

智慧高速数据挖掘平台设计及其实现

胡静¹, 张广浩², 冀金科², 孙晓亮³

(1. 郑州市规划勘测设计研究院, 郑州 450052; 2. 河南交通投资集团有限公司, 郑州 450016; 3. 交通运输部公路科学研究院, 北京 100088)

摘要:“智慧高速”以高速公路出行者为最终服务对象,立足路段范围内的主动化管理和精细化服务;在服务过程中产生了一系列多源异构的信息数据;如何消除信息孤岛现象,打通各种数据源之间的关联,发现潜在事件链是目前亟待解决的问题;为更好地探讨、解决智慧高速数据挖掘问题,以京港澳高速驻马店至信阳段为例,在较完善的高速公路基础设施之上,将先进的信息技术、通信技术、控制技术、传感器技术以及系统综合技术有效集成并应用于高速公路地面交通数据服务;在规划思路和内容、现有数据资源分类、数据资源架构、数据流程设计、数据共享设计等方面对智慧高速公路数据挖掘平台进行设计,对建立实时、准确、高效的数据共享挖掘机制有一定的借鉴意义。

关键词:智慧高速;数据挖掘;数据共享;数据中心

Design and Implementation of Data Mining Platform for Intelligent Expressway

Hu Jing¹, Zhang Guanghao², Ji Jinke², Sun Xiaoliang³

(1. Zhengzhou Urban Planning Design & Survey Research Institute, Zhengzhou 450052, China;
2. Henan Transport Investment Group Co., Ltd, Zhengzhou 450016, China;
3. Research Institute of Highway, Ministry of Transport, Beijing 100088, China)

Abstract: “Intelligent Expressway” takes expressway travelers as the ultimate service object and bases itself on active management and refined service within the range of road sections. During the service process, a series of multi-source heterogeneous information data have been generated. How to eliminate the phenomenon of information isolated island, get through the correlation between various data sources, and find the potential event chain is an urgent problem to be solved at present. In order to discuss and solve the data mining problem of intelligent expressway, this paper take the G4 (Beijing—Hongkong—Macao) expressway, Zhumadian—Xinyang section as an example, on the basis of relatively perfect expressway infrastructure, to integrate the advanced information technology, communication technology, control technology, sensor technology and system integration technology and apply to expressway ground traffic data service. Through designing the data mining platform for intelligent expressway in terms of planning ideas and contents, classification of existing data resources, data resource architecture, data flow design, data sharing design, etc. to provide a reference for the establishment of a real-time, accurate and efficient data sharing and mining mechanism.

Keywords: Intelligent expressway; data mining; data sharing; data centre

0 引言

“智慧高速”以高速公路出行者为最终服务对象,立足路段范围内的主动化管理和精细化服务。同时,兼顾路网层面的信息服务,逐步实现与交通运输部、所在省份路网中心、周边省份高速路网、地方道路管理部门、气象部门等之间的数据共享与交换。在日常管理方面,能够为途前、途中出行者提供衔接路网的交通运行信息^[1]。在应急救援方面,能够面向衔接路网提供统一的救援服务等。在服务过程中产生了一系列多源异构的信息数据。如何消除信息孤岛现象,打通各种数据源之间的关联,发现潜在事件链。为“智慧高速”的安全运行与指挥管理提供决策支持,提升高速公路运营效率,实现资源最优化配置是目前亟待解

决的问题^[2]。

为更好地探讨、解决智慧高速数据挖掘问题,以京港澳高速驻马店至信阳段(以下简称“驻信段”)为例,在较完善的京港澳高速公路基础设施之上,将先进的信息技术、通信技术、控制技术、传感器技术以及系统综合技术有效集成并应用于高速公路地面交通数据服务,从而建立起发挥作用的、实时、准确、高效的数据共享挖掘机制。

1 数据架构设计

1.1 规划思路和内容

1.1.1 统一规划,建立科学、完整的数据资源体系结构

“智慧高速”建设除满足所在路段现有业务管理、领导决策的需求之外,须着眼于上级高速公路联网收费机构的业务变革的需要,建设基础性、战略性交通行业数据资源。

1.1.2 采用分布采集、集中处理、集中交换、集中管理、全局应用的建设模式

“智慧高速”的数据采集既涉及到所在路段现有及后期

收稿日期:2019-05-20; 修回日期:2019-06-10。

基金项目:河南省基础与前沿技术研究项目(162300410214)。

作者简介:胡静(1987-),女,河南社旗人,硕士,工程师,主要从事道路规划设计方向的研究。

新建的高速公路管控系统，也涉及到省级高速公路联网收费机构共享的相关系统，各类信息系统数据存储方式、数据质量不尽相同^[3]。项目建设应利用数据交换和共享平台实现相关数据交换与共享，通过建立所在路段“智慧高速”的云数据中心，对形成的基础数据和业务数据库进行统一管理维护，满足各级相关管理部门和社会公众对不同层次数据的需求。

1.2 数据资源规划工程化方法模型

“智慧高速”管控系统的信息化建设规划应包括需求分析、系统建模、通信—计算机网络规划、应用系统开发计划等方面，与常规方法相同，在此不再赘述。

2 现有数据资源分类

通常情况下，所在路段管理机构及上级单位应在早期建设了若干业务系统，具有一定数据基础，但由于分布于不同应用系统数据库中，不便于统一管理与维护，难于保证数据的完整性、实时性，也无法满足共享信息资源的需求，对业务管理十分不利^[4]。为了合理规划现有的信息资源，需要梳理并分类各相关单位现有数据资源。通常情况下现有公路管理数据资源，包括但不限于：

1) 高速公路基本属性数据

高速公路基本属性数据包括但不限于：公路地理数据、环境情况数据、技术状况等相关描述数据，它包括路线概况、路基、路面、主要构筑物、沿线设施、交通量和沿线环境等基础信息资源^[5]。

2) 所在省份交通地理信息系统数据

包括但不限于：基础地形图空间数据、省高速公路及其结构建筑物的空间数据和属性数据、省国、道骨干交通网及其附属设施的空间数据和属性数据。

3) 高速公路养护、路政系统数据

高速公路养护系统的数据主要包括但不限于：大修、中小修过程管理数据、生态绿化工程数据、交通工程建设工程数据、雨雪应急保障工程数据等、养护人员上报数据等。

高速公路路政系统的数据主要包括：路政巡查车 GPS 数据、路政巡查车视频信息、路政人员上报数据等。

4) 高速公路交通流检测、气象检测系统、收费系统数据

高速公路交通流检测系统数据主要包括：微波交通流检测器采集的交通流量、密度、速度等相关数据。

高速公路气象检测系统数据主要包括：能见度、路面状况、大气温度、风速风向、大气湿度等。

高速公路收费系统数据主要包括：依托所在路段全线收费站，采集获取收费车型、行程时间等信息。

5) 高速公路应急资源数据

高速公路应急资源数据主要依托养护路政巡查道班获取的国家投资拨付的高速公路应急资源名称、位置、保管机构、维护人员等。

数据共享的核心思想是通过建立统一的数据标准和数据结构，各参与数据交换的接入节点则分别建立起内部数

据结构与公共数据结构之间映射关系，通过数据交换服务平台与远端适配器之间自动的交换任务调度，实现在两点之间的端对端数据共享^[6]。

3 数据资源架构

通常情况下，“智慧高速”管控系统数据资源库可以划分为基础数据库、业务数据库、专题数据库三类，形成全线数据集聚、跨业务共享、跨部门服务的数据资源架构（见图 1）。

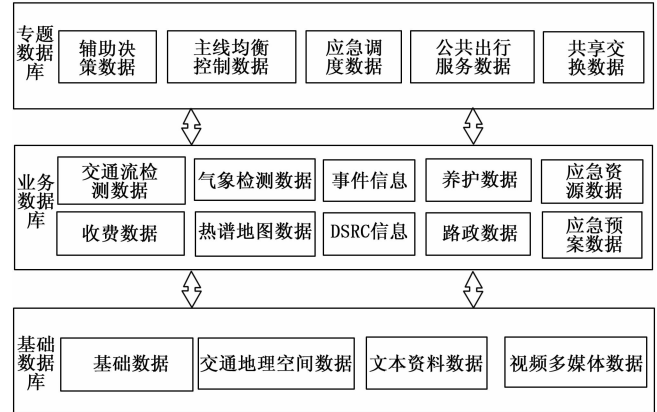


图 1 “智慧高速”数据中心数据资源总体架构图

3.1 基础数据库

基础数据库是支撑“智慧高速”系统内的各业务系统的具有共性特征的数据集合，通过对全线数据资源的管理对象和数据共享需求分析，确定若干满足于各类业务应用需要的、具有基础性、全局性的数据库^[7]。通过与省级高速公路联网收费机构、公安、气象等单位进行数据行抽取转换等一系列操作，在“智慧高速”云数据中心形成开放的全线基础数据库，供所在路段、上级管理部门、省级高速公路联网收费机构等单位共享使用。

基础数据库来源于省级高速公路联网收费机构的公路网基础数据、交通地理空间数据，省高速公路联网收费公司、公安、气象等上传下达的信息，以及所在路段等产生的文本资料数据、视频多媒体数据等（见表 1）。

3.2 业务数据库

业务数据库来源于所在路段监控中心、省级高速公路联网收费机构的业务数据库，通过对全线业务系统的结果数据提炼，形成与各应用系统业务相关的数据库，用于支撑全线各业务系统的应用^[8]。主要包括但不限于：路网相关管理数据、事件发生的相关信息、养护相关数据、路政数据、收费数据、应急数据等（见表 2）。

4 数据流程设计

4.1 数据流程总图

目前通行的“智慧高速”管控软件系统，一般包括高速公路运行状态监测与预警系统、高速公路运行状态综合评估与决策支持系统、高速公路主线交通均衡控制系统、高速公路应急指挥调度和处置系统、高速公路一体化出行信息发布系统、高速公路数据交换与共享平台等分系统^[9]。

表1 基础数据库

序号	数据名称	数据内容	数据来源
1	公路网基础数据	路线基础属性数据	省级高速公路联网收费机构
		路面基本信息数据	
		桥隧涵基本信息数据	
		路口属性数据	
		公路沿线设施数据	
		收费站基本数据	
2	交通地理空间数据	基础地理背景数据	省级高速公路联网收费机构
		地名及 POI 数据	
		路网数据	
		交通规则信息	
		遥感影像数据	
3	文本资料数据库	公告公示数据库、政策法规数据库、党委工作数据库等。	所在路段管理部门、省级高速公路联网收费机构
		政务图像、视频、多媒体培训教材及网络课件等信息,设备信息等。	所在路段管理部门、省级高速公路联网收费机构

表2 业务数据库

序号	数据名称	数据内容	数据来源
1	事件信息	人工报送事件信息	社会公众
		视频事件检测信息	视频事件检测器
		周边路网联动事件信息	省级高速公路联网收费机构
		基于移动终端的事件信息	社会公众
2	交通流检测数据	微波交通流检测数据(主线)(如断面交通流等)	微波交通流检测器
		微波交通流检测数据(服务区)(如服务区流量等)	微波交通流检测器
		DSRC 交通数据(如路段平均速度等)	DSRC 路侧交通流采集设备(RSU)
		收费数据(如行程时间等)	收费站
3	气象检测数据	视频能见度数据	视频能见度检测器
		能见度数据	能见度检测器
		路面状态数据等	气象站(含路面状况检测器、能见度检测器等)
4	视频信息	视频信息	视频摄像机
		移动巡查车车载视频信息	移动巡查车车载视频摄像机
5	移动巡查车数据	GPS 数据等	路政巡查车
6	移动终端数据	GPS 数据等	移动单兵
7	热谱地图数据	路面温度、温度谱等	路面温度检测车
8	养护数据	公路大修工程过程管理数据	所在路段管理机构
		公路中小修工程过程管理数据	
		公路生态绿化工程数据	
		公路交通工程建设工程数据	
9	路政数据	公路雨雪应急保障工程数据	所在路段管理机构
		行政许可数据	
		行政处罚数据	
		行政强制数据	
10	收费管理数据	路政巡查车数据	所在路段管理机构
		收费站数据	
11	应急数据	收费费用数据	所在路段管理机构
		应急物资数据	
		应急人员数据	
		应急预案信息等	

通过京港澳高速驻马店至信阳段为例,以数据为原始驱动,以系统为业务流导向,构建了一套应用系统和两大应用支撑平台的系统数据流程总图(见图2)。

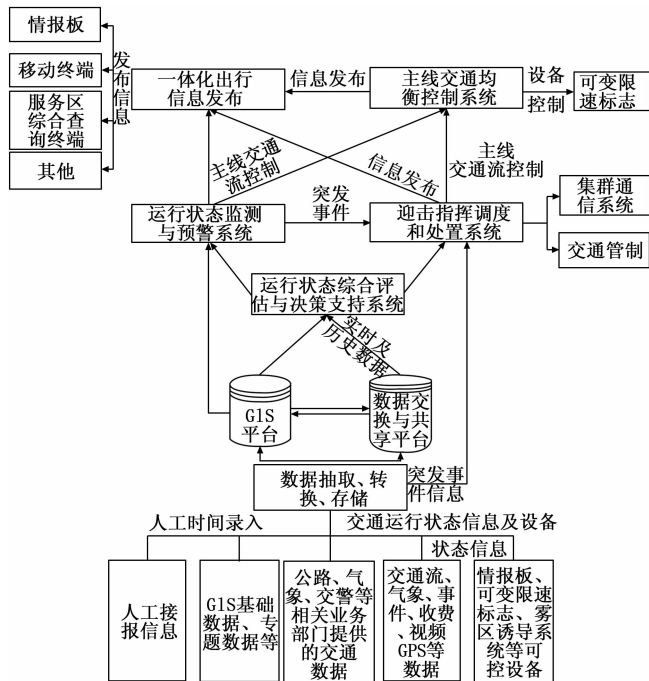


图2 智能高速数据流程总图(以京港澳高速公路驻马店至信阳段为例)

4.2 数据采集设计

以京港澳高速驻马店至信阳段为例,利用信息路由的先进技术,建立数据智能采集模型(见图3),与高速相关行业的数据以及高速业务数据通过相关应用平台系统进行抽取、整合,根据数据采集映射规则与驻信段数据中心进

行数据采集、聚合、挖掘和信息发布等^[10]。

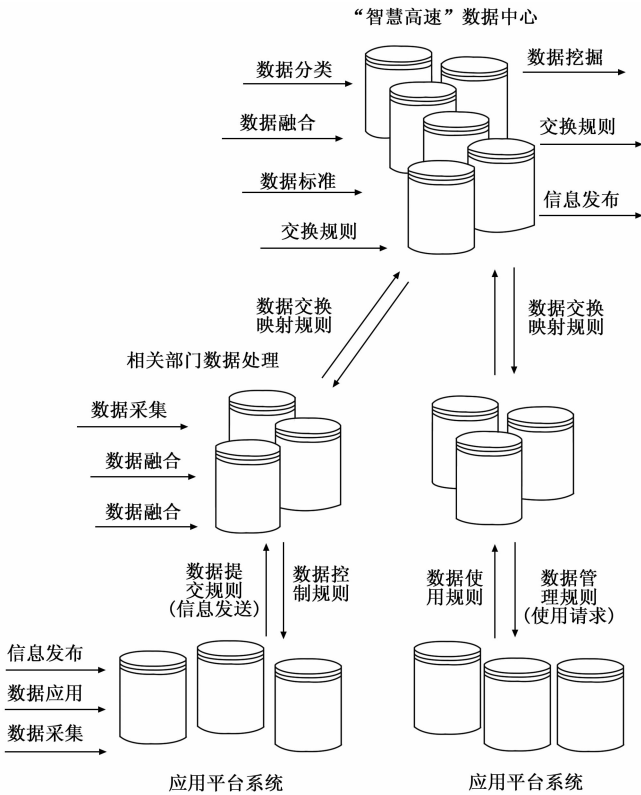


图 3 数据采集设计模型 (以京港澳高速公路驻信段为例)

4.3 数据分析处理流程设计

以京港澳高速驻马店至信阳段为例，依照既定数据标准体系，在源数据录入数据中心前对其进行分析处理，消除数据冗余和冲突（见图 4）。数据整合依照数据状况不同而有不同的整合方法，首先从与京港澳高速驻信段示范工程建设相关数据源那里获取原始数据，然后对获取的多源异构数据进行相应的预处理，预处理的方法有很多，但主要包括数据的清洗、数据集成、数据选择、数据转换等。其后，是数据分析阶段^[11]。

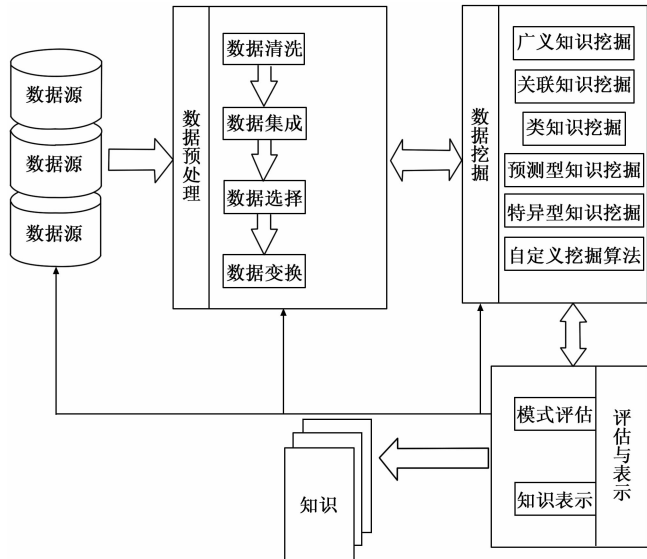


图 4 数据分析处理流程图

这里主要采用数据分析中常用的数据挖掘方法来对预处理之后的数据进行挖掘分析，可以使用的数据挖掘方法很多，主要包括广义知识挖掘、关联知识挖掘、类知识挖掘、预测性知识挖掘、特异性数据挖掘和自定义挖掘算法等，将进行数据分析以后得到的模型和规则等内容形成高速公路知识库，为智能高速的运营提供决策支持。

5 数据共享设计

5.1 数据共享机制框架

基于云平台架构，搭建数据交换与共享平台，在整合所在路段监控中心现有系统的数据资源，接入本工程外场设备采集的数据，建设智能高速公路数据中心^[12]。并将明确数据资源定义，建立完善公路基础信息、交通阻断、交通流量、公路气象、视频图像、应急资源、操作指令、高速公路运行指标等信息的采集、更新、交换与共享机制，制定高速公路运行数据共享与交换技术规范，确定信息交换方式，实现信息交换共享与协同运行。

基于以上设计目标和对高速公路业务的分析，“智慧高速”信息共享机制框架（见图 5），“智慧高速”数据共享模块之间关联为：共享平台从所在路段各种数据采集设备（主要为路段工程内装配的设备）获取多源异构数据信息；然后通过对多源异构数据的组织、存储、分析等发现数据之间的关联以及隐含的交通事件链；最后，通过平台提供的各种发布手段及时发布信息，为所在路段的车流和交通事故处理能力的提供强有力的支持。

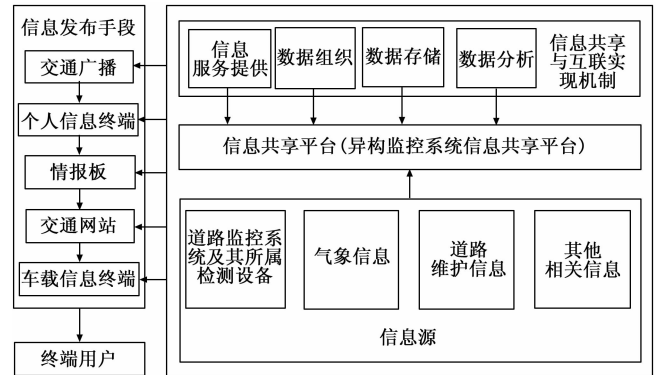


图 5 智慧高速公路信息共享机制框架

5.2 数据共享机制逻辑结构

根据以上具体的“智慧高速”信息共享框架图所示。从高速公路现场的数据采集，然后将采集后的数据发到信息共享平台进行整合分析，最后将分析报告发布或推送出去，进而完成信息共享机制的各个环节^[13]。数据交换与共享平台由交通信息管理、数据交换、数据管控、数据存储、数据分析、数据服务接口、数据挖掘与综合应用等几部分组成。（见图 6）。

6 工程实例

通过对河南省高速信息化建设和京港澳高速驻信段现状的调研，在现阶段京港澳高速驻信段基础设施改善的基础上，重点解决诸多信息系统的部署与采集设备的精度的提高。以京港澳高速驻信段拥有的大量交通流量数据，

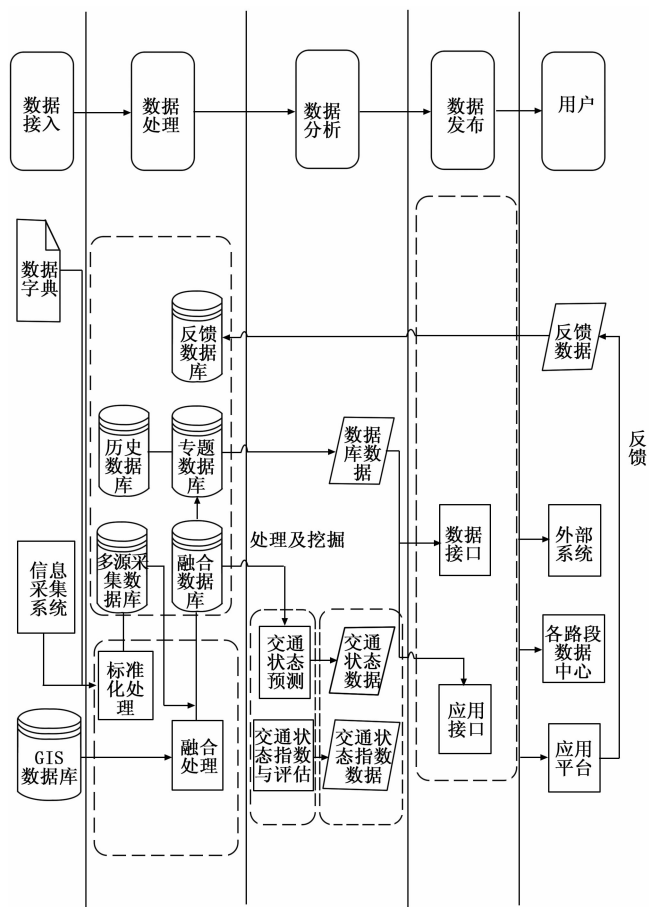


图 6 智慧高速信息共享机制逻辑结构

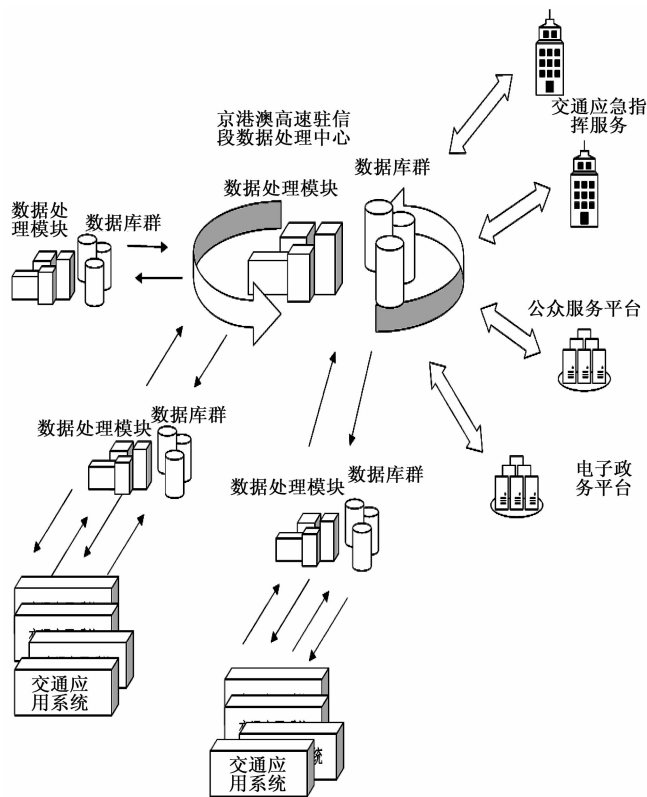


图 7 京港澳高速驻信段数据挖掘共享平台

为“智慧高速”数据挖掘共享提供基础的数据资源(见图 7)。解决现有系统没有很好地挖掘分析这些多源异构数据信息之间关联的问题,促进高速公路管理的精准化、智慧化发展,为河南省高速公路信息化的建设提供了有力的保障。其成功实施后将对京港澳全段乃至全国高速公路的数据共享机制建设起到示范作用。

7 结论

数据挖掘共享平台可以为“智慧高速”运行状态监测与预警系统提供充足的数据条件,可定期融合交通流量数据、气象数据及交通事件数据进行交通运行状态的综合展现和突发事件的预警。数据处理可以完成异构数据的格式转换和融合处理,通过数据清洗与校验步骤,为原型系统功能展示提供规范、稳定的基础资料。

数据挖掘共享平台负责对系统采集的多源异构数据进行清洗、合并、转换、剔除无效数据、数据装载等;完成对交通信息的优化、分析、加工、挖掘等工作。数据处理平台具有海量数据处理能力、强大的数据存储能力、良好的稳定性、可扩展性和先进性。每个查询分析功能可以提供定制的若干图、表,以实现特定的统计分析需求,方便用户直接使用。为管理者提供基于地理信息系统图层展现手段,相关分析结果数据可以通过地理信息系统展现。

数据挖掘共享平台是“智能公路”的核心组成部分,可以有力提升高速公路运营效率,实现资源最优配置。

全面提升通道型高速公路的智能化管理和服务水平,为下一阶段“智慧交通”的建设提供有益探索。

参考文献:

- [1] Martin J L, Gadegbeku B, Wu D, et al. Cannabis, alcohol and fatal road accidents [J]. Plos One, 2017, 12 (11): e0187320.
- [2] 刘犇莹. 国外高速公路信息化发展现状简析及其启示 [J]. 中国管理信息化, 2018, 21 (15): 134-136.
- [3] 张纪升, 李 斌, 王笑京, 等. 智慧高速公路架构与发展路径设计 [J]. 公路交通科技, 2018 (1): 88-94
- [4] 刘犇莹, 黄金科, 张广浩, 等. 京港澳高速驻信段智能管控需求分析 [J]. 中国交通信息化, 2018 (22): 77-78.
- [5] 张广浩. 智能高速公路理念在京港澳高速驻信段的应用 [J]. 中国交通信息化, 2019 (5): 92-94
- [6] PANG-NING TAN. 数据挖掘导论 [M]. 北京: 人民邮电出版社, 2011
- [7] 王笑京, 公路网运行监测与服务暂行技术要求 [M]. 北京: 人民交通出版社出版, 2012
- [8] 刘云霞. 数据预处理 [M]. 厦门: 厦门大学出版社, 2011.
- [9] 杨兆升, 于德新. 智能运输系统概论. 第 3 版 [M]. 北京: 人民交通出版社, 2015.
- [10] 李凯伦, 王建军. 高速公路多匝道远程协同控制技术研究 [J]. 计算机测量与控制. 2019 (4): 81-84
- [11] 刘 宇, 蒋 涛, 李建明. 基于马尔算法的高速公路视频监控图像能见度研究 [J]. 计算机测量与控制, 2017 (9): 60-63.
- [12] 马 艳. 战术导弹数据挖掘平台设计及其关键技术 [J]. 计算机测量与控制. 2017 (1): 96-98
- [13] 薛 蓓, 周延怀, 王晓兰. 基于云平台的大数据资源挖掘技术研究 [J]. 计算机测量与控制, 2017 (12): 275-278.