

数据挖掘在智慧交通领域的应用分析

宋川 谢金江

重庆市交通规划研究院, 重庆 401147

[摘要] 智慧交通系统是以解决城市用地资源紧张、缓解城市交通拥堵为目标的信息系统。为了进一步挖掘智慧交通的潜力,提升其与城市发展需求的契合度,本文将围绕数据的应用进行讨论分析,探索大数据在智慧交通领域应用的相关方法。

[关键词] 智慧交通; 城市发展建设; 数据挖掘

DOI: 10.33142/sca.v4i4.4290

中图分类号: F49

文献标识码: A

Application Analysis of Data Mining in Intelligent Transportation

SONG Chuan, XIE Jinjiang

Chongqing Communication Research Institute, Chongqing, 401147, China

Abstract: Intelligent transportation system is an information system aiming at solving the shortage of urban land resources and alleviating urban traffic congestion. In order to further tap the potential of intelligent transportation and improve its fit with the needs of urban development, this paper will discuss and analyze the application of data and explore the relevant methods of big data application in the field of intelligent transportation.

Keywords: intelligent transportation; urban development and construction; data mining

引言

目前,智慧交通系统的主要功能包括交通的实时监控、公共车辆的运输管理、旅行信息服务的收集、车辆辅助控制等。上述功能以数据参数作为应用基础,不断加强数据的深入挖掘,才能更准确地甄别出影响道路运营的相关因素,实现交通运输的精确制导。为确保大数据在智慧交通领域的高质量、高效率应用,需要研究人员不断提升对数据属性的挖掘。

1 数据挖掘分析

数据挖掘是从大量、复杂、模糊类的数据中提炼具有应用价值信息的过程,随着智慧型城市建设的快速推进,城市发展过程的数据量呈指数增长,包含个人、家庭、企业、政府各类型海量的信息数据,使有效数据的提取难度大幅度增加。因此,寻找高效获取信息的工具成了智慧交通领域的迫切需求,而数据挖掘工作正是为了解决这一问题而生的互联网时代产物。具体工作方法可大致可分为以下几类:一是关联分析,即指若多个对象间存在一定的联系性,则可根据其中一个对象完成整体信息的推算,目的在于挖掘对象间的隐藏关系,能够通过兴趣度、可信度、支持度等关键指标完成相关性的度量,确保数据规则更符合真实情况;二是分类分析,即是寻找一个可以描述与区分信息类型的模块,用以预测未知内容的特点,能够利用特定的构造模型推导出训练集的分类;三是聚类分析,是以数据对象不存在预定类别作为应用前提,将数据依照相似性原则划分成多个类别,以此使同一类的信息对象保持极高的相似性;四是时间序列分析,是指按照时间顺序排列相关数字,并对此类内容加以分析,计算出重复产生几率较高的模式,进而推断未来的参数范围,该任务环节主要是对数据序列存在的依赖关系进行探究^[1]。

2 数据挖掘在智慧交通领域的具体应用研究

智慧数据中心需要构建可扩展的应用平台,将交通作为重点领域,实现大量、复杂、无规则数据的有效计算,并提炼一部分规则,用以挖掘数据中的隐藏规律与联系。为了使讨论内容更具有方向性,本文将以某城市的智慧型建设作为分析案例,根据该城市各部门提取的相关数据,结合城市发展特点、规模、人口数量进行深入研究,明确解决交通问题的实际方法,改造城市的管理系统,充分发挥数据挖掘技术的优势。

2.1 建立数据模型

根据相关数据信息采集,构建相应的数据模型,并保证信息具有较高的代表性,实现平台良好的管理能力,确保过程的稳定与流畅,一般情况下,管理闭环流程为:信息的实时获取、分析数据内容、提出管理决策、针对决策绩效进行研究、信息采集,这种不间断的重复类循环,能够确保数据模型的真实性和完整性。其整体框架可分为:数据准备,是指构建数据库,将监控信息有效收集;数据理解,通过从数据源中完成信息提炼,将其用于交通局、环保局、电信局,

经过业务专家分析并完成抽取转换后,建立预测模型和关联模型,之后投入到实际应用场景中,完成相关应用的评估,再根据需求与目标,完成内容优化,作为城市部件和政务信息的应用基础^[2]。

2.2 交通数据的处理

不管是人工输入还是机械设备进行自动采集,都可能在数据获取时出现一定误差,产生此类问题的原因在于:一是车辆检测器自身存在设计缺陷,在抓拍画面时,没有起到精准捕捉的作用,影响计量的准确性;二是人员个人的操作失误,在处理大量数据时没有准确输入相应参数;三是恶劣天气干扰,使电能设备无法有效维持高质量的运转,或是下雨天导致路面状况湿滑,影响车辆的正常行驶,造成数据出现缺失、不完整的问题。因此为了保证数据的真实、可靠,需要在使用前对交通数据预先完成相应处理,删除掉无用的、错误的、重复的信息内容。首先,要分析各个数据字段,判断数据内容是否符合业务含义要求,可通过使用阈值法完成错误信息的识别。其次,要使用正态分布原则将识别出的隐含异常函数进行彻底清除,并使用同时间段、相邻周期的历史平均值进行数据补偿。最后为了保证样本数据的全面性,需要将采集间隔时间设置为五分钟,以此获得全天约300个采样抓拍数据^[3]。

2.3 构建挖掘模型

将该城市某路段区域的抓拍数据按照时间序列进行聚类分析,例如:3月份的车流量为每小时810辆,人流量为每小时1240个;4月份的车流量为每小时800辆,人流量为每小时1400个;5月份的车流量为每小时1300辆,人流量为每小时1100个。将以上数据作为车流量变化趋势研究对象,并完成车流量模式包含时段特性的内容识别,为之后的车流量预测打下理论基础,同时在数据挖掘的聚类分析过程中分类的个数需要根据变化的统计量进行自动校对,保证算法的准确性。

2.4 实验结果分析

通过聚类法观察数据特点,并运用SPSS软件构建数据源、视图和挖掘结构,找到数据内部存在的时间特性,之后以聚类算法作为分析方式,将数据完成有效划分,从而进行类别间的关系判断,根据处理后的信息参数制定流量模式曲线图,帮助研究人员进行高质量的挖掘结果分析。其曲线图内容大致为:一类车流量曲线初始值为150辆,在4时左右达到最低值30辆,在8时达到最高峰500辆;二类车流量曲线初始值为140辆,在5时左右达到最低值40辆,在8时达到最高峰480辆;三类车流量曲线初始值为130辆,在4时左右达到最低值50辆,在9时达到最高峰450辆;四类车流量曲线初始值为120辆,在5时左右达到最低值60辆,在16时达到最高峰440辆;五类车流量曲线初始值为200辆,在5时左右达到最低值50辆,在14时达到最高峰430辆。

2.5 分析挖掘结果

根据上述内容表达的车流量模式曲线图和近三个月的车流量进行样本分布情况的研究,可得出以下结论:一,三、四月数据分布相对均匀,在各个类别中的分布量大致相同,而五月则主要集中在前三个类别中,总体趋势的差异性较大,主要是受天气情况影响;二,工作日的车流量呈现第一、第二类曲线分布情况,大体上相对接近,在全天24小时内有4~5个高峰点,与居民的上下班高峰一一对应;三,近三个月的周末流量走势一致,峰值不够明显,并且走势相对平缓,数值波动较小。与工作日不同,居民周末出行的时间规律性不明显。通过上述内容分析可以帮助城市中心平台更好的进行未来一段时间某区域的车流量、人流量预测,从而制定良好的应对措施,避免交通堵塞问题的发生。

3 结论

综上所述,通过对数据挖掘技术进行分析讨论,提出数据模型建立、交通数据预处理、挖掘模型构建、挖掘结果分析等数据挖掘在智慧交通领域的具体应用内容,从而提高城市交通管理与控制效果,加强智慧交通预测性低、滞后性强的问题解决效率,保证数据的真实、有效以及准确度,推动相关功能的进一步优化,维持城市稳定、可持续化的发展趋势。

[参考文献]

[1]魏万旭,方勇,丁泓十.基于视频数据挖掘的城市轨道交通车站行人交通行为特征提取系统研究[J].铁道运输与经济,2021,43(8):119-125.

[2]林丙梅.论数据挖掘技术在计算机网络病毒防范中的应用[J].网络安全技术与应用,2021(8):66-67.

[3]梁金如.基于数据挖掘的公共管理部门绩效评价研究[J].现代电子技术,2021,44(15):168-172.

作者简介:宋川(1986.1-),毕业院校:重庆大学,所学专业:建筑与土木工程领域工程硕士专业,当前就职单位:重庆市交通规划研究院,职务:道路所主任工程师,职称级别:高级工程师;谢金江,重庆市交通规划研究院,助理工程师。