

文章编号: 1005-8451 (2016) 06-0065-05

城市轨道交通行为分析与数据挖掘决策系统研究

杨迎卯

(温州市铁路与轨道交通投资集团有限公司, 温州 325000)

摘要: 本文对智慧城市轨道交通行为分析与数据挖掘系统进行探讨, 从系统需求分析入手, 对系统功能、运行流程进行设计, 并结合城市轨道交通实际应用场景进行分析。

关键词: 城市轨道交通; 行为分析; 数据挖掘; 辅助决策

中图分类号: U231.92 : TP39 **文献标识码:** A

Behavior analysis and Data Mining Decision Support System for Urban Transit

YANG Yingmao

(Wenzhou Mass Transit Railway Investment Group Co. LTD., Wenzhou 325000, China)

Abstract: This article discussed the behavior analysis and Data Mining Decision Support System for the Urban Transit of smart city, designed the System from the system requirements, function and running process, combined with the practical application scenarios of Urban Transit to analyze them.

Key words: Urban Transit; behavior analysis; data mining; decision support

随着我国城市轨道交通的飞速发展、智慧城市理念的深化和实施, 智慧轨道交通建设提上日程。智慧轨道交通是以未来人类社会为理想目标, 以信息化的人类智慧为指导, 以轨道交通的规划设计、建设运营、控制管理为指导, 集“人、事、物”为一体的物联网为信息采集、交换与服务的基础支撑平台; 以智能的信息化决策和处理技术为基本手段, 通过对海量轨道交通信息的梳理、过滤、挖掘和利用, 构建“高效、便捷、安全、可视、环保和可预测”的现代轨道交通运输系统。

本文就智慧轨道交通中“如何通过对乘客的行为分析和数据挖掘, 实现对地铁商业和地铁广告业务的拉动, 达到便利乘客消费、增加地铁商业开发收入”的议题展开讨论。

1 系统需求分析

为了使地铁附加销售额增长, 根据地铁业务的现实情况, 提出两个系统对营销策略进行支持与决

策: (1) 行为分析与数据挖掘决策系统; (2) 个性化及广告推荐精准营销系统。本文主要介绍行为分析与数据挖掘决策系统。该系统是从运营者及商家的角度, 对客户的宏观行为进行统计与决策, 制定相应的营销政策、预警客户流失、推动销售额增长, 从而对传统业务与新业务经营进行分析和决策支持、优化成本。

1.1 地铁业务相关系统的需求

1.1.1 数据信息挖掘内容

- (1) 出行信息。各车站进出客流量实时数据信息, 各车站客流量历史数据信息;
- (2) 电子商务物流配送信息;
- (3) 消费偏好信息;
- (4) 生活需求、服务需求信息。

1.1.2 大数据挖掘目的

- (1) 挖掘市民的出行信息, 为地铁公司列车开行方案、行车组织提供决策支持; 为城市轨道交通规划提供决策支持。
- (2) 调整地铁商业服务内容, 便于商业服务更贴近用户。

收稿日期: 2015-11-10

作者简介: 杨迎卯, 工程师。

(3) 精准推送地铁公益信息、商业广告等内容。

1.2 地铁业务大数据获取途径

根据地铁业务的场景和特点,地铁运营过程中数据的采集主要为乘客使用的市民卡。除了乘坐地铁闸机刷卡外,市民卡还可以在以下的服务中使用市民卡:

- (1) 地铁商铺消费;
- (2) 地铁站提供的物流服务(快递自提点等);
- (3) 互动型广告、自动售卖机、报刊售卖、生活费用缴纳;
- (4) 手机充电、雨伞借用、针线包、急救药箱等便民服务;
- (5) 综合查询机(宾馆预订信息、景点信息、文化场所信息、消费场所信息、旅游线路信息、交通出行信息、火车票飞机票预订信息等)。

除市民卡(闸机、消费)以外,还可以考虑引入以下数据采集方式:

- (1) WIFI 上网热点;
- (2) 手机 NFC 技术;
- (3) 移动通信天线、手机定位;
- (4) 摄像头视频数据分析;
- (5) 拾音器;
- (6) 与地铁相关的智能穿戴设备;
- (7) 地铁里其他信息数据采集装置。

2 系统功能设计

- (1) 客户细分。
- (2) 客户流失预警。
- (3) 增量销售。销售策略制定与决策,提高营销活动利润。
- (4) 营销效果统计。根据主要收益来源,统计新产品营销效果,分析如何激活沉默用户;学习数据,对下一步营销策略进行预测。
- (5) 高价值客户维系。通过数据学习,完善高价值差异化服务。比如设立VIP、积分卡进行促销,回馈利润。
- (6) 地铁站客流量统计。
- (7) 商家及运营者盈利模型制定,财务经营状况预测。

(8) 商家及运营者内部企业投资及新业务数据推荐。通过不断学习数据,及数据走向预测,制定新的投资策略,开发新的业务,以数据驱动业务,增加企业收入及可持续科学发展。

(9) 其他相关传统的商业智能功能。

3 系统结构

基于上述系统需求分析和功能设计,城市轨道交通行为分析与数据挖掘决策系统结构如图1所示。

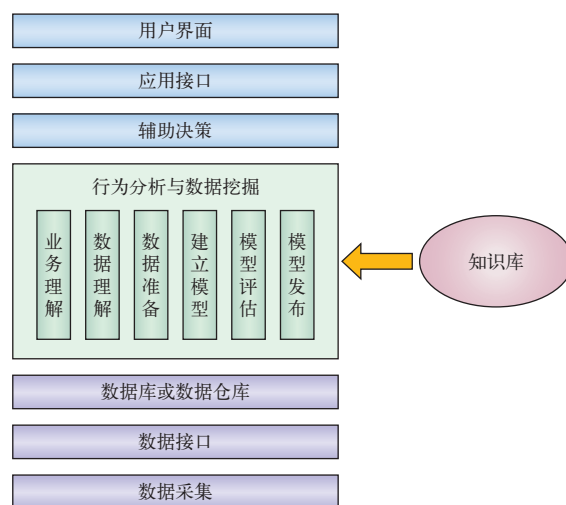


图1 系统结构图

4 系统实现流程

城市轨道交通行为分析与数据挖掘决策系统的实现过程分为6个阶段,如图2所示。

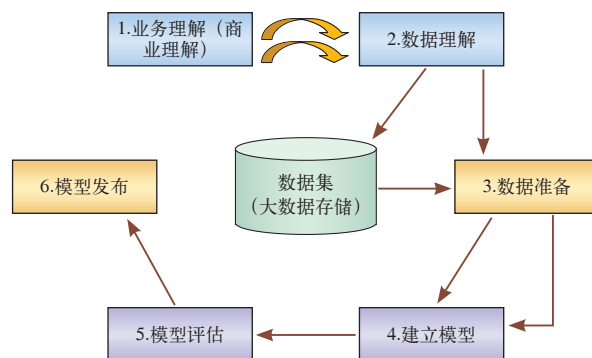


图2 行为分析与数据挖掘过程

(1) 业务理解:商业理解,项目目的。确定从数据中获得支持决策的信息。业务理解阶段是了解相关领域的有关情况,熟悉背景知识,弄清要求。在确定需求后,应对现有资源如已有的历史数据进行

评估, 确定是否能够通过数据挖掘技术来确定需求, 进一步确定数据挖掘的目标和制定数据挖掘计划。

(2) 数据理解: 通过数据发现经营状况。理解数据的统计学规律, 根据具体的商业理解确定决策方向, 进行数据分析。

(3) 数据准备: 根据数据理解(分析), 进行数据准备, 从数据集、数据仓库等数据存储系统中获取需要进行数据分析的数据。数据挖掘所处理的数据集通常不仅是海量数据, 而且可能存在大量的噪声数据、冗余数据、稀疏数据或不完全数据等。数据准备包括数据抽取、清洗、转换和加载, 具体包括数据的清洗、集成、选择、变换、规约, 以及数据的质量分析等步骤。

(4) 建立模型: 数据挖掘中的建模实际是利用已知的数据和知识建立一种模型, 并能将该模型有效地应用到未知的数据或相似情况中。在数据挖掘中, 可以使用多种不同的模型: 关联规则模型、决策树模型、神经网络模型、粗糙集模型、数理统计模型、时间序列分析模型等。

(5) 模型评估: 根据商业需求、模型数据学习后的结果, 及数据命中率, 修改完善模型, 达到更加接近想要得到的数据训练结果。数据挖掘得到的模式有可能没有实际意义或没有实用价值, 也可能不能准确反映数据真实意义, 甚至在某些情况下是与事实相反的, 因此对于数据挖掘的结果需要进行评估, 确定数据挖掘是否存在偏差, 挖掘结果是否正确, 哪些是有效的、有用的模式, 能够满足需求。

(6) 模型发布: 将评估后的数据模型发布, 供商业、决策人员使用, 为公司进行客户行为分析和决策支持。

5 应用场景实例分析

城市轨道交通行为分析与数据挖掘决策系统的整体应用如下。

5.1 业务理解

根据地铁各站点客流量及用户在地铁站内消费数据, 商家应进行产品销售策略的变动, 对高价值客户进行维系。产品销售策略如图3所示。

5.2 数据理解

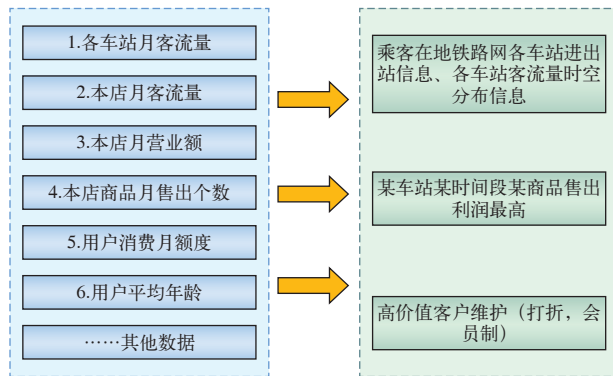


图3 大数据整体应用场景

(1) 客户消费倾向度分析如图4所示。

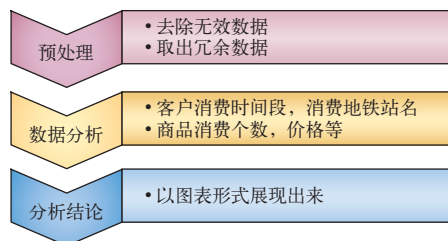


图4 客户消费倾向度分析示意图

(2) 高价值客户数据分析如图5所示。

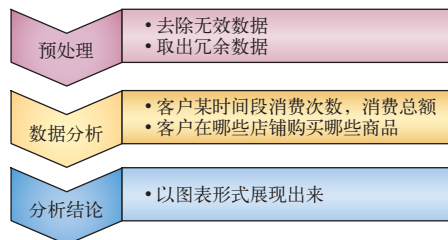


图5 高价值客户数据分析示意图

(3) 客户消费倾向度因素分析。

时间阶段: 上班、下班时间;
地铁站: 几号线、是否为换乘站;
商品: 品牌、价格。

(4) 高价值客户倾向度因素分析。

服务质量;
商品价格。

5.3 数据准备

从数据仓库, 或者其他数据存储系统中, 获取数据理解中需要的数据, 如图6所示。

5.4 建立模型

客户消费倾向度可采用分类模型, K-mean 算法, 如图7所示。

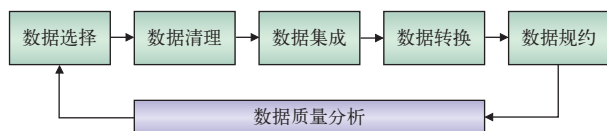


图6 数据准备示意图

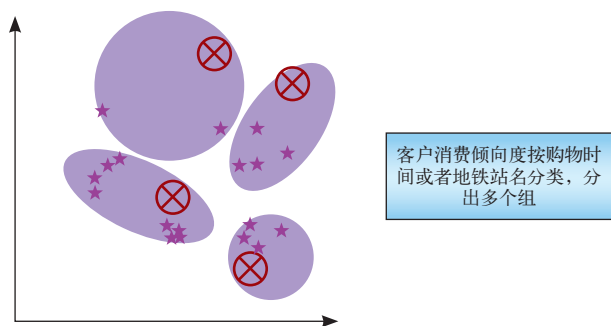


图7 K-mean算法示意图

5.5 模型评估

K-mean 算法对历史数据进行模型校验，得出是否可以刷卡这个选项对客户消费倾向度有较大影响，因此在分组中需加入此分组。

5.6 模型发布

以图、表形式展现数据学习结果，并向相关关系者提供分析数据。

6 关键技术

6.1 并行数据挖掘技术

当数据挖掘对象是一个庞大的数据集或是许多广泛分布的数据源时，效率就成为数据挖掘的瓶颈。随着并行处理技术的快速发展，用并行处理的方法来提高数据挖掘效率的需求越来越大。

并行数据挖掘涉及到了一系列体系结构和算法方面的技术，如硬件平台的选择（共享内存或分布式）、并行的策略（任务并行、数据并行或任务并行与数据并行结合）、负载均衡的策略（静态负载均衡或动态负载均衡）、数据划分的方式（横向的或纵向的）等。处理并行数据挖掘的策略主要涉及3种算法：并行关联规则挖掘算法、并行聚类算法和并行分类算法。

6.2 数据挖掘隐私保护技术

数据挖掘在产生社会、经济效益的同时也出现了乘客隐私泄露的问题。如何在防止乘客隐私泄露的前提下进行数据挖掘，是系统迫切需解决的问题。

基于隐私保护的数据挖掘是指采用数据扰乱、数据重构、密码学等技术手段，能够在保证足够精度和准确度的前提下，使数据挖掘者在不触及实际隐私数据的同时，仍能进行有效的数据挖掘。

受数据挖掘技术多样性的影响，隐私保护的数据挖掘方法呈多样性。基于隐私保护的数据挖掘技术可从4个层面进行分类：从数据分布情况，分为原始数据集中式和分布式；从数据隐藏情况，分为对原始数据进行扰动、替换和匿名隐藏等；从数据挖掘技术层面，分为分类挖掘、聚类挖掘、关联规则挖掘；从数据隐藏内容层面，分为原始数据隐藏和模式隐藏。

6.3 上网行为分析技术

乘客在地铁范围内的行为信息，除了地铁商铺消费、物流服务、出行 OD 信息、便民服务使用、综合查询服务和 NFC、无线定位、视频分析、智能穿戴等技术手段采集到的数据信息外，通过地铁站内公共 WIFI 采集到乘客上网行为也是非常重要的。因此，采用先进的上网行为分析技术非常必要。

上网行为分析技术包括 Web 页面内容识别技术和业务识别技术。

Web 页面识别技术包括无效页面过滤、内容清洗、文本切词、无效关键词过滤、文档语义识别等。

常用的业务识别技术主要有端口识别技术、流量特征识别技术、深度包检测（DPI, DeepPaket Inspection）识别技术、IP 地址识别技术、会话判别技术、拓扑特征识别技术等。

6.4 内容推荐技术方法

城市轨道交通行为分析与数据挖掘决策系统的主要目标之一就是广告等内容的精准推送。因此，如何通过行为分析和数据挖掘，实现精准的内容推荐是系统的重要目标之一。

目前，内容推荐的方法主要有：基于分类 TOP 排名直接推荐，基于用户喜好的个性化推荐，基于用户行为进行关联推荐等。其中，行为关联推荐法是通过大量乘客的数据进行分析挖掘，找出乘客的身份、特征、出行、消费等不同行为之间的关联关系，根据乘客已有的行为，与分析出的不同行为之间的关系进行比较，从而确定用户下一步的可能

行为。行为关联推荐法需要大量数据支持,对数据分析挖掘的要求较高,但这种方法在进行内容推荐时往往成功率较高。

7 结束语

随着智慧城市和城市轨道交通的发展,智慧轨道交通将进入快速发展的时期。数据挖掘决策在智慧轨道交通的数据智慧采集—数据智慧融合—数据智慧挖掘—智慧决策的链条中处于关键位置。随着网络技术和智能穿戴等数据采集技术的发展,智慧轨道交通系统所需要的数据将越来越丰富,而系统的数据挖掘和辅助决策将成为系统的核心。

责任编辑 王 浩

(上接 P54)

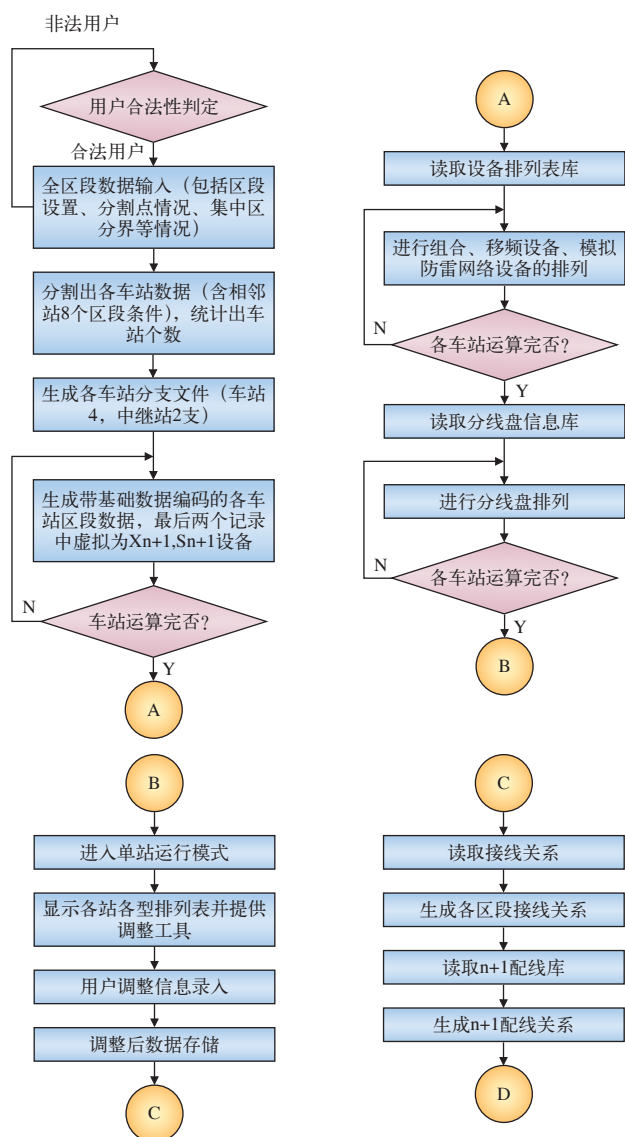


图2 室内辅助设计软件流程示意图

向运行时更多区段发送有效码的工程,可对基础代码第2位和第3位进行扩容,并对相关图纸、数据配置库内相关电路及配线信息等进行扩充定义,软件原则上可以进行适应性升级,但最高码序增加、反向编码区段增多时,由于编码、站联及与车站内的结合关系变得更为复杂,分区间关联度增加,使得零散组合定义、电路图库建设等工作量成级数倍增加,进而将增加软件图库和数据配置库的容量及软件处理的复杂度。

参考文献:

- [1] 肖彩霞. ZPW-2000R 无绝缘移频轨道电路系统技术说明书—系统描述 [R]. 黑龙江: 黑龙江瑞兴科技股份有限公司, 2012.

责任编辑 陈 蓉