文章编号：1002-0268（2005）03-0108-03

基于数据挖掘的城市机动车停车调查分析研究

顾志康，李旭宏，杭 文
（东南大学交通学院运输与物流工程系，江苏 南京 210096）

摘要：进行城市机动车停车场规划前需要开展停车场调查，以得到停车场的基本特征参数、机动车进出记录和停车者的停车行为等信息。通过数据挖掘中的一些方法可以在数据库中挖掘出有用的信息。根据机动车进出停车场的记录计算一系列的停车指标，并借助多元统计分析中典型相关分析的方法，分析停车指标与停车场基本特征参数之间的关系。然后利用关联规则挖掘停车问询调查数据中的潜在信息，分析停车者的行为特征，以掌握更多的停车信息。

关键词：停车指标；停车行为；数据挖掘；典型相关分析；关联规则挖掘
中图分类号：U491.7 文献标识码：A

Analysis of City Motor Vehicles Parking Survey Based on Data Mining

GU Zhi-kang, LI Xu-hong, HANG Wen
(Department of Transport and Logistics Engineering, Transportation College
of Southeast University, Jiangsu Nanjing 210096, China)

Abstract: Parking survey should be carried out to obtain parking data before city motor vehicles park planning. Parking data include basic parameters of parks, in-and-out records of vehicles and parking behavior of drivers. Useful information can be found out through some methods of data mining. Calculating parking indexes according to in-and-out records of vehicles, analysis was conducted between parking indexes and basic features of parks with canonical correlation analysis. Then potential information were mined out in parking enquiry survey data by association rule mining. Having analyzed parking behavior, planners can learn more information and make right decisions in parking planning.

Key words: Parking indexes; Parking behavior; Data mining; Canonical correlation analysis; Association rule mining

0 引言

随着中国经济的不断发展，城市机动车拥有量也在快速增长。然而由于种种原因，城市机动车停车场的建设速度往往落后于机动车的增长速度，停车泊位不足，因此产生了停车供求严重不平衡的矛盾。

解决机动车停车难的问题必须对机动车停车场进行合理规划，而在规划前首先要进行调查，收集停车场的资料。借助数据挖掘中的一些方法，可以在停车调查数据库中挖掘出有用的信息。数据挖掘技术包括统计分析、分类预测、关联规则和神经网络等。本文的思路是结合停车指标与停车场基本特征参数，借助多元统计中典型相关分析的方法来研究影响停车指标的外部因素，同时利用关联规则挖掘停车者问询调查数据中的信息，分析停车者的停车行为，以此为城市机动车停车场规划提供帮助。

1 停车场调查数据的分析方法

停车场调查主要包括 3 类：停车场基本特征调查、进出车辆调查和停车者问询调查。

停车场基本特征可分为停车场自身特征、主体建筑特征和区位特征。停车场自身特征有车位数、停车

收稿日期：2004-01-13
作者简介：顾志康（1976—），男，江苏常州人，博士生，主要研究方向为交通运输规划与管理.

场面积、停车场建造形式等；主体建筑特征有主体建筑面积和就业（居住）人数等。

进出车辆调查是记录车辆进出停车场的时间，由此可以计算一系列的停车指标，比如周转率、泊位利用率和饱和度等。

停车者问询调查是抽取一定比例的停车者发放问卷，以此来了解停车者的出行目的和停车后步行距离等情况。

1.1　典型相关分析

根据实际经验，单项停车指标与单个停车场参数之间通常没有显著的相关性。但如果同时考虑 M 个停车指标与 N 个停车场参数的相关性，情况可能会有所变化。

在普通回归分析中，考察若干个自变量对某一个或某些因变量之间的联系，其优点是因变量与自变量的关系明确，缺点是整体性不够，不能反映一些因变量作为一个整体与另外一部分自变量作为一个整体之间的内在联系。而典型相关分析研究两组变量作为两个整体之间的相互依赖关系，同时它又保留了变量间两两相互关系。

假设 $x=(x_1, x_2, \cdots, x_p)$，$y=(y_1, y_2, \cdots, y_p)$是两个相互关联的随机变量。利用主成分思想，在两组变量中选取若干代表性的综合变量 u_i、v_i，每一变量都是原变量的一个线性组合

$$\begin{cases} u_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p \equiv a'x \\ v_i = b_{i1}y_1 + b_{i2}y_2 + \cdots + b_{iq}y_q \equiv b'y \end{cases}$$

只考虑方差为 1 的 x、y 的线性函数 $a'x$ 和 $b'y$，求使它们相关系数达到最大的一组，如果存在 a_1、b_1 使 $\rho(a'_1x, b'_1y)=\max\rho(a'x, b'y)$，则称 a'_1x、b'_1y 是 x、y 的第一对典型相关变量，同理可求第 2 对、第 3 对等。这些典型相关变量就反映了 x 与 y 之间线性相关的情况。

设停车指标为第 1 组变量 $x=(x_1, x_2, \cdots, x_i)$，停车场基本特征参数为第 2 组变量 $y=(y_1, y_2, \cdots, y_i)$。通过典型相关分析可以研究它们间的关系。

1.2　关联规则挖掘

停车问询调查的数据蕴含了停车者的行为特征，运用关联规则可以挖掘其中潜在的信息。关联规则挖掘就是从给定的数据集中搜索数据项间存在的有价值联系。设 $I=\{i_1, i_2, \cdots, i_m\}$是项的集合，任务相关的数据 D 是数据库事务的集合，其中每个事务 T 是项的集合，使得 $T\subseteq I$。设 A 是一个项集，事务 T 包含 A 当且仅当 $A\subseteq T$。关联规则是形如 $A\Rightarrow B$ 的蕴涵式，其中 $A\subset I$，$B\subset I$，且 $A\cap B=\varnothing$。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。如果 D 中包含 A 的事务同时也包含 B 的百分比是 c 的话，那么规则 $A\Rightarrow B$ 在事务集 D 中具有置信度 c。同时满足最小支持度阈值和最小置信度阈值的规则称作强关联规则，这些阈值可事先设定。

停车问询调查包括停车者出行目的、期望停车时间等数据，这些数据之间存在着一些可以指导停车规划决策的关联规则，比如停车目的为上班的人所能容忍的步行距离是多少、停车目的为公务的人停车后步行距离在 100m 内的占多大比例等。

停车问询调查数据库中所有记录组成事务集 D。假设停车者出行目的有 a 种不同取值，则项集 $I_1=\{i_1, i_2, \cdots, i_a\}$，同理可得其他字段的项集 I_2、I_3、I_4 和 I_5，那么 $I=\{I_1, I_2, I_3, I_4, I_5\}$。设 A 与 B 是两个不同字段的取值，规则 $A\Rightarrow B$ 有支持度 s 反映了所有记录中具有 A 属性的停车者所占百分比，置信度 c 反映了在这些具有 A 属性的停车者中同时还具有 B 属性的停车者所占百分比。

2　实例

2002 年合肥市进行了一次城市机动车停车场调查，主要包括配建和公共停车场，时间为调查日的 8∶00～20∶00。根据调查所得资料，进行停车状况分析研究。

2.1　停车指标与停车场基本特征间的关系

选取日周转率（x_1）、高峰小时周转率（x_2）、高峰小时饱和度（x_3）、日泊位利用率（x_4）、平均停车时间（x_5）5 个停车指标组成 1 组变量 X。

考虑停车场特征变量 Y 时首先选取每个停车场的车位数（y_1）、停车场面积（y_2）、主体建筑面积（y_3）和主体建筑就业（居住）人数（y_4）。因为部分配建停车场的车位向公众开放，所以把公用车位比例（y_5）归入变量 Y。根据调查，合肥停车场的建造形式（y_6）主要有地面停车场、地下停车库和停车楼，考虑它们的可达性不同，给 3 种停车场类型赋值，将其量化。合肥市区可分为护城河内、一环路内、二环路内以及二环路外 4 个部分，这里规定区位系数从城里向城外逐渐变小，由此得到区位变量（y_7）。停车场主体建筑类型也是影响停车指标的主要因素，但难以量化，所以在此处的变量 Y 里未加考虑。实际操作中可以先按主体建筑类型将停车场分类，然后在每一类里分别进行典型相关分析。

上述变量全部计算完毕后，将其标准化，以消除不同量纲的影响。变量 X 与 Y 做典型相关分析后，第 1

对典型变量显著相关，其似然率 χ^2 检验值为0.001 1＜α=0.05，典型相关系数为0.762 9，其他典型变量相关都不显著。第1对典型变量的线性组合为

$$\begin{cases} v_1 = -0.3304x_1 - 0.0751x_2 - 0.4732x_3 - \\ \quad 0.2774x_4 + 0.1845x_5 \\ w_1 = 1.0181y_1 + 0.7217y_2 + 0.1449y_3 + \\ \quad 0.3315y_4 + 0.1979y_5 + 0.0385y_6 + 0.3721y_7 \end{cases}$$

每个变量前系数的绝对值大小（载荷）反映了这个变量在典型变量中的影响力大小。在 v_1 中，高峰小时饱和度（x_3）的载荷最大，然后是日周转率（x_1），高峰小时周转率（x_2）的载荷最小。而且日周转率、高峰小时周转率、高峰小时饱和度、日泊位利用率与平均停车时间的系数前的符号相反，即前4个变量增长时平均停车时间下降，这也是符合常理的。

在 w_1 中车位数（y_1）的载荷最大，停车场面积（y_2）次之，这是因为周转率、饱和度、泊位利用率均与车位数有较大的关系，而停车场面积与车位数是正相关的。

接下来载荷较大的是区位变量（y_7），揭示了区位系数对停车指标有相当大的影响。所以在规划停车场时应考虑区位因素，同样类型的停车场在不同的区域应该有不同的建设标准。城市中心区采用高标准，因为停车难的问题集中体现在中心区；外围区域采用的标准可相对降低，避免土地资源的浪费。

主体建筑就业（居住）人数（y_4）的载荷排第四，因为有停车需求的人数与建筑物里的总人数成正比例关系。停车场建造形式（y_6）的载荷最小，这是因为目前合肥市的停车场主要以地面停车场为主，在此次调查中地下停车库和停车楼只占总数的10%，所以影响很小。

总体看来，变量 X 与 Y 相关主要是由于 x_1、x_3、x_4 与 y_1、y_2、y_4、y_7 相关，因此用高峰小时饱和度、日周转率和日泊位利用率作为停车场的评价指标较为合适，而车位数、停车场面积、区位和主体建筑人数是影响停车指标的主要因素。

进一步考察各个变量间的两两相互关系，相关系数在0.7以上的有日周转率与高峰小时周转率（0.878 4）、日周转率与高峰小时饱和度（0.702 6）、高峰小时周转率与高峰小时饱和度（0.837 6）、高峰小时周转率与日泊位利用率（0.727 0）、车位数与停车场面积（0.772 5）。

2.2 停车行为特征分析

本次停车者问询调查共调查了5个方面：出行目的、期望停车时间、停车后步行距离、能容忍的步行距离和选择停车场所考虑的首要因素。关联规则挖掘时采用Apriori算法，先找出满足最小支持度阈值的频繁项集，然后由频繁项集产生满足最小置信度阈值的强关联规则。这里最小支持度阈值设为5%，最小置信度阈值设为20%，阈值设的较小实际上是放宽了约束条件，这样可以了解到更为全面的信息。根据本次停车调查的特点，在所有的规则里应重点考察与停车目的有关的强关联规则。以下是挖掘出的部分规则，其中 X 代表停车者。

（1）停车目的⇒停车时间

停车目的（X，上班）⇒停车时间（X，1～2h）（s=24%，c=22%）

停车目的（X，上班）⇒停车时间（X，2～4h）（s=24%，c=37%）

停车目的（X，公务）⇒停车时间（X，小于1h）（s=40%，c=48%）

停车目的（X，公务）⇒停车时间（X，1～2h）（s=40%，c=31%）

停车目的（X，餐饮）⇒停车时间（X，1～2h）（s=12%，c=28%）

停车目的（X，餐饮）⇒停车时间（X，2～4h）（s=12%，c=30%）

从中可以看出，停车目的为上班的支持度为24%，表明在调查的所有人中上班目的占了24%，其停车时间为1～2h的有22%，2～4h的有37%。公务与餐饮目的支持度分别为40%和12%。很明显，上班、公务和餐饮是目前合肥市区机动车停车的主要目的，三者共占所有停车目的的76%。这3种停车目的相互比较，上班和餐饮目的的停车时间较长，而公务目的的停车时间相对较短。

（2）停车目的⇒步行距离

挖掘出的规则显示，无论是什么停车目的，停车者步行距离在100m内的置信度均超过40%，最高达到86%。说明大多数停车者都将车停放在离目的地近的停车场，这符合人们的停车心理。

（3）停车目的⇒能容忍的步行距离

出行目的为上班、餐饮的停车者所能容忍的步行距离在100m内的置信度是75%左右，出行目的为公务的停车者对应的置信度是66%，其他的则在55%以下。这从一方面反映了停车者希望停车后的步行距离较短，另一方面说明了不同停车目的的停车者能容忍的步行距离是有差别的，在停车（下转第118页）

步、换档综合控制试验。试验结果如图 4、图 5 所示。图 4 为加速踏板位置（即节气门开度）为 50% 的起步曲线，起步过程发动机转速维持在 2 000 ~ 2 500r/min 左右，且波动小；离合器接合时间短（约 2.2s），滑磨较小；起步平稳。图 5 为节气门全开时 1 档升 2 档的动力性换档曲线，在发动机转速为 5 500r/min 左右开始换档，在整个换档过程中综合控制发动机、离合器和变速器，换档时间短（约 0.9s），发动机功率中断时间短，充分发挥了动力性。

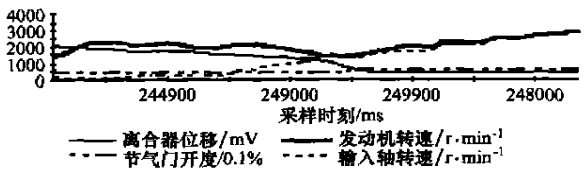

图 4 起步曲线

4 结论

提出了 TCU 与 ECU 通过信息共享来实现起步、换档过程中车辆动力传动系统综合控制的构想。分析

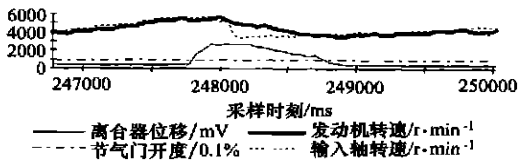

图 5 动力性换档曲线

了 CAN 总线协议和特点，将其定为 TCU 与 ECU 之间的通信总线。通过分析 TCU 和 ECU 共用的信息，定义了二者之间通信信息，并提出了基于 CAN 总线的 AMT 起步、换档过程的综合控制策略。试验表明，该方法在实现基本功能的同时，提高了 AMT 系统的起步、换档性能。

参考文献：

[1] 葛安林. 车辆自动变速理论与设计 [M]. 北京：机械工业出版社，1993.

[2] Robert Bosch GmbH. CAN Specification (Version 2.0) [S]. Stuttgart：Robert Bosch GmbH，1991.

[3] 邬宽明. CAN 总线原理和应用系统设计 [M]. 北京：北京航空航天大学出版社，1996.

（上接第 110 页）

场选址时应充分考虑到这些因素。

（4）停车目的⇒选择停车场考虑的首要因素

无论是什么停车目的，考虑停车场收费因素的置信度都没超过 30%，大部分在 20%以下，说明目前合肥市停车收费价格比较低，另外市区内私家车的出行比例不高。而停车场收费价格是一个重要的经济杠杆，将来可以用其调节停车需求的水平。考虑步行距离的置信度中，以上班、购物和餐饮目的为最高，超过 50%，文娱和家居最低，只有 30%左右。考虑安全因素的置信度中，家居目的最高，达到了 44%。

3 结语

本文运用典型相关分析的多元统计方法研究了停车场基本特征与停车指标间的相互关系，并利用关联规则挖掘停车者的行为特征，为停车场调查数据的分析提供了一种新的思路。当停车调查数据繁多时，借助目前已有的数据挖掘工具，用文中的方法进行分析研究，可以方便快捷地挖掘出数据库中蕴含的信息，为停车规划决策提供参考。

文中以合肥市机动车停车调查数据的分析为例来说明方法的运用。在该调查中未收集停车场收费价格的数据，这也是影响停车指标的重要因素。另外在停车者问询调查部分只调查了停车者出行目的和停车时间等几个方面的特征，在今后的停车调查中可进一步调查停车者的年龄、性别、职业等信息，为深入分析停车者的停车行为特征提供帮助。

参考文献：

[1] 徐吉谦，过秀成. 交通工程学基础 [M]. 南京：东南大学出版社，1995.

[2] 何晓群. 现代统计分析方法与应用 [M]. 北京：中国人民大学出版社，1998.

[3] 朱道元，吴诚鸥，秦伟良. 多元统计分析与软件 SAS [M]. 南京：东南大学出版社，1999.

[4] Jiawei Han，Micheline Kamber. 数据挖掘概念与技术 [M]. 北京：机械工业出版社，2001.