DOI:10.3969/j.issn.1009-6868.2017.04.001
网络出版地址:http://kns.cnki.net/kcms/detail/34.1228.TN.20170706.1044.004.html

机器学习在大视频运维中的应用

Applications of Machine Learning in Big Video O&M

屠要峰/TU Yaofeng
吉锋/JI Feng
文韬/WEN Tao
(中兴通讯股份有限公司,江苏 南京 210012)
(ZTE Corporation, Nanjing 210012, China)

中图分类号:TN929.5 文献标志码:A 文章编号:1009-6868 (2017) 04-0002-007

摘要: 通过对中兴通讯大视频运维系统整体架构和关键模块的介绍,以及机器学习技术在大视频运维系统中端到端异常检测、根因分析与故障预测等场景的具体应用的分析,并结合硬盘故障预测的实例,认为随着人工智能在运维领域的应用发展,从基于规则的自动化运维转向基于机器学习的智能运维必然成为趋势。中兴通讯适时采用了机器学习方法来提取历史巡检数据中蕴含的故障特征,并构建集成预测模型来提升大视频运维的精度和效率,目前取得了较好的效果。

关键词: 大视频;大数据;机器学习;人工智能

Abstract: In this paper, the overall architecture and key modules of ZTE big video operation and maintenance (O&M) system are introduced. Then the application scenes of machine learning technology in this system, including the end-to-end anomaly detection, root cause analysis and fault prediction are analyzed, as well as the instances of hard disk breakdown prediction. With the development of artificial intelligence technologies in the field of O&M, the trend which is from rule-based automatic to machine learning-based intelligent will be formed. To improve the precision and efficiency of the big video O&M, ZTE has adopted a machine learning method to extract the fault features contained in the historical inspection data, and has built the integrated prediction model, which has achieved good performances.

Key words: big video; big data; machine learning; artificial intelligence

随着移动互联网和宽带网络的快速发展,视频业务以广泛的受众、高频次的使用、较高的付费意愿,已经具备成为“杀手应用”的潜质。越来越多的电信运营商将视频业务视为发展的新机遇,并作为与宽带、语音并列的基础业务。据Conviva用户视频报告的数据,35%的用户把视频观看体验作为选择视频服务的首要条件[1]。因此,运维保障成为视频业务的关键。

当前视频业务发展已进入“大内容”“大网络”“大数据”“大生态”的大视频时代。业务形态多样,包括交互式网络电视(IPTV)、基于互联网应用服务(OTT)的TV、移动视频等;组网复杂,视频在多屏之间的无缝衔接、码率格式适配等需求对网络提出了更高的要求;数据多样性大大增加,需要从视频码流、终端播放器、内容分发网络(CDN)、业务平台、网络设备等各个环节获取数据,既有结构化数据,又有半结构化、非结构化数据;数据实时性要求大大提高,传统网管采集数据的粒度是5 min,而大视频业务要求秒级的数据采集和分析,数据量和计算量增加了百倍。

这些都对传统的运维模式和技术方案带来很大的挑战。如何在大视频背景下客观评价和度量终端用户的体验质量,如何界定视频业务系统故障和网络故障,如何快速诊断网络中的故障并提前发现网络隐患,如何发掘视频业务运营和利润的增长点,成为各大运营商对大视频业务运维的关注重点。

1 大视频智能运维系统的架构及关键技术

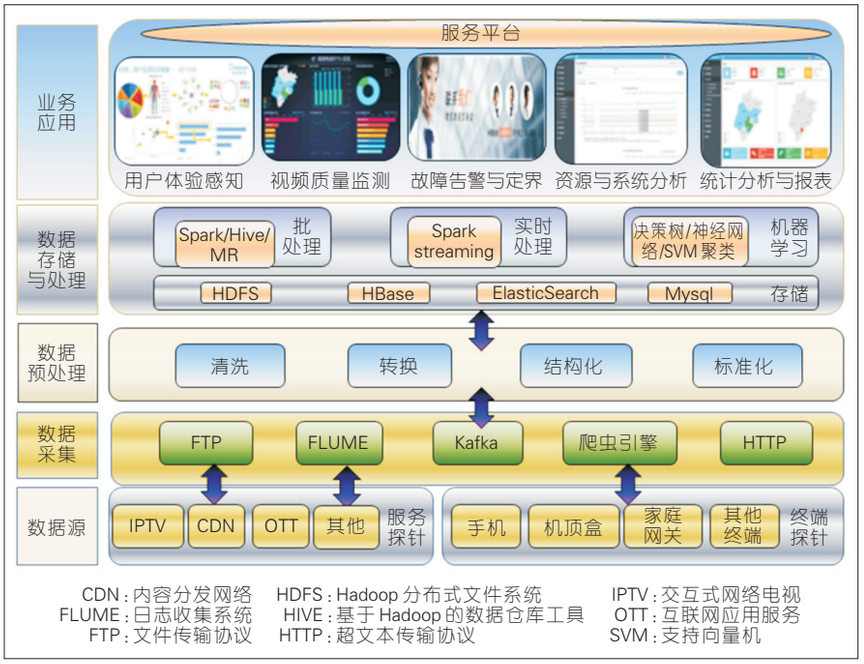
在原有运维技术手段基础上,进一步依托大数据及人工智能技术,对大视频业务系统产生的各类信息进行汇聚、分析、统计、预测等,中兴通讯形成了智能化的大视频运维系统,其系统架构如图1所示。

大视频运维系统有以下几个部分组成:

(1)数据源。数据源主要指大视频业务智能运维所需要采集的数据,包括终端的播放记录、关键绩效指标(KPI)数据;接入网络的用户宽带信息、资源拓扑数据;CDN的错误日志、告警、链路状态、码流信息等;IPTV业务账户、频道/节目信息等。

(2)数据采集及预处理。数据采集层主要是Kafka、文件传输协议(FTP)、超文本传输协议(HTTP)等用

收稿日期:2017-05-28
网络出版日期:2017-07-06



▲图 1 大视频运维系统架构

于数据采集的组件；数据预处理是指对各种异构日志数据进行解析、转换、清洗、规约等操作，主要完成数据使用前的必要处理及数据质量保证。

（3）数据分析处理。数据分析处理主要包括流式计算处理框架 Spark、离线批处理 MR 框架、人工智能计算框架、数据存储及检索引擎等。业务组件包括批处理、数据实时分析、机器学习等模块。批处理模块主要是对时效性要求不高的业务模块的处理及数据的离线分析，包含但不限于故障及异常的根源分析、故障及特定规则阈值的动态预测、事件的依赖分析及关联分析、异常及重要时序模式发现、多事件的自动分类等；数据实时处理主要是对于时效性要求较高的安全事件进行监测控制、异常检测与定位、可能引发严重故障的预警、对已知问题的实时智能决策等；机器学习模块包括离线的机器学习训练平台、算法框架和模型。

（4）业务应用层。业务应用层主要提供智能业务监测控制、端到端故障定界定位、用户体验感知、统计分析与报表等主要业务场景的分析及应用。

大视频运维系统涉及的关键技术包括：

（1）大数据技术。该技术可以构建基于大数据的处理平台，实现数据的采集、汇聚、建模、分析与呈现。

（2）探针技术。该技术可以实现全网探针部署，包括机顶盒探针、直播源探针、CDN 探针、无线探针、固网视频探针等，通过探针技术实现全面的视频质量实时监测控制以及数据采集。

（3）视频质量分析指标。该指标以用户体验为依据建立视频质量评估体系，对视频清晰度、流畅度、卡顿等多项用户体验质量（QoE）指标进行分析。

（4）人工智能技术。机器学习本身有很多成熟的算法和系统，以及大量的优秀的开源工具。如果成功地将机器学习应用到运维之中，还需要 3 个方面的支持：数据、标注的数据和应用[2]。大视频系统本身具有海量的日志，包括从终端、网络、业务系统多方面的数据，在大数据系统中做优化存储；标注的数据是指日常运维工作会产生标注的数据，比如定位一次现网事件后，运维工程师会记录下过程，这个过程会反馈到系统之中，反过来提升运维水平；应用指运维工程师是智能运维系统的用户，用户使用过程发现的问题可以对智能系统的优化起正向反馈作用。

2 人工智能技术在大视频运维系统中的应用

2.1 基于人工智能的端到端智能运维

传统电信网络、业务系统的运维模式通常是在故障发生后，运维、开发人员被动地进行人工故障的定位与修复。技术专家通过分析系统日志，依据事先制订的系统运行保障规则、策略和依赖模型，判断故障发生的原因并进行修复。这一过程不仅工作量巨大，操作繁琐，代价高昂，容易出错，且不能满足持续、快速变化的复杂系统环境需求。

大视频业务系统的故障定界定位尤其复杂且耗时耗力，原因在于：大视频系统中网元众多且业务流程复杂，如包括 IPTV 管理系统、电子节目菜单（EPG）、CDN、机顶盒、直播源编码器等众多网元，发现问题需要各个网元一起定位排查，对人员技能的要求很高。大视频系统对网络要求比较高，机顶盒经过光网络单元（ONU）、光线路终端（OLT）、宽带远程接入服务器（BRAS）、核心路由器（CR）等，从接入设备、承载设备到 CDN 服务器，中间任何一个网络设备出现丢包、抖动等问题都会导致用户的观看体验受影响，对这种卡顿分析是一个大难题。随着视频业务的快速发展和业务量不断增长，如何快速定位问题，降低运维门槛变得越来越迫切。

端到端智能运维系统就是利用大数据采集分析、人工智能与机器学习等技术提升系统运维智能化能力，从智能化的故障定位、智能化的根因分析机制入手，覆盖从被动式事后根

源追溯到主动式事中实时监测控制及事前提前预判的各种业务场景(如图2所示),提供从数据收集分析,故障预判到定位,再到故障自动修复的端到端保障能力。

面向历史的事后追溯主要有历史故障根因分析、系统瓶颈分析、业务热点分析等;面向实时的事中告警主要有异常监测、异常告警、事件关联关系挖掘、实时故障根因分析等;面向未来的事前预判主要有故障预测、容量预测、趋势预测、热点预测等。其中,事后追溯更多面向离线、非实时的运维故障分析,事中告警和事前预判更多面向实时或准实时的运维故障检测、分析及预测。

机器学习技术在端到端的智能运维系统中有几个应用点。

(1)日志预处理模块

预处理的核心问题是将半结构、非结构化的日志转换为结构化的事件对象。事件被定义为一种现实世界系统状态的体现,通常涉及到系统状态的改变。本质上,事件是时序的且经常以日志的方式进行存储,例如:业务事务日志、股票交易日志、传感器日志、计算系统日志、HTTP请求、数据库查询和网络流量数据等。捕获这些事件体现了随着时间变化的系统状态和系统行为以及它们之间的时序关系。事件对象可以简单定义为:Event={时间戳,事件类型,<属性1:属性值1,属性2:属性值2…>}。事件挖掘是一系列从历史事件和日志数据中自动、高效地获取有价值知识的技术,正确提取事件才能后续从时间、空间等多角度挖掘事件之间的关联、依赖等关系。将文本日志集转换成系统事件的典型技术方案包括:基于日志解析器、基于分类和基于聚类的方法。

最为直接的解决方案是采用日志解析器,该方法为每一种特定的系统日志实现对应的日志解析器,每种类型的日志采用正规表达式或预定义模板进行抽取。这种方式需要用户了解系统日志,一个日志解析器难以适配不同格式的多种系统日志,需要大量的人力来开发定制的日志解析器软件。从机器学习辅助人工完成日志解析的角度,可以采用分类或聚类的方式。

日志分类方法是一种直接从日志数据中识别事件类型的方法,它通过分类器模型将一条条日志消息划分成若干个预定义的事件类型,如图3所示。

一种简单的分类方法是为每一个事件类型预先定义一种对应的正则表达式模式(如前所述的解析器或称为过滤器);另一种更为通用的日志消息分类方法是基于机器学习分类模型的方法,即用户提供一些标记过的日志消息,每个消息的事件类型已被明确标注;然后,机器学习算法根据标记的数据建立一个分类模型,利用这个模型对新的日志消息进行分类。虽然这种方式带来一定的泛化性,其主要问题在于需要大量的带标记日志消息,需要一定数据积累与人力消耗。

另外一种基于聚类的方法,不需要大量人力且适用于多种系统日志,虽并非十分准确,但可以应用于能够容忍一些错误或噪音事件的事件挖掘应用中。日志消息聚类采用无监督方法将日志划分为各种类型事件,因为日志消息聚类并不要求准备一系列标记过的训练数据,所以这种方法更加实用。业界典型的日志聚类方式包括基于日志消息签名[3]和基于树状结构的聚类[4]等算法。

在实际应用场景中,需要根据业务的复杂度和数据积累情况,综合选择日志预处理解决方案:准确性要求非常高的场景需要使用专业日志解析器;而分类、聚类的机器学习方式更适合容忍一些错误或噪音事件的事件挖掘应用。

(2)日志离线分析模块

日志离线分析的核心问题是通过机器学习算法发现事件之间的关联、依赖关系。离线分析负责从历史日志数据中获得事件间关联性和依赖性知识并构建知识库。事件挖掘综合利用数据挖掘、机器学习、人工智能等相关技术去发现事件之间的隐藏模式、未来的趋势等关系。分析人员可以利用已发现的事件模式对未来事件的行为作预测,同时挖掘出的事件依赖关系也可以用于系统故障的诊断,帮助运维人员找出问题的根因,达到解决问题的目的。

事件的关联分析,本质上根据日志文件中的每个消息事件的时间戳,发现时序事件之间的关联性。我们

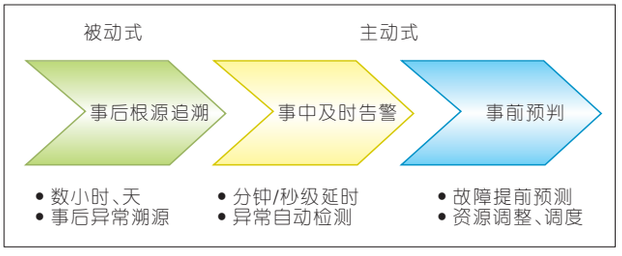
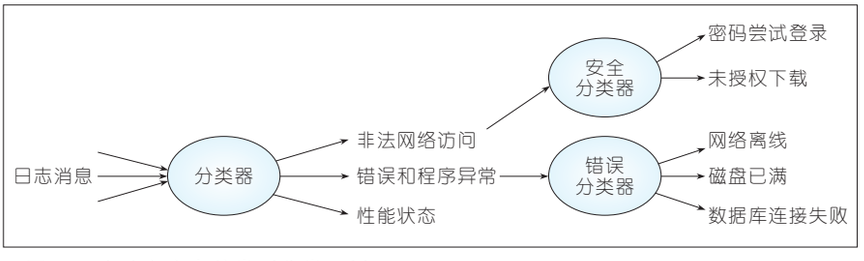


图2▶ 基于人工智能的智能运维典型应用场景



▲图3 日志消息中事件类型分类示例

重点挖掘两种类型的相关性：基于时序连续值数据的相关性和基于离散事件数据的相关性。以大视频系统为例：中央处理器（CPU）使用率、内存使用情况、磁盘读写数据量、网络接收或发送数据量都可以表示为时间序列的连续值；而应用程序服务器上请求和应答序列被视作事件数据，因为每个数据项的值都是属于某个类别的离散值，例如：CPU 使用率的时序图与磁盘读写数据量时序图（如横轴时间、纵轴数值的可视化表示）就会有很强的相关性；网络异常事件与应用程序服务器上请求和应答异常事件，通过离散事件数据分布图（横轴时间、纵轴事件类型的可视化展示）分析就存在一定相关性，如两类事件基本在同一时间点同时出现，具备一定的关联性。

事件之间存在关联性，不一定表明事件之间一定存在依赖或因果关系，比如事件 A 和 B 具备相关性，并不代表 A 引发 B 或 B 引发 A，因此需要基于关联关系基础上进一步挖掘相互依赖关系。所谓事件依赖分析，发现类似 A→B 的依赖关系，最终形成一个事件依赖的动态概率模型图。如图 4 所示，A、B、C 对应于大视频运维中不同事件，通过基于时间窗的事件依赖算法[5]挖掘出各种故障事件之间的依赖并形成相应的依赖图。

总之，离线分析主要是通过机器学习算法形成关联、依赖的规则或概率图模型，另外还包括利用历史时序故障数据进行传统机器学习特征工程建模或深度学习端到端的时序建模，为接下来的在线实时故障分析、定位与预测等提供支撑。

（3）实时分析模块

实时分析模块负责实时处理新产生的日志数据并根据离线分析获得的知识模型完成在线运维的管理操作。典型的实时分析技术主要有异常检测、故障根因分析、故障预判和问题决策等。

对于实时的异常检测，可选择的方案有两种：基于监督学习和基于无监督的方案。前者利用基于离线训练出来的检测模型进行判断，这种方式不如后者使用普遍。

故障预判更多的是基于历史的数据进行分析建模，如下文即将讲述的基于 Smart 硬盘故障预测示例。

问题决策则更为全面，对前述检测出的异常，预判即将要出现的故障以及定位已知故障的原因，进行高层的资源调度，或发出设备替换的决策指令，最终避免可能出现的故障，自动修复已知的故障（若可以修复）或者发出告警通知运维人员进行人工修复。

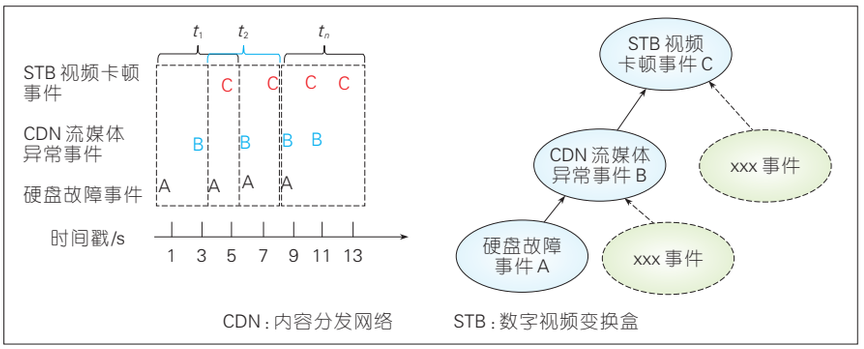
（4）智能故障定位及根源分析

故障智能定位是模拟人工排查故障的流程，对可疑的故障检查点进行逐一排查，通过采集各业务模块的告警、性能指标、错误和异常日志，组织生成故障定位的基础事件数据，针对故障现象配置对应的检查点及处理建议。

在故障定位时，从故障现象出发，通过中序遍历方式遍历整个故障树。前一个节点的出参是后一个节点的入参，检查点调用应用程序编程接口（API）检查本节点的故障原因是否存在，通过 API 来分别从各种网元获取对应现象的证据信息，直至分析到叶子节点。然后将所有满足条件的节点进行回归，根据权重返回现象的原因。遍历结束后综合各个节点的检查结果形成本次故障定位的诊断结论。

在用户报障时可能对故障产生的时间、触发的位置、观看的节目等信息记忆模糊不清。在故障定位过程中首先需要从用户的行为记录里筛选出故障记录，这个筛选的过程采用前述的日志聚类算法，对影响用户感知的 KPI 进行聚类，对聚类结果根据预定的规则或分类器判断出属于故障类的记录。如果有多条故障记录，任选一条故障记录进行定位。除通过算法筛选故障记录外，另提供人工辅助筛选功能提高准确性。

智能化的根因分析，主要根据前期分析出来的事件依赖概率图模型，建立基于历史故障定位及处理经验集的专家知识库，利用机器学习的理论与技术，在多维变量间因果关系做出权重的动态调整，调整各个检查点的权重。故障定位方法较之前的传统方法具有更精确的错误定位效果和更显著的定位效率。



▲图 4 基于时间窗的依赖关系挖掘（左）与依赖概率（右）

2.2 基于人工智能的硬盘故障预测实例

当前大视频运维过程中遇到的难题之一就是 CDN 故障硬盘的置换。为了规避软硬件风险，提升数据中心管理效率，制订合理的数据备份迁移计划，业界各大主流 IT 企业均展开针对硬盘故障预测的研究工作。研究者认为：在此预测技术的支撑下，可以极大地提升服务/存储系统的整体可用性。我们接下来将列举一个基于机器学习实现的 CDN 硬盘故障预判的实例。

当前，自我监测分析和报告技术（SMART）已经成为工业领域中硬盘驱动状态监测和故障预警技术的事实标准[6]。研究表明：硬盘的一些属性值如温度、读取错误率等，和硬盘是否发生故障有一定的关系。如果被检测的属性值超过预先设定的一个阈值，则会发出警报。然而，硬盘制造商估计，这种基于阈值的算法只能取得3% ~ 10%的故障预测准确率和低预警率[7]。学术界和工业界在采用机器学习方法提升SMART硬盘故障预测精度方面的工作由来已久，但受限于数据集规模，现有方法取得的预测模型效果不佳。近年来，随着越来越多厂商的关注，基于SMART巡检数据的硬盘故障预测研究有了很好的数据支撑，一方面体现在硬盘规模快速增长，另一方面体现在采样工作正规化。在以上高质量数据支撑下，基于SMART巡检数据的故障预测水平得到了显著提升。

我们在大视频运维中基于SMART数据进行硬盘故障预测，采用了基于旋转森林的集成预测模型方案，基本流程如图5所示。

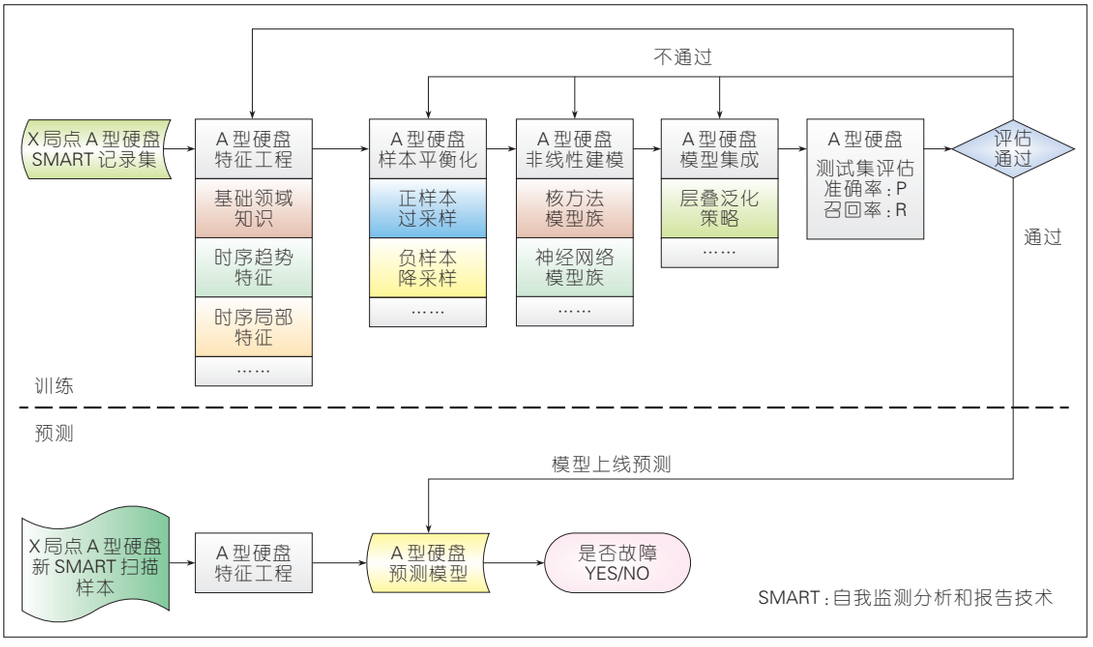
将SMART扫描数据集按照局点和硬盘型号进行细分，每个局点每个硬盘型号的数据分别建立预测模型，每个预测模型的构建过程为：

（1）特征工程。特征工程是决定预测效果的关键步骤。我们不但需要考虑观测点当时的SMART取值，也需要考虑该SMART取值的历史变化趋势、震荡幅度、跳变频率等因素，主要策略包括取高价值属性和衍生时序特征。取高价值属性，即采用“数据驱动和领域知识相结合”的策略，一方面和相关硬件专家交流，听取他们的领域指导意见；另一方面，从故障硬盘的历史SMART记录集出发，找出“故障硬盘和健康硬盘在该属性上统计性质存在不一致”的SMART属性。专家知识和数据驱动结果都作为特征工程结论的一部分，宁多勿少。衍生时序特征，即在找出具有提示性效果的高价值SMART属性后，对其时序特征做进一步衍生、调整。以上两种特征工程策略相互补充，共同组成了模型训练需要的特征空间。

（2）模型训练。模型选择和训练、优化是构造预测模型的直接步骤，由于基于SMART记录集做硬盘预测是一个高维分类问题，同时正负数据严重不平衡，采用线性分类模型往往没有很好结果，因此考虑采用构造非线性模型来解决问题，主要分两大步骤：重新平衡正负样本和非线性建模。重新平衡正负样本，即采用“过采样+降采样”结合的策略，对于负样本（健康硬盘），考虑采用聚类方法提取聚类质心，将质心附近的样本按比例提取作为该聚类的代表，从而实现降采样，而聚类算法和聚类质量评价准则需要根据实际数据分布来决定；对于正样本（故障硬盘），考虑采用过采样方法来提升正样本数据分布。以上降采样和过采样策略结合，把正负样本的比率从1:50重构到1:5以内，重构训练集。非线性建模，即利用旋转森林技术对以上训练集进行降维，并选择核方法、神经网络来构造分类超平面，择优选择其中有代表性的模型，然后再将这些模型利用层叠泛化技术组合形成最后的预测模型。

（3）模型评估。模型评估即将模型训练阶段生成的模型，在测试集上进行测试，重点关注预测准确率和故障覆盖率（召回率），直到选出符合要求的模型。

（4）模型上线。模型上线即将通过模型评估的最终模型部署到现网



◀图5
硬盘故障预测流程

环境后，每次SMART扫描得到的新样本均需要输入本模型，得到“未来一段时间该硬盘是否发生故障”的预测结果。

我们在Backblaze数据集（2016年Q1—Q2）[8]上选取了某型号希捷硬盘的SMART扫描记录做验证，其中时间跨度为6个月182天，数据粒度为每天扫描，涉及到11 890块硬盘（连续3天扫描找不到则视为故障盘，这种盘共242块），共2 118 925条扫描记录。按照7:3的大致比例划分训练集和测试集：测试集共3 560块硬盘（故障盘共83块），共634 950条扫描记录。

我们采用如表1所示的SMART基础属性来进行建模，共有12个基础属性。

考虑到SMART属性前后变化趋势也可能昭示着后续硬盘故障，因此我们在以上基础属性上衍生了时序率属性，包括每个基础采集之前9天内的相对变化率。

将以上基本SMART属性和衍生属性融合，作为SMART故障预测的特征参与模型构建。我们采用的子分类器如表2所示，共4类18个。

随着旋转森林特征子集分块参数的变化，生成的子分类器对故障的预测能力也在不断调整，最终生成不同分块参数对应的模型在同一测试集下不同的预测效果（如图6所示）。当旋转森林特征子集分块参数为6时，能够取得98.8%的最高覆盖率，同时达到5.75%的误报率；当旋转森林特征子集分块参数为5时，能够取得3.6%的最低误报率，同时达到97.6%的覆盖率。

▼表1 采用的SMART基础属性表

ID	SMART属性名称
1	重映射扇区计数
2	重映射扇区计数（原始值）
3	当前待映射扇区计数
4	当前待映射扇区计数（原始值）
5	底层数据读取错误率
6	主轴起旋时间
7	寻道错误率
8	通电时间
9	（磁头）高飞写入
10	温度（摄氏度）
11	无法校正的错误
12	硬件ECC校正

ECC：错误检查和校正技术
SMART：自我检测分析和报告技术

▼表2 旋转森林备选异构子分类器

类型编号	学习器	备注
①	多层感知机MLP	单隐层节点数目分别为6/7/8/9/10/11/12/13/14
②	支持向量机SVM	高斯核函数，Cost =10，σ分别为0.6/0.7/0.8
③	支持向量机SVM	高斯核函数，Cost =100，σ分别为0.6/0.7/0.8
④	支持向量机SVM	高斯核函数，Cost =1 000，σ分别为0.6/0.7/0.8

MLP：多层感知机　　SVM：支持向量机

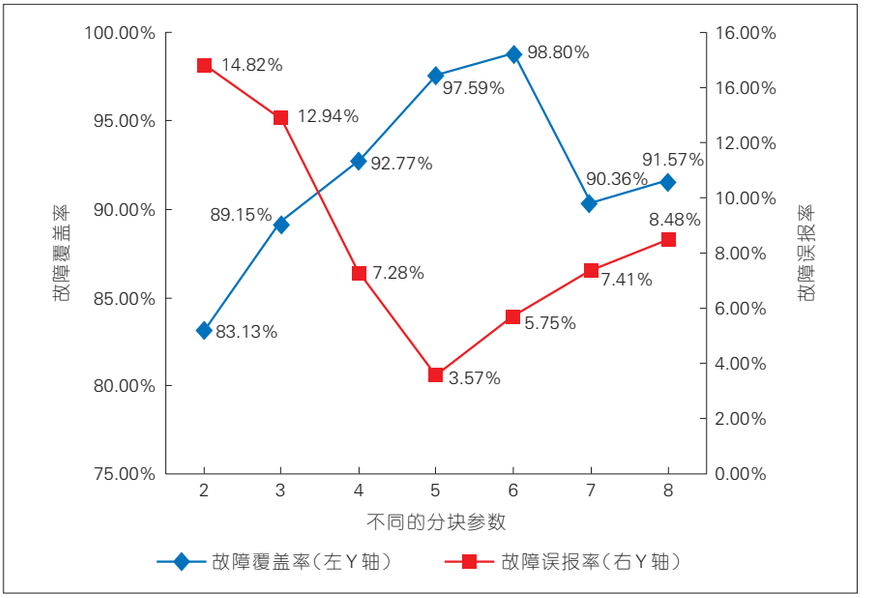
此外，在当前测试中可以发现：绝大部分故障在预报30天内可以被证实，图7是预警提前天数的分布累计情况。

综上所述，基于SMART的故障预测技术在当前智能运维领域已经有了长足的进步和发展，中兴通讯在大视频运维中也适时采用了机器学习方法来提取历史巡检数据中蕴含的故障特征，并构建集成预测模型来提升大视频运维的精度和效率。从当前Backblaze数据集的测试情况来看，也取得了较好的效果。

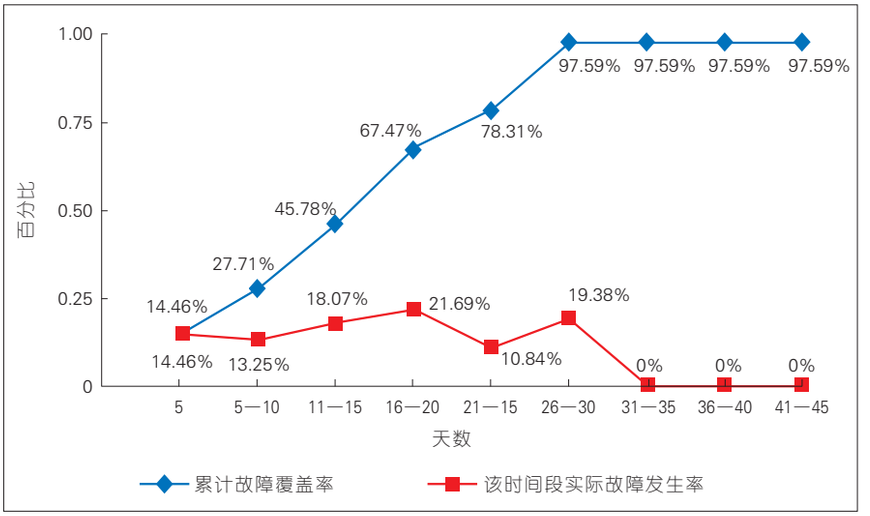
在当前工作的基础上，我们后续将进一步提升人工智能在大视频运维中的落地效果，包括采用半监督学习来提高模型的数据利用率，采用迁移学习来加速模型在新局点的训练部署进度，使用强化学习来优化大视频运维的策略和流程等。

3 结束语

信息通信技术（ICT）时代，无论对于运营商网络还是业务系统的运维支撑，都需要加速与人工智能技术的落地实践，提供高度自动化和智能化的运维解决方案。人工智能、机器学习技术在大视频运维的智能化提升重点体现在运维模式从被动式事后分析转为积极主动预测、分析及决策。随着人工智能技术的加速发展，



▲图6 随旋转森林分块系数变化的故障覆盖率和误报率（预警提前量为30天情况）



▲图7 预警提前天数分布累计(最低误报率情况下)

大视频运维与人工智能技术的结合会越来越紧密,大视频运维技术将朝着更加智能化的方向演进,实现更加自动化和精准的故障预测和排查,主动发现业务系统中的故障或薄弱环节并加以修复。在实现智能运维基础上,通过对视频业务使用者的行为分析、家庭及用户画像等一系列的建模分析,充分挖掘海量数据的价值,衍生出新的业务形态,实现智能化的运营系统,为运营商创造新的商机。

参考文献

[1] 黄珂,李锐,姜春鹤.基于大数据的视频体验保障[J]. 中兴通讯技术(简讯), 2017 (3): 22–25
[2] 基于机器学习的智能运维[EB/OL].(2017–04–22)[2017–06–25].https://zhuanlan.zhihu.com/p/26216857
[3] TANG L, LI, T, PERNG C S. LogSig: Generating System Events from Raw Textual Logs[C]//In Proceedings ACM International Conference on Information and knowledge Management. UK:ACM, 2011:785–794
[4] TANG L, LI T. LogTree: A Framework for Generating System Events from Raw Textual Logs[C]//In Proceedings of IEEE International Conference on Data Mining (ICDM). USA: IEEE, 2010:491–500
[5] LUO C, LOU J G, LIN Q W, et al. Correlating Events with Time Series for Incident Diagnosis[C]//In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM, 2014: 1583–1592
[6] HAMERLY G, ELKAN C. Bayesian Approaches to Failure Prediction for Disk Drives[EB/OL].[2017–06–28]. http://cseweb.ucsd.edu/~elkan/smart.pdf ICML
[7] ECKART B, CHEN X, HE X, et al. Failure Prediction Models for Proactive Fault Tolerance within Storage Systems[C]// Modeling, Analysis and Simulation of Computers and Telecommunication Systems 2008, IEEE International Symposium on. USA: IEEE, 2008. DOI:10.1109/MASCOT.2008.4770560
[8] ECKART B, CHEN X, HE X, et al. Failure Prediction Models for Proactive Fault Tolerance within Storage Systems[J]. IEEE International Symposium on Modeling, 2009, 1(3):1–8. DOI:10.1109/MASCOT.2008.4770560
[9] Hard Drive Reliability Statistics [EB/OL]. [2017–06–28].https://www.backblaze.com/b2/hard–drive–test–data.html

作者简介

屠要峰,中兴通讯股份有限公司云计算及IT研究院副院长;长期从事电信业务及云计算产品的研发工作,主要研究方向云计算、大数据及人工智能;已发表论文10余篇。

吉锋,中兴通讯股份有限公司云计算及IT研究院项目经理;先后从事过IPTV/OTT/CDN、移动互联网相关的产品研发、标准&技术的研究工作,目前研究方向为大数据与人工智能;已发表论文3篇。

文韬,中兴通讯股份有限公司云计算及IT研究院高级系统工程师;研究方向为机器学习与智能运维、推荐系统的结合。