

基于数据挖掘的大学生网络购物行为研究

张晓芳

(福州理工学院, 福州 350014)

摘要:以福州高校大学生为调查对象,研究大学生网络购物的行为特征,采用数据挖掘中的聚类和关联进行实证分析,得出网络购物年限、购买频数、购买金额、信誉度等之间的关系,最后提出应该针对90后大学生进行精准营销、加强网络购物的安全诚信体系、社会各阶层应引导大学生培养理性网购行为等方面的建议。

关键词:大学生;网络购物;数据挖掘;对策建议

中图分类号:F713.36 **文章标志码:**B **文章编号:**1671-1807(2017)03-0115-04

随着科技的发展和时代的进步,互联网在全球的普及越来越广,互联网与经济社会也得到深度的融合。截至2015年12月,我国网络购物用户规模达到4.13亿,从年龄结构来看,20-29岁的网民人数占总网民人数的29.9%,位居最高比例。从职业结构来看,学生人数比例最高,为25.2%,较2014年提高了5.8个百分点。大学生群体已经成为推动网络购物发展的重要力量^[1]。本文针对福州高校大学生,通过问卷调查的形式,了解大学生网络购物的影响因素和行为特征,不仅有利于引导大学生的健康消费理念,更有利于网络商家改善网络营销水平,提升我国网络购物的整体环境。

1 基于SPSS Clementine12.0数据挖掘技术

表1 网购年龄与网购频次的部分对应数据表(前8个样本)

网购年龄	5年以上	5年以上	1~3年	1~3年	3~5年	3~5年	5年以上	3~5年
网购频次	5次以上	5次以上	2~3次	5次以上	4~5次	5次以上	5次以上	2~3次

由表1抽取的前8个观察值大致可以看出,网购年龄越大,月网购频次似乎也就越多,但这种判断并不十分准确。为了更清楚的挖掘出二者之间的关系,我们通过聚类分析来进行处理。通过数据处理,我们筛选出网购年龄和月网购频次这两个字段。

用聚类方法分析网购年龄与月网购频次的目的是分析出购买频率与网购年龄之间的关系,从而确定出每个网购年龄所对应的网购频次,以区别出高频次网购人群和低频次网购人群。由于聚类簇数不易确

的统计分析

此次针对福州高校大学生进行网络问卷调查,涉及福州大部分公办和民办高校,范围包括高职高专和本科院校,回收的有效样本共计711个,其中没有过网购经历的样本数为13个,充分表明绝大部分的被调查对象已经成为网购的一员,由于本次调查主要是针对有网购体验的人群,因此接下来的分析中我们将剔除13个没有网购体验的样本,仅保留698个有过网购体验的样本。

1.1 购买频次与网购年龄的聚类分析

首先我们来分析一下,网购年龄与网购频次的关系。

定,因此,采用Twostep算法进行聚类分析,该算法能在指定范围内自动确定聚类簇数。Twpstep算法采用层次凝聚法进行聚类,通过两步完成聚类。第一步预聚类,即逐条扫描拟聚类的数据,根据距离判定是归于已形成的Clusters,还是形成一个新的Cluster,该步通常采用的是一些传统的聚类方法;第二步聚类,将预聚类形成的Sub-Clusters作为输入,该步采用“层次凝聚法”,递归合并输入的Sub-Clusters,一直到合并到最后一个Cluster包含了所有的记

收稿日期:2016-11-03

基金项目:2015年福建省中青年教育科研社科项目(JAS150742)。

作者简介:张晓芳(1983-),女,福建沙县人,福州理工学院,讲师,硕士,研究方向:市场营销。

录^[2]。

我们在 Clementine12.0 平台选择 Twostep 算法,结果如下:



图 1 购买频次与网购年龄的聚类分析结果

由图 1 可以看出, Twostep 算法进行聚类分析得到了四个 Cluster。第一个聚类结果包含 194 个样本,这 194 个样本的人均网购年龄为 3~5 年,月均网购频次为 4~5 次,无论是人均年龄还是月均网购频次,都大于其它 3 个聚类,表明我们的推断是合理的,网购年龄越大,一般情况下月均网购频次也就越多。

艾瑞咨询公司 2009 年中国互联网用户网购购物行为及满意度调查报告中指出,“越早接触电子商务的用户,网购次数越频繁。”其调查数据还显示“培养用户频繁网购习惯至少需要 3 年”^[3]。因此,本次调查的结果与艾瑞咨询公司得出的是一致的。当消费者接触网购并养成习惯后,网购的频率并会随着时间的推移而增加,这里除了消费者对网购的依赖,还源于消费者对网购的忠诚。对于年轻一代的大学生而言,这种依赖和忠诚是能够通过行之有效的培养而形成的。

此外,引人注意的是第二个聚类结果中 111 个样本的人均网购年龄不到 3 年,到月均网购频次却达到了 4~5 次,本文认为这一群体受到电商的影响更大,网购能力、网购意识与网购意愿较强,是未来网购群体的重要力量。这一类大学生属于意识相对超前,更容易接受新鲜事物的一代,对于这类大学生,网络商家应该有针对性的采取营销策略进行维护和培育,提高忠诚度。

第三个聚类和第四个聚类中样本量虽然达到了 424,但月均网购频次却不到 3 次。四个聚类结果表明网购频次与网购年龄存在一定的正向关系,而且调查样本中,月均网购频次的分布并不是均匀的,差异十分明显。

1.2 网购年龄与网购金额的聚类分析

我们探讨了网购年龄与网购频次之间的关系,即网购时间越长,网购次数越频繁。那么网购年龄与网购金额之间是否也存在类似的关联呢?本节将采用聚类分析方法对这一问题进行探讨,以区别出高消费网购人群和低消费网购人群所对应的网购年龄。同样我们还是采用 Twostep 算法,结果如下:



图 2 网购金额与网购年龄的聚类分析结果

由图 2 可以看出, Twostep 算法进行聚类分析得到了 3 个 Cluster。由于篇幅限制,这里我们进行简单的分析,由分析结果可以看出,第一个聚类结果中包含 246 个样本,样本平均网购年龄为 3~5 年,月均支出金额为 300~400,与其它两个聚类结果相比,支出金额最高,网购年龄也比较大,表明网购年龄与网购支出方面确实存在着正向的关系。这也说明,一个人的网购年龄越大,对网购的“忠实感”越强。第三个聚类结果表明有高达 292 个样本的人均年网购年龄不足 3 年,月均网购支出也不到 300,也表明了网购年龄与网购支出方面确实存在着正向联系。

这启示电商企业一定要提供优质的服务,培养顾客的忠诚感,让顾客真正喜欢网购,这样才能让顾客愿意在网购上支出,才能促进电商企业的发展。

在 Twostep 算法的基础上,我们采用 K-Means 算法, K-Means 算法作为 Twostep 算法的补充,是一种常用的基于划分的聚类方法,“它根据最终分类的个数 k 随机地选取 k 个初始的聚类中心,不断地迭代,直到达到目标函数的最小值,即得到最终的聚类

结果。该算法主要用来处理连续的数值变量,最终分类的个数 k 由用户主观确定”^[4]。由于 Twostep 算法已经将聚类结果分为了三种,故在此我们将 k 设定为 3, K-Means 算法的结果如下:



图3 网购金额与网购年龄的 K-Means 聚类分析结果

由图 3 的聚类结果可以看出,与 Twostep 算法相比,两者的基本结论并无显著差别,以第 2 个聚类结果为例,包含 349 个样本,无论是人均网购年龄还是网购支出都偏低,这与 Twostep 算法的结论基本一致,这里我们不再过多阐述。

1.3 信誉度与网购频次、网购年限的关联分析

我们知道,网购过程中,买家十分在意卖家的信誉,那么,信誉度在大学生们的购买决策中到底发挥了多大作用,本研究将对这一问题进行探讨。

1) 数据处理。由表 1 的统计结果可以看出,网购年限多在 1 年以上,占比达 90.11%,因此具有代表性的样本的网购年限都在一年以上,我们将网购年限在 1 年以下的设置为 F,其余的设置为 T。而由表 1 的结果可以看出,月网购次数大部分都在 2 次以上,1 次以下的仅占 9.31%,不具代表性,因此我们将月网购次数在 2 次以上的设置为 T,将 1 次以下的设置为 F。对于信誉这一指标的处理,我们依据“对于同款商品在多家网店同时有卖的情况下,您选择的依据是?”这个问题的回答,如果被调查对象选择的是信誉,则为 T,其它的为 F。

2) 关联分析。在 SPSS Clementine12.0 中选择 Carma 算法,将“月网购频次”、“网购年龄”和“信誉”做为输入变量,将数据方向设为“双向”,设置最低条件支持度为“20%”,设置最低规则置信度为“20%”,最大前项数为“10”。然后执行数据流,查看执行结果,得到条关联规则,从中挑选出后项只包含信誉字

条的规则共 3 条,具体数据如表 3 所示。

表 2 关联分析训练集部分数据表

月网购频次	网购年龄	信誉
T	T	T
T	T	T
T	T	T
T	T	F
T	F	F
F	T	F
T	F	F

表 3 关联分析表

后项	前项	支持度%	置信度%	提升
信誉	月网购频次 网购年龄	61.19	63.42	1.003
信誉	网购年龄	66.57	63.1	0.998
信誉	网购频次	93.02	62.8	0.993

用 Carma 算法分析得到的三条规则中,第二条和第三条规则的提升值分别为“0.998”、“0.993”,均小于 1,没有意义,因此舍弃。下面我们只对第一条规则进行讨论分析。第一条规则的前项为月网购频次和网购年龄两个属性,后项为信誉属性,其置信度为 61.19%,支持度也达到了 63.42%。由此可以分析得知,同时满足网购年龄 1 年以上、月网购频次 2 次以上的大学生最容易受电商信誉的影响,并且其占比还是相当大的(通过数据统计表明,同时满足这两个条件的占比达到了 70%以上)。这个结果给我们的启示是,有过一年以上并月购次数在 2 次以上的大学生对网络购物的感受相对比较成熟,也了解通过何种方法能够在网络中购买到物美价廉的心仪产品,对于他们来说,信誉度是能够实现上述目的的最佳选择。

2 相关对策建议

2.1 网络商家应针对大学生的特性做精准营销

以上我们结合问卷数据对大学生网购的行为进行了分析,他们作为新生代的网购力量,对网购市场的贡献不可小觑,也已经受到越来越多商家的追捧和青睐。网络商家可以分析大学生网络购物行为特征,有针对性的采取营销策略,做精准营销。首先,网购年限越长的的大学生,每个月网购次数越多,支出也越多,针对这类大学生,更多考虑的是通过会员等方式留住他们,长期做好客户关系的维护,使其成为忠诚用户。针对网购年限较短的学生,采取各种大学生青睐的优惠措施让利给他们,主要的目的是吸引大学生

的注意力并购买;其次,商家应该调查大学生平时喜欢逛的网站有哪些,使用哪些 APP,搭建网络销售渠道,进行重点推广;第三,做好关键字搜索的优化,减少大学生在线选购商品的时间,降低搜索成本;第四,管理好买家评价,提高卖家销量和信誉度,形成良性循环。我们希望能够对大学生做比较准确的定位,在大学生关注的各方面下功夫。

2.2 网络平台及商家应加强网络购物的安全诚信体系建设

由此可以分析得知,同时满足网购年龄 1 年以上、月网购频次 2 次以上的大学生最容易受电商信誉的影响,并且其占比还是相当大的(通过数据统计表明,同时满足这两个条件的占比达到了 70% 以上)。这个结果给我们的启示是,有过一年以上并月购次数在 2 次以上的大学生网络购物的感受相对比较成熟,也了解通过何种方法能够在网络中购买到物美价廉的心仪产品,对于他们来说,信誉度是能够实现上述目的的最佳选择。因此,网络购物的安全诚信体系建设是大学生长期选择网购的保障。首先,网络商家应该重视商品信息的诚信建设。网络商品虚假信息问题成为阻碍大学生网络购物的重要因素,很多商家虚构商品信息,以假乱真,通过托人购买、朋友互刷、多号自刷等手段来积累信誉,但是一旦大学生上当受骗一次,那么很容易带来负面影响,让大学生对网络购物失去信心,甚至杜绝再次网购。其次,完善退换货服务,真正落实 7 天无理由退换货,让他们放心。最后,确保网络支付安全。大学生每月用于购物的资金有限,所以更加重视网络支付安全和购物安全,为了给他们长期的依赖网络购物,支付安全是必须的支

撑体系之一^[5]。

2.3 社会各阶层应引导大学生培养理性网购行为

网络购物已经是当代大学生群体购物的主要形式,未来将会有越来越多大学生成为网购大军中的一员。大学生网购行为的不合理不仅会影响自身健康的学习和生活,还会影响到自身理性价值观和消费者的形成,因此,社会各阶层应想方设法引导大学生培养合理的网购行为。首先,国家应从宏观层面的政策对大学生的购物行为进行正确的引导,如通过立法对网购行为的引导和规范。其次,网购市场自身的规范化管理。在网购中遇到的各种售后问题,应该进一步明确行业规范,加强网络商家的规范管理。最后,需要从学校、家庭的角度在微观层面对大学生进行教育引导,学校和家庭应给大学生灌输文明消费和理性消费的观念,与大学生多沟通交流,了解大学生的日常消费情况,通过调整生活费的供给,帮忙大学生养成理性的消费习惯。

参考文献

- [1] CNNIC 发布第 37 次中国互联网络发展状况统计报告[EB/OL]. (2016-01-22)[2016-07-28]. <http://tech.sina.com.cn/i/2016-01-22/doc-ifxnuvvh5133709.shtml>.
- [2] 西蒙(SOMAN K P),等.数据挖掘基础教程[M].范明,牛常勇,译.北京:机械工业出版社,2009:1-6.
- [3] 彭剑芳.基于数据挖掘的网络购物行为分析[D].昆明:云南大学,2011.
- [4] 龚曙明.应用统计学[M].2版.北京:清华大学出版社,2005:20-25.
- [5] 陈鸿坤.大学生网络消费行为与营销策略[J].农村经济与科技,2016,27(6):130-132.

Research on College Students' Online Shopping Behavior Based on Data Mining

ZHANG Xiao-fang

(Fuzhou Institute of Technology, Fuzhou 350014, China)

Abstract: College students in Fuzhou as research object, research on behavior characteristics of college students online shopping, analyzes the clustering and correlation of data mining, the online shopping experience, purchase frequency and purchase amount, the relationship between reputation, finally should needle of precision marketing, strengthen the safety and integrity of the system, the network shopping classes the society should guide students to cultivate rational online shopping behavior and other aspects of the proposal of the post-90s College students.

Key words: college students; online shopping; data mining; countermeasures and suggestions