

智测 · 工业设备故障预测分析系统

一、企业简介

智擎信息技术（北京）有限公司专注于工业大数据，利用物联网和机器学习技术为工业企业提供设备数据采集、智能监控、故障预警、自诊断、产线效能优化、运维管理分析、大部件库存优化分析、销量预测等产品和服务。产品能够提供基于公有云及私有云的落地实施。

智擎信息提供针对工业领域深度定制的分析模型软件套件产品，涵盖了设备预测性运维、自诊断、设备性能优化、成本分析、供应链分析等，并具有机器学习自建模等一系列功能和服务。为企业管理者提供战略性的决策支持。智擎信息致力于成为中国乃至全球领先的工业设备及生产线智能分析平台和预测服务提供商。

我们工业 APP 产品是基于 Hadoop 和 Spark 开源生态建立而来。在机器学习建模和分析方面，我们将深度学习和机器学习同行业知识图谱相融合形成了针对工业领域的动态阈值深度学习算法。我们优化了深度神经网络（例如：LSTM 模型），并将故障树、故障诊断机理等信息植入到深度神经网络的增量、增强学习之中。我们独创了工业领域的自动化建模方法。此自动化建模机制针对工业领域中的温度、振动、压力、转速，以及大部件的失效进行了优化，实现了高准确率的自动化调参和建模机制。

二、工业 APP 简介

(一)、问题定位

近些年随着国内工业制造业飞速的发展对设备管理提出了更高的要求。在设备管理和生产线优化方面，随着设备老化程度的持续提高和对生产效率的新要求，很多客户需要一套完整的工业大数据平台来支撑他们从设备运行监控管理、预测分析和运营决策支撑。

本产品覆盖了数据的采集、数据的机器学习建模训练、预测分析模型运行环境、模型库管理、设备健康度分析功能、故障预测模型创建和分析功能、故障树分析、处理措施推荐及窗口期功能等。主要解决了如下问题：降低故障频次及非计划性停机时间、提升设备产能、对设备进行全生命周期的管理。

(二)、创新点

本产品可以通过公有云、私有化部署和混合云的方式部署实现。同时，可以通过移动端的方式为用户提供服务。产品特性涵盖了历史数据的机器学习自动化建模（故障预测模型、关联性故障模型、大部件生存分析模型）、SaaS 化的设备故障预测 APP 应用落地等。主要优势如下：

- 1、针对故障和诊断的自动化建模和调优机制。产品内置了针对行业具体应用场景而优化的自动化建模和模型调优机制，融合了经过行业认可的模型参数库和知识图谱机制，可以根据更新的参数和知识图

谱进行自动化优化从而创建新模型。让设备专家针对模型的创建和优化时间缩短到数小时内。

2、可以针对不同的设备进行多样性模型的部署和管理。也可以针对位于不同工况下的同类设备部署不同版本的模型，便于更好的模型适配和预测准确度。

3、可靠性：考虑到同工业互联网平台的适配，在整个数据仓库的架构设计中，引入 Hadoop 生态系统的多个组件，对于整个 Hadoop 集群，以及每个生态组件，都设计的故障转移机制或者集群。最大程度保证服务的连续性，以及出现故障的自动转移。

4、可扩展性：在 Hadoop 生态系统中各个组件自身均支持分布式部署，在可扩展性方面每个组件均支持在线的扩展性，非常容易进行新的节点和资源的增加，对于集群的管理引入 Apache Ambari 进行管理，可以方便地部署组件以及进行节点的扩展。在负载和主节点的冗余机制中引入 Apache Zookeeper，更易于管理主节点的可靠性。

5、部署灵活性：在设计方案中可采用云服务及私有化部署的方式，在 Hadoop 集群的搭建时采用 Hadoop 以及各个开源组件。

6、易用性：在易用性的设计上面，针对用户可见的操作均设计为简单操作使用，用户可以按照操作手册简单的学习后即可操作。数据模型的创建也都是通过可视化的配置来完成的。

7、安全性：在数据展示端的 B/S 架构系统中采用 Apache Shiro 安全框架进行安全和身份认证的管理，对于不同角色的人员进行功能使用的控制，对于数据的访问按照区域进行数据隔离。

(三)、功能介绍

产品功能包含了数据接入、数据管理、模型管理、故障预测、仪表盘管理、模型调度、诊断和自动化等功能。

数据接入：通过数据采集终端从设备上把传感器的数据采集出来，或者从客户方已经采集出来的数据进行数据转发，最终进行大数据平台的数据接入，其中传输的过程包含了数据的压缩和加密。

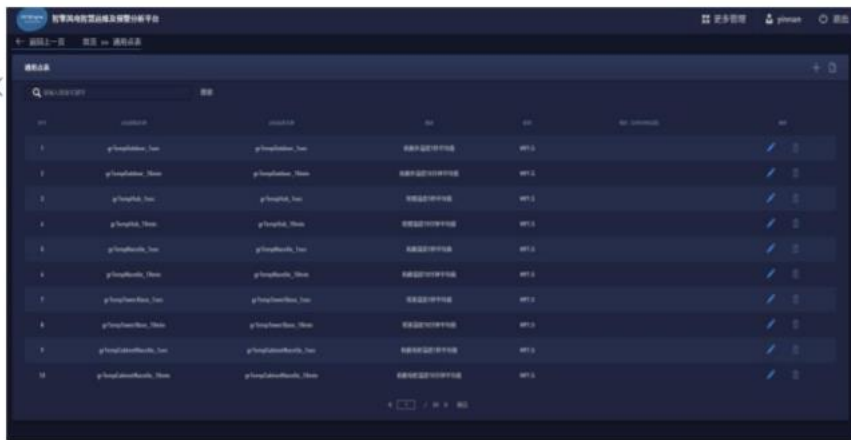


图 19 数据接入界面

数据管理：针对实时接入的设备传感器数据、批量导入的历史数据、以及用户自由上传的文件数据进行统一接入、处理和存储等一体化的管理。

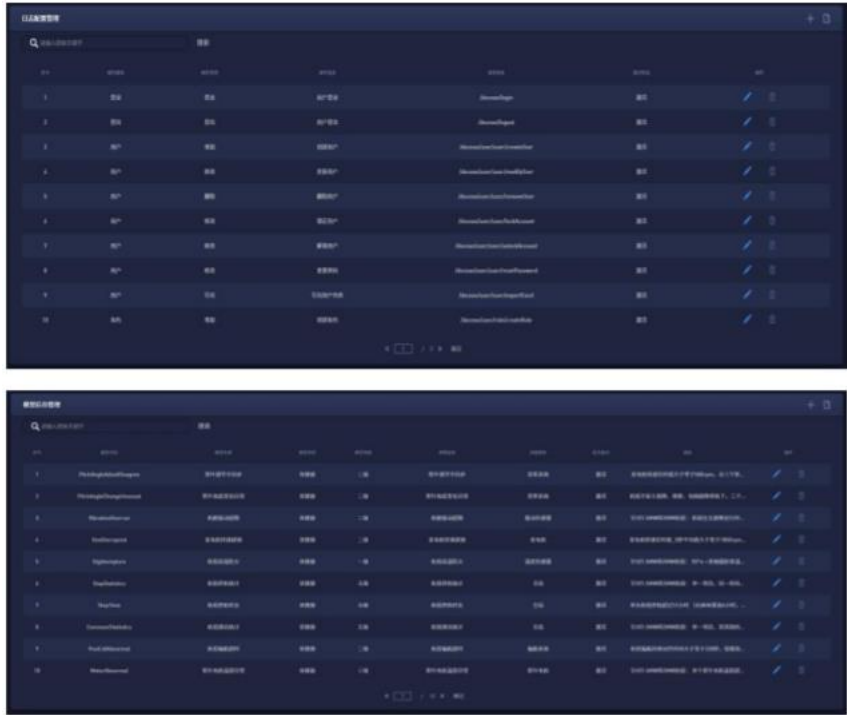


图 20 数据管理界面

模型管理：包含模型的建模过程管理、模型测试、模型评估、模型上线部署、模型调度运行等，从模型的创建、训练、测试、部署、上线全链路流程功能的覆盖。



图 21 模型管理界面

故障预测：通过特定故障场景进行模型的建模训练，完成训练后的模型上线部署，可以针对该故障进行故障发生的提前预测，在预测到故障的发生，即生成故障的预警，针对故障的预警可以由专业的人员进行审核并下发到现场进行故障预警的排查与检修。



图 22 故障预测界面

仪表盘管理：针对数据的可视化，仪表盘是数据最终呈现的方法，可以支持历史数据、实时数据、上传的文本数据，以及故障数据，主数据、故障预警数据等进行可视化展示，操作方便，支持多种类型的图标构建。



图 23 仪表盘管理界面

模型调度：模型调度是针对不同设备的实时数据，针对不同的模型进行定时调度执行的管理功能。



图 24 模型调度界面

诊断和自动化：故障预测模型针对不同的故障场景或者不同的设

备部件进行预测后，生成的预警会关联到特定的设备和部件上，并且针对不同模型的排查与诊断提供方案。针对模型的自动化包括自动化建模和模型上线后的增量自动化优化。



图 25 故障诊断及运维指导界面

(四)、功能和技术指标优势

1、提升模型分析和预测的准确率：通过自动化参数调整和增量/增强学习将预测准确度指标提升 5%。同时，对迁移学习的应用，在未进行预训练状况下，也可以达到较为准确的预测值；

2、统一数据总线层：统一设计数据接入及数据存储层，并形成统一的接口总线，从而建立标准化的数据管理流程。

3、统一业务分析层：实现统一的针对业务的机器学习算法/模型调用层，最终实现从数据接入和处理、模型算法分析预测到展示的集中管理。

4、产品提供多种编程语言接口：Java/Scala/Python/R 等。支持基于分布式 R 和 Python 等主流数据分析软件。

三、技术方案说明

(一)、工业 APP 架构

本产品基于 Hadoop 大数据平台基础之上，利用 Hadoop 构建分布式集群，进行数据的分布式存储，利用 spark 构建分布式计算框架，进行数据和模型的分布式计算，利用 Hive 和 HBase 搭建数据仓库，自研发数据处理和管理组件。

机器学习算法模型利用 tensorflow 构建深度学习模型训练框架，自研发自组织自动化数据标记算法，自研发分布式模型执行运算调度框架，自研发自动化故障建模框架。产品架构设计如下：



图 26 工业设备故障预测分析系统 APP 架构图

智能终端做为智擎工业设备数据采集运算终端，可以与设备 PLC 通信，把传感器的实时数据，录波数据采集出来，并且作为端的运算终端，针对实时的数据进行初步的运算，然后进行大数据平台的

数据接入，可以支持数据的断点续传。WYEngine SDK Apollo 作为产品数据接口的 SDK 组件，运行与 Hadoop 大数据平台执行，按照工业设备的数据应用场景进行功能组件的封装，实现从数据的接入、数据处理、数据存储、数据运算、模型管理和任务调度等功能，为上层应用提供平台级的支撑。

此外，APP 产品功能覆盖了针对数据的管理，模型管理，模型库，应用仪表盘，模型评估和模型调度，并且封装一系列针对工业场景优化的算法和业务场景。

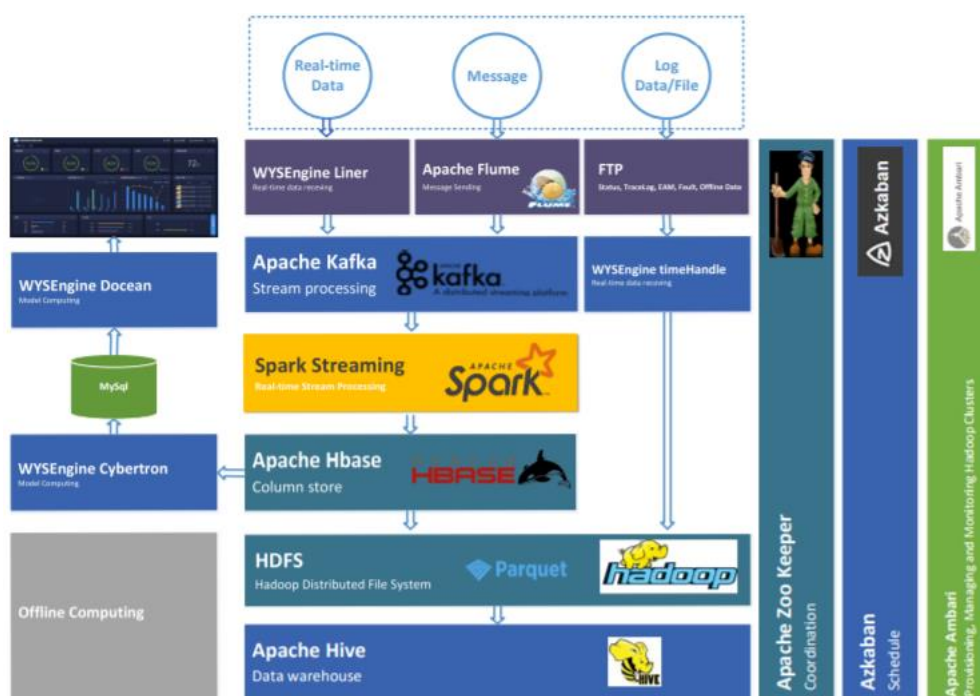


图 27 工业设备故障预测分析系统 APP 组件图

HDFS 作为实时数据和历史数据的存储平台，使用服务器集群进行部署，HDFS 中的文件存储采用 Apache Parquet 的存储格式进行存储，最大程度上进行数据的压缩存储，并且保证数据使用时真实性。 Apache Hadoop Yarn 作为整个 Hadoop 集群的资源管理器，对所有计算任务所需要的资源进行调度管理，并针对集群资源进行

管理。

Apache Zookeeper 是一个分布式的分布式应用程序协调服务，可以为分布式应用提供一致性服务，作为 HDFS 的 NameNode 主备的管理，针对 Kafka 集群的协调管理，针对 Spark 集群的协调管理。

数据仓库的实时数据采集部分由 Flume 完成，Flume 对外提供 API，由外部程序调用 API 进行数据的传输，Flume 接收到的数据放入 Apache Kafka 的消息队列中进行缓存，以同步数据采集与数据处理存储的速度不一致性；

文件日志数据的采集方位为开放 FTP 共享文件夹，由外 10 部程序进行日志数据的写入，Flume 监控文件夹，对新写入的数据进行处理并存入 HDFS 中。

数据处理 ETL 部分的内容由 Spark Streaming 进行处理，处理完成后将数据存储到 HDFS 中，另外，处理完成的数据进行后续的计算。

分析部分在数据 ETL 清洗之后进行特征选择、知识图谱特征的融合、机器学习模型训练、模型评估、模型保存和基于效果机制的增量增强学习，最终开放成为 API 供调用。

整个运算部分的内容由 Spark 集群进行处理，处理后的结果数据写入 MySQL 数据库，以供展示系统抽取数据并进行展示。展示端的直接短接 MySQL 数据库，需要确保 MySQL 数据库中的数据保持的是小量级，汇总或者处理后的数据，以确保最终的展示端应用的响应速度。对于数据的查询和使用主要提供 Spark SQL、Hive 组件

支持类传统 SQL 的数据查询。

实时数据指从设备现场的工业设备中采集的实时数据，按照点位进行采集，每个点位包含点位的名称、内容、时间戳等字段，不同机型的工业设备采用的点位配置表不同，不同的风场的工业设备，或者相同机型的不同的工业设备所使用的点位配置表也不同。

(二)、工业 APP 关键技术

在关键技术路线方面，主要引入了在线、离线相融合的复杂调度方式。这种方式将比较高效的解决数据接入、建模计算和实时预测各个方面的资源调度难题。尤其是针对深度学习的建模和自学习过程，将有效的避免波峰波谷过于明显的问题。

在模型创建和分析预测方面，我们使用了深度学习、机器学习、知识图谱相结合的方式。在深度学习方面，我们使用 LSTM 等算法，以及使用了增量和增强学习用于自动化建模和调优。

对于准确率的判别方面，我们利用了混淆矩阵来进行，主要技术方面利用内存技术、MPP 存储共同优化大数据平台的瓶颈问题。

此外，冷热数据存储和对机器学习自学习方面都提供强有力的支撑，尤其是对增量学习和迁移学习方面，平台可以存储相关的可变参数模板来提升算法模型准确率。

四、应用情况描述

（一）、应用场景描述

开展设备运行现场精益运维管理，对现场运维期的业务进行细化，将数据分析、资产管理、故障预测和诊断、设备 KPI 管理、物资保障和作业窗口等集中在一起，作为基础数据，结合现场运维策略，将现场工作任务进行综合管理，降低设备运维成本（减小备品备件损失和更换频次）、降低设备故障频次和停机时间，提高设备可利用率及收益。其中，重点实现机组故障自诊断分析，推送合适作业窗口、备件位置、数量信息及故障解决方案；实现对现场 KPI 的管理，给现场运维计划提供数据支撑。以上信息均以工单的形式下发至设备现场，解决了现有的工单不能自动创建、备件位置和数量查询繁琐等问题，改善了故障处理效率，降低了设备损失及运维成本。

（二）商业化情况

本产品主要应用在发电行业和石化领域，在发电行业的客户有金风、明阳、上海电气，在石化领域的客户有中海油。

本产品帮助用户降低故障频次及非计划性停机时间、提升设备产能、对设备进行全生命周期的管理。帮助用户降低非计划性停机时间达到 19%（平均），降低故障频次达到 21%（平均）。降低直接成本消耗。此外，促进客户的信息化平台的统一性，数据管理的一致性都起到非常重要的作用。

本产品帮助公司针对不同客户的服务投入减小 40%，并且可以提升运算和模型自优化准确度。针对不同行业客户的项目交付成本将减少 40%，利润增加 15%以上。

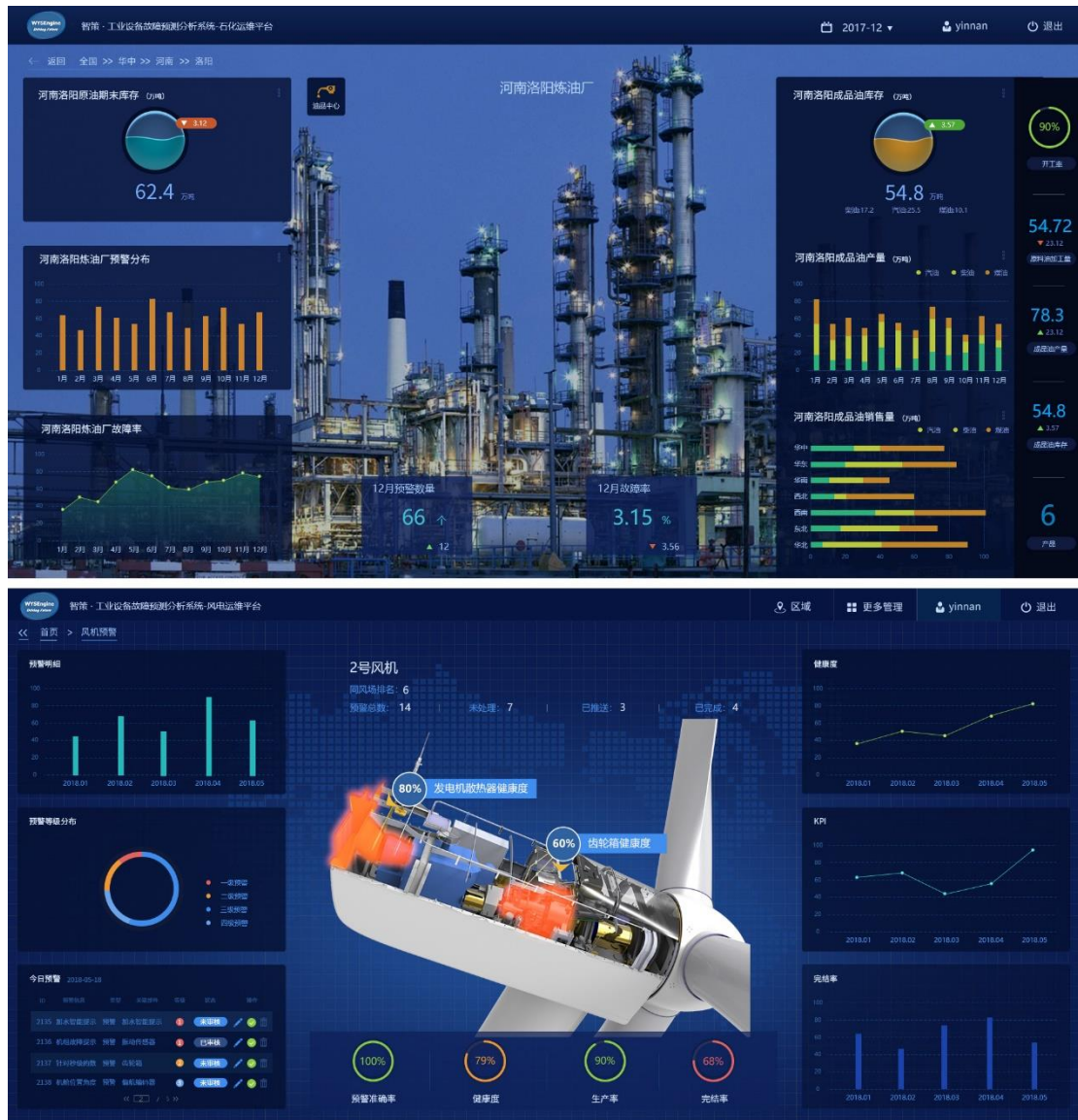


图 28 工业设备故障预测分析系统 APP 效果图