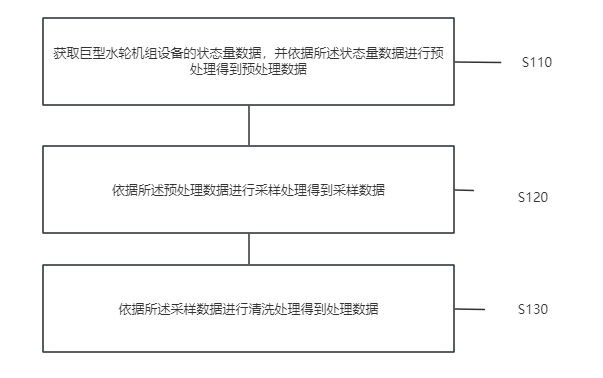
在本申请的实施例中，获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；依据所述预处理数据进行采样处理得到采样数据；依据所述采样数据进行清洗处理得到处理数据。本申请通过运用新研发的数据采样方法，对各种正常运行工况下的正样本数据进行采样，构建了正样本空间，便于查看数据趋势和对运行时状态进行偏离度计算。采样结果能符合原始数据趋势，时间序列满足概率统计上的无偏估计，能提高预警预测算法相关算法运行效率和结果精度。本申请通过运用新研发的数据处理方法，能够自适应地选择合适算法，实现自动去重功能，实现延迟数据、丢失数据、无效数据的填充，实现特征数据的去噪或光滑处理，实现对多维度的冗余信号进行降维处理。



1、一种基于巨型水轮机组的数据处理方法，其特征在于，所述方法用于对基于巨型水轮机组设备的状态量进行大数据处理，包括：

获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

依据所述预处理数据进行采样处理得到采样数据；

依据所述采样数据进行清洗处理得到处理数据。

2、根据权利要求1所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述状态量数据进行预处理得到预处理数据的步骤，包括：

依据所述状态量数据进行时序性排序得到排序数据；

依据所述排序数据进行同源压缩处理得到所述预处理数据。

3、根据权利要求1所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述预处理数据进行采样处理得到采样数据的步骤，包括：

依据所述预处理数据根据获取平台进行采集管理得到采集数据；

依据所述采集数据进行采样分析得到所述采样数据。

4、根据权利要求3所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述采集数据进行采样分析得到所述采样数据的步骤，包括：

依据所述采集数据判断所述采集数据的数据性质；其中，所述数据性质包括时间性质、异常性质和不均匀性质；

依据所述数据性质匹配采样方法得到所述采样数据。

5、根据权利要求4所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述数据性质匹配采样方法得到所述采样数据的步骤，包括：

若所述数据性质为时间性质，则通过广义平滑加权采样得到所述采样数据；

若所述数据性质为异常性质，则通过空间采样得到所述采样数据；

若所述数据性质为不均匀性质，则通过不规则采样得到所述采样数据。

6、根据权利要求1所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述采样数据进行清洗处理得到处理数据的步骤，包括：

依据所述采样数据判断所述采样数据的数据类型；其中，所述数据类型包括离群数据、延迟数据、残缺数据以及无效数据；

依据所述数据类型进行清洗得到所述处理数据。

7、根据权利要求6所述的基于巨型水轮机组的数据处理方法，其特征在于，所述依据所述数据类型进行清洗得到所述处理数据的步骤，包括：

若所述数据类型为离群数据，则通过离群算法得到所述处理数据；

若所述数据类型为延迟数据，则通过加权平均算法得到所述处理数据；

若所述数据类型为残缺数据，则通过前序数据补正得到所述处理数据；

若所述数据类型为无效数据，则通过前序数据修正得到所述处理数据。

8、一种基于巨型水轮机组的数据处理系统，所述系统用于对基于巨型水轮机组设备的状态量进行大数据处理，其特征在于，包括：

预处理数据模块，用于获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

采样模块，用于依据所述预处理数据进行采样处理得到采样数据；

处理数据模块，用于依据所述采样数据进行清洗处理得到处理数据。

9、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7中任一项所述的基于巨型水轮机组的数据处理方法的步骤。

10、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7中任一项所述的基于巨型水轮机组的数据处理方法的步骤。

**一种基于巨型水轮机组的数据处理方法及系统**

**技术领域**

本发明涉及大数据领域，具体涉及一种基于巨型水轮机组的数据处理方法及系统。

**背景技术**

水轮机组是一套复杂的机电系统，具有故障原因复杂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，对水轮机组进行实时状态监测及设备健康管理就很有必要。水轮机组系统设备复杂，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。

针对电站运行过程中生产数据越来越多、历史数据库中存储的数据量急剧膨胀的情况，现有的数据处理方法在效率和效果上表现不佳，主要依靠专家经验进行处理和判断。

在此情况下，数据无法进行精准采样，存在采样偏差大、采样不够全面等缺点，对后期数据分析的准确度造成影响；同时，容易出现无效数据无法进行补全、删除、修正等问题。

**发明内容**

鉴于此，提出了本申请“一种基于巨型水轮机组的数据处理方法及系统”，以便解决或者至少部分解决所述问题，包括：

一种基于巨型水轮机组的数据处理方法，所述方法用于对基于巨型水轮机组设备的状态量进行大数据处理，包括：

获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

依据所述预处理数据进行采样处理得到采样数据；

依据所述采样数据进行清洗处理得到处理数据。

优选地，所述依据所述状态量数据进行预处理得到预处理数据的步骤，包括：

依据所述状态量数据进行时序性排序得到排序数据；

依据所述排序数据进行同源压缩处理得到所述预处理数据。

优选地，所述依据所述预处理数据进行采样处理得到采样数据的步骤，包括：

依据所述预处理数据根据获取平台进行采集管理得到采集数据；

依据所述采集数据进行采样分析得到所述采样数据。

优选地，所述依据所述采集数据进行采样分析得到所述采样数据的步骤，包括：

依据所述采集数据判断所述采集数据的数据性质；其中，所述数据性质包括时间性质、异常性质和不均匀性质；

依据所述数据性质匹配采样方法得到所述采样数据。

优选地，所述依据所述数据性质匹配采样方法得到所述采样数据的步骤，包括：

若所述数据性质为时间性质，则通过广义平滑加权采样得到所述采样数据；

若所述数据性质为异常性质，则通过空间采样得到所述采样数据；

若所述数据性质为不均匀性质，则通过不规则采样得到所述采样数据。

优选地，所述依据所述采样数据进行清洗处理得到处理数据的步骤，包括：

依据所述采样数据判断所述采样数据的数据类型；其中，所述数据类型包括离群数据、延迟数据、残缺数据以及无效数据；

依据所述数据类型进行清洗得到所述处理数据。

优选地，所述依据所述数据类型进行清洗得到所述处理数据的步骤，包括：

若所述数据类型为离群数据，则通过离群算法得到所述处理数据；

若所述数据类型为延迟数据，则通过加权平均算法得到所述处理数据；

若所述数据类型为残缺数据，则通过前序数据补正得到所述处理数据；

若所述数据类型为无效数据，则通过前序数据修正得到所述处理数据。

为实现本申请还包括一种基于巨型水轮机组的数据处理系统，所述系统用于对基于巨型水轮机组设备的状态量进行大数据处理，包括：

预处理数据模块，用于获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

采样模块，用于依据所述预处理数据进行采样处理得到采样数据；

处理数据模块，用于依据所述采样数据进行清洗处理得到处理数据。

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巨型水轮机组的数据处理方法的步骤。

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巨型水轮机组的数据处理方法的步骤。

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“数据无法进行精准采样，存在采样偏差大、采样不够全面等缺点，对后期数据分析的准确度造成影响；同时，容易出现无效数据无法进行补全、删除、修正等问题”，本申请提供了“通过均匀采样、分层随机采样、OCSVM采样、非均匀采样方法，实现在庞大数据中样本的精确采样，构建正样本空间。通过运用Kmeans离群算法、KNN离群算法、LOF局部离群因子算法、孤立森林算法、四分位数算法等算法，构建数据清洗模型，实现对离群数据、丢失数据、延迟数据、无效数据的清洗，还原设备运行数据的真实性”的解决方案，具体为：获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；依据所述预处理数据进行采样处理得到采样数据；依据所述采样数据进行清洗处理得到处理数据。本申请通过运用新研发的数据采样方法，对各种正常运行工况下的正样本数据进行采样，构建了正样本空间，便于查看数据趋势和对运行时状态进行偏离度计算。采样样本覆盖了所有工况下的状态量模式，采样结果能符合原始数据数据趋势，时间序列满足概率统计上的无偏估计，同时能提高预警预测算法相关算法运行效率和结果精度。本申请通过运用新研发的数据处理方法，能够自适应地选择合适算法，实现自动去重功能，实现延迟数据、丢失数据、无效数据的填充，实现对突变、离群、死点、噪声等数据的自动清洗；实现对含高频噪声的特征数据的去噪或光滑处理，避免造成误报，提升模型的准确性和鲁棒性；能够对多维度的冗余信号进行降维处理，消除冗余信息给预测结果带来的不确定性。

**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性的前提下，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巨型水轮机组的数据处理方法的步骤流程图；

图2是本申请一实施例提供的一种基于巨型水轮机组的数据处理系统的结构框图；

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巨型水轮机组的数据处理方法的数据源管理图；

图5是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的测点管理图；

图6是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的数据集管理图；

图7是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的采样示例图；

图8是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法广义平滑加权采样效果图；

图9是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的不规则采样图；

图10是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的样本标准差原理图；

图11是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的置信水平与置信区间关系图；

图12是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的置信区间估计；

图13是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的上导瓦温原始数据图；

图14是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法OCSVM采样的正常样本上导瓦温数据图；

图15是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的OCSVM采样后的样本图；

图16是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的切片采集数据取均值处理图；

图17是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法清洗模拟数据图；

图18是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法离群数据清洗图；

图19是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法延迟数据清洗图；

图20是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法残缺数据清洗图；

图21是本申请一实施例提供的一种基于巨型水轮机组的数据处理方法无效数据清洗图。

**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：水轮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，为了避免设备故障导致严重的后果，对水轮机组进行实时状态监测及设备健康管理就很有必要。水轮机组系统设备复杂，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。

目前基于大数据的水轮机组设备健康管理方法存在数据急剧膨胀和分析挖掘方法相对滞后的矛盾。①数据急剧膨胀：随着传感器技术、数据采集、处理及传输技术等计算机网络技术的飞速发展，水电站运行过程中产生数量越来越多、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。②分析挖掘方法相对滞后:目前行业现有的分析平台自动分析方法和自动分析系统相对滞后，无法精准的从大量生产数据中自动提取有价值的分析数据和预测信息，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，无法及时发现机组轴承油槽油混水、调速器管路漏油等设备安全隐患。

由于水电站生产环境复杂，数据清洗任务复杂，一个业务模型落地需要花费大量的时间，然而，对于水轮机组设备的健康诊断模型而言，建立上千个业务模型经常发生，实施周期较长、难度较大。

专家经验是有限的，不可能穷尽所有经验，同时专家经验随着工况，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。

水轮机设备健康管理是非常交叉的学科，涉及电气、机械、化学、数据等，对人员素质要求极高。

对于水轮机设备健康分析而言，需要实现利用数据说话，通过大数据分析技术结合AI算法实现隐患的挖掘，尽量少用专家经验参与，节约模型落地时间，真正实现水电站智能化、智慧化应用。随着大数据计算能力、存储能力、数据压缩能力的提高，以及AI算法非线性拟合能力的增强，大规模的数据计算、高价值的数据分析挖掘成为可能。虽然大数据存储压缩能力强，但是在实际应用中，数据采样仍然必不可少，原因如下：一方面，站在业务角度，可视化展示状态量的数据趋势时，假如不采样，那么仅仅只能查询较短时间内的趋势，比如1小时，假如要分析一天、一周、一月的数据时，卡顿非常明显，带来不好的用户体验；另一方面，站在大数据分析的角度，构建正样本空间时，不同工况下的正样本非常多、数据相似性较大，假如不采样，会带来很大的冗余计算。综合以上原因，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，同时结合水轮机组各种工况特点，设计采样方法。

目前数据采样常用方法有广义平滑加权采样(GSW)法、蓄水池采样算法、不规则采样法、LTTB采样等。但这些方法存在采样偏差大，采样不够全面等缺点，对后期数据分析的准确度造成影响。因此需探索研发更加先进、更高精度的采样方法，以实现数据样本的精准采样。

电站运行过程中产生数量越来越多、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。在如此大量生产数据中存在离群数据、丢失数据、延迟数据、无效数据等情况，需对这些数据进行清洗，还原设备真实的运行过程。目前对数据的清洗方法较为匮乏，主要依靠专家经验进行判断，但专家经验是有限的，不可能穷尽所有经验，而且仅靠专家经验去判断，不仅费时费力，效果还不佳。因此需探索研发智能的数据清洗方法，以实现对离群数据、丢失数据、延迟数据、无效数据等情况的补全、删除、修正。

目前大数据采样存在以下问题：因数据源的不稳定会导致数据采样失败或者采样样本不具代表性；数据量大，导致采样工作费事费力，速度缓慢：各种原因造成的错误和不完整的低质量数据，导致数据采样质量低下：不同来源的数据可能存在各种格式差异，导致难以进行统一处理，造成采样困难：数据来源存在不确定性，需要从多个渠道进行数据采样，这些渠道的数据质量和有效性也存在巨大差异，造成采样工作困难：大规模数据采样需要投入大量的人力、物力和财力，造成采样成本高昂。

目前数据清洗存在问题，数据来源和应用场景多，需要花费大量的时间和精力去了解数据的来源、目的和意义；数据清洗的标准和规则制定也较困难：如何确定哪些是离群数据、丢失数据、延迟数据、无效数据，以及如何删除无效数据，补全丢失数据，还原数据的真实性，也是一个较复杂且难以处理的问题：由于错误或异常情况导致的异常值，如何确定一个合适的标准来区分正常和异常的数据，如何识别和处理异常值也是一个挑战；对于测点数据中的非平稳性数据，采用什么样的方法，把非平稳数据转化为平稳数据，进行智能算法分析也是一大难题；对于测点数据的噪声情况，如何降噪，还原时序数据的真实信号，方便精准预测也是一大挑战。

针对电站运行过程中产生越来越多的生产数据，历史数据库中存储的数据量急剧膨胀。在此情况下，如何对数据进行精准采样，如何对存在的离群数据、丢失数据、延迟数据、无效数据进行清洗的问题。提出了先进解决方法。对于数据采样，提出了均匀采样、分层随机采样、OCSVM采样、非均匀采样方法。对于数据清洗，提出了离群数据的清洗采用Kmeans离群算法、KNN离群算法、LOF局部离群因子算法、孤立森林算法、四分位数算法；延迟数据的清洗采用依据数据的强相关性，利用前序数据规律经过简单的换算（比如：加权平均法）的方法；数据丢失的清洗采用SVR、CONVLSTM算法；无效数据的清洗采用SVR、CONVLSTM算法。

通过均匀采样、分层随机采样、OCSVM采样、非均匀采样方法，实现在庞大数据中样本的精确采样，构建正样本空间。通过运用Kmeans离群算法、KNN离群算法、LOF局部离群因子算法、孤立森林算法、四分位数算法等算法，构建数据清洗模型，实现对离群数据、丢失数据、延迟数据、无效数据的清洗，还原设备运行数据的真实性。

在本申请的实施例中，相对于现有技术中的“数据无法进行精准采样，存在采样偏差大、采样不够全面等缺点，对后期数据分析的准确度造成影响；同时，容易出现无效数据无法进行补全、删除、修正等问题”，本申请提供了“通过均匀采样、分层随机采样、OCSVM采样、非均匀采样方法，实现在庞大数据中样本的精确采样，构建正样本空间。通过运用Kmeans离群算法、KNN离群算法、LOF局部离群因子算法、孤立森林算法、四分位数算法等算法，构建数据清洗模型，实现对离群数据、丢失数据、延迟数据、无效数据的清洗，还原设备运行数据的真实性”的解决方案，具体为：获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；依据所述预处理数据进行采样处理得到采样数据；依据所述采样数据进行清洗处理得到处理数据。本申请通过运用新研发的数据采样方法，对各种正常运行工况下的正样本数据进行采样，构建了正样本空间，便于查看数据趋势和对运行时状态进行偏离度计算。采样样本覆盖了所有工况下的状态量模式，采样结果能符合原始数据趋势，时间序列满足概率统计上的无偏估计，同时能提高预警预测算法相关算法精度。实现了延迟数据，丢失数据，无效数据的填充；能对测点在时间轴上重复的数据进行自动去重任务；实现对突变，离群，死点，噪声等数据进行清洗，根据不同场景采用不同算法进行对比清洗效果；实现对含高频噪声的特征数据进行去噪或光滑处理，避免造成误报，提升模型的准确性和鲁棒性；实现基于采用频率的自定义设置，实现降采样任务，完成数据转换功能；能够对多维度的冗余信号进行降维处理，消除冗余信息给预测结果带来的不确定性。

参照图1，示出了本申请一实施例提供的一种基于巨型水轮机组的数据处理方法的步骤流程图，具体包括如下步骤：

S110、获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

S120、依据所述预处理数据进行采样处理得到采样数据；

S130、依据所述采样数据进行清洗处理得到处理数据。

下面，对本示例性实施例中基于巨型水轮机组的数据处理方法作进一步地说明。

如上述步骤S110所述，获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据。

在本发明一实施例中，可以结合下列描述进一步说明步骤S110所述“获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据”的具体过程。

如下列步骤所述，依据所述状态量数据进行时序性排序得到排序数据；依据所述排序数据进行同源压缩处理得到所述预处理数据。

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“依据所述状态量数据进行时序性排序得到排序数据”的具体过程。

作为一种示例，依据所述状态量数据中的时间戳对每个水轮机组设备数据按照时间顺序进行排序；若无时间戳的状态量，则将水轮机组设备数据进行删除。所述状态量是包括多个水轮机组设备数据的集合。

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“依据所述排序数据进行同源压缩处理得到所述预处理数据”的具体过程。

在一具体实施例中，根据所述排序数据中各数据的状态、时间和幅值，确定所排序数据中的重复数据；剔除所述排序数据中的所述重复数据，得到第一数据；根据所述第一数据中各数据的所述状态和所述幅值，确定所述第一数据列表中的相邻同幅同状态数据，压缩所述相邻同幅同状态数据，得到所述预处理数据。其中，此处的压缩可以为随机性剔除。

如上述步骤S120所述，依据所述预处理数据根据获取平台进行采集管理得到采集数据；依据所述采集数据进行采样分析得到所述采样数据。

在本发明一实施例中，可以结合下列描述进一步说明步骤S120所述“依据所述预处理数据根据获取平台进行采集管理得到采集数据；依据所述采集数据进行采样分析得到所述采样数据”的具体过程。

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“依据所述预处理数据根据获取平台进行采集管理得到采集数据”的具体过程。

在一具体实施例中，测点数据采集管理，数据来源主要来自水电站工业互联网平台，通过接口方式实现数据接入。完成数据源的管理，数据测点的管理，数据集的管理。整个应用具体的功能满足如下：通过MQTT、Kafka、Restful等接口完成与水电站工业互联网平台接口交互。同时具备数据缓存机制，保证数据的完整性交互。数据源管理如下图4所示，测点管理如下图5所示，数据集管理如下图6所示。

如下列步骤所述，依据所述采集数据判断所述采集数据的数据性质；其中，所述数据性质包括时间性质、异常性质和不均匀性质；依据所述数据性质匹配采样方法得到所述采样数据。

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“依据所述采集数据判断所述采集数据的数据性质”的具体过程。

如下列步骤所述，依据所述采集数据得到数据的时间权重、异常权重以及不均匀权重；依据所述时间权重、异常权重以及不均匀权重得到所述数据性质。

在一具体实施例中，在时间权重、异常权重以及不均匀权重中，比较三者的数值占比大小，占比高的为数据性质，比如，若时间权重的占比大于所述异常权重和所述不均匀权重则得到所述数据性质为时间性质；同理可得异常性质和不均匀性质。

在一具体实施例中，将采集数据中每个数据的单个性质，得到整个数据性质，若所述采集数据中的数据有异常，则为异常数据；异常性质优先于不均匀性质，不均匀性质优先大于时间性质，再依据每个数据的单个性质，得到时间权重、异常权重以及不均匀权重。

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“依据所述数据性质匹配采样方法得到所述采样数据”的具体过程。

如下列步骤所述，若所述数据性质为时间性质，则通过广义平滑加权采样得到所述采样数据；若所述数据性质为异常性质，则通过空间采样得到所述采样数据；若所述数据性质为不均匀性质，则通过不规则采样得到所述采样数据。

在一具体实施例中，对于水轮机设备健康分析而言，需要实现利用数据说话，通过大数据分析技术结合AI算法实现隐患的挖掘，尽量少用专家经验参与，节约模型落地时间，真正实现水电站智能化，智慧化应用。随着AI算法非线性拟合能力的增强，对于挖掘数据隐含规律提供了模型可能，大数据计算能力，存储能力，数据压缩能力的提高，为大规模的数据计算提供可能，根据技术经验，目前能支持到PB级数据。尤其是数据压缩能力，可以加大数据存储和备份周期。虽然大数据存储压缩能力强，但是在实际应用中，数据采样仍然必不可少，原因如下：

一方面，站在业务角度，可视化展示状态量的数据趋势时，假如不采样，那么仅仅只能查询较短时间内的趋势，比如1小时，假如要分析一天、一周、一月的数据时，卡顿非常明显，带来不好的用户体验；

另一方面，站在大数据分析的角度，构建正样本空间时，由于不同工况下的正样本非常多、数据相似性高，假如不采样，会带来很大的冗余计算。

综合以上原因，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，同时结合水轮机组各种工况特点，设计采样方法，下面举例说明采样过程存在的难点和问题：从图7可知，由于数据点数太密，且波动非常频繁，将所有数据点通过折线依次连接后出现重叠，可视化效果差，此时就需要对原始的数据点进行采样，比如将10000个原始数据点采样为200个数据点。对于采用具体的采样方式至关重要。

在一具体实施例中，若所述数据性质为时间性质，则通过广义平滑加权采样得到所述采样数据；具体的，广义平滑加权采样(GSW)，水电站每天都会产生大量的高维数据，这样的数据自然与时间戳相关。实时预测分析可以分两步进行。首先通过对数据进行切片、分割和聚合来指定要关注的数据部分和要预测的度量；第二步，基于聚合结果训练预测模型，用于理解和预测指定量度的趋势。其中第一步是提供实时响应的性能瓶颈。GSW采样获得原始数据的近似集合作为训练预测模型的输入，并分析了利用GSW抽样估计聚集的误差界。

正常样本空间的时间序列数据量巨大，通常需要使用几个月的历史数据来预测未来的数据点。其次，时间序列数据是非常高维的，可能有数百个属性组合。现有的改进方法主要通过对历史数据进行采样来减少输入数据的大小，其中采样算法分为均匀采样和加权采样。均匀采样为聚合提供了无偏估计，但估计误差可能与度量范围成比例；加权抽样方案提供了最优估计，具有更好的误差界，但必须为每个度量单独抽取一个加权样本，如果存在多个不同的度量，总的空间消耗可能是巨大的。为了生成可以兼顾多种度量并且仍然有准确度保证的紧凑样本，广义平滑加权GSW采样方法：每个满足约束的样本的采样概率为,其中△为正常数，w为采样权重。进一步优化度量为：同时满足为的无偏估计。广义平滑加权采样(GSW)采样效果如图8所示。

在一具体实施例中，若所述数据性质为异常性质，则通过空间采样得到所述采样数据；空间采样(OCSVM)，OCSVM算法是一种经典的异常检测算法，基本原理与 SVM 类似，与 SVM 关注的二分类问题不同的是，OCSVM 只有一个分类，这也正是异常检测所需要的，不关注异常的数据点。OCSVM 的基本思想是: 将数据映射到高维空间中，使得正常样本在该空间中有较高的密度，而异常样本则较为孤立。具体而言，OCSVM通过找到一个超平面来划分正常样本和异常样本，而这个超平面应尽可能地与正常样本接近。OCSVM的目标是最大化支持向量到超平面的边距，即最大化正常样本的密度。通过这种方式，OCSVM可以有效地识别出异常样本，因为异常样本往往远离正常样本区域。

OCSVM算法可以分为以下几个步骤：数据预处理：对输入数据集进行标准化处理，使得数据在各个维度上具有相同的尺度，避免某一个特征对结果产生较大影响。建立模型：根据预处理后的数据集，使用OCSVM算法建立支持向量机模型。模型的训练过程就是寻找最优的超平面，使得正常样本尽可能地靠近该超平面。寻找异常样本：在训练完模型后，通过计算每个样本到超平面的距离，可以确定哪些样本是异常样本。距离超平面较远的样本很可能是异常样本，而距离较近的样本则是正常样本。阈值确定：根据距离的分布情况，可以选择一个适当的阈值来判断样本是否为异常。通常情况下，距离超过阈值的样本被认为是异常。正常样本空间，OCSVM使用均匀采样；异常样本空间，OCSVM使用加权采样。

在一具体实施例中，若所述数据性质为不均匀性质，则通过不规则采样得到所述采样数据。具体的， 不规则抽样也叫不均匀抽样：一种将抽样权重与个体本身结合的抽样方法。样本不均衡时：某些情况下占比较小的样本种类可能更具备研究价值。例如：在设备健康管理方面，传感器脱落、松动等原因造成的数据样本可能占大部分，但是实际上相比于其他水轮机机组缺陷价值并不大。对于不规则采样，除了采用业务规则以外，同时可以采用聚类的方式来实现不规则采样，先将数据集中元素分组聚类构建成有总体代表性的样本组，再随机选择一个或者多个样本组作为样本集。聚类采样与分层采样不同之处在于，聚类所区分出的类别，每组都可以一定程度上代表整体，无需分别对每组再进行采样和汇总，直接完整选取几组即可。通过聚类实现权重的分配，做到不规则采样。如图9所示。

作为一种示例，采样过后的数据还需要进一步的检验，为样本时序数据评估，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。首先要确定，估计量：指任何一个对总体参数给出估计值的样本统计量，例如样本均值。估计值：指从某一样本计算得到的估计量的一个具体数值。通过对样本的一系列计算去预估总体，而这一系列的计算就是统计量的计算。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10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（例如：样本的均值、方差或标准差）算得一个具体的数值（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（例如：总体的均值、方差或标准差）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，使得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。

样本点估计值；

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）；

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）。

置信水平，显著性水平估计

通过对水电机组设备时序数据进行正样本采样，对于时序采样数据通过置信水平估计和置信区间估计，采样正样本空间满足样本总体的分布规律，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11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步走可以实现，如图12所示。

作为一种实施例中，样本空间筛选采样实例，对于水电机组数据样本采样，实例部分主要采用以下方式进行采样，分别是：OCSVM采样和非均匀采样。整体如下：上导瓦温原始数据如图13所示。

在一具体实施例中，OCSVM采样，以发电工况下，选定正常样本上导瓦温数据如图14所示，通过OCSVM采样后的样本如图15所示。

在一具体实施例中，不均匀采样，以发电工况下，按时间进行不均匀切片，开始凌晨0点-早上8点，2.5小时间隔为切片，后续分别以3小时，4小时，5小时进行切分。以切片采集数据取均值处理，如图16所示。

最终通过各种采样方法实现了正样本空间的采样，同时通过概率置信度估计。

如上述步骤S130所述，依据所述采样数据进行清洗处理得到处理数据。

在本发明一实施例中，可以结合下列描述进一步说明步骤S130所述“依据所述采样数据进行清洗处理得到处理数据”的具体过程。

如下列步骤所述，依据所述采样数据判断所述采样数据的数据类型；其中，所述数据类型包括离群数据、延迟数据、残缺数据以及无效数据；依据所述数据类型进行清洗得到所述处理数据。

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“依据所述数据类型进行清洗得到所述处理数据”的具体过程。

如下列步骤所述，若所述数据类型为离群数据，则通过离群算法得到所述处理数据；若所述数据类型为延迟数据，则通过加权平均算法得到所述处理数据；若所述数据类型为残缺数据，则通过前序数据补正得到所述处理数据；若所述数据类型为无效数据，则通过前序数据修正得到所述处理数据。

在一具体实施例中，数据清洗算法选取，对于水电站而言，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。对于水电站主要存在离群数据、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进行划分，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。在此场景下，该方法不能完全的去除离群点，容易删除正常数据。

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。

在一具体实施例中，SVR算法整体过程解释如下：SVR算法是机器学习算法，非线性特征拟合能力较强，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有依赖，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，整个数据清洗过程在历史数据中确定清洗规则和算法参数，当算法稳定后用于实时数据清洗，以此更新整个数据的清洗，此处理过程自动化进行，人工参与少，整个过程可视化展示具有很强的业务解释性和清洗的正确性判断，用户可自定义操作过程控制。平台采用Spark内存计算模式，清洗效率极高，满足数据清洗能力不低于 70 万条/次，算法时长不高于30 分钟要求。

作为一种示例，数据清洗过后得到清洗数据，依据所述清洗数据进行非平稳性处理得到平稳数据。

在一具体实施例中，数据非平稳性处理，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居多，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，移动平均是一种最简单的数据[平滑方法](https://so.csdn.net/so/search?q=%E5%B9%B3%E6%BB%91%E6%96%B9%E6%B3%95&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/fengdu78/article/details/_blank)，用于平滑时间序列数据。它通过计算一定窗口内数据点的平均值来减少噪音，同时保留数据的趋势。移动平均可以是简单移动平均或指数加权移动平均。

简单移动平均（SMA）： [简单移动平均](https://so.csdn.net/so/search?q=%E7%AE%80%E5%8D%95%E7%A7%BB%E5%8A%A8%E5%B9%B3%E5%9D%87&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/fengdu78/article/details/_blank)是一种通过计算数据点在一个固定窗口内的平均值来平滑数据的方法。窗口的大小决定了平滑程度。较大的窗口将导致更平滑的曲线，但会减缓对趋势的反应，而较小的窗口将更敏感地跟随数据的波动。

指数加权移动平均：指数加权移动平均是一种通过对数据点应用指数权重来平滑数据的方法。它对最近的数据点给予较高的权重，而对较早的数据点给予较低的权重。使得EMA更适合用于追踪快速变化的数据。

指数平滑，一种常用的时间序列数据平滑和预测方法，用于处理具有趋势和季节性的数据。它通过分配不同权重给历史数据点，将较高权重分配给较新的数据，以捕获数据的变化趋势。指数平滑通常用于生成预测，特别是在需要对未来时间点进行预测的情况下。

局部加权散点平滑，一种非参数的数据平滑方法，通常用于拟合或平滑散点数据，以捕获数据的局部趋势和模式。Loess平滑是一种局部回归方法，它对于不同区域的数据采用不同的权重，以确保在数据的每个局部区域上都能获得较好的拟合。

作为一种示例，依据所述平稳数据进行噪声处理。

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，是一种信号处理工具，用于去除信号中高频成分，从而保留低频成分。在数据平滑的上下文中，低通滤波器用于减小或去除数据中的高频噪声或快速变化，以使数据趋于平滑，保留慢变化的趋势或周期性特征。

Kalman滤波，卡尔曼滤波是一种用于状态估计和数据融合的强大数学和统计工具。卡尔曼滤波是一种递归的、最优的状态估计方法，通常用于处理含有噪声的时间序列数据，以估计数据的状态，同时提供估计的不确定性。

Savitzky-Golay滤波器，Savitzky-Golay滤波是一种信号处理技术，用于对离散数据序列进行平滑和去噪。它是一种线性平滑滤波器，通过拟合多项式来估计数据点的平均值，以减小噪声和突发波动。

在一具体实施例中，数据清洗实例，模拟真实数据如图17所示。对于离群数据清洗，以K-means离群算法数据清洗为例，如图18所示。延迟数据清洗如图19所示；残缺数据清洗，如图20所示；无效数据清洗，如图21所示。

通过以上的清洗过程，均在Spark平台上运行，运行速度快。每分钟处理30-50万数据。

在一具体实施例中，数据采样的有益效果及发明要点：运用新研发的数据采样方法，对各种正常运行工况下的正样本数据进行采样，构建了正样本空间，便于查看数据趋势和对运行时状态进行偏离度计算。采样样本覆盖了所有工况下的状态量模式，采样结果能符合原始数据趋势，时间序列满足概率统计上的无偏估计，同时能提高预警预测算法相关算法精度。

在一具体实施例中，数据清洗的有益效果及发明要点：实现了延迟数据，丢失数据，无效数据的填充；能对测点在时间轴上重复的数据进行自动去重任务；能实现对突变，离群，死点，噪声等数据进行清洗，根据不同场景采用不同算法进行对比清洗效果；能实现对含高频噪声的特征数据进行去噪或光滑处理，避免造成误报，提升模型的准确性和鲁棒性实现基于采用频率的自定义设置，实现降采样任务，完成数据转换功能；能够对多维度的冗余信号进行降维处理，消除冗余信息给预测结果带来的不确定性。

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巨型水轮机组的数据处理系统，具体包括如下模块：

预处理数据模块210，用于获取巨型水轮机组设备的状态量数据，并依据所述状态量数据进行预处理得到预处理数据；

采样模块220，用于依据所述预处理数据进行采样处理得到采样数据；

处理数据模块230，用于依据所述采样数据进行清洗处理得到处理数据。

在本发明一实施例中，所述预处理数据模块210包括：

排序数据子模块：用于依据所述状态量数据进行时序性排序得到排序数据；

压缩处理子模块：用于依据所述排序数据进行同源压缩处理得到所述预处理数据。

在本发明一实施例中，所述采样模块220包括：

采集数据子模块：用于依据所述预处理数据根据获取平台进行采集管理得到采集数据；

采样子模块：用于依据所述采集数据进行采样分析得到所述采样数据。

在本发明一实施例中，所述采样子模块包括：

数据性质子模块：用于依据所述采集数据判断所述采集数据的数据性质；其中，所述数据性质包括时间性质、异常性质和不均匀性质；

匹配采样方法子模块：用于依据所述数据性质匹配采样方法得到所述采样数据。

在本发明一实施例中，所述匹配采样方法子模块包括：

时间性质子模块：用于若所述数据性质为时间性质，则通过广义平滑加权采样得到所述采样数据；

异常性质子模块：用于若所述数据性质为异常性质，则通过空间采样得到所述采样数据；

不均匀性质子模块：用于若所述数据性质为不均匀性质，则通过不规则采样得到所述采样数据。

在本发明一实施例中，所述处理数据模块230包括：

数据类型子模块：用于依据所述采样数据判断所述采样数据的数据类型；其中，所述数据类型包括离群数据、延迟数据、残缺数据以及无效数据；

清洗数据子模块：用于依据所述数据类型进行清洗得到所述处理数据。

在本发明一实施例中，所述清洗数据子模块包括：

离群数据子模块：用于若所述数据类型为离群数据，则通过离群算法得到所述处理数据；

延迟数据子模块：用于若所述数据类型为延迟数据，则通过加权平均算法得到所述处理数据；

残缺数据子模块：用于若所述数据类型为残缺数据，则通过前序数据补正得到所述处理数据；

无效数据子模块：用于若所述数据类型为无效数据，则通过前序数据修正得到所述处理数据。

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巨型水轮机组的数据处理方法的计算机设备，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巨型水轮机组的数据处理方法。

也即，上述处理单元16执行上述程序时实现：获取影像集，依据所述影像集得到数据库；其中，所述影像集包括检查数据、影像数据以及报告数据；依据所述数据库进行数据清洗生成规范数据库；当收到调度数据任务时，依据所述调度数据任务在所述规范数据库中进行调度生成调度影像。

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巨型水轮机组的数据处理方法。

也即，给程序被处理器执行时实现：获取影像集，依据所述影像集得到数据库；其中，所述影像集包括检查数据、影像数据以及报告数据；依据所述数据库进行数据清洗生成规范数据库；当收到调度数据任务时，依据所述调度数据任务在所述规范数据库中进行调度生成调度影像。

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++，还包括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，或者还是包括为这种过程、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巨型水轮机组的数据处理方法及系统，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。

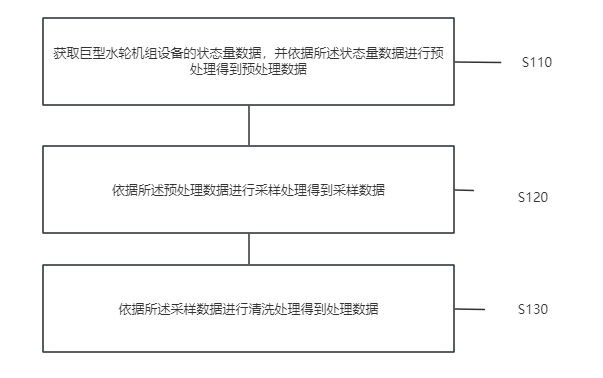


图1

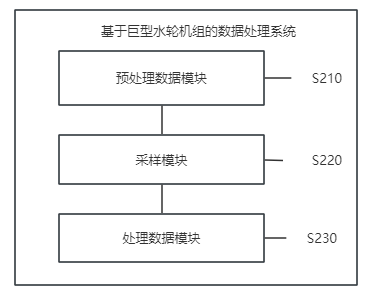


图2

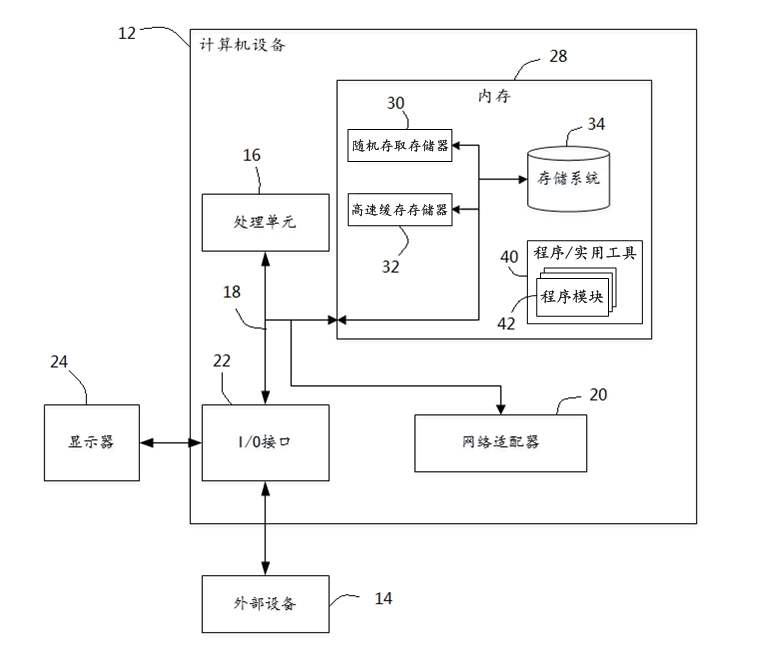


图3

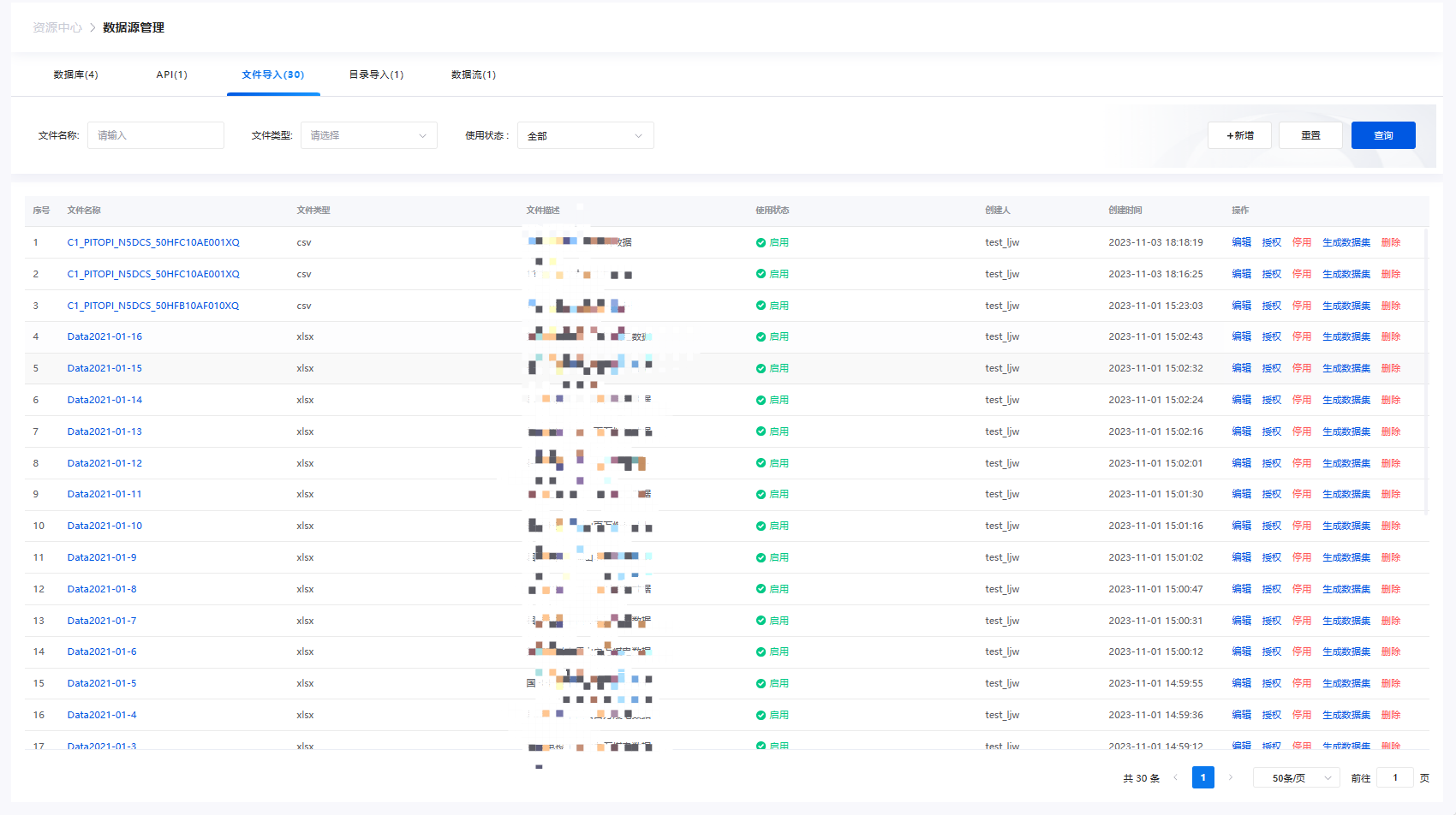


图4

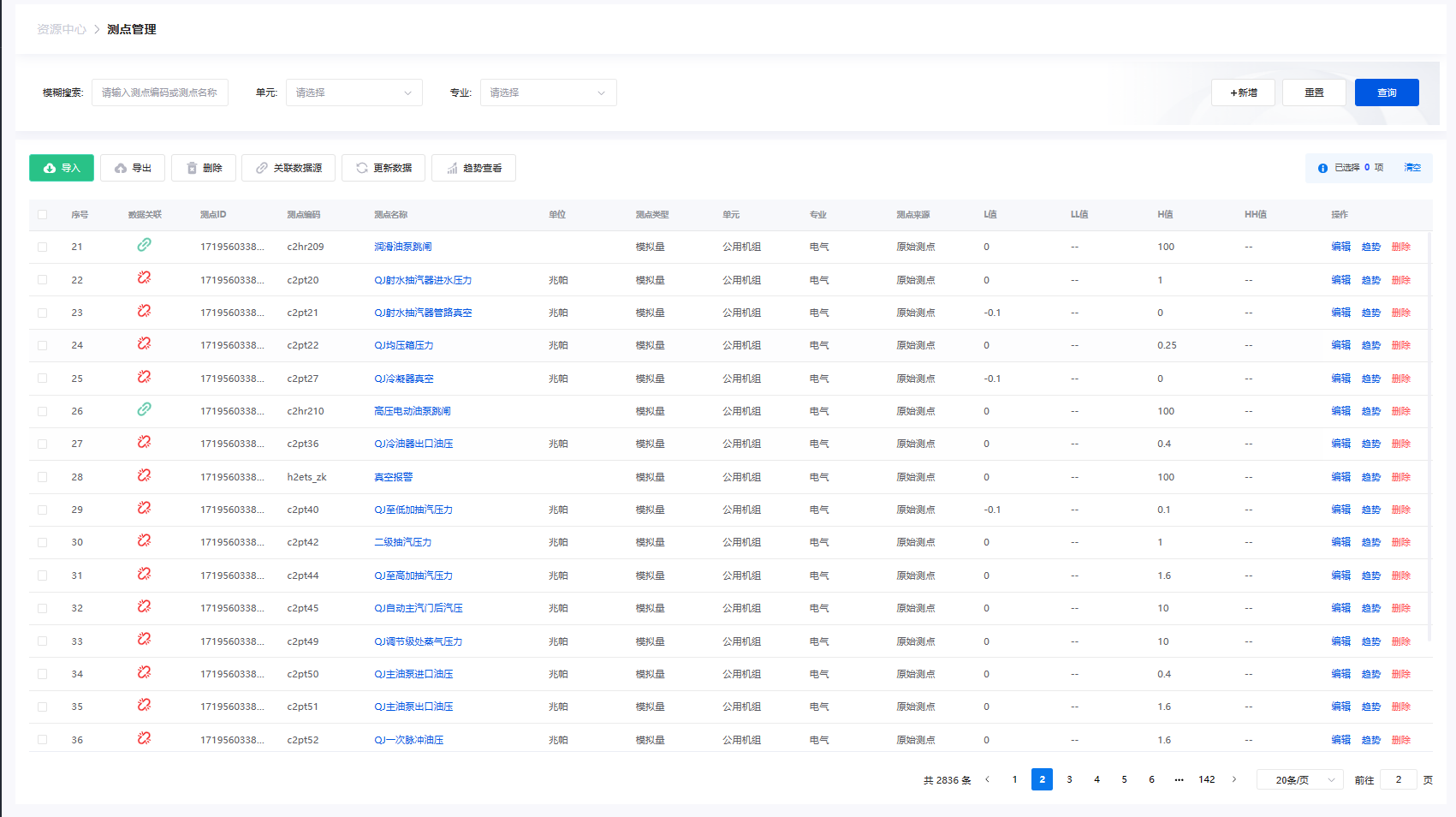


图5

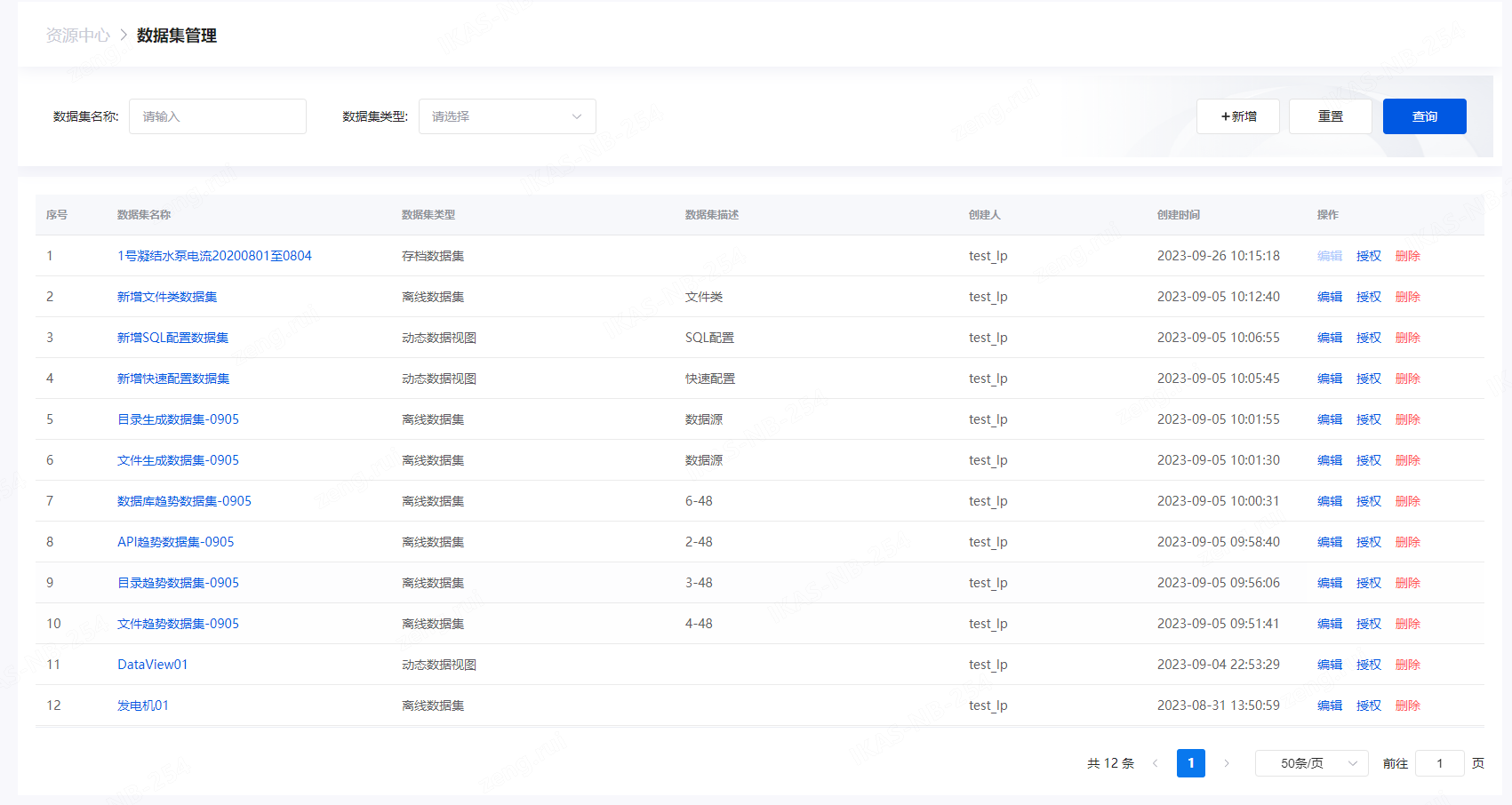


图6

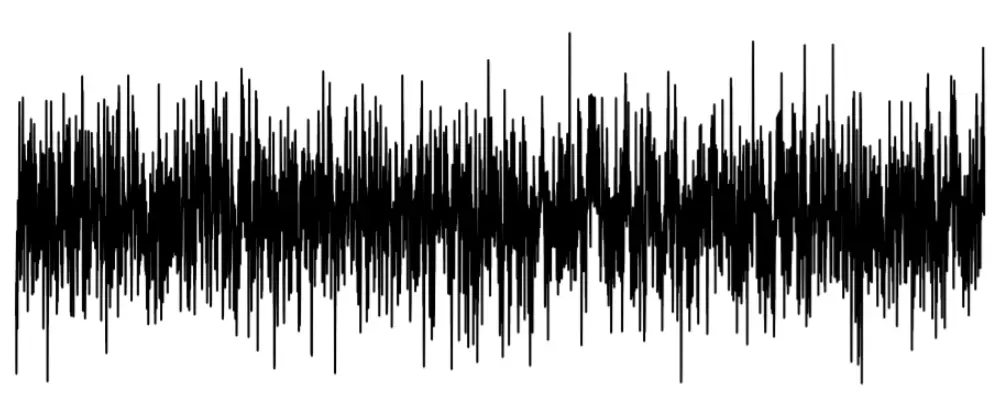


图7

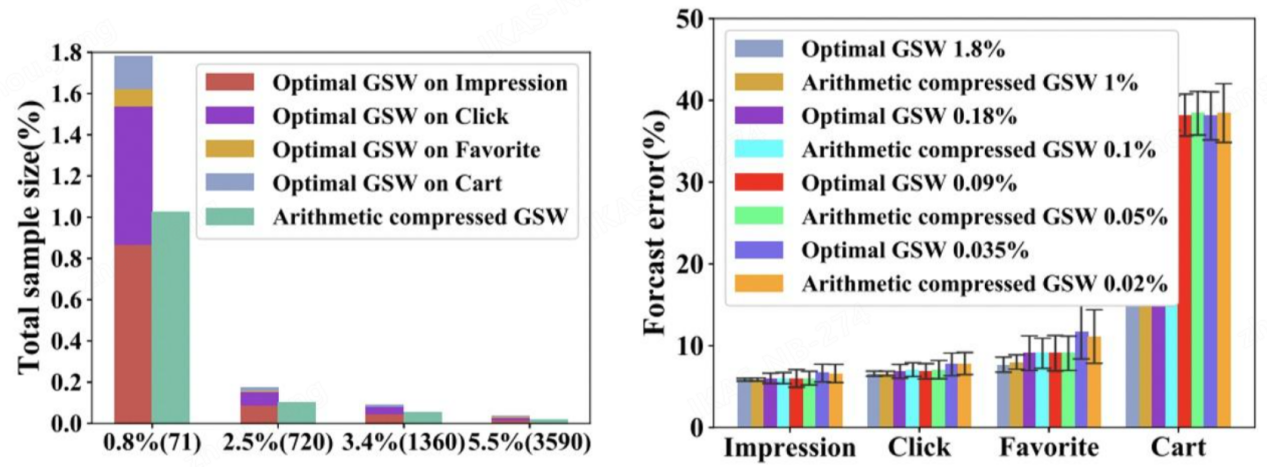


图8

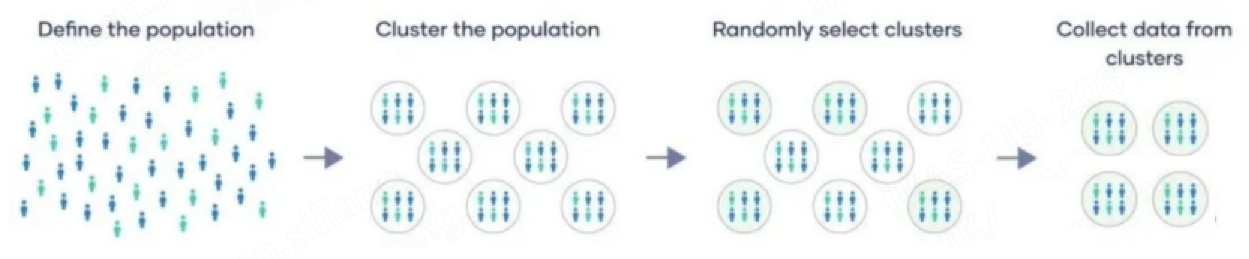


图9

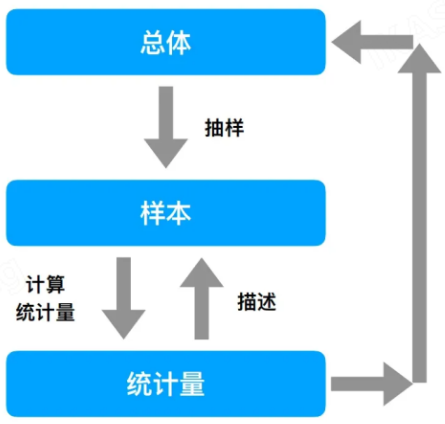


图10

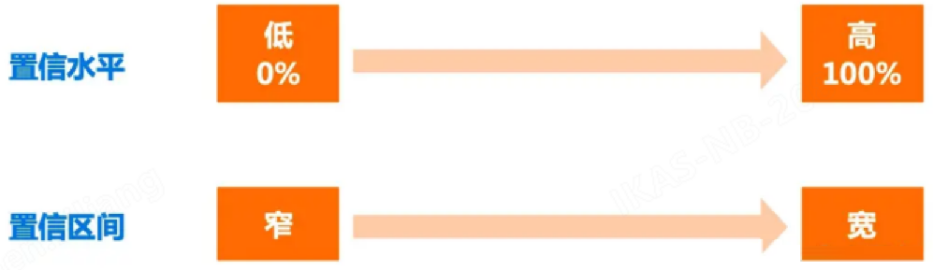


图11

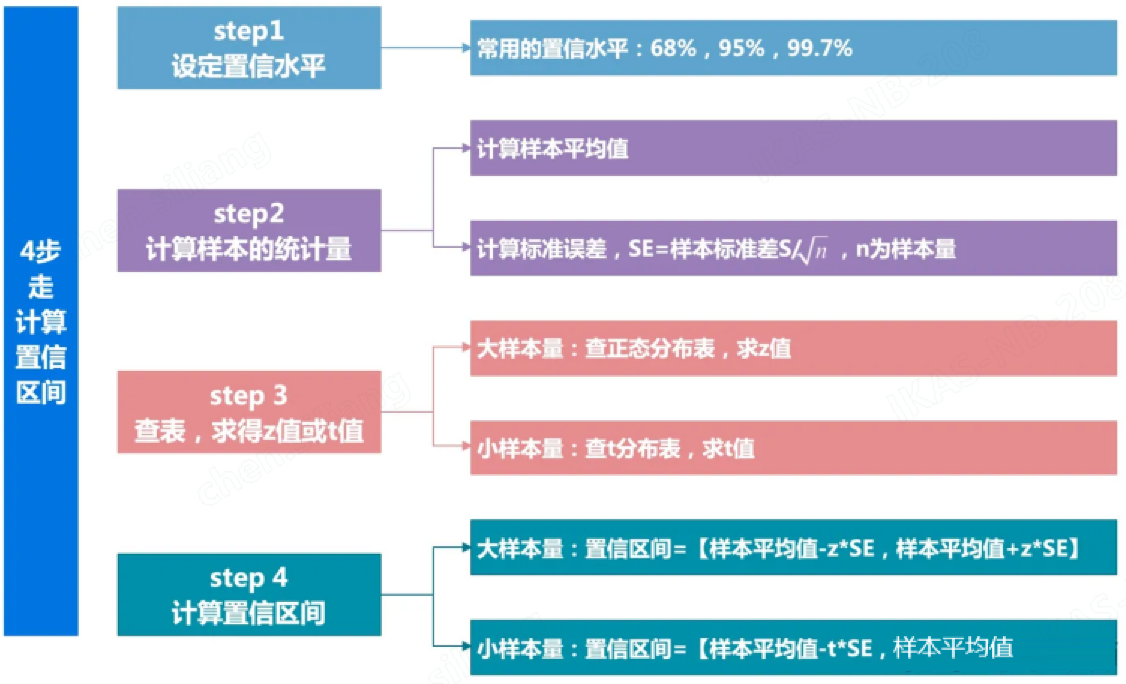


图12

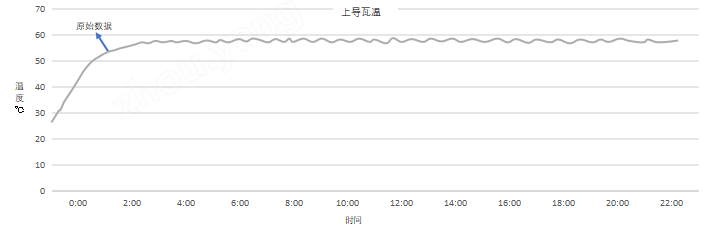


图13

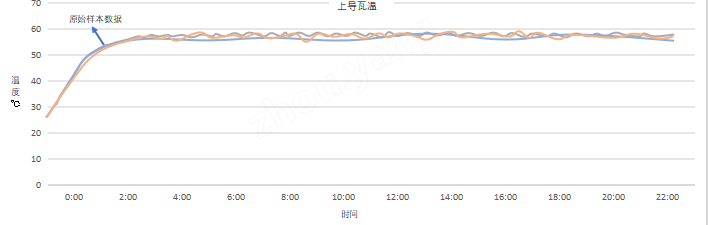


图14

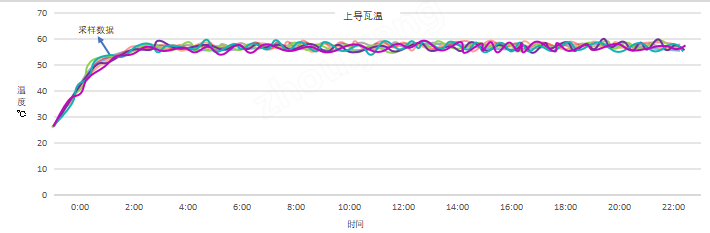


图15

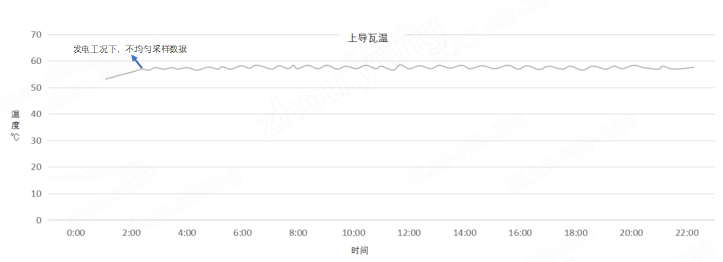


图16

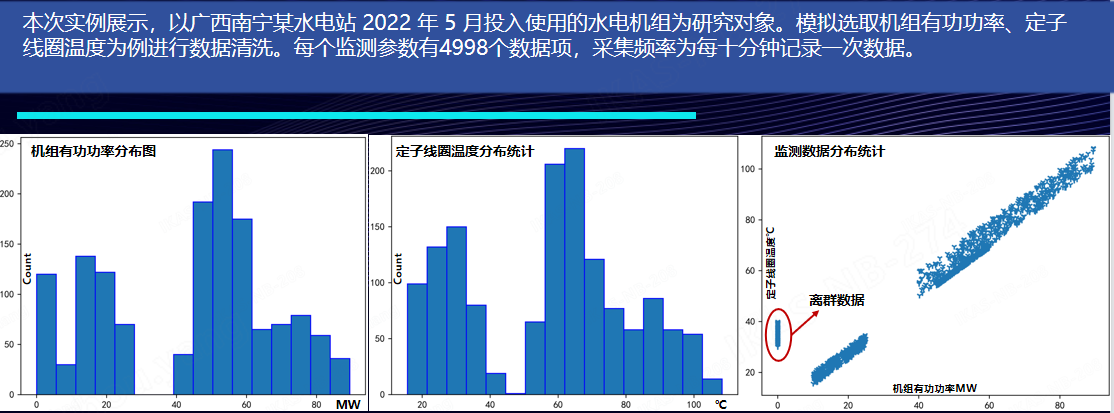


图17

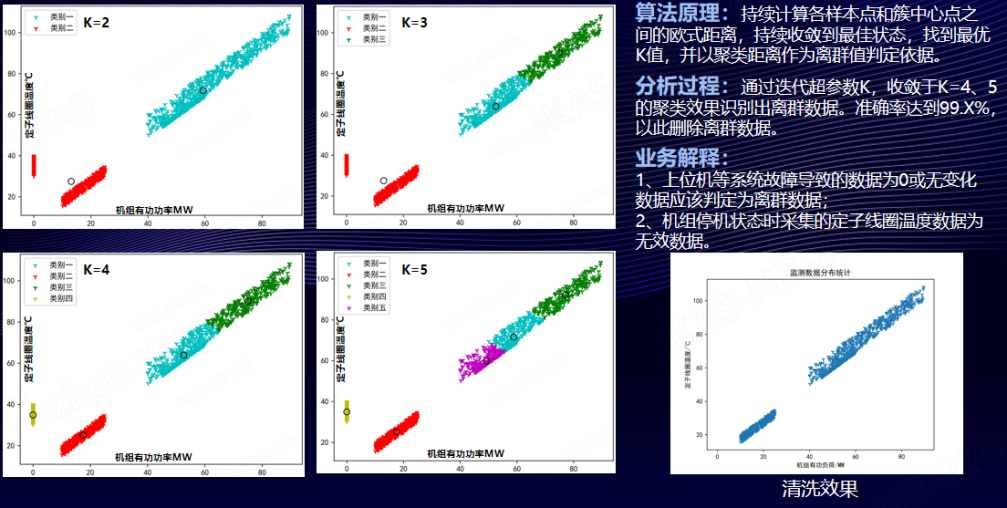


图18

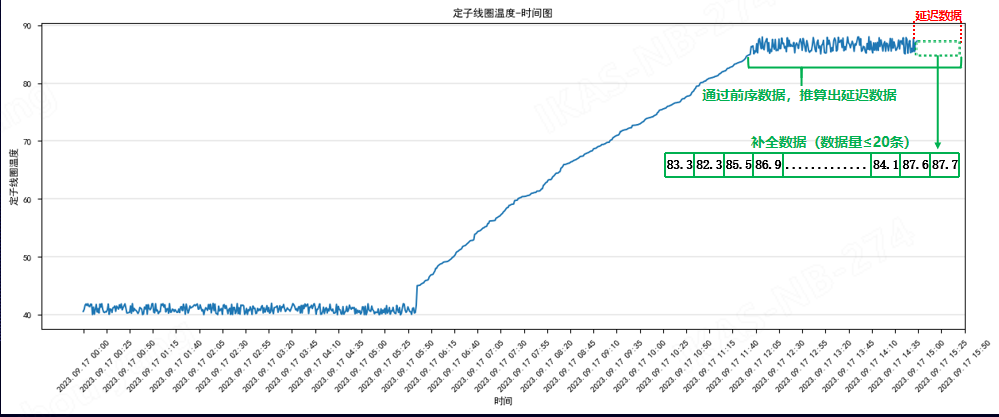


图19

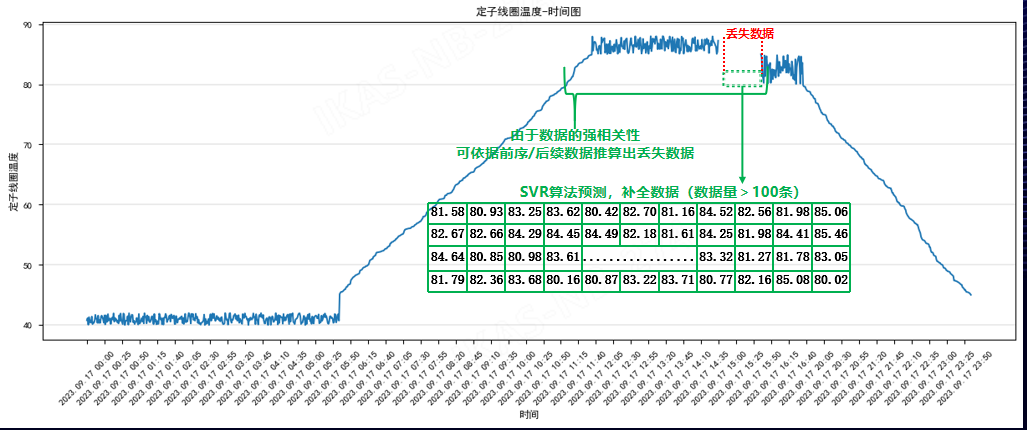


图20

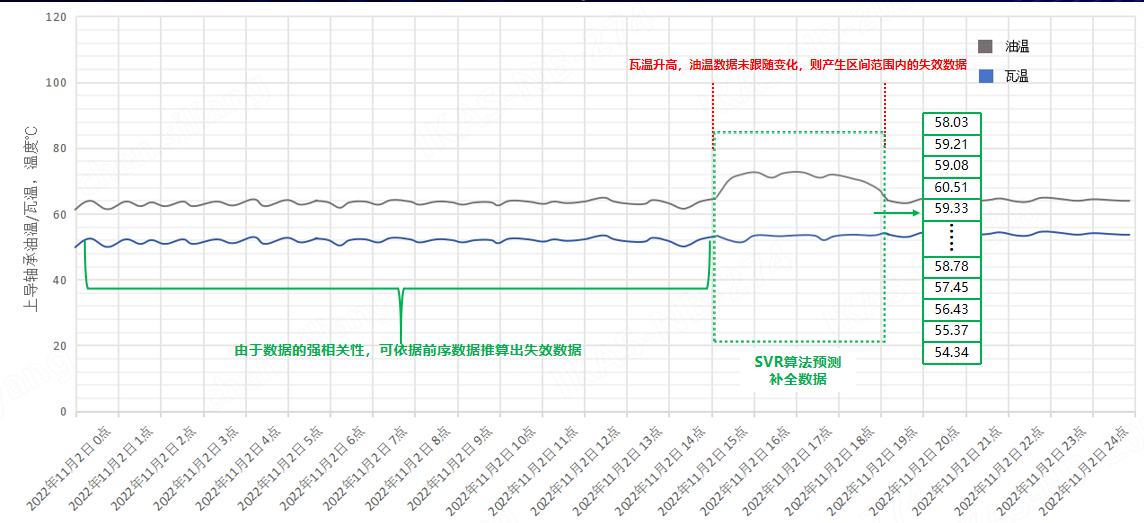


图21