

基于马尔科夫模型的就餐人数预测^①

徐彭娜¹, 林志兴², 林 劫¹, 江育娥¹

¹(福建师范大学 软件学院, 福州 350108)

²(福建三明学院 现代教育技术中心, 三明 365004)

摘 要: 准确预测就餐人数能够降低学校食堂的运行成本, 提高学生对食堂的满意度. 根据校园一卡通的消费情况, 提出一种基于马尔科夫模型的就餐人数预测研究方法. 首先, 通过计算早餐就餐行为得到初始概率; 其次, 分别通过计算早、午餐和午、晚餐就餐行为得到早餐午餐概率转移矩阵和午餐晚餐概率转移矩阵; 最后, 根据初始概率和概率转移矩阵构建的模型预测三餐的就餐人数. 该方法的就餐人数预测的平均预测误差率为 1.31%, 具有良好的预测效果. 实验结果表明, 该方法能够反映学生的就餐行为, 从而可以为学校后勤部门提供一些参考意见, 有助于学校的建设和管理也有助于满足学生的需要.

关键词: 数据挖掘; 马尔科夫模型; 就餐人数; 预测; 概率转移矩阵

Repast Number Prediction Based on Markov Model

XU Peng-Na¹, LIN Zhi-Xing², LIN Jie¹, JIANG Yu-E¹

¹(Faculty of Software, Fujian Normal University, Fuzhou 350108, China)

²(Center of Modern Education Technology, Sanming University, Sanming 365004, China)

Abstract: To predict the repast number accurately can reduce the cost of school canteen and improve students' satisfaction. A novel method based on Markov model to predict repast number is proposed according to the consumption situation of campus card system. Firstly, an initial probability is obtained by calculating the eating behavior of breakfast. Secondly, two transfer probability matrices are computed, one is the transfer probability between the behaviors of students having breakfast and having lunch; the other is the transfer probability between the behaviors of students having lunch and having supper. Finally, a Markov model is constructed according to the initial probability and the two probability transfer matrices to forecast the number of diners. The average prediction error of the proposed method is 1.31%, which has a good prediction performance. The experimental results show that the proposed Markov method can capture the students' dining behavior accurately. It may provide valuable information for the school logistics department, contribute to the construction and management of school and meet the needs of students better.

Key words: data mining; Markov models; repast number; prediction; transfer probability matrix

数字化校园发展迅速, 其中校园一卡通系统是其重要的组成部分^[1]. 校园一卡通凭借一卡多用、圈存消费、实时管理、数据共享等特征, 已成为学生、老师不可或缺的使用工具, 成为学校不可或缺的管理工具. 大量用户使用一卡通就餐、洗浴、出勤、借书等, 每天都产生大量流水数据. 其中就餐数据是最易获取且最具有代表性的数据. 校园一卡通还与校园的相关

系统相结合^[2], 如校园网的缴费、消费等. 所以, 如何从海量就餐数据中挖掘出能给予学校、老师、后勤部门有用的信息具有实际意义.

目前一卡通数据多采用聚类^[3]、关联规则、决策树、数据仓库等方法, 对数据进行统计分析处理. 但是一卡通数据在预测方面多数处于理论层面的描述, 实际的具体实现比较少. 本文通过对传统一卡通数据进

① 基金项目:福建省自然科学基金(2014J01220);三明学院科研基金(B201201/G);福建省教育厅科技基金(JB13187)

收稿时间:2016-08-01;收到修改稿时间:2016-08-29 [doi: 10.15888/j.cnki.csa.005685]

行数据统计分析,在挖掘过程中采用基于马尔科夫模型(Markov Model)对就餐人数进行预测研究.即通过计算早餐的初始概率、早餐午餐概率转移矩阵和午餐晚餐概率转移矩阵,在已有数据的基础上预测校园内早、中、晚餐的就餐总人数.马尔科夫模型能较好的预测就餐人数,为学校后勤提供决策支持,有效的防止供餐不足或过多剩余的现象;对早、中、晚的人数预测,能够结合用户或学校的其他数据,为之后更具体的学生行为分析提供借鉴和实验基础.

本文总共分为四个部分.第一部分介绍了校园一卡通的相关工作和近年来的成果;第二部分详细介绍了马尔科夫的概念和用于就餐预测研究的马尔科夫模型;第三部分介绍了使用某高校的校园一卡通就餐流水数据的预测实验和实验结果;第四部分是对本文的总结.

1 相关研究

数据挖掘在校园一卡通数据上的理论分析和技术研究日益成熟.研究的相关问题有:消费行为如食堂就餐、热水使用情况、水电费管理等;教学辅助如贫困生补助、励志奖学金评比、出勤率等;行为特征分析如根据消费情况分析学生行为.

在消费行为方面,王德才^[4]使用支持向量机分类器将消费者分成三类,在分类基础上使用关联规则分析消费行为,了解学生的消费模式和消费类别,对贫困生管理提供决策支持,但是算法的效率不高,尚待优化.张硕^[5]利用 WEKA 数据挖掘平台,使用 K-Means 算法和 J48 决策树算法对在校就餐流水数据分析就餐周期性情况、就餐高峰、学生消费水平等情况,有助于学校对学生的了解,在励志奖学金评比中做出相应的决策,但是使用的数据较单一,只针对消费数据进行分析.罗华群^[6]对校园一卡通的消费数据在消费时间和消费金额上的差异进行统计、分析,对学校后勤部门提供决策支持,但是在数据预处理阶段没有对异常情况进行预先处理,这对统计结果产生一定的影响.

在教学影响方面,欧阳圣^[7]基于数据挖掘构建一个校园行为分析系统,结合聚类方法对学生行为统计分析,将理论应用于实践,并反作用于理论,但是分析结果的准确性不够理想.陈建兵^[8]使用关联规则的方法对数据进行分析贫困学生的消费情况、参与勤工助学活动、学习成绩之间的关系,给学校和学生部门

提供了决策支持,但是在分析过程中只使用少量数据,有一定的局限性.李齐光^[9]发明了一种基于一卡通消费行为分析识别经济困难生的方法和装置,通过数据统计等技术手段识别经济困难生,以便于为经济困难生提供生活帮助,提高经济困难生判断的客观性,减少人为主观判断,并减少复杂的申请程序,真正把帮助落实到经济有困难的学生上.

在行为分析方面,Fan^[10]根据所收集的信息数字化校园卡,提出一个基于网络探索学生成就和其研究伙伴关系的方法,结果表明,优秀学生的学习有更多样化的成果,更擅长与同学们合作学习,通过优秀学生的行为给学校的教学决策提供支持.徐剑^[11]对大学生的消费习惯进行聚类,将消费水平分为三类,从早餐就餐情况判定学生的消费习惯是否具有不规律性,将学生的成绩,消费习惯和学生的消费水平进行关联性分析,结论表明学生成绩与消费习惯有一定关系,为教学老师提供决策支持,但由于学生的行为多样化,仅仅分为三类不能够充分体现学生行为的特征.

综上所述,在校园一卡通数据的统计分析以及系统的理论和实现上都颇有成果,但是在数据预测方面的研究还是比较少的.本文提出基于马尔科夫模型,预测高校就餐人数,给学校相关部门提供决策支持.

2 基于马尔科夫模型的就餐预测研究

2.1 马尔科夫模型

给定离散时间 $t_{i+1} \in T, \{0 \leq t_1 < t_2 < \dots < t_n \in T\}$ 和离散状态空间 $I_{t+1}, \{I_0, I_1, \dots, I_{t+1} \in I\}$, 设有随机过程 $X_{t+1}, \{X_t, t \in T\}$, 其 X_{t+1} 时刻的条件概率如公式(1)所示:

$$\begin{aligned} P\{X_{t+1} = I_{t+1} | X_0 = I_0, X_1 = I_1, \dots, X_t = I_t\} \\ = P\{X_{t+1} = I_{t+1} | X_t = I_t\} \end{aligned} \quad (1)$$

符合公式(1)的随机过程被称为马尔科夫链.马尔科夫链主要用于离散变量的随机模拟,在实际中有着广泛的应用^[12].

公式(2)表示了马尔科夫链 $\{X_t, t \in T\}$ 在时刻 t 向时刻 $t+1$ 变化的一阶转移概率.

$$P_{ij}(t) = P\{X_{t+1} = j | X_t = i\} \quad (2)$$

其中, $i, j \in I$, 条件概率 $P\{X_{t+1} = j | X_t = i\}$ 表示在时刻 t 发生 i 事件的条件下(当前状态), $t+1$ 时刻发生 j 事件的概率(将来状态), 这里假设 $t+1$ 时刻事件 j 发生概率只与 t 时刻事件 i 发生概率有关, 与已过去的其他时刻的事件无关. 假设已知 t 时刻过程概率为 $S_i(t)$, 那么第 $t+1$ 时刻

发生 j 事件的概率为 $S_j(t+1)$, 公式如下所示:

$$S_j(t+1) = \sum_{i=1}^k S_i(t) P_{ij}, i=1,2,\dots,k \quad (3)$$

公式(3)中, k 表示集合 I 的事件个数. 对于 t 时刻的事件集合 I 的过程概率向量 $S(t)=(S_1(t), S_2(t), \dots, S_k(t))$, 有 $S(t+1)=S(t)P$, 其中 P 为一阶概率转移矩阵^[13].

使用马尔科夫模型的前提条件是, 在任何期间或任何状态下, 变量的下一个期间或下一个状态的转移概率是符合某一概率分布的随机值, 下一个状态仅仅依赖于最近期间的状态, 而不依赖于在此之前的状态.

2.2 马尔科夫模型就餐预测研究

马尔科夫模型体现马尔科夫性质, 满足 $t+1$ 时刻系统状态的概率分布只与 t 时刻的状态有关, 与 t 时刻以前的状态无关. 文献[14]证实了午餐消费和晚餐消费具有较强的正相关性, 即午餐是否就餐的行为只与早餐有关, 晚餐是否就餐的行为只与午餐的行为有关. 就餐类型分为三类: 早餐、午餐、晚餐, 用 A 、 B 、 C 依次表示早、午、晚有就餐行为, a 、 b 、 c 依次表示早、午、晚的无就餐行为. 马尔科夫过程中, 某过程的状态根据一定的概率分布在各个状态间转换, 未来时间的状态是不确定的, 一卡通用户就餐行为可以认为在特定餐(如早中晚餐)的就餐状态(用餐/不用餐)与下一餐的就餐状态之间的转移, 这样的状态转移是完全随机的, 因此应用马尔科夫模型可以较好地表示一卡通用户的就餐行为.

早餐就餐行为的初始概率 $P_u^i(A)$ 和 $P_u^i(a)$ 分别表示了一个用户(u)在一个星期中的某天(星期 i)的用餐(A)和不用餐(a)的概率, 其中 u 代表用户; i 表示星期中的某天; A 代表用餐, a 代表不用餐. 如公式(4)所示. 其中 $N_u^i(A)$ 表示用户 u 星期 i 早餐有就餐行为的总天数, N_u^i 为数据中用户 u 星期 i 的总天数, 前提假设是一个用户一天吃一次早餐.

$$\begin{cases} P_u^i(A) = \frac{N_u^i(A)}{N_u^i} \\ P_u^i(a) = \frac{N_u^i(a)}{N_u^i} \end{cases} \quad (4)$$

公式(5)和公式(6)是一阶概率转移矩阵. 其中公式(5)中表示训练集中每个用户星期 i 的早餐午餐概率转移矩阵, 公式(6)中表示训练集中每个用户星期 i 的午餐晚餐概率转移矩阵.

$$P_u^i(L) = \begin{bmatrix} P_u^i(AB) & P_u^i(Ab) \\ P_u^i(aB) & P_u^i(ab) \end{bmatrix} = \begin{bmatrix} \frac{N_u^i(AB)}{N_u^i(A)} & \frac{N_u^i(Ab)}{N_u^i(A)} \\ \frac{N_u^i(aB)}{N_u^i(a)} & \frac{N_u^i(ab)}{N_u^i(a)} \end{bmatrix} \quad (5)$$

$$P_u^i(S) = \begin{bmatrix} P_u^i(BC) & P_u^i(Bc) \\ P_u^i(bC) & P_u^i(bc) \end{bmatrix} = \begin{bmatrix} \frac{N_u^i(BC)}{N_u^i(B)} & \frac{N_u^i(Bc)}{N_u^i(B)} \\ \frac{N_u^i(bC)}{N_u^i(b)} & \frac{N_u^i(bc)}{N_u^i(b)} \end{bmatrix} \quad (6)$$

例如, 甲用户前三周星期一的就餐行为为 ABc 、 ABc 、 abC , 那么 $N_{甲}^1(A) = 2$, $N_{甲}^1(a) = 1$, $N_{甲}^1 = 3$, $N_{甲}^1(AB) = 1$, $N_{甲}^1(Ab) = 1$, $N_{甲}^1(aB) = 0$, $N_{甲}^1(ab) = 1$, $N_{甲}^1(B) = 1$, $N_{甲}^1(b) = 2$, $N_{甲}^1(BC) = 0$, $N_{甲}^1(Bc) = 1$, $N_{甲}^1(bC) = 2$, $N_{甲}^1(bc) = 0$. 由公式(4)、(5)、(6)可得:

$$\begin{cases} P_{甲}^1(A) = \frac{N_{甲}^1(A)}{N_{甲}^1} = \frac{2}{3} \\ P_{甲}^1(a) = \frac{N_{甲}^1(a)}{N_{甲}^1} = \frac{1}{3} \end{cases}$$

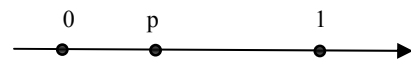
$$P_{甲}^1(L) = \begin{bmatrix} \frac{N_{甲}^1(AB)}{N_{甲}^1(A)} & \frac{N_{甲}^1(Ab)}{N_{甲}^1(A)} \\ \frac{N_{甲}^1(aB)}{N_{甲}^1(a)} & \frac{N_{甲}^1(ab)}{N_{甲}^1(a)} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$

$$P_{甲}^1(S) = \begin{bmatrix} \frac{N_{甲}^1(BC)}{N_{甲}^1(B)} & \frac{N_{甲}^1(Bc)}{N_{甲}^1(B)} \\ \frac{N_{甲}^1(bC)}{N_{甲}^1(b)} & \frac{N_{甲}^1(bc)}{N_{甲}^1(b)} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

以上介绍的条件概率和概率转移矩阵是预测模型的基础. 选取模拟随机值 r , 应用 r 对是否就餐进行模拟预测.

模拟方法如下:

(1) 在直线上将 $[0,1]$ 区间划分成两个子区间, 其分点记为 p , p 为就餐概率. 如下所示.



(2) 在计算机上取随机模拟值 r .

(3) 判定 r 在 x 轴上所落的位置, 若 $r \in [0,p]$, 则可以认为有就餐行为.

早餐预测方法: $p = P_u^i(A)$ 为早餐有就餐行为的概率, 则早餐为无就餐行为的概率为 $1-p$. 若 $r \in [0,p]$, 则预测早餐有就餐行为.

午餐预测方法: 根据预测的早餐就餐行为和随机模拟值 r_2 来预测午餐的就餐行为, 若早餐的就餐类型为 A , 此时 $p = P_u^i(AB)$, 当 $r_2 < p$ 时预测午餐有就餐行为,

反之则无, 若早餐的就餐类型为 a, 此时 $p=P_u^i(aB)$, 当 $r_2 < p$ 时预测午餐有就餐行为, 反之则无.

晚餐预测方法: 根据预测的午餐就餐行为和随机模拟值 r_3 来预测晚餐的就餐行为, 若午餐的就餐类型为 B, 此时 $p=P_u^i(BC)$, 当 $r_3 < p$ 时预测晚餐有就餐行为, 反之则无, 若午餐的就餐类型为 b, 此时 $p=P_u^i(bC)$, 当 $r_3 < p$ 时预测晚餐有就餐行为, 反之则无.

统计真实值与预测集的早、午、晚就餐人数数量, 计算就餐总人数误差率.

3 实验与结果分析

3.1 数据来源

本实验采用的数据是某高校 2015 年 12 月的食堂就餐流水数据共 76 万条. 经过数据预处理, 把每个用户的系列就餐行为提取为一个时间序列. 这些时间序列数据总条数为 12442. 也即, 总人数为 12442 人, 其中性别为男的人数为 5651, 性别为女的人数为 6791. 教职工人数为 753, 学生 11689.

3.2 数据预处理

在实验开始之前, 需将获取到的 76 万条就餐流水数据使用数据库和 R 工具对数据进行预处理. 预处理步骤如下:

① 清洗数据: 将学号缺失、就餐时间不在食堂营业时间的错误数据剔除.

② 为早午晚三餐分段: 将 6:00:00-10:00:00 置为早餐时间, 如有刷卡记录则将就餐行为置为 A, 将 10:00:01-15:00:00 置为午餐时间, 如有刷卡记录则将就餐类型置为 B, 将 15:00:01-20:00:00 置为晚餐时间, 如有刷卡记录则将就餐类型置为 C.

③ 记录合并: 将同一个学号的日期相同且就餐类型也相同的记录合并, 记作一次就餐行为(如同一天的多次早餐刷卡行为合并为一次早餐刷卡行为).

④ 形成时间序列: 将每个用户 u 的无就餐行为用小写字母(a, b, c)补全后按照时间排序形成就餐类型序列 $Y_u, Y_u = \{y_{u,1}, y_{u,2}, \dots, y_{u,b}, \dots, y_{u,m}\}$. 其中 m 表示用户 u 共有 m 次就餐行为, $y_{u,t}$ 表示用户 u 的第 t 次就餐行为.

预处理结果为 n 个用户就餐信息 $MESS = \{mess_1, mess_2, \dots, mess_u, \dots, mess_n\}$, 其中 n 为用户数量, $mess_u$ 表示用户 u 的就餐信息. $mess_u = (std_u, Y_u)$, 其中 std_u 表示用户 u 的学号, Y_u 表示用户 u 的就餐行为序列.

3.3 预测模型和预测方法

本实验使用软件 R 对数据建模进行预测. 本实验使用一卡通的就餐流水数据, 基于马尔科夫模型对每个用户 u 在星期 i 对就餐人数进行预测研究. 将数据的前三周的数据作为训练集, 第四周的数据作为测试集.

使用训练集建立模型, 具体过程如下:

① 对每一个用户 u 在星期 $i(1 \leq i \leq 7)$ 的基础上建立 7 天的训练模型;

② 从就餐行为序列 Y_u 中提取用户 u 前三周的星期 i 的总就餐行为构成的时间序列集合, 作为用户 u 的训练集 $train_{ui} = \{y_{u,1}, y_{u,2}, \dots, y_{u,z}\}$, $y_{u,z}$ 表示前三周中星期 i 的第 z 次就餐行为;

③ 根据公式(4)计算用户 u 在训练集中的初始概率;

④ 根据公式(5)、公式(6)计算概率转移矩阵, 建立训练模型.

具体流程如图 1 所示.

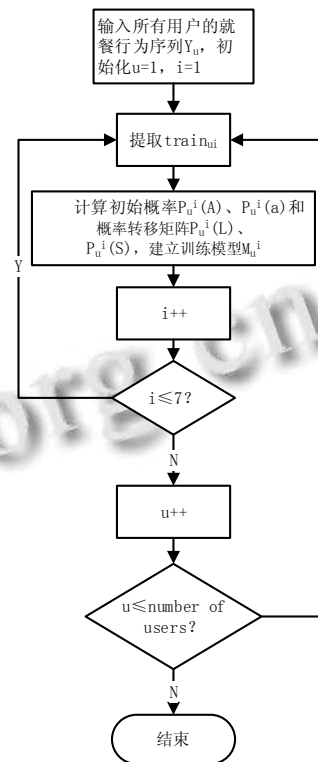


图 1 基于马尔科夫模型预测流程图

在已有的马尔科夫模型上, 对用户的早餐、午餐、晚餐进行预测, 具体过程如下:

① 早餐预测: 选取一个 0-1 之间的随机模拟值 r_1 , 若 $r_1 < P_u^i(A)$, 则早餐有就餐行为, 反之则无就餐行为;

② 午餐预测: 选取一个 0-1 之间的随机模拟值 r_2 , 根据早餐的就餐行为结果和 r_2 来预测午餐的就餐行为, 若早餐的就餐类型为 A 且 $r_2 < P_u^i(AB)$, 则预测午餐有就餐行为, 反之则无, 若早餐的就餐类型为 a 且 $r_2 < P_u^i(aB)$, 则预测午餐有就餐行为, 反之则无;

③ 晚餐预测: 选取一个 0-1 之间的随机模拟值 r_3 , 根据午餐的就餐行为结果和 r_3 来预测晚餐的就餐行为, 若午餐的就餐类型为 B 且 $r_3 < P_u^i(BC)$, 则预测晚餐有就餐行为, 反之则无, 若午餐的就餐类型为 b 且 $r_3 < P_u^i(bC)$, 则预测晚餐有就餐行为, 反之则无.

3.4 实验结果

在实验过程中使用随机模拟值预测三餐就餐行为, 进行 20 次实验, 分别形成 20 组不同实验结果, 最终的算法评价是基于这 20 组实验结果的平均值.

本实验采用就餐人数预测误差率 err 作为预测评估指标:

$$err = \frac{|P_{ik} - T_{ik}|}{N} \quad (7)$$

其中, P_{ik} 表示预测值, T_{ik} 表示真实值, i 表示星期中的某天($i=1,2,\dots,7$), k 表示就餐类型($k=A,B,C$), N 表示就餐总人数. 平均误差值表示真实值和预测值之间的误差. 平均误差率越小, 表示预测效果越准确. 就餐人数预测误差率如表 1 所示.

表 1 就餐人数预测误差率表

	早餐	午餐	晚餐	平均误差率(%)	
周一	真实值	3972	6764	6499	2.31
	预测值	3778	6369	6225	
	误差率(%)	1.56	3.17	2.20	
周二	真实值	4765	6924	6628	0.76
	预测值	4612	6881	6538	
	误差率(%)	1.23	0.35	0.72	
周三	真实值	4048	6204	5529	2.19
	预测值	4415	6461	5725	
	误差率(%)	2.94	2.07	1.57	
周四	真实值	4162	6698	6565	0.47
	预测值	4190	6644	6472	
	误差率(%)	0.22	0.43	0.75	
周五	真实值	4209	6528	5511	0.80
	预测值	4276	6642	5392	
	误差率(%)	0.53	0.92	0.96	
周六	真实值	2008	4545	4355	1.08
	预测值	1946	4600	4070	
	误差率(%)	0.50	0.44	2.29	

周	真实值	1413	4607	5006	
日	预测值	1448	4371	4696	1.56
	误差率(%)	0.28	1.90	2.50	
	平均值(%)	1.04	1.33	1.57	1.31

表 1 描述了某高校 2015 年 12 月的星期 i 的三餐就餐人数的真实值和预测值, 并计算真实值和预测值之间的误差率. 为了直观描述在星期 i 的就餐人数的分布情况, 如图 2 所示.

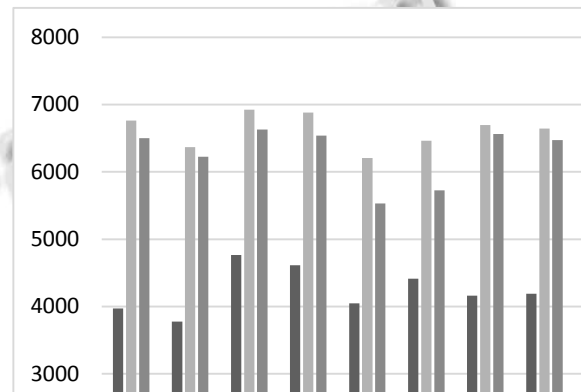


图 2 基于马尔科夫模型预测人数对比图

图 2 中横坐标表示星期一至星期日的真实值和预测值; 纵坐标表示就餐人数. 从图 2 中可以发现早餐的就餐人数普遍比午餐和晚餐的就餐人数低, 周末的三餐就餐人数均偏低; 预测人数与真实人数相近. 为了直观描述预测误差情况, 如图 3 所示.

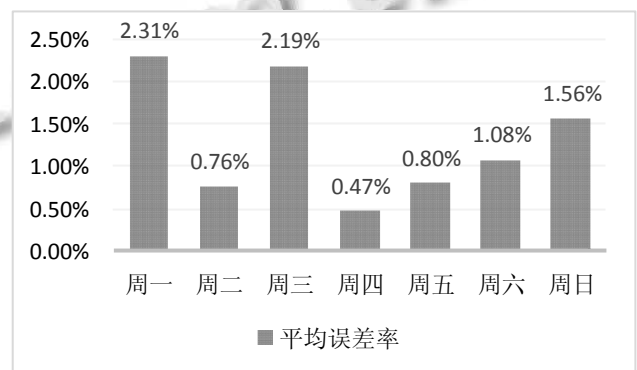


图 3 基于马尔科夫模型预测误差率

图 3 中横坐标表示星期 i , 纵坐标表示平均误差率. 由图 3 可知, 星期一和星期三的预测误差率偏大, 均大于 2%. 经调查表明, 星期一请假和旷课情况较多, 周三下午没有课程安排, 学生在此阶段行为比较随机, 这说明周末和不排课的客观情况会对就餐情况产

生影响(也影响实验结果).

实验结果表明,星期四的人数预测误差率最小,为 0.4%,一周的平均人数预测误差率为 1.31%. 周末的就餐人数相对普遍较低,尤其是早餐的就餐人数大大减少,但人数稳定,早餐预测误差率低,说明周末早起人数少,外出就餐人数多. 除去周末,周三、周五晚餐的就餐人数相对较少,调查表明周三下午没有安排课程,周五晚上课程安排的较少,部分学生在周五晚上开启“周末模式”,即和周末行为相似.

本文对就餐人数的总体预测,在用户总数为 12442 人的情况下,平均误差总人数仅为 160 人左右,这可以为学校食堂的后勤提供很大的备货依据和备餐依据,有效的防止供餐不足或过多剩余的现象;对早、中、晚的人数预测,能够结合用户的其他属性,为之后更具体的学生行为分析提供借鉴和实验基础.

4 总结

本文对一卡通数据采用了基于马尔科夫模型的就餐人数预测研究. 通过对用户的前三周就餐数据计算概率建立训练模型,预测第四周的就餐行为,预测就餐总人数和预测早、中、晚餐人数. 实验结果表明,基于马尔科夫模型的就餐数据预测模型平均误差率为 1.31%,平均早餐预测误差率为 1.04%,平均午餐预测误差率为 1.33%,平均晚餐预测误差率为 1.57%,预测结果较为准确. 准确的预测能更好地为学校食堂的后勤提供很大的备货依据和备餐依据,有效的防止供餐不足或过多剩余的现象;对早、中、晚的人数预测,能够结合用户的其他属性,为之后更具体的学生行为分析提供借鉴和实验基础. 但由于只有就餐数据,存在数据单一的缺点^[15],如何进行更全面的学生行为分析是下一步要做的工作.

参考文献

- 1 Feng JW. Study on the campus smart card system based on SaaS. International Conference on Information Sciences.
- 2 Shi Q, Qian Y, Gong Y, et al. The analysis and data mining of students' online data based on digital campus. Applied Mechanics & Materials, 2013, 397-400: 2326-2329.
- 3 Shi Q, Xu L, Shi Z, et al. Analysis and research of the campus network user's behavior based on k-means clustering algorithm. International Conference on Digital Manufacturing & Automation. IEEE Computer Society. 2013. 196-201.
- 4 王德才.数据挖掘在校园卡消费行为分析中的研究与应用[学位论文].哈尔滨:哈尔滨工程大学,2010.
- 5 张硕.基于 WEKA 的校园一卡通数据挖掘与分析[学位论文].武汉:华中师范大学,2014.
- 6 罗华群.校园一卡通数据的挖掘与应用[学位论文].上海:华东师范大学,2009.
- 7 欧阳圣.数据挖掘在消费行为分析中的应用[学位论文].长沙:湖南大学,2011.
- 8 陈建兵.利用校园一卡通数据优化高校贫困生认定系统[学位论文].成都:成都电子科技大学,2012.
- 9 李齐光,张平,赵鑫.基于一卡通消费行为分析识别经济困难生的方法和装置,2014.
- 10 Fan S, Li P, Liu T, et al. Population behavior analysis of Chinese university students via digital campus cards. IEEE International Conference on Data Mining Workshop. IEEE. 2015. 72-77.
- 11 徐剑.基于一卡通数据的消费行为与成绩的关联性研究分析[硕士学位论文].南昌:南昌大学,2011.
- 12 Ching WK, Ng MK. Markov Chains: Models, Algorithms and Applications. Springer Berlin, 2012, 83(483): xiv.
- 13 韩忠明,张晨,李斌.基于 Markov 模型的异常用户检测.计算机仿真,2014,31(6):316-320.
- 14 马千.基于校园卡消费流水数据的大学生就餐行为分析[硕士学位论文].长春:东北师范大学,2015.
- 15 金培莉,王晓震.校园一卡通系统决策支持实例分析.华东师范大学学报(自然科学版),2015(S1):525-529.