

基于车联网大数据的 UBI 系统研究

韩家群^{1,2}, 刘南杰^{1,2}, 黄波^{1,2}, 赵海涛^{1,2}

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003;
2. 南京邮电大学 网络基因工程研究所, 江苏 南京 210003)

摘要:在大数据和车联网时代背景下,提出了基于大数据的车联网保险系统的研究,即基于驾驶行为的车辆保险系统(Usage-Based Insurance, UBI)。该系统在智能车载终端 OBD 的应用、车辆数据收集、驾驶行为信息存储及处理、数据分析建模的基础上给出了合理的车险预测方案,并针对用户个性化服务进行了模块化的系统分析和处理。此外,在驾驶行为分析研究的基础上,给出了车险预测模型和 UBI 车险定价策略。系统的分析结果表明,在车联网大数据时代下的 UBI 系统在车险行业有很好的应用前景。

关键词:车联网;大数据;车辆保险系统;车保险

中图分类号: TN911

文献标识码: A

文章编号: 1673-629X(2016)12-0026-04

doi:10.3969/j.issn.1673-629X.2016.12.006

Research on UBI System Based on Big Data in IOV

HAN Jia-qun^{1,2}, LIU Nan-jie^{1,2}, HUANG Bo^{1,2}, ZHAO Hai-tao^{1,2}

(1. College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

2. Network Gene Engineering Research Institute, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Under the era background of Big Data and IOV (Internet of Vehicle), the research on the insurance system for IOV based on Big Data is carried out and the UBI (Usage-Based Insurance) system is proposed. The proper scheme of forecasting automobile insurance is presented through equipping the vehicle with OBD (On-Board Diagnostics), collecting the vehicle data, storing and processing the information about driving behavior, modeling the analysis on data by UBI system. The modular system is analyzed and processed in terms of personalized service of users. In addition, on the basis of analyzing the driving behavior, the model of forecasting the automobile insurance and the pricing policy of UBI are proposed. The analysis results show that the UBI system is of broad and potential application prospects in the field of automobile insurance in the era of Big Data.

Key words: IOV; Big Data; UBI; automobile insurance

0 引言

2013年国内的财险行业突破了亿万元大关,比2009年增加了21.3%,尽管如此,但保险行业的盈利仍然不理想^[1]。由于传统的机动车辆保险只考虑车辆购置价、购车类型等,车辆保险模式极其单一,没有考虑驾驶行为对机动车辆保险的影响,导致大部分优质的车险用户为少数因恶劣的驾驶行为造成高额理赔的用户买单,因而使得投保人的车险保费设定存在严重不合理的现象^[2]。

相比之下,国外的保险费率更为灵活。例如,美国

未婚低龄保险费率最高(缺乏责任感,易出现车辆事故);德国新手费率高(出险概率高);加拿大周末用车比上班用车费率低(出险概率低)。国外积极推广的UBI保险^[3],取得了一定的成效,未来UBI的车联网保险模式也将被持续推广与应用。

随着互联网时代的到来和技术全球化的发展,移动互联网正在不断渗透到社会、经济各个领域,同样地互联网下的车联网也正向着汽车保险行业渗透,因而基于车联网的汽车保险行业有巨大的发展前景。其中,车联网技术、大数据技术等是未来保险行业发展的

收稿日期:2016-01-19

修回日期:2016-05-11

网络出版时间:2016-11-22

基金项目:国家(青年)自然科学基金(61201162);政策引导类计划(产学研合作)—前瞻性联合研究项目(BY2015011-01)

作者简介:韩家群(1991-),女,硕士研究生,研究方向为车联网大数据;刘南杰,博士,教授,研究方向为泛在通信、车联网、智能交通。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20161122.1231.048.html>

核心驱动力^[4]。在这样的时代背景下,对车联网保险进行了研究,并创新性提出了大数据时代下的 UBI 系统研究。该系统从车主的驾驶行为习惯、行车里程、购置价格及车辆类型等方面进行综合分析,在车联网保险的第一代基于按里程付费(Pay As You Drive, PAYD)的车保险到第二代考虑驾驶安全(Pay How You Drive, PHYD)的车保险基础上,提出车和人相结合模式厘定车险方案,打破传统的只对车或者人单一的分析模式。文中分析处理的数据均是由车载终端 OBD 收集的真实驾驶行为数据^[5-6]。

1 智能车载终端

车联网(Internet of Vehicles, IOV)是通过 OBD、GPS 等装置,完成车自身状态和环境信息数据的采集^[7],通过互联网将采集的数据传输到中央处理器并对数据进行分析处理,并对不同需求的车辆进行有效监管和提供综合服务的系统,实现车辆的智能化控制。

车载诊断(On-Board Diagnostics, OBD)是车联网的核心技术,融合了汽车智能感知模块、汽车与互联网的连接模块、汽车系统和部件(发动机、排放控制系统等)的监测模块,实现车辆状况的实时记录和报告。OBD 模式的车联网系统是由 OBD 终端、后台系统、手机 APP 这三个主要部分组成。图 1 为 OBD 模式下的车联网模型,车辆内置的传感器具有智能感知功能,车载诊断 OBD 通过控制局部网(Controller Area Network, CAN)与总线相连,获取电控单元(Engine Control Unit, ECU)中的车辆状态信息。该模式系统与物联网的逻辑组成类似,由数据采集、数据分析处理、数据报告等组成。

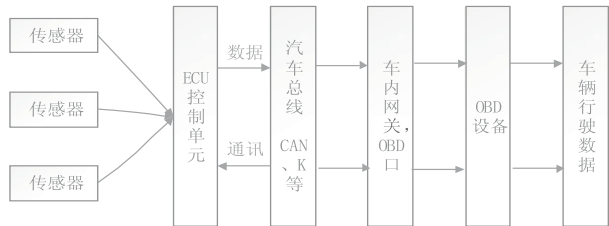


图 1 OBD 模式下的车联网模型

2 大数据

大数据是来源于人类活动,通过记录人类某些行为而得到的数据。在人类发现数据其他价值以前,数据只是一个数量上的理解,由于互联网的快速发展,数据出现井喷式发展,使任何数据背后都有其自身的信息价值。文中提出的 UBI 系统研究正是在大数据分析 and 处理的基础上,针对车保用户专业化和个性化服务给出了合理车险模式。

大数据是新一代对大量的各种样式的数据进行

高速捕获、提取、分析和处理后得到数据规律,从而取得额外价值的技术。它具有海量性(Volume)、快速变动(Variety)、多样化(Velocity)、信息价值(Value)和真实性(Veracity)五大特征。大数据不在于数据本身的信息意义,而是它能衍生出多维度潜在的信息价值^[8-9]。

大数据处理的生命周期包含数据源、收集、存储、分析处理和预测等过程,生命周期体现了不同阶段对数据实施不同的处理策略。

3 车联网大数据时代的 UBI 系统

如图 2 所示,大数据时代的 UBI 系统主要由数据源、数据的处理、数据的分析和预测模型等部分组成。

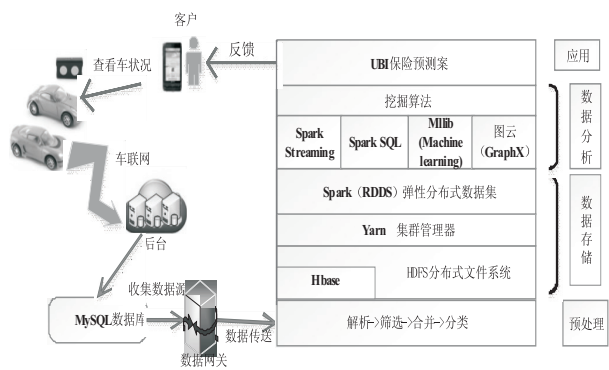


图 2 基于大数据的 UBI 车险系统

3.1 数据源

机动车辆中安装的 OBD 对车辆的各个系统进行实时监测,车联网的应用实现了从客户端-服务器(Client/Server)成功连接,服务器是整个应用系统的资源中心,客户端发送的数据传送到数据库服务器,客户端也可以对数据库进行访问。文中数据源存储在关系数据库 MySQL 中,通过数据网关传输到分布式数据库管理系统中。MySQL 具有体积小、速度快、成本低等特点,适用于车况中快速产生数据,及时更新数据库中的数据,去除了冗余的数据信息,减少了网络资源的浪费。

3.2 数据处理

数据处理包含数据预处理和数据存储两部分,数据预处理可以获取对车保险预测方案有价值的信息。通过对驾驶行为有关的数据解析,筛选出文中提出的 UBI 系统所需的数据,如每日四急(急刹车、急加速、急减速、急转弯)次数、行驶里程、出行时间、超速次数等数据,然后对这些数据进行分类、合并,并存储到分布式数据库 HBase 中。HBase 是一种基于 Hadoop 的项目,也称 Hadoop 分布式文件系统(Hadoop Distributed File System, HDFS)^[10]。它是一个非结构化数据存储的分布式数据库,使用 Zookeeper 管理集群,在架

构层面上分为 Master(Zookeeper 中的 leader)和多个区域服务器(Region Server,RS)。基本架构如图 3 所示。其中,RS 是集群中的一个节点,每个 RS 可以负责管理多个 Region,每个 Region 只能由一个 RS 提供服务。HBase 中需要多个 Region 存储数据,HBase 给每个 Region 定义一定的范围,落在规定范围的数据,就会分配给规定的 Region,从而把负载分到各个节点上,这就是分布式存储的过程及优点。

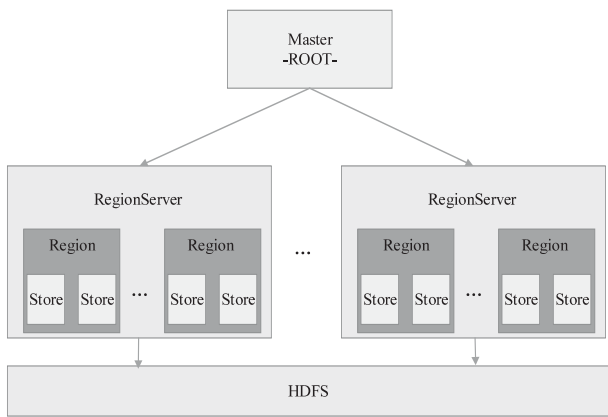


图 3 HBase 基本架构

YARN(Yet Another Resource Negotiator)是分布式集群的资源管理器。MapReduce1 架构在整个集群上执行 Map 和 Reduce 任务并报告结果,但在大型集群中,当集群节点超过一定量时,就会出现级联故障,级联故障通过网络泛洪形式导致整个集群严重恶化。为了克服 MapReduce1 的这种缺陷,采用 YARN 分层集群管理框架的技术,能使集群共享、可伸缩和更可靠。YARN 分层结构是资源管理程序(ResourceManager)将各部分资源传给基础节点代理程序(NodeManager),NodeManager 启动和监视基础应用程序执行和资源管理(CPU、内存等资源分配)。

Spark 是一个基于内存计算的集群计算系统,它的核心是弹性分布式数据集(Resilient Distributed Datasets,RDD)。Spark 的所有操作基于 RDD,RDD 是容错的、并行的数据结构,RDD 是一个不可修改的分布的对象集合。每个 RDD 由多个分区组成,每个分区可以同时分布在集群中的不同节点上计算。RDD 的分区特性与并行计算能力,使得 Spark 可以更好地利用可伸缩的硬件资源。若将分区与持久化二者结合起来,就能更加高效地处理海量数据^[11]。

文中收集了 1 000 辆汽车数据,并分析处理驾驶行为相关数据信息,如四急、行驶里程、最大瞬时速度和出行的时间。图 4 是基于驾驶行为分别从每天驾驶的距离、每天四急的次数总和、最大速度和最晚出行时间四个方面所得数据的柱状图。通过这些数据的分析,得出相应的驾驶行为处理结果,为文中大数据时代下的 UBI 车险方案提供有力证据。

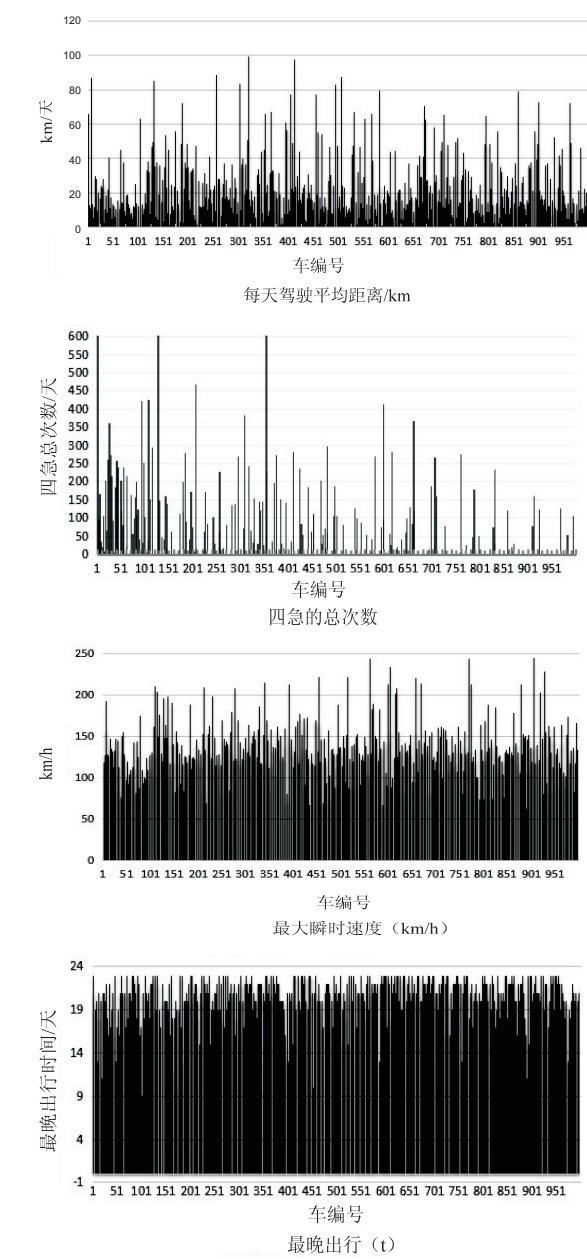


图 4 基于驾驶行为的数据分析

3.3 数据建模分析

数据建模分析是针对预处理提取的数据特征,得到想要的结果。在数据提取后,常使用的是 Spark 算法。Spark 常用的应用有 Spark SQL、Spark Streaming、MLLib、Graph 等。Spark SQL 使用 RDD 实现 SQL 查询;Spark Streaming 流式计算,提供实时计算功能;GraphX 图计算框架,实现了基本的图计算功能,常用图算法和 pregel 图编程框架;MLLib 机器学习库,提供常用分类、聚类、回归、交叉检验等机器学习算法并行实现,如朴素贝叶斯、逻辑回归、决策树、神经网络、TFIDF、协同过滤等算法,在 MLLib 中已经存在,只需将数据带入,调用比较方便。

3.4 车保险的预测方案

文中提出的 UBI 系统对不同的驾驶行为给予不

同的保险费率,并提供个性化的增值服务。在大数据分析处理后,该系统提供的机动车辆保险的实施方案如下:

给每个用户每天设置一个基总分数值(如 100 分),四急/每日行驶总里程/每日超速次数/每日夜间

行驶时间按 5:2:2:1 分配总分值,即 50 分/20 分/20 分/10 分。

表 1 是根据驾驶行为制定的评分规则,通过累计的分数,判断一个人的驾驶行为的优良性^[12]。

表 1 评分规则

驾驶行为	评分规则
四急	两公里内发生一次扣 20 分,两公里到十公里发生一次扣 15 分,十公里以后发生一次扣 10 分;急刹车、急转弯按 100% 比例扣分,急加速、急减速按以下规则 90% 比例扣分。市区高峰按 100% 扣分,市区非高峰按 75% 扣分,非市区高峰按 75% 扣分,非市区非高峰按 60% 扣分
行驶总里程	总里程阈值为 50 km,超过安全阈值 20~40 km 扣 2 分,40~60 km 扣 6 分,60~70 km 扣 15 分,70 km(不含)及以上得 0 分
超速次数	设置城区最高速度 80 km/h,高速最高速度 120 km/h。满分 20 分,发生 1~2 次扣 5 分,3~4 次扣 10 分,5~6 次扣 15 分,6 次以上不得分
夜间行驶时间	夜间行车时间 1~2 h 扣 2 分,2~4 h 扣 5 分,4 h(不含)及以上得 0 分,夜间时间判断为 0:00 至 06:00,23:00~0:00

根据方案累计一年的得分情况记为 Sum,驾驶的天数即算入计算分数的天数为 Day,平均得分记为 Avg:

Avg=Sum/Day

为了防止恶意做假行为,天数 Day 有一定的规定:若 Day<100,视为最低等级,100≤Day<250,则在原来的 Sum 上乘一定比例 50%,若 Day≥250 则按照原 Sum 计算。

根据 Avg 分析将不同客户分为不同的等级,Avg≥80 为五星级客户,60≤Avg<80 为四星级客户,40≤Avg<60 为三星级客户,20≤Avg<40 为二星级客户,0≤Avg<20 为一星级客户。

4 结束语

不同星级的客户可以承担不同车保险费率,保险公司应奖励优质客户(即星级高的客户),在下一年的保险中给予优惠活动,同时,惩罚劣质用户(即星级低的客户),可以提高来年投保车辆的保险费率。此外,获取的数据还可以为客户提供个性化服务,如根据驾驶习惯和经常去的地方,适时为其推荐地方特色和商店活动信息,对于驾驶行为不良的用户给予及时提醒等服务。

大数据时代下的 UBI 系统从大数据的获取、存储、分析、建模等方面进行了详细的描述,以四急、驾驶里程等为依据制定了合理的 UBI 的车险费率模型。该系统具有真实性和实际价值意义,在车险行业具有很好的应用前景^[3,13]。

参考文献:

[1] 迪纳科技. 保险行业车联网解决方案白皮书[EB/OL]. [2014-04-01]. <http://www.cpsdna.com/article-545.htm>

ml.

[2] 彭江琴,刘南杰,赵海涛,等. 智能 UBI 系统研究[J]. 计算机技术与发展,2016,26(1):142-146.

[3] Kusek G,Kilic I. Project-based application on big data usage [C]//2015 fourth international conference on agro-geoinformatics. [s. l.]:[s. n.],2015:89-92.

[4] 刘文鹏. 大数据时代的汽车保险[J]. 经营者,2015(2):166-167.

[5] 乔 木. 大数据语境下 UBI 发展现状及趋势研究[J]. 现代商业,2015(1):53-54.

[6] 梁小英,朱园丽,赵一衡. 科技引领未来,专业创造价值——大数据时代下的 UBI 产品探索[J]. 金融电子化,2014(9):28-29.

[7] 刘南杰. 崛起的车联网[J]. 音响改装技术,2013(11):50.

[8] Demchenko Y,de Laat C,Membrey P. Defining architecture components of the Big Data Ecosystem [C]//2014 international conference on collaboration technologies and systems. [s. l.]:[s. n.],2014:104-112.

[9] Tekiner F,Keane J A. Big data framework [C]//2013 IEEE international conference on systems,man and cybernetics. [s. l.]:IEEE,2013:1494-1499.

[10] Pandey S,Tokekar V. Prominence of MapReduce in big data processing [C]//2014 fourth international conference on communication systems and network technologies. [s. l.]:[s. n.],2014:555-560.

[11] Riggins F J,Wamba S F. Research directions on the adoption, usage, and impact of the internet of things through the use of big data analytics [C]//2015 48th Hawaii international conference on system sciences. [s. l.]:[s. n.],2015:1531-1540.

[12] 彭江琴,刘南杰,仲 浩,等. 基于 GID 的 UBI 系统研究[J]. 微型机与应用,2014,33(22):51-53.

[13] 郁佳敏. 车联网大数据时代汽车保险业的机遇和挑战[J]. 南方金融,2013(12):89-95.