余额查询、消费记录查询，并且可以通过手机进行空中充值的“武汉通智能手环”。同时该手环还具备运动、睡眠、膳食管理功能。由此可见，智能手环这种新兴电子产品在武汉市有一定的接受度和发展潜力。为了此次调查方便有效以及样本的统一，本次调查将调查范围限定在武汉地区。

调查对象：武汉市城镇常住居民总体

调查单位：武汉市每一个城镇常住居民

（三）调查项目

本次调查着眼于智能手环的客户群体和潜在客户的特征分析，关注用户体验，并发掘最有价值潜在客户。调查中包含有四个方面的内容：调查单位基本信息、

智能手环现有客户分析、智能手环潜在客户分析、智能手环改善分析。对于这四个层次，我们设计了相对应的调查项目以进行考量。

（四）调查方式与方法

由于智能手环这一商品普及度还不是很高，我们主要希望通过线下调查了解智能手环用户的满意度，并发掘潜在用户，进而探究其发展前景，因此我们需要调查的不仅仅是使用智能手环的用户，而是更加需要调查潜在用户群体对智能手环的了解及看法。所以我们不用预先知道智能手环使用者的地理分布特征而去专门抽取使用智能手环的用户，而是将全部武汉市的居民视为我们的调查对象。

为了保证调查的科学性，我们采用概率抽样的方式，使总体中的每个单位都有一个事先已知的非零概率被抽中，这种抽样遵循了随机原则，排除了主观随意性或目的性。并且为了减小抽样误差，我们首先进行分层抽样，又为了使每个调查单位的入样概率均相等，我们采取了三阶段不等概率 PPS 抽样。因此，本次调查采用分层抽样和三阶段抽样相结合的概率抽样调查方式，其中三阶段抽样分别是从武汉市中抽取行政区，再从行政区中抽取社区，最后从中抽取常住居民。抽样保证了前两阶段进行等概率抽样，使得各最终单元入选样本的概率相同。并且考虑了人力因素和问卷回收效果，我们主要采取在社区周边拦访调查的方式，并采取问卷调查法和文案调查法相结合的调查方法。

1. 调查方式

（1）分层抽样

首先运用分层抽样的方法将总体划分为两层。武汉市有 13 个辖区，其中江岸区、汉江区、硚口区、武昌区、洪山区、青山区 7 个为中心城区，东西湖区、蔡甸区、江汉区、黄陂区、新洲区、汉南 6 个为新城。因此我们将总体分为两个层次，即中心城区和新城。

表 4-1 分层抽样样本比重表

分层	所含社区数	所占比重
中心城区	884	0.6837
远城区	409	0.3163
总计	1293	1

（2）不等概率三阶段抽样

第一阶段的 PPS 抽样：在分得的两层的每一层内独立地进行三阶段抽样。以第一层为例，第一阶段的初级抽样单元的抽取采取概率比例规模抽样方法，该方法是放回的不等概率抽样，即每个行政区的入样概率是不等的，是与该区所含最终单元个数大小成正比的，即与该区的常住人口数成正比。根据每一层的财力、

物力、人力等多重现实因素的考虑下确定每一层的总体中抽取多少初级抽样单元，利用 PPS 法进行初级抽样单元的抽取时运用代码法进行实施。即赋予每个行政区与该辖区人口数相同的代码数，将代码数依次进行累加，利用计算机产生 5 个 1~634200 的随机数。随机数所属的代码范围对应的行政区入样，则这就构成了中心城区层的初级抽样单元，具体如表 2 所示。远城区层的初级抽样单元的选取和第一层的选取方法完全相同，最终共抽出 7 个区，按照比例，中心城区抽出 5 个区，远城区抽出 2 个区。

表 4-2 中心城区代码法抽样表

序号	行政区	常住居民户数	累计	代码范围	随机产生数	抽中	抽中再编码
1	江岸区	926879	926879	1~926879	201343	是	1
2	江汉区	713106	1639985	926880~1639985	1548907	是	2
3	硚口区	848349	2488334	1639986~2488334	201343	是	3
4	汉阳区	616730	3105064	2488335~3105064		否	
5	武昌区	1246814	4351878	3105065~4351878	3903962	是	4
6	洪山区	1477475	5829353	4351879~5829353	4522149	是	5
7	青山区	512647	6342000	5829354~6342000		否	

第二阶段的分层抽样：第一层的第二阶段为从第一层的初级抽样单元中抽取二级单元，即从每个被抽取的中心城区中抽取入样社区。第二阶段采用分层抽样的方式进行抽样，由于每层的人口数都是不同的，权数由每个行政区人数比例决定将第一阶段抽出的行政区的所有社区列进 EXCEL，乱序后，从 1 开始编号排序，基于实施的便利性以及调查成本的考虑，我们小组决定从中心城区抽取 10 个社区，远城区抽取 4 个社区展开调研。

具体在每层中使用简单随机抽样，利用 Excel 生成一个 0-9 十个数字组成的随机数表，从第 4 行第 7 列开始抽取，即设置(4,7)为起点，依次向右一共抽取 10 个样本。第二层的二级单元的取法也和第一层的第二阶段的抽取相同。

表 4-3 随机数表抽样

序号	行政区	所含社区数	入样号码	对应社区名
1	江岸区	136	115、133、006	同福、滨江、辅仁
2	江汉区	112	173	琴台
3	硚口区	129	374、268	额头湾、汉宜
4	武昌区	138	385、443、489	宝通寺、解放桥、珞军
5	洪山区	143	611	尤李、

第三阶段系统抽样：第三阶段是要从入样社区中抽取被访者，由于预调查过程中我们遇到高档小区无法进入、业主名册无法获取等困难，考虑了调查难度和人力资源限制，同时也为了在有限的时间内获取足够的样本，我们采取系统抽样的方式，在每个社区内拦访行人调查。在实施过程中，我们制定了如下调查规则：以社区居委会为中心，在其东南西北四个方位上按照实际环境分别固定四个发放调查问卷的地点；每隔五分钟拦访一位路过的行人邀请其填写调查问卷，若该位被访者拒绝调查则邀请此刻距离他最近的行人填写；每隔两个小时更换一次发放调查问卷的地点。

2. 调查方法

（1）文案调查

在大规模问卷调查前期，我们采取文案调查法，利用计算机检索，查阅国内外可穿戴设备的发展情况、智能手环使用情况、智能手环的主要品牌和功能、现状及相应文献，获取对本小组调研有一定指导意义的二手资料。

（2）问卷调查

问卷调查法是进行本次调查的主要调查方法，可以获得符合课题研究目的的一手数据。我们对每个社区按照分配好的最佳样本量在社区周边拦访调查，收集样本的信息。

（五）抽样调查基本步骤

1.编制抽样框

首先我们将总体分为中心城区和远城区两层，每层的所有行政区作为一级单元的抽样框，然后分别在这两层中独立使用三阶段抽样方法，第一阶段抽出行政区作为一级抽样单位。二级单元的抽样框是第一阶段入样的行政区的所有社区，第二阶段，通过分层抽样抽出社区作为二级抽样单位。第三阶段的三级单元的抽样框为各层中每个入样社区的所有居民。具体抽样框如表 4-4 所示。

2.设计调查问卷

针对本次问卷调查的目的，我们设计了相应的调查问卷。问卷主要分为五个方面第一方面是调查单位基本信息，通过对被调查者性别、年龄、就业情况、年收入、对于可穿戴设备的认知度等信息了解被调查者的基本特征，有利于我们对样本进行分类分析。第二方面是对于客户是否是智能手环用户的调查，以此区分出现有用户合潜在用户，并与第一部分中的基本信息相结合得出现有客户人群和潜在客户人群的特征。第三方面是对于用户使用的智能手环的价格和原因、满意度的分析。第四方面是非智能手环用户人群对智能手环的价格及功能预期的调查，与第一部分的基本信息想结合，可针对不同的人群进行不同的营销推广方式。

3.试抽样调查

在正式调查之前，先根据抽样方案，从抽样框中抽取小样本进行试调查，以发现各个方面可能存在的不完善之处，检验所设计的抽样方案及抽样框、调查表的科学性和可操作性。我们的试抽样调查范围仍是抽样框中的各级抽样单元，对入样的 14 个社区附近的居民发放共 140 份（每处各 10 份）问卷，共回收 140 份问卷，进行试抽样调查检验分析。根据回收的数据检测初始问卷的信度和效度，然后对检测效果较差的项目进行调整修改。

表 4-4 居民社区抽样框

总体 分层	一级单元 的抽样框	入样 行政区	二级单元的抽样 框	入样 社区	三级元 的抽样框
中心城 区	中心城区的所 有行政区	江岸区	江岸区的 所有社区	同福社区	同福社区所有居民
				滨江社区	滨江社区所以居民
				辅仁社区	辅仁社区所有居民
		江汉区	江汉区的所有社 区	琴台社区	琴台社区所有名居 民
		硚口区	硚口区的所有社 区	额头湾社 区	额头湾社区所有居 民
				汉宜社区	汉宜社区所有居民
		武昌区	武昌区的所有社 区	宝通寺社 区	宝通寺社区的所有 居民
				解放桥社 区	解放桥社区的所有 居民
				珞军社区	珞军社区的所有居 民
		洪山区	洪山区的所有社 区	尤李社区	尤李社区的所有居 民
远城区	远城区的所有 行政区	江夏区	江夏区的所有社 区	流芳社区	流芳社区的所有居 民
				长咀社区	长咀社区的有居民
		东西湖区	东西湖区的所有 社区	北湖区	北湖区的所有居民
				金口区	金口区的所有居民

4.样本容量的确定

根据预调查问卷填写的智能穿戴设备认知比例作为估计对象,因此关注总体认知比例的样本方差。修正前最佳样本量 n_0 的计算公式为 (4.1)

$$n_0 = \frac{t^2 PQ / d^2}{1 + \frac{1}{N} [\frac{t^2 PQ}{d^2} - 1]} \quad (4.1)$$

N 为总体数量,取置信度为 95%时的 t 值, $t=1.96$, p 为样本比例, d 为绝对允许误差, $d=0.04$,根据预调查结果 $P=0.52$, 由于在实践中, 如果 p 在 0.5 附近取值,可根据总体方差在 $p=0.5$ 时达到极大值来对样本量进行计算,故取 $p=0.5$ 。武汉常住人口总数 $N=1022$ 万, 则可以近似得出最佳样本量为:

$$n_0 = \frac{t^2 p(1-p)}{d^2} = \frac{1.96^2 \cdot 0.5 \cdot 0.5}{0.04^2} = 600$$

即 600 为以估计简单随机抽样比例 P 时的样本量为基础,在 95%的置信度下按抽样绝对误差不超过 4%的要求得到的最佳样本量。由于我们的抽样方案比较复杂,难以计算实际设计效应 $deff$ 。根据预调查和文献资料的结合,假设我们采取的多阶段抽样的设计效应为 1.8, 则应回收的有效样本量为 1080 份。

$$n = n_0 \cdot deff = 600 \cdot 1.8 = 1080$$

考虑到被抽中的居民中途放弃填写问卷,或其自身在调查公司等相关单位工作不宜作为样本等原因导致的样本无效问题,在征集指导老师的意见和经验后,我们假设无效比例为 20%, 则实际应调查的样本量为 1350 份。

$$n' = n \div (1 - 0.2) = 1350$$

按照比例新城区和中心城区的实际应调查的样本量分别为 923 和 427,根据调查力量的情况确定中心城区抽取的样本量为 930,远城区样本量为 428。中心城区入样的 10 个社区中,每个社区随机拦访 93 名行人,远城区抽中的 4 个社区,每个社区随机拦访 107 名行人。

5.正式抽样调查

本次共发放 1358 份问卷,有效问卷为 1120 份,有效问卷回收率 80.26%。其中琴台社区、同福社区、汉宜社区等 10 个社区有效问卷大于最佳样本量,我们采取截尾的方式删除部分有效问卷以控制问卷数量。对于未达到最佳样本量的社区,依据测算的最佳样本量进行问卷补发。在正式调查阶段结束后,我们仅需前往武昌区的宝通寺社区补发两份问卷,有效的避免了再次大规模前往各个社区进行问卷补发的工作,调查方案制定得比较合理。

五、调查实施

（一）调查组织工作

在熟悉了调查项目性质、目的、要求等方面内容之后，我们首先系统地学习调查过程中应掌握的调查知识以及访问技巧，提升调查人员素质。由于我们团队的四个队员都是女生，基于安全性和任务完成效率的考量，我们邀请了各自的亲人朋友协助问卷的发放工作。

表 5-1 调查人员记录表

调查社区	调查人员	调查时间
同福社区	黄菡、施贺青	3 月 8 号、3 月 9 号
滨江社区	黄菡、陈曦	3 月 12 号
辅仁社区	黄菡、严志明	3 月 13 号
琴台社区	魏羽璠、黄竞超	3 月 7 号、3 月 8 号
额头湾社区	魏羽璠、方瑞	3 月 12 号
汉宜社区	魏羽璠、杨天歌	3 月 13 号
宝通寺社区	孔慧琳、贺苏迅	3 月 5 号
解放桥社区	孔慧琳、陈琦轩	3 月 6 号
珞军社区	孔慧琳、陈琦轩	3 月 12 号、3 月 13 号
尤李社区	孔慧琳、贺苏迅 万文娇、刘昭锦 黄菡、施贺青	3 月 16 号
流芳社区	万文娇、唐昊	3 月 5 号、3 月 6 号
长咀社区	万文娇、张烁	3 月 9 号、3 月 10 号
北湖区	万文娇、刘昭锦	3 月 12 号
金口区	万文娇、刘昭锦	3 月 13 号

在发放正式调查问卷阶段，我们从 3 月 5 号开始到 3 月 16 号截止，共花费了 12 天的时间。3 月 19 号前往宝通寺社区进行了一次问卷补发。在发放问卷的过程中，每个团队成员都邀请了自己的朋友进行协助，在整个调查过程中，严格规定团队成员必须和自己的朋友同步行动，两人同时为一个被调查者服务，及时回答被调查的疑问。当发现被调查者填写问卷速度较快、随意性较高时，及时提示被调查者，若其不配合则放弃该位被调查者，并将其问卷标记为无效问卷，按照事先制定好的规则抽取下一个样本。

（二）调查时间和工作期限

我们调查的工作期限为 2015 年 12 月 5 日-2016 年 4 月 6 日。

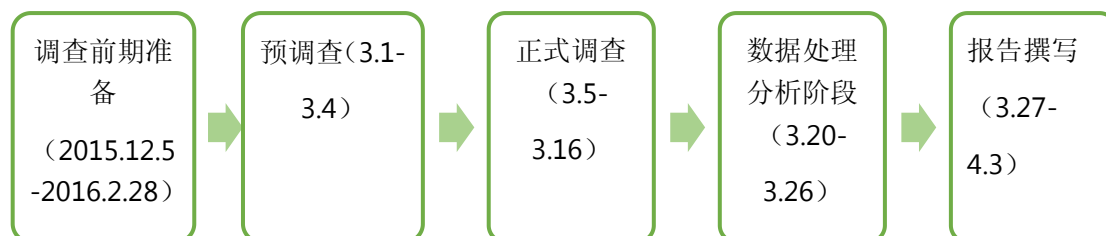


图 5-1 调查工作流程图

（三）质量控制

本次实地调研过程中，我们“云调研电子问卷”的形式，以智能客户端为展现形式，在在平板电脑上运行电子问卷页面，调研过程中只需要将平板电脑给被调查者填写，替代传统纸质问卷形式。这种调研方式带来以下优点：

1.更高的数据回收质量

电子问卷内置问卷跳问控制和质量控制，如果被调查者发生了漏答或是填写形式不正确，程序会自动提醒被调查者错误，直至访问员填写正确才能继续答题并成功提交。在问卷的录入环节就保证被调查者录入的数据格式符合要求，完全避免数据缺失情况的发生，为我们的实地调查带来方便。

2.更短的调研周期

使用平板电脑装载电子问卷进行实地调研的方式，让我们省去了数据录入环节，控制该环节产生的录入错误缩短了项目周期和时间。一般传统纸质问卷都需要人工将数据录入系统，为了减少录入错误，一般还需要双重录入，数据录入环节需要耗费大量时间。采用电子问卷可以及时将数据导出，更为方便快捷。

3.更低的调查费用

电子问卷省去了纸质问卷的打印成本和装订费用。让我们这支由四个大学生组成的团队不用负担过大的费用，在武汉市范围内调查 1000 余人的方案能够顺利执行。

在问卷调查过程中，我们随时随地检查回收问卷的有效度。当发现被调查者填写问卷速度较快、随意性较高时，及时提示被调查者，若其不配合则放弃该位被调查者，并将其问卷标记为无效问卷，按照事先制定好的规则抽取下一个样本。此外，我们在问卷中设置了两道过滤题，过滤掉被调查者及其亲人朋友在咨询调查公司工作的人群和六个月内接受过智能可穿戴产品相关调查的人群，因为此类人群较为熟悉我们的问卷流程，容易选择那些使调查尽快结束的选项，其填写的

问卷反映真实情况的程度较差，因此我们统一规定将这部分问卷视为无效问卷。

（四）抽样效果控制

我们的抽样过程保证了调查全过程均为概率抽样，即随机抽样。为判断抽样效果的好坏，控制抽样估计的精度，我们对样本设计的效果进行了测定，计算本次抽样调查的设计效果系数 $deff$ 。

1. 抽样效果控制

首先，对三阶段的总体比例的方差进行无偏估计，记 N 为初级单元数， M_j 为每个初级单元含有的二级单元数， K_{ij} 为每个二级单元含有的三级单元数， n, m, k 依次为总体抽取初级单元的样本量，初级单元抽取二级单元的样本量和二级单元抽取三级单元的样本量。 Z_i, Z_{ij}, Z_{ijk} 分别为第 i 个初级单元入样的概率，第 i 个初级单元下第 j 个二级单元被抽中的概率和第 i 个初级单元第 j 个二级单元下的第 k 个三级单元入样的概率。记：

$$y_{ijk} = \begin{cases} 1 & \text{听说过智能手环} \\ 0 & \text{没有听说过智能手环} \end{cases}$$

比例估计为总体单位中具有某种属性的单位的比例，样本比例为该样本的均值。总体总值估计量的无偏估计为式（5.1）

$$\hat{Y} = \frac{1}{nmk} \sum_{i=1}^n \frac{1}{Z_i} \sum_{j=1}^m \frac{1}{Z_{ij}} \sum_{u=1}^k y_{iju} \quad (5.1)$$

其总体总值的无偏估计为式（5.2）

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^N \frac{Y_i^2}{Z_i} - Y^2 = \frac{1}{nm} \sum_{i=1}^N \sum_{j=1}^{M_i} \frac{Y_{ij}^2}{Z_{ij}^2} - Y_i^2 = \frac{1}{nmk} \sum_{i=1}^N \frac{1}{Z_i} \sum_{j=1}^{M_i} \frac{1}{Z_{ij}} \sum_{u=1}^{K_{ij}} \frac{Y_{iju}^2}{Z_{iju}} - Y_{ij}^2 \quad (5.2)$$

其中 $Y_{ij} = \sum_{u=1}^{K_{ij}} Y_{iju}$ ，即为第 i 个初级单元下第 j 个二级单元下抽中的三级单元的
标志值总和。 $Y_i = \sum_{j=1}^{M_i} \sum_{u=1}^{K_{ij}} Y_{iju} = \sum_{j=1}^{M_i} Y_{ij}$ 即为第 i 个初级单元下所抽取的所有二级单元分别包含的三级单元的标志值总和。

而对于三阶段抽样而言，由于初级单元和二级单元被抽中的概率均由各单元包含三级单元的个数的比例决定的，第三阶段为等概率抽样,可得式（5.3）

$$Z_i = \frac{\sum_{j=1}^{M_i} K_{ij}}{N \cdot M_i} = \frac{\sum_{j=1}^{M_i} K_{ij}}{M_0}, \quad Z_{ij} = \frac{K_{ij}}{M_i}, \quad Z_{iju} = \frac{1}{K_{ij}} \quad (5.3)$$

因此可以得到总体总值的估计量为公式（5.4）

$$\hat{Y} = \frac{M_0}{nmk} \sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k y_{iju} \quad (5.4)$$

其中 $M_0 = \sum_{i=1}^N \sum_{j=1}^{M_i} K_{ij}$ ，那么总体总值的方差估计量就为公式（5.5）

$$\bar{\bar{Y}} = \frac{\hat{Y}}{M_0} = \frac{1}{nmk} \sum_{i=1}^n \sum_{j=1}^m \sum_{u=1}^k y_{iju} \quad (5.5)$$

所以总体均值的方差为公式（5.6）

$$v(\hat{Y}) = \frac{M_0^2}{n(n-1)} \sum_{i=1}^n \bar{y}_i - \bar{\bar{Y}}^2 \quad (5.6)$$

其中 $\bar{\bar{y}} = \frac{1}{mk} \sum_{j=1}^m \sum_{u=1}^k y_{iju}$ ，则总体均值的方差即为式（5.7）：

$$v(\bar{\bar{y}}) = \frac{1}{n(n-1)} \sum_{i=1}^n \bar{y}_i - \bar{\bar{y}}^2 \quad (5.7)$$

而总体比例的估计相当于总体均值的估计，则类比可得式（5.8）

$$v(p) = \frac{1}{n(n-1)} \sum_{i=1}^n p_i - p^2 \quad (5.8)$$

根据该三阶段的方差公式（5.7）可分别求出每一层内的总体比例方差。则根据分层抽样的方差估计可得式（5.9）

$$v(p_{prop}) = \sum_{h=1}^L W_h^2 v(p_h) \quad (5.9)$$

在求出第 i 层的总体比例的方差的无偏估计 $v(p_h)$ 后，两层总共的总体比例方差的无偏估计即可由式（5.9）算得。其中 W_h 为每层的层权，在本次调查中即为第 h 层拥有的人口数占总人口的比重。

2.设计系数效应计算

假设这 1080 名被调查者是通过简单随机抽样方式直接抽取的，那么可以计算出简单随机抽样方式下总体均值方差的无偏估计量，见公式（5.10）

$$v(p_{srs}) = \frac{1}{n} \frac{f}{1-f} p(1-p) \quad (5.10)$$

其中，f 表示随机抽样方式下的抽样比；n 表示样本总量； S^2 表示样本方差。根据调查数据计算可得 $v(p_{srs}) = 0.00023$

$$deff = \frac{v(p_{prop})}{v(p_{srs})} \quad (5.11)$$

最终计算设计效果系数式为式（5.11），根据计算表 5-2，可以得到我们的分层不等概率三阶段抽样的设计系数。

表 5-2 抽样效果计算表

行政区		回收有效问卷数	认知人数	比例 $p_{ij}(\%)$	每层比例 $p_i(\%)$	$(p_{ij}-p_i)^2$	每层方差	总方差	Deff
中心城区	江岸区	222	117	52.7	51.9	0.000064	0.0001532	0.00042	1.83
	江汉区	74	38	51.4		0.000025			
	硚口区	148	74	47.9		0.0016			
	武昌区	222	119	53.6		0.000289			
	洪山区	74	36	48.6		0.001089			
远城区	江夏区	170	89	52.4	51.2	0.000144	0.00144		
	东西湖区	170	85	50.0		0.000144			

则可得到我们的分层三阶段抽样的设计系数为 1.83，抽样方案的设计效果较好，达到了预期目标。

六、试调查的数据初步分析

在完成问卷设计后,我们依据抽样框在入样的 14 个社区进行问卷的试发放。每个社区发放 10 份问卷,共发放了 140 份问卷,回收 140 份问卷,经过筛选后,得到了 128 份有效问卷。根据回收问卷对数据进行项目分析以及信度、效度检验,以检查问卷的题目的鉴别度,测量结果的可靠性以及问卷的有效性。

(一) 区分度分析

在整理回收的问卷后,首先对量表进行各个题项的区分度分析。主要目的在于检验编制的量表或检验个别题项的可靠程度。本次调查采用高低组平均数差异检验法进行检验。低区分度的题目不能有效的鉴别被调查者。

首先对使用过智能手环的被调查者填写的 Q13(智能手环满意度调查题)量表进行项目分析,将每个被调查者所填的量表的 10 项的各得分相加求得总分,如果被调查者该项未回答,则将该项的平均得分赋予给被调查者此项的得分。将每个样本的总得分从高到低排序,将总得分前 27%的视为高分组,后 27%的视为低分组。然后对高低两组的各个条目得分进行独立样本 t 检验以剔除差异不显著的项目。得到前 27%的得分为 39 分以上,后 27%的得分为 27 分以下。运用样本差异性 t 检验法对两组的 10 个题项分别进行检验,结果中的 t 值全部显著(p 值小于 0.05),表明量表中的 10 个题项均具有鉴别度,量表中所有题都能鉴别不同受试者的反应程度。因此,对于问卷中用于测试使用过智能手环的调查者的部分通过了项目分析检验。

同理,我们对于非智能手环用户的被调查者填写的 Q16-Q18 题也进行了项目分析,得到高分组的得分为 32 分以上,低分组的得分为 19 分以下。运用样本差异性 t 检验法对两组的 3 个题项分别进行检验,结果中的 t 值全部显著(p 值小于 0.05),表明这 3 个题项均具有鉴别度。因此,对于问卷中用于测试非智能手环用户的问题也通过了项目分析检验。

综上所述,调查问卷通过了检验,可以用于正式调查。

(二) 信度检验

问卷设计质量的信度检验指的是对问卷测量结果准确性的分析,即对设计的问卷在多次重复使用下得到的数据结果的精确性和可靠性检验。通过信度分析,可以反映被测特征的真实程度。我们采用 Cronbach 信度测量问卷项目的内在一致性系数,信度系数的取值范围为 [0,1]。Cronbach 系数的计算公式如下式所示:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right) \quad (6.1)$$

其中， k 为量表的总题数， Si^2 为第 i 题得分的题内方差， ST^2 为全部题项总得分的方差。从公式中可以看出，Cronbach 系数评价的是量表中各调查项目得分的内部一致性。信度系数越大，说明测量的可信度越大。一份设计较为良好的问卷，信度系数最好在 0.80 以上，0.70 至 0.80 之间算是可以接受的范围，若分量表的内部一致性系数在 0.60 以下或者总量表的信度系数在 0.80 以下，应考虑重新修订量表或增删题项。

我们对问卷中 Q13 的量表计算的两个维度的 Cronbach 系数的结果如下表所示：

表 6-1 信度分析表

层面	Cronbach 系数	项数	信度评价
外形和材质满意度	0.744	3	较好
功能与服务满意度	0.723	4	较好
使用体验满意度	0.721	3	较好

各层面的 Cronbach 系数均大于 0.7，总体量表信度 0.741，由此说明问卷结构与题项设计的科学性及合理性。

（三）效度检验

效度是指能够准确测出所需研究的事物的程度，即有效性。效度分为三种类型：内容效度、准则效度和结构效度。我们主要采取内容效度与结构效度对预调查的数据进行效度检验。

1. 内容效度

内容效度又称表面效度或逻辑效度，是指所设计的题项能否代表所要测量的内容或主题。用各分量表与总量表之间的相关性作为考察量表内容效度的指标，检测某量表所能代表的主题的多少。

表 6-2 单项与总和和相关效度分析

	外形和材质满意度	功能与服务满意度	使用体验满意度
相关系数	0.652	0.786	0.749
Sig（双侧）	0.000	0.000	0.000

由表 6-2 可以看出，各因子与总量表得分之间的相关性很显著（ p 值均小于 0.01），这表明量表具有较好的内部一致性。

2.结构效度

结构效度是指测量结果体现出来的某种结构与测值之间的对应程度。结构效度可采用相关分析、因子分析、结构方程模型来评价，而结构效度评价的常用统计方法是因子分析。通过因子分析可以从量表中提取一些公因子，各公因子分别与特定变量高度相关，这些公因子代表了量表的基本结构，通过因子分析可以考察问卷是否能测量我们问卷设计时假设的某种结构。

在进行因子分析前需要进行采样充足性检验(Kaiser-Meyer-Olkin, KMO) 和 Bartlett 球状检验。KMO 取值在 0 和 1 之间。当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时，KMO 值接近 1，意味着变量间的相关性越强，原有变量越适合作因子分析，并且因子分析的结果越好。Bartlett 球度检验值越大则表明变量之间的独立性越高，各个条目存在共享公因子的可能性越低，越适合做因子分析。

KMO>0.9 非常适合因子分析； $0.8 < \text{KMO} < 0.9$ 适合；0.7 以上尚可，0.6 时效果很差，0.5 以下不适宜作因子分析。利用 SPSS 计算问卷的 KMO 系数为 0.757，表明适合进行因子分析，P 值为 0.000，因此问卷的结构设计较好。

上述检验结果表明，预调查所采用的问卷能够达到本次调查目的。

七、正式调查数据处理与检验

（一）数据处理

1.数据的接收与导出

由于我们使用实地电子问卷调研的形式，避免了被调查者漏答及回答内容形式不正确的问题。调查过程中，我们将被调查者配合度较低、填写速度较快、较随意的问卷、以及中途放弃填写的问卷标注为无效问卷，在筛除被调查者及其亲朋好友在咨询调查工作的、以及六个月内接受过智能可穿戴产品调研的样本后，回收的有效样本整体质量较高。没有缺失数据、问卷不完整的情况。实地发放调查问卷阶段结束后，我们直接将数据导出、无需人工录入数据、避免了人工录入阶段出错的可能性，缩短了项目执行周期。

2.数据的净化

我们的调查问卷没有设置开放性问题，所有调查问题只需点击相应选项即可。因此不存在数据超过范围、出现异常值的情况。经分析，我们的调查问卷中第7题（每周锻炼频次）和第9题（最钟爱的休闲活动）可以作为测试被调查者作答是否真实可信的依据。若被调查者第7题选择“我不喜欢运动”，第9题却选择“运动健身”，则将该问卷视为前后矛盾逻辑不一致的无效问卷。

（二）数据分析方法

1.定性分析

（1）归纳分析法

问卷调查前，我们查阅了相关文献了解了可穿戴设备、智能手环的基本情况和现状。通过问卷调查，我们获取了14个社区附近的居民信息，以及他们对智能手环的看法和了解程度。

（2）比较分析法

本次调查我们比较了两个方面的信息：第一，智能手环用户群和非智能手环群的个人信息特征；第二，智能手环使用者对其使用体验各方面的不同满意程度。

（3）结构分析法

我们的问卷调查主要从现有用户和潜在用户这两个角度进行剖析，旨在发现潜在用户特点并结合已有用户对智能手环的使用体验提出产品改进方面的建议。

2.定量分析

(1) 检查分析

在对调查资料数据进行了相关处理之后，我们进行检验分析，主要对问卷的量表设计部分进行信度和效度检验。

(2) 描述性统计

在对搜集到的数据出入计算机进行初步处理后，我们进行了描述性统计分析。我们通过频数分析方法计算各描述统计量指标，绘制了频率分布条形图和饼图，并对调查对象的人群特征分布、使用满意度、改善意见、不使用原因等情况有了一个整体的印象。

(3) Logistic 回归分析

建立二元选择模型，定量分析性别、年龄、学历、职业、收入、兴趣爱好、性格倾向、锻炼频率等因素对是否会购买智能手环的显著影响因素，进而确定智能手环用户群所具有的显著特征。

(4) 聚类分析

针对智能手环潜在用户采用聚类分析法来对潜在用户进行分类，发掘重要潜在客户、重要发展客户、次要和一般潜在客户、低价值客户，提取出每个类别群体的特征。

(三) 数据检验

1.信度检验

正式调查共发放问卷 1325 份，回收有效问卷 1080 份，总回收率为 80.26%。对问卷中第 13 题的 10 道小题组成的智能手环使用满意度量表进行信度检验

问卷信度的概念和基本理论，在预调查分析中已有论述，此处直接采用 Cronbach 系数作为检验标准，展示检验结果显示我们的问卷在主观量表结构中，各层面的 Cronbach α 系数均大于 0.7，属于好的情况，由此说明问卷结构与题项设计的科学性及合理性。

2.效度分析

(1) 内容效度

表 7-1 单项与总和相关度分析

	外形和材质满意度	功能与服务满意度	使用体验满意度
相关系数	0.748	0.764	0.728
Sig (双侧)	0.000	0.000	0.000

注:基于问卷第 13 题各项与整体相关程度分析, 结果在 5%显著性水平下均显著。

由表 7-1 可以看出, 相关性都很显著, 这表明问卷具有良好的内容效度

(2) 结构效度

我们利用因子分析的方法对问卷结构效度进行检测。利用 SPSS 计算问卷的 KMO (Kaiser-Meyer-Olkin) 系数为 0.742, P 值为 0.000, 问卷的结构设计较好。

3.随机性检验

在调查过程中, 我们需要保证抽样的随机性, 因此我们对其中的分类变量进行游程检验。根据样本标志排列所形成的游程的多少进行判断的检验。将样本的观察值按从小到大顺序排列, 找出中位数(或平均数), 分为大于中位数(平均值)的与小于中位数的两个部分。用上下交错形成的游程个数来检验样本是否是随机的。在本调查中, 我们采取单样本变量值的随机性检验, 即对某变量的取值出现是否随机进行检验。单样本变量值零假设为: 总体某变量的变量值出现是随机的。数来实现。所谓游程是样本序列中连续出现的变量值的次数。例如在对性别进行单样本的游程检验时, 由于这个分类变量中有两类取值, 则我们将男性取值为 0, 女性取值为 1。我们得到取值为 0 的数目为 508, 取值为 1 的数目为 572。0 的游程数为 256, 1 的游程为 287, 则总共的游程即 R 为 543。为进行游程检验, 我们可以构造出 Z 统计量:

$$Z = \frac{R - \left(\frac{2n_0n_1}{n_0 + n_1} + 1 \right)}{\sqrt{\frac{2n_0n_1(2n_0n_1 - n_0 - n_1)}{(n_0 + n_1)^2(n_0 + n_1 + 1)}}} \sim N(0,1) \quad (7.1)$$

在 z 统计量中, n0 为 0 出现的次数, n1 为 1 出现的次数, R 为总游程数。代入式求得 Z=0.238。取 a=0.05 时, -1.96<0.238<1.96, 即没有充分理由拒绝原假设, 认为性别序列数据顺序随机化程度较高。

借助 SPSS 软件对问卷中各分类变量进行样本随机性检验, 结果表明大部分变量序列的样本数据顺序不违反随机性, 可认为本次调查问卷的调查数据随机化程度较高, 是一次成功的调查。

4.多选题² 独立性检验

我们希望通过² 独立性检验，探究变量与变量之间是否存在显著的关联关系，调查的大部分项目是否会随着属性变量的变化而变化，以便有利于对调查项目的深层次分析。

以问题“您希望智能手环具有以下哪种功能”与年龄阶段是否相关为例，我们首先做出两变量的列联表如表 7-2：

表 7-2 用户选择偏好与年龄阶段列联表

年龄 功能选择偏好	19 岁以 下	20-29 岁	30-39 岁	40-49 岁	50 岁以 上
提供心率、血压监测数据	47	168	167	198	201
运动计步、计算消耗的卡路里	89	167	178	186	98
消息通知	168	123	87	81	43
GPS 安全定位，以防儿童意外走失	56	115	175	78	89
具有交通卡、门卡、签到卡等磁卡的 刷卡功能	122	134	168	88	65
提供噪音、温度、空气湿度等监测数 据	78	41	124	101	24
远程控制家电、汽车等各种设备	103	87	167	56	43
这些功能都不需要	2	9	37	4	23

借助 SPSS，计算出 $\chi^2 = 367.302$ ，以及对应 P 值=0.000，取显著性水平 0.05，有² ² (20)，因此拒绝原假设，即认为智能手环功能偏好偏好与年龄阶段两个变量是不独立的，说明了二者之间具有相关关系，不同年龄阶段的人群智能手环功能选择偏好显著不一致。除此以外，问卷中所涉及的大部分客观信息两两之间都存在相关关系，这为我们后续的建模分析提供了基础。

以上检验结果表明本次调查是一次成功的调查。

八、武汉市智能手环市场及用户使用现状分析

（一）智能手环市场现状

1.智能手环市场认知率、使用率

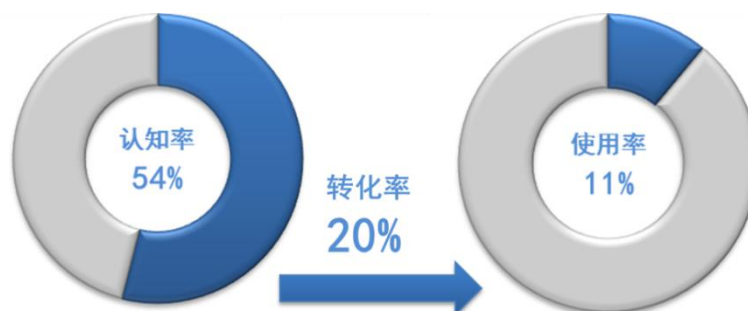


图 8-1 智能手环市场认知率与使用率

根据问卷结果，在所有被调查者中，有 54%的人听说过智能手环，在这部分人群中，有 84%的人表示虽然听说过，但了解程度不高；有 46%的人完全没有听说过智能手环。智能手环市场已经具备了一定的用户认知基础，但从认知到购买的转化率还有待提高。

2.智能手环认知渠道

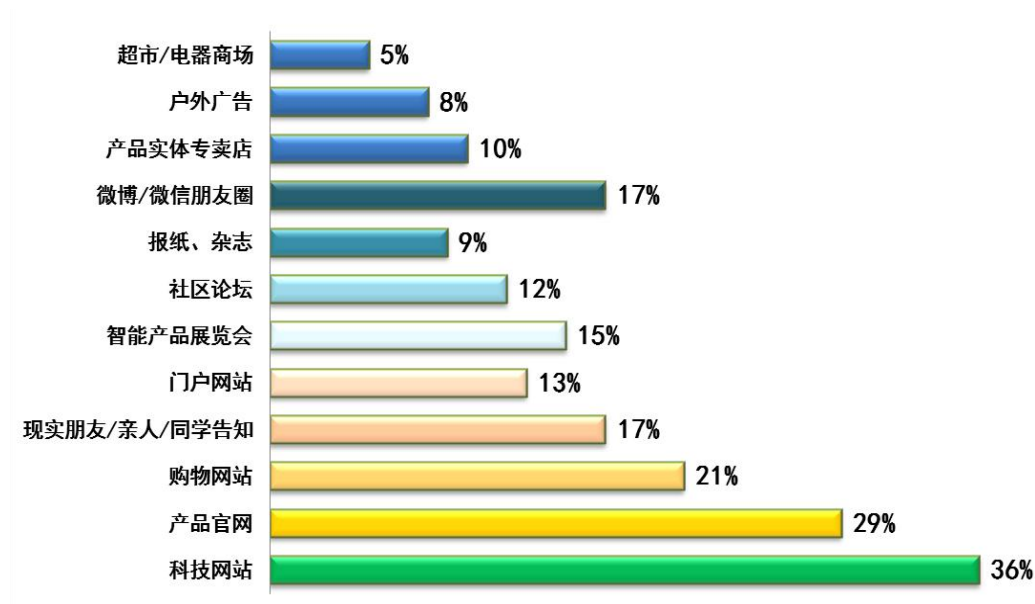


图 8-2 智能手环认知渠道

当前，纸媒的宣传力量已经被新媒体大大削弱，通过报纸、杂志了解智能手环的用户仅仅占到 9%可见一斑。在 583 个听说过智能手环的被调查者中，认知渠道占比最高的为科技网站，其次为产品官网和购物网站。厂商首先应该建设好

自己的官网，注重与消费者的线上互动，将产品的形象、特性、功能通过网络页面的表现方式传递给消费者，加深消费者对产品的理解，增进其对产品的认可度和品牌的忠诚度和支持度。其次，科技网站作为排名第一的认知渠道，其对消费者的引导作用不容小视，厂商应该注重这方面的广告投放，树立品牌的形象。

3.智能手环品牌占有率

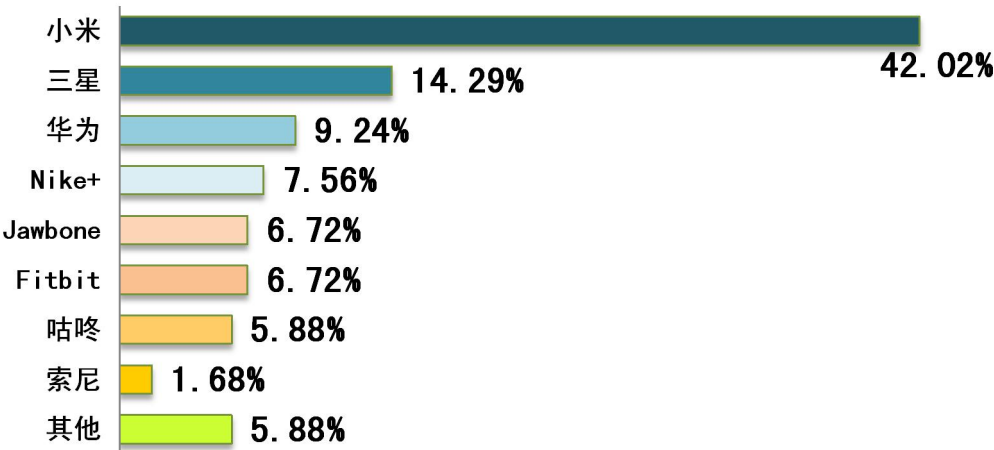


图 8-3 智能手环市场认知率与使用率

由图 8-2 可知，售价合理、供货稳定的小米手环在我国智能手环行业占据领导地位，就同时佐证了本文文本挖掘部分选择小米手环作为代表是比较合理的。三星 Gear 系列由于手机产品份额在适配机型独占的情况下排名第二，国产品牌中，华为及咕咚均有上榜。

4.智能手环来源方式

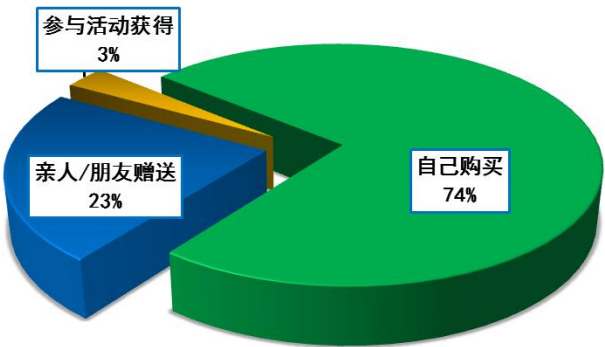


图 8-4 智能手环来源

在智能手环的使用人群中，自购产品的比例较高，占据 74%。值得注意的是，亲人朋友以礼物的形式赠送智能手环也占据相当一部分比例。在文本挖掘的 LDA 模型结果显示，“家人”一词是购买者评价的特征词，这说明可能存在相当一部分比例的消费者会选择为家人购买手环。智能手环作为一种新兴智能可穿戴设备，具备科技感、时尚感，作为礼物赠送是一种大方、得体的选择，其次，智能

手环具备的健康检测等功能对亲朋好友的关心牵挂，同时也符合绿色生活、健康身体的文化趋势。厂商可以从此处着手，打造“智能手环是送礼佳品”的营销概念，参考“脑白金”的营销理念，推广类似于“送礼就送脑白金”的广告词，将以智能手环作为礼物送给亲友变为一种潮流化趋势，带动产品销售数量。

（二）智能手环用户特征

1.性别分布

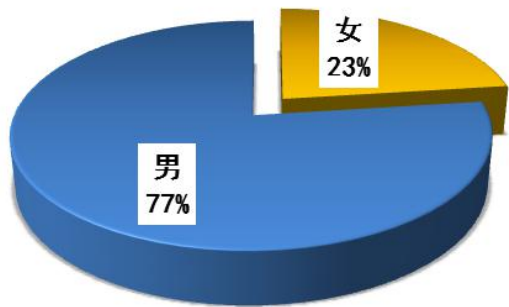


图 8-5 智能手环用户性别分布

在被调查的智能手环用户中，男性所占比例为 77%，明显高于女性。而在整体被调查群体中，53%的被调查者为男性，47%为女性，男女比例基本持平，符合 1:1 的比例。由此可以看出，男性对智能手环的购买欲望及需求比率大于女性。厂商在保持男性用户的同时，可以注重女性用户需求，开拓女性用户市场。

2.年龄分布

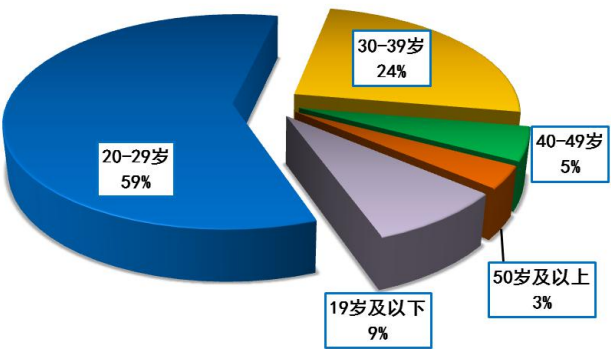


图 8-6 智能手环用户年龄分布

由图 8-6，在使用智能手环的用户中，年龄在 39 岁及以下的智能手环用户占全体用户的 94%。其中 20-29 岁的用户最多，达 59%。这个年龄段的人多为在校大学生或者是刚步入社会的职场新人，对新鲜事物的接受程度和接受速度较高，更加关注智能科技新产品，追逐时髦前卫，且自身具备一定经济条件支撑其购买智能手环。值得关注的是，年龄在 50 岁及以上的用户所占比例仅为 3%，这一部分用户尽管对新型产品的接受速度步入年轻人快，但他们对健康、养生的需求

是最大的。随着经济的发展，人们生活水平的不断提高，以及我国迈入老龄化社会的事实，对关注、检测老年人健康状况的科技产品的需求将会逐渐膨胀。如何占据这部分市场以充分挖掘市场潜力是一个值得思考的问题。

3.学历分布

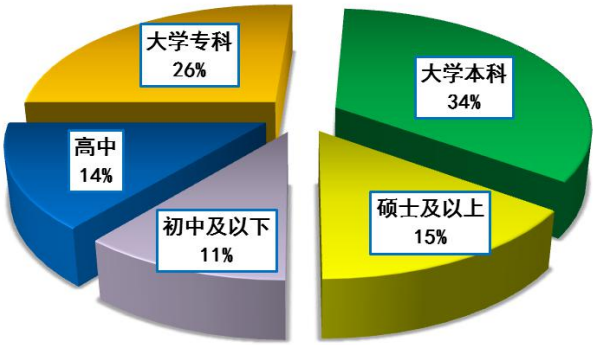


图 8-7 智能手环用户学历分布

在智能手环的用户群中，拥有专科及以上学历的用户所占比例为 75%，其中本科学历占 34%，专科学历占 26%，硕士学历为 15%。在目前的国内市场中，接受过高等教育的用户是智能手环的主要使用群体，而受教育程度较低的用户智能手环使用率并不高。因此，厂商在保持智能手环性价比抓住年轻人市场的同时，可关注高端市场，令智能手环成为社会地位社会身份的象征，占据一部分传统高端手表市场。

4.职业分布

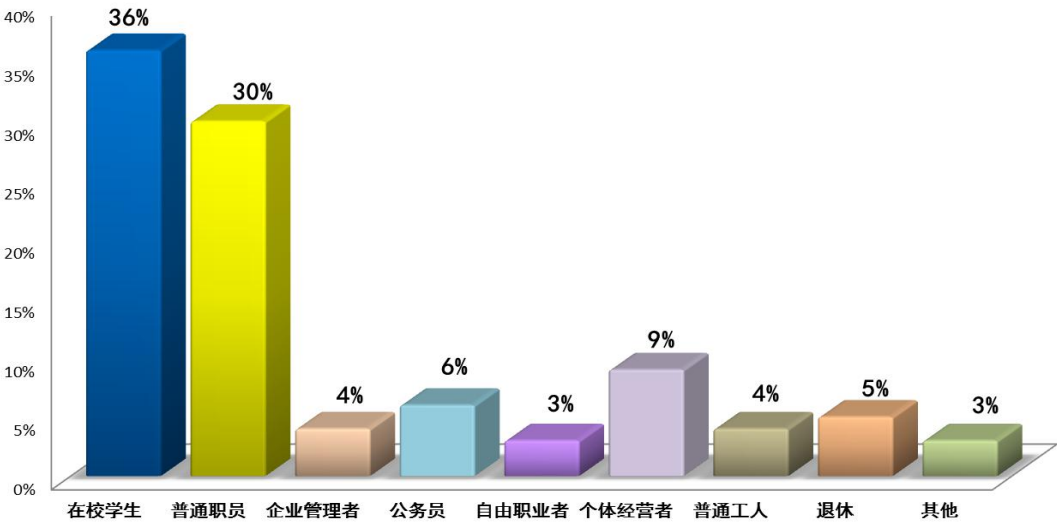


图 8-8 智能手环用户职业分布

由图 8-8，在使用智能手环的用户中，在校学生占 36%，普通职员占 30%。这说明在校学生及普通职员对智能手环的兴趣较高。在 1080 个被调查者中，在校学生、普通职员、公务员、个体经营者、普通工人、退休六个类群体所占比例

大致相当，企业管理者、自由职业者、其他三个类别的数量较少。这说明购买能力较强的企业管理者、公务员对智能手环的使用率较低，但并不意味着这部分人群对智能手环的热情较低，而是产品没有打动他们，激起他们的消费欲望。最具购买力的、愿意支付产品功能溢价的这部分群体的市场潜力有待开发，在产品未来的研发、改善阶段，应给予更多关注。

5.年收入分布

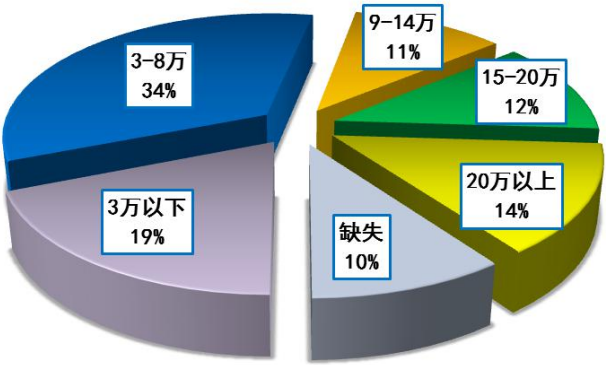


图 8-9 智能手环用户年收入分布

由图 8-9，在被调查的智能手环用户中，34%的用户收入在 3-8 万，19%的用户收入在 3 万以下，收入为 20 万以上的用户为 14%。在年收入的每个等级下，用户数量分布较为均衡，其中 3-8 万的群体占比最高，这和整体被调查对象的年收入分布差异不大。智能手环作为入门级的智能可穿戴设备，尽管品牌较多，功能不尽相同，但整体而言售价并不高。从用户的年收入分布来看，该变量或许不是影响消费者购买的显著因素，为得到可信的结论还需要进一步的统计分析。

6.智能手环用户性格自我评价

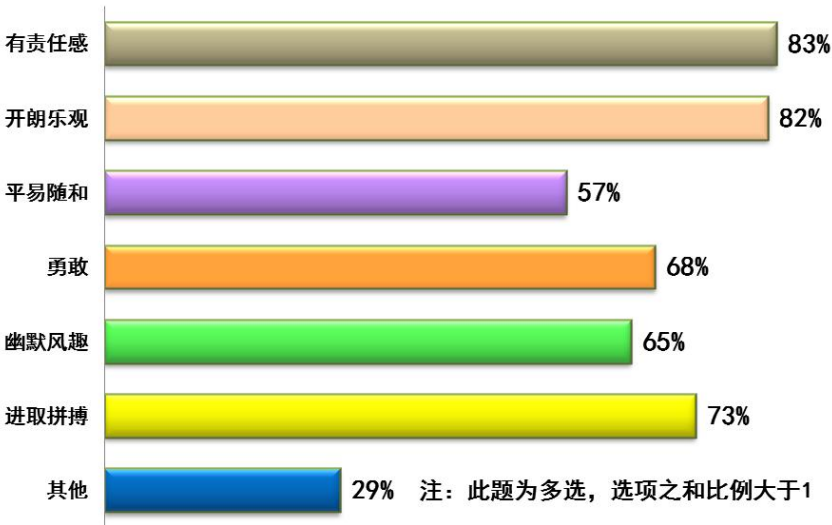


图 8-10 智能手环用户性格自我评价

由图 8-10，智能手环用户对自己的评价是具有拼搏精神、开朗进取、勇敢

无畏，但比较自我、不那么随和。我们认为，在产品的宣传、包装上，应抓住这一属性，为产品贴上“勇敢无畏”、“做自己”等情感标签，使产品具有独特的性格，让消费者产生情感共鸣。

（三）智能手环使用现状

1.智能手环用户流失情况

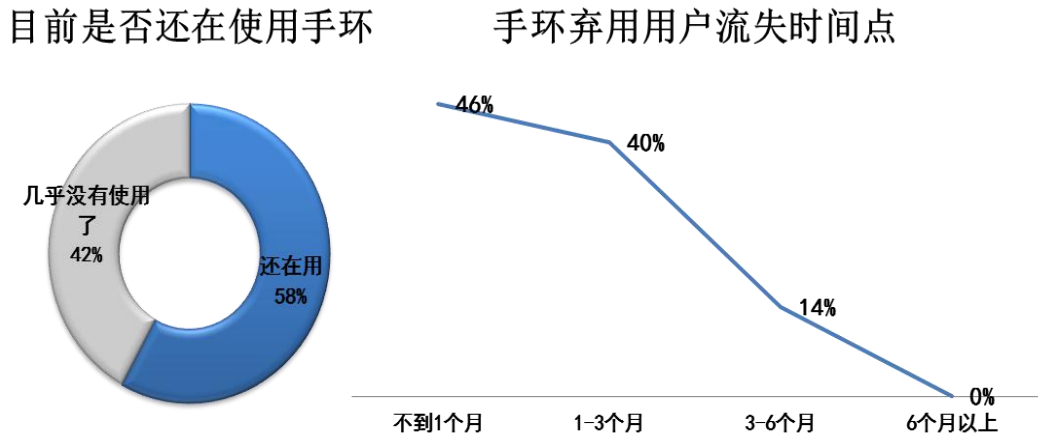


图 8-11 智能手环用户流失情况

在电商平台评价文本的数据挖掘中，我们已经发现用户中对产品的负面评价占据相当一部分的比例。为了研究智能手环用户的使用率和弃用率，我们在问卷中设计了专门的问题。结果显示，42%的手环用户表示目前已经不再使用产品，这部分用户中，将近一半的流失用户在 1 个月内就流失，3 个月内流失的比例高达 86%。面对这样的调查结果，我们不禁要追问流失率为什么这么高？

我们认为其中的原因有以下几点：第一，产品实用性不足，用户没有找到适合的使用场景和养成使用习惯；第二，产品没有有效地融入生活，智能手环与手机连接传导数据的过程不够流畅；第三，周围使用智能手环的朋友不够多，难以坚持。第四，功能不够完善，智能手环的功能还比较单一，不能满足大部分人的需求。

习惯的养成，需要足够的重复次数，当习惯回路越短、越简单，用户使用的次数就会越多，同时，操作越简单，越容易形成条件反射，缩短习惯养成时间。以小米手环为例，用户读取数据需要通过手机中的小米运动 APP，打开手机的蓝牙与设备连接才可传输数据，不可独立使用，让大量的消费者难以坚持使用。

2.智能手环用户产品满意度

消费者在评价智能手环时一般需要综合考虑其外形材质、功能服务和使用体验等多方面因素。我们在调查问卷中设计了顺序量表，分别对智能手环外形材质、功能服务和使用体验的满意程度按照非常满意、满意、一般、不满意、非常

不满意进行等级排序，本文选择中位数统计分析方法进行用户满意度的评价分析。

结果显示，智能手环用户对产品整体满意度较好，约 61%的用户对智能手环整体较为满意。其中满意度较差的二级指标是“使用方便程度”和“数据质量”，满意度较高的是“颜值外观”，“手环重量”，“材质舒适程度”，“功能种类”。

表 8-1 智能手环用户满意度评价中位数分布

一级指标	二级指标	非常不满意	不满意	一般	满意	非常满意
外形和材质	颜值外观				√	
	手环重量				√	
	材质舒适程度				√	
功能与服务	功能种类			√		
	续航能力			√		
	价格水平			√		
	售后服务			√		
使用体验	界面与操作体验			√		
	使用方便程度		√			
	数据质量		√			

智能手环作为一个实物穿戴在人们身上能被人看到，不仅具有实用价值，同时具有象征意义，象征使用者的身份地位、格调品味。人们能够通过可以直观感知的物品外观、材质、品牌等，而不仅仅是从功能质量来判断产品的特性和档次。对于任何显露在外部的可穿戴设备来说，厂商要始终紧跟流行趋势、甚至引领潮流，给予外观设计足够的重视以保证产品销量、打造时尚的品牌形象。

消费者对智能手环的功能种类不满意，不单单是手环的功能不够齐全不够多。产品可以追求功能上的广泛，但这相对较难实现。更现实更为可行的选择是有选择的着重某一种功能。而当前消费者之所以对功能种类不满意，一是产品没有满足消费者的实际产品使用场景的需求，二是激发、培养用户需求的市场工作有待加强。因此产品的功能与消费者联系得更为紧密，数据准确性得以提升，智能手环的前景将会更加开阔。

综上，智能手环目前并不能称得上是一个实用产品，整体满意度差强人意。产品在用户体验、性能等方面亟需提升。值得一提的是 2015 年武汉城市一卡通有限公司开发了既可以用于乘坐公交、地铁、轮渡，还能在自动贩卖机、电影院、大型超市等场所进行消费，同时可以进行武汉通余额查询、消费记录查询，并且可以通过手机进行空中充值的“武汉通智能手环”。如此多样化的功能对手环的功能性，数据准确性等方面具有较高要求，厂商需给予足够的重视。

九、智能手环用户特征的模型分析

通过第八节我们运用描述统计的方法，分析了智能手环用户的特征。但实际上，消费者是否会购买智能手环是多种因素综合影响的结果。有必要建立计量模型，把众多因素综合在一起进行考察。本节通过建立二元选择模型，定量分析性别、年龄、学历、职业、收入、兴趣爱好、性格倾向、锻炼频率等因素对购买智能手环的影响，进而确定智能手环用户群所具有的显著特征。

（一）模型的选择

二元选择模型为由两种备选方案中必择其一的个体抉择行为所建立的模型。考虑到是否使用智能手环只涉及两种回答，是一个二分类因变量，我们建立二元选择模型。

二元选择模型一共分为线性概率模型、Probit 模型、Logit 模型，其中线性概率模型假定被解释变量与解释变量间是线性关系，采用普通最小二乘法（OLS）或加权最小二乘法（WLS）进行参数估计。尽管在实际应用中，对于同一资料用 Probit 回归和 Logistic 回归的结果非常接近，但本文采用的是应用更为广泛的 Logistic 回归^{3M6}，原因如下：

第一，Logistic 回归中的偏回归系数可以计算其 OR 值（Odds Ratio，优势比），具有更贴近实际的解释意义，而 Probit 回归中的偏回归系数含义为其他自变量取值保持不变时自变量每改变 1 个单位，出现阳性结果的概率密度函数值的改变量，这种解释远不如前者直观有用；

第二，Probit 回归是在正态分布的理论基础上进行的，Logistic 回归则是基于二项分布的理论。由于发放的调查问卷中分类变量较多，连续变量极少，因此我们认为采用 Logistic 回归更为合理。

（二）二元选择模型的建立

将第 i 个个体“是否拥有智能手环”这种二元选择行为表示为因变量，当选择“是”时， y_i 取值为 1，当选择“不是”时， y_i 取值为 0。

由于本次设置的问卷调查中大量数据为分类资料，例如年收入分成了五档，如果直接编码为 5、4、3、2、1，令其作为自变量纳入分析，就等价于是假设这五档间的差距完全相等，或者说他们对因变量的数值影响程度是均匀的，这样的假设会过于简单武断、与实际情况不符。另外对于无序多分类变量，如职业，它们之间不存在数量上的高低，因此不可能为其给出一个单独的回归系数估计值，

^{3M6} 王济川、郭志刚.Logistics 回归模型-方法与应用[M].北京高等教育出版社，2001

来表示职业每上升一个单位时因变量的变化趋势。因此，本文采用统计上标准的做法，采用虚拟变量进行拟合，然后根据分析结果对模型进行简化。

（三）对智能手环用户群体特征的研究

对基准类的设定上，为方便录入，原则上将最后一个选项设为基准类，当最后一个选项为其他或该选项的被选中频次小于 30 次时，则以倒数第二个选项设定为基准类。纳入所有需要考虑的变量，建立形如下式的二元选择 Logit 模型：

$$y = \beta_0 + \beta_1 Male + \beta_2 Age19 + \beta_3 age20 + \beta_4 age30 + \beta_5 age40 + \beta_6 Mid\ sch + \beta_7 High\ school + \beta_8 Ycollege + \beta_9 Student + \beta_{10} Clerk + \beta_{11} Executive + \beta_{12} Cov + \beta_{13} Freelance + \beta_{14} Worker + \beta_{15} Income3 + \beta_{16} Income8 + \beta_{17} Income14 + \beta_{18} Income20 + \beta_{19} Photo + \beta_{20} Sports + \beta_{21} Music + \beta_{22} Book + \beta_{23} TV + \beta_{24} Hm + \beta_{25} Pedometer + \beta_{26} Reminder + \beta_{27} GPS + \beta_{28} Magocard + \beta_{29} Em + \beta_{30} Rm + \beta_{31} Watch + \beta_{32} Band + \beta_{33} Camera + \beta_{34} Glasses + \beta_{35} Headphone + \beta_{36} Aggressive + \beta_{37} Funny + \beta_{38} Brave + \beta_{39} Amiable + \beta_{40} Optimistic + \beta_{41} 3or4 + \beta_{42} 1W + \beta_{43} 1M$$

选用向后逐步剔除方法，对初步建立的回归方程根据极大似然估计的统计量的概率值删除对因变量影响不显著的自变量，显著性水平给定为 0.05。经过多轮筛选，排除不显著的变量后，得到的回归结果如下：

表 9-1 logit 模型最终回归结果（剔除后）

变量	回归系数	Wals	P 值	OR{Exp(B)}
常数项	-0.303	1.843	0.175	0.738
性别(female)	-0.289	8.627	0.003	0.749
年龄(Age20)	0.282	17.112	0.000	1.325
最高学历(College)	0.228	10.554	0.001	1.095
爱好(Sports)	0.124	12.185	0.000	1.132
运动频率(3or4-1w)	0.222	19.854	0.000	1.248
$R^2 = 0.2176$		$P(\text{likelihood Ratio test}) = 0.002$		

模型中使用似然比检验中的 χ^2 统计量为 25.001， $p(\chi^2 > 25.001) = 0.002$ ，远小于 0.05 的显著性水平，可以认为解释变量的整体显著性良好。

（四）模型结果分析

$OR\{EXP(B)\}$ 为发生比率，当消费者购买智能手环的概率较小时（一般认为小于0.1），可近似认为 $EXP(B)$ 与发生概率之比非常接近。因此，本文可以将 $EXP(B)$ 值解释为消费者购买智能手环的发生概率为相应的自变量变化前的 $EXP(B)$ 倍数。

1.性别

在5%的显著性水平上，性别通过显著性检验，系数为-0.289。这说明，在其他条件不变的情况下，消费者购买智能手环这一行为存在性别差异。从 $EXP(B)$ 值来看，女性购买智能手环的概率是男性的0.749倍，男性购买智能手环的积极性相对较高。故模型结果显示消费者的性别与购买智能手环的概率呈现负相关。

2.年龄

模型估计结果显示，消费者的年龄状况通过模型系数的显著性检验，系数为0.282，这说明不同的年龄段影响消费者购买智能手环的意愿。从 $EXP(B)$ 值来看，20-29岁的年轻群体购买智能手环的积极性是基准类50岁以上的老龄人群的1.325倍。

3.学历

学历的回归系数为正，且通过了显著性检验，说明具有大学本科学历的群体对智能手环的购买意愿更强。从 $EXP(B)$ 值来看，大学本科学历的人群与基准类（高中/中专/技校）的人群相比，其购买意愿是后者的1.095倍。

4.兴趣爱好

爱好的回归系数为正，且通过了显著性检验，说明喜爱运动的群体相比与基准类（看电视）人群，对智能手环的购买意愿更强。也即，使用智能手环的用户更喜爱户外运动，没有一般人那么喜欢宅在家里。

5.运动频率

运动频率的回归系数为正，且通过了显著性检验，说明每周运动3-4次的群体相比与基准类（我不爱运动）的人群，对智能手环的购买意愿更强。也即，智能手环更贴近热爱运动健身的消费的需求。

十、智能手环潜在用户挖掘

（一）模型选择

1.潜在用户定义

进行实地调查最主要的目的是要挖掘潜在客户，以便智能手环厂商能更有目的性和针对性地开展研发和宣传活动，拓展更多的用户，占有更多的市场份额。对于潜在客户的定义，是指那些未购买或使用过智能手环，但有可能在未来成为智能手环用户的非用户群体。

我们通过分析问卷调查结果发现，有一部分被调查者已经具备了智能手环的易用人特征，但最终却没有购买智能手环，我们将这一部分人群视为目前最有可能挖掘到的潜在用户群体。

2.潜在客户价值模型的建立

在上文中，我们已经对智能手环用户的特征做了一定的分析。接下来，我们通过建立潜在客户价值模型，利用聚类分析，识别、挖掘出最有价值的潜在客户。

传统的 RFM 模型是衡量客户价值和客户创利能力的重要工具和手段。该模型通过客户的近期购买行为、购买的总体频率以及花费金额三项指标描述该客户的价值状况^{3M6}。但是在本次调查中，这些指标并不适用，因此我们从 RMF 模型中衍生出一个类似的模型来适应本次调查项目。我们通过研究潜在客户的数据资料，分析潜在用户在购买智能手环目标和动机方面的共有行为模式。利用调查问卷中的调查项目，选取五个相关项目，将调查结果量化，并把五个量化项目作为研究潜在客户价值的指标。

（二）模型运用与分析

1.聚类因子的选取

我们针对潜在用户进行聚类分析，因此先选取问卷中针对非用户的问题。理论上来说，所有非用户群体均应视为潜在用户，但为了挖掘出更有可能购买智能手环的群体，我们针对非智能手环用户填写的第 16 题进行潜在用户筛选。我们发现，针对“您目前没有购买智能手环的原因”这个问题，第五个选项（我不喜欢智能手环，不考虑购买）的非用户态度比较坚决，成为用户的可能性很小。因此，我们将没有选择这个选项的其余被调查者，定义为潜在客户，共有 563 人。

进行聚类前，我们需要从问卷的调查项目中选取五个因子，作为模型的五个指标。结合上文的结论，我们认为，在消费者特征中有四个变量（性别、年龄、

^{3M6} 美国数据库营销研究所 Arthur Hughes 的研究报告

学历、爱好）与成为潜在客户的关联比较大。再结合第 21 题，我们就可以进行聚类分析。由于第 18 题为多选题，我们假设勾选的选项数目越多的被调查者，购买意愿越大，成为潜在客户的机率越大。对该题的量化处理方法如下：将选择其中一个选项的赋值“1”，否则赋值为“0”，最后将所有选项赋值相加，最后得到第 21 题的得分，也就是该项的指标值。

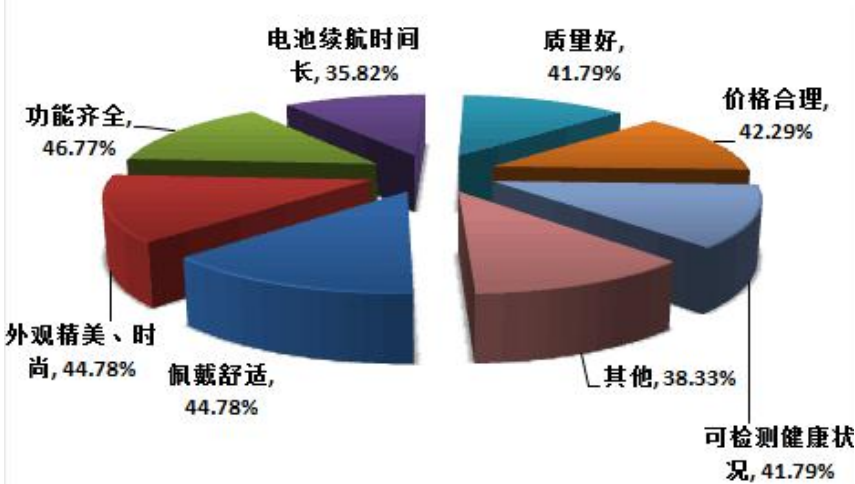


图 10-1 用户关注功能分布表

我们最终确定进行聚类分析的指标有：Q2、Q1、Q3、Q9、Q12 和 Q20，对应因子为：年龄、性别、学历、爱好、年收入和购买意愿（您会因为智能手环具备的哪些特点而购买智能手环）。这六个指标分别用 A、E、S、J、I、P 表示作为识别客户价值的指标（见表 10-1），记为 AESHIP 模型。

表 10-1 潜在客户人物指标含义表

指标	A	E	S	h	I	P
含义	年龄	性别	学历	爱好	年收入	购买意愿（会因几种特点而购买智能手环）

2.潜在客户价值模型的构建

该模型的构建主要由两个部分构成，第一部分根据上述六个指标的数据，对客户做聚类分群；第二部分结合具体项目对客户群进行特征分析，分析其客户价值，并对每个客户群进行排名。

(1) 潜在客户聚类

采用聚类分析的方法对客户进行客户群分，聚成五类，利用 R 软件聚类结果如下表。

表 10-2 潜在客户类型聚类中心表

潜在客户类型	I	II	III	IV	V
年龄（A）	20-29 岁	30-39 岁	40-49 岁	50 岁以上	20-29 岁
学历（E）	大学本科	大学专科	硕士及以上	高中/中专/职 校	大学本科
性别（S）	男	男	男	女	女
爱好（H）	运动健身	看电视	看书看报	听音乐	运动健身
年收入（I）	3 万以下	3-8 万	15-20 万	9-14 万	3 万以下
购买意愿（P 考虑购买情 况数量）	7 类	4 类	3 类	2 类	5 类

(2) 潜在客户价值分析

针对聚类结果进行特征分析，如图 10-2，分别分析五个客户群在各个属性上的情况，从而总结出每个群的特点，

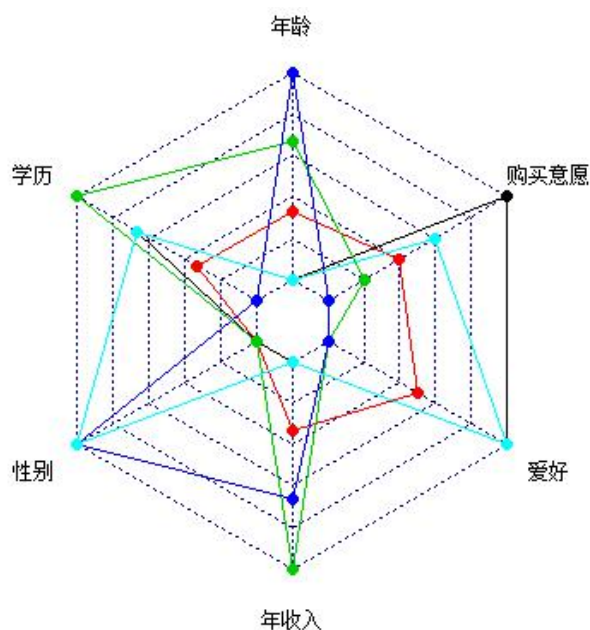


图 10-2 特征分析图

据潜在客户类型聚类中心表和特征分析图说明每个客户群都有显著不同的表现特征，基于该特征描述，我们将潜在用户分成五个类别：重要潜在客户、重要发展客户、次要潜在客户、一般和低价值潜在客户。其中每种潜在客户类别的特征如下：

重要潜在客户：这类客户是第 I 类潜在客户（黑色线条表示），这类群体主要是爱好运动健身的男性，年龄在 20-29 岁之间，他们具有大学本科学历，而且对智能手环非常感兴趣，购买意愿最高。结合这类人的问卷分析，这一类人是非常重要的潜在客户，在未来最有可能购买智能手环的人群之一。但是，这类人中大多数还是学生或者刚参加工作不久，所以年收入较低。针对这一人群，需要产品开发和营销公司在产品性价比方面着手开展研发和营销活动

重要发展客户：这类客户是第 V 类潜在客户（浅蓝色线条表示），这类客户主要是爱好运动健身的女性，年龄在 20-29 岁之间，大学本科学历，对智能手环比较感兴趣。“全民健身”的热潮使得这一类人是非常重要的潜在客户，是在未来非常有可能购买智能手环的人群之一。但是，与第一类群体类似，他们中大多数还是学生或者刚参加工作不久，所以收入较低。所以针对这一人群，不仅需要关注产品性价比方面，还需要有针对性的策划营销方案，例如投放广告鼓励这类人群购买智能手环。

次要和一般潜在客户：第 III 类（绿色线条表示）和这类客户由第 II 类（红色线条表示），第 III 类主要是高学历，高收入，年龄在 40-49 之间的男性，但是对健身没有特别喜好，爱好看书读报，在购买意愿这一属性上处于中间水平。第 II 类主要是年龄在 30-39 岁之间的男性群体，学历和收入均两个属性上均处于中间地位，这类人群有一定的意愿购买智能手环，但是对运动健身没有特别的喜好。这两类人群队长智能手环的需求不强烈，智能手环公司可以适当针对这些人群的特点，增加手环其他方面的功能，来刺激他们的需求。

低价值潜在客户：第 IV 类人群（宝蓝色线条表示），是年纪较大（50 岁以上）的女性，这类人群学历不高，对运动和手环都没有特别的关注，价值不大。

十一、调查结论与产品发展策略

（一）调查结论

1.智能手环已具备一定的市场认知率，功能实用性是集中关注点

通过问卷调查发现，有 54% 的人听说过智能手环，智能手环市场已经具备了一定的用户认知基础，但从认知到购买的转化率还有待提高。消费者最为看重的功能是睡眠监测、心率测量以及运动记录步数、充当闹钟等提醒功能，非客户中愿意为健康检测功能买单的比重最大。从电商平台评论文本挖掘的高频词云中也可以看出购买者最关注的还是手环的功能方面，并且他们对于手环数据的准确性也有一定的期待。另外，问卷调查的结果显示，被调查者有很大一部分希望智能手环能具有 GPS 安全定位，以防儿童意外走失，和交通卡、门卡、签到卡等磁卡的刷卡功能，这说明这些功能对使用者是十分具实用性有吸引力的。

尽管智能手环作为一个实物穿戴在人们身上能被人看到，高颜值的外观设计固然重要。然而要建立良好的口碑，提高消费者对品牌的认可度、忠诚度，打造好评度高销量高的明星产品，最为核心、重要的还是产品功能的实用性。

2.消费者对智能手环总体的满意度差强人意

通过抓取天猫和京东两个电商网站的消费者评价数据，进行一系列的文本挖掘后发现：19.9%的消费者购买的智能手环给予了高度满意的评价，10.7%的用户做出了负面消极的评价。看似矛盾的结果，细想一下并不矛盾。以小米手环为代表的智能手环因为价格便宜、功能种类多、商家发货速度快等原因让消费者感觉到物超所值，对商品较为满意。但由于产品实用性不足，用户没有找到适合的使用场景，养成持续使用的习惯，特别是智能手环与手机相连传导数据的过程不够流畅，导致类似“失望”、“垃圾”、“不准”、“玩玩”等负面评价出现的频次很高。

通过实地进行的问卷调查，我们进一步发现，约 61% 的用户对购买的智能手环整体较为满意。其中满意度较高的指标是“颜值外观”，“手环重量”，“材质舒适程度”，“功能种类”，满意度较差的是“使用方便程度”和“数据质量”，这直接导致 36% 的用户在三个月内弃用智能手环。

我们认为，智能手环产品尚存在较大的改进发展空间，针对消费者实际需求对产品做出改进，达到优化用户体验的目的，能够激发出更大的市场潜力。

3.易用人群特征明确

从二元模型的结果及潜在用户挖掘分析可以得到，变量中性别、年龄、学历、爱好、运动频率可以用于界定易用人群。最终界定出的易用人群特征为男

性、20-29 岁的年轻群体、具备大学本科学历、喜爱运动且运动频率较高的用户。此外，由于市面上大部分智能手环普遍价格不高，所以职业、年收入等变量对消费者购买智能手环的影响不大。另外结合面描述性分析部分，可以看出被调查中，对自身性格评价是“进取拼搏”、“开朗乐观”、“勇敢”的人群更倾向于使用或者购买智能手环。

通过实地调查后的建模分析，我们发现智能手环目前的消费者主要是有健身需求的男性，对健康检测有需求的消费者市场有待开发。

4.接近 45%潜在用户有较大可能成为智能手环用户

将潜在用户从非客户中剥离开后，通过潜在用户群特征分析，我们发现接近 45%的潜在用户为重要潜在客户和重要发展客户，是比较容易发展为客用户的，还有 15%是潜在用户中相对容易成为用户的。所有潜在用户可以被分为五类，每一类也通过聚类中心有了更明确的定位。所以，智能手环开发商可以通过用户群特征分析对潜在客户细化定位，也可以针对潜在客户制定发展方向，更好地满足潜在客户的需求。

5.小米手环占据了近一半的智能手环市场份额

在描述性分析中，性价比高、供货稳定的小米手环在我国智能手环行业占据领导地位，市场份额达到了 42%。另外三星 Gear 系列由于手机产品份额在适配机型独占的情况下排名第二。说明小米手环的低价营销策略和品牌效应策略十分成功，其产品也适宜于智能手环易用人群的特点。

（二） 产品发展策略

1. 宣传策略

（1）在宣传方案中注入情感化标签

亲人朋友以礼物的形式赠送智能手环也占据相当一部分比例。智能手环作为一种新兴智能可穿戴设备，具备科技感、时尚感、易穿戴、易使用，作为礼物赠送是一种大方、得体、贴心的选择，其次，智能手环具备的健康检测等功能对亲朋好友的关心牵挂，同时也符合绿色生活、健康身体的文化趋势。厂商可以从此处着手，打造“智能手环是送礼佳品”或者“或者送礼就送智能手环”的营销概念，抓住消费者心理适当利用“名人效应”、“从众心理”，将以智能手环作为礼物送给亲友变为一种潮流化趋势，并且同时提高智能手环认知度，以带动产品销售数量。

智能手环用户对自己的评价是具有拼搏精神、开朗进取、勇敢无畏，但比较自我、不那么随和。我们认为，在产品的宣传、外形、包装设计上，应抓住这一属性，为产品贴上“勇敢无畏”、“做自己”等情感标签，使产品具有独特的性格，参考可口可乐不同饮料瓶配以不同广告标语来吸引不同性格客户群这一营销方式，为不同性格或者爱好的客户配以个性化的设计，这一让消费者产生情感共鸣。

(2) 加大新媒体渠道的宣传力度

当前，纸媒的宣传力量已经被新媒体大大削弱，由图 8-2 可知，通过报纸、杂志了解智能手环的用户仅仅占到 9%便可见一斑。在 583 个听说过智能手环的被调查者中，认知渠道占比最高的为科技网站，其次为产品官网和购物网站。厂商首先应该建设好自己的官网，注重与消费者的线上互动，将产品的形象、特性、功能通过网络页面的表现方式传递给消费者，加深消费者对产品的理解，从而增进其对产品的认可度和品牌的忠诚度和支持度。其次，科技网站作为排名第一的认知渠道，其对消费者的引导作用不容小视，厂商应该注重这方面的广告投放，树立品牌的形象。

2. 产品研发方向

(1) 挖掘中老年人市场

非客户群体中，大部分为年龄在 50 岁及以上的中老年人群。这一部分用户尽管对新兴产品的接受速度不如年轻人快，但他们对健康、养生的需求是最大的。随着经济的发展，人们生活水平的不断提高，以及我国迈入老龄化社会的事实，对能够监测健康状况的科技产品的需求将会逐渐膨胀。厂商可以更多的关注中老年人群体的需求，占据这块有潜力的市场。针对中老年人群体厂商应该考虑尝试电视宣传，以增加智能手环认知率，扩大智能手环用户群体。

(2) 注重高端市场

在目前的国内市场中，大部分的智能手环价格不高，尤其是市场份额最大的小米手环一直以来关注的都是平价市场。但是，根据我们的调查发现，智能手环的主要使用群体和有较大可能在未来购买智能手环的潜在用户都是接受过高等教育的，而受教育程度较低的用户的手环使用率并不高。因此，厂商在保持智能手环性价比以求抓住年轻人市场的同时，可以关注高端市场，将智能手环的材质、外观、性能等方面想更商务更高端的方向发展，致力于使智能手环占据一部分传统高端手表市场份额，令智能手环也成为社会地位、社会身份的象征。

(3) 重视产品情感化设计

重视产品情感化设计，给用户带来一种高度沉浸其中的流体验，具体而言就是重视智能手环与手机连接的流畅性，这将有助于增加用户粘性，培养用户的忠诚度，与用户形成持久关系。

问卷调查结果显示，智能手环用户的弃用率很高。习惯的养成，需要足够的重复次数，当习惯回路越短、越简单，用户使用的次数就会越多，同时，操作越简单，越容易形成条件反射，缩短习惯养成时间。重视产品情感化设计，让用户在使用产品的过程中产生愉悦的情绪，才能培养用户的使用习惯，降低产品弃用率，打造良好的口碑。

(4) 完善现有功能，开发更具实用性的功能

作为一种实用性的可穿戴设备，人们重点关注的是其数据的准确性和功能的实用性。除了对现在手环普遍现有的睡眠监测、心率测量以及运动记录步数、充当闹钟等提醒功能进行完善，使手环的数据结果更有参考性外，厂商需要适当增加功能的种类，满足更多人的需求。例如，GPS 安全定位、以防儿童意外走失、普通支付、交通卡、门卡、签到卡等磁卡的刷卡等功能，不仅实用性很高，而且这些功能能够满足更多人群（例如儿童和老人）的生活需求，这些功能在被调查者中呼声也很高，说明开发这些功能是很有必要的，而且这些功能能扩大智能手环的受众群，刺激消费。

参考文献

- [1]赵妍妍,秦兵,刘挺. 文本情感分析[J]. 软件学报,2010,08:1834-1848.
- [2]PANG BO, LEE L, VAITHYANATHAN S.Thumbs up Sentiment classification using machine learning techniques [C] // Proc of Conference on Empirical Methods in Natural Language Processing. Morristown, NJ: Assuciation for Computational Linguistics, 2002: 79-86.
- [3]Ni X,Xue G,Ling X ,et al. Exploring in the Weblog space by detecting informative and affective articles [C] //Proc.of the16th Int'l. Conf .on World Wide Web,2007:281-290.
- [4]蔡健平, 王琳琳, 林世平. 基于机器学习的词语和句子极性分析[c]t / 中国人工智能学会第 12 届全国学术年会论文集: 上集. 北京: 北京邮电大学出版社, 2007.
- [5]Hu M, Liu B. Mining and summarizing customer reviews[C]// Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, Usa, August. 2004:168-177.
- [6]朱嫣岚,闵锦,周雅倩,等.基于 HowNet 的词汇语义倾向计算[J].中文信息学报, 2006, 20(1): 14-20.
- [7]闻彬,何婷婷,罗乐,宋乐,王倩. 基于语义理解的文本情感分类方法研究[J]. 计算机科学,2010,06:261-264.
- [8]石晶,范猛,李万龙. 基于 LDA 模型的主题分析[J].自动化学报, 2009,12:1586-1592
- [9]侯志强、刘喜波.分层三阶段及以上抽样的抽样设计田.数学的实践与认识, 2007,: 114-117.
- [10]金勇进、候志强.中国人口变动调查的抽样设计田.统计与信息论坛, 2007, 2 (4):12-2_5.
- [11]SPSS Missing Value AnalysisTM 7. 5 User Manual. SPSS Inc. Chicago, Illinois,1997
- [12]张文彤主编.SPSS 11 统计分析教程(高级篇).北京:北京希望电子出版社, 2002
- [13]封顺天, 可穿戴设备发展现状及趋势, [J], 中国工业经济, 2014 年(第三期)
- [14]赵斌.可穿戴设备设计趋势及策略研究[J].包装工程, 2015 年, 第二期.
- [15]李金昌.应用抽样技术「M」.北京:科学出版社, 2010.
- [16]贾俊平, 何晓群, 金勇进.统计学「M」.北京:中国人民大学出版社, 2009.

- [17]柯慧欣, 丁立宏.市场调查与分析「M」.北京:中国统计出版社, 2000.
- [18]王学民.应用多元分析「M」.上海:上海财经大学出版社, 2009.
- [19]达摩达尔.N.古扎拉第, 唐.C.波特.计量经济学基础「M」.北京:中国人民大学出版社, 2011.

附录一

随机数表

1	4	7	2	2	7	2	6	7	7	5	9	8	3	3	1
2	1	2	1	4	5	7	0	1	2	5	4	8	3	6	5
7	2	7	4	8	7	7	5	8	6	2	5	4	5	1	5
6	1	8	0	3	1	6	1	1	7	7	3	1	1	5	7
6	5	3	8	5	4	4	3	3	7	4	1	3	3	9	3
9	7	5	6	7	5	8	7	3	1	2	6	8	0	0	6
4	8	9	1	7	3	6	5	8	3	9	9	1	5	4	5
9	0	1	4	4	7	7	8	0	1	6	3	2	7	4	3
0	5	4	5	4	4	6	5	4	6	3	6	7	5	8	5
4	6	9	6	2	9	6	7	9	3	9	4	5	8	2	3
2	7	4	1	1	1	0	9	1	9	8	8	9	3	0	7
4	7	1	8	3	0	5	2	4	6	6	9	4	5	3	3
2	4	2	6	7	2	7	6	8	7	3	6	7	9	6	9
0	4	8	6	2	6	3	8	2	8	9	7	7	4	8	1
1	7	2	0	3	5	0	4	8	5	7	0	7	1	7	6
9	2	8	1	2	7	6	4	4	4	2	5	6	4	3	1

附录二

智能手环用户群及其发展前景调查问卷

尊敬的朋友:

您好!我们是中南财经政法大学的学生,为了参加“第六届全国大学生市场调查与分析大赛”,我们希望了解您关于智能手环的一些看法。本次调查将对您提供的信息进行严格保密,不用写调查者姓名,所有回答用于学术研究。您只需根据自己的实际情况填写本问卷。衷心感谢您的支持与配合!

注:无特别说明,请直接在选中答案前的“□”内打√。

问卷编号: _____ 调研员: _____ 时间: _____ 调研地点: _____

您的性别: ☐男 ☐女

您的年龄: ☐19岁及以下 ☐20-29岁 ☐30-39岁 ☐40-49岁 ☐50岁以上

最高学历: ☐初中及以下 ☐高中/中专/职校 ☐大学专科 ☐大学本科
☐硕士及以上

您的职业: ☐在校学生 ☐普通职员 ☐企业管理者 ☐政府/机关干部/公务员 ☐
☐自由职业者 ☐个体经营者/承包商 ☐普通工人 ☐商业服务业职工
☐农林牧渔劳动者 ☐退休 ☐其他

01.请问您或您的家人及亲朋好友是否有人在以下地方工作:

☐市场研究/广告公司 ☐电视台/电台/报纸/杂志等媒介机构
☐社情民意调查机构/咨询公司 ☐销售/商品研发部门 ☐无

02.请问您最近6个月内是否接受过智能可穿戴产品的市场研究访问?

☐是 ☐否

03.请问您是否听说过智能手环?

☐比较了解 ☐听说过,但具体不了解 ☐没听说过

04.请问您通过哪些渠道获得智能手环的相关信息?(多选)

☐科技网站 ☐产品官网 ☐购物网站 ☐现实朋友/亲人/同学告知
☐门户网站 ☐智能产品展览会 ☐社区论坛 ☐报纸/杂志 ☐电视
☐微博/微信朋友圈 ☐产品实体专卖店 ☐户外广告 ☐超市、电器商场

05.您目前的年收入范围是?

☐3万以下 ☐3-8万 ☐9-14万 ☐15-20万 ☐20万以上

06.您对自己的性格评价(多选)

☐冲劲拼搏 ☐幽默风趣 ☐勇敢 ☐平易随和 ☐开朗爱笑 ☐有责任感
☐其他

07.您的锻炼频率?

☐一周3-4次 ☐一周1次 ☐一个月1次 ☐我不喜欢运动

08.您希望智能手环具有以下哪种功能？（多选）

- ☐提供心率、血压监测数据
- ☐运动计步、计算消耗的卡路里
- ☐消息通知，如短信、来电、邮件、日期、闹钟等提醒功能
- ☐GPS 安全定位，以防儿童意外走失
- ☐具有交通卡、门卡、签到卡等磁卡的刷卡功能
- ☐提供噪音、温度、空气湿度等监测数据
- ☐远程控制家电、汽车等各种设备
- ☐这些功能我都不需要

9.您最感兴趣的休闲活动是？

- ☐摄影 ☐上网 ☐听音乐 ☐看书看报 ☐看电视 ☐运动健身

●以下问题为智能手环用户填写，非智能手环用户不用填这一部分

10.请问您的智能手环是通过何种渠道获得的？

- ☐自己购买 ☐亲人/朋友赠送的 ☐参与活动获得的

11.您的智能手环品牌是？

- ☐小米 ☐华为 ☐三星 ☐Fitbit ☐Nike+ ☐Jawbone ☐索尼 ☐咕咚
- ☐OPPO ☐其他

12.您对智能手环的满意度评价（请在相应的等级下打钩）：

		您对这些项目的满意程度				
		非常不满意	比较不满意	一般	比较满意	非常满意
外形和材质	颜值外观	1	2	3	4	5
	手环重量	1	2	3	4	5
	材质舒适程度	1	2	3	4	5
功能与服务	功能种类	1	2	3	4	5
	续航能力	1	2	3	4	5
	价格水平	1	2	3	4	5
	售后服务	1	2	3	4	5
使用体验	界面与操作体验	1	2	3	4	5
	使用方便程度	1	2	3	4	5
	数据质量	1	2	3	4	5

13.您目前的智能手环使用情况是？

- ☐还在使用中 ☐几乎没有用了

14.您是何时放弃使用智能手环的？（还在使用的用户不用填）

- ☐不到 1 个月就几乎没有用了 ☐使用 1-3 个月后就几乎没有用了
- ☐使用 3-6 个月后就几乎没有用了 ☐6 个月后就几乎没有用了

●以下问题为非智能手环用户填写，智能手环用户不用填这一部分

15.您目前没有购买智能手环的原因？

- ☐未来打算买，还在观望中 ☐价格高 ☐功能不实用
☐不能独立于手机使用 ☐我不喜欢智能手环，不考虑购买 ☐其它

16.对于智能手环，定价多少你会考虑购买？

- ☐200 元以下 ☐200-500 元 ☐501-1000 元 ☐1001-2000 元 ☐2000 元以上

17.您会因为智能手环具备下列的哪些特点而选择购买？（多选）

- ☐佩戴舒适 ☐外观精美，时尚 ☐功能齐全
☐电池续航时间长 ☐质量好 ☐价格合理
☐可监测健康状况 ☐其他

我们的问卷到此结束，衷心感谢您的参与！

附录三

问卷编码表

变量序号	变量名	变量类型	变量所	取值范围	选项	对应题号	对应问题
1	Q1	数值型	1	0-2	1-男 2-女 0-缺失	0	性别
2	Q2	数值型	1	0-5	1-19 岁及以下 2-20-29 岁 3-30-39 岁 4-40-50 岁 5-50 岁以上 0-缺失	0	年龄
3	Q3	数值型	1	0-5	1-初中及以下 2-高中/中专/职校 3-大学专科 4-大学本科 5-硕士及以上 0-缺失	0	最高学历
4	Q4	数值型	1	0-11	1-在校学生 2-普通职员 3-企业管理者 4-政府/机关干部/公务员 5-自由职业者 6-个体经营者/承包商 7-普通工人 8-商业服务业职工 9-农林牧渔劳动者 10-退休 11-其他 0-缺失	0	职业
5	Q5	数值型	1	0-5	1-市场研究/广告公司 2-电视台/电台/报纸/杂志	1	家人及亲朋好友是否有人在以下地方工作

		型			等媒介机构 3-社情民意调查机构/咨询公司 4-销售/商品研发部门 5-无 0-缺失		
6	Q6	数值型	1	0-3	1-是 2-否 0-缺失	2	最近 6 个月内是否接受过智能可穿戴产品的市场研究访问
7	Q7	数值型	1	0-4	1-比较了解 2-听说过,但具体不了解 3-没听说过 0-缺失	3	是否听说过智能手环
8	Q8	数值型	1	0-13	1-科技网站 2-产品官网 3-购物网站 4-现实朋友/亲人/同学告知 5-门户网站 6-智能产品展览会 7-社区论坛 8-报纸/杂志 9-电视 10-微博/微信朋友圈 11-产品实体专卖店 12-户外广告 13-超市、电器商场 0-缺失	4	哪些渠道获得智能可穿戴设备的相关信息
9	Q9	数值型	1	0-5	1-3 万以下 2-3-8 万 3-9-14 万 4-15-20 万 5-20 万以上 0-缺失	5	年收入范围
10	Q10	数	1	0-7	1-冲劲拼搏	06	对自己的性格评价

		值 型			2-幽默风趣 3-勇敢 4-平易随和 5-开朗爱笑 6-有责任感 7-其他 0-缺失		
11	Q11	数 值 型	1	0-4	1-一周 3-4 次 2-一周一次 3-一个月 1 次 4-我不喜欢运动 0-缺失	07	锻炼频率
12	Q12	数 值 型	1	0-8	1-提供心率、血压监测数据 2-运动计步、计算消耗的卡路里 3-消息通知，如短信、来电、邮件、日期、闹钟等提醒 4- GPS 安全定位，以防儿童意外走失 5-具有交通卡、门卡、签到卡等磁卡的刷卡功能 6-提供噪音、温度、空气湿度等监测数据 7-远程控制家电、汽车等各种设备 8-这些功能我都不需要 0-缺失	08	希望智能手环具有哪种功能
13	Q13	数 值 型	1	0-7	1-摄影 2-运动健身 3-听音乐 4-看书看报 5-看电视 6-上网	09	最感兴趣的休闲活动

					7-其他		
					0-缺失		
14	Q14	数值型	1	0-4	1-自己购买 2-亲人/朋友赠送的 3-参与活动获得的 0-缺失	10	用户的智能手环是通过何种渠道获得的
15	Q15	数值型	1	0-10	1-小米 2-华为 3-三星 4-Fitbit 5-Nike+ 6-Jawbone 7-索尼 8-咕咚 9-OPPO 10-其他 0-缺失	11	用户智能手环品牌
16	Q16.1	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环颜值外观满意度
17	Q16.2	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环重量满意度
18	Q16.3	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环材质舒适度满意度

19	Q16.4	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环功能种类满意度
20	Q16.5	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环续航能力满意度
21	Q16.6	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环价格水平满意度
22	Q16.7	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环售后服务满意度
23	Q16.8	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环界面与操作体验满意度
24	Q16.9	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意	12	对智能手环使用方便程度满意度

					5-非常满意		
					0-缺失		
25	Q16.10	数值型	1	0-5	1-非常不满意 2-不满意 3-一般 4-满意 5-非常满意 0-缺失	12	对智能手环数据质量满意度
26	Q17	数值型	1	0-2	1-还在使用 2-几乎没有用了 0-缺失	13	目前的智能手环使用情况
27	Q18	数值型	1	0-4	1-不到 1 个月就几乎没有用了 2-使用 1-3 个月后就几乎没有用了 3-使用 3-6 个月后就几乎没有用了 4-6 个月后就几乎没有用了 0-缺失	14	何时放弃使用智能手环
28	Q19	数值型	1	0-6	1-未来打算买，还在观望中 2-价格高 3-功能不实用 4-不能独立于手机使用 5-其他 6-我不喜欢智能手环，不考虑购买 0-缺失	15	目前没有购买智能手环的原因
29	Q20	数值型	1	0-5	1-200 元以下 2-200-500 元 3-500-1000 元 4-1001-2000 元 5-2000 元以上 0-缺失	16	对于智能手环，定价多少你会考虑购买

30	Q21	数值型	1	0-8	1-佩戴舒适 2-外观精美，时尚 3-功能齐全 4-电池续航时间长 5-质量好 6-价格合理 7-可监测健康状况 8-其他 0-缺失	17	会因为智能手环具备的哪些特点而选择购买
----	-----	-----	---	-----	--	----	---------------------

附录四

小米手环客户群二元选择模型变量表

变量名	变量符号	变量定义
是否拥有智能手环	Y	$Y = \begin{cases} 1 & \text{拥有} \\ 0 & \text{没有} \end{cases}$
性别	①女 Female ②男	$\text{Female} = \begin{cases} 1 & \text{女} \\ 0 & \text{男} \end{cases}$
年龄	①19岁及以下 Age19 ②20-29岁 Age20 ③30-39岁 Age30 ④40-49岁 Age40 ⑤50岁及以上	$\text{Age19} = \begin{cases} 1 & 19 \text{岁及以下} \\ 0 & \text{其他} \end{cases}$ $\text{Age20} = \begin{cases} 1 & 20 - 29 \text{岁} \\ 0 & \text{其他} \end{cases}$ $\text{Age30} = \begin{cases} 1 & 30 - 39 \text{岁} \\ 0 & \text{其他} \end{cases}$ $\text{Age40} = \begin{cases} 1 & 40 - 49 \text{岁} \\ 0 & \text{其他} \end{cases}$ 50岁以上为基础类型
最高学历	①初中及以下 Mid-schl ②高中/中专/技校 High-schl ③大学专科 3Ycollege ④大学本科 College ⑤硕士及以上	$\text{Mid-schl} = \begin{cases} 1 & \text{初中及以下} \\ 0 & \text{其他} \end{cases}$ $\text{High-schl} = \begin{cases} 1 & \text{高中/中专/技校} \\ 0 & \text{其他} \end{cases}$ $\text{3Ycollege} = \begin{cases} 1 & \text{大学专科} \\ 0 & \text{其他} \end{cases}$ $\text{College} = \begin{cases} 1 & \text{大学本科} \\ 0 & \text{其他} \end{cases}$ 硕士及以上为基础类型
职业	①在校学生 Student ②普通职员 Clerk ③企业管理者 Executive ④政府/机关干部/公务员 Gov ⑤自由职业者 Freelance ⑥个体经营者/承包商	$\text{Student} = \begin{cases} 1 & \text{在校学生} \\ 0 & \text{其他} \end{cases}$ $\text{Clerk} = \begin{cases} 1 & \text{普通职员} \\ 0 & \text{其他} \end{cases}$ $\text{Executive} = \begin{cases} 1 & \text{企业管理者} \\ 0 & \text{其他} \end{cases}$ $\text{Gov} = \begin{cases} 1 & \text{政府/机关干部/公务员} \\ 0 & \text{其他} \end{cases}$ $\text{Freelance} = \begin{cases} 1 & \text{自由职业者} \\ 0 & \text{其他} \end{cases}$ $\text{Businessman} = \begin{cases} 1 & \text{个体经营者/承包商} \\ 0 & \text{其他} \end{cases}$

	Businessman		
	⑦普通工人	Worker =	$\begin{cases} 1 & \text{普通工人} \\ 0 & \text{其他} \end{cases}$
	Worker	Service clerk =	$\begin{cases} 1 & \text{商业服务业职工} \\ 0 & \text{其他} \end{cases}$
	⑧商业服务业职工	Worker =	$\begin{cases} 1 & \text{农林牧渔劳动者} \\ 0 & \text{其他} \end{cases}$
	Service clerk		
	⑨农林牧渔劳动者		退休为基础类型
	farmer		
	⑩退休		
	Retir		
年收入	①3 万以下	Income3 =	$\begin{cases} 1 & \text{3 万以下} \\ 0 & \text{其他} \end{cases}$
	Income3	Income8 =	$\begin{cases} 1 & \text{3 - 8 万} \\ 0 & \text{其他} \end{cases}$
	②3-8 万	Income14 =	$\begin{cases} 1 & \text{9 - 14 万} \\ 0 & \text{其他} \end{cases}$
	Income8	Income20 =	$\begin{cases} 1 & \text{15 - 20 万} \\ 0 & \text{其他} \end{cases}$
	③9-14 万		
	Income14		
	④15-20 万		20 万以上为基础类型
	Income20		
	⑤20 万以上		
兴趣爱好	①摄影	Photo =	$\begin{cases} 1 & \text{摄影} \\ 0 & \text{其他} \end{cases}$
	Photo	Sports =	$\begin{cases} 1 & \text{运动健身} \\ 0 & \text{其他} \end{cases}$
	②运动健身	Music =	$\begin{cases} 1 & \text{听音乐} \\ 0 & \text{其他} \end{cases}$
	Sports	Book =	$\begin{cases} 1 & \text{看书看报} \\ 0 & \text{其他} \end{cases}$
	③听音乐	TV =	$\begin{cases} 1 & \text{看电视} \\ 0 & \text{其他} \end{cases}$
	Music		上网为基础类型
	④看书看报		
	Book		
	⑤看电视		
	TV		
	⑥上网		
	Internet		
期望手环的功能	①提供心率、血压监测数据	HM =	$\begin{cases} 1 & \text{提供心率、血压监测数据} \\ 0 & \text{其他} \end{cases}$
	HM(Health-monitoring)	Pedometer =	$\begin{cases} 1 & \text{运动计步、计算消耗的卡路里} \\ 0 & \text{其他} \end{cases}$
	②运动计步、计算消耗的卡路里		
	Pedometer		
	③消息通知，如短信、来电、		

	邮件、日期、闹钟等提醒 reminder	Reminder = $\begin{cases} 1 & \text{消息通知, 如短信、} \\ & \text{来电、} \\ & \text{邮件、日期、闹钟等提醒} \\ 0 & \text{其他} \end{cases}$
	④ GPS 安全定位, 以防儿童意外走失	
	GPS	GPS = $\begin{cases} 1 & \text{GPS 安全定位, 以防} \\ & \text{儿童意外走失} \\ 0 & \text{其他} \end{cases}$
	⑤ 具有交通卡、门卡、签到卡等磁卡的刷卡功能 Magcard	
	⑥ 提供噪音、温度、空气湿度等监测数据	Magcard = $\begin{cases} 1 & \text{具有交通卡、门卡、签} \\ & \text{到卡等磁卡的刷卡功能} \\ 0 & \text{其他} \end{cases}$
	EM(Environmental monitoring)	
	⑦ 远程控制家电、汽车等各种设备	EM = $\begin{cases} 1 & \text{提供噪音、温度、空气} \\ & \text{湿度等监测数据} \\ 0 & \text{其他} \end{cases}$
	RM(Remote-control)	
	⑧ 这些功能我都不需要	RM = $\begin{cases} 1 & \text{远程控制家电、汽车等} \\ & \text{各种设备} \\ 0 & \text{其他} \end{cases}$
	这些功能我都不需要为基础类型	
拥有的智能 可穿戴设备	① 智能手表	Watch = $\begin{cases} 1 & \text{智能手表} \\ 0 & \text{其他} \end{cases}$
	Watch	
	② 智能手环	Band = $\begin{cases} 1 & \text{智能手环} \\ 0 & \text{其他} \end{cases}$
	Band	
	③ 运动相机	Camera = $\begin{cases} 1 & \text{运动相机} \\ 0 & \text{其他} \end{cases}$
	Camera	
	④ 智能眼镜	Glasses = $\begin{cases} 1 & \text{智能眼镜} \\ 0 & \text{其他} \end{cases}$
	Glasses	
	⑤ 智能耳机	Headphone = $\begin{cases} 1 & \text{智能耳机} \\ 0 & \text{其他} \end{cases}$
	Headphone	
	⑥ 暂无	暂无为基础类型
	Nothing	
性格	① 冲劲拼搏	Aggressive = $\begin{cases} 1 & \text{冲击拼搏} \\ 0 & \text{其他} \end{cases}$
	Aggressive	
	② 幽默风趣	Funny = $\begin{cases} 1 & \text{幽默风趣} \\ 0 & \text{其他} \end{cases}$
	Funny	
	③ 勇敢	Brave = $\begin{cases} 1 & \text{勇敢} \\ 0 & \text{其他} \end{cases}$
	Brave	
	④ 平易随和	Amiable = $\begin{cases} 1 & \text{平易随和} \\ 0 & \text{其他} \end{cases}$

	Amiable ⑤ 开朗爱笑 optimistic ⑥ 有责任感 responsible	Optimistic = $\begin{cases} 1 \\ 0 \end{cases}$	开朗爱笑 其他
锻炼频率	① 一周 3-4 次 3or4-1W ② 一周 1 次 1-1W ③ 一个月 1 次 1-1M	$3\text{or}4-1W = \begin{cases} 1 \\ 0 \end{cases}$ $1-1W = \begin{cases} 1 \\ 0 \end{cases}$ $1-1M = \begin{cases} 1 \\ 0 \end{cases}$	一周 3-4 次 其他 一周 1 次 其他 一个月 1 次 其他
		我有责任感为基础类型 我不喜欢运动为基础类型	

致 谢

在准备第六届全国大学生市场调查分析大赛的比赛过程中，我们团队以“智能手环使用现状及产品发展分析报告——基于电商平台消费者购买评价的文本挖掘与武汉市的抽样调查”为题，对天猫和京东电商平台小米手环用户评论和武汉市的智能手环用户和潜在用户状况进行了调查研究，历时近两个月。在此，我们团队成员向调查研究过程中给予帮助的各位老师和所有的受访者表示由衷的谢意。

张海波老师在我们整个调查过程中，包括前期选题、调查实施、后期编写报告时都给予了许多有用的信息与指导，指出我们的方案所存在的盲点和漏洞，帮助我们完善了研究思路；并且在问卷设计过程中，多次对我们问卷的措辞、完整性等方面提出了专业性的修改意见；感谢张海波老师的悉心指导！在调查过程中也得到了大多数居民的支持，他们耐心地填写问卷并给我们提供很多关于智能手环认知和使用情况的信息，让我们更好的完成本次调查！。

在本次调研过程中，我们团队成员收获颇丰，无论是对理论知识和实际应用都有了更深的了解！谢谢本次比赛给我们带来的收获与成长！