

物联网、大数据分析和机器学习技术在灾备中的应用研究

刘伟佳¹, 李博权²

(1 农信银资金清算中心有限责任公司, 北京 100013; 2 中国科学院 国家空间科学中心, 北京 100094)

摘 要: 物联网、大数据分析、机器学习是智能时代的三驾马车, 对于科技的智能化发展具有重大意义. 本文先对灾备的发展趋势进行了分析, 提出“智慧+”灾备的思路, 在此基础上, 简要分析了物联网、大数据分析、机器学习等关键技术, 最后将这些技术应用到灾备系统设计中, 提出一种物联网环境下基于大数据分析和机器学习的“智慧+”灾备系统架构, 并对该架构的实现原理和功能进行了分析.

关键词: 灾备; 物联网; 大数据分析; 机器学习; 智能

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2018)12-0055-04

Application Research on Internet of Things, Big Data Analysis and Machine Learning Technology in Backup for Disaster Recovery

LIU Wei-jia¹, LI Bo-quan²

(1 Rural Credit Banks Funds Clearing Center, Beijing 100013, China;

2 National Space Science Center, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: Internet of things, big data analysis and machine learning are the troika of the intelligent age, which is of great significance to the intelligent development of science and technology. This paper first analyses the development trend of the backup for disaster recovery, then provides the thinking of "wisdom +" backup system for disaster recovery, on this basis, analys briefly the key technologies of the Internet of things, big data analysis and machine learning, finally uses these techniques to the design of backup system for disaster recovery, and provides a kind of "wisdom+" architecture of backup system for disaster recovery based on the big data analysis and machine learning in the environment of Internet of things, also analys the implementation principle and function of this system.

Key words: backup for disaster recovery; internet of things; big data analysis; machine learning; intelligence.

1 引言

随着科技的日新月异, 基于互联网的产品及业务需求与日俱增, 信息系统的超快节奏运行模式已开启. 这是一个呈现“随时、随地、随享”的“智慧+”业务时代, 又是一个体现“高速、高效、高质”的“智慧+”科技时代. 一旦发生因自然、人为等因素造成的系统运行中断, 后果不堪设想, 有可能造成数据丢失、业务长时间中断、服务体验差、客户失去信任而大量流失等各种负面影响, 甚至会终结业务或企业的发展. 所以, 建立与信息系统“匹配”的灾备系统是这个时代的不可或缺. 近年来, 灾备系统的建设和应

用普遍受到重视, 是企业加强业务连续性的重要表现. 但很少有报道提到灾备系统真正的实现接管生产业务的案例. 原因就在于, 当生产系统运行中断时, 要各种收集数据、分析判断、掌握情况, 才能做出各项决策, 数据是否全面将直接影响决策的方向. 如果这些环节全靠人为手工进行, 将大大占用决策的时间, 甚至错过接管业务的最佳时机. 所以, 在这个“智慧+”新时代, 灾备系统也要向“智慧+”灾备转型, 真正的将灾备作为业务连续性最后一道防线的作用凸显出来, “数据、预测、智能”三要素将是“智慧+”灾备的特点. 数据, 意味着能实时感知并“全、精、准”收集汇总灾备系统运转的各项数据, 可以引入物

联网技术,将灾备系统物物相连,点点互联,实现数据的智慧感知. 预测,意味着能分析判断,包括分析接管等等,可以引入大数据分析技术,对物联网感知数据进行汇总分析并挖掘信息价值. 智能,意味着学习能力,即能学习流程、解决方案并能分析新情况、推荐方案等,可以通过机器学习技术实现.

本文将物联网、大数据、机器学习技术和灾备系统相结合,分析设计一种物联网环境下基于大数据和机器学习技术的“智慧+”灾备系统,为灾备运维、管理和接管决策效率的提升提供解决思路,实现了业务系统的应急智能接管.

2 关键技术分析

2.1 物联网

物联网是近年来国内外学者及企业的研究热点,综合相关文献[1-3]中的介绍,其概念总结为:人与人、人与物、物与物间可通过 N 信息感知设备实现信息有效交互的一种智能网络. “一切皆感知”和“一切皆交互”是物联网的核心价值体现. 一切皆感知,就是通过射频识别、传感器、二维码等信息感知设备和信息采集技术,实现各类物体的信息随时随地采集;一切皆交互,就是通过物联网实现人与人、人与物、物与物间的数据传输、信息相互传递,即在物联网中物品是可以“讲话”表达信息的. 正因如此,物联网在各类信息系统的智能化发展方面起到关键的推动作用.

2.2 大数据分析

大数据分析技术近两年逐步得到商业化应用^[4-7]. 随着信息科技的快速发展,数据作为信息系统运转的产物,已呈现规模大、种类多、不规则、速度快等特点,如何从这些数据中挖掘出需要的信息,如何让数据能“讲有用的话”,为决策提供参考,是大数据分析技术的价值所在. 可用于大数据分析的技术包括数据挖掘、可视化、文本法分析、预测及统计分析等技术.

2.3 机器学习

机器学习是实现人工智能的核心技术,在智能领域应用范围较广^[8-10]. 机器学习是指计算机具备学习和判断能力,在没有人类明确指令下,其可以自行处理数据并输出有效结果. 机器学习涉及的理论包括概率统计理论、算法复杂性理论、近似理论,凸分析等. 机器学习在系统预测预警方面发挥很大作用,典型的机器学习方法包括主成分分析法、随机森林算法、朴素贝叶斯算法、支持向量机算法等,尤其

是近年来神经网络在人工智能算法领域成效显著,其预测分析能力大大推动了大数据等技术的应用.

3 物联网、大数据分析和机器学习技术在灾备中的应用研究

预测分析是实现一切智能体系的前提. 智能化和自动化最大的区别就是能进行信息反馈. 通过对数据的全面分析预测,生成“智能情景”(如图 1 所示),提供给系统管理及维护人员,能帮助快速决策、科学指挥和高效维护.

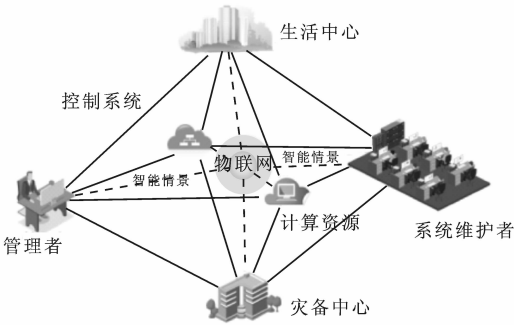


图 1 基于物联网环境的灾备场景

这里,智能是指把实时感知、识别的数据,通过大数据分析平台进行信息挖掘,经机器学习平台进行知识学习,实现智能知识库的建立,达到灾备系统监测的信息反馈和机制调整的目的,从而可以实现对系统接管时机、用时、事件处理和影响度等方面的准确预测. 所以,本文将物联网、大数据分析和机器学习技术与灾备技术相结合,设计出一种物联网环境下基于大数据分析和机器学习的“智慧+”灾备系统架构,主要由数据整理、大数据分析、机器学习和智能知识库等主要平台组成,如图 2 所示.

3.1 实现原理分析

(1)数据整理

数据整理平台可以为大数据分析和机器学习提供数据素材,由数据采集、数据汇总、数据存储、目标设定和数据处理模块组成. 物联网环境下,灾备系统的各项资源,包括机房环境、安全防护、计算资源、存储资源、网络传输、控制系统、同步系统、监控系统、巡检系统、维护系统等,都可以通过射频识别、传感器、二维码等感知技术和智能装置,随时随地对这些资源高速运转的各项数据进行感知、识别. 数据采集模块负责和这些感知设备对接,接收数据,并将数据转到数据汇总模块,数据汇总模块将数据分类转存到数据存储模块. 数据存储模块需要配备关系型和非关系型两种数据库,以存储结构化和非结构化数

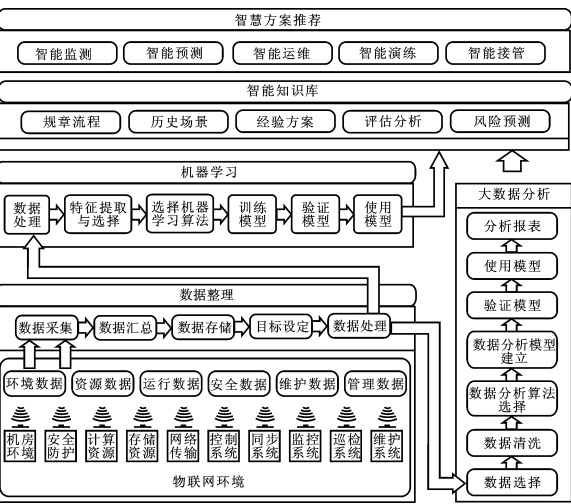


图2 物联网环境下基于大数据分析和机器学习的“智慧+”灾备系统架构

据。数据存储模块将数据提供到数据处理模块进行加工,这里加工是指对数据进行筛选、整合等处理。目标设定模块负责每次数据分析和机器学习的任务设定,如要得到存储空间趋势分析、接管决策分析、学习运维解决方法等。数据处理模块需要同时对接大数据分析平台和机器学习平台,因为大数据分析和机器学习各自采用不同的算法和处理过程,所以数据处理模块需针对不同算法的要求和需求,并按照本次任务目标导向来进行数据初步加工。

(2)大数据分析

大数据分析平台由数据选择、数据清洗、数据分析算法选择、数据分析模型建立、验证模型、使用模型和分析报表等模块组成。数据选择模块对数据整理模块的数据按照本次任务目标进行选择提取,精选数据、目标明确有利于提高数据分析效率。数据清洗模块是对已选定的数据格式、信息等进行处理。然后结合本次任务选取合适匹配的数据分析算法并建立数据分析模型,再通过测试数据验证模型的准确性和可靠性,直至模型符合要求,最终针对本次任务形成对决策与服务定位有精确导向的数据分析报表。

(3)机器学习

机器学习平台由数据处理、特征提取与选择、机器学习算法选择、训练模型、使用模型和验证模型等模块组成。这里的数据处理模块是指按照机器学习所使用的算法进行数据处理。从经处理的数据中提取并选择选取有用的数据特征,形成样本数据集。针对不同的样本数据集选择恰当的机器学习算法,经训练模型利用收集样本数据集让计算机学习如何对

系统情景进行分类的过程。验证模型模块抽取部分训练数据作为验证数据,评判不同参数取值下测试模型执行和迭代的程度,最终选择测试结果好的、符合实际需求的模型使用。

(4)智能知识库

智能知识库包含规章制度、历史场景、经验方案、评估分析、风险预测等模块,实时更新灾备系统管理运维的知识储备。规章制度模块是指存储更新来自行业监管、单位内部关于管理及运维的制度流程,当出现运维事件或变更维护等情况时,通过大数据分析及机器学习,和智能知识库“规章制度”模块自动匹配对接,分析当前情况、流程阶段和处理措施推荐。“历史场景”模块是指以往解决的运维事件、变更维护、应急处置、演练、接管等实际发生的场景。“经验方案”模块存储历史场景的实际解决方案。“评估分析”模块是指系统当前现状和“规章制度”、“历史场景”、“经验方案”模块之间的评估分析,以运维事件为例,发生时通过大数据分析和机器学习平台,掌握当前情况,先和智能知识库“历史场景”模块握手匹配,如果匹配成功,则接着和“经验方案”模块匹配,按此经验方案推荐;如不匹配,则再和“规章制度”模块握手匹配,分析当前情况处于流程的哪个阶段,需要完成的任务处置事项和操作。“风险预测”模块是指通过大数据分析和机器学习得到的系统现状和已有的风险点场景匹配,预测现有及可能出现的风险趋势。

3.2 系统功能分析

(1)智能监测,“智慧+”预测:数据的价值挖掘

通过全面地获取灾备系统各项数据,并对数据进行分析和学习,可以实现智能监测,实现灾备系统各项监测指标信息反馈和机制调整。监测到系统运行指标有异常,如存储空间不足、设备告警等情况时,可以分析对系统运行状况的影响,判断是否需要重点监测并就这些指标增加巡检频率,分析按照目前数据产生情况剩余存储空间可以支撑多久,告警如不及时处理可能会产生的后果等,形成分析导向图给系统维护人员。

(2)智能分析,“智慧+”运维:运维的智能助手

通过大数据分析及学习平台,把运维事件状态描述、内容描述、处理过程及结果,整个流水线的数据都进行整合分析,提炼、记忆、学习运维事件的处理方法和相关流程,对每个环节的操作进行智能分析评估,将学习结果反馈给智能知识库。做到出现运维事件时,系统可以从智能知识库中调出最优最准

确的运维事件解决方案,推送解决方案参考建议给运维及管理人员,全面提高运维效率.之后系统再智能分析学习本次运维事件解决方法和过程,补充更新智能知识库.

(3)智能场景,“智慧+”演练:演练的智能辅导

很多演练都头疼场景的设置,无论怎样穷举,都罗列不完.本文设计的灾备系统可以智能推送演练策略和场景,实现“智能场景库”,实现从“人找场景”到“场景找人”的演练模式的转变,实现从“人人参演”到“智能陪练”的演练方式的转变.物联网感知灾备的一切元素,整合、联合分析所有数据,精准查找出薄弱环节,优先针对这些环节涉及的角色、人物、资源环境,综合评估推荐一套智能演练场景.

(4)智能决策,“智慧+”接管:接管的智能指挥

灾备实现接管时需要对接管各项环境资源、时机、影响的准确判断,难度非常大.在短时间内快速掌握大量数据难.还需要对大量数据进行分析,难度更高.这也是大多机构在生产系统有突发情况时不愿意启动灾备系统的原因.本文设计的灾备系统可以实时掌握灾备系统的全部数据,通过大数据分析和机器学习,可以为决策层快速提供科学参考依据.

4 结束语

本文将物联网、大数据分析、机器学习技术应用到灾备系统中,设计一种物联网环境下基于大数据分析和机器学习的“智慧+”灾备系统架构,提出灾备系统的智能监测、智能预测、智能运维、智能演练、智能接管等实现思路,为灾备运维、管理和接管决策效率的提升提供解决方案.

参考文献:

- [1] Yan B, Huang G. Application of RFID and internet of things in monitoring and anti-counterfeiting for products [J]. International Seminar on Business & Information Management, 2009, 1 (1):392-395.
- [2] Yan T, Limin L. Application of internet of things technology in enterprise production safety [C]//Computer Technology & Development, CDAIT Board

meeting. Atlanta, Georgia. 2012.

- [3] Alam S, MMR Chowdhury, Noll J. SenaaS: an event-driven sensor virtualization approach for Internet of Things cloud [C]//IEEE International Conference on Networked Embedded Systems for Enterprise Applications. Suzhou, China, 2010: 1-6, DOI: 10.1109/NESEA.2010.5678060.
- [4] Alyass A, Turcotte M, Meyre D. From big data analysis to personalized medicine for all: challenges and opportunities [J]. BMC Medical Genomics, 2015, 8 (1):1-12.
- [5] Ahuja S P, Moore B. State of big data analysis in the cloud [J]. Network & Communication Technologies, 2013, 2 (1):62-68.
- [6] Chen S, BaiducomInc. Big data applications and practices of baidu [J]. Big Data Research, 2015 (11): 104-114.
- [7] Liu X, CR Center. The strategy of big data for alibaba group and its practices of credit reference [J]. Credit Reference, 2014, 32(10):10-14.
- [8] Rosten E, Drummond T. Machine learning for high-speed corner detection [C]// European Conference on Computer Vision, Austria, Graz, 2006, 3951: 430-443.
- [9] iang H, Zhang Y. Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine [C]// Power & Energy Society General Meeting. IEEE, Boston, USA, 2016.
- [10] Lemley J, Bazrafkan S, Corcoran P. Deep learning for consumer devices and services: pushing the limits for machine learning, artificial intelligence, and computer vision [J]. IEEE Consumer Electronics Magazine, 2017, 6 (2):48-56.

作者简介:

刘伟佳 女,(1986-),硕士.研究方向为云灾备技术.

E-mail:lwjaa@126.com.

李博权 男,(1987-),硕士,助理研究员.研究方向为卫星运控技术.