

数据挖掘在校园超市商品销售中的应用

谢瑾

(兰州大学数学与统计学院,甘肃兰州 730000)

摘要:为了提高超市的销售率,对校园超市POS系统的数据做了调查分析。首先介绍了关联规则挖掘算法的相关概念,其次运用条件概率等方法,分析了大学生日常在超市购买多种商品时的选择概率,最后得出相应的关联规则及营销策略。

关键词:数据挖掘;关联规则;条件概率;营销策略

中图分类号:TP311

文献标识码:A

文章编号:1007-9416(2020)09-0043-03

0 引言

随着科技的发展,运用现代技术将掌握的资料转变为数据,通过存储及分析后,为相关部门或企业给出正确的决策方案。在购物篮分析(Market Basket Analysis)中,“关联规则是比较重要的一种规则模式,也是知识模式最活跃的分支之一”^[1]。关联规则反映事物之间的相互依存性和关联性,用于从大量数据中挖掘出有价值的数据项之间的相关关系。

随着超市的信息技术如条形码、电子收款机和POS系统在超市的应用,POS系统在大学校园超市中的收款、仓储等方面的应用已经非常普及,同时POS系统也积存了大量的商业运行数据,利用每天有效销售数据得到的准确信息,利用数据挖掘分析出大学生消费的购物习惯和行为,挖掘商品之间隐含的关联性。为校园超市的管理者提供科学的营销策略,提高决策的高效性和科学性。

购物篮分析的主要目的是在购买交易中分析出能够同时购买一类产品或一组产品的可能性(相互关联),通过商品销售记录的关联规则,挖掘学生经常购买商品的规律。利用得到的规律采取不同的销售策略,例如商品的放置区域以及布局设置,利用物理空间提高学生逗留时间和购买其它商品的概率。

1 基本理论

1.1 数据挖掘

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程^[2]。常用的挖掘算法有关联分析、决策树、遗传算法等。

1.2 关联规则

关联规则是寻找在同一个事件中出现的不同项的相关性,关联分析被广泛应用于购物篮或事务数据分析^[3]。关联规则是一种常用技术,它为了挖掘数据之间的相互关系,从而促进信息的显化。

关联规则中的基本概念^[4]:

(1)项集:设 $I=\{i_1, i_2, \dots, i_m\}$ 是项的集合,则 I 称为项集(itemset)。

(2)事务:事务是项的集合。

(3)事务集:事务的集合称为事务集。

(4)关联规则:关联规则是形如 $A \Rightarrow B$ 的蕴含式,规则 $A \Rightarrow B$ 在事务集 D 中成立,具有支持度 S ,其中 S 是 D 中事务包含 $A \cup B$ 的百分比,它是概率 $P(A \cup B)$ 。

(5)频繁项集:它是为满足最小支持度的项集,最小支持度是以获取对用户有用的规则,摒弃没用的规则,需要设定最小支持度和最小置信度两个阈值。

1.3 购物篮分析

购物篮是指超市供顾客购物时使用装商品的篮子,顾客付款时购物篮内的商品通过收款机登记结算并记录。所谓购物篮分析就是通过这些购物篮所显示的交易信息来研究顾客的购买行为^[5]。购物篮分析是数据挖掘最直接、简单和有效的应用领域,它能在潜在的顾客群体中发现新顾客,并建立相应的模型,能确定销售策略从而增加销量;能发现并防止顾客的流失,能更好地为社会提供更有价值的服务。

关联规则反映事物间的关联性,挖掘有价值的数据项之间的相关关系,其中某项属性值可根据其它属性值进行

收稿日期:2020-07-13

作者简介:谢瑾,女,甘肃兰州人,本科,研究方向:基础数学。

预测。经典案例“啤酒与尿布”反映了啤酒和尿布之间依赖或关联。这种现象就是卖场中商品之间的关联性。

1.4 概率概念

(1)条件概率:条件概率是指事件A在事件B发生的条件下发生的概率。条件概率表示为: $P(A/B)$ 。若只有两个事件A、B,则 $P(A/B) = \frac{P(AB)}{P(B)}$ 。

(2)联合概率:联合概率表示两个事件共同发生的概率。A与B的联合概率表示为 $P(AB)$ 或 $P(A, B)$ 或 $P(A \cap B)$ 。

(3)置信度:置信度是对支持度进行衡量的主要指标,用于衡量支持度的可信度及数据强度。这项指标将商品同时出现在购物篮中的概率进行多次运算,因此这是衡量商品相关性的主要指标。

2 具体应用

对兰州大学榆中校区天猫超市的商品销售情况做了调查和分析,利用关联规则挖掘学校超市的具体应用,由于校园超市和消费群体的局限性,该调查旨在分析大学生日常在超市购买多种商品时的选择概率,讨论超市商品捆

绑销售策略,给出具体的分析过程和营销建议。

2.1 数据来源

数据来自兰州大学“天猫超市”的实际数据。收集得到的几天不同时段随机收集的购物小票,对销售数据中的缺失值和噪声数据进行预处理。销售数据以小时为单位存储在Excel表格中,最终得到的有效数据为96份。利用Excel表格工具汇总出数据统计记录如表1(表格中的数字“1”仅代表消费记录中包含此类商品,不代表数量)。

从POS系统中整理出数据表数据,不同商品出现的频数如表2。

从图1中可以看出,96份消费记录中面包出现的频数最高,超过了一半,其次是乳制品。饮料、方便面和卤制香肠肉类销售数量也较多。这一统计情况基本符合大学生日常生活的普遍消费需求。

2.2 购买面包的同时购买的其它商品

将出现面包类的消费记录单做进一步统计,发现购买面包的51人在选择面包的同时,选择其他的商品主要有(图2)。

记“购买面包”为事件A;“购买乳制品”为事件B;“购买饮料”为事件C;“购买膨化食品”为事件D。

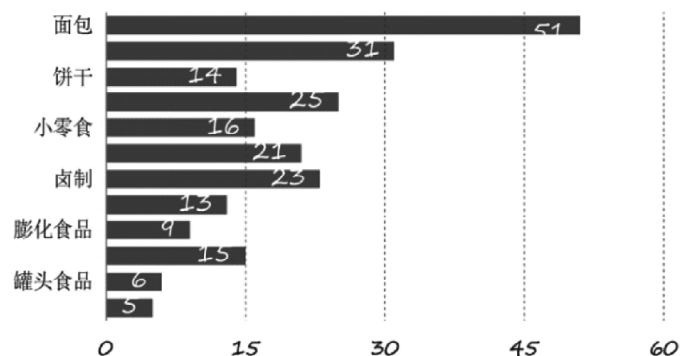


图1 不同商品的销售量

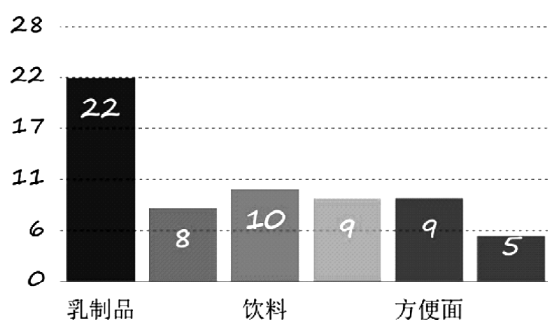


图2 购买面包的同时选择其他的商品

表1 超市的商品销售记录样例

商品名 顾客	面包类	乳制品	饼干	饮料	小零食	方便面	卤制香肠类	糖果巧克力	膨化食品	日用品	文具	罐头食品	咖啡
顾客1					1	1	1						
顾客2	1		1		1				1				
顾客3					1		1			1			
顾客4	1	1											
顾客5	1										1		

注:样例数据第一行所表示:学生1同时买了面包1个、饼干1个、小零食1个、方便面1个和卤制香肠类1个,下同。

表2 不同商品出现的频数

商品名	面包类	乳制品	饼干	饮料	小零食	方便面	卤制香肠肉类	糖果巧克力	膨化食品	日用品	罐头食品	咖啡
总和	51	31	14	25	16	21	23	13	9	15	6	5

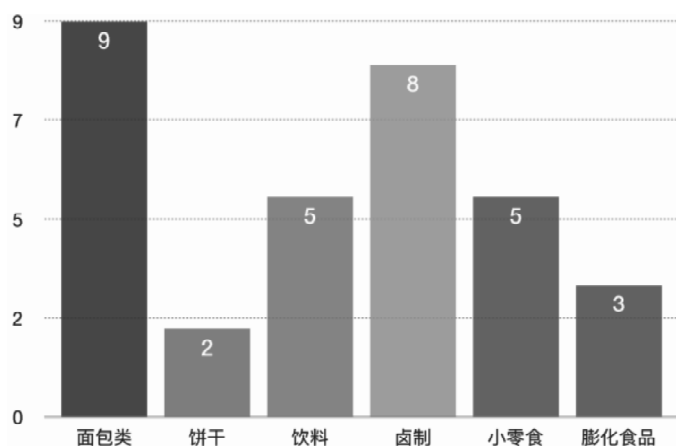


图3 购买方便面的同时购买其他产品

$$\text{则 } P(A) = \frac{51}{96}, P(AB) = \frac{22}{96}, P(AC) = \frac{10}{96}, P(AD) = \frac{5}{96}$$

$$\text{由条件概率: } P(B|A) = \frac{P(AB)}{P(A)} = 43\%$$

$$P(C|A) = \frac{P(AC)}{P(A)} = 20\% \quad P(D|A) = \frac{P(AD)}{P(A)} = 9.8\%$$

由此得知,仅有9.8%的学生会在购买面包的同时购买膨化食品,这个比例较低,即只有较少的学生买面包的同时会购买膨化食品。但却有43%的学生会在购买面包的同时购买乳制品,这表明在购买面包的消费群体中,同时购买其它商品时选择乳制品的概率最大。

2.3 购买方便面的同时购买的其它商品

记“购买方便面”为事件A;“购买面包”为事件B;“购卤制香肠肉类”为事件C。

$$\text{则 } P(A) = \frac{21}{96}, P(AB) = \frac{9}{96}, P(AC) = \frac{8}{96}$$

同理,由条件概率:

$$P(B|A) = \frac{P(AB)}{P(A)} = 43\%, P(C|A) = \frac{P(AC)}{P(A)} = 38\%$$

在购买方便面的同时会购买面包和卤制香肠类食品的概率较大,这通常也符合大多数人的购物选择。

3 结语

校园超市应制定出更好的营销策略,为学生提供更便捷和优质的服务。研究商品之间的关联程度能让管理者制定出更好的营销策略,比如商品的摆放位置,摆放数量,捆绑销售等。通过分析,商品相关性的规则都有高置信度,由于商品销售关联度大,可通过摆放货品在超市货架的位置,进一步提高整体销售量。通过购物调查分析,针对兰州大学学生这一特定的消费群体,运用条件概率的计算,调查得出的结果也基本符合我们的认知,超市货物摆放以及数量基本符合这一规律。但真正的商品分析要复杂得多,需要挖掘的数据也较复杂。

参考文献

- [1] 焦李成,刘芳.智能数据挖掘与知识发现[M].西安:西安电子科技大学出版社,2006.
- [2] Han Jiawei,Micheline Kamber.数据挖掘概念与技术[M].范明,孟小峰,译.北京:机械工业出版社,2007.
- [3] 陈玉婷,王斌,刘博,等.关联规则挖掘算法介绍[J].计算机技术与发展,2006(5):21-25.
- [4] 蒋盛益.商务数据挖掘与应用案例分析[M].北京:电子工业出版社,2014(1):104-109.
- [5] Qu Yang,Weimin Cai,Qingsheng,Researches on discovery of association rules,Computer Science in Chinese,1999(3):41-44.

Application of Data Mining in Comm Odity Sales of Campus Supermarket

XIE Jin

(School of Mathematics and Statistics, Lanzhou University, Lanzhou Gansu 730000)

Abstract:In order to improve the sales rate of supermarkets, the data of POS system in campus supermarkets were investigated and analyzed.Firstly, this paper introduces the relevant concepts of association rule mining algorithm, then analyzes the selection probability of college students when they buy a variety of goods in the supermarket by using conditional probability and other methods, and finally obtains the corresponding association rules and marketing strategies.

Key words:data mining; association rules; conditional probability; the marketing strategy