

基于OLAP的道路交通事故影响因素分析

惠宝锋, 冯桂莲, 王咏宁

(青海民族大学物理与电子信息工程学院, 青海 西宁 810007)

摘要: 本文通过数据挖掘相关基础进行数据库搭建, 选取2002—2016年数据进行OLAP分析, 通过数据切片及数据向下钻取, 将二维数据转换成三维数据进行深度剖析, 结果显示全国道路交通事故影响因素主要集中在机动车超速行驶及逆向行驶中, 其中驾驶人行为尤为突出。在Eviews软件中对数据进行预测分析发现道路交通事故存在非线性关系, 并且预测了未来三年的道路交通事故数量及伤亡数量, 数据显示出道路交通事故呈逐年下降趋势。

关键词: 数据挖掘; 联机分析处理; 交通事故; 预测分析

中图分类号: TP319 **文献标识码:** A

A Study on the Influencing Factors of Road Traffic Accidents Based on OLAP

HUI Baofeng, FENG Guilian, WANG Yongning

(School of Physics and Electronic Information Engineering, Qinghai University for Nationalities, Xining 810007, China)

Abstract: The paper constructs the database by means of related data mining, and selects the data from 2002 to 2016 for OLAP analysis. Through data slicing and data drilling down, the two-dimensional data are converted into three-dimensional data for further analysis. The results show that the main influencing factors of domestic road traffic accidents are mainly speeding and reverse driving of motor vehicles, especially the behaviors of drivers. The data analysis in Eviews shows that there is a non-linear relationship between the predictions and road traffic accidents. The data of accidents and casualties in the next 3 years is predicted which shows a declining trend in road traffic accidents.

Keywords: data mining; OLAP; traffic accidents; prediction analysis

1 引言(Introduction)

随着社会的发展与进步, 汽车已经成为人们生活中必不可少的交通工具, 目前汽车总量已经达到2017年中国大概有2.45亿辆, 随着汽车与交通运输给我带了便捷的同时, 也带来了事故频发的交通事故。早在1994年瑞典推出了交通“零伤亡愿景”计划, 力争在2020年前实现道路交通事故零死亡和零重伤率^[1]。国外的许多学者运用线性回归、Logistic回归、神经网络等手段分析交通事故严重程度与人的特性、车辆速度、道路条件和交通环境之间的关联性^[2]。

本文通过数据挖掘基本理论^[3], 建立道路交通事故数据库^[4,5], 通过OLAP(OnLine Analytical Processing, 联机分析处理)及经济计量视图软件Eviews, 对2002—2016年交通事故数据进行抽取分析, 切块与切片等数据处理, 试图通过交通

事故数量及类型分析^[6], 来揭示对道路交通事故影响的主要因素, 通过制定相关法律、法规等政策, 最终进行道路交通事故预测及分析^[7,8], 试图从数据挖掘的角度来预测交通事故未来的走向。

2 道路交通事故分析系统构建(System construction)

通过《中国统计年鉴》中的“交通事故”相关内容分别对全国各要省及直辖市从2002—2016年道路交通事故数据进行收集与整理, 并建立了交通事故基本数据库(Traffic Accident), 其内容主要涉及内容包含地区表及交通事故分类表, 地区表包含32个省及直辖市的交通事故总发生数、总死亡人数、总直接财产损失, 如图1所示。类表是全国数据汇总数据包括事故总数、特大事故、重大事故、机动车、非机动车, 以及行人等导致的交通事故总和, 如图2所示。

地区	发生数(起)	死亡人数(人)	受伤人数(人)	损失折款(万...)
北京	17645	1447	10424	6287.1
天津	9883	1119	6017	5313.1
河北	40693	6150	28929	14121.6
山西	25779	3029	15757	6285.7
内蒙古	9772	1943	7934	2247.2
辽宁	24785	4293	16992	12633.1
吉林	16266	2424	9600	4872.8
黑龙江	12378	2488	10933	4721.8
上海	42078	1503	15749	23883.0
江苏	53633	7184	30421	24714.8
浙江	68759	5901	32902	42145.2
安徽	27845	4333	20386	8639.2
福建	35867	4093	26977	14712.4
江西	17696	3355	14754	7150.5
山东	64786	9339	52006	18961.0
河南	40734	5893	33933	14026.8
湖北	20369	3024	15995	7178.4
湖南	23517	3355	23136	8691.0
广东	69555	10801	64788	30098.9
广西	16070	3554	15300	6884.4
海南	2334	373	1986	1166.1
重庆	14221	1146	11103	4186.3
四川	20980	5727	21972	12483.2
贵州	4064	1655	2704	2067.8
云南	19815	2964	10182	10739.8
西藏	803	354	732	568.1
陕西	13194	2285	10528	3998.9
甘肃	7929	2110	6360	3170.7
青海	1525	558	1344	631.9
宁夏	13315	971	7366	1860.0
新疆	10519	2539	9295	4366.4

图1 traffic accident数据库交通事故细分图

Fig.1 Detailed data in traffic accident database

类别	发生数(起)	死亡人数(人)	受伤人数(人)	损失折款(万...)
总计	754919	105930	546485	308787.3
重大事故	99327	98463	53894	52638.9
特大事故	3495	7467	9399	25538.7
机动车	705396	93289	508308	303185.6
汽车	540046	62737	342379	273089.9
摩托车	116488	20068	124847	16403.6
拖拉机	17095	3552	14304	3753
非机动车	24694	4701	20533	2827
自行车	19940	3856	16407	2279.8
行人和乘客	22888	7196	16447	2583.7
其他	1941	764	1197	191

图2 traffic accident数据库全国数据汇总分类

Fig.2 National data summary in traffic accident database

本文根据Traffic Accident数据库所建立的基本数据模型,分别建立表之间的关系试图,将全国各地区所发生的道路交通事故数据进行对比和分析,将其中交通事故较高的地区的具体数据引入SPSS统计软件进行线性分析,如图3所示。

从图3结果中可以看出大部分地区的道路交通事故发生频率在逐渐减少,只有个别地区故数量在上下波动,但大致呈现出递减趋势,且经济发达地区高于不发达地区,死亡人数也更多,北京地区受到严格的交通法规及政策的影响,交通事故发生率相对较低。

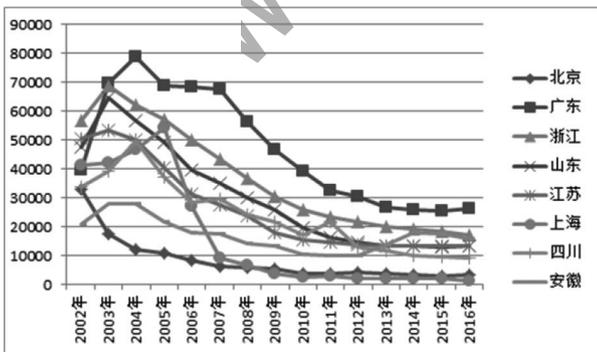


图3 主要城市交通事故分析对比结果图

Fig.3 Comparison results of traffic accident analysis in major cities

3 道路交通事故数据分析(Analysis of road traffic accident data)

我们通过OLAP对所获得、加工后的数据进行切片和钻取,对交通事故的各个影响因素进行分析和构建,从中得到影响交通且造成交通事故的影响因素主要包括:行人和乘客的因素、环境因素、驾驶人及车辆因素等。其中主要是驾驶人的因素最为直接。通过对驾驶人因素进行切块与切片分析,影响原因包含了超速、酒后驾驶、违法变道、不按规定路线行驶、违反交通信号灯、违法占道、占用公交专用道、无证驾驶、毒驾等多种因素,我们将选取一些典型的因素进行三维数组分析。如图4所示。

2014-2016年机动车造成交通事故类别

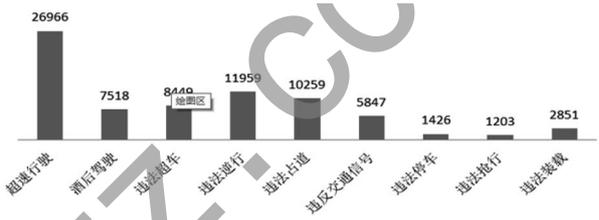


图4 2014—2016年机动车造成交通事故类别

Fig.4 2014—2016 traffic accidents caused by motor vehicles

在进行分析处理上我们首先对驾驶人数据进行了二位数组的采集,在建立的二位表的基础上向下进行钻取数据,通过在多层数据中来进行更详细的数据分析,最终得到以车类型(机动车与非机动车)、时间(2014—2016)、事故具体因素为三维数组图。如图5所示,从图中可以看出超速行驶与违法逆行造成的交通事故较多,监管部门应当加强监管。

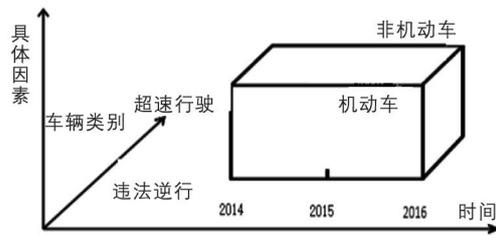


图5 驾驶者因素的三维切块

Fig.5 Three dimensional module of the driver's factors

在行人及乘客因素中通过切片可以看出行人主要是由于违反交通信号和其他影响安全的行为所导致的交通事故,环境因素中导致交通事故的主要因素是道路总体路况,气候环境及行人为按信号灯过马路和道路信号灯故障等。

4 道路交通事故预测分析(Prediction analysis)

由于道路交通事故数量随时间变化是非线性相关的,我们将数据引入Eviews中,对给定区间的交通事故数据进行非线性测量,预测2018—2020年道路交通事故结果如表1和表2

所示。

表1 2018—2020全国交通事故因素预测表

Tab.1 2018—2020 national traffic accident factor forecast

预测	机动车事故数	非机动车事故数	行人和乘客事故数
2018年	131562	9637	1564
2019年	117326	8766	1148
2020年	101424	8562	1004

表2 2018—2020全国道路交通事故预测表

Tab.2 2018—2020 traffic accident prediction table

预测	事故数	受伤人数	死亡人数	损失折款(万元)
2018年	127148	218132	60181	61078.1
2019年	114621	211543	58748	62944.6
2020年	102706	205305	56865	55067.1

根据表1和表2预测数据在Eviews进行校验，在具体应用中我们利用所获得的数据对数据进行分析，再此基础上利用间接代换方法对对数函数进行非线性模型预测，分别得出预测期内的道路交通事故数量分布图、全国道路交通事故受伤人数数量分布图、全国道路交通事故死亡人数数量分布图及预测数据图。如图6—图9所示。

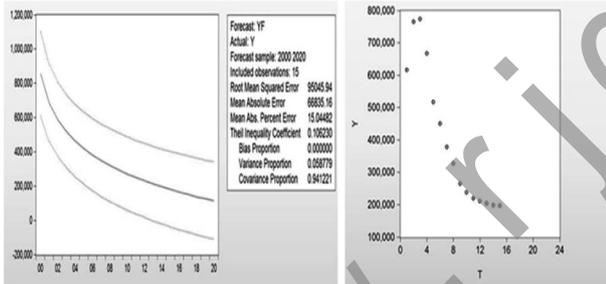


图6 2018—2020年全国道路交通事故数量预测分布及关系图
Fig.6 2018—2020 prediction of the number distribution of traffic accidents

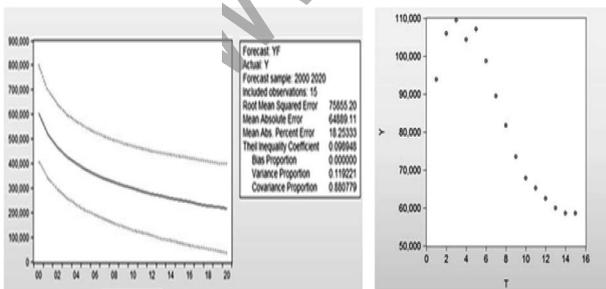


图7 2018—2020年全国道路交通事故受伤人数数量预测分布及关系图

Fig.7 2018—2020 prediction and distribution of injured persons in traffic accidents and their relationship diagrams

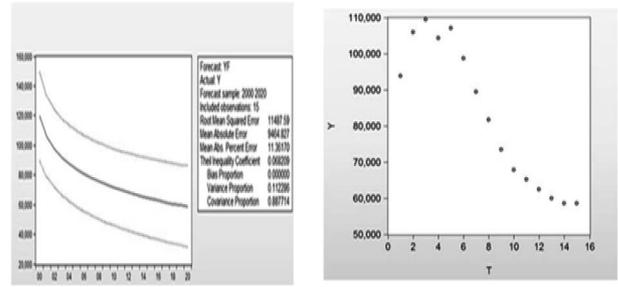


图8 预测2018—2020年全国道路交通事故死亡人数数量预测分布及关系图

Fig.8 2018—2020 prediction and distribution of death toll in road traffic accidents

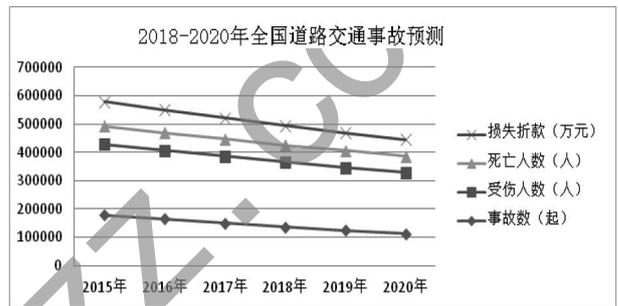


图9 2018—2020年全国道路交通事故预测分析

Fig.9 2018—2020 prediction and analysis of road traffic accidents in China

从以往数据及进行OLAP分析图中我们明显看出，随着社会发展与道路交通规则及相应的法律约束，及其宣传等工作，交通道路事故的总体事故发生率、交通事故死亡认识、其受伤人数及财产损失都有明显下降趋势，在进行预测的基础上对分析结果进行了间接替换法进行模型校验，表明符合非线性相关关系，证明所用分析方法可以进行道路交通事故数据的预测分析。

5 结论(Conclusion)

本文利用数据挖掘基础进行联机数据处理，建立相关数据库，并全国各个主要省份交通事故进行统计，在此基础上通过OLAP进行系统分析，取得影响交通安全的主要影响因素，通过对数据的切块及切片处理，向下钻取了三维数据表，显示出影响交通安全的因素主要来自机动车，在机动车分类中驾驶人成为主要因素，而且形成交通事故因素为超速驾驶和逆向行驶，在最后我们利用经济计量视图软件Eviews对全国交通事故进行了三年的预测分析，通过间接代换方法进行预测，结论符合非线性对数函数模型，符合预测分析要求。预测发现在未来三年随着时间的推移，交通事故发生数量在逐渐缩短，同时伤亡和死亡数量也呈下降趋势。

(下转第20页)