

基于校园一卡通消费行为分析 ——以宣城职业技术学院为例

刘训星 姜峰 龚勇

(宣城职业技术学院, 安徽 宣城 242000)

摘要: 随着智慧校园的建设, 一卡通在校园中发挥重要的作用, 在一卡通中存放大量数据, 本文中使用了 weka 调用 Apriori 算法, 分析我校贫困生消费情况, 同时对全校消费额占比分析和刷卡次数占比分析, 为学校管理层在降低学生生活成本和提供高效服务两方面提供数据支持。

关键词: 智慧校园; 一卡通; 消费行为分析

一、引言

宣城职业技术学院经过这几年的快速发展, 信息化技术得到长足的发展, 随着德处项目不断推进, 学校在基础建设、管理多方面将得到很大提高, 其中智慧校园建设是德处项目的重要组成部分, 校园一卡通将是智慧校园建设实现的基础, 包含学生大量信息, 具有消费功能和身份识别功能。

本文通过对宣城职业技术学院校园一卡通产生的数据获取、处理、分析, 获取当前学生消费行为, 通过关联性因素挖掘分析, 获得有效数据, 可以依据对数据挖掘分析的结果, 有助于学生管理、有助于提高服务水平, 为我校智慧校园的建设提供更有价值的信息。

二、消费行为分析

Apriori 算法能够通过数据的分析挖掘出强关联, 在 Apriori 算法中支持度是重要指标用来表示集中项一同出现的概率大小, 当概率小时说明两个数据集间的关联小, 当概率为百分之百时前一个项集直接影响后面一个项集。Apriori 算法在关联规则挖掘重要算法, 主要任务有频繁项集产生和关联规则产生。

三、算法介绍

(一) Apriori 关联规则指标

从样本中能挖掘大量的关联规则, 现实中在大量关联规则中只有部分关联规则是有效的。造成无效的原因主要在两个方面: 一个是没有达到足以信服程度, 另一个是所产生的规则只能很小范围使用。所以在判断一条挖掘的规则是否有价值, 有不少衡量指标, 在这其中支持度和置信度是两条重要衡量指标。

支持度:

$$Support(X \rightarrow Y) = P(X \cap Y) / P(I) \quad (1)$$

$P(X \cap Y)$ 表示发生 X 同时发生 Y 的个数, $P(I)$ 表示总共发生的个数, $Support(X \rightarrow Y)$ 表示 X 和 Y 同时发生占总发生数的比例。

置信度:

$$Confidence(X \rightarrow Y) = P(X \cap Y) / P(X) \quad (2)$$

$P(X \cap Y)$ 表示在项集同时发生 X 和 Y 的个数。 $P(X)$ 表示是在项集中只发生 X 的个数。 $Confidence(X \rightarrow Y)$ 表示在项集中发生 X 的个数中同时发生 X 和 Y 所占比例大小。

(二) Apriori 关联规则操作步骤

(1) 频繁项集的产生与发现。步骤一: 构建单元项集列表, 可以通过构造相应的数据, 然后将数据元素转换位不可变的集合, 并将其存储到列表里; 步骤二选择满足支持度要求数据集合,

创建频繁项集同时记录任意元素支持度并存放字典中, 删除不符合支持度要求的元素; 步骤三寻找符合支持度要求组合集合, 创建频繁我们所需项集列表, 找出符合支持度要求集合列表。

(2) 频繁项集发现关联规则。

(三) 实验的数据获取

从宣城职业技术学院校园一卡通数据存放数据库中最新数据导出, 从而用于实验。打开存放数据服务器的 sql server 2005, 从中找到存放数据的数据库。

从数据库找到存放数据的表, 从中我们就能看到一卡通在使用过程中产生的数据了。

为了使得获取的数据有助于数据挖掘过程中的使用, 将存放在服务器中一卡通的数据以 .csv 格式导出数据。在 SQL Server Management Studio 窗口中找到新建查询命令, 打开新建查询窗口, 在窗口中输入查询 SQL 语句; 执行查询命令; 将查询结果另存为制定格式文件 (.csv)。

三、一卡通中学生消费数据的分析和挖掘

在服务器中获取 9257 条一卡通原始消费数据。

(一) 按消费场所的角度分析

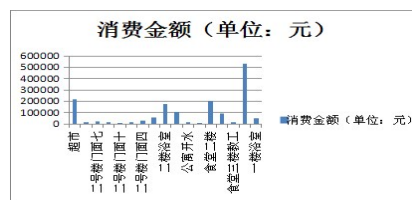


图1 消费金额

通过获取 2018 年 9 月 1 日 - 2018 年 12 月 31 日一卡通数据, 通过分析处理: ①在此期间一共消费 1539678.41 元; ②消费额占总额前三位的是食堂一楼、超市、食堂二楼。为此供学校管理层提供基础数据, 严格管理食堂与超市, 有助于降低学生生活成本, 特别是生活困难的学生; ③一卡通共刷卡次数为 1048575 次。④刷卡次数前三位是二楼浴室、一楼浴室、开水房。有助于管理层, 重视热水供应过程中安全问题, 同时注意浴室设施设备是否安全够用, 设备是否需要更新, 改善学生生活水平; ⑤二楼门面八和二楼门面十二刷卡次数和消费额都较低, 有助于管理层引进更有竞争力服务商, 从而更好为学校师生提供服务。

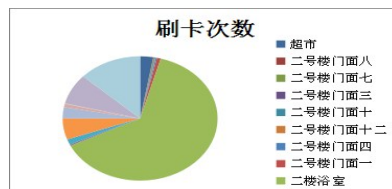


图2 刷卡次数

(二) 使用软件 weka 3.8.0 挖掘数据

(1) 数据预处理。在众多数据中有少数一卡通消费次数和消费额都非常低, 可能是走读生或卡片丢失重新补办, 同时目前数

(下转第 33 页)

多消费者购买。

(4)口碑相传,最典型的就是微信朋友圈里做微商的朋友。他们通过购买和使用某种产品,带动身边的某些朋友参与到互动中。“粉丝经济”的命脉就是“粉丝应援互动平台”,平台里的每位人员,大多是具有强消费能力和高活跃度的“粉丝团”,除具有消费能力外,甚至可以在平台里对接粉丝站,与明星面对面交流。

(5)开启驱动新模式,是“粉丝经济”的主要助推力。家电行业不甘落伍,率先在电商行业做好宣传工作,而后以派对形式等“粉丝经济”拉动群众消费。例如,“格兰仕”在冬季联合乐视跨界举办超级品牌高端家电狂欢派对,与5万名“西红柿粉丝”如期“约会”,将免费参观工厂、举办游园派对和买家电密切结合在一起,家电企业对玩跨界游戏驾轻就熟,但联合业界精英共同拓展家电营销新领域的活动,实属罕见。该活动以强大的阵容,实现了节能消费。

(6)搭建平台。从老牌的贴吧、粉丝网平台算起,目前市场上的粉丝互动平台如雨后春笋般纷至沓来。“阿里”重金入局,通过近两年的“集五福、送红包”活动,将阿里星球的事业推向了新高度,支付宝软件的注入,使得“粉丝经济”的大船承载重金。在如火如荼的网络经济“粉丝”大战中,获益最多的是商家。

四、偶像粉丝经济未来发展趋势

随着国内养成内偶像的不断地发展,未来“偶像+电商”的商业模型也会走进视野,实际上,目前网红早已是“抖音”“微博”的主流,而养成型偶像更像是一种高端性网红,通过对偶像的引流,相信以后偶像也会逐渐在电商上占有一席之地。现在就已经有

很多明星在小红书上分享产品,为不少品牌带来盈利。

粉丝也会逐渐成为一种职业,粉丝也会带来偶像的经济效应,不少粉丝由于有着良好的拍照技术拥有大量偶像粉丝的关注,饭圈称之为站子,不少站子通过跑行程、拍图、拍视频然后放在网络媒体上吸引偶像粉丝的点击观看,同时将自己拍的图制作,如PB(photobook)手幅、台历等,甚至还会以展览的形式进行贩卖赚钱。

目前,我国偶像粉丝经济处于发展阶段,粉丝经济背后群体是“90后”,其购买力会越来越强,粉丝经济市场空间仍然很大,新媒体时代也会对偶像的传播宣传有很大的帮助,未来粉丝经济可能会带来更大的效应和价值,选择正确的营销模式,会给粉丝和偶像甚至企业家带来共赢。从论述中可以看出,随着互联网与现代技术的发展,粉丝经济产生了不同的模式特色引发新的经济浪潮。然而,粉丝经济在不断的发展变化中也必然存在一定的风险,因为任何营销模式的成功都必须建立在优质产品或服务的基础上。因此企业应注重产品的品质和客户良好的产品或服务体验,这样才能推动企业的长久发展。

参考文献:

- [1]傅潇莹,张杨.浅谈互联网时代的偶像产业和粉丝经济[J].现代经济信息,2018(10):339.
- [2]李佳倩.国内养成类真人秀的产业模式及困境[J].西部广播电视,2018(12):92-93.

(上接第31页)

据量比较大,为了防止这些数据对后面挖掘结果的影响,我们刷卡次数低于等于100,同时消费额低于300元数据删除,处理之后还保留2546条有效数据。

(2)数据离散化处理。由于在进行关联规则分析时,不能处理为Numeric型数据,所以必要的Numeric型数据进行离散化处理,将刷卡次数分为B_H、B_M、B_L,将消费金额分为C_H、C_M、C_L。

(3)将经过预处理后的数据转换为.csv格式存放,用weka软件打开.csv格式数据,并保存为.arff格式数据。在weka中打开.arff格式数据,由于一卡通卡号对后面数据处理没有意义,故在进行关联规则分析前将其删除。

(4)Apriori算法的使用。将参数metrictype设为Confidence,此处共三个可选择分别是杠杆率、提升度、确信度。将参数numRules设为6,目的是通过管理规则算法产生6条规则,其它参数设置如图3所示。

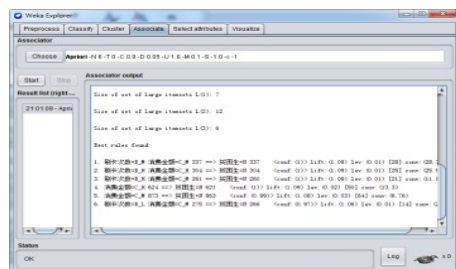


图3 关联规则分析结果

(5)关联规则分析结果。通过上述6条规则,我们能获得以下结论:第一个:消费水平为C_H的有624张卡,有621张卡不是贫困生,可信度是100%;第二个:消费水平为C_M的有873张卡,有863张卡不是贫困生,可信度是99%;通过对规则4、5分析我校贫困生比较节约,非常珍惜来之不易的补助。第三个:刷卡次数为

B_H且消费金额为C_H有304张卡,有304张卡不是贫困生,可信度是100%;第四个:刷卡次数为B_H且消费金额为C_M有261张卡,有260张卡不是贫困生,可信度是100%。通过对规则2、3分析我校贫困生消费比较控制。

四、结语

关联规则能够真实反应学生消费情况,通过数据挖掘不难发现存在规律,这些规律有助于管理者作出恰当决策。

学生贫困等级划分时,有清晰界限,然而在消费时界限还比较模糊。前期对数据处理进行大量处理工作,虽然食堂、浴室和超市都使用便捷Pos终端,但是依然有不少同学使用微信和支付宝支付。在消费额会存在一定程度的偏差。本文只在数据挖掘一个方面进行研究,在数据挖掘还有许多挖掘算法,能进一步提高挖掘数据的准确性。在未来寻找性能更加优秀的挖掘模型。

参考文献:

- [1]丁荣,孙晓辉,李智勇.基于校园一卡通的高校贫困生消费行为分析[J].网络与信息工程,2016(18):78-79.
- [2]张红蕾.数据挖掘在校园卡消费中的研究与应用[D].兰州:兰州交通大学,2016.

作者简介:

刘训星(1981.11-),男,安徽宣城人,讲师,硕士,主要研究方向:数据挖掘和教育技术信息化。

姜峰(1986.10-),男,安徽池州人,初级职称,大学本科。

龚勇(1982-),男,安徽宣城人,硕士,讲师,主要研究方向:人工智能。

基金项目:

2018年度安徽省高等学校省级质量工程计算机教研室项目,编号:2018jyssf087,负责人:刘训星。