基于大数据的电网信息运维主动监控预警系统

基于减少电力信息网络运维故障的需求，设计实现了一种基于大数据的电网信息运维主动监控预警系统，系统包含采集监控、预警分析和决策辅助三个应用模块。提出了包含状态预警、阈值预警、快变预警、趋势预警、评价预警和关联预警六种方法的电网信息运维主动预警方法体系，提供了对运维故障进行提前预警的理论基础。系统基于大数据技术，通过对基础设施数据、信息系统运维数据和外部数据的高效分析，实现了对电力信息运维态势的全面感知和对运维风险的实时预警，智能辅助用户决策。

陈龙[1] 万明[1] 李世锋[2]/1. 国家电网公司南瑞集团 2. 国家电网公司华东分部

陈龙/工程师

关键词/Keywords
大数据·
预警·
辅助决策·

0 引言

电力信息网络是电网稳定运行的重要基础[1]，近年来，随着国家电网公司信息化建设的快速发展，公司信息系统数量不断增加，对日常运维提出了更高的标准和需求。目前主要采取在故障发生之后进行告警和抢修的被动运维模式，这种模式导致运维人员将日常大部分时间和精力都花在处理简单且重复的“被动救火”问题上，不但事倍功半而且常会出现恶性连锁反应。公司缺乏在故障发生之前对信息运维体系进行提前预警的能力，缺乏对运维隐患进行定位和分析的能力，亟需实现以预防为主的主动运维模式。

此外，随着公司信息运维业务的发展，大量数据不断生成和积累[2-3]：从数据类型上看，除传统的结构化数据外，还产生了大量日志、工单等半结构化数据和视频等非结构化数据；从数据处理速度上看，传统计算框架大体满足业务需要，但随着信息运维数据的不断增加，数据计算性能问题逐渐显现，需要基于大数据技术扩展计算能力。

通过对试点单位信息运维故障的统计情况分析发现，内存泄漏问题、归档日志问题以及数据库连接问题等绝大部分信息运维故障可在发生前获取相关信息，通过风险评估可以对信息运维体系进行预警，阻止故障发生。

1 系统总体设计

电网信息运维主动监控预警系统由数据采集层、集成计算层、数据分析层和预警应用层组成，通过与信息通信一体化调度运行支撑平台（SG－I6000）的集成，实现了对电网信息资源的统一管理；对海量电网信息运维数据的高效分析挖掘；在故障发生之前的故障定位、故障分析、故障预警和故障解决，从而全面提升电网信息运维整体水平。系统总体架构如图1所示。

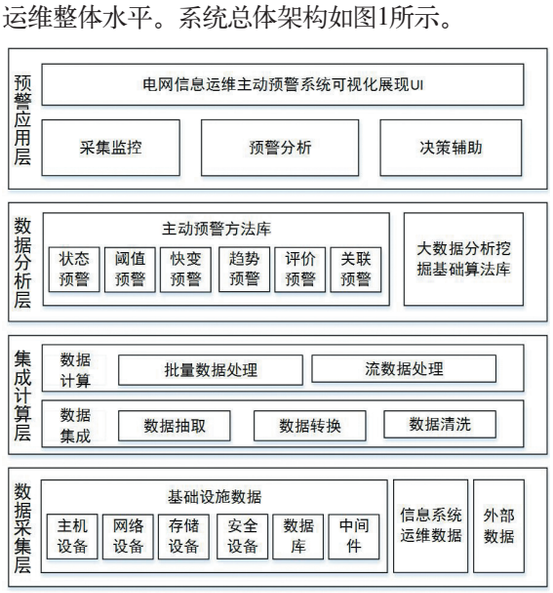

图1 电网信息运维主动监控预警系统总体架构图

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1）数据采集层。基于SNMP等数据采集协议实现对电网信息资源的运行和应用数据进行采集。系统采集的信息资源主要包括：基础设施数据、信息系统运维数据和外部数据。

2）集成计算层。实现对多源异构数据的汇聚、存储和清洗，并提供批量数据和实时流数据的计算服务。系统采用HDFS＋Oracle存储框架和Hadoop＋Spark streaming计算框架。

3）数据分析层。通过大数据分析挖掘基础算法库和主动预警方法库，提供数据分析支持。大数据分析挖掘基础算法库包含回归分析、聚类分析、分类及关联规则挖掘等基础算法；主动预警方法库提供六种主动预警方法。

4）预警应用层。基于底部支撑，实现上层应用，并提供前台交互服务。系统共包含采集监控、预警分析和决策辅助三个功能模块。

2　预警方法设计

实现对电力信息网络的主动预警，对电网信息系统的稳定运行具有积极意义。近年来，王阳光、丁真真和姚继军等专家学者[4-6]对预警理论进行了深入研究，提出了相关领域的预警方法，且具有指导意义。

根据国家电网公司信息领域运维现状，设计了包含状态预警、阈值预警、快变预警、趋势预警、评价预警和关联预警六种方法的电网信息运维主动预警方法体系，如图2所示。

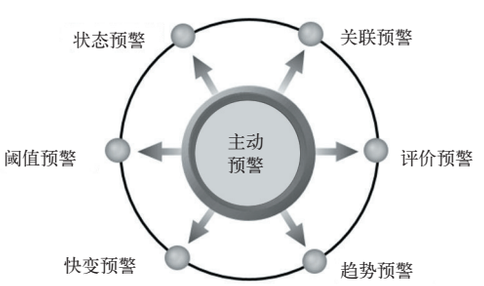

图2　电网信息运维主动预警方法体系

2.1 状态预警

状态预警是通过硬件或者服务是否可达来判断信息资源的状态，通过对其状态的监控实现预警。

实现状态预警，首先要判断信息资源是否可达，目前系统每5 min探测一次，确定信息资源是否响应。如果接收到信息资源的应答数据，则认为其可达，否则，如果等待应答超时，则判定其不可达。

在确定信息资源是否可达的前提下，将其状态分为三种：正常状态、失联状态和不稳定状态。正常状态表示信息资源每次都可达，失联状态表示信息资源连续两次不可达，不稳定状态表示信息资源经常不可达，但是达不到失联标准。目前的不稳定状态判断原则是：一个工作日探测的288次中至少有3次不可达，且不连续，则认为其处于不稳定状态。如果信息资源处于不稳定状态，需要对其进行状态预警。

2.2 阈值预警

通过对信息资源设置风险预警阈值，将当前监测数据与预警阈值进行比较，如果监测数据不在相应的预警阈值范围内，则认为监测对象符合预警条件，生成预警事件。

传统的阈值是靠运维人员手动设置的固定值，极易造成阈值与真实值不符。阈值预警方法充分利用大数据技术，通过对历史数据的分析自适应设置动态阈值。同时，系统提供接口供用户对阈值进行手动修改，弥补系统的不足。

2.3 快变预警

信息资源监测数据需要与同类数据进行比较，若变化过于剧烈，差值大于一定比例，则认为被监测资源可能处于较大变动中，需要生成预警事件。快变预警包括横向预警和纵向预警两种方式。

横向预警表示与同类资源相比较形成的预警，主要针对负载均衡集群，如果集群中某台设备的负载远远大于其他设备，则认为需要对其进行风险预警。

纵向预警示与自身历史数据进行比较形成的预警，主要是基于大数据统计分析技术，对历史数据进行挖掘分析，并将分析结果应用到信息运维主动预警模型。

2.4 趋势预警

趋势预警通过对信息资源监测数据的趋势分析来判断资源是否会达到预警触发条件。趋势预警通过预警触发阈值、指标与告警阈值的差值、指标增长率这三个指标来实现短期的信息资源的

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

风险预警。趋势预警算法

$$\frac{\alpha(V-V_0)}{t(T_w-N)}+\beta\geqslant S \quad (1)$$

式中，当前点的值为V，上一个点的值为V_0，两个点之间的时间差为t；α、β为修正值；S为短期趋势预警阈值，其中，α，β和S均可由用户配置。当指标值低于预警触发阈值时，不会进行预警；指标值达到甚至超过预警触发阈值时，通过短期趋势预警算法，可以实现信息资源的短期风险预警。

2.5 评价预警

评价预警通过打分来对信息资源进行评价，通过分值来判断信息资源的健康度。评价算法

$$S=S_0-\left(\sum_{i=1}^{m}\alpha_i f_1^i+\sum_{j=1}^{n}\alpha_j f_2^j+\sum_{k=1}^{p}\alpha_k f_3^k\right) \quad (2)$$

式中，S_0为可由用户自定义设置的初始得分；m、n、p分别为监测周期内的告警事件级别数量、预警事件级别数量、检修事件级别数量；α和f分别为各级事件的权重和次数；S为最终评价得分，可以由用户自定义设置阈值，实现评价预警。

2.6 关联预警

关联分析又称关联挖掘，就是在信息系统及其基础设施的海量运行和应用数据中，查找存在于项目集合或对象集合之间的频繁模式、关联、相关性或因果结构。关联充分利用大数据关联分析手段，对信息系统及其基础设施的各类指标进行关联分析，挖掘相关模式，从而实现信息资源的主动预警。

3 系统应用设计

电网信息运维主动监控预警系统包含以下三个功能模块：采集监控模块、预警分析模块和决策辅助模块。

3.1 采集监控模块

采集监控模块通过多采集协议（SNMP、SSH、Telnet、JMX、WMI、JDBC…）实现对内部信息资源的实时监控和数据采集，通过Web Service方式实现外部数据的集成，通过flume实现分布式海量日志的采集、聚合和传输，通过sqoop实现oracle和Hadoop分布式文件系统HDFS的数据传输，最后将所有数据集成到HDFS，实现全量数据的统一管理和应用。

采集监控模块的采集频率为5 min/次，它采集的内部信息资源数据包括两类：第一类是基础设施数据，包括主机设备、网络设备、存储设备、安全设备、数据库和中间件的实时和历史数据；第二类是信息系统运维数据，包括信息系统各个主要页面是否有响应以及响应时长等指标。

采集监控模块实现对各类信息资源的监控和采集，实现内外部数据的高效传输和汇聚，是上层应用的基础。

3.2 预警分析模块

预警分析模块通过对数据的分析，实现预警信息的生成、压缩和归并。预警分析模块分为主动预警引擎和压缩归并子模块。

主动预警引擎基于大数据技术，集成了状态预警、阈值预警、快变预警、趋势预警、评价预警和关联预警六种主动预警方法，通过对各类信息资源运行异常的监测和分析，实现了各类缺陷的预判和预警信息的生成。

压缩归并子模块通过对预警信息的压缩和归并，从而实现运维故障的根源定位。压缩指的是将多次发生的同一个预警事件合并为一个单独的事件，例如某台设备出现故障，它会持续的产生网络事件，此时可以通过压缩方法来剔除掉无用的网络事件，只需要保留一条预警事件；归并指的是将多个局部预警事件综合为一个整体的预警事件，例如信息网络某个关键节点出现故障，则与该节点相连的所有设备均无法正常运行，此时可以通过归并方法来剔除掉无用的大量预警事件，只需要保留关键节点设备的预警事件。

3.3 决策辅助模块

决策辅助模块基于预警分析模块的输出结果，实现预警事件的全面分析挖掘，辅助用户决策。决策辅助模块主要有以下功能：

1）预警信息分级。将预警事件分为指标级、基础设施级和信息系统级。其中最基础的预警事件均为指标级；如果某台基础设施有多个指标出现预警，则将相关预警事件归并为一个基础设施级预警事件；如果有信息系统核心节点或多个普通节点出现预警，则将相关预警事件归并为一个信息系统级预警事件。预警事件的分级可以辅助用户对预警事件重要性进行研判和分析。

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2）信息系统评价。从基础架构和可靠性两个方面对信息系统进行全面评价。基础架构评价指的是对信息系统所依托的软件基础设施进行评价，并根据各个节点的评价结果及各个节点在信息系统网络拓扑架构中的重要性得到信息系统基础架构运行状况的评价；信息系统可靠性评价指的是根据其主要页面的实时探测指标，包括页面是否有响应以及响应时间，从而对信息系统本身的运行状况进行评价。评价结果实现了对各个信息系统的全面分析，辅助用户决策。

3）信息系统排名。根据评价预警模块结果，实现对各个信息系统的评分和排名，用户根据分析结果对各个信息系统一段时间内的运行状况进行全方面了解，从而有效实现下一步运维决策。

本文基于国家电网公司信息运维现状，基于大数据技术，研究实现了一种电网信息运维主动监控预警方案。系统建立了包含状态预警、阈值预警、快变预警、趋势预警、评价预警和关联预警六种方法的电网信息运维主动预警方法体系，提供对运维故障提前预警的依据。系统采用模块化设计，实现了对基础设施数据、信息系统运维数据和多类型外部数据的高效分析挖掘，实现了运维态势的全面感知、运维风险的实时预警及预警事件的及时处置。

参考文献

[1] 郭创新，陆海波，俞斌，等. 电力二次系统安全风险评估研究综述[J]. 电网技术，2013(1): 112-118.

[2] 张东霞，苗新，刘丽平，等. 智能电网大数据技术发展研究[J]. 中国电机工程学报，2015(1): 2-11.

[3] 邓炜瑛. 智能电网大数据处理技术现状与挑战[J]. 中外企业家，2015.

[4] 王阳光，尹项根，游大海，等. 基于无线传感器网络的电力设施冰灾实时监测与预警系统[J]. 电网技术，2009(7): 14-19.

[5] 丁真真. 基于BP神经网络的快速消费品上市公司财务风险预警系统研究[D]. 哈尔滨：东北林业大学，2012.

[6] 姚继军. 基于数据挖掘的安全风险预警系统技术研究[D]. 北京：中国地质大学，2010.

[7] 郭金根，刘晔. 基于IEC 61970的智能电网数据采集平台设计[J]. 浙江电力，2010，29(8): 41-44.

[8] 颜雪松，蔡之华. 一种基于Apriori的高效关联规则挖掘算法的研究[J]. 计算机工程与应用，2004，38(10): 209-211.

[9] 冯永青，孙宏斌，朱成骐，等. 基于信息理论与技术的地区电网辅助决策系统设计[J]. 电力系统自动化，2004，28(4): 58-62.

[10] 纪佩宇，印杰，武旭红. 面向物联网的大数据预警信息高速处理规则引擎[J]. 计算机与数字工程，2013(9): 193-193.

作者简介

陈龙（1986-），男，博士，工程师，主要研究方向为大数据、云计算和电力信息化。

万明（1982-），男，硕士，工程师，主要研究方向为自动化运维。

李世锋（1977-），男，本科，高级工程师，主要研究方向为信息系统运维管理。

（收稿日期：2015－12－10）EA

（编辑：崔文静）

(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net