

# 基于校园一卡通数据的大学生用餐消费相关性分析

王金鑫

(延边大学信息化中心,吉林 延吉 133002)

**摘要** 基于某高校在校本、专科生校园一卡通午餐和晚餐消费的流水数据,采用统计学方法对数据进行处理、描述、分析和建立模型,对学生午餐和晚餐消费的相关性、消费金额与刷卡次数的相关性进行研究。

**关键词** 大学生 校园一卡通 大数据 分析 SPSS

中图分类号: TP393

文献标识码: A

文章编号: 2096-4390(2018)15-0081-02

校园一卡通系统作为数字校园建设重要组成部分,已经在各高校全面推广和普及,通过对校园一卡通数据的挖掘可以了解学生生活、学习、消费情况等。关于校园一卡通数据挖掘和分析的研究主要从系统架构和数据库关联、数据处理方法等角度讨论。罗华群<sup>[1]</sup>以校园一卡通日常消费数据为研究对象,利用语言 Matlab 对数据进行编程,对学生每月消费额进行分析,判断是否是贫困生,张丽丽<sup>[2]</sup>以校园卡信息资源数据为研究对象,对校园卡信息系统建立了模型,仅仅属于理论上进行了描述;王德才<sup>[3]</sup>用支持向量机、关联规则的数据挖掘方法将消费按群体分类等角度进行建模,只将所有本科生消费额进行分类,并没有更细致的考虑其他情况,曾秋凤<sup>[4]</sup>利用 K-means 聚类方法对学生消费水平、用餐高峰时期进行了分析,初步实现了校园卡数据挖掘学生消费行为。

本文是以学生校园一卡通食堂就餐消费数据为研究对象,分别从年龄、性别、出生地、学院的角度进行了描述性分析,对午餐与晚餐的消费金额、用餐消费金额与刷卡次数进行了相关性分析,对学生管理和饮食就餐等方面有重要意义。

## 1 数据来源与挖掘

本文数据是来自于吉林省某高校 2017 年 7 月到 12 月份在校本、专科生校园一卡通食堂就餐消费数据,包括学号(idserial)、

身份证号码(identity)、学院(deptstr)、消费时间(jdatetime)、消费金额(tranamt)等。

利用 Toad for Oracle、SQL 语句和 exp 系统命令 exp system/\*\*\*\*@ table file=e:\daochu.dmp full=y 将所研究数据从 Oracle 数据库中导出,形成 Access 文件,同时为了提高数据利用率和导出的时效性,利用 BUFFER=( )命令扩展缓存空间。在 Access 数据库信息筛选过程中,为了保证数据质量,去除冗余数据记录,保留学号、身份证号码、学院、年级、消费时间、消费金额等有效数据,再将 Access 文件转换成 Excel 表格。

由于 Excel 表格行数的局限性,将数据存放在多个 Excel 子表格中,对消费时间进行“分列”、“排序”,筛选出午餐消费数据和晚餐消费数据分别存放在 Excel 表格中,利用函数 IF(MOD(MID("身份证号码所在位置",17,1),2),"男","女")从身份证号码中提取学生性别,利用 MID 函数获取学生出生地和出生日期,出生日期与 2017 做差可获取学生年龄,再利用 COUNTIF、SUMIF、SUM 等函数获取学生出生地人数、学院人数、年级人数、用餐次数、消费总额等,最终生成 SPSS 文件<sup>[5-6]</sup>。

筛选午餐消费人数 15109 人,晚餐消费人数 13899 人,合计数据 1478903 条消费记录。去除中午在食堂就餐晚上不在食堂就餐和中午不在食堂就餐晚上在食堂就餐的同学消(转下页)

通信。此外还编写了 Http Client Util 工具类来提高开发效率,对 Http Client 操作的常用方法进行封装,如实例的创建、GET、POST 请求等,还提供判断网络是否可用等基本方法。

**数据的请求。**在 Activity 类或 Adapter 类中,开启请求数据异步线程,由服务层中相应类调用 Http Client Util 工具类向远端服务发送 GET 请求。GET 方法获取的数据格式为 JSON,再由服务层中的相应类解析 JSON 格式的数据,最后返回数据。

**数据上传。**在 Activity 类或 Adapter 类中,开启数据上传异步线程,由服务层中相应类对上传的数据进行封装(JSON 格式),再由相应类调用 Http Client Util 工具类向远端服务发送 POST 请求,提交数据。

**数据保存与同步。**本系统在进行数据上传和请求时是以 JSON 格式进行,使用 Jackson 框架可完成数据的封装与解析。在数据保存与同步时,可调用 execute Post 方法使用 Http Client 访问网络资源并上传数据,上传成功后更新上传服务器的时间和状态。

**3.5 数据分析模块。体征数据异常提醒。**本系统会实时对用户测量的各类体征数据进行监测分析,当体征数据出现异常或接近阈值时,会通过特殊文字或图形提醒用户,引起用户的注意。

**健康建议推送。**本系统会在移动终端推送符合用户的健康建议,这些建议是由与健康服务平台合作的医疗机构对用户的健康数据进行分析后所提出的。

**3.6 数据展示模块。**用户可通过健康体征数据的综合展示来快速了解自己的体征状况,也可了解某一类健康体征的数据详情,还可通过健康体征数据的图形化展示,更为直观的了解自己近期的健康体征数据走势。

## 4 结论

本系统虽然能够为人们提供一种高效、便捷的个人健康管理手段,但系统不是很成熟,还有很多有待完善的地方,一些关键技术还有待于进一步研究。不过,目前整个学术界和产业界已经充分认识到移动健康的意义,随着研究的深入,移动健康管理技术必然会促进医疗事业的全面发展。

## 参考文献

- [1]徐倩,赵文龙.基于移动医疗 App 的用户健康信息需求分析[J].现代情报,2015,35(11):79-82.
- [2]徐莉莉,刘建明.国内健康咨询网站用户需求调查分析[J].医学信息学杂志,2014,35(7):50-53.
- [3]李奇.移动服务支持下健康管理系统的设计与实现[D].哈尔滨 哈尔滨工业大学,2015.

费记录,最终保留了中午和晚上同时在食堂就餐来自 18 个学院 13484 名学生的 1448934 条消费记录。

### 2 数据的描述性分析

从年龄角度分析,就餐学生中,年龄最大是 29 岁,年龄最小是 15 岁,用餐最多为 19 岁,中数为 20 岁;从就餐学生性别角度分析,各年级食堂女生就餐人数明显高于男生,总体超出近 75%,超出最高是 2014 级,女生就餐人数是男生的 1.8 倍,最低是 2016 级,女生就餐人数超出男生 67%;从年级角度分析,低年级学生在食堂就餐人数高于高年级学生。

从学生出生地角度分析,来自各省市自治区的学生午餐单笔刷卡消费均低于 8 元,其中来自上海、吉林和黑龙江的学生午餐单笔刷卡消费高于自其他省份的学生消费金额;学生晚餐单笔刷卡消费基本情况显示,来自吉林、黑龙江和辽宁的学生单笔刷卡消费略高于来自其他省份的学生。

汉语、体育和艺术 3 个学院的消费均值略高于其他学院,通过数据图表构建程序,学生午餐刷卡次数在 300 次以内较多,晚餐刷卡次数主要集中在 250 次以内,多数学生午餐刷卡次数与晚餐刷卡次数成正比,说明多数学生就餐选择食堂;男生单笔刷卡消费金额高于女生。

### 3 数据的相关性分析

#### 3.1 午餐消费与晚餐消费的相关性分析

由于用餐学生习惯不同,对午餐和晚餐消费方式也不同,下面我们对两餐的消费数据进行分析,首先进行相关性分析,结果如表 1 所示。表 1 表明,学生午餐消费与晚餐消费的相关性显著,两者具有较强的正相关性,即午餐消费多,晚餐消费也多。

表 1 午餐消费和晚餐消费相关性

		午餐消费	晚餐消费
午餐消费	Pearson 相关性	1	.738**
	显著性(双侧)		0.000
	N	13484	13484
晚餐消费	Pearson 相关性	.738**	1
	显著性(双侧)	0.000	
	N	13484	13484

\*\* . 在 .01 水平(双侧)上显著相关。

如果将午餐消费设成变量 X,晚餐消费看成变量 y,午餐消费与晚餐消费大体呈线性关系。下面用线性回归分析再进行研究。

表 2 表示,常量和解释变量的 p 值都小于 0.001,及其显著,因此晚餐消费和午餐消费的线性方程可以表示为  $Y=14.004+0.6X$ ,说明学生午餐多消费 1 元时,晚餐将多消费 0.6 元。

表 2 晚餐消费与午餐消费模型系数<sup>a</sup>

模型		非标准化系数		标准系数	t	Sig.
		B	标准 误差	试用版		
1	(常量)	14.004	2.676		5.232	.000
	午餐消费	.600	.005	.738	126.860	0.000

a. 因变量: 晚餐消费

#### 3.2 消费额与刷卡次数的相关性分析

在进行数据整理期间,发现存在单笔刷卡金额由不足 1 元到几十元,并且存在较多的小金额刷卡记录,通过 POS 机交易记录和食堂窗口定价的分析,大金额单笔刷卡记录是几位同学在同一窗口使用同一张校园卡消费,小金额刷卡记录是购买主食时消费的,比如购买二两两米饭或一个馒头等。

我们知道刷卡次数增加,消费总额一定增加,那么,刷卡次数的明显增加是否会带来总消费额明显增加呢?下面我们对刷卡次数与消费总金额进行相关性分析<sup>[7]</sup>。

表 3 午餐消费与午餐刷卡次数相关性

		午餐消费	午餐刷卡次数
午餐消费	Pearson 相关性	1	.907**
	显著性(双侧)		0.000
	N	13484	13484
午餐刷卡次数	Pearson 相关性	.907**	1
	显著性(双侧)	0.000	
	N	13484	13484

\*\* . 在 .01 水平(双侧)上显著相关。

由表 3 可知,午餐消费与刷卡次数的相关性十分显著,说明即使有小金额单笔刷卡消费存在,不影响整体数据关系走向。

我们将刷卡次数看成变量 X,午餐消费金额看成变量 Y,可以得到两个变量之间存在正相关的线性关系,下面用线性回归分析再进行研究。

表 4 午餐消费与午餐刷卡次数模型系数<sup>a</sup>

模型		非标准化系数		标准系数	t	Sig.
		B	标准 误差	试用版		
1	(常量)	63.176	1.991		31.732	.000
	午餐刷卡次数	5.711	.023	.907	249.503	0.000

a. 因变量: 午餐消费

从表 4 可以得到常量和解释变量的 p 值都小于 0.001,及其显著,因此午餐消费与午餐刷卡次数的关系可以表示为  $Y=63.176+5.711X$ ,说明午餐消费每刷一次卡将带来平均 5.711 元的就餐消费。

对晚餐消费与晚餐刷卡次数进行相关性分析,二者也具有较强的线性关系,将晚餐刷卡次数看成变量 X,晚餐消费金额看成变量 Y,得出晚餐消费与晚餐刷卡次数的关系为  $Y=26.632+6.581X$ ,这说明晚餐消费每刷一次卡将带来平均 6.581 元的就餐消费。

本文的研究结果,对食堂用餐窗口的设置、管理与服务的改善提供了理论依据,同时也为贫困学生的认定与资助工作带来了帮助,由于部分学生不在食堂就餐,因此本文研究结果只能作为贫困生认定工作的参考,不能作为依据。

### 参考文献

- [1]罗华群.校园一卡通数据的挖掘与应用[D].上海:华东师范大学软件学院,2009
- [2]张丽丽.数据挖掘技术在校园卡信息管理系统中的应用研究[J].中国管理信息化,2013,16(4):79-80.
- [3]王德才.数据挖掘系统在校园卡消费行为分析中的研究与应用[D].哈尔滨:哈尔滨工程大学,2010.
- [4]曾秋凤.数据挖掘在校园一卡通数据库中的应用研究[D].南昌:江西师范大学,2008.
- [5]张文彤,邝春伟.SPSS 统计分析基础教程(第 2 版)[M].北京:高等教育出版社,2011.
- [6]薛薇.统计分析与 SPSS 的应用(第三版)[M].北京:中国人民大学出版社,2011
- [7]马千.基于校园卡消费流水数据的大学生就餐行为分析[D].长春:东北师范大学,2015.

作者简介:王金鑫(1986-)男,延边大学信息化中心,助理工程师。