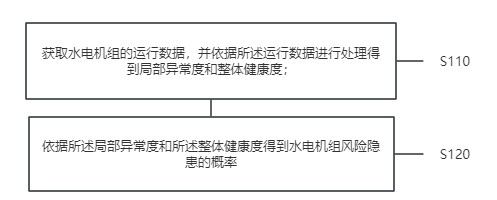
在本申请的实施例中，本申请提供了“为实现水电设备健康预警分析智能化应用，首先以单变量测点、多变量测点构建健康度偏离度评估设备风险趋势模型，然后在此基础上构建融合集成模型，形成最终异常预警结论，最后使用概率图模型分析潜在隐患”的解决方案，具体为：获取水电机组的运行数据，依据所述运行数据进行处理得到局部异常度和整体健康度；依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。本申请通过基于大数据和AI的方式研究实现了水电机组设备的健康管理体系，通过水电机组设备健康预警预测模型体系、设备测点预警预测模型、劣化趋势模型、数据趋势分析模型，构建了完整的基于大数据的机组健康风险管理体系。



1、一种基于水电机组的风险预测方法，其特征在于，所述方法用于对水电机组设备进行风险隐患预测，包括：

获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；

依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

2、根据权利要求1所述的基于水电机组的风险预测方法，其特征在于，依据所述运行数据进行处理得到局部异常度和整体健康度的步骤，包括：

依据所述运行数据进行处理得到初始数据；其中，所述初始数据包括初始水轮发电机数据、初始负荷数据以及初始环境温度数据；

依据所述初始数据得到局部异常度和整体健康度。

3、根据权利要求2所述的基于水电机组的风险预测方法，其特征在于，依据所述初始数据得到局部异常度和整体健康度的步骤，包括：

依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度；

依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度。

4、根据权利要求3所述的基于水电机组的风险预测方法，其特征在于，依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度的步骤，包括：

依据所述初始水轮发电机数据和初始负荷数据和预设单测点模型得到目标状态量；

依据所述目标状态量和标准状态量进行比对得到所述局部异常度。

5、根据权利要求3所述的基于水电机组的风险预测方法，其特征在于，依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度的步骤，包括：

依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据输入预设多测点模型得到预测健康度；

依据所述预测健康度和标准健康度进行比对得到整体健康度。

6、根据权利要求1所述的基于水电机组的风险预测方法，其特征在于，依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率的步骤，包括：

将多个模型的所诉局部异常度和所述整体健康度输入集成模型，得到整个水电机组健康状态；

发现异常后，使用概率图模型对测点数据进行处理，分析可能造成的风险隐患及其概率。

7、一种基于水电机组的风险预测系统，所述系统用于对水电机组设备进行风险隐患预测，其特征在于，包括：

获取模块，用于获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；

隐患概率模块，用于依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

8、一种电子设备，其特征在于，包括处理器、存储器及存储在所述存储器上并能够在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现如权利要求1至6中任一项所述的基于水电机组的风险预测方法的步骤。

9、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质上存储计算机程序，所述计算机程序被处理器执行时实现如权利要求1至6中任一项所述的基于水电机组的风险预测方法的步骤。

**一种基于水电机组的风险预测方法及系统**

**技术领域**

本发明涉及大数据，人工智能领域，具体涉及一种基于水电机组设备的风险预测方法及系统。

**背景技术**

水电机组是一套复杂的机电系统，具有故障率高和故障危害性大的特点。因此，为了避免设备出现严重的后果，对水电机组进行实时状态监测及设备健康管理很有必要。水电机组系统设备复杂，各子系统之间以及子系统与系统之间关联性强，各功能部件相互协作，任何一个功能失效都会造成整个系统瘫痪。

目前行业主流的数据挖掘分析方法相对滞后，大多数判断依靠专家经验，无法发挥数据本身的价值以挖掘其中的数据规律，弥补专家经验考虑不充分，无法穷举所有情况的不足。一个专业稳定的业务模型需要花费大量的时间去梳理，建模和验证，并没有通过大数据技术将人力资源释放，同时对于大型水电站而言，设备数量多，设备复杂度高，需要建立大量的业务模型，因此对运维人员而言数据分析工作量巨大。

**发明内容**

鉴于所述问题，提出了本申请，一种基于水电机组的风险预测方法及系统，包括：

一种基于水电机组的风险预测方法，所述方法用于对水电机组设备进行风险隐患预测，包括：

获取水电机组的运行数据，依据所述运行数据进行处理得到局部异常度和整体健康度；

依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

优选地，所述依据所述运行数据进行处理得到局部异常度和整体健康度的步骤，包括：

依据所述运行数据进行处理得到初始数据；其中，所述初始数据包括初始水轮发电机数据，初始负荷数据以及初始环境温度数据；

依据所述初始数据得到局部异常度和整体健康度。

优选地，所述依据所述初始数据得到局部异常度和整体健康度的步骤，包括：

依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度；

依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度。

优选地，所述依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度的步骤，包括：

依据所述初始水轮发电机数据、初始负荷数据和预设单测点模型得到目标状态量；

依据所述目标状态量和标准状态量进行比对得到所述局部异常度。

优选地，所述依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度的步骤，包括：

依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据输入预设多测点模型得到预测健康度；

依据所述预测健康度和标准健康度进行比对得到整体健康度。

优选地，所述依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率的步骤，包括：

将单测点预警预测模型的局部异常度、多测点预警预测模型的局部异常度和整体健康度输入集成模型，得到整个水电机组健康状态；

发现异常后，使用概率图模型对测点数据进行处理，分析可能造成的风险隐患及其概率。

为实现本申请还包括一种基于水电机组的风险预测的系统，所述系统用于对水电机组设备进行风险隐患预测，包括：

获取模块，用于获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；

隐患概率模块，用于依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

为实现本申请还包括一种电子设备，包括处理器、存储器及存储在所述存储器上并能够在所述处理器上运行的计算机程序，所述计算机程序被所述处理器执行时实现如所述的基于水电机组的风险预测方法的步骤。

为实现本申请还包括一种计算机可读存储介质，所述计算机可读存储介质上存储计算机程序，所述计算机程序被处理器执行时实现如所述的基于水电机组的风险预测方法的步骤。

本申请具有以下优点：

在本申请的实施例中，相对于现有技术中的“大多数判断依靠专家经验，无法发挥数据本身的价值以挖掘其中的数据规律，弥补专家经验考虑不充分，无法穷举所有情况的不足。一个专业稳定的业务模型需要花费大量的时间去梳理，建模和验证，并没有通过大数据技术将人力资源释放，同时对于大型水电站而言，设备数量多，设备复杂度高，需要建立大量的业务模型，因此对运维人员而言数据分析工作量巨大。”，本申请提供了“为实现水电设备健康预警分析智能化应用，首先以单变量测点，多变量测点构建健康度偏离度评估设备风险趋势模型，然后在此基础上构建融合集成模型，形成最终异常预警结论，最后使用概率图模型分析潜在隐患”的解决方案，具体为：获取水电机组的运行数据，依据所述运行数据进行处理得到局部异常度和整体健康度；依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。本申请通过基于大数据和AI的方式研究实现了水电机组设备的健康管理体系，通过水电机组设备健康预警预测模型体系、设备测点预警预测模型、劣化趋势模型、数据趋势分析模型，构建了完整的基于大数据的机组健康风险管理体系，完成提高水电机组早期缺陷劣化方向的精准研判，提高水电机组健康管理智能化水平；机组的设备健康预警预测体系，利用人工智能结合大数据分析，真正做到挖掘数据规律，从数据角度说话，推断设备风险状态隐患的概率，做到了事前防范；机组预警预测分析体系，计算部分系统自动执行，通过数据解释性，模型解释性和少量人员经验参与，减少人工经验判断带来的偏差，提高趋势判断依据，减少分析工作。

**附图说明**

为了更清楚地说明本申请的技术方案，下面将对本申请的描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1是本申请一实施例提供的一种基于水电机组的风险预测方法的步骤流程图；

图2是本申请一实施例提供的一种基于水电机组的风险预测的系统的结构框图；

图3是本发明一实施例提供的一种计算机设备的结构示意图；

图4是本申请一实施例提供的一种基于水电机组的风险预测方法的逻辑架构图；

图5是本申请一实施例提供的一种基于水电机组的风险预测方法大数据架构图；

图6是本申请一实施例提供的一种基于水电机组的风险预测方法整体数据流架构图；

图7是本申请一实施例提供的一种基于水电机组的风险预测方法的样本原理图；

图8是本申请一实施例提供的一种基于水电机组的风险预测方法置信水平与置信区间关系图；

图9是本申请一实施例提供的一种基于水电机组的风险预测方法的置信区间图；

图10是本申请一实施例提供的一种基于水电机组的风险预测方法的相关系数图；

图11是本申请一实施例提供的一种基于水电机组的风险预测方法的风险异常概率原理图；

图12是本申请一实施例提供的一种基于水电机组的风险预测方法优化图；

图13是本申请一实施例提供的一种基于水电机组的风险预测方法单变量偏离度健康估计图；

图14是本申请一实施例提供的一种基于水电机组的风险预测方法孤立随机森林对残差点的分数分布图；

图15是本申请一实施例提供的一种基于水电机组的风险预测方法的多测点流程图；

图16是本申请一实施例提供的一种基于水电机组的风险预测方法的SBM算法流程图；

图17是本申请一实施例提供的一种基于水电机组的风险预测方法超球算法算法训练原理图；

图18是本申请一实施例提供的一种基于水电机组的风险预测方法LSTM算法训练原理图；

图19是本申请一实施例提供的一种基于水电机组的风险预测方法隐患预测整体逻辑思路图；

图20是本申请一实施例提供的一种基于水电机组的风险预测方法贝叶斯网络图；

图21是本申请一实施例提供的一种基于水电机组的风险预测方法贝叶斯网络模型业务解释性图。

**具体实施方式**

为使本申请的所述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本申请作进一步详细的说明。显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

技术背景：水电机组是一套复杂的机电系统，具有故障率高和故障危害性大的特点。因此，为了避免设备出现严重的后果，对水电机组进行实时状态监测及设备健康管理很有必要。水电机组系统设备复杂，各子系统之间以及子系统与系统之间关联性强，各功能部件相互协作，任何一个功能失效都会造成整个系统瘫痪。

当前水电站普遍开展水电机组设备健康趋势的分析工作，在设备早期隐患发现、设备异常预测等方面取得一定成效，保障了机组的安全稳定运行。但在诊断机组健康运行方面，仍然以报警定值为辅，以人工分析为主，主要依靠运维人员用肉眼从浩瀚的设备数据中提取特征数据，发现设备异常，或者依赖专家经验对设备进行流程式建模，根据在模型中设置的阈值或条件来判断设备是否存在异常，此种模式太过于直接，也忽略了设备各部件之间的相关性。

目前基于大数据的水电机组设备健康管理存在的问题，随着传感器技术、数据采集、传输及处理技术等ITC技术的飞速发展，水电站运行过程中产生越来越多、精度越来越高的生产数据，历史数据库中存储的数据量急剧膨胀，分析并利用这样海量的数据已经远远超出了运维人员的能力，目前行业主流的数据挖掘分析方法相对滞后，无法精准的从大量生产数据中自动提取分析有价值的数据和预测信息，最大限度发挥数据价值，出现了“数据非常丰富，信息相当贫乏”的“数据坟墓”现象，导致分析平台在提前预警、早期预防、及时干预等方面存在功能缺陷，无法及时发现水轮发电机组轴承油槽油混水、轴承油槽油位下降等设备的安全隐患。

由于水电站生产环境的复杂性，数据清洗任务的复杂性，一个业务模型落地需要花费大量的时间，对于水电机组设备的健康诊断模型而言，通常需要建立上千个业务模型，实施周期长且难度较大，并且专家经验是有限的，不可能穷举所有场景案例。

运行数据随着水电机组工况变化，环境变化会有变动，比如机组C修后，数据波动较大，特征空间影响较大。

水电机组设备健康管理是非常交叉的学科，涉及电气、机械、化学、大数据等，对人员素质要求极高。

大型水电站一方面设备数量众多，数据量巨大，运维人员数据分析工作繁重，另一方面现代管理制度要求水电站实施精益化运维管理，人员配置越来越精简。海量的数据分析工作和趋于精简的人员配置之间的矛盾日益突出。

通过大数据和人工智能技术覆盖水电机组以及水电站其他设备的状态诊断和趋势分析，主要包括：特征数据的对比分析，场景的相似度计算，隐患模式的识别分析和特征统计规律分析。目的在于评估风险，趋势预测，决策计算，可靠性指标分析，模型系统框架的自学习过程。最终实现设备隐患的预测挖掘分析，设备特征数据的趋势分析，设备健康整体评价，劣化指标分析等，形成完整闭环，提高大数据管理水电机组健康的能力，减少专家经验的干预，从数据规律本身推出初步建议，专家辅助审核，提高智能化管理能力。对于水电站的数字化建设而言，其核心本质是降低成本，提高效率，只有通过利用大数据结合AI算法挖掘数据隐患规律和趋势，才能减少人员和时间的投入，才能真正做到单体水电站的智能化，拓展到流域级别的智慧化应用。逻辑架构图如4所示。

针对目前水电机组设备健康管理所面临的数据情况而言，所涉及的数据大致每台机组每天30GB+，整体每年属于TB级别数据，同时基于水电特殊的环境情况而言，需要考虑时序数据，原始数据，特征数据等大量的数据分析的需求，以及数据清洗，异常数据溯源等需求，所以建立的大数据架构需要满足强大的基准性，综合以上需求，整体的大数据架构图如5所示。

整体数据流架构以工业互联网平台为基础，实现智能应用在本身数据流向上的架构划分，不改变工业互联网架构主体。首先从大数据中心采集数据，经过数据清洗后，放入大数据平台，最终分析数据价值，应用数据。整体数据流架构如图6所示。

水电机组健康预警预测分析是一个非常复杂的系统工程场景，为了高效执行TB级特征数据查询、特征关系分析、异常回溯复盘、数据清洗等任务，数据库设计必须非常灵活。具体的数据库设备包括如下方面：原始数据库：原始数据库用于存储与备份，主要针对原始没有清洗过的数据；历史数据库：历史数据库主要存储清洗后的数据，作为正常运行工况数据；异常数据库：主要存储健康评价后且告警后30分钟左右的数据，应用于异常复盘；时序实时数据库：用于存储时序实时状态量数据；特征数据库：用于存储特征数据；中间结果数据库：主要存储一些告警数据或者一些中间结果数据。

在本申请的实施例中，相对于现有技术中的“大多数判断依靠专家经验，无法发挥数据本身的价值以挖掘其中的数据规律，弥补专家经验考虑不充分，无法穷举所有情况的不足。一个专业稳定的业务模型需要花费大量的时间去梳理，建模和验证，并没有通过大数据技术将人力资源释放，同时对于大型水电站而言，设备数量多，设备复杂度高，需要建立大量的业务模型，因此对运维人员而言数据分析工作量巨大。”，本申请提供了“为实现水电设备健康预警分析智能化应用，首先以单变量测点，多变量测点构建健康度偏离度评估设备风险趋势模型，然后在此基础上构建融合集成模型，形成最终异常预警结论，最后使用概率图模型分析潜在隐患”的解决方案，具体为：获取水电机组的运行数据，依据所述运行数据进行处理得到局部异常度和整体健康度；依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。本申请通过基于大数据和AI的方式研究实现了水电机组设备的健康管理体系，通过水电机组设备健康预警预测模型体系、设备测点预警预测模型、劣化趋势模型、数据趋势分析模型，构建了完整的基于大数据的机组健康风险管理体系，完成提高水电机组早期缺陷劣化方向的精准研判，提高水电机组健康管理智能化水平；机组的设备健康预警预测体系，利用人工智能结合大数据分析，真正做到挖掘数据规律，从数据角度说话，推断设备风险状态隐患的概率，做到了事前防范；机组预警预测分析体系，计算部分系统自动执行，通过数据解释性，模型解释性和少量人员经验参与，减少人工经验判断带来的偏差，提高趋势判断依据，减少分析工作。

参照图1，示出了本申请一实施例提供的一种基于水电机组的风险预测方法的步骤流程图，具体包括如下步骤：

S110、获取水电机组的运行数据，依据所述运行数据进行处理得到局部异常度和整体健康度；

S120、依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

下面，对本示例性实施例中基于水电机组的风险预测方法作进一步说明。

如上述步骤S110所述，获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度。

如下列步骤所述，依据所述运行数据进行处理得到初始数据；其中，所述初始数据包括初始水轮发电机数据，初始负荷数据以及初始环境温度数据；依据所述初始数据得到局部异常度和整体健康度。

在本发明一实施例中，可以结合下列描述进一步说明步骤S110所述“获取水电机组的运行数据，依据所述运行数据进行处理得到初始数据”的具体过程。

如下列步骤所述，依据所述运行数据进行采样得到采样数据；其中，所述运行数据包括水轮发电机数据、负荷数据以及环境温度数据；依据所述采样数据进行清洗得到清洗数据；依据所述清洗数据进行特征处理得到所述初始数据；其中，所述初始数据包括初始水轮发电机数据，初始负荷数据以及初始环境温度数据。

在本发明一实施例中，可以结合下列描述进一步说明步骤所述“依据所述运行数据进行采样得到采样数据”的具体过程。

在一具体实施例中，数据采样算子的开发及部署，是基于设备正常健康样本数据进行模型搭建、模型训练、模型评估，从而建立并获得正常样本的空间分布规律；再通过构建一个时间序列样本数据评估模型，自适应地选择均匀采样、广义平滑加权采样(GSW)、空间采样(OCSVM)方法或不规则采样方法，以得到适用于模型训练的样本数据。应提供至少四种采样算子(包括但不限于均匀采样、广义平滑加权采样、空间采样、不规则采样等)及至少一种时间序列样本数据评估模型。

需要说明的是，由于数据较为稠密，不同工况下的正样本数据极具类似，假如不采样，会带来很大的冗余计算，浪费大量的计算资源和等待时间。因此，构建正样本空间必须采取合理的采样办法，通过对正样本空间分布进行分析，同时结合水电机组各种工况特点，设计采样方法。

作为一种示例，本申请可以采用均匀采样、广义平滑加权采样(GSW)、空间采样(OCSVM)、蓄水池采样算法、不规则采样、LTTB采样、分层均匀随机抽样中任意一种采样方法，再使用时序模型对采样后的数据进行评估。

在一具体实施例中，对于采样后的样本时序数据需要进行初步检验，以确定采样后的样本数据和总体数据在统计意义上属于无偏估计。常用估计量是：样本平均值，样本方差，样本标准差。原理图以及一般采用的方法如图7所示。

点估计：对于来自一个测量总体的任何随机样本，如果对随机量计算出一个具体的数值（某个样本的均值、方差或标准差），用以估计总体的参数（例如：总体的均值、方差或标准差），则该数值称为总体参数的一个点估计。用点估计反映总体参数时，应该给出尽可能多的附加信息，以便于评价估计值的准确度和精度。准确度受度量方法和抽样设计影响；精度则由固定容量n的样本标准差决定，标准差越小越精确。

区间估计：尽管有点估计及其准确度和精度的一些信息，但是仍然未能从样本跳跃到总体，即未能把点估计与待估总体参数联系起来，给出估计对参数的接近程度或确定在估计值中存在多大的可能误差。为了从样本信息推断总体参数，需要用到区间估计。区间估计是一个从样本到总体的推断，区间估计将总体参数置于一个实区间上。区间的边界值由三个因素决定：①样本点估计值；②联系总体参数和样本点估计的样本统计量（如Z统计量，做正态变换得到）；③该统计量的抽样分布（例如，样本均值的理论抽样分布服从正态分布，则Z统计量的抽样分布是标准正态分布）。

通过对水电机组设备时序数据进行正样本采样，并对时序采样数据计算置信水平估计和置信区间估计，发现采样正样本空间符合样本总体的分布规律，可作为无偏差估计。大致步骤如下：

置信区间估计，预估的总体某个指标的范围；

置信水平估计，是对预估结果准确概率的要求，也称为置信度；

显著性水平设置，为1-置信水平，通常用 1700114753672(1)表示。置信水平与置信区间关系如图8所示。

置信区间估计：估计置信区间，通过抽样样本估计总体样本概率特征，通过如下4步骤可以实现，如图9所示。

在本发明一实施例中，可以结合下列描述进一步说明步骤所述“依据所述采样数据进行清洗得到清洗数据”的具体过程。

在一具体实施例中，数据清洗算子的开发及部署，是针对水电设备原始传感器信号数据类型多样、数据尺度差异大的特点，同时考虑到部分测点在时间维度上存在较多的缺失数据及由于传感器异常等原因存在大量死点和数据漂移的现象，开发的专用数据清洗算子。针对异常值(离群)问题：设计机器学习算子和深度学习算子进行无监督检测。针对缺失值问题：应根据不同的需求设计机器学习算子和深度学习算子。针对非平稳序列问题：采用多种差分算子相结合(包括且不限于一阶差分、二阶差分及季节差分)的方式对数据进行平稳化，使原始非平稳的时序数据转化为可预测的平稳数据以便后续建模。针对时序噪声问题：为了使时序数据更加平滑以实现准确预测，设计多种滤波算子，包括并不限于移动平均算子、高斯过程平滑算子以及指数平滑算子等。

作为一种示例，对于水电机组设备而言，数据清洗要占据大量的时间，目的在于还原真实的运行过程。对于水电机组设备主要存在离群数据，丢失数据，延迟数据，无效数据等情况的处理，通过历史数据结合台账缺陷等数据进行数据清洗，完成设备运行数据的真实还原，此时采用的算法如下：

离群数据采用如下算法，进行综合验证，适应不同的业务场景：Kmeans离群算法：持续计算各样本点和簇中心点之间的欧式距离，持续收敛到最佳状态，找到最优K值，并以聚类距离作为离群值判定依据；KNN离群算法：利用相邻的K个样本点稀疏程度或蜂窝程度，判断是否为离群数据；LOF局部离群因子算法：根据样本局部周围密度稀疏情况，判断离群数据；孤立森林算法：孤立森林将类似于决策树对数据进行递归划分，划分次数越多则紧密程度越强，划分次数越少则判定为离群数据的概率越高。四分位数算法：基于数据分布3sigma原则去除离群点。

延迟数据处理采用如下算法：对于少量的延迟数据，依据数据的强相关性，利用前序数据规律经过简单的换算（如：加权平均法等），可直接实现对数据序列的补全。

数据丢失处理采用如下算法：对于大量的丢失数据，依据数据的强相关性，用SVR、CONVLSTM算法利用前序数据预测后续数据，实现对数据序列的补全。

无效数据处理采用如下算法：根据数据的强相关性，用SVR、CONVLSTM算法利用前序数据预测后续数据，实现对数据序列的修正。

SVR算法整体过程解释如下：SVR算法是机器学习算法，对于非线性特征数据拟合能力较强，也常常用于线性回归场景，拟合精度较高。在实际应用中一般先用SVR先进行模型试验，假如试验过程中发现特征数据在时间和空间上存在依赖性，且数据量较大时，采用CONVLSTM进行拟合，目前两个算法一般配合使用，算法精度较高。

CONLSTM算法整体过程解释如下：对于CONLSTM算法而言，不仅仅可以学习时间维度和空间维度的特征相关性记性训练，同时对于前序数据的波动具有记忆功能，往往用于复杂数据的拟合，在实施过程中一般和SVR配合使用。

在一具体实施例中，在历史数据中确定清洗规则和清洗算法参数，当算法稳定后，用于实时数据清洗，以完成整个数据集的清洗功能。此处理过程自动化进行，人工参与少，整个过程可可视化展示，具有很强的业务解释性，同时便于清洗的正确性判断，用户可自定义操作过程控制。

在一具体实施例中，对于测点数据中的非平稳性问题，采用移动平均，加权平均，多阶差分算子等对数据进行平稳化，根据实际情况选择合适算子，一般在水电领域，常用多阶差分和加权平均，把非平稳数据转化为平稳数据，进行智能算法分析。常见的平滑处理算法如下：移动平均、简单移动平均（SMA）、指数加权移动平均、指数平滑、局部加权散点平滑、数据噪声处理。

在一具体实施例中，对于测点的噪声情况而言，一般采用滤波算子进行处理，比如移动平均算子，指数平均算子，高斯滤波平滑算子，还原时序数据的真实信号，方便精准预测。常见的噪声处理技术如下：低通滤波器、Kalman滤波、Savitzky-Golay滤波器。

在本发明一实施例中，可以结合下列描述进一步说明步骤所述“依据所述清洗数据进行特征处理得到所述初始数据”的具体过程。

在一具体实施例中，特征工程算子开发及部署，通过分析设备多个测点特征的相关性，开发特征工程算子有效削减冗余信息，并对数据的维度进行规约和变换，解决测点数据之间由于高度相关、尺度不同对预测结果的不利影响。特征变换(缩放)：针对数据尺度差异问题，设计时序数据正则化及标准化算子，以使不同设备的数据范围更加统一。特征降维：针对特征稀疏问题，需开发PCA等传统降维算子，并研究基于神经网络自编码器的非线性降维算子。

作为一种示例，将清洗数据依次进行冗余特征消除、特征变换、特征降维以及特征选择得到所述初始数据。

在一具体实施案例中，冗余特征消除：主要用于消除状态量之间的高相关性影响以及删除无效特征。在水电的物理系统中，独立系统的测点数据，比如发电机、水轮机等数据都是具有相关性，只是相关性的强弱不同，一般会先采用相关系数进行初步分析，如图10所示，对于相关性很强的特征一般不会删除其中某一特征，而是对这些特征进行加权平均成为新的特征后输入模型进行训练，这样既保留了原始状态量的物理意义，又考虑了两者相关性的影响。对于无效特征，即对预测目标几乎无用的特征，可以使用特征筛选方法进行过滤，比如常见的卡方检验、方差分析、信息增益等，在实际使用中往往多种方法相互印证，选择最优方法。

在一具体实施例中，特征变换（缩放）：主要消除特征间量纲不同带来的不利影响，使不同量纲的数据转换到同一量纲，即无量纲化。常见的无量纲化方法有标准化和区间缩放法。标准化的前提是特征值服从正态分布，标准化后，其转换成标准正态分布。区间缩放法利用了边界值信息，将特征的取值区间缩放到某个特点的范围，例如[0,1]等。

在一具体实施例中，特征降维：由于水电机组设备涉及测点相对较多，输入模型的特征量维度一般较高，将导致计算量大、训练时间长的问题，因此降低特征矩阵维度必不可少。常见的降维方法除了基于L1惩罚项的模型、自编码神经网络以外，还有主成分分析法(PCA)和线性判别分析(LDA)。PCA和LDA有很多的相似点，其本质是要将原始的样本映射到维度更低的样本空间中，但是PCA和LDA的映射目标不一样：PCA是为了让映射后的样本具有最大的发散性；而LDA是为了让映射后的样本有最好的分类性能。所以说PCA是一种无监督的降维方法，而LDA是一种有监督的降维方法。

在一具体实施例中，特征选择，当数据预处理完成后，需要选择有意义的特征输入机器学习模型进行训练。通常来说，从两个方面考虑来选择特征：①特征是否发散：如果一个特征不发散，例如方差接近于0，也就是说样本在这个特征上基本上没有差异，这个特征对于样本的区分并没有什么用。②特征与目标的相关性：这点比较显见，与目标相关性高的特征，应当优先选择。可以将特征选择方法分为3种：过滤法、包装法、嵌入法。过滤法：按照发散性、相关性或者其他指标对各个特征进行评分，设定评分阈值或者特征数量阈值以选择特征，主要的方法有方差选择法、相关系数法、卡方检验法、互信息法等。包装法：根据目标函数(通常是预测效果评分)，每次选择若干特征，或者排除若干特征，主要的方法有递归特征消除法。嵌入法：先使用数据训练机器学习模型，得到各个特征的重要程度系数，根据系数从大到小选择特征，主要的方法为基于惩罚项的特征选择法。

如上述步骤S120所述，依据所述初始数据得到局部异常度和整体健康度。

在本发明一实施例中，可以结合下列描述进一步说明步骤S120所述“依据所述初始数据得到局部异常度和整体健康度”的具体过程。

如下列步骤所述，依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度；依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度。

在本发明一实施例中，可以结合下列描述进一步说明步骤所述“依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度”的具体过程。

如下列步骤所述，依据所述初始水轮发电机数据和初始负荷数据和预设单测点模型得到目标状态量；依据所述目标状态量和标准状态量进行比对得到所述局部异常度。所述标准状态量为预先设置。

在一具体实施例中，开发基于单测点输出的异常检测和预测模型，基于设备重要测点及其相关测点数据，分别利用基于不同原理的智能算法开发多个单测点预测模型以预测单测点的健康状态。对数据挖掘领域应用较多的提升树算法与自编码器算法进行改进，使其适用于水电设备信息挖掘与故障预警。研发包括但不限于可自动调参的 LightBGM和XGBoost算法、增强学习的改进Autoencoder算法等，分别利用所研究的智能算法开发多个单测点预测模型。预测模型应构建正常状态下测点分布规律，并能预测单测点的状态，根据预测状态与实际状态进行单个测点异常的预测。预测测点为水轮发电机组上导轴承系统、推导轴承系统和水导轴承系统设备的常用测点。

水电机组单测点的设备健康预警预测分析所使用的主要分析手段是采用预测状态量与标准状态量的偏差进行评估分析，同时通过偏差度建立异常风险的评估概率，此方法是较好的一种从单变量角度去思考偏差的方法。然而，目前对于偏离程度所导致的风险分布情况尚不明确，且在不同工况下，尤其是C级维修后，测点的偏离程度存在差异。为了进一步提升建模的有效性，根据水电机组的经验和数据规律，提出了在此基础上增加正常区间估计模型：使用统计模型对不同工况下的健康区间进行估计，以了解偏离度对设备缺陷风险和劣化风险的影响。

对于单测点状态量的风险预警预测分析主要根据输入相关状态量，利用机器学习或者深度学习算法预测目标状态量的输出，根据目标状态量的输出与真实值进行对比确定健康偏离度。针对水电机组相关特征数据的整理，确定输入特征和目标特征，分别利用基于不同原理的智能算法开发多个单测点预测模型以预测单测点的健康状态。对数据挖掘领域应用较多的提升树算法与自编码器算法进行改进，使其适用于水电机组设备信息挖掘与故障预警。研发包括但不限于可自动调参的 LightBGM和XGBoost算法、增强学习的改进Autoencoder算法等。具体的实现过程如下：

模型输入：

目标变量：上导轴承8#瓦温；

解释变量：有功功率，水头，上导轴承油位，上导轴承油温，上导轴承X向摆度，上导轴承Y向摆度；

训练数据：原始数据前80%，约630000条；

测试数据：原始数据后20%，约160000条。

模型训练：对于单测点的预警预测模型而言，根据不同模型原理特点，分别进行建模，对于目前的模型我们主要采用 LightBGM算法、XGBoost算法、Autoencoder算法进行对比分析，下面对三类算法原理的优缺点进行综合分析。综合分析如下：

算法训练原理：

LightBGM算法训练原理：LightBGM以决策树为基础，是GBDT算法的一个实现，支持高效率的并行训练，并且具有更快的训练速度、更低的内存消耗、更好的准确率、支持分布式可以快速处理海量数据等优点。在每次的训练结果的基础上继续迭代下一个模型，提高正确率，最终达到最优的精度。GBDT在每一次迭代的时候，都需要遍历整个训练数据多次。如果把整个训练数据装进内存则会限制训练数据的大小；如果不装进内存，反复地读写训练数据又会消耗非常大量的时间。尤其面对工业级海量的数据，普通的GBDT算法占用内存较大、训练效率较低，LightGBM有效解决了GBDT在海量数据上遇到的问题，可以更好更快地用于电力实践。LightGBM使用控制量类型的状态量作为输入特征数据进行训练，最终输出上导瓦温目标预测状态量，通过目标预测状态量和目标真实状态量的偏差量来估计健康风险程度以及异常风险概率。

XGBoost算法训练原理：XGBoost是基于提升树的加法模型，是GBDT的一种，其基本思想和GBDT一样，但是增加了许多优化。其使用的每一个基分类器都是回归树，训练方式是采用前向分步算法逐步优化里面的每个基学习器。算法基本思想是：首先，对所有特征都按照特征值进行预排序；其次，在遍历分割点的时候用O(#data)的代价找到每一个特征上的最好分割点；最后，比对选择全局的最好分割点，将数据分裂成左右子节点。XGBoost仍然是使用控制量类型的特征量进行训练，通过预测上导瓦温与实际上导瓦温来估计健康偏差度以估计风险。

Autoencoder算法训练原理；Auto-Encoder自编码器算法，是一种无监督式学习模型。它基于反向传播算法与最优化方法，利用输入上导瓦温本身作为监督，来指导神经网络尝试学习一个映射关系，得到重构输出的上导瓦温预测值。在时间序列异常检测场景下，异常样本数量通常远小于正常样本数量，当使用自编码器重构出来的上导瓦温预测值跟原始输入的差异超出一定偏差时，即认为原始时间序列存在异常，进一步可以估计风险异常概率。整个原理如图11所示。整体的优化策略如图12所示。

在一具体实施例中，Auto-Encoder，XGBoost，LightGBM本质上都是一种回归预测方法，可以通过MAE，MSE，RMSE进行评价。

平均绝对误差MAE：；

均方误差MSE：；

均方根误差RMSE:。

综上所述：对不同算法原理下、不同数据分布下的单变量偏离度健康估计，最终分析结果如下图13所示。Anomaly scores为经过ARIMA拟合后利用孤立森林对残差点计算的异常得分值，分数分布图如图14所示；Standardized residual 展示了经过ARIMA模型拟合后的残差曲线；Histogram estimated density 中展示残差分布曲线；Noramal QQ图用于检查残差序列是否符合正态分布；Correlogram它展示了时间序列观测值与其滞后（lag）版本之间的相关性。

在本发明一实施例中，可以结合下列描述进一步说明步骤所述“依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度”的具体过程。

如下列步骤所述，依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据输入预设多测点模型得到预测健康度；依据所述预测健康度和标准健康度进行比对得到整体健康度。所述标准健康度是预先规定的值。

在一具体实施例中，开发基于多测点输出的设备整体健康度预测模型。

基于平台中设备对象上的多维度监测数据，结合人工智能算法开发基于多种不同原理的设备/部件整体健康度预测模型，以综合评估设备状态的整体健康度。对设备信息挖掘与故障诊断中常用的SBM算法、超球算法、LSTM算法等进行优化，研发适用于水电领域的智能预警算法，结合所研究的智能算法开发设备/部件整体健康度预测模型，基于设备整体健康度模型进行设备的综合监测和整体预测。整体健康度预测模型应具备故障测点自诊断功能，在发生故障预测时自动对异常测点进行提取并基于故障风险进行排序，指导运维人员更快地定位故障、排除故障。设备对象包含水轮发电机组的上导轴承系统、推导轴承系统和水导轴承系统。其中每个多测点输出的设备整体健康度预测场景均需开发不少于文中所述的三种算法模型。

作为一种示例，对于多测点的预警预测分析，完成设备部件或者设备整体的健康评估，通过预测未来健康度与实际健康度之间的偏差来确定风险概率。首先，对历史样本空间中与部件或者整体设备健康相关的特征进行相关性分析，据此选择合适特征并确定其权重，生成历史样本空间的设备状态健康值；接着，输入其他特征量和健康评价值，训练预测模型；最终，计算预测健康值和实际健康值的偏离度，根据预警策略实现预警功能。整体如下几个关键步骤：

数据输入：

目标变量：上导轴承8#瓦温；

解释变量：上导轴承相邻瓦温、有功功率、水头、上导轴承油位、上导轴承油温、上导轴承X向摆度、上导轴承Y向摆度、冷却水流量等30个变量；

训练数据：原始数据前80%，约810000条；

测试数据：原始数据后20%，约200000条。

模型训练：对于多测点的设备健康预警预测算法，通过SBM算法、超球算法、LSTM算法等算法进行综合验证，每个算法有自己独特原理，同时有不同的适用场景，对于水电机组健康管理而言，在不同工况下，特征数据的波动较大，不确定性和模糊性较强。所以应采用不同算法进行验证，发挥出大数据价值。整个核心算法流程如图15所示。

在一具体实施例中，基于SBM算法优化训练原理如下：多测点预警采用基于SBM的优化算法，主要包含以下几点：①针对SBM相似性表征能力差的问题，优化的SBM算法基于权重最大和贝叶斯优化的原则，动态实现保留最相似的向量，并去除剩余的向量，可以有效地删除噪音数据的影响；②针对SBM测量信号的相似性冗余的问题，优化的SBM算法基于自动关联的原则，将实际的向量与状态矩阵中的每个向量进行比较，状态矩阵仅为独立的标签分配一个相似性值，消除了相似性信号的溢出效应，提高了算法的鲁棒性；③针对SBM对标签缺失值难以处理的问题，优化的SBM算法基于缺失值判定原则，将缺失值进行有效的区别，动态实现对状态矩阵的选择；④针对SBM只能使用静态状态矩阵的问题，优化的SBM算法基于自学习规则，动态的实现状态矩阵的删减，有效的保留价值最大的数据作为状态矩阵。针对改进的SBM算法，最终形成工业级稳定的智能预警VBM算法。整个判断流程图如图16所示。

为了尽量减少错报，误报，漏报等情况，实现工业级稳定性强的智能预警算法，整体的智能预警策略如下：

滑动窗口比例（WindowRatio）

|  |  |
| --- | --- |
| 名称 | WindowRatio |
| 输入 | 一个测点一连串时序数值对应的残差（如：最近18个残差）；窗口大小，残差阈值，正常点个数阈值 |
| 输出 | true 或 false |
| 功能/算法 | 此函数查看数据窗口并将其与阈值进行比较。如果正常点（残差低于残差阈值）个数小于阈值，则windowRatio将返回true，否则将返回false； |

平滑残差 (Smooth Residual)

|  |  |
| --- | --- |
| 名称 | Smooth Residual |
| 输入 | 一个测点一连串时序数值对应的残差（默认最近6个残差） |
| 输出 | 一个测点当前数值的平滑残差 |
| 功能/算法 | Residual=Actual-Estimated；  以当前测点残差在内的前6个残差为计算样本，去除最高值、最低值后，计算余下的4个残差的平均值作为Smooth Residual的输出值 |

残差高 (H)

|  |  |
| --- | --- |
| 名称 | 残差高 |
| 输入 | 一个测点一连串的时序数值对应的残差（如：最近18个残差） |
| 输出 | True 或 false；High fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | (WindowRatio(16, Tag.ResidualIndicationPositive(18),"=",1,9)) And (Tag.Residual>= Tag.ResidualPositiveThreshold)； |
| 备注 | 残差=观察值-预测值 |

残差低 (L)

|  |  |
| --- | --- |
| 名称 | 残差低 |
| 输入 | 一个测点一连串的时序数值对应的残差（如：最近18个残差） |
| 输出 | True 或 false；Low fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | (WindowRatio(16, Tag.ResidualIndicationNegative(18),"=",1,9)) And (Tag.Residual>= Tag.ResidualPositiveThreshold)； |
| 备注 | 残差=观察值-预测值 |

向上阶跃 (Step High)

|  |  |
| --- | --- |
| 名称 | Step High |
| 输入 | 一个测点一连串（默认7个）的时序数值及对应的残差；正向偏差；负向偏差；  A set of tag history containing at least 4 observations E.g. Tag.Actual(7) |
| 输出 | True 或 false；High fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | 如果提供的数值发生向上阶跃变化，则此函数返回true。  将对提供的数值（不包括最近的观测值）进行奥林匹克平滑处理，以获得最近观测值之前的平均值。将最近的两个观测值与评估差异的平均值进行比较。  当观测值偏差（当前观测值-当前观测值之前的平均值）大于2.5倍正向偏差（2.5\* Tag.ResidualPositiveThreshold）且当前残差（观察值-预测值）须大于1.5倍正向偏差（1.5\* Tag.ResidualPositiveThreshold），函数返回true，报警信息：“Step High fired for [Source Tag] - [Description] on [Asset]” |
| 备注 | value – averageValue > |stepUpThreshold| |

向下阶跃 (Step Low)

|  |  |
| --- | --- |
| 名称 | Step Low |
| 输入 | 一个测点一连串（默认7个）的时序数值及对应的残差；正向偏差；负向偏差；  A set of tag history containing at least 4 observations E.g. Tag.Actual(7) |
| 输出 | true 或 false；Low fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | 如果提供的数值发生向下阶跃变化，则此函数返回true。  将对提供的数值（不包括最近的观测值）进行奥林匹克平滑处理，以获得最近观测值之前的平均值。将最近的两个观测值与评估差异的平均值进行比较。  当观测值偏差（当前观测值-当前观测值之前的平均值）小于2.5倍负向偏差（2.5\* Tag.ResidualNegativeThreshold）且当前残差（观察值-预测值）须小于1.5倍负向偏差（1.5\* Tag.ResidualNegativeThreshold），函数返回true，报警信息：“Step Low fired for [Source Tag] - [Description] on [Asset]” |
| 备注 | value – averageValue < -|stepDownThreshold| |

长时间高 (LH)

|  |  |
| --- | --- |
| 名称 | 残差长时间高 |
| 输入 | 一个测点一连串的时序数值及对应的残差；正向偏差；最近60个残差， |
| 输出 | true 或 false；Low fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | 一个测点一连串的时序数值对应的残差，当最近60个测点的残差有55个满足要求，且当前残差（观察值-预测值）须大于1倍正向偏差（1\* Tag.ResidualPositiveThreshold），函数返回true，报警信息：“Long High fired for [Source Tag] - [Description] on [Asset]” |

长时间低 (LL)

|  |  |
| --- | --- |
| 名称 | 残差长时间低 |
| 输入 | 一个测点一连串的时序数值及对应的残差；负向偏差；最近60个残差， |
| 输出 | true 或 false；Low fired for [Source Tag] - [Description] on [Asset] |
| 功能/算法 | 一个测点一连串的时序数值对应的残差，当最近60个测点的残差有55个满足要求，且当前残差（观察值-预测值）须小于1倍负向偏差（1\* Tag.ResidualNegativeThreshold），函数返回true，报警信息：“Long Low fired for [Source Tag] - [Description] on [Asset]” |

在一具体实施例中，超球算法训练原理：在仅有水电机组正常运行数据的样本情况下， OneClassSVM能通过对多维特征的融合实现水电机组设备的健康状态评估，但特征向量空间分布的复杂程度会直接影响OneClassSVM的效果。为此基于超球优化支持向量的水轮机设备异常检测方法，通过超球优化改善特征向量的空间分布以降低数据描述任务的难度，进而使得超球优化能更有效地识别水电机组状态异常。在不同工况、不同测点，超球优化算法效果更好。整体的分析过程如图17所示。

在一具体实施例中，LSTM算法训练原理：水电机组设备健康往往通过报警阈值进行报警，对于告警后的状态，相关评价导则有相关风险结论，但是对于设备早期的隐患，没有告警或者接近告警值时的健康状况难以及时捕捉分析。基于统计的传统检测方法在解决时间序列异常检测上存在很大挑战，因此提出基于LSTM时间序列重建的方法进行水电机组设备健康风险异常检测。该算法首先引入一层LSTM网络对多测点数据的时间序列进行向量表示，采用另一层LSTM网络对时间序列进行逆序重建，然后利用重建值与实际值之间的误差，通过极大似然估计方法对该时段序列进行异常概率估计，最终通过学习异常报警阈值实现时间序列异常检测。算法过程如下图18所示。

综合以上分析，通过以上综合模型分析，主要通过重构偏离度来估计设备健康趋势。

将单测点预警预测模型的结果、多测点预警预测模型的结果和整体健康度输入集成模型，结合多个模型的学习能力，形成最终的集成预测结果，得到整个水电机组健康状态。开发集成预测模型能有效降低单模型预测方差，提升预测模型准确性和可靠性。模型集成决策所研究的算法模型包括但不限于投票机制、权重融合、层次化方法、特征融合和模型堆叠、基于密度的融合等方法。

作为一种示例，基于水电站实时多维度传感器时序数据，结合人工智能算法开发针对水轮发电机组上导轴承、推导轴承、水导轴承的健康状态和性能监测的预测应用。该预测应用可在大量正常样本的基础上进行建模，获取设备的实时健康度，对上导、推导、水导轴承系统的早期异常进行识别和实现分级预测。涉及到的相关模型算子经过拆分后分别封装部署在工业互联网平台。开发基于部件显示的模型数据统计仪表盘，能实时展示检测数据和模型预测结果，便于运维人员查看每个系统部件和测点的问题严重程度及偏离度，以达到对机组运行状态的整体监测和设备异常及时处理的目标。

在一具体实施例中，对于水电机组设备健康状态预警预测分析而言，主要采用单测点偏离度和多测点偏离度进行异常风险估计，多测点偏离度更偏向于设备部件或设备整体。从这个角度出发，对于不同偏离度和异常发生概率之间的非线性映射，一方面，利用预测值与实际值之间的偏离度，通过极大似然估计方法进行异常概率估计；另一方面，通过预测值与真实值之间概率分布关系进行异常估计；最终根据实际情况进行验证确认。

发现异常后，使用概率图模型对可能造成的隐患进行挖掘。在一具体实施例中，异常分析、隐患预测整体逻辑思路如图19所示。

当多个测点组合已经分析出在T时刻存在异常偏离度，表明存在潜在异常风险。考虑到异常可能会在未来T+1时刻继续传播，特征量之间会产生依赖关系，导致设备出现隐患，此时需要对未来设备的隐患状态趋势进行预测。整个模型的逻辑原理如下图20所示。

贝叶斯网络隐患预测分析，原始数据：导叶开度、转速、上导瓦温、上导油温、轴承冷却水流量、油槽油混水信号、上导X向摆度、上导Y向摆度、油槽油位、其他。

数据探索分析：通过对原始测点数据进行EDA探索分析，探索分析主要结合设备缺陷数据，业务经验，以及设备状态评价导则进行综合分析，对测点数据进行离散化处理。

最终得到模型输入数据，模型主要采用概率图模型、贝叶斯网络对数据特征数据进行处理。

模型输入数据：

上导瓦温℃: (45~49)、(50~60)、(60~70)、(72~85)

上导X向摆度μm: (100~190)、(200~300)、(320~400)

导叶开度：全关、空转、空载、全开；

转速：0%、25%、50%、95% 、100%；

信号量：油循环油泵启动信号、冷却水示流信号。

贝叶斯网络模型业务解释性，如图21所示。

整个贝叶斯网络模型本身是对现在设备评价导则的升级，消除了单一状态量风险的估计，结合多个装下组合下的隐患估计，符合运维运行思维，因此模型本身具有较强的解释性，同时结合数据趋势分析、劣化趋势分析角度完成数据层面的解释，最终为推理过程符合业务场景进行业务解释，此是其他算法难以比拟的地方。

机组设备隐患预测优化策略，对于贝叶斯网络的优势非常明显，主要集中在如下几点：

不依赖于负样本：在早期样本缺少情况下，专家经验少量参与，对特征量进行离散化。概率软距离判断隐患：相对于偏离度等算法，利用特征之间的概率依赖来推断隐患发生概率，适合不确定性和模糊性建模。业务解释性强：由于模型基于特征之间的相互依赖性建模，符合运维经验，具有解释性。自动更新学习：通过内存计算模式，不断更新特征划分区间，进行模型自学习更新。

贝叶斯网络的优化策略，针对贝叶斯网络的优化，主要是计算问题，早期的贝叶斯网络一直都是解决设备类故障诊断或者预测的优秀方法之一，与神经网络、专家知识库并列，但由于特征量的各种组合带来的组合爆炸问题，导致贝叶斯网络的应用受到限制。基于此情况，存在以下优化点：①利用业务规则对无效特征进行过滤后，再使用模型进行训练；②对强相关的特征进行加权弱化处理，对特征依赖之间的推理更加的平滑，而不是过度依赖少数强相关特征。

在一具体实施例中，根据业务规则搭建测点关系图，作为贝叶斯网络的数据逻辑结构图；将测点数据及潜在隐患进行离散化处理后，输入给贝叶斯模型进行训练；最后，针对输入的测点数据，进行离散化后，使用贝叶斯模型可以预测出可能出现的隐患及其发生概率。

在一具体实施例中，本发明通过基于大数据+AI的方式研究实现了一种水电机组设备的健康管理体系，明显的有益效果如下：通过水电机组设备健康预警预测模型体系、设备测点预警预测模型、劣化趋势模型、数据趋势分析模型，构建了完整的基于大数据的水电机组健康风险管理体系，提高水电机组早期缺陷劣化方向的精准度，提高水电机组健康管理智能化水平；

水电机组的设备健康预警预测体系，利用人工智能结合大数据分析，真正做到挖掘数据规律，依据数据推断设备风险状态隐患的概率，不仅做到了事前防范，还为水电机组基于大数据运行运维管理标准提供指导支撑，形成企业标准；

水电机组预警预测分析体系，核心计算任务系统自动执行，通过采用数据解释性和模型解释性技术，结合领域专家的少量经验，有效减少因人工经验判断带来的偏差，提高趋势判断的准确性，能在一个页面完成对分析结果的可视化复盘，减少查询验证工作量，真正提高了数字化支撑作用，提升了整体运维效率。

通过建立正确的水电机组设备的预警预测分析体系准确找到设备劣化隐患方向，根据历史项目经验，通过台账数据统计，早期隐患发现率提高40%，节约消缺时间大约30%，大致节约10-20天的隐患处理时间，减少了20%~30%检修及运维费用支出；

通过实施项目的经验推断，早期水电机组的预警预测分析全过程可视化复盘，减少48%左右、大致2-5天的查询验证时间，提高运维效率；

根据设备健康学验证，通过建立早期的设备预警预测体系，能有效的提高设备的使用寿命，保障设备安全运行，根据历史项目经验推算，设备使用寿命提高20%-40%。

对于系统实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

参照图2，示出了本申请一实施例提供的一种基于水电机组的风险预测的系统，具体包括如下模块，

获取模块210，用于获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；

隐患概率模块220，用于依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

在本发明一实施例中，所述获取模块210包括：

初始数据模块，用于依据所述运行数据进行处理得到初始数据；其中，所述初始数据包括初始水轮发电机数据，初始负荷数据以及初始环境温度数据；

偏离度模块，用于依据所述初始数据得到局部异常度和整体健康度；

在本发明一实施例中，所述偏离度模块220包括：

局部异常度子模块：用于依据所述初始水轮发电机数据和初始负荷数据进行单测点分析得到局部异常度；

整体健康度子模块：用于依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据进行多测点分析得到整体健康度。

在本发明一实施例中，所述局部异常度子模块包括：

目标状态量子模块：用于依据所述初始水轮发电机数据和初始负荷数据和预设单测点模型得到目标状态量；

局部异常度子模块：用于依据所述目标状态量和标准状态量进行比对得到所述局部异常度。

在本发明一实施例中，所述整体健康度子模块包括：

预测健康度子模块：用于依据所述初始水轮发电机数据，所述初始负荷数据以及所述初始环境温度数据输入预设多测点模型得到预测健康度；

整体健康度子模块：用于依据所述预测健康度和标准健康度进行比对得到整体健康度。

在本发明一实施例中，所述隐患概率模块230包括：

概率子模块：用于将所述局部异常度和所述整体健康度进行非线性映射计算得到所述水电机组风险隐患的概率。

在本发明一实施例中，所述概率子模块包括：

贝叶斯网络子模块：用于将离散化后的测点数据输入贝叶斯网络，计算得到所述水电机组风险隐患的概率。

需要说明的是，对于方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明实施例并不受所描述的动作顺序的限制，因为依据本发明实施例，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作并不一定是本发明实施例所必须的。

在本具体实施例与上述具体实施例中有重复的操作步骤，本具体实施例仅做简单描述，其余方案参考上述具体实施例描述即可。

对于系统实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

参照图3，示出了本申请的一种基于水电机组的风险预测方法的计算机设备，具体可以包括如下：

上述计算机设备12以通用计算设备的形式表现，计算机设备12的组件可以包括但不限于：一个或者多个处理器或者处理单元16，内存28，连接不同系统组件（包括内存28和处理单元16）的总线18。

总线18表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构（ISA）总线，微通道体系结构（MAC）总线，增强型ISA总线、音视频电子标准协会（VESA）局域总线以及外围组件互连（PCI）总线。

计算机设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备12访问的可用介质，包括易失性和非易失性介质，可移动的和不可移动的介质。

内存28可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器30和/或高速缓存存储器32。计算机设备12可以进一步包括其他移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统34可以用于读写不可移动的、非易失性磁介质（通常称为“硬盘驱动器”）。尽管图3中未示出，可以提供用于对可移动非易失性磁盘（如“软盘”）读写的磁盘驱动器，以及对可移动非易失性光盘（例如CD-ROM，DVD-ROM或者其他光介质）读写的光盘驱动器。在这些情况下，每个驱动器可以通过一个或者多个数据介质界面与总线18相连。存储器可以包括至少一个程序产品，该程序产品具有一组（例如至少一个）程序模块42，这些程序模块42被配置以执行本申请各实施例的功能。

具有一组（至少一个）程序模块42的程序/实用工具40，可以存储在例如存储器中，这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其他程序模块42以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本申请所描述的实施例中的功能和/或方法。

计算机设备12也可以与一个或多个外部设备14（例如键盘、指向设备、显示器24、摄像头等）通信，还可与一个或者多个使得操作人员能与该计算机设备12交互的设备通信，和/或与使得该计算机设备12能与一个或多个其他计算设备进行通信的任何设备（例如网卡，调制解调器等等）通信。这种通信可以通过I/O接口22进行。并且，计算机设备12还可以通过网络适配器20与一个或者多个网络（例如局域网（LAN）），广域网（WAN）和/或公共网络（例如因特网）通信。如图3所示，网络适配器20通过总线18与计算机设备12的其他模块通信。应当明白，尽管图3中未示出，可以结合计算机设备12使用其他硬件和/或软件模块，包括但不限于：微代码、设备驱动器、冗余处理单元16、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统34等。

处理单元16通过运行存储在内存28中的程序，执行各种功能应用以及数据处理，例如实现本申请实施例所提供的一种基于水电机组的风险预测方法。

也即，上述处理单元16执行上述程序时实现：获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

在本申请实施例中，本申请还提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如本申请所有实施例提供的一种基于水电机组的风险预测方法。

也即，给程序被处理器执行时实现：获取水电机组的运行数据，并依据所述运行数据进行处理得到局部异常度和整体健康度；依据所述局部异常度和所述整体健康度得到水电机组风险隐患的概率。

可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是但不限于电、磁、光、电磁、红外线或半导体的系统、系统或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦可编程只读存储器（EPROM或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、系统或者器件使用或者与其结合使用。

计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、系统或者器件使用或者与其结合使用的程序。

可以以一种或多种程序设计语言或其组合来编写用于执行本申请操作的计算机程序代码，上述程序设计语言包括面向对象的程序设计语言（诸如Java、Smalltalk、C++）和常规的过程式程序设计语言（诸如“C”语言或类似的程序设计语言）。程序代码可以完全地在操作人员计算机上执行、部分地在操作人员计算机上执行、作为一个独立的软件包执行、部分在操作人员计算机上部分在远程计算机上执行或者完全在远程计算机或者服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（LAN）或广域网（WAN）——连接到操作人员计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

尽管已描述了本申请实施例的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例做出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

以上对本申请所提供的一种基于水电机组的风险预测方法及系统，进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，本说明书内容不应理解为对本申请的限制。

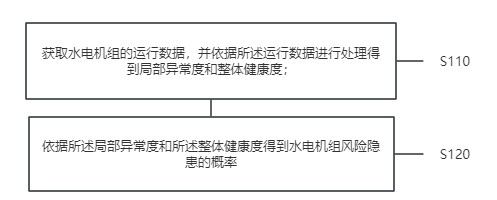


图1

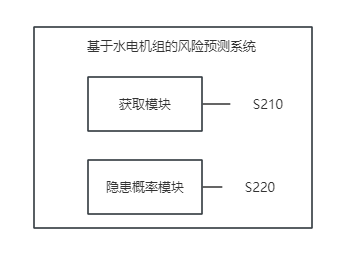


图2

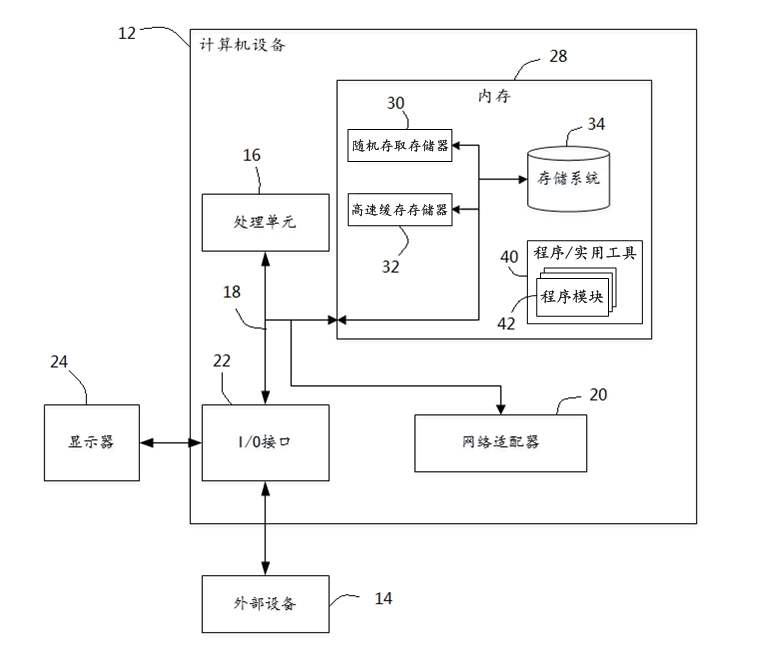


图3

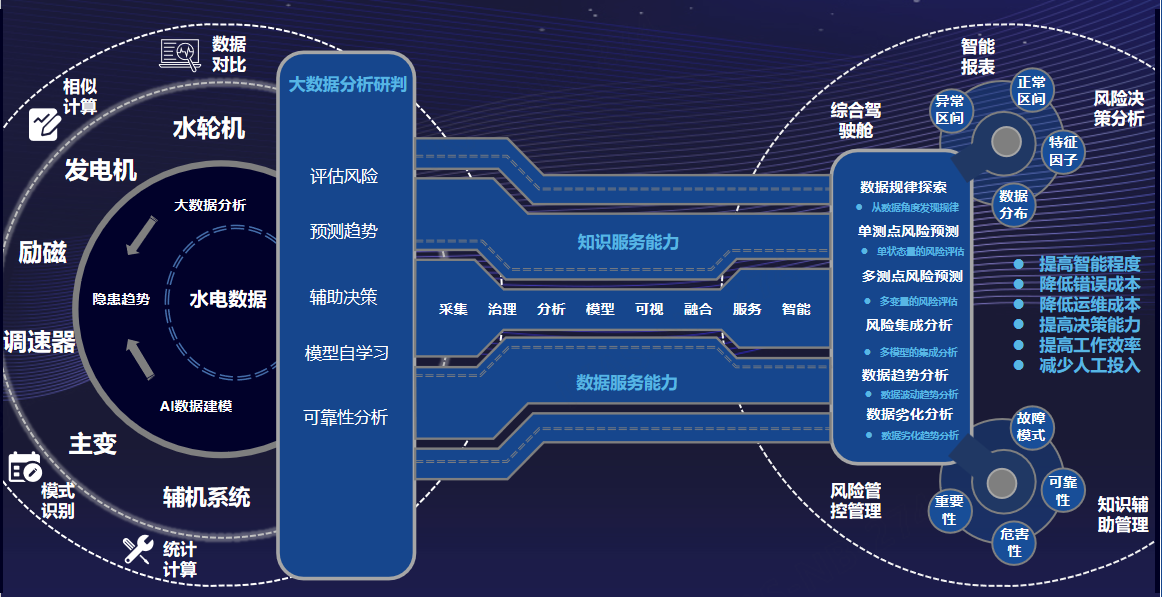


图4

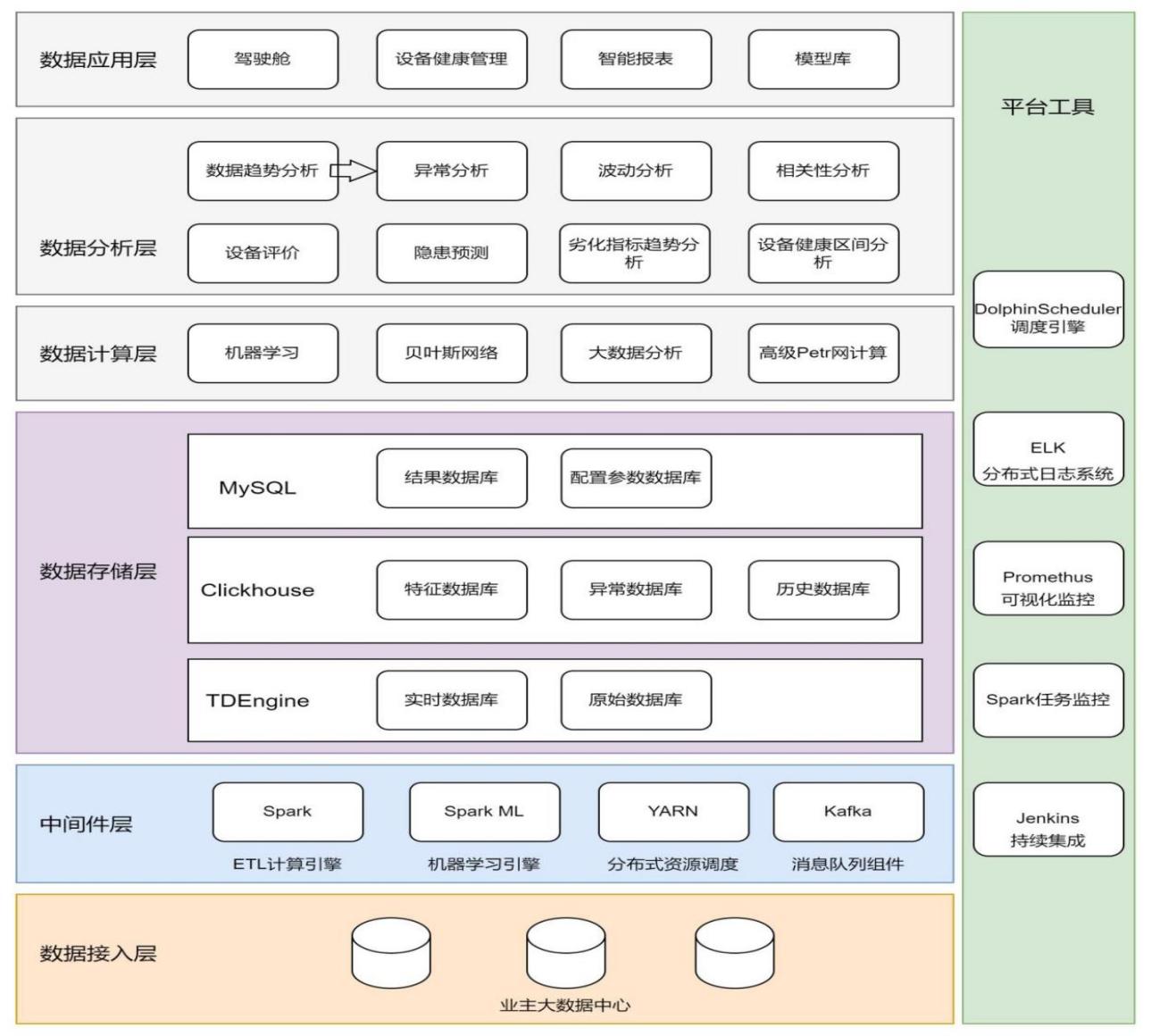


图5

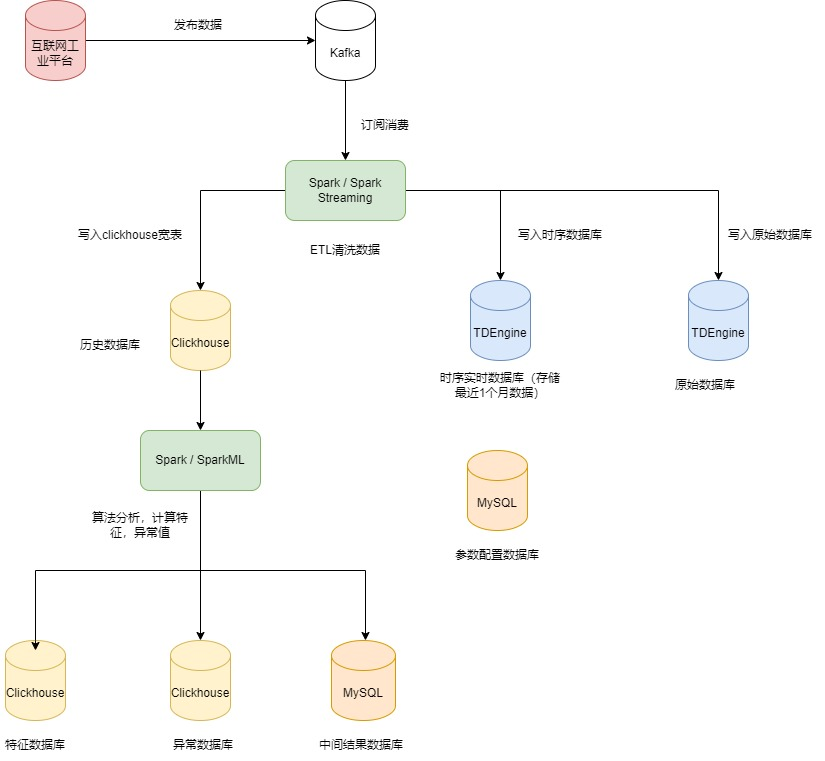


图6

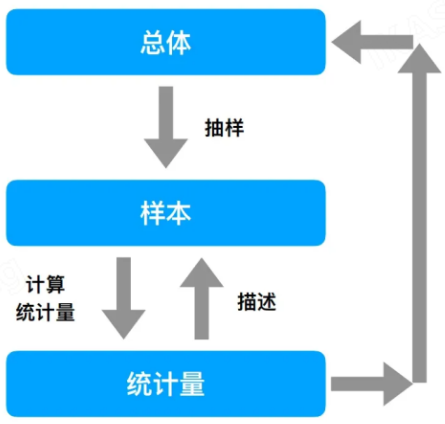


图7

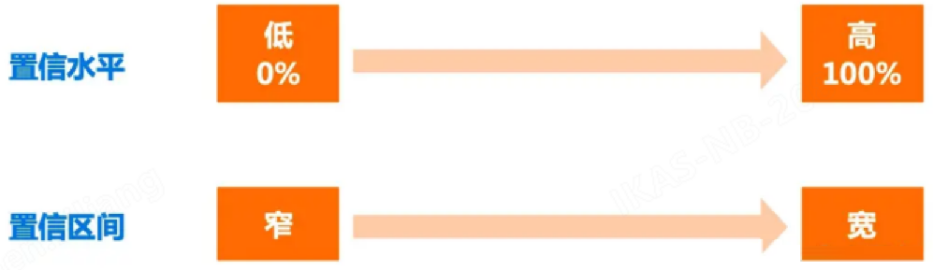


图8

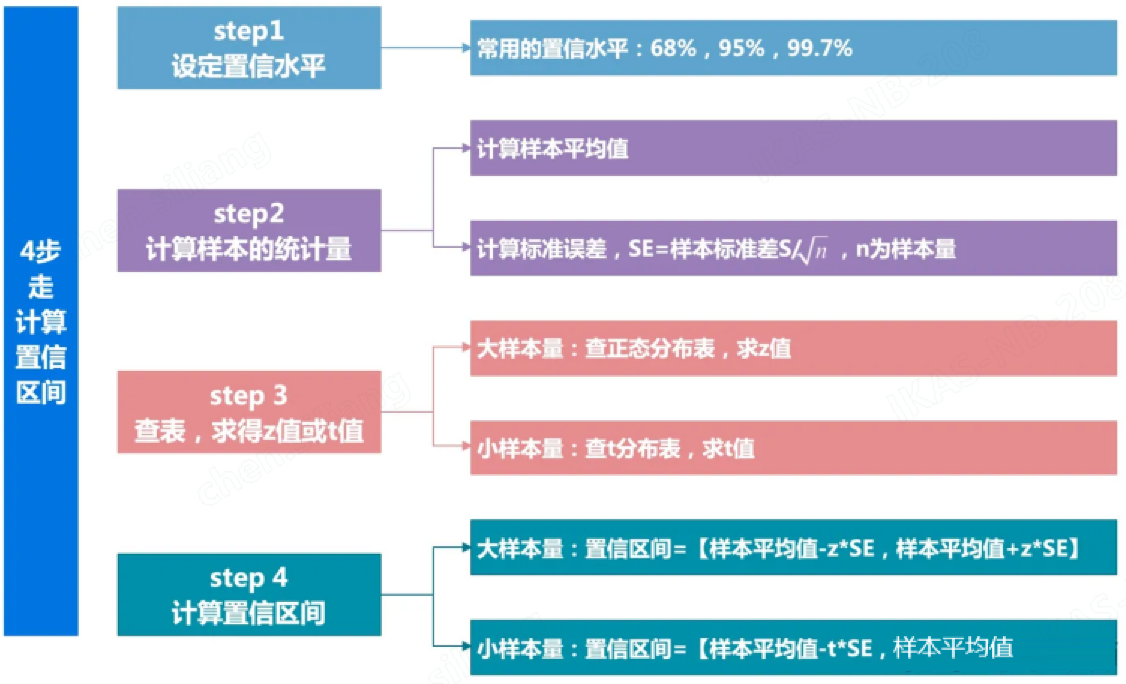


图9

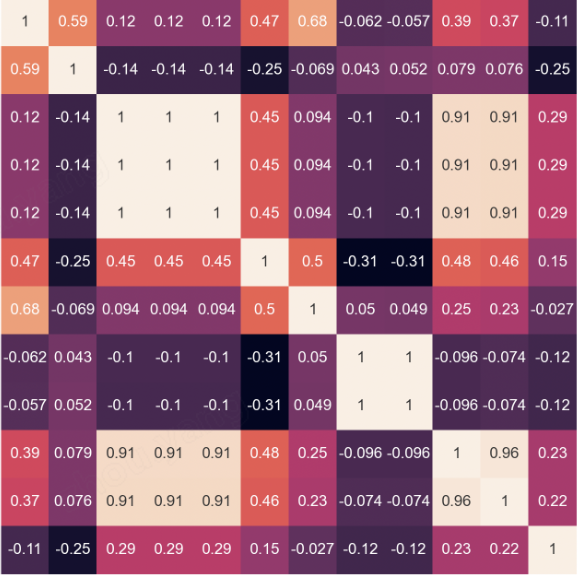


图10

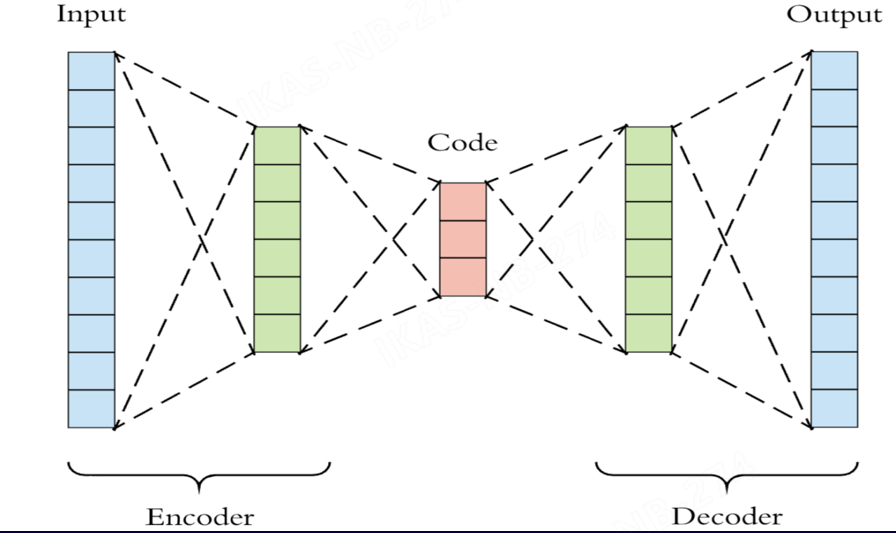


图11



图12

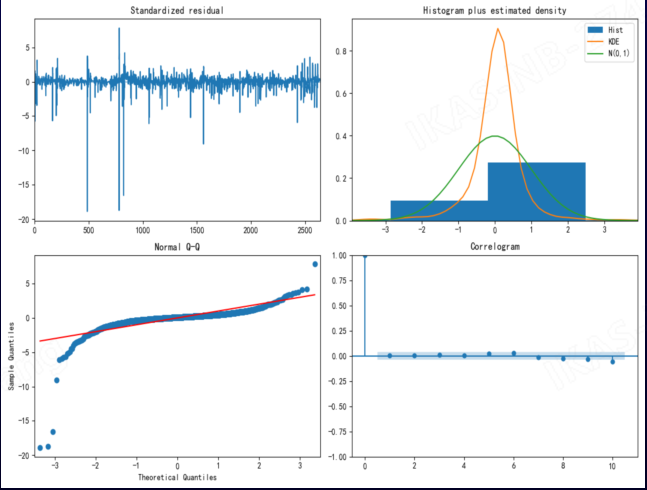


图13



图14

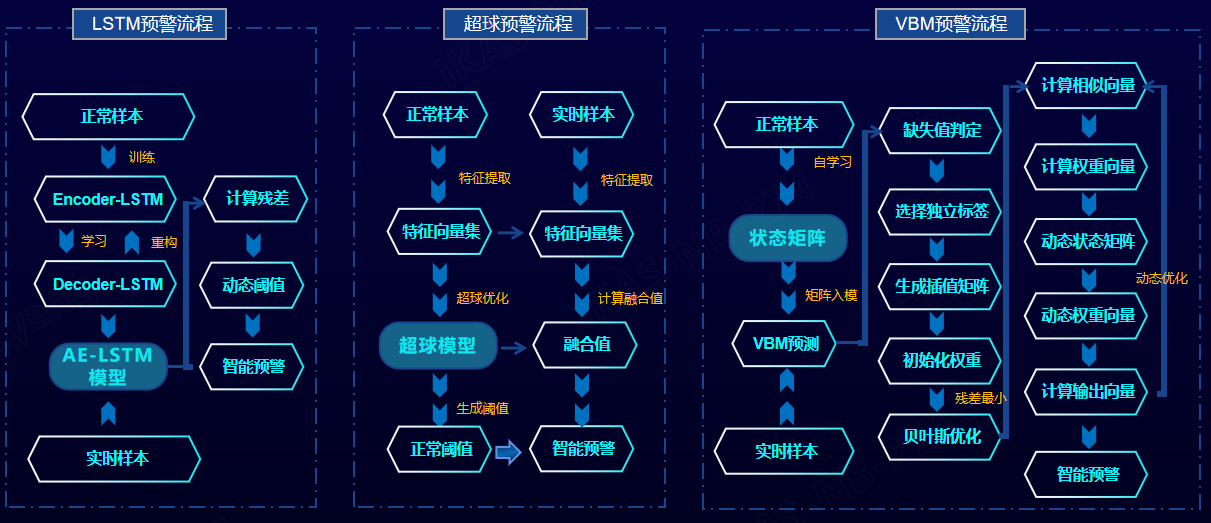


图15

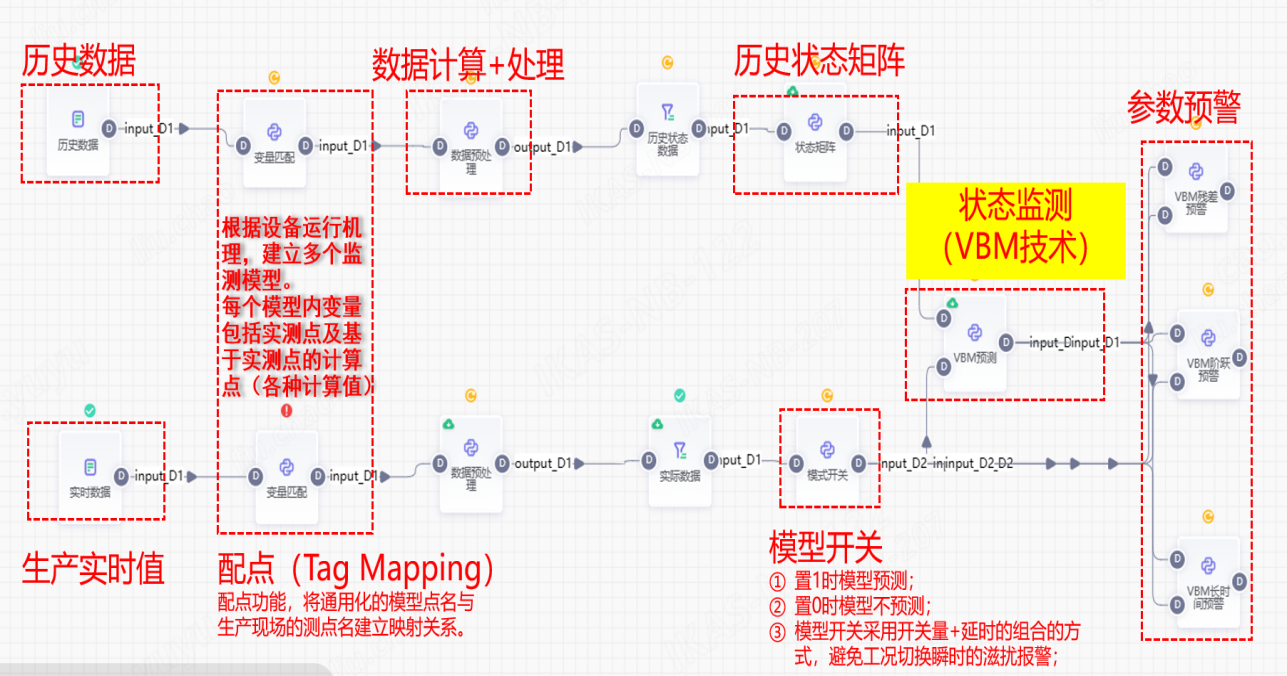


图16

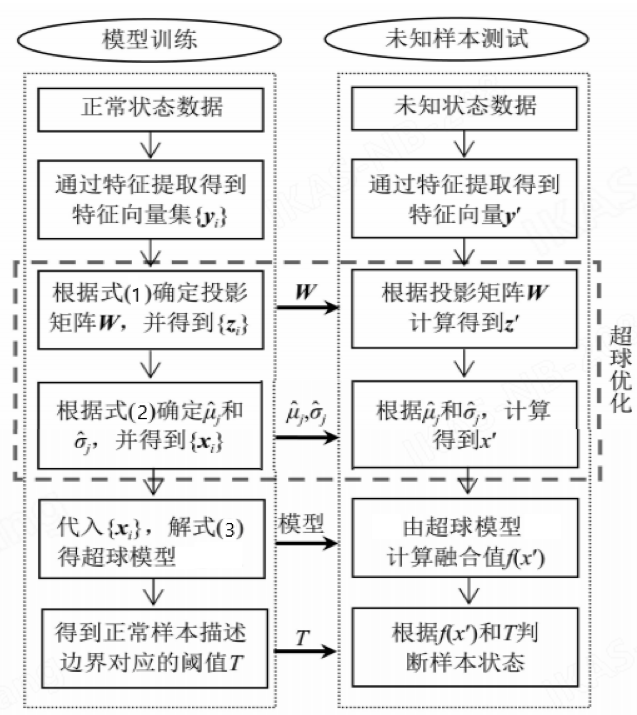


图17

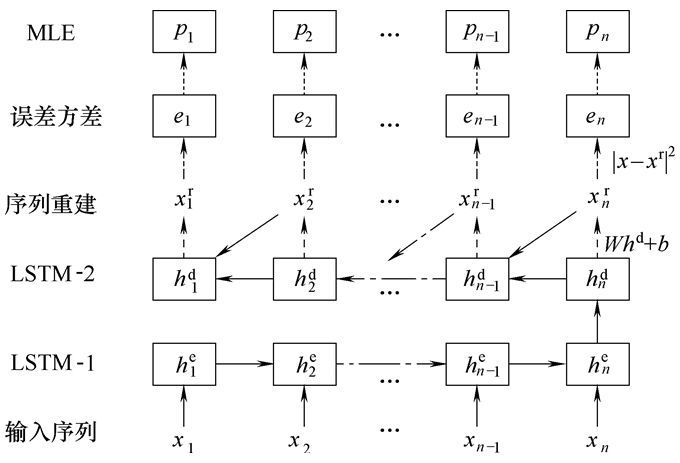


图18

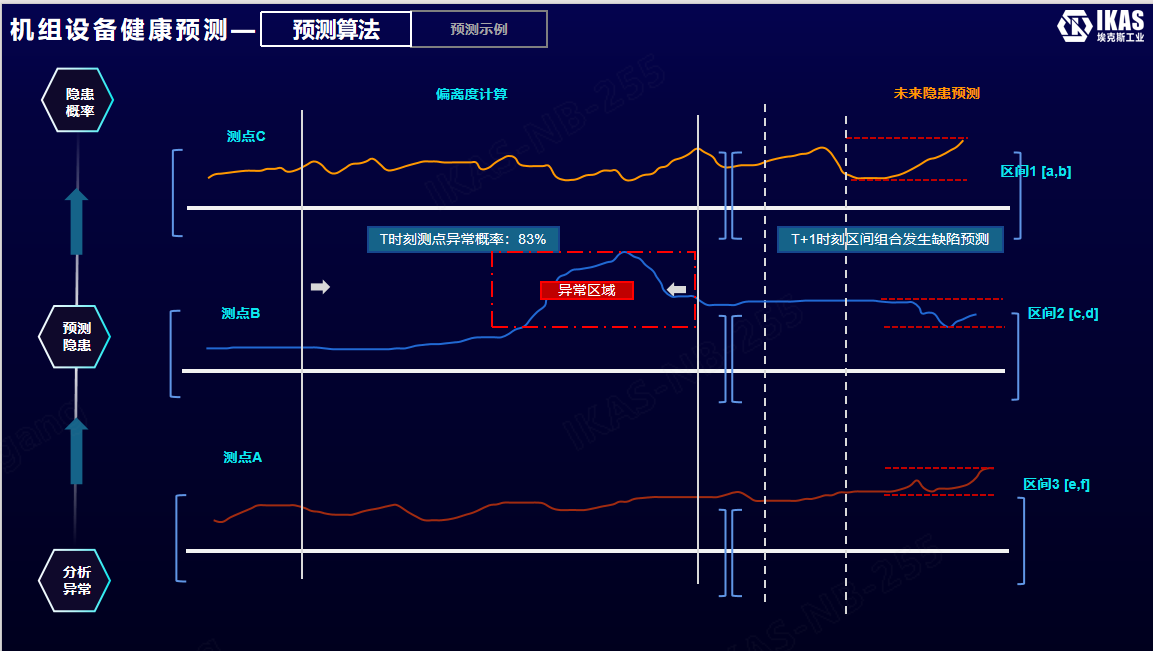


图19

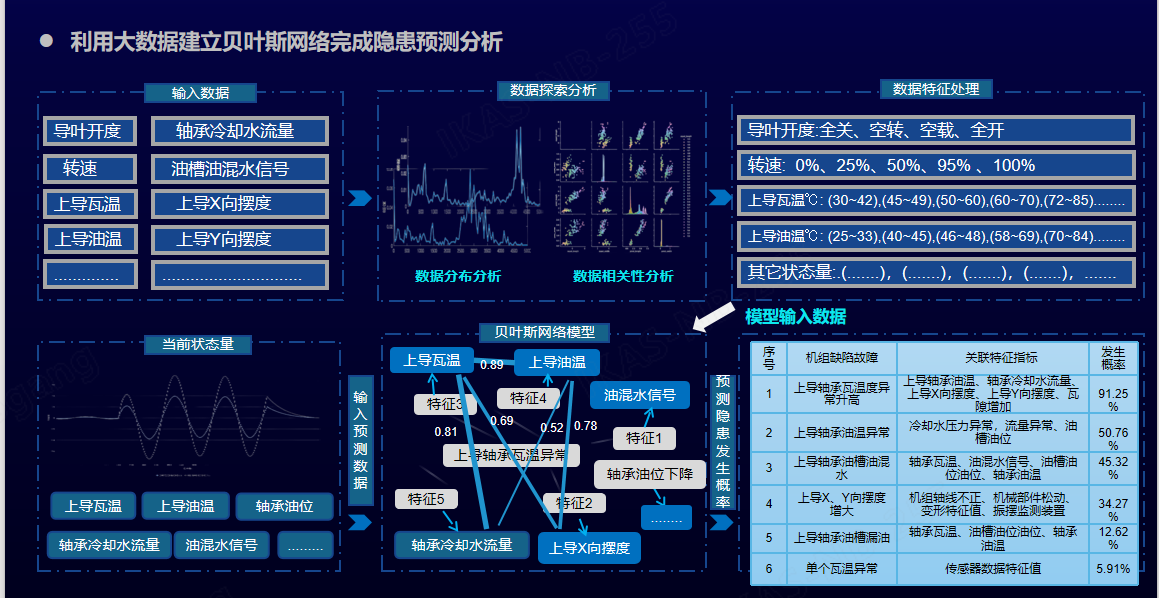


图20



图21