

An Elevator Early Warning System Based on Hadoop and Pig

Jianbing Lin¹, Zhongfu Xu²

¹Department of Computer Science and Technology, Information Engineering College, Putian University, Putian Fujian

²Putian Branch of Fujian Special Equipment Inspection and Research Institute, Putian Fujian

Email: liniss@126.com, 1024154695@qq.com

Received: Feb. 16th, 2019; accepted: Feb. 28th, 2019; published: Mar. 7th, 2019

Abstract

The basic structure of elevator early warning system platform is introduced, and the relationship and specific role of each component of the early warning system are analyzed in detail. The function modules of early warning system are analyzed and introduced, the principle of analyzing and filtering data process using Hadoop and Pig technology is described particularly, and some key parameters involved are also analyzed. Through the elevator early warning system, it can provide pre-supervision measures for the safe use of elevators by further using large data of elevators, and the information level of elevator management has been improved.

Keywords

Elevator Early Warning System, Hadoop, Pig, Neural Network

一种基于Hadoop和Pig的电梯预警系统

林建兵¹, 许忠福²

¹莆田学院信息工程学院计算机系, 福建 莆田

²福建省特种设备检验研究院莆田分院, 福建 莆田

Email: liniss@126.com, 1024154695@qq.com

收稿日期: 2019年2月16日; 录用日期: 2019年2月28日; 发布日期: 2019年3月7日

摘要

介绍了电梯预警系统平台架构的基本组成, 详细分析了预警平台各组成部分之间的关系和具体作用。对预警系统的各个功能模块进行了详细的分析介绍, 并详细阐述了利用Hadoop和Pig技术进行电梯相关数

据的分析过滤原理和流程, 并对涉及到的一些关键参数也进行了分析。通过电梯预警系统, 为进一步运用电梯大数据进行电梯的安全使用提供事前监管措施, 提升了电梯管理的信息化水平。

关键词

电梯预警系统, Hadoop, Pig, 神经网络

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着经济社会的发展, 电梯的使用日益增加, 电梯安全是是民众出行安全的重要组成部分。近年来随着信息技术特别是大数据和云计算技术的广泛运用, 开展城市智慧电梯综合预警信息平台的建设技术也越来越成熟。应用最新信息技术监管电梯安全, 实现对其故障的提前预警和不间断监控成为一项重要的研究任务。通过对电梯故障数据的长期持续采集, 通过特定的数据分析技术发现电梯运行状况存在的一些异常特征, 并进一步通过科学的计算分析实现对电梯故障的预警, 使得相关单位提前采取措施, 规避可能出现的安全问题。

2. 电梯预警系统平台架构设计

电梯预警系统平台分为三层: 分别为数据采集层、数据处理分析层、数据输出层, 其平台架构如图 1 所示。

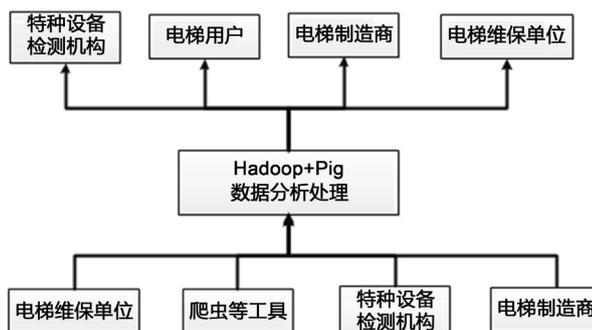


Figure 1. Architecture: elevator early warning system

图 1. 预警系统体系架构图

数据采集层包括电梯维保单位、各种特种设备检查检测研究院、电梯制造商等机构, 也包含利用爬虫等网络工具从因特网上采集得到的数据; 数据处理分析层结构比较复杂, 包含 Hadoop 分布式基础架构, 内含相应的工具以及分析平台 Pig, 用于提高经过优化处理的 MapReduce 运算[1]。Hadoop 是一种实现了一个分布式文件的架构系统, 简称 HDFS。HDFS 有高容错性的特点, 可以用来部署在低廉的硬件上, 而且它提供高吞吐量来访问应用程序的数据。Pig 是一种操作 Hadoop 的轻量级脚本语言, 以非常方便的处理 HDFS 和 HBase 的数据。数据分析层有包含相应的硬件基础设施, 这些设施包括服务器、存储及其虚拟化、资源池、云管理要素, 存储资源虚拟化设施, 基础设施管理平台配套设施, 以及为预警系统

信息服务平台服务的通讯设备, 为数据分析应用中心、数据加工中心以及数据共享和交换中心服务的配套硬件设施等; 数据输出层也叫数据输出单位, 主要包括电梯用户、电梯维保单位、各种特种设备检查检测研究院和电梯及其相关设备制造商。其中电梯维保单位、各种特种设备检查检测研究院和电梯及其相关设备制造商既是相关数据提供者同时也是分析处理国的数据受益者[2]。

3. 电梯预警系统功能设计

电梯预警系统功能划分为五大功能模块, 分别为: 数据采集模块、数据集成模块、数据预处理模块、数据过滤分析模块和数据输出模块, 如图 2 所示。

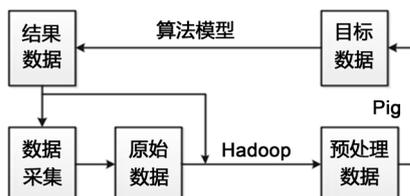


Figure 2. Function module of early warning system

图 2. 预警数据功能模块图

数据采集模块提供两种接口, 一种是针对电梯设备相关机构的接口, 另外一个接口是爬虫等网络工具的接口。目前阶段, 由于各个电梯机构提供的数据没有统一的标准, 因此电梯设备相关数据的接口的数据格式规定比较灵活, 能够满足各类不同的格式多样的数据输入。利用爬虫工具从因特网上搜索到的数据也是形式多样, 没有统一的规范, 因此该接口的格式要求也比较宽泛, 唯一和前一接口不同的是该接口的数据录入项的数目是固定的[3]。数据集成模块的功能是对数据采集模块得到的数据进行初步的格式化和规范化, 对从采集模块的两个接口得到的数据进行合并, 并去掉重复的数据, 对一些明显的非相关数据进行初步的剔除。数据预处理模块主要是通过 Hadoop 对大量数据进行分布式的预处理, 这种处理是以比较可靠、高效和可伸缩的方式进行; 该模块通过 HDFS 进行的文件的分布式存储和文件的增删改等常规操作。数据过滤分析模块是对数据预处理模块的数据进行加工分析, 这一模块用到了 Pig 技术。Pig 是一个基于 Hadoop 平台的海量数据分析技术, 它提供类似 SQL 的查询语言, 该语言的编译器会把类 SQL 的数据分析请求转换为一系列经过优化处理的 MapReduce 运算, Pig 能够为复杂的海量数据并行计算提供了一个简单的操作和编程接口, 数据的分析和过滤通过此接口得以完成[4]。Pig 分析过滤完数据后, 形成目标数据, 然后根据机构用户和电梯用户的日常使用情况采取特点的算法形成结果数据, 准备输出给用户。最后一个模块为数据输出模块, 数据输出模块分为两个部分, 分别针对机构用户和普通电梯用户, 机构用户是指电梯维保单位、各种特种设备检查检测研究院和电梯及其相关设备制造商[5]。输出给机构用户的数据包括一些专业的和各种电梯设备参数相关的数据, 输出给普通电梯用户的数据主要是日常电梯的一些基本数据和预警数据。

4. 数据分析过滤过程

数据分析过滤环节是电梯预警系统中最为关键的一环, 预警功能是否能够准确达到预定的标准和数据分析与过滤环节的关键参数设计紧密相关。在此过程中, 涉及两个具体操作流程: 一是将预处理的数据通过 Pig 技术生成相应的目标数据, 另外是将目标数据通过特点的算法模型进行相应的计算生成结果数据, 再通过数据输出模块将结果数据呈现给最终用户。Pig 分析数据时, 通过提取电梯及其相关设备的关键字如“设备代码”、“器件品牌”、“首次投用日期”、“器件故障开始时间年数”和“故障持续时间”、“器件预期寿命”以及一些可以查询到的“平均故障率”等, 采取类似 SQL 语言的分析方法进

行相应的分析查询操作, 并将分析的结果存入目标数据集中, 在通过 HDFS 进行存放。对目标数据采取的算法模型主要有深度前馈神经网络模型(DFN)和霍普菲尔网络模型(HN)。

利用深度前馈神经网络在训练目标数据时, 算法只是向上一层传递了少量的误差信息。考虑到计算时间的限制, 该算法在处理目标数据时堆叠的层次不多, 从而导致预警的精度有时会不够精确, 这是和时间因素综合权衡的结果。利用霍普菲尔网络对电梯及其相关设备的样本目标数据进行训练时, 每一批样本都作为 HN 的输入样本, 在训练过程中被隐藏起来, 训练过后恢复原来的状态并作为结果数据并将被输出。

5. 结论与说明

电梯预警系统涉及到电梯系统的各个方面, 各个相关机构的数据既互相关联又有所区别。经常出现同一个参数在不同的部门有不同的描述的情况。此外, 一些电梯设备产商对一些敏感的数据不愿意公开, 这些对预警系统的精确运行带来不利的影响。因此未来电梯预警系统涉及的电梯运行参数和环境参数、故障类型、频率、故障程度及运行寿命等有价值的数据的大量准确获取, 可以为电梯预警系统的进一步完善提供更加科学、完整的数据分析方法, 也为下一步利用大数据技术进行的电梯故障风险预警系统深度发展奠定坚实的基石。

基金项目

本文受莆田市科技项目(2018GP2004)资助。

参考文献

- [1] 徐斌, 李琳, 钟珞. 面向大数据的智慧电梯分析预警平台[J]. 武汉理工大学学报(交通科学与工程版), 2017, 12(2): 359-362.
- [2] 李树国, 张磊, 杨俊, 仇景泽. 电梯风险预警平台数据的噪声消除与分类挖掘[J]. 起重运输机械, 2017, 3(1): 21-23.
- [3] 王胜贤, 高天生. 基于 ZigBee 和 GPRS 的电梯远程监控系统的设计[J]. 测控技术, 2016, 25(3): 149-156.
- [4] 邓晓宇, 杨俊峰. 基于大数据应用的智慧电梯平台设计初探[J]. 信息系统工程, 2018, 11(4): 13-16.
- [5] 陈家焱, 洪涛, 刘钢海, 等. 物联网技术的电梯安全监控系统研究[J]. 中国计量学院学报, 2013, 24(3): 31-25.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>
期刊邮箱: csa@hanspub.org